

Gleichmäßige Konfidenzbänder für Erwartungswertfunktional und Kovarianzkern funktionaler Daten

Von der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

von

Christoph Reihl

aus Marktredwitz

1. Gutachter: Prof. Dr. Melanie Birke
2. Gutachter: Prof. Dr. Hajo Holzmann

Tag der Einreichung: 06.12.2023

Tag des Kolloquiums: 18.03.2024

Kurzfassung

Die vorliegende Arbeit befasst sich mit zwei Schätzern bzgl. diskreter Beobachtungen innerhalb des funktionalen Regressionsmodells

$$Y_{ij} = \mu(x_{ij}) + Z_i(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p$$

wobei $\hat{\mu}$ ein Schätzer für den Trend μ , und \hat{R} ein Schätzer für die Autokovarianz der Zufallsfunktion Z darstellen soll. Für die Schätzer wird asymptotische Erwartungstreue, asymptotische Normalität, sowie Konsistenz, jeweils im funktionalen Sinne nachgewiesen. Darauf aufbauend, werden gleichmäßige, oder auch simultane Konfidenzbänder konstruiert und zusätzlich eine (naive) Bootstrap-Methode vorgestellt - inklusive zugehörigem Beweis der fast sicheren Konsistenz, bedingt auf die Stichprobe. Weiterhin wird durch eine Simulation die Gültigkeit der Konfidenzbänder für die Autokovarianz mittels mehrerer Datensettings überprüft, sowohl für den normalen Fall, als auch für den Bootstrap.

Bzgl. der Messpunkte x_{ij} werden zwei Fälle unterschieden: Synchronen Messpunkte innerhalb jeder Stichprobeneinheit, d. h. $x_{1j} = x_{2j} = \dots = x_{nj}$, und asynchrone Messpunkte, d. h. $x_{i_1j} \neq x_{i_2j}$ für mindestens ein $j = 1, \dots, p$ und $i_1 \neq i_2$. Auch wenn die Beweisstrukturen von letzterem Fall auf ersterem aufbauen, gibt es doch wesentliche Unterschiede, weshalb die Beweise, wenn auch teilweise stark verkürzt, erneut vollzogen werden.

Abstract

The present work deals with two estimators regarding discrete observations within the functional regression model

$$Y_{ij} = \mu(x_{ij}) + Z_i(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p$$

with $\hat{\mu}$ being an estimator for the trend μ , and \hat{R} being an estimator for the autocovariance of the random function Z . For the estimators, asymptotic unbiasedness, asymptotic normality and consistency will be established, everything within a functional point of view. Based on this, simultaneous confidence bands will be constructed and additionally a (naive) bootstrap version for these bands will be proposed - including a proof of consistency, conditionally on the random sample. Furthermore, the validity of the confidence bands for the autocovariance will be checked with a simulation, considering several data settings, both for the normal and the bootstrapped bands.

There will be two cases, regarding the observation points x_{ij} : Synchronous points within every sample unit, i. e. $x_{1j} = x_{2j} = \dots = x_{nj}$, and asynchronous points, i. e. $x_{i_1j} \neq x_{i_2j}$ for at least one $j = 1, \dots, p$ and $i_1 \neq i_2$. Even if the proofs for the asynchronous case are pretty much based on the proofs for the synchronous case, there are some crucial differences, hence the proofs will be accomplished once again, however not that detailed.

Inhaltsverzeichnis

1. Einleitung	6
1.1. Ein Überblick über funktionale Daten	6
1.2. Analyse und Interpretation der Daten	8
1.3. Kurzzusammenfassung der Arbeit	9
2. Mathematische Grundlagen	10
2.1. Konvergenz- und Stetigkeitsarten reellwertiger Funktionen	10
2.2. Metrische Entropien	12
2.3. Stochastische Grundbegriffe	14
2.3.1. Gleichmäßige Konvergenz in \mathbb{P}	18
2.3.2. Funktionaler zentraler Grenzwertsatz	23
2.3.3. Bootstrap-Techniken	25
3. Das Modell und die Schätzer	30
3.1. Die Modellgleichung und deren Eigenschaften	30
3.2. Die allgemeinen Schätzer	31
3.2.1. Trendschätzer $\hat{\mu}$	31
3.2.2. Kovarianzschätzer \hat{R}	32
3.3. Annahmen	33
3.3.1. Annahmen an das Modell	33
3.3.2. Annahmen an die Schätzer	37
3.4. Konkrete Beispiele	44
3.4.1. Fehlerterme	44
3.4.2. Der lokal lineare Schätzer	51
4. Eigenschaften der Schätzer	52
4.1. Bias und Varianz	52
4.1.1. Trendschätzer	52
4.1.2. Kovarianzschätzer	55
4.2. Asymptotische Normalität	62
4.2.1. Trendschätzer	62
4.2.2. Kovarianzschätzer	69
4.3. Konsistenz	90
4.3.1. Trendschätzer	90
4.3.2. Kovarianzschätzer	92

5. Konfidenzbänder	98
5.1. Trendschätzer	98
5.2. Kovarianzschätzer	105
6. Bootstrap-Methode der Konfidenzbänder	116
6.1. Trendschätzer	116
6.2. Kovarianzschätzer	132
7. Erweiterung auf asynchrone Messpunkte	152
7.1. Annahmen (asynchrones Modell)	153
7.1.1. Annahmen an das asynchrone Modell	153
7.1.2. Annahmen an die asynchronen Schätzer	154
8. Eigenschaften der asynchronen Schätzer	159
8.1. Bias und Varianz	160
8.1.1. Trendschätzer	160
8.1.2. Kovarianzschätzer	160
8.2. Asymptotische Normalität	164
8.2.1. Trendschätzer	164
8.2.2. Kovarianzschätzer	164
8.3. Konsistenz	166
8.3.1. Trendschätzer	166
8.3.2. Kovarianzschätzer	168
9. Konfidenzbänder (asynchroner Fall)	171
9.1. Trendschätzer	171
9.2. Kovarianzschätzer	172
10. Bootstrap-Methode der Konfidenzbänder (asynchroner Fall)	176
10.1. Trendschätzer	176
10.2. Kovarianzschätzer	187
11. Simulationen	202
11.1. Settings und Aufbau der Simulationen	202
11.2. Fall (hinreichend) große Stichprobe	204
11.3. Fall kleine Stichprobe	210
12. Fazit	215
12.1. Resultate und Interpretation	215
12.2. Ausblick für weitere Nachforschungen	217
A. Anhang	220
A.1. Beweisergänzung zu Satz 2.22	220
A.2. ε aus Setting (S.4) erfüllt Annahme (A.4) nicht	223
A.3. Erfüllbarkeit der Annahmen von Satz 4.4 und Satz 4.5	224

1. Einleitung

1.1. Ein Überblick über funktionale Daten

Die Entstehung und Ausarbeitung von mathematischen Funktionen hat eine lange Geschichte hinter sich. Mit hoher Wahrscheinlichkeit wurde das Konzept von Funktionen bereits in der Vorzeit, wie dem alten Orient oder Ägypten, sicher aber in der Antike unter den ersten berühmten Mathematikern wie Pythagoras, Archimedes oder Euklid verwendet, wenn auch vermutlich in stark vereinfachter Tabellenform. Richtig formalisiert wurden Funktionen jedoch erst im 17. Jahrhundert von René Descartes und Pierre de Fermat¹, was auch schnell die Untersuchung wichtiger Beschaffenheiten - wie Stetigkeit, Differenzierbarkeit und Integrierbarkeit - nach sich zog. Gottfried Wilhelm Leibniz und Isaac Newton beschäftigten sich als erstes mit den Grundlagen der Infinitesimalrechnung, was schließlich über Leonard Euler in der Analysis mündete². Während die Analysis damals noch das reine Studium von Eigenschaften eindimensionaler Funktionen darstellte, hat sie sich heute zu einem Oberbegriff verschiedenster Teilgebiete entwickelt. Eines der wohl wichtigsten davon ist die sogenannte Funktionalanalysis, welche ebenfalls Konzepte wie Stetigkeit oder Differenzierbarkeit ausarbeitet, jedoch nicht von Funktionen $f : D \rightarrow Z$ per se, sondern von Funktionalen und Operatoren, welche wiederum selbst auf eben genannte Funktionen angewendet werden. Es ist also eine noch viel allgemeinere Herangehensweise, bei welcher die Funktionen quasi selbst als einzelne Variablen innerhalb eines bestimmten Raumes (wie eines Banach- oder Hilbertraumes) betrachtet werden. Die Funktionalanalysis entstand im Laufe der 20er Jahre, wobei vor allem Stefan Banach, Frigyes Riesz und Maurice René Fréchet als Begründer gelten³, und ist heutzutage die Basis vieler weiterer Forschungsgebiete; bspw. dient sie als allgemein anerkannte Basis zur Formalisierung der Quantenmechanik⁴.

Verglichen mit obiger Teildisziplin, ist die Stochastik noch ein relativ junges Fachgebiet der Mathematik. Auch wenn durchaus davon ausgegangen werden kann, dass bereits im Altertum einige Berechnungen durchgeführt wurden, welche den Begriff des Zufalls mit einfließen haben lassen, wurde die moderne Wahrscheinlichkeitsrechnung erst 1933 von Andrei Kolmogorow in seinem Lehrbuch „Grundbegriffe der Wahrscheinlichkeitsrechnung“ in umfassender Weise beschrieben und, soweit möglich, axiomatisiert. Eines der wichtigsten Konzepte in allen stochastischen Bereichen sind die Zufallsvariablen, welche erlauben, mathematisch schwer bzw. nicht zu verarbeitende Ereignisse innerhalb eines abstrakten Wahrscheinlichkeitsraumes $(\Omega, \mathcal{A}, \mathbb{P})$ mittels einer messbaren Abbildung $X : \Omega \rightarrow \Omega'$ in Objekte zu überführen, für welche allerlei mathematische Eigenschaf-

¹ [You76] ² [Dun99] ³ [Pie07] ⁴ [Dav15]

ten gelten oder anwendbar sind. Prinzipiell sind dabei dem Bildraum Ω' keine Grenzen gesetzt, solange man ob der Messbarkeit eine dazu passende σ -Algebra \mathcal{A}' sinnvoll definieren kann. In vielen Fällen ist $\Omega' = \mathbb{N}$ oder $\Omega' = \mathbb{R}$ die passende Wahl, seien hier Torzahlen von Sportmannschaften, welche einer Poisson-Verteilung folgen¹, oder Temperaturschwankungen, welche einer Normalverteilung folgen², als zwei von unzähligen Beispielen erwähnt. In der etwas weiterführenden Stochastik ist es allerdings auch gang und gäbe, für Ω' einen Funktionenraum zu wählen, was zu den sogenannten **Zufallsfunktionen**, oder auch **stochastischen Prozessen** führt. Hierbei wird nun auch die Brücke zu oben erläuterten Funktionalanalysis geschlagen, da wir nun (Zufalls)Variablen haben, die Funktionen sind, und damit natürlich auch andere Eigenschaften für sie gelten. Stochastische Prozesse sind mittlerweile ebenfalls in nahezu allen Anwendungsgebieten zu finden, sei es im Finanzsektor, bei der Modellierung von Aktienverläufen, oder aber auch in der Physik, bei der Angabe der Aufenthaltswahrscheinlichkeit von Elementarteilchen.

Aufgrund dieser Tatsache hat sich innerhalb der letzten Jahrzehnte ein mehr oder weniger neues Fachgebiet entwickelt: Die **Funktionale Datenanalyse**, oder kurz **FDA**. Es ist ein eigenständiger Bereich innerhalb der induktiven Statistik und beinhaltet ganz klassische Methoden, wie Schätztheorie und Hypothesentests, allerdings mit dem gravierenden Unterschied, dass die Input Daten bzw. die zugrunde liegenden Modelle nicht aus reellen Zahlen oder endlichdimensionalen Vektoren, sondern aus Funktionen bestehen. Dieses Vorgehen ist nur konsequent, denn liegt in der Natur von vornherein eine Funktion mit stetigem Träger vor, wie z. B. ein zeitabhängiger Temperaturverlauf, so würde eine Vereinfachung in die diskrete Ebene immer Genauigkeitsverluste beinhalten. Auch wenn in der Praxis logischerweise immer nur an endlich vielen Stellen beobachtet werden kann, so wird die zugrunde liegende Entstehung der Daten als Funktion mit stetigem Definitionsbereich, wie z. B. Zeit oder Raum, angesehen. Dies wird dadurch erzielt, indem man annimmt, dass nahe beieinander liegende Werte einen hinreichend kleinen Abstand haben, also den Daten eine einigermaßen glatte Funktion zugrunde liegt, und dass diese theoretisch beliebig fein ausgewertet werden kann. Dadurch können zum Einen sehr hilfreiche Funktionseigenschaften - Stetigkeit, Differenzierbarkeit, Integrierbarkeit - angenommen und ausgenutzt werden, zum Anderen müssen aber natürlich vollkommen andere Ansätze zur Überprüfung wichtiger stochastischer Aussagen verwendet werden. Bestes Beispiel ist der *Satz von Donsker*³, eine funktionale Erweiterung zum zentralen Grenzwertsatz, welcher in dieser Arbeit auch noch Anwendung finden wird. Ein breites Spektrum an wissenschaftlichen Arbeiten zu funktionalen Daten bietet vor allem James Ramsay, häufig in Zusammenarbeit mit Bernard Silverman, siehe z. B. [Ram82], [RS07] oder [FV06]. Bevor wir in medias res gehen, werfen wir einen kurzen Blick darauf, wie die funktionalen Daten in unserem Fall vorliegen und verarbeitet werden.

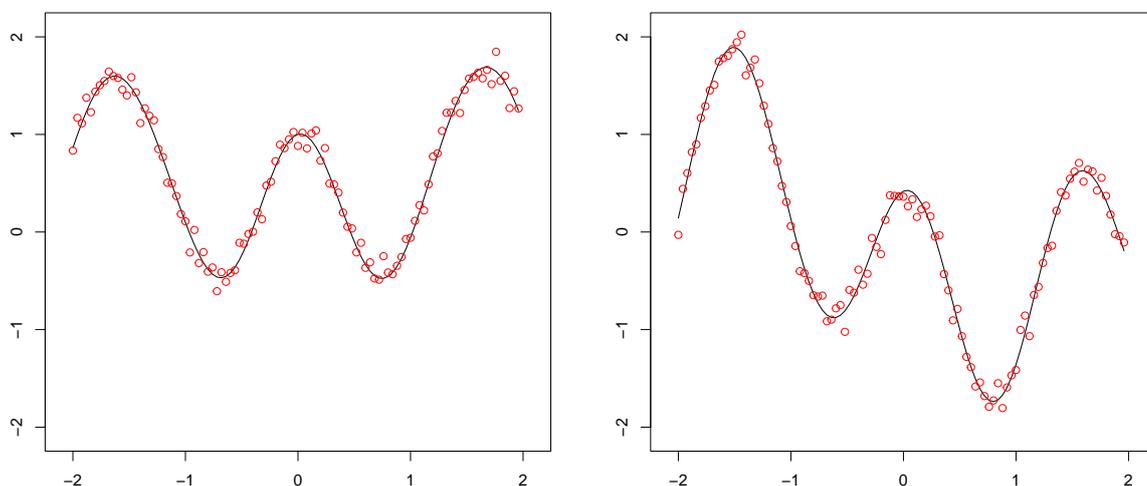
¹ [Dam10] ² [cda] ³ [Don51]

1.2. Analyse und Interpretation der Daten

Wie bereits erwähnt, liegt den Daten, die beobachtet und analysiert werden sollen, eine Funktion mit stochastischem Charakter zugrunde. Wir nehmen an, dass sich diese durch eine deterministische Komponente $\mu : T \rightarrow \mathbb{R}$, und eine stochastische Komponente $Z : T \rightarrow \mathbb{R}$ darstellen lässt, welche unbekannt sind, aber ihrerseits wieder ein gewisses Mindestmaß an Eigenschaften erfüllen, darauf wird später noch genauer eingegangen. Wir beobachten also Daten, die der Funktionsgleichung $Y = \mu(x) + Z(x)$ folgen, wobei jede Beobachtung ebendieser Daten eine andere Ausprägung annehmen kann - denn Z ist eine Zufallsfunktion und jede Beobachtung eine Realisierung ebendieser.

Da es in der Praxis schwierig bis unmöglich ist, die Daten beliebig fein, quasi überabzählbar oft zu messen, ist es zweckmäßig, sich hier auf $p \in \mathbb{N}$ viele Messstellen x_1, \dots, x_p zu beschränken. Was dabei auch noch berücksichtigt werden muss, ist der stets auftretende und nicht beeinflussbare Messfehler ε , was zu Daten der Form $Y_j = \mu(x_j) + Z(x_j) + \varepsilon_j$, $j = 1, \dots, p$ führt. Abbildung 1.1 soll dies ein klein wenig veranschaulichen. Die schwarzen Linien zeigen zwei verschiedene Ausprägungen der Zufallsfunktion $\mu(x) + Z(x)$ (μ ist hier eine Sinusfunktion), die roten Punkte sind die in der Praxis durchführbaren Beobachtungen der Form $Y_j = \mu(x_j) + Z(x_j) + \varepsilon_j$, wobei an $p = 100$ äquidistanten Punkten aus dem Intervall $[-2, 2]$ gemessen wurde.

Abbildung 1.1.: Zwei verschiedene Realisierungen und die dazugehörigen Messungen der Daten.



Unser Ziel ist es nun, eine möglichst gute Schätzung für die wahren Parameter, also μ und bspw. die Kovarianzfunktion von Z zu kreieren. Die Umsetzung davon werden wir in Kürze erfahren. Zunächst aber noch ein paar Worte zum Aufbau der Arbeit.

1.3. Kurzzusammenfassung der Arbeit

Auch wenn, wie in Abschnitt 1.2 erwähnt, die Beobachtungen diskret vorliegen, und es sich somit streng genommen um reellwertige Zufallsvariablen, also ein klassisches Regressionsmodell handelt, sollen ja nach wie vor Aussagen für den komplett funktionalen Fall getroffen werden, was durch eine Grenzwertbildung gelingt, indem immer mehr - im Grenzfall unendlich viele - Messpunkte miteinbezogen werden und somit eine Verstetigung erfolgt. Damit dies gelingt, werden zunächst einmal in Kapitel 2 grundlegende stochastische Konzepte rekapituliert bzw. auf den funktionalen Fall erweitert; allen voran der funktionale zentrale Grenzwertsatz in Abschnitt 2.3.2 spielt dabei eine entscheidende Rolle, jedoch in einer leicht anderen, für unsere Zwecke zugeschnittenen Version, als in der klassischen Formulierung nach Donsker.

Kapitel 3 führt das konkrete Modell und die Schätzer von Interesse ein. Für beides werden mehrere Annahmen formuliert, welche für die erarbeiteten Ergebnisse unabdingbar sind, und weil das Modell und die Schätzer sehr allgemein formuliert sind, werden auch konkrete Beispiele angeführt, um das Ganze etwas anschaulicher zu gestalten.

Kapitel 4 bildet den Hauptteil der Arbeit, da dort die wichtigsten Eigenschaften - allen voran die asymptotische Normalität - der Schätzer ausgearbeitet werden. Aus diesen Eigenschaften werden in Kapitel 5 simultane Konfidenzbänder für die tatsächlichen Funktionen abgeleitet und Kapitel 6 liefert eine (konsistente) Bootstrap-Methode für diese Konfidenzbänder, für den Fall einer zu kleinen Stichprobe.

In Kapitel 7, 8, 9 und 10 werden prinzipiell alle Aussagen und Beweise aus Kapitel 4, 5 und 6 erneut durchgeführt, allerdings für den noch allgemeineren Fall, dass sich die Messpunkte innerhalb jeder Stichprobe unterscheiden können, was zuvor ausgeschlossen wurde. Die Gültigkeit der Aussagen für diesen Fall mag auf den ersten Blick offensichtlich erscheinen, allerdings bedarf dies durchaus einer detaillierteren Überprüfung, da des Öfteren eine Mittelung über die Stichprobenwerte, also bezüglich der Messpunkte stattfindet. Unterscheiden sich diese in jeder Einheit, unterscheiden sich natürlich auch die Funktionswerte, somit ist erst einmal nicht gesichert, dass die erörterten Methoden noch funktionieren.

Kapitel 11 bestätigt zum Einen die Gültigkeit der Konfidenzbänder und der dazugehörigen Bootstrap-Methode für einen Schätzer und verschiedene Datensettings in einer Simulationsstudie, zum Anderen werden die verschiedenen Methoden miteinander verglichen und vorgeschlagen, in welcher Situation welche Methode vorzuziehen ist. Abschließend werden in Kapitel 12 alle Resultate noch einmal kurz zusammengefasst und zudem ein kleiner Ausblick für mögliche weitere Forschungsaufgaben, diese Thematik betreffend, angegeben.

2. Mathematische Grundlagen

Auch wenn diese Arbeit im Wesentlichen dem Fachgebiet Stochastik zuzuschreiben ist, spielen funktionale Daten, also letztlich Funktionen eine wesentliche Rolle, daher werden wir kurz auf ein paar funktionalanalytische Konzepte eingehen, bevor in Abschnitt 2.2 und 2.3 detailliertere Aussagen folgen. Wir werden uns aber generell auf einer etwas fortgeschritteneren Ebene bewegen und absolut Grundlegendes als bekannt voraussetzen, wie z. B. die Definition eines stochastischen Prozesses oder die populäre *Cauchy-Schwarz-Ungleichung*.

2.1. Konvergenz- und Stetigkeitsarten reellwertiger Funktionen

Viele der in diesem Abschnitt auftretenden Definitionen sind wohl in den meisten Einführungsbüchern zu Analysis, wie [Kön03] und [Kön06] zu finden; gegen Ende wird es jedoch etwas spezifischer, wobei man auf weiterführende Fachliteratur wie [CR87] oder [MV13] zurückgreifen kann.

Seien $(f_n)_{n \in \mathbb{N}}$ und f mit $f, f_n : D \rightarrow \mathbb{R}$ reellwertige Funktionen. Da in dieser Arbeit Konvergenzaussagen von stochastischen Prozessen, also letztlich von Funktionenfolgen, eine große Rolle spielen, wollen wir diese noch einmal kurz rekapitulieren. Für eine Funktionenfolge gibt es zwei Konvergenzbegriffe, die für uns von Bedeutung sein werden.

Definition 2.1. (*Punktweise Konvergenz*)

f_n konvergiert genau dann **punktweise** gegen f , wenn:

$$\forall x \in D, \forall \gamma > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0 : |f_n(x) - f(x)| < \gamma$$

Äquivalente Formulierung: $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ für alle $x \in D$.

Definition 2.2. (*Gleichmäßige Konvergenz*)

f_n konvergiert genau dann **gleichmäßig** gegen f , wenn:

$$\forall \gamma > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \forall x \in D : |f_n(x) - f(x)| < \gamma$$

Äquivalente Formulierung: $\lim_{n \rightarrow \infty} \sup_{x \in D} |f_n(x) - f(x)| = 0$.

Die Definitionen mögen sehr ähnlich aussehen, sind jedoch nicht äquivalent, erkennbar an der Reihenfolge der Quantoren. Gleichmäßige Konvergenz ist stärker als punktweise, insbesondere folgt aus gleichmäßiger Konvergenz auch immer die punktweise Konvergenz, aber nicht umgekehrt. Ein bekanntes Beispiel hierfür ist die Funktionenfolge $f_n(x) = x^n$ definiert auf dem halboffenen Intervall $D = [0, 1)$. Diese konvergiert zwar punktweise gegen die Nullfunktion $f \equiv 0$, allerdings nicht gleichmäßig. Die gleichmäßige Konvergenz bietet im Gegensatz zur punktweisen Konvergenz den Vorteil, wichtige Eigenschaften der Funktionenfolge, wie z. B. Stetigkeit, auf die Grenzfunktion zu übertragen, daher ist in dieser Arbeit generell die gleichmäßige Konvergenz die erstrebenswerte Aussage.

Die Stetigkeit einer Funktion ist ebenfalls ein hinlänglich bekannter Begriff, aber auch hier gibt es verschiedene Arten.

Definition 2.3. (*Stetigkeit*)

f ist genau dann **stetig**, wenn:

$$\forall x_0 \in D, \forall \gamma > 0, \exists \delta > 0, \forall x \in D : |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \gamma$$

Äquivalente Formulierung: $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ für alle $x_0 \in D$.

Definition 2.4. (*Gleichmäßige Stetigkeit*)

f ist genau dann **gleichmäßig stetig**, wenn:

$$\forall \gamma > 0, \exists \delta > 0, \forall x, x_0 \in D : |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \gamma$$

Auch diese Definitionen sind nicht äquivalent, bspw. ist die Funktion $f(x) = x^2$ zwar stetig, aber nicht gleichmäßig stetig. Die gleichmäßige Stetigkeit ist also wieder stärker und aus ihr folgt die Stetigkeit.

Betrachtet man nun Funktionenfolgen, so kann der Stetigkeitsbegriff noch einmal anders festgelegt werden.

Definition 2.5. (*Gleichgradige Stetigkeit*)

f_n ist genau dann **gleichgradig stetig**, wenn:

$$\forall x_0 \in D, \forall \gamma > 0, \exists \delta > 0, \forall n \in \mathbb{N}, \forall x \in D : |x - x_0| < \delta \Rightarrow |f_n(x) - f_n(x_0)| < \gamma$$

Die gleichmäßige Stetigkeit besagt also im Prinzip, dass die Stetigkeit in jedem Argument der Funktion, die gleichgradige Stetigkeit wiederum, dass die Stetigkeit für jede Funktion der Folge „gleich stark“ ausgeprägt ist. Hätte man nur die reine Stetigkeit, so kann bei vorgegebenem γ für jedes $x_0 \in D$ bzw. jedes $n \in \mathbb{N}$ eine separate Konstante δ_{x_0} bzw. δ_n existieren. Bei gleichmäßiger bzw. gleichgradiger Stetigkeit muss dieses δ global bzgl. x_0 bzw. n konstant sein.

Diese Stetigkeitsbegriffe können zu guter Letzt auch noch zusammengefasst werden.

Definition 2.6. (*Gleichmäßige gleichgradige Stetigkeit*)

f_n ist genau dann **gleichmäßig gleichgradig stetig**, wenn:

$$\forall \gamma > 0, \exists \delta > 0, \forall n \in \mathbb{N}, \forall x, x_0 \in D : |x - x_0| < \delta \Rightarrow |f_n(x) - f_n(x_0)| < \gamma$$

Bei vorgegebenem γ muss also ein δ existieren, so dass die Implikation für alle Funktionen der Folge und für alle Punkte der Definitionsmenge gilt. δ hängt also nur von γ ab, weder von n noch x_0 .

Die gleichmäßige gleichgradige Stetigkeit wird später noch eine sehr wichtige Rolle spielen, weil sich dadurch über ein paar Umwege die erwünschte gleichmäßige Konvergenz herleiten lässt. Allerdings ist es im Allgemeinen extrem schwierig, diese mittels obiger Definition zu überprüfen. Wir führen daher ein weiteres mathematisches Werkzeug ein, welches den Nachweis oftmals deutlich vereinfacht.

2.2. Metrische Entropien

Metrische Entropien und deren Abwandlungen sind mittlerweile ein wichtiger Baustein in der Stochastik, aber auch in anderen Bereichen wie der Chaostheorie. Grob formuliert sind sie eine Art Kennzahl für den Grad der Kompaktheit eines metrischen Raumes. Dadurch lassen sich viele problematische Berechnungen abschätzen, wie z. B. Suprema von stochastischen Prozessen, was später noch eine Rolle für die gleichmäßige gleichgradige Stetigkeit ebendieser Prozesse spielen wird. Erstmals definiert wurden metrische Entropien von Andrei Kolmogorow und weiter ausgearbeitet von Jakow Sinai in den 50er Jahren, sie werden daher zum Teil auch als *Kolmogorow-Sinai-Entropien* bezeichnet. Einen exzellenten Überblick über metrische Entropien, deren Eigenschaften und Geschichte bietet [Tik93], eine Zusammenfassung verschiedenster Arbeiten von Kolmogorow et al. Generell sind aber metrische Entropien heutzutage absolut keine Seltenheit mehr in fundierten mathematischen Lehrbüchern.

Für die metrischen Entropien benötigen wir zunächst noch das Konzept (pseudo)metrischer Räume, welches wir kurz wiederholen.

Definition 2.7. (*Pseudometrik und pseudometrischer Raum*)

Sei M eine beliebige Menge. Eine Abbildung $d : M \times M \rightarrow \mathbb{R}_0^+$ heißt **Pseudometrik**, wenn für beliebige Elemente $x, y, z \in M$ die folgenden Bedingungen erfüllt sind.

i) $x = y \Rightarrow d(x, y) = 0$

ii) $d(x, y) = d(y, x)$

iii) $d(x, z) \leq d(x, y) + d(y, z)$

(M, d) heißt dann **pseudometrischer Raum**.

Der einzige Unterschied zur Definition einer Metrik ist, dass i) nicht in Äquivalenz gilt. Das heißt, es kann Elemente aus M geben, die verschieden sind, deren metrischer Abstand aber dennoch 0 ist. Gibt es solche Elemente in M , dann ist d eine *echte* Pseudometrik. Gibt es sie nicht, dann ist d sogar eine Metrik. Das bedeutet, jede Metrik ist automatisch auch eine Pseudometrik, aber nicht umgekehrt.

Definition 2.8. (*Kapazitäts-, Überdeckungszahl und metrische Entropie*)

Sei (M, d) ein pseudometrischer Raum, $T \subseteq M$ und $\delta > 0$.

$$N_{\text{cap}}(\delta, d, T) := \max \{ \#T_0 \mid T_0 \subseteq T, \forall t, t' \in T_0, t \neq t' : d(t, t') > \delta \}$$

heißt **Kapazitätzahl von T für δ** . Sie kann interpretiert werden als die maximale Anzahl von abgeschlossenen Hyperkugeln der Dimension $D := \dim(T)$ mit Radius δ bezüglich d , welche man in T packen kann, ohne dass die Kugeln sich überschneiden. Sie wird daher auch als **packing Number** bezeichnet.

$$N_{\text{cov}}(\delta, d, T) := \min \{ \#M_0 \mid M_0 \subseteq M, \forall t \in T : \inf_{t' \in M_0} d(t, t') \leq \delta \}$$

heißt **Überdeckungszahl von T für δ** . Sie kann interpretiert werden als die minimale Anzahl von abgeschlossenen Hyperkugeln der Dimension $D := \dim(T)$ mit Radius δ bezüglich d , welche T vollständig überdecken. Sie wird daher auch als **covering Number** bezeichnet.

Die **metrische Entropie von T für δ** ist schließlich definiert durch $\log(N_{\text{cov}}(\delta, d, T))$.

Die Kapazitätzahl kann mittels der Überdeckungszahl beschränkt werden, und umgekehrt.

Lemma 2.9.

Sei (M, d) ein pseudometrischer Raum, $T \subseteq M$ und $\delta > 0$. Dann gilt:

$$N_{\text{cap}}(2\delta, d, T) \leq N_{\text{cov}}(\delta, d, T) \leq N_{\text{cap}}(\delta, d, T) \leq N_{\text{cov}}\left(\frac{\delta}{2}, d, T\right)$$

Beweis.

Siehe [Düm17], Seite 52 und [Wai19], Seite 124. □

Es gibt noch einige weitere Möglichkeiten, die Kennzahlen aus Definition 2.8 festzulegen. Bspw. kann man die Überdeckungszahl auch noch in **äußere** und **innere** Überdeckungszahl unterteilen, wobei erstere unserer Definition entspricht und bei letzterer die Mittelpunkte der Hyperkugeln in der betrachteten Menge T liegen müssen. Da es aber für all diese Definitionen ebenfalls gegenseitige Abschätzungen gibt, sind die Definitionen alle mehr oder weniger äquivalent, ähnlich wie bei der Äquivalenz von p -Normen, siehe dafür z. B. [Tao14].

Beispiel 2.10.

Gegeben sei der (pseudo)metrische Raum (\mathbb{R}^n, d^p) , wobei d^p die durch eine p -Norm erzeugte Metrik ist. Sei zudem $T \subset \mathbb{R}^n$ eine Menge innerhalb eines D -dimensionalen Unterraums von \mathbb{R}^n und $T_{\max} := \max_{t \in T} \|t\|_p$. Dann gilt:

$$N_{\text{cov}}(\delta, d^p, T) \leq \left(\frac{2T_{\max} \sqrt{D}}{\delta} \right)^D$$

Diese Abschätzung ist aus [SSBD14], Seite 388. Hier bewiesen für $p = 2$, also der euklidischen Metrik, was sich jedoch für jedes $p \in [1, \infty]$ äquivalent durchführen lässt.

2.3. Stochastische Grundbegriffe

Wir kommen nun zu den stochastischen Grundlagen, welche für diese Arbeit benötigt werden. Vieles davon dürfte als klassische Aussage bereits bekannt sein, hier nun erweitert auf eine funktionale Ebene. Vorab sei noch erwähnt, dass wir künftig für reellwertige Zufallsvariablen $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ und reellwertige stochastische Prozesse $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathcal{C}(T), \mathcal{C})$ - \mathcal{B} ist die borelsche σ -Algebra und \mathcal{C} eine geeignete σ -Algebra auf dem Raum der stetigen Funktionen; wir beschränken uns von vornherein größtenteils auf stetige Prozesse - auf die konkrete Angabe eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, \mathbb{P})$, sowie des Arguments ω verzichten werden, da es uns hauptsächlich um Konvergenz- bzw. Pfadeneigenschaften gehen wird, wofür der dazugehörige Wahrscheinlichkeitsraum nicht wirklich eine Rolle spielt. Zudem sei generell für alle $\omega \in \Omega$ und $t \in T$ die unabdingbare Messbarkeitseigenschaft der Zufallsobjekte erfüllt.

Definition 2.11. (Stochastische Konvergenzarten)

Seien $X, (X_n)_{n \in \mathbb{N}}$ mit $X := \{X(t) \mid t \in T\}$ und $X_n := \{X_n(t) \mid t \in T\}$ stetige reellwertige stochastische Prozesse.

- a) X_n konvergiert genau dann **punktweise** bzw. **gleichmäßig fast sicher** (f. s.) gegen X , wenn gilt:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} |X_n(t) - X(t)| = 0\right) = 1, \forall t \in T \quad \text{bzw.} \quad \mathbb{P}\left(\lim_{n \rightarrow \infty} \sup_{t \in T} |X_n(t) - X(t)| = 0\right) = 1$$

Schreibweise: $X_n \xrightarrow{\text{f.s.}} X$.

- b) X_n konvergiert genau dann **punktweise** bzw. **gleichmäßig in \mathbb{P}** gegen X , wenn für alle $\gamma > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(|X_n(t) - X(t)| > \gamma\right) = 0, \forall t \in T \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{t \in T} |X_n(t) - X(t)| > \gamma\right) = 0$$

Man spricht hier auch von **Konvergenz in Wahrscheinlichkeit** bzw. **stochastischer Konvergenz**.

Schreibweise: $X_n \xrightarrow{\mathbb{P}} X$.

c) Sei $p \in \mathbb{R}^+$. X_n konvergiert genau dann **punktweise** bzw. **gleichmäßig im p -ten Mittel gegen X** , wenn gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n(t) - X(t)|^p] = 0, \quad \forall t \in T \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} \mathbb{E}\left[\sup_{t \in T} |X_n(t) - X(t)|^p\right] = 0$$

Schreibweise: $X_n \xrightarrow{\mathcal{L}^p} X$.

d) X_n konvergiert genau dann **schwach gegen X in $(\mathcal{C}(T), \|\cdot\|_\infty)$** , wenn für alle $f \in \mathcal{C}_{u,b}(\mathcal{C}(T), \|\cdot\|_\infty)$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

Es handelt sich also um die beschränkten, bzgl. der durch die Supremumsnorm induzierten Metrik gleichmäßig stetigen, reellwertigen Funktionale $f : \mathcal{C}(T) \rightarrow \mathbb{R}$.

Schreibweise: $X_n \xrightarrow{w} X$ in $(\mathcal{C}(T), \|\cdot\|_\infty)$.

Teil d) bedarf ggf. noch einer etwas ausführlicheren Erklärung. Im Wesentlichen handelt es sich hierbei um die relativ geläufige **Konvergenz in Verteilung**, formuliert im maßtheoretischen Sinne. Die Beschaffenheit von f ist deshalb so komplex, weil es sich bei $X, (X_n)_{n \in \mathbb{N}}$ nicht um „normale“ Zufallsvariablen, sondern um stochastische Prozesse handelt, also um Zufallsvariablen, welche Werte im Vektorraum $\mathcal{C}(T)$ annehmen. Der asymptotische Grenzprozess muss demnach wieder in $\mathcal{C}(T)$ liegen, wobei $\|\cdot\|_\infty$ wiederum die dazu korrespondierende Norm angibt. Theoretisch könnte man hier auch andere Normen verwenden, allerdings stellt $(\mathcal{C}(T), \|\cdot\|_\infty)$ im Falle eines kompakten Trägers T einen *Banachraum* dar, innerhalb dessen die betrachteten stochastischen Prozesse messbar bzgl. $t \in T$ und $\omega \in \Omega$ sind. Wäre das nicht der Fall, müsste mit *äußeren Wahrscheinlichkeitsmaßen und Erwartungswerten* gearbeitet werden, was noch einmal komplizierter wäre.

Die schwache Konvergenz für Zufallsvariablen mit Werten im \mathbb{R}^n ist äquivalent definiert, nur dass in diesem Fall $f \in \mathcal{C}_{u,b}(\mathbb{R}^n, \|\cdot\|)$ und $\|\cdot\|$ in aller Regel eine p -Norm ist. Hier lässt sich die Konvergenz über das *Portmanteau-Theorem* oftmals deutlich leichter durch die punktweise Konvergenz der Verteilungsfunktionen oder der charakteristischen Funktionen zeigen. Um diese beiden Fälle etwas voneinander abzugrenzen, sprechen wir im weiteren Verlauf von schwacher Konvergenz für stochastische Prozesse, und Konvergenz in Verteilung für reellwertige Zufallsvektoren, Schreibweise: $X_n \xrightarrow{D} X$. Streng genommen handelt es sich bei beiden aber um die gleiche Konvergenzart. Zwischen der schwachen Konvergenz von stochastischen Prozessen und der Konvergenz in Verteilung von Zufallsvektoren gibt es noch eine weitere Verknüpfung, welche wir später brauchen werden.

Definition 2.12. (Konvergenz der endlichdimensionalen Randverteilungen)

Seien $X, (X_n)_{n \in \mathbb{N}}$ mit $X := \{X(t) \mid t \in T\}$ und $X_n := \{X_n(t) \mid t \in T\}$ reellwertige stochastische Prozesse. X_n konvergiert genau dann bzgl. der **endlichdimensionalen Randverteilungen gegen X** , falls für alle $t_1, \dots, t_j \in T$ mit $j \in \mathbb{N}$ gilt:

$$(X_n(t_1), X_n(t_2), \dots, X_n(t_j))^T \xrightarrow{D} (X(t_1), X(t_2), \dots, X(t_j))^T$$

Schreibweise: $X_n \xrightarrow{\text{fdd}} X$.

Die verschiedenen Konvergenzarten stehen in folgender Beziehung zueinander:

Proposition 2.13.

Seien $X, (X_n)_{n \in \mathbb{N}}$ mit $X := \{X(t) \mid t \in T\}$ und $X_n := \{X_n(t) \mid t \in T\}$ stetige reellwertige stochastische Prozesse.

- a) $X_n \xrightarrow{\text{f.s.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$, jeweils punktweise, als auch gleichmäßig.
- b) $X_n \xrightarrow{\mathcal{L}^p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$ für alle $p \in \mathbb{R}^+$, jeweils punktweise, als auch gleichmäßig.
- c) $X_n \xrightarrow{\mathbb{P}} X$ punktweise $\Rightarrow X_n \xrightarrow{\text{fdd}} X$.
- d) $X_n \xrightarrow{\mathbb{P}} X$ gleichmäßig $\Rightarrow X_n \xrightarrow{w} X$ in $(\mathcal{C}(T), \|\cdot\|_\infty)$.
- e) $\forall t \in T : X_n(t) \xrightarrow{D} a \in \mathbb{R} \Rightarrow X_n \xrightarrow{\mathbb{P}} a$ punktweise.
- f) $X_n \xrightarrow{w} a \in \mathbb{R}$ in $(\mathcal{C}(T), \|\cdot\|_\infty) \Rightarrow X_n \xrightarrow{\mathbb{P}} a$ gleichmäßig.
- g) $X_n \xrightarrow{\mathbb{P}} X$ punktweise und $\forall t \in T, \forall \gamma > 0 : \sum_{n=1}^{\infty} \mathbb{P}(|X_n(t) - X(t)| > \gamma) < \infty$
 $\Rightarrow X_n \xrightarrow{\text{f.s.}} X$ punktweise.

Proposition 2.14.

Seien $(X_n)_{n \in \mathbb{N}}$ und $(Y_n)_{n \in \mathbb{N}}$ Folgen reellwertiger Zufallsvariablen und $f : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion.

- a) $X_n \xrightarrow{\text{f.s.}} X$ und $Y_n \xrightarrow{\text{f.s.}} Y \Rightarrow X_n + Y_n \xrightarrow{\text{f.s.}} X + Y$ und $X_n Y_n \xrightarrow{\text{f.s.}} XY$.
- b) $X_n \xrightarrow{\mathbb{P}} X$ und $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$ und $X_n Y_n \xrightarrow{\mathbb{P}} XY$.
- c) $X_n \xrightarrow{\text{f.s.}} X \Rightarrow f(X_n) \xrightarrow{\text{f.s.}} f(X)$, $X_n \xrightarrow{\mathbb{P}} X \Rightarrow f(X_n) \xrightarrow{\mathbb{P}} f(X)$, $X_n \xrightarrow{D} X \Rightarrow f(X_n) \xrightarrow{D} f(X)$.

Die Aussagen und Beweise von Proposition 2.13 und 2.14 sind in zahlreichen stochastischen Lehrbüchern zu finden. Anzumerken ist hierbei, dass die Aussagen meist nur für reellwertige Zufallsvariablen, also bzgl. stochastischer Prozesse punktweise formuliert sind. Da sich für die Gleichmäßigkeit bei den Konvergenzarten nach Definition 2.11 aber lediglich die dazugehörige Norm ändert, sind die Beweise dazu äquivalent.

Zu guter Letzt noch zwei der vermutlich bekanntesten Lemmata, was Konvergenzeigenschaften in der Stochastik angeht.

Lemma 2.15. (Satz von Slutsky)

Seien $(X_n)_{n \in \mathbb{N}}, (A_n)_{n \in \mathbb{N}}, (B_n)_{n \in \mathbb{N}}$ mit $X_n := \{X_n(t) \mid t \in T\}$, $A_n := \{A_n(t) \mid t \in T\}$ und $B_n := \{B_n(t) \mid t \in T\}$ stetige reellwertige stochastische Prozesse.

a) Falls $\forall t \in T : X_n(t) \xrightarrow{D} X(t)$, $A_n \xrightarrow{\mathbb{P}} a$ punktweise und $B_n \xrightarrow{\mathbb{P}} b$ punktweise, dann gilt für alle $t \in T$:

$$A_n + B_n X_n(t) \xrightarrow{D} a + bX(t)$$

b) Falls $X_n \xrightarrow{w} X$ in $(\mathcal{C}(T), \|\cdot\|_\infty)$, $A_n \xrightarrow{\mathbb{P}} a$ gleichmäßig und $B_n \xrightarrow{\mathbb{P}} b$ gleichmäßig, dann gilt:

$$A_n + B_n X_n \xrightarrow{w} a + bX, \text{ in } (\mathcal{C}(T), \|\cdot\|_\infty)$$

Beweis.

Siehe [vdVvdVW96], Seite 32. Hier wird das Lemma für $A_n \xrightarrow{D} a$ und $B_n \xrightarrow{D} b$ bzw. $A_n \xrightarrow{w} a$ und $B_n \xrightarrow{w} b$ formuliert, was aber nach Proposition 2.13 e) bzw. f) jeweils aus der Konvergenz in \mathbb{P} folgt. \square

Lemma 2.16. (Kolmogorows starkes Gesetz der großen Zahlen)

Sei $(X_n)_{n \in \mathbb{N}}$ eine stochastisch unabhängige Folge reellwertiger Zufallsvariablen.

(1. SLLN) Falls $\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} < \infty$, so genügt $(X_n)_{n \in \mathbb{N}}$ dem starken Gesetz der großen Zahlen, also:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{\text{f.s.}} 0$$

Dies ist insbesondere erfüllt, wenn $\text{Var}(X_n)$ für alle $n \in \mathbb{N}$ durch eine Konstante $C \in \mathbb{R}^+$ beschränkt ist, denn dann gilt nach dem Basler Problem:

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} \leq C \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{C\pi^2}{6} < \infty$$

(2. SLLN) Falls $(X_n)_{n \in \mathbb{N}}$ identisch verteilt ist und zudem $\mathbb{E}[|X_1|] < \infty$, so genügt $(X_n)_{n \in \mathbb{N}}$ dem starken Gesetz der großen Zahlen, also:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{f.s.}} \mathbb{E}[X_1]$$

Die beiden Aussagen sind auch als Kolmogorows 1. bzw. 2. Gesetz der großen Zahlen bekannt, daher (1. SLLN) bzw. (2. SLLN) (1. bzw. 2. Strong Law of Large Numbers).

Beweis.

Siehe [Sch11], Seite 345 - 349. \square

Prinzipiell hätten wir nun alle Werkzeuge beisammen, die für die Beweise der Aussagen dieser Arbeit nötig sind. Das Problem ist, dass es in vielen Fällen schwierig bis unmöglich ist, die gleichmäßigen Konvergenzen direkt, also so wie sie in Definition 2.11 angegeben sind, zu zeigen; hauptsächlich wegen des Supremums innerhalb der Wahrscheinlichkeiten bzw. Erwartungswerte. Wir lernen daher nun eine andere Methode kennen, um zumindest die gleichmäßige Konvergenz in \mathbb{P} nachzuweisen.

2.3.1. Gleichmäßige Konvergenz in \mathbb{P}

Die Grundidee hierbei ist, die Konvergenz an sich nur punktweise nachzuweisen, und an den stochastischen Prozess noch zusätzliche Anforderungen zu stellen, aus denen so - in Verbindung mit der punktweisen Konvergenz - die gleichmäßige Konvergenz folgt. Hierbei handelt es sich um die gleichmäßige gleichgradige Stetigkeit aus Definition 2.6, allerdings nun wieder in einem stochastischen Kontext.

Definition 2.17. (*Gleichmäßige gleichgradige Stetigkeit in \mathbb{P}*)

Sei $(X_n)_{n \in \mathbb{N}}$ mit $X_n := \{X_n(t) \mid t \in T\}$ eine Folge von reellwertigen stochastischen Prozessen. Falls für alle $\gamma, \eta > 0$ ein $\delta > 0$ und ein $n_0 \in \mathbb{N}$ existiert, so dass für alle $n \geq n_0$:

$$\mathbb{P}\left(\sup_{d(t,t') \leq \delta} |X_n(t) - X_n(t')| > \gamma\right) < \eta$$

dann heißt $(X_n)_{n \in \mathbb{N}}$ **gleichmäßig gleichgradig stetig in \mathbb{P}** . d ist dabei eine geeignete Metrik auf T . Eine alternative Formulierung dazu wäre, dass

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{d(t,t') \leq \delta} |X_n(t) - X_n(t')| > \gamma\right) = 0$$

für alle $\gamma > 0$ gelten soll.

Lemma 2.18.

Seien $X, (X_n)_{n \in \mathbb{N}}$ mit $X := \{X(t) \mid t \in T\}$ und $X_n := \{X_n(t) \mid t \in T\}$ reellwertige stochastische Prozesse, wobei T kompakt ist. Falls für $(X_n)_{n \in \mathbb{N}}$ die beiden folgenden Bedingungen erfüllt sind:

- i) X_n konvergiert punktweise in \mathbb{P} gegen X .
- ii) $(X_n)_{n \in \mathbb{N}}$ ist gleichmäßig gleichgradig stetig in \mathbb{P} .

Dann konvergiert X_n gleichmäßig in \mathbb{P} gegen X .

Beweis.

Siehe [New91] oder [FP07]. □

Für den Nachweis der punktweisen Konvergenz in \mathbb{P} gibt es zahlreiche Verfahren, dieser ist also in aller Regel nicht das Problem. Der Nachweis der gleichmäßigen gleichgradigen Stetigkeit ist allerdings bei weitem komplexer, weil nach wie vor das Supremum innerhalb der Wahrscheinlichkeit extreme Probleme bereitet. Wir formulieren daher zwei Lemmata, mit welchen sich gleichmäßige gleichgradige Stetigkeit einer Folge von stochastischen Prozessen folgern lässt.

Lemma 2.19.

Sei $(X_n)_{n \in \mathbb{N}}$ mit $X_n := \{X_n(t) \mid t \in [0, 1]^D\}$, mit $D \in \mathbb{N}$, eine Folge von reellwertigen stochastischen Prozessen. Falls gilt:

i) Es existiert ein messbares $\Phi : [0, 1]^D \rightarrow \mathbb{R}$, welches für alle $n \in \mathbb{N}$ die Ungleichung $|X_n| \leq \Phi$ und zugleich $(\mathbb{E}[|\Phi|^2])^{\frac{1}{2}} < \infty$ erfüllt.

ii) Für alle $\gamma, \eta > 0$ existiert ein $\delta > 0$, so dass

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_0^\delta \left(\log \left(\frac{N_{\text{cov}}(u, d_n, [0, 1]^D)}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du > \gamma \right) < \eta$$

dann ist $(X_n)_{n \in \mathbb{N}}$ gleichmäßig gleichgradig stetig in \mathbb{P} . $N_{\text{cov}}(u, d_n, [0, 1]^D)$ ist dabei die Überdeckungsanzahl von $[0, 1]^D$ bzgl. der kanonischen Metrik

$$d_n(t, t') = \left(\mathbb{E}[(X_n(t) - X_n(t'))^2] \right)^{\frac{1}{2}}$$

Das Integral im Inneren wird auch als **Überdeckungsintegral** bezeichnet.

Im Wesentlichen steht im Inneren die metrische Entropie von $[0, 1]^D$ für d_n , nur ist das Argument im Logarithmus den Faktor $\frac{1}{\sqrt{u}}$ vergrößert. Dieses Integral wird oft als *Dudley's Entropie Integral* formuliert. Generell geht das Lemma sehr in Richtung *Dudley's Theorem* bzw. *Dudley's Bound*, welches in verschiedenen Varianten gleichmäßige Stetigkeit und Beschränktheit von (sub)Gaußprozessen impliziert, siehe z. B. [Dud67], [Adl90], [Tal14] oder [Wai19]. Die Voraussetzungen sind in unserem Fall jedoch stärker, da wir zur Gleichmäßigkeit auch noch die Gleichgradigkeit benötigen. Genau deswegen auch die Vergrößerung des Logarithmus Arguments: Die Entropie wird künstlich aufgebauscht, um eine „stärkere“ Stetigkeit zu garantieren.

Beweis.

Siehe [Pol84], Seite 150. Hierzu noch einige Anmerkungen.

Pollard beschreibt die Stetigkeit, wie viele andere Autoren auch, schlichtweg als „Equicontinuity“, was streng genommen nur die gleichgradige Stetigkeit darstellt. Die Gleichmäßigkeit ist jedoch trotzdem gegeben, da die Funktionsvariable t im daraus resultierenden ε - δ -Kriterium keine Rolle spielt. Um Missverständnisse zu vermeiden, werden wir in dieser Arbeit immer von gleichmäßiger gleichgradiger Stetigkeit sprechen.

Generell ist der Satz ein klein wenig anders bzw. etwas allgemeiner als in der hier vorliegenden Arbeit formuliert. Zum einen gibt er den Integranden mit $\left(2 \log \left(\frac{N_{\text{cov}}(u, d_n, [0, 1]^D)^2}{u} \right) \right)^{\frac{1}{2}}$ an, was sich jedoch leicht aus unserem ergibt, wenn wir $\gamma = \frac{\gamma'}{2}$ mit $\gamma' > 0$ setzen. Darüber hinaus wird die Überdeckungsanzahl dort nicht mit der Metrik d_n und der Menge $[0, 1]^D$, also der Indexmenge der stochastischen Prozesse X_n , indiziert, sondern mit dem empirischen Maß \mathbb{P}_n und sogenannten *zulässigen Klassen* \mathcal{F} . Nichtsdestoweniger erfüllt der hier formulierte Satz Pollards Voraussetzungen, denn:

- Die Überdeckungszahl wird bezüglich der $\mathcal{L}^2(\mathbb{P}_n)$ -Halbnorm induzierten Pseudometrik bestimmt. Das ist aber gerade d_n .
- Die Definition der zulässigen Klassen findet sich auf Seite 196. Sie besteht aus der Messbarkeit bzgl. $\mathcal{A} \otimes \mathcal{B}(T)$, welche nach unseren Voraussetzungen gilt, und der Eigenschaft, dass die Indexmenge T ein *Suslin-Raum* sein soll, was eine Verallgemeinerung eines polnischen Raumes ist. $[0, 1]^D$ ist aber als abgeschlossene Teilmenge bzw. Unterraum des polnischen Raumes \mathbb{R}^D selbst ein polnischer Raum und somit auch ein Suslin-Raum, wonach diese Voraussetzung ebenfalls gilt, siehe dafür z. B. [Bau11], Seite 157.

Die Folgerung der gleichmäßigen gleichgradigen Stetigkeit, also das ε - δ -Kriterium, ist ebenfalls bzgl. zulässiger Klassen formuliert, was aber wieder nur eine Verallgemeinerung der hier vorliegenden Situation ist. Ebendiese Verallgemeinerung garantiert uns auch die Gleichmäßigkeit. \square

Lemma 2.20.

Sei $(X_n)_{n \in \mathbb{N}}$ mit $X_n := \{X_n(t) \mid t \in [0, 1]^D\}$, mit $D \in \mathbb{N}$, eine Folge von reellwertigen stochastischen Prozessen und d eine Metrik auf $[0, 1]^D$. Falls für alle $t, t' \in [0, 1]^D$ Konstanten $C, a \in \mathbb{R}^+$ und $b \in (D, \infty)$ und ein $n_0 \in \mathbb{N}$ existieren, so dass für alle $n \geq n_0$:

$$\mathbb{E} \left[|X_n(t) - X_n(t')|^a \right] \leq C(d(t, t'))^b$$

dann ist $(X_n)_{n \in \mathbb{N}}$ gleichmäßig gleichgradig stetig in \mathbb{P} .

Dieses Lemma ist eine Abwandlung des *Stetigkeitssatz von Kolmogorow-Tschenow*, welcher lokale Hölderstetigkeit eines stochastischen Prozess impliziert, siehe z. B. [Kle13], Seite 470. Da wir hier aber eine stärkere Folgerung unter leicht anderen Voraussetzungen haben, werden wir das Lemma noch einmal beweisen. Prinzipiell müssen nur die Fälle $t \neq t'$ betrachtet werden, denn für $t = t'$ ist die Ungleichung trivialerweise für jedes C, a und b erfüllt. Dieser Fall wird daher in den Beweisen von vornherein ignoriert. Die strikte Bedingung $b > D$ ist wesentlich für den Nachweis der Gleichmäßigkeit der Stetigkeit, wie wir gleich im Beweis sehen werden.

Beweis.

Der Beweis beruht auf der Idee, das stetige Intervall $[0, 1]^D$ in eine abzählbare Menge $T \subset [0, 1]^D$ dyadischer Brüche aufzuteilen. Da T dicht in der Menge $[0, 1]^D$ liegt, genügt es die Stetigkeit nur bezüglich T zu zeigen, weil $X_n(t)$ dann für ein beliebiges $t \in [0, 1]^D$ und alle $n \geq n_0$ mit $n_0 \in \mathbb{N}$ durch $\lim_{r_t \rightarrow t} X_n(r_t)$ dargestellt werden kann, wobei $r_t \in T$. Man spricht hier auch von der *stetigen Fortsetzung*. Sei für $m \geq 1$ und $k_1, \dots, k_D = 0, 1, \dots, 2^{m+1} - 1$:

$$A_n^m(k) := \left\{ |X_n((k_1, \dots, k_D) \cdot 2^{-(m+1)}) - X_n((k_1 + 1, \dots, k_D + 1) \cdot 2^{-(m+1)})| \leq q(2^{-m}) \right\}$$

$$A_n^m = \bigcap_{l=m}^{\infty} \bigcap_{k_1, \dots, k_D=0}^{2^{l+1}-1} A_n^l(k)$$

$q : [0, 1] \rightarrow \mathbb{R}$ ist dabei eine monoton wachsende nichtnegative Funktion, so dass

$$\sum_{m=1}^{\infty} q(2^{-m}) < \infty$$

und für alle $n \geq n_0$

$$\sum_{m=1}^{\infty} \sup_{d(t,t') \leq 2^{-m}} \mathbb{P}\left(|X_n(t) - X_n(t')| > q(2^{-m})\right) 2^{D \cdot m} < \infty \quad (2.1)$$

Eine mögliche Funktion wäre $q(x) = x^c$ mit $0 < c < \frac{b-D}{a}$, denn dann gilt mit der geometrischen Reihe

$$\sum_{m=1}^{\infty} q(2^{-m}) = \sum_{m=1}^{\infty} \underbrace{\left(\frac{1}{2^c}\right)^m}_{< 1} < \infty$$

und der *Markov-Ungleichung* bzw. der Voraussetzung aus Lemma 2.20 für alle $n \geq n_0$:

$$\begin{aligned} & \sum_{m=1}^{\infty} \sup_{d(t,t') \leq 2^{-m}} \mathbb{P}\left(|X_n(t) - X_n(t')| > q(2^{-m})\right) 2^{D \cdot m} \\ & \leq \sum_{m=1}^{\infty} \sup_{d(t,t') \leq 2^{-m}} \frac{\mathbb{E}\left[|X_n(t) - X_n(t')|^a\right] 2^{D \cdot m}}{q(2^{-m})^a} \leq \sum_{m=1}^{\infty} \sup_{d(t,t') \leq 2^{-m}} \frac{C(d(t,t'))^b 2^{D \cdot m}}{q(2^{-m})^a} \\ & < \sum_{m=1}^{\infty} \frac{C(2^{-m})^b 2^{D \cdot m}}{2^{-m \cdot c \cdot a}} = C \sum_{m=1}^{\infty} \underbrace{\left(2^{a \cdot c - b + D}\right)^m}_{< 1, \text{ da } c < \frac{b-D}{a}} < \infty \end{aligned}$$

Es ist

$$\begin{aligned} \mathbb{P}((A_n^m)^c) &= \mathbb{P}\left(\bigcup_{l=m}^{\infty} \bigcup_{k_1, \dots, k_D=0}^{2^{l+1}-1} (A_n^l(k))^c\right) \leq \sum_{l=m}^{\infty} \sum_{k_1, \dots, k_D=0}^{2^{l+1}-1} \mathbb{P}((A_n^l(k))^c) \\ &= \sum_{l=m}^{\infty} \sum_{k_1, \dots, k_D=0}^{2^{l+1}-1} \mathbb{P}\left(|X_n((k_1, \dots, k_D) \cdot 2^{-(l+1)}) - X_n((k_1+1, \dots, k_2+1) \cdot 2^{-(l+1)})| > q(2^{-l})\right) \\ &\leq \sum_{l=m}^{\infty} \sum_{k_1, \dots, k_D=0}^{2^{l+1}-1} \sup_{d(z,t') \leq 2^{-l}} \mathbb{P}\left(|X_n(t) - X_n(t')| > g(2^{-l})\right) \\ &= \sum_{l=m}^{\infty} (2^{l+1})^D \sup_{d(t,t') \leq 2^{-l}} \mathbb{P}\left(|X_n(t) - X_n(t')| > g(2^{-l})\right) \\ &= 2^D \sum_{l=m}^{\infty} \sup_{d(t,t') \leq 2^{-l}} \mathbb{P}\left(|X_n(t) - X_n(t')| > g(2^{-l})\right) 2^{D \cdot l} \end{aligned}$$

Da diese Reihe nach Gleichung (2.1) konvergiert, konvergiert der Ausdruck für $m \rightarrow \infty$ gegen 0, wonach auch

$$\mathbb{P}((A_n^m)^c) \xrightarrow{m \rightarrow \infty} 0, \text{ bzw. } \mathbb{P}(A_n^m) \xrightarrow{m \rightarrow \infty} 1$$

gilt. Für alle $\omega \in A_n^m$ und $k_1, \dots, k_D = 0, 1, \dots, 2^{m+1} - 1$ mit $k_* 2^{l-1} \leq r_* \leq (k_* + 1)2^{l-1}$, $r_* \in \mathbb{R}$ gilt

$$|X_n((r_1, \dots, r_D) \cdot 2^{-(m+l)}) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+l)})| \leq \sum_{j=m}^{m+l-1} q(2^{-j})$$

was sich für alle $n \geq n_0$ durch vollständige Induktion über l beweisen lässt. Für $l = 1$ soll für alle $k_1, \dots, k_D = 0, 1, \dots, 2^{m+1} - 1$ und $k_* \leq r_* \leq k_* + 1$ für alle $n \geq n_0$

$$|X_n((r_1, \dots, r_D) \cdot 2^{-(m+1)}) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+1)})| \leq q(2^{-m})$$

gelten, was aber wegen $\omega \in A_n^m \subset A_n^m(k)$ klar ist. Für $l+1$ betrachten wir die Ungleichung

$$\begin{aligned} & |X_n((r_1, \dots, r_D) \cdot 2^{-(m+l+1)}) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+l+1)})| \\ & \leq |X_n((r_1, \dots, r_D) \cdot 2^{-(m+l+1)}) - X_n(\langle r_1/2 \rangle, \dots, \langle r_D/2 \rangle \cdot 2^{-(m+l)})| \\ & \quad + |X_n(\langle r_1/2 \rangle, \dots, \langle r_D/2 \rangle \cdot 2^{-(m+l)}) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+l)})| \end{aligned}$$

wobei $\langle r_*/2 \rangle$ den Integer Part von $r_*/2$ bezeichnet. Der erste Betrag ist wegen $r_* - 2\langle r_*/2 \rangle = 0$ oder 1 und $\omega \in A_n^{m+l}(k)$ kleiner oder gleich $q(2^{-(m+l)})$. Beim zweiten Betrag greift die Induktionsannahme, wonach er kleiner oder gleich $\sum_{j=m}^{m+l-1} q(2^{-j})$ ist. Damit folgt schließlich die Behauptung

$$\begin{aligned} & |X_n((r_1, \dots, r_D) \cdot 2^{-(m+l+1)}) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+l+1)})| \\ & \leq q(2^{-(m+l)}) + \sum_{j=m}^{m+l-1} q(2^{-j}) = \sum_{j=m}^{m+l} q(2^{-j}) \end{aligned}$$

Ein D -Tupel von dyadischen Brüchen $(t_1, \dots, t_D) \in T$, wobei $k_* 2^{-(m+1)} \leq t_* \leq (k_* + 1)2^{-(m+1)}$, ist darstellbar durch $t_* = r_* 2^{-(m+1)}$, wobei $k_* 2^{l-1} \leq r_* \leq (k_* + 1)2^{l-1}$. Es folgt damit für $\omega \in A_n^m$ mit $n \geq n_0$:

$$\sup_{\substack{t \in T \\ k_* 2^{-(m+1)} \leq t_* \leq (k_* + 1)2^{-(m+1)}}} |X_n(t_1, \dots, t_D) - X_n((k_1, \dots, k_D) \cdot 2^{-(m+1)})| \leq \sum_{j=m}^{\infty} q(2^{-j}) \quad (2.2)$$

Es bleibt nun noch zu zeigen, dass für beliebige $\gamma, \eta > 0$ ein $\delta > 0$ und ein $n_0 \in \mathbb{N}$ existiert, so dass für alle $n \geq n_0$:

$$\mathbb{P}\left(\sup_{\substack{t \in T \\ d(t, t') \leq \delta}} |X_n(t) - X_n(t')| > \gamma\right) < \eta$$

Sei dafür $m_0 \in \mathbb{N}$ so gewählt, dass $\sum_{j=m_0}^{\infty} q(2^{-j}) < \frac{\gamma}{3}$ und $\mathbb{P}((A_n^{m_0})^c) < \eta$ und n_0 dasjenige aus der Voraussetzung des Satzes. Für $\delta = 2^{-(m_0+1)}$ folgt mit Gleichung (2.2) auf $A_n^{m_0}$ für alle $n \geq n_0$

$$\sup_{\substack{t \in T \\ d(t, t') \leq \delta}} |X_n(t) - X_n(t')| \leq \gamma$$

oder anders formuliert

$$\begin{aligned} & \left\{ \sup_{\substack{t \in T \\ d(t, t') \leq \delta}} |X_n(t) - X_n(t')| > \gamma \right\} \subset (A_n^{m_0})^c \\ \Rightarrow & \mathbb{P}\left(\sup_{\substack{t \in T \\ d(t, t') \leq \delta}} |X_n(t) - X_n(t')| > \gamma\right) \leq \mathbb{P}((A_n^{m_0})^c) < \eta \end{aligned}$$

für alle $n \geq n_0$.

□

Kommen wir nun zum wichtigsten Satz für die Resultate dieser Arbeit.

2.3.2. Funktionaler zentraler Grenzwertsatz

Wie bereits erwähnt, wird der hier verwendete funktionale zentrale Grenzwertsatz sich etwas von der klassischen Formulierung nach Donsker unterscheiden. Wir brauchen dafür zunächst noch eine weitere Definition.

Definition 2.21. (*Umhüllende*)

Sei $B \subset \mathbb{R}^n$, $n \in \mathbb{N}$. Der Vektor $\Phi \in \mathbb{R}^n$ heißt **Umhüllende** von B , falls für alle $b = (b_1, \dots, b_n) \in B$ die Bedingung $|b_i| \leq \Phi_i$ für alle $i = 1, \dots, n$ erfüllt ist.

Nun können wir den funktionalen zentralen Grenzwertsatz in der für uns am brauchbarsten Gestalt aufsetzen.

Satz 2.22. (Funktionaler zentraler Grenzwertsatz)

Für ein Dreiecksschema $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ von stochastischen Prozessen mit $X_{n,i} := \{X_{n,i}(t) \mid t \in T\}$, wobei $T \subseteq \mathbb{R}^D$, $D \in \mathbb{N}$ und die Prozesse in jeder Zeile stochastisch unabhängig sind, seien für alle $t, t' \in T$:

$$S_n(t) := \sum_{i=1}^{k_n} \left(X_{n,i}(t) - \mathbb{E}[X_{n,i}(t)] \right)$$

$$d_{\Sigma_n}(t, t') := \left(\sum_{i=1}^{k_n} \mathbb{E} \left[|X_{n,i}(t) - X_{n,i}(t')|^2 \right] \right)^{\frac{1}{2}}$$

Falls die folgenden Bedingungen erfüllt sind:

i) $\mathcal{R}(t, t') := \lim_{n \rightarrow \infty} \mathbb{E}[S_n(t)S_n(t')]$ existiert in \mathbb{R} für alle $t, t' \in T$.

ii) Der Grenzwert $d(\cdot, \cdot)$ ist für alle $t, t' \in T$ wohldefiniert durch $d(t, t') = \lim_{n \rightarrow \infty} d_{\Sigma_n}(t, t')$ und für alle deterministischen Folgen $(t_n)_{n \in \mathbb{N}}, (t'_n)_{n \in \mathbb{N}} \subseteq T$ gilt: $\lim_{n \rightarrow \infty} d(t_n, t'_n) = 0 \Rightarrow \lim_{n \rightarrow \infty} d_{\Sigma_n}(t_n, t'_n) = 0$.

iii) $\limsup_{n \rightarrow \infty} \sum_{i=1}^{k_n} \mathbb{E}[\Phi_{n,i}^2] < \infty$, wobei $\Phi_{n,i}$ der i -te Eintrag einer Umhüllenden Φ_n von $(X_{n,i}(t))_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ nach Definition 2.21 ist.

iv) $\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \mathbb{E}[\Phi_{n,i}^2 \mathbf{1}_{\{\Phi_{n,i} > \delta\}}] = 0$ für alle $\delta > 0$.

v) Es gelte $T_{\max} := \max_{t \in T} \|t\|_1 < \infty$ und für alle $\delta \in (0, 1]$ existieren Konstanten $\gamma, C \in \mathbb{R}^+$, sodass für alle $t, t' \in T$, $n \in \mathbb{N}$ und $i = 1, \dots, k_n$ gilt:

$$d^1(t, t') \leq \delta^\gamma \Rightarrow |X_{n,i}(t) - X_{n,i}(t')| \leq C\delta\Phi_{n,i} \quad (2.3)$$

wobei d^1 die Manhattan-Metrik, also die von der Summennorm erzeugte Metrik darstellt.

Dann gilt:

$$S_n \xrightarrow{w} \mathcal{G}(0, \mathcal{R}), \text{ in } (\mathcal{C}(T), \|\cdot\|_\infty)$$

wobei \mathcal{G} ein zentriertes, D -dimensionales Gauß-Feld mit Kovarianzfunktion \mathcal{R} ist.

Beweis.

Der Beweis ist in [Pol90], Seite 53 ff. zu finden. Dort ist das Resultat ein wenig anders formuliert als hier: Aus den Voraussetzungen resultiert zum einen Totalbeschränktheit von T bzgl. d , zum anderen konvergieren die endlichdimensionalen Randverteilungen von S_n

gegen mehrdimensionale Normalverteilungen mit Erwartungswertvektor $\vec{0}$ und Kovarianzmatrizen, welche durch \mathcal{R} festgelegt sind, und somit ihrerseits wiederum das Gauß-Feld eindeutig durch $\mathcal{G}(0, \mathcal{R})$ festlegen; das entspricht im Wesentlichen unserer Ausführung. Zudem unterscheidet sich dort Bedingung v) dahingehend, dass $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ handhabbar sein muss, was durch unsere Bedingung aber impliziert wird, siehe dafür Anhang A.1. \square

Bevor wir uns endlich dem konkreten Modell widmen, verschaffen wir uns zum Abschluss noch einen groben Überblick über einige Techniken des *Bootstrappings*, da auch dies einen wesentlichen Teil der Arbeit darstellen wird.

2.3.3. Bootstrap-Techniken

Bootstrapping ist eine, von [Efr79] erstmals vorgestellte, Methode des Resamplings, um die Verteilung eines Schätzers bei kleiner Inputmenge möglichst gut anzunähern. Sie ging aus Verbesserungen der *Jackknife-Methode* hervor, ein Verfahren zur Schätzung der Verzerrung, also des Bias eines Schätzers. Insbesondere wird Bootstrapping dann verwendet, wenn der Stichprobenumfang verhältnismäßig klein und die theoretische Verteilung der interessierenden Statistik nicht bekannt ist, wobei die tatsächliche Verteilung F durch die empirische Verteilung F_n ersetzt wird. Diese Verteilung ist im Allgemeinen ein zufälliges Maß, welches in aller Regel äußerst schwer handelbar ist. Man kann diese Komplexität einigermaßen umgehen, indem man nicht die empirische Verteilung per se, sondern die aus einer Stichprobe resultierende Verteilung, also die bedingte empirische Verteilung betrachtet. Wenn künftig von „empirischer Verteilung von X_1, \dots, X_n “ die Rede ist, dann ist damit die empirische Verteilung bedingt auf ebendiese Zufallsvariablen gemeint. Für den Anwendungsfall ist das unkritisch, da man hier so oder so mit Stichproben arbeitet. Allerdings benötigen wir für die bevorstehenden Beweise noch etwas mathematisches Vorwissen.

Definition 2.23. (*Bedingte Wahrscheinlichkeit und bedingter Erwartungswert*)

Sei X eine reellwertige Zufallsvariable und $Y := (Y_i)_{i \in \{1, \dots, n\}}$ eine Folge von reellwertigen stochastischen Prozessen mit $Y_i := \{Y_i(t) \mid t \in T\}$. Die bedingte Wahrscheinlichkeit bzw. der bedingte Erwartungswert von X bzgl. Y ist für alle $A \in \sigma(X)$ definiert durch:

$$\mathbb{P}(X \in A \mid Y) := \mathbb{P}(X \in A \mid \sigma(Y)) \quad \text{bzw.} \quad \mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)]$$

wobei $\sigma(X)$ bzw. $\sigma(Y) := \sigma(Y_i, i \in \{1, \dots, n\})$ die Initial- σ -Algebra¹ von X bzw. Y ist. Sowohl die bedingte Wahrscheinlichkeit, als auch der bedingte Erwartungswert, sind von der mathematischen Struktur her ebenfalls Zufallsvariablen, daher können alle Aussagen darüber maximal fast sicher getroffen werden. Für bedingte Wahrscheinlichkeiten bzw. Erwartungswerte schreiben wir künftig abkürzend $\mathbb{P}^*(X \in A) := \mathbb{P}(X \in A \mid Y)$ bzw. $\mathbb{E}^*[X] := \mathbb{E}[X \mid Y]$. Worauf bedingt wird, wird sich bei uns immer aus dem Kontext erschließen. Über die bedingten Erwartungswerte lässt sich zudem - analog zum normalen Fall - die bedingte Varianz Var^* und die bedingte Kovarianz Cov^* definieren.

¹ Die von den Zufallsfunktionen erzeugte (kleinste) σ -Algebra, siehe z. B. [Els13], Seite 126.

Für den bedingten Erwartungswert gelten natürlich alle gängigen Eigenschaften, wie die *Linearität*, *Monotonie*, *Turmeigenschaft*, *iterierte Erwartung* etc. (siehe dafür z. B. [Kle13]). Da wir in den Bootstrapverfahren Eigenschaften der Unabhängigkeit von Zufallsvariablen ausnutzen möchten, jedoch immer mit bedingten Wahrscheinlichkeiten und Erwartungswerten arbeiten müssen, definieren wir die stochastische Unabhängigkeit noch einmal im bedingten Sinne und halten eine weitere wichtige Eigenschaft fest.

Definition 2.24. (*Bedingte stochastische Unabhängigkeit*)

Seien X_1, X_2 reellwertige Zufallsvariablen und $Y := (Y_i)_{i \in \{1, \dots, n\}}$ eine Folge von reellwertigen stochastischen Prozessen mit $Y_i := \{Y_i(t) \mid t \in T\}$. X_1 und X_2 sind genau dann bedingt auf Y stochastisch unabhängig, wenn für alle $A_1 \in \sigma(X_1)$ und $A_2 \in \sigma(X_2)$ gilt:

$$\mathbb{P}^*(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}^*(X_1 \in A_1)\mathbb{P}^*(X_2 \in A_2)$$

Lemma 2.25.

Seien X_1, X_2 reellwertige Zufallsvariablen und $Y := (Y_i)_{i \in \{1, \dots, n\}}$ eine Folge von reellwertigen stochastischen Prozessen mit $Y_i := \{Y_i(t) \mid t \in T\}$. Falls X_1 und X_2 bedingt auf Y stochastisch unabhängig sind, so gilt:

$$\mathbb{E}^*[X_1 X_2] = \mathbb{E}^*[X_1]\mathbb{E}^*[X_2]$$

Beweis.

Siehe [CT97], Seite 234. □

Bemerkung 2.26.

Definition 2.23 und 2.24 und Lemma 2.25 sind für den sehr allgemeinen Fall formuliert, in welchem die Folge Y , auf welche bedingt wird, aus stochastischen Prozessen besteht. Die Formulierungen gelten allesamt genauso, falls Y reellwertige Zufallsvektoren oder Zufallsvariablen sind.

Es gibt eine Vielzahl von Bootstrapverfahren, wie *naiver Bootstrap*, *bayesianischer Bootstrap*, *smooth Bootstrap*, *wild Bootstrap* etc., für mehr Informationen zu diesen Methoden, siehe z. B. [ET93], [Das08] oder [ST12]. In dieser Arbeit spielt nur der naive oder i.i.d. Bootstrap eine Rolle. Hier werden aus der ursprünglichen Stichprobe immer wieder neue Stichproben durch Ziehen mit Zurücklegen erzeugt, mit welchen schließlich die Verteilung des Schätzers von Interesse approximiert wird. [BF81] und [Fre81] lieferten für diesen klassischen Fall bereits früh gute Resultate innerhalb von Regressionsmodellen, indem die zentrierten Residuen gebootstrap werden.

Wie bei allen statistischen Schätzmethoden stellt sich natürlich die Frage, wie man die Qualität der Schätzung am besten überprüfen kann. Beim Bootstrapping hat es sich herauskristallisiert, dass die Konsistenz des Schätzers eine sehr erstrebenswerte Eigenschaft ist. Das bedeutet, der Schätzer soll, je nach starker oder schwacher Konsistenz, fast sicher bzw. in \mathbb{P} gegen den wahren Wert konvergieren. Nun wird beim Bootstrapping wie bereits erwähnt kein Parameter im herkömmlichen Sinne, sondern eine Verteilung geschätzt. Wir brauchen also noch eine sinnvolle Methode, um den „Abstand“ zweier Verteilungen zu messen. Zu diesem Zweck führen wir nun eine Metrik ein, welche in [Mal72] erstmals beschrieben wurde und sich gut dafür eignet.

Definition 2.27. (*Mallows' Distanz*)

Es seien $F, G \in \mathcal{F}_{r,s}$ zwei Verteilungen, wobei $\mathcal{F}_{r,s} = \{F \in \mathcal{F}_{\mathbb{R}^s} : \int \|\vec{x}\|^r dF(\vec{x}) < \infty\}$ die Menge aller Verteilungen auf \mathbb{R}^s mit $s \in \mathbb{N}$, $r > 0$ und $\|\cdot\|$ eine geeignete Metrik ist. Dann ist die Mallows' Distanz dieser beiden Verteilungen definiert durch

$$\rho_r(F, G) := \inf_{\mathcal{F}_{\vec{X}^1, \vec{X}^2}} \left(\mathbb{E}[\|\vec{X}^1 - \vec{X}^2\|^r] \right)^{\frac{1}{r}}$$

wobei $\mathcal{F}_{\vec{X}^1, \vec{X}^2}$ die Menge aller möglichen gemeinsamen Verteilungen des Tupels (\vec{X}^1, \vec{X}^2) mit Randverteilungen $F_{\vec{X}^1} = F$ und $F_{\vec{X}^2} = G$ darstellt. Für zwei Zufallsvariablen \vec{X}^1 und \vec{X}^2 ist die Mallows' Distanz naheliegenderweise über deren Verteilungen $F_{\vec{X}^1}$ und $F_{\vec{X}^2}$ definiert, also

$$\rho_r(\vec{X}^1, \vec{X}^2) := \rho_r(F_{\vec{X}^1}, F_{\vec{X}^2})$$

Für den Fall, dass es sich um, auf weitere Zufallsvariablen $Y := (Y_i)_{i \in \{1, \dots, n\}}$, bedingte Verteilungen handelt, schreiben wir

$$\rho_r^*(F, G) := \inf_{\mathcal{F}_{\vec{X}^1, \vec{X}^2}} \left(\mathbb{E}^*[\|\vec{X}^1 - \vec{X}^2\|^r] \right)^{\frac{1}{r}}$$

Worauf bedingt wird, wird sich wieder aus dem Kontext erschließen. Die bedingte Mallows' Distanz ρ_r^* ist mathematisch betrachtet erneut eine Zufallsvariable, also können auch hier alle Aussagen höchstens fast sicher getroffen werden.

Wir setzen ab hier $r = 2$, was für unsere Zwecke völlig ausreicht. Nun können wir die Konsistenz eines Bootstrap-Schätzers sinnvoll definieren.

Definition 2.28. (*Konsistenz von Bootstrap-Schätzern*)

Sei $X := (X_i)_{i \in \{1, \dots, n\}} \stackrel{\text{iid}}{\sim} F$, $n \in \mathbb{N}$ eine Folge reellwertiger Zufallsvariablen und $T_n := T(X_1, \dots, X_n)$ eine Folge messbarer reellwertiger Funktionen. Sei weiter $X_1^*, \dots, X_n^* \mid X \stackrel{\text{iid}}{\sim} F_n$, wobei F_n die empirische Verteilung von X ist, und $T_n^* := T(X_1^*, \dots, X_n^*)$. T_n^* ist genau dann ein stark bzw. schwach konsistenter Bootstrapschätzer für T_n , wenn

$$\rho_2^*(T_n, T_n^*) \xrightarrow{\text{f.s.}} 0 \quad \text{bzw.} \quad \rho_2^*(T_n, T_n^*) \xrightarrow{\mathbb{P}} 0$$

Dabei ist ρ_2^* die auf X bedingte Mallows' Distanz aus Definition 2.27.

Die Mallows' Distanz ist nicht die einzige Art, mit welcher sich die Konsistenz definieren lässt. [Das08] schlägt bspw. auch die *Kolmogorov Metrik* $\rho(F, G) := \sup_{x \in \mathbb{R}} |F(x) - G(x)|$ vor. Tatsächlich wäre hierfür jede denkbare Metrik zweier Verteilungen möglich. Für die Mallows' Distanz gelten allerdings einige sehr brauchbare Eigenschaften, welche im weiteren Verlauf der Arbeit noch des Öfteren benötigt werden.

Lemma 2.29.

Seien \vec{X}^1, \vec{X}^2 Zufallsvariablen und $(\vec{X}_n^1)_{n \in \mathbb{N}}, (\vec{X}_n^2)_{n \in \mathbb{N}}$ Folgen von Zufallsvariablen mit Werten in \mathbb{R}^s als separablen Banachraum mit äquivalenten p -Normen $\|\cdot\|_p$ und ρ_2 die Mallows' Distanz aus Definition 2.27.

a) Für alle $a \in \mathbb{R}$ gilt:

$$\rho_2(a\vec{X}^1, a\vec{X}^2) \leq |a| \cdot \rho_2(\vec{X}^1, \vec{X}^2)$$

Falls T ein linearer Operator auf \mathbb{R}^s mit zugehöriger Operatornorm $\|\cdot\|$ ist, dann gilt:

$$\rho_2(T\vec{X}^1, T\vec{X}^2) \leq \|T\| \cdot \rho_2(\vec{X}^1, \vec{X}^2)$$

b) Falls $\|\mathbb{E}[\vec{X}^1]\|_p < \infty$ und $\|\mathbb{E}[\vec{X}^2]\|_p < \infty$, dann gilt:

$$\rho_2^2(\vec{X}^1, \vec{X}^2) = \left(\rho_2(\vec{X}^1 - \mathbb{E}[\vec{X}^1], \vec{X}^2 - \mathbb{E}[\vec{X}^2]) \right)^2 + \|\mathbb{E}[\vec{X}^1] - \mathbb{E}[\vec{X}^2]\|_p^2$$

c) Falls $(\vec{X}_n^1)_{n \in \mathbb{N}}$ und $(\vec{X}_n^2)_{n \in \mathbb{N}}$ unabhängig sind und $\mathbb{E}[\vec{X}_n^1] = \mathbb{E}[\vec{X}_n^2]$ für alle $n \in \mathbb{N}$, dann gilt:

$$\rho_2^2\left(\sum_{i=1}^n \vec{X}_i^1, \sum_{i=1}^n \vec{X}_i^2\right) \leq \sum_{i=1}^n \rho_2^2(\vec{X}_i^1, \vec{X}_i^2)$$

d) Falls $X := (\vec{X}_i^1)_{i \in \{1, \dots, n\}} \stackrel{\text{iid}}{\sim} F$, $n \in \mathbb{N}$ und $\vec{X}_1^{1*}, \dots, \vec{X}_n^{1*} \mid X \stackrel{\text{iid}}{\sim} F_n$ für alle $n \in \mathbb{N}$, wobei F_n die empirische Verteilung von $\vec{X}_1^1, \dots, \vec{X}_n^1$ ist, dann gilt:

$$\rho_2^*(\vec{X}_1^1, \vec{X}_n^{1*}) \xrightarrow{\text{f.s.}} 0$$

e) $\rho_2(\vec{X}^1, \vec{X}_n^1) \rightarrow 0$ gilt genau dann, wenn $\vec{X}_n^1 \xrightarrow{D} \vec{X}^1$ und $\mathbb{E}[\|\vec{X}_n^1\|_p^2] \rightarrow \mathbb{E}[\|\vec{X}^1\|_p^2]$.

f) Falls $\vec{X}_n^1 \xrightarrow{D} \vec{X}^1$ und $\rho_2(\vec{X}_n^1, \vec{X}_n^2) \rightarrow 0$, dann gilt $\vec{X}_n^2 \xrightarrow{D} \vec{X}^1$.

Alle Aussagen gelten f. s. für die bedingte Mallows' Distanz ρ_2^* .

Beweis.

Wir beweisen an dieser Stelle nur f), die Beweise für a) - e) sind in [BF81] bzw [ST12] nachzulesen.

Wir schreiben zunächst die Mallows' Distanz aus Definition 2.27 ein wenig um. Nach [BF81] gibt es unter den gegebenen Voraussetzungen ein W-Maß π auf $\mathbb{R}^s \times \mathbb{R}^s$, für welches $\pi \circ (\vec{X}^1)^{-1} = F_{\vec{X}^1}$, $\pi \circ (\vec{X}^2)^{-1} = F_{\vec{X}^2}$ und zudem $\int \|\vec{X}^1 - \vec{X}^2\|^2 d\pi$ minimal wird.

Wegen der Monotonie der Wurzelfunktion lässt sich die Mallows' Distanz von \vec{X}^1 und \vec{X}^2 demnach darstellen durch

$$\begin{aligned} \rho_2(\vec{X}^1, \vec{X}^2) &= \inf_{\mathcal{T}_{\vec{X}^1, \vec{X}^2}} \left(\mathbb{E}[\|\vec{X}^1 - \vec{X}^2\|^2] \right)^{\frac{1}{2}} = \inf_{\mathcal{T}_{\vec{X}^1, \vec{X}^2}} \left(\int \|\vec{X}^1 - \vec{X}^2\|^2 dF_{\vec{X}^1, \vec{X}^2} \right)^{\frac{1}{2}} \\ &= \left(\int \|\vec{X}^1 - \vec{X}^2\|^2 d\pi \right)^{\frac{1}{2}} =: \left(\mathbb{E}_{\pi}[\|\vec{X}^1 - \vec{X}^2\|^2] \right)^{\frac{1}{2}} \end{aligned}$$

Nach dem *Portmanteau-Theorem* gilt $\vec{X}_n^2 \xrightarrow{D} \vec{X}^1$ genau dann, wenn für alle beschränkten und lipschitzstetigen Funktionen f

$$\lim_{n \rightarrow \infty} \left| \int f dF_{\vec{X}_n^2} - \int f dF_{\vec{X}^1} \right| = 0$$

Seien nun π und $(\pi_n)_{n \in \mathbb{N}}$ Maße, für welche $\pi \circ (\vec{X}^1)^{-1} = F_{\vec{X}^1}$, $\pi_n \circ (\vec{X}_n^1)^{-1} = F_{\vec{X}_n^1}$, $\pi \circ (\vec{X}^2)^{-1} = F_{\vec{X}^2}$ und $\pi_n \circ (\vec{X}_n^2)^{-1} = F_{\vec{X}_n^2}$. Dann folgt mit dem *Transformationssatz für Maße* (*) und der *Jensenschen* (**) bzw. der *Cauchy-Schwarzschen Ungleichung* (***):

$$\begin{aligned} \left| \int f dF_{\vec{X}_n^2} - \int f dF_{\vec{X}^1} \right| &= \left| \int f dF_{\vec{X}_n^2} - \int f dF_{\vec{X}_n^1} + \int f dF_{\vec{X}_n^1} - \int f dF_{\vec{X}^1} \right| \\ &\stackrel{\Delta\text{-Ungl.}}{\leq} \left| \int f dF_{\vec{X}_n^2} - \int f dF_{\vec{X}_n^1} \right| + \underbrace{\left| \int f dF_{\vec{X}_n^1} - \int f dF_{\vec{X}^1} \right|}_{= o(1), \text{ da } \vec{X}_n^1 \xrightarrow{D} \vec{X}^1} \\ &= \left| \int f d(\pi_n \circ (\vec{X}_n^2)^{-1}) - \int f d(\pi_n \circ (\vec{X}_n^1)^{-1}) \right| + o(1) \\ &\stackrel{(*)}{=} \left| \int f(\vec{X}_n^2) d\pi_n - \int f(\vec{X}_n^1) d\pi_n \right| + o(1) = \left| \int f(\vec{X}_n^2) - f(\vec{X}_n^1) d\pi_n \right| + o(1) \\ &= \left| \mathbb{E}_{\pi_n} [f(\vec{X}_n^2) - f(\vec{X}_n^1)] \right| + o(1) \stackrel{(**)}{\leq} \mathbb{E}_{\pi_n} [|f(\vec{X}_n^2) - f(\vec{X}_n^1)|] + o(1) \\ &\stackrel{f \text{ lipschitz}}{\leq} L_f \cdot \mathbb{E}_{\pi_n} [|\vec{X}_n^2 - \vec{X}_n^1|] + o(1) \stackrel{(***)}{\leq} L_f \cdot \left(\mathbb{E}_{\pi_n} [|\vec{X}_n^2 - \vec{X}_n^1|^2] \right)^{\frac{1}{2}} + o(1) \\ &= L_f \cdot \underbrace{\rho_2(\vec{X}_n^2, \vec{X}_n^1)}_{\rightarrow 0} + \underbrace{o(1)}_{\rightarrow 0} \rightarrow 0 \end{aligned}$$

Die Beweise können alle analog für, auf beliebige feste Realisierungen $Y(\omega) = y$ bedingte Verteilungen durchgeführt werden, danach gelten sie auch f. s. für ρ_2^* . \square

3. Das Modell und die Schätzer

Wir lernen nun das grundlegende Modell bzw. dessen Regressionsgleichung kennen und gehen kurz auf dessen Eigenschaften und die zu schätzenden Komponenten ein. Anschließend definieren wir die jeweiligen Schätzer in einer relativ allgemeinen Form und legen Annahmen für das Modell, wie auch die Schätzer selbst fest, die im weiteren Verlauf der Arbeit benötigt werden, um die gewünschten Eigenschaften der Schätzer zu erzielen. Das Modell und ein Teil der Annahmen entstammt [Deg11].

3.1. Die Modellgleichung und deren Eigenschaften

Wir betrachten eine Stichprobe der Größe n , von der jede Einheit p verschiedene Messpunkte miteinbezieht. Es gibt also insgesamt $p \cdot n$ Messpaare (Y_{ij}, x_j) mit $1 \leq i \leq n$ und $1 \leq j \leq p$. Dabei ist Y_{ij} der Wert der Zielvariable der i -ten Stichprobeneinheit, beobachtet an einer festen Stelle x_j . x_j entspricht der Kovariable, welche wiederum dynamisch von der Stichprobengröße und der Anzahl der Messpunkte abhängen soll, d. h. es gilt $x_j := x_j(n, p)$. Im Verlauf der Arbeit werden meistens asymptotische Aussagen getroffen, und zwar immer für den Fall $n \rightarrow \infty$. Zur Abkürzung werden wir künftig darauf verzichten, das jedes Mal explizit zu erwähnen, d. h. für jede von n abhängige Folge $a(n)$ gilt $a(n) \rightarrow a \Leftrightarrow \lim_{n \rightarrow \infty} a(n) = a$. Zudem soll o. B. d. A. $x_j \in [0, 1]$ gelten, weil gerade für dieses Intervall viele hilfreiche Aussagen und Sätze existieren und sich jedes andere Intervall $[a, b]$, $a, b \in \mathbb{R}$ leicht zu $[0, 1]$ transformieren lässt. Für die Modellgleichung gilt:

$$Y_{ij} = \mu(x_j) + Z_i(x_j) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p \quad (3.1)$$

$\mu : \mathbb{R} \rightarrow \mathbb{R}$ ist eine unbekannte, aber hinreichend glatte Funktion, die Z_i 's sind unabhängige Kopien eines Zufallsprozesses $Z = \{Z(x) \mid x \in [0, 1]\}$, welcher zentriert ist und eine Kovarianzfunktion R besitzt, und die ε_{ij} 's sind zentrierte reellwertige Zufallsvariablen. μ stellt die systematische Komponente bzw. den Trend dar. Z kann als individuelle (zufällige funktionale) Abweichung des Trends interpretiert werden. ε stellt wiederum den Messfehler selbst dar, welcher bei jeder Datenerhebung in der Praxis unvermeidbar und nicht beeinflussbar ist.

Es handelt sich also um diskrete, fehlerbehaftete Beobachtungen, deren zugrunde liegende Struktur jedoch aus dem Banachraum $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ kommt. Durch eine Erhöhung der Anzahl der Messpunkte p , im Grenzfall unendlich viele, liegen die Beobachtungen schließlich dicht im kontinuierlichen Intervall $[0, 1]$, wodurch der funktionale Charakter der Daten wiederhergestellt bzw. ausreichend gut angenähert wird. Die Anzahl der Messpunkte soll dynamisch von der Stichprobengröße abhängen, d. h. es gilt $p := p(n)$, und insbesondere bei wachsender Stichprobengröße ebenfalls ansteigen, also $p \rightarrow \infty$.

Welche Komponenten sind für eine Schätzung nun überhaupt sinnvoll? Logischerweise interessiert man sich für die Trendfunktion, welche im Prinzip dem Erwartungswert des Modells entspricht, also der Funktion, welcher die Beobachtungsdaten im Mittel folgen. Daher benötigen wir einen Schätzer $\hat{\mu}$ für μ (der funktionale Schätzer des Trends). Weil die Beobachtungen von Grund auf durch die Funktion Z in jedem noch so kleinen Punkt zufällig beeinflusst werden, ist auch der Einflussfaktor ebendieser Funktion von großem Interesse. Dieser entspricht gerade der Kovarianzfunktion von Z , wodurch ein weiterer Schätzer \hat{R} für R relevant wird (der funktionale Schätzer der Kovarianz).

ε ist ein Messfehler, welcher im besten Fall gar nicht auftauchen sollte, was aber wie bereits erwähnt in der Realität nicht funktioniert. Die Schätzer sollten also so konzipiert sein, dass dieser Messfehler möglichst gut ausgeglichen wird, und im Grenzfall einer unendlich großen Stichprobe bzw. unendlich vielen Messpunkten gar nicht mehr auftaucht.

3.2. Die allgemeinen Schätzer

Wir formulieren zunächst einen allgemeinen Schätzer für μ , und bauen auf diesem einen weiteren Schätzer für R auf.

3.2.1. Trendschätzer $\hat{\mu}$

$$\begin{aligned}\hat{\mu}_i(x) &= \sum_{j=1}^p W_j(x) Y_{ij} \\ \Rightarrow \hat{\mu}(x) &= \frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij}\end{aligned}\tag{3.2}$$

Y_{ij} sind die beobachteten Daten aus Abschnitt 3.1. $W_j(x)$ sind klassische Gewichtsfunktionen, wie sie in zahlreichen anderen Glättungs- und Schätzverfahren auftreten, bspw. basierend auf *Kernfunktionen* und darauf aufbauender *Kerndichteschätzung*, siehe z. B. [Nad65]. Prinzipiell ist man bei deren Wahl ziemlich frei, wichtig ist, dass die Beobachtungspunkte x_j enthalten sind, und zudem eine Bandbreite h , welche festlegt, wie stark die jeweiligen Beobachtungspunkte x_j bei einer Schätzung an einer konkreten Stelle $x \in [0, 1]$ zu Buche schlagen. Die Bandbreite soll dynamisch von der Stichprobengröße

und der Anzahl der Messpunkte abhängen, d. h. es gilt $h := h(n, p)$, und bei wachsendem n und p abfallen, also $h \rightarrow 0$. An die Gewichtsfunktionen werden später auch noch weitere Annahmen gestellt. Da für Y_{ij} innerhalb jeder Stichprobeneinheit an den gleichen Punkten x_1, \dots, x_p beobachtet wird, kann Gleichung (3.2) noch wie folgt umgeformt werden.

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij} = \sum_{j=1}^p W_j(x) \frac{1}{n} \sum_{i=1}^n Y_{ij} = \sum_{j=1}^p W_j(x) \bar{Y}_j \quad (3.3)$$

\bar{Y}_j ist das arithmetische Mittel von Y an der Beobachtungsstelle x_j . $\hat{\mu}$ ist also insbesondere ein **linearer Schätzer** oder auch **linearer Glätter**.

Kommen wir nun zum Schätzer für die Kovarianzfunktion R der Zufallsfunktion Z .

3.2.2. Kovarianzschätzer \hat{R}

$$\hat{R}(s, t) = \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) \quad (3.4)$$

Die Umformung ist durch den *Verschiebungssatz* möglich. Der Schätzer ist im Prinzip nichts anderes, als die klassisch berechnete *Stichprobenkovarianz*, nur dass anstatt der nackten Stichprobendaten Y die Trendschätzer, also mehr oder weniger die geglätteten Versionen, an den Stellen s und t verwendet werden. Prinzipiell wäre es auch möglich, zuerst die Kovarianz mit Y zu berechnen und danach eine Linearkombination mit zweidimensionalen Gewichtsfunktionen $W_{jk}(s, t)$ zu bilden. Das Problem hierbei ist jedoch, dass durch das Produkt im Schätzer Mischterme der Zufallsfunktion Z und des Messfehlers ε entstehen, welche sich mit zweidimensionalen Gewichten nicht mehr ausglätten lassen. Man hat in diesem Fall also immer eine (zu) starke Verzerrung bzgl. des Messfehlers. Im Fall multiplikativ trennbarer Gewichtsfunktionen - also falls $W_{jk}(s, t)$ durch $W_j(s) \cdot W_k(t)$ darstellbar ist - würde der Schätzer wieder in der ursprünglichen Version aus Gleichung (3.4) resultieren.

Zudem weist dieser Schätzer den Verzerrungsfaktor $\frac{1}{n}$ auf, was man aber leicht durch die korrigierte Version

$$\hat{R}(s, t) = \frac{1}{n-1} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) \quad (3.5)$$

ausmerzen kann. Da ansonsten aber alle Rechnungen äquivalent funktionieren, und der Vorfaktor $\frac{1}{n-1}$ manche Rechenschritte zum Teil deutlich in die Länge ziehen würde, arbeiten wir in der gesamten Arbeit mit dem Schätzer aus Gleichung (3.4) und werden nur an manchen Stellen auf das korrigierte Pendant hinweisen.

Kommen wir nun zu einigen weiteren Annahmen, welche für Kernaussagen dieser Arbeit erforderlich sind.

3.3. Annahmen

Wir benötigen sowohl Annahmen für das Modell bzw. dessen Komponenten aus Abschnitt 3.1, als auch die Schätzer selbst aus Abschnitt 3.2. In beiden Fällen sollen diese dazu dienen, eine möglichst optimale Glättung zu erzielen. Wir werden gleich etwas genauer auf die jeweiligen Annahmen eingehen.

3.3.1. Annahmen an das Modell

Annahme 3.1.

- (A.1) μ hat auf $[0, 1]$ hölderstetige Ableitungen zum Exponenten $\alpha_\mu \in (0, 1]$ mit Konstante $H_\mu > 0$, schreibe $\mu \in \mathcal{C}^{1, \alpha_\mu}([0, 1])$.
- (A.2) Für alle $x, x' \in [0, 1]$ sei $|Z(x) - Z(x')| \leq K|x - x'|^\beta$ f. s., wobei K eine Zufallsvariable mit $\mathbb{E}[K^4] < \infty$ und $\beta \in (0, 1]$ eine Konstante ist.
- (A.3) Der Zufallsprozess $Z = \{Z(x) : x \in [0, 1]\}$ ist zentriert (also $\mathbb{E}[Z(x)] = 0$ für alle x) und hat die Kovarianzfunktion R und das vierte Verbundmoment M mit den Abbildungsvorschriften:

$$R : \begin{cases} [0, 1]^2 \rightarrow \mathbb{R} \\ (s, t) \mapsto \mathbb{E}[Z(s)Z(t)] \end{cases}, \quad M : \begin{cases} [0, 1]^4 \rightarrow \mathbb{R} \\ (s, t, s', t') \mapsto \mathbb{E}[Z(s)Z(t)Z(s')Z(t')] \end{cases}$$

Diese haben auf $[0, 1]^2$ bzw. $[0, 1]^4$ jeweils hölderstetige partielle Ableitungen zu den Exponenten $\alpha_R, \alpha_M \in (0, 1]$ mit Konstanten $H_R, H_M \in \mathbb{R}^+$, schreibe $R \in \mathcal{C}^{1, \alpha_R}([0, 1]^2)$ und $M \in \mathcal{C}^{1, \alpha_M}([0, 1]^4)$. Weil es sich bei den Z_i 's nach Modellvoraussetzung um unabhängige Kopien von Z handelt, sind diese bezüglich i auch unabhängig identisch verteilt.

- (A.4) Die Zufallsvektoren $\vec{\varepsilon}_i := (\varepsilon_{i1}, \dots, \varepsilon_{ip})^T$, $i = 1, \dots, n$ sind bezüglich i unabhängig identisch verteilt, unabhängig von Z (und damit natürlich auch unabhängig von allen Z_i 's) und zentriert (also $\mathbb{E}[\vec{\varepsilon}_1] = \vec{0}$). Seien zudem $\lambda_{\max,1}$ bzw. $\lambda_{\max,2}$ die maximalen Eigenwerte der Kovarianzmatrizen $\text{Cov}(\vec{\varepsilon}_1) \in \mathbb{R}^{p \times p}$ bzw. $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) \in \mathbb{R}^{p^2 \times p^2}$. Dann soll $\lambda_{\max,1} = \mathcal{O}(p^{d_1})$, $d_1 \in [0, 1]$ und $\lambda_{\max,2} = \mathcal{O}(p^{d_2})$, $d_2 \in [0, 2]$ gelten, d. h. es sollen endliche Konstanten $C_1, C_2 \in \mathbb{R}^+$ und $p_1, p_2 \in \mathbb{N}$ existieren, so dass $\lambda_{\max,1} \leq C_1 p^{d_1}$ für alle $p \geq p_1$ und $\lambda_{\max,2} \leq C_2 p^{d_2}$ für alle $p \geq p_2$. Wir bezeichnen im weiteren Verlauf zur Vereinfachung die Einträge von $\text{Cov}(\vec{\varepsilon}_1)$ größtenteils mit σ_{jk} , $j, k = 1, \dots, p$.
- (A.5) Die Punkte x_j , $j = 1, \dots, p$ werden durch die Vorschrift $\int_0^{x_j} f(t)dt = \frac{j - 0.5}{p}$ erzeugt, wobei $f : [0, 1] \rightarrow \mathbb{R}^+$ eine lipschitzstetige positive Lebesgue-Dichte ist.

Bemerkung 3.2.

(A.1) - (A.3) sind Glattheitsannahmen - insbesondere (A.2) eine Art stochastische Hölderstetigkeit - und stellen sicher, dass sowohl die Trendfunktion μ , als auch deren individuelle Abweichung Z einfach zu handhaben sind. Sie erlauben viele nützliche Abschätzungen, bspw. mittels Taylorpolynome. Man kann theoretisch auch annehmen, dass μ , R und M hölderstetige (partielle) Ableitungen bis zur Ordnung m haben, was zu schnelleren Konvergenzraten führen würde, wie wir später sehen werden. (A.4) zwingt den Messfehler ε dazu, nicht zu viel Einfluss auf den Schätzer zu haben, so dass er bzgl. der später ermittelten asymptotischen Normalität ausgeglistet werden kann. Tatsächlich schließt (A.3) dabei nicht explizit aus, dass die Kovarianzfunktion bzw. das vierte Verbundmoment von Z an den betrachteten Datenpunkten die gleichen Eigenschaften wie der Fehler aus (A.3) erfüllt. Das ist aber eher ein pathologisches Beispiel, weil man dann im weiteren Verlauf die Zufallsfunktion äquivalent zum Fehler behandeln könnte, und der standardisierte Grenzprozess eine entartete Zufallsfunktion $f(x) \equiv 0$ ergeben würde; für unsere Zwecke ist eine derartige Zufallsfunktion Z also komplett irrelevant. (A.5) sorgt schließlich dafür, dass x_1, \dots, x_p äquidistant sind und ein gleichmäßiges Gitter auf $[0, 1]$ darstellen. Zudem lassen sich mit dieser Annahme über die Extremwerte der Dichte Abschätzung über die Lage bzw. die Abstände der Punkte durchführen.

Aus (A.3) und (A.4) folgt eine weitere wichtige Eigenschaft für Z und ε , welche wir in folgender Proposition festhalten.

Proposition 3.3.

Sei Annahme 3.1 erfüllt und $m \in \{1, \dots, 4\}$. Dann existieren für alle $x_1, \dots, x_m \in [0, 1]$ und alle $j_1, \dots, j_m \in \{1, \dots, p\}$, $p \in \mathbb{N}$ Konstanten $E_{Z,1}, \dots, E_{Z,m}, E_{\varepsilon,1}(p), \dots, E_{\varepsilon,m}(p) \in \mathbb{R}^+$, für welche folgende Ungleichungen gelten:

$$\mathbb{E} \left[\left| \prod_{k=1}^m Z(x_k) \right| \right] \leq E_{Z,m}, \quad \mathbb{E} \left[\left| \prod_{k=1}^m \varepsilon_{1j_k} \right| \right] \leq E_{\varepsilon,m}(p)$$

Durch die Jensensche Ungleichung gelten diese Schranken automatisch auch, falls der Betrag außerhalb des Erwartungswertes steht.

Die Verbundmomente von Z und ε lassen sich also bis zum Grad 4 durch universelle Konstanten beschränken. Für ε hängt diese Konstante noch von der Anzahl der Messpunkte p ab, was aber weiter kein Problem darstellt, da die Schranken für alle $p \in \mathbb{N}$ existieren.

Beweis.

Die Schranken für ε lassen sich relativ schnell über (A.4) herleiten.

Logischerweise gelten für alle $j_1, j_2, j_3, j_4 = 1, \dots, p$ die vier folgenden Ungleichungen:

$$\begin{aligned}\mathbb{E}[|\varepsilon_{1j_1}|] &\leq \max_{j_1 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1}|] \\ \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2}|] &\leq \max_{j_1, j_2 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2}|] \\ \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3}|] &\leq \max_{j_1, j_2, j_3 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3}|] \\ \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3} \varepsilon_{1j_4}|] &\leq \max_{j_1, j_2, j_3, j_4 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3} \varepsilon_{1j_4}|]\end{aligned}$$

Wegen $\text{Cov}(\vec{\varepsilon}_1) \in \mathbb{R}^{p \times p}$ bzw. $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) \in \mathbb{R}^{p^2 \times p^2}$, also der vorausgesetzten Existenz der Kovarianzmatrizen, müssen für alle $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$ die Erwartungswerte von ε_{1j_1} , $\varepsilon_{1j_1} \varepsilon_{1j_2}$ und $\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3} \varepsilon_{1j_4}$ existieren. Wir können daher sofort drei der vier Schranken festlegen:

$$\begin{aligned}E_{\varepsilon,1}(p) &:= \max_{j_1 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1}|] < \infty, \quad E_{\varepsilon,2}(p) := \max_{j_1, j_2 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2}|] < \infty \\ E_{\varepsilon,4}(p) &:= \max_{j_1, j_2, j_3, j_4 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3} \varepsilon_{1j_4}|] < \infty\end{aligned}$$

Für die vierte verwenden wir die *Cauchy-Schwarz-Ungleichung*, die für alle $j_1, j_2, j_3 = 1, \dots, p$ gleichermaßen gilt:

$$\mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3}|] \leq \left(\mathbb{E}[(\varepsilon_{1j_1} \varepsilon_{1j_2})^2] \right)^{\frac{1}{2}} \left(\mathbb{E}[(\varepsilon_{1j_3})^2] \right)^{\frac{1}{2}} = \left(\mathbb{E}[\varepsilon_{1j_1} \varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_2}] \right)^{\frac{1}{2}} \left(\mathbb{E}[\varepsilon_{1j_3} \varepsilon_{1j_3}] \right)^{\frac{1}{2}}$$

Da die Wurzelfunktion monoton wachsend und zudem ≥ 0 ist, folgt:

$$\begin{aligned}\max_{j_1, j_2, j_3 \in \{1, \dots, p\}} \mathbb{E}[|\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3}|] &\leq \left(\max_{j_1, j_2 \in \{1, \dots, p\}} \mathbb{E}[\varepsilon_{1j_1} \varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_2}] \right)^{\frac{1}{2}} \left(\max_{j_3 \in \{1, \dots, p\}} \mathbb{E}[\varepsilon_{1j_3} \varepsilon_{1j_3}] \right)^{\frac{1}{2}} \\ &\leq \underbrace{(E_{\varepsilon,4}(p))^{\frac{1}{2}}}_{< \infty} \underbrace{(E_{\varepsilon,2}(p))^{\frac{1}{2}}}_{< \infty} =: E_{\varepsilon,3}(p) < \infty\end{aligned}$$

Für Z verwenden wir die folgenden Ungleichungen, die für alle $x_1, x_2, x_3, x_4 \in [0, 1]$ gelten:

$$\begin{aligned}\mathbb{E}[|Z(x_1)|] &\leq \sup_{x_1 \in [0,1]} \mathbb{E}[|Z(x_1)|] \\ \mathbb{E}[|Z(x_1)Z(x_2)|] &\leq \sup_{x_1, x_2 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)|] \\ \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)|] &\leq \sup_{x_1, x_2, x_3 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)|] \\ \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)Z(x_4)|] &\leq \sup_{x_1, x_2, x_3, x_4 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)Z(x_4)|]\end{aligned}$$

Nach (A.3) gilt $R \in \mathcal{C}^{1,\alpha_R}([0, 1]^2)$ und $M \in \mathcal{C}^{1,\alpha_M}([0, 1]^4)$, R bzw. M sind also auf den kompakten Gebieten $[0, 1]^2$ bzw. $[0, 1]^4$ stetig partiell differenzierbar. Das impliziert über den *Schrankensatz* für beide Funktionen die Lipschitzstetigkeit und damit insbesondere die Stetigkeit. Aufgrund der Stetigkeit der Funktion $f(x) = x^r$, $r \in \mathbb{R}$, sind auch R^r bzw. M^r stetig. Mit dem *Satz vom Minimum und Maximum* existiert dadurch

$$R_{\max}^r := \max_{x \in [0,1]} (R(x, x))^r \stackrel{f \text{ mon. w.}}{=} \left(\max_{x \in [0,1]} R(x, x) \right)^r$$

$$M_{\max}^r := \max_{x \in [0,1]} (M(x, x, x, x))^r \stackrel{f \text{ mon. w.}}{=} \left(\max_{x \in [0,1]} M(x, x, x, x) \right)^r$$

und zudem entspricht in diesem Fall die Supremumsnorm der Maximumsnorm. Aufgrund der Definitionen von R und M gilt $R(x, x) \geq 0$ und $M(x, x, x, x) \geq 0$ für alle $x \in [0, 1]$, R_{\max}^r und M_{\max}^r sind daher für alle $r \in \mathbb{R}$ wohldefiniert. Mit der *Cauchy-Schwarz-Ungleichung*, die für alle $x_1, x_2, x_3, x_4 \in [0, 1]$ gleichermaßen anwendbar ist, gilt nun:

$$\begin{aligned} \sup_{x_1 \in [0,1]} \mathbb{E}[|Z(x_1)|] &\leq \sup_{x_1 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^2] \right)^{\frac{1}{2}} = \sup_{x_1 \in [0,1]} \left(R(x_1, x_1) \right)^{\frac{1}{2}} = \max_{x_1 \in [0,1]} \left(R(x_1, x_1) \right)^{\frac{1}{2}} \\ &= \underbrace{R_{\max}^{\frac{1}{2}}}_{< \infty} =: E_{Z,1} \end{aligned}$$

$$\begin{aligned} \sup_{x_1, x_2 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)|] &\leq \sup_{x_1, x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^2] \right)^{\frac{1}{2}} \left(\mathbb{E}[(Z(x_2))^2] \right)^{\frac{1}{2}} \\ &= \left(\sup_{x_1 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^2] \right)^{\frac{1}{2}} \right) \left(\sup_{x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_2))^2] \right)^{\frac{1}{2}} \right) = \underbrace{R_{\max}}_{< \infty} =: E_{Z,2} \end{aligned}$$

$$\begin{aligned} \sup_{x_1, x_2, x_3 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)|] &\leq \sup_{x_1, x_2, x_3 \in [0,1]} \left(\mathbb{E}[(Z(x_1)Z(x_2))^2] \right)^{\frac{1}{2}} \left(\mathbb{E}[(Z(x_3))^2] \right)^{\frac{1}{2}} \\ &= \left(\sup_{x_1, x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^2(Z(x_2))^2] \right)^{\frac{1}{2}} \right) \left(\sup_{x_3 \in [0,1]} \left(\mathbb{E}[(Z(x_3))^2] \right)^{\frac{1}{2}} \right) \\ &\leq \left(\sup_{x_1, x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^4] \mathbb{E}[(Z(x_2))^4] \right)^{\frac{1}{4}} \right) R_{\max}^{\frac{1}{2}} \\ &= \left(\sup_{x_1 \in [0,1]} \left(\mathbb{E}[(Z(x_1))^4] \right)^{\frac{1}{4}} \right) \left(\sup_{x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_2))^4] \right)^{\frac{1}{4}} \right) R_{\max}^{\frac{1}{2}} \\ &= \left(\sup_{x_1 \in [0,1]} \left(M(x_1, x_1, x_1, x_1) \right)^{\frac{1}{4}} \right) \left(\sup_{x_2 \in [0,1]} \left(M(x_2, x_2, x_2, x_2) \right)^{\frac{1}{4}} \right) R_{\max}^{\frac{1}{2}} \\ &= \left(\max_{x_1 \in [0,1]} \left(M(x_1, x_1, x_1, x_1) \right)^{\frac{1}{4}} \right) \left(\max_{x_2 \in [0,1]} \left(M(x_2, x_2, x_2, x_2) \right)^{\frac{1}{4}} \right) R_{\max}^{\frac{1}{2}} \\ &= \underbrace{M_{\max}^{\frac{1}{2}}}_{< \infty} \underbrace{R_{\max}^{\frac{1}{2}}}_{< \infty} =: E_{Z,3} \end{aligned}$$

$$\begin{aligned}
\sup_{x_1, x_2, x_3, x_4 \in [0,1]} \mathbb{E}[|Z(x_1)Z(x_2)Z(x_3)Z(x_4)|] &\leq \sup_{x_1, x_2, x_3, x_4 \in [0,1]} \left(\mathbb{E}[(Z(x_1)Z(x_2))^2] \mathbb{E}[(Z(x_3)Z(x_4))^2] \right)^{\frac{1}{2}} \\
&= \left(\sup_{x_1, x_2 \in [0,1]} \left(\mathbb{E}[(Z(x_1)Z(x_2))^2] \right)^{\frac{1}{2}} \right) \\
&\quad \cdot \left(\sup_{x_3, x_4 \in [0,1]} \left(\mathbb{E}[(Z(x_3)Z(x_4))^2] \right)^{\frac{1}{2}} \right) \leq \underbrace{M_{\max}}_{< \infty} =: E_{Z,4}
\end{aligned}$$

□

3.3.2. Annahmen an die Schätzer

Da es sich um lineare Schätzer handelt, die mit Gewichtsfunktionen arbeiten, betreffen die diesbezüglichen Annahmen ausschließlich ebendiese Gewichte. Sie sollen für alle $x, x' \in [0, 1]$ und $j = 1, \dots, p$, $p \in \mathbb{N}$ gelten.

Annahme 3.4.

$$(A.6) \quad |W_j(x)| \leq \frac{C_W}{ph} \text{ mit } C_W \in \mathbb{R}^+.$$

$$(A.7) \quad |W_j(x) - W_j(x')| \leq \frac{C_{W_-}|x - x'|}{ph^2} \text{ mit } C_{W_-} \in \mathbb{R}^+.$$

$$(A.8) \quad \text{Falls } |x - x_j| > h, \text{ dann soll } W_j(x) = 0 \text{ gelten.}$$

$$(A.9) \quad \sum_{j=1}^p W_j(x) = 1 \text{ und } \sum_{j=1}^p W_j(x)(x - x_j) = 0.$$

Bemerkung 3.5.

(A.6) gewährleistet die Beschränktheit der Gewichte bzgl. der Anzahl der Messpunkte p und der Bandbreite h . (A.7) ist eine Art Lipschitzstetigkeit, ebenfalls bzgl. p und h . (A.8) ist logisch, da h den Bereich der Punkte festlegt, welche für einen bestimmten Punkt x eine Rolle spielen; alle Messpunkte, welche zu weit von x entfernt sind, sollen keinen Einfluss mehr haben, daher wird ihr Gewicht auf 0 gesetzt. (A.9) impliziert, dass der Schätzer Polynome vom Grad 1 reproduziert. Auch diese Annahme kann auf Polynome vom Grad m erweitert werden, was in Kombination mit den erweiterten Hölderstetigkeiten der Ableitungen von μ , R und M , siehe Bemerkung 3.2, zu schnelleren Konvergenzraten führt.

Auch aus diesen Annahmen folgen unmittelbar nützliche Eigenschaften für die Gewichte bzw. für deren Anwendung, welche wir in drei weiteren Propositionen festhalten.

Proposition 3.6.

Für die Gewichte und Messpunkte seien Annahme 3.1 und 3.4 erfüllt. Dann gilt:

- a) Für alle $x \in [0, 1]$ existiert eine Konstante $C_{W_0} \in \mathbb{R}^+$, sodass die maximale Anzahl der von 0 verschiedenen Gewichte $W_1(x), \dots, W_p(x)$ ab einem hinreichend großen p höchstens $C_{W_0}ph$ beträgt.
- b) Für alle $x \in [0, 1]$ und $r \in \mathbb{R}$ gilt:

$$\sum_{j=1}^p |W_j(x)|^r \leq \frac{C_{W_r}}{(ph)^{r-1}} \text{ mit } C_{W_r} \in \mathbb{R}^+$$

Beweis.

- a) Siehe [Her17] Seite 32.
- b) Aus Teil a) und (A.6) folgt:

$$\sum_{j=1}^p |W_j(x)|^r \leq C_{W_0}ph \cdot \left(\frac{C_W}{ph}\right)^r = \frac{\overbrace{C_{W_0} C_W^r}^{=: C_{W_r}}}{(ph)^{r-1}}, \text{ mit } C_{W_0}, C_W \in \mathbb{R}^+ \Rightarrow C_{W_r} \in \mathbb{R}^+$$

□

Proposition 3.7.

Sei $f \in \mathcal{C}^{1,\alpha}(\mathbb{R}^D)$, mit $D \in \mathbb{N}$ und $\alpha \in (0, 1]$, eine deterministische Funktion und die Gewichte des Schätzers erfüllen Annahme 3.4. Dann gilt für alle $(x_{j_1}, \dots, x_{j_D}), (x_1, \dots, x_D) \in [0, 1]^D$:

$$\begin{aligned} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(x_{j_1}, \dots, x_{j_D}) &= \sum_{j_1, \dots, j_D=1}^p W_{j_1}(x_1) \cdot \dots \cdot W_{j_D}(x_D) f(x_{j_1}, \dots, x_{j_D}) \\ &= f(x_1, \dots, x_D) + \mathcal{O}(h^{1+\alpha}) \end{aligned}$$

Beweis.

Wir untersuchen für diese Aussage die betragsmäßige Differenz

$$\left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}_j) - f(\vec{x}) \right| \tag{3.6}$$

wobei $\vec{x}_j := (x_{j_1}, \dots, x_{j_D})$ und $\vec{x} := (x_1, \dots, x_D)$.

Nach dem *Satz von Taylor* kann $f(\vec{x}_j)$ für $m \in \mathbb{N}_0$ wie folgt dargestellt werden:

$$f(\vec{x}_j) = T_m f(\vec{x}_j; \vec{x}) + R_m f(\vec{x}_j; \vec{x})$$

wobei T_m das m -te Taylorpolynom am Entwicklungspunkt \vec{x} und R_m das dazugehörige Restglied ist. Dieses Restglied hat verschiedene Darstellungsformen, wir entscheiden uns hierbei für die *Lagrangesche Darstellung* mit Multiindex-Notation:

$$R_m f(\vec{x}_j; \vec{x}) = \sum_{|\mathbf{l}|=m+1} \partial^{\mathbf{l}} f(\vec{x} + \tau_j(\vec{x}_j - \vec{x})) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!}$$

wobei $\tau_j \in [0, 1]$ und ∂ der Differentialoperator sein soll. Für $m = 0$ folgt:

$$\begin{aligned} f(\vec{x}_j) &= T_0 f(\vec{x}_j; \vec{x}) + R_0 f(\vec{x}_j; \vec{x}) = f(\vec{x}) + \sum_{|\mathbf{l}|=1} \partial^{\mathbf{l}} f(\vec{x} + \tau_j(\vec{x}_j - \vec{x})) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} \\ &= f(\vec{x}) + \sum_{|\mathbf{l}|=1} \partial^{\mathbf{l}} f(\vec{x}) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} - \sum_{|\mathbf{l}|=1} \partial^{\mathbf{l}} f(\vec{x}) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} + \sum_{|\mathbf{l}|=1} \partial^{\mathbf{l}} f(\vec{x} + \tau_j(\vec{x}_j - \vec{x})) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} \\ &= f(\vec{x}) + \sum_{|\mathbf{l}|=1} \partial^{\mathbf{l}} f(\vec{x}) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} + \sum_{|\mathbf{l}|=1} \left(\partial^{\mathbf{l}} f(\vec{x} + \tau_j(\vec{x}_j - \vec{x})) - \partial^{\mathbf{l}} f(\vec{x}) \right) \frac{(\vec{x}_j - \vec{x})^{\mathbf{l}}}{\mathbf{l}!} \\ &= f(\vec{x}) + \sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} (x_{j_l} - x_l) + \sum_{l=1}^D \left(\frac{\partial f(\vec{x} + \tau_j(\vec{x}_j - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right) (x_{j_l} - x_l) \end{aligned}$$

und damit wiederum

$$\begin{aligned} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}_j) &= \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}) \\ &+ \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) \left(\sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} (x_{j_l} - x_l) \right) \\ &+ \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) \left(\sum_{l=1}^D \left(\frac{\partial f(\vec{x} + \tau_j(\vec{x}_j - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right) (x_{j_l} - x_l) \right) \end{aligned}$$

$f(\vec{x})$ beinhaltet keinen Index, der für die Summe bzw. das Produkt bzgl. der Gewichte relevant wäre. Daher folgt für den ersten Summanden mit (A.9):

$$\sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}) = f(\vec{x}) \cdot \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) = f(\vec{x}) \cdot \prod_{k=1}^D \underbrace{\left(\sum_{j=1}^p W_j(x_k) \right)}_{=1} = f(\vec{x})$$

Der zweite Summand lässt sich umformen zu:

$$\begin{aligned} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) \left(\sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} (x_{j_l} - x_l) \right) &= \sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} \left(\sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) (x_{j_l} - x_l) \right) \\ &= \sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} \left(\left(\prod_{k=1}^D \left(\sum_{j=1}^p W_j(x_k) \right) \right) (x_{j_l} - x_l) \right) \end{aligned}$$

Für den Fall $l = k$ gilt nun für den jeweiligen Teilsummanden nach (A.9):

$$\frac{\partial f(\vec{x})}{\partial x_k} \left(\sum_{j=1}^p W_j(x_k) (x_{j_k} - x_k) \right) = 0$$

Da sowohl die äußere Summe, als auch das innere Produkt von 1 bis D laufen, entsteht in jedem Produkt ein Faktor, der 0 ist, wodurch auch jeder einzelne Teilsummand und somit der komplette zweite Summand 0 ist. Damit folgt für Gleichung (3.6):

$$\begin{aligned} \left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}_j) - f(\vec{x}) \right| &= \left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) \right. \\ &\quad \cdot \left(\sum_{l=1}^D \left(\frac{\partial f(\vec{x} + \tau_j(\vec{x}_j - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right) (x_{j_l} - x_l) \right) \left. \right| \\ &\stackrel{\Delta\text{-Ungl.}}{\leq} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{j_k}(x_k)| \right) \\ &\quad \cdot \left(\sum_{l=1}^D \left| \frac{\partial f(\vec{x} + \tau_j(\vec{x}_j - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right| |x_{j_l} - x_l| \right) \end{aligned} \quad (3.7)$$

Da $f \in C^{1,\alpha}(\mathbb{R}^D)$, gilt außerdem für alle $l = 1, \dots, D$:

$$\left| \frac{\partial f(\vec{x} + \tau_j(\vec{x}_j - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right| \leq H_f \|\vec{x} + \tau_j(\vec{x}_j - \vec{x}) - \vec{x}\|^\alpha = H_f \tau_j \|\vec{x}_j - \vec{x}\|^\alpha \stackrel{\tau_j \in [0,1]}{\leq} H_f \|\vec{x}_j - \vec{x}\|^\alpha$$

wobei H_f die Hölderkonstante von f bezeichnet. Wir verwenden für $\|\cdot\|$ wieder die *Manhattan-Metrik*, also $\|\vec{x}_j - \vec{x}\| = \sum_{k=1}^D |x_{j_k} - x_k|$. Daraus folgt über Gleichung (3.7):

$$\left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}_j) - f(\vec{x}) \right| \leq \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{j_k}(x_k)| \right) \left(\sum_{l=1}^D H_f \left(\sum_{k=1}^D |x_{j_k} - x_k| \right)^\alpha |x_{j_l} - x_l| \right)$$

Nach (A.8) gilt, dass $|W_{j_k}(x_k)| = 0 \Rightarrow \prod_{k=1}^D |W_{j_k}(x_k)| = 0$, falls $|x_{j_k} - x_j| > h$. Das heißt, $|x_{j_k} - x_j|$ muss für alle $k = 1, \dots, D$ kleiner oder gleich h sein. Damit folgt schlussendlich mit Proposition 3.6 über obige Gleichung:

$$\begin{aligned} \left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(\vec{x}_j) - f(\vec{x}) \right| &\leq \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{j_k}(x_k)| \right) \left(\sum_{l=1}^D H_f \left(\sum_{k=1}^D h \right)^\alpha h \right) \\ &= D^{1+\alpha} H_f h^{1+\alpha} \left(\prod_{k=1}^D \left(\sum_{j=1}^p |W_j(x_k)| \right) \right) \\ &\leq D^{1+\alpha} H_f h^{1+\alpha} \left(\prod_{k=1}^D C_{W_1} \right) = D^{1+\alpha} H_f C_{W_1}^D h^{1+\alpha} \\ &= \mathcal{O}(h^{1+\alpha}) \end{aligned}$$

□

Bemerkung 3.8.

Erweitert man (A.9) dahingehend, dass für ein $m \in \mathbb{N}$ und alle $k = 1, \dots, m : \sum_{j=1}^p W_j(x)(x_j - x)^k = 0$, dann gilt Proposition 3.7 für $f \in \mathcal{C}^{m,\alpha}(\mathbb{R}^D)$ ebenfalls in leicht veränderter Form:

$$\sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{j_k}(x_k) \right) f(x_{j_1}, \dots, x_{j_D}) = f(x_1, \dots, x_D) + \mathcal{O}(h^{m+\alpha})$$

Genau diese Eigenschaft würde zu schnelleren Konvergenzraten führen, wie in Bemerkung 3.2 und 3.5 erwähnt. Der Beweis dafür ist deutlich komplizierter - gerade weil man es im Taylor-Polynom dann mit verschiedenen m -ten Ableitungen zu tun hat - funktioniert vom Prinzip her aber genauso.

Proposition 3.9.

Sei $A \in \mathbb{R}^{p^D \times p^D}$, mit $D \in \mathbb{N}$, eine symmetrische, positiv semidefinite Matrix, für die, für $p \rightarrow \infty$, die folgende Eigenschaft gilt: Der maximale Eigenwert ist von der Ordnung $\mathcal{O}(p^d)$ mit $d \in \mathbb{R}$. Anders formuliert: Es existiert eine endliche Konstante $C \in \mathbb{R}^+$ und ein $p_0 \in \mathbb{N}$, so dass für den maximalen Eigenwert λ_{\max} von A gilt $\lambda_{\max} \leq Cp^d$ für alle $p \geq p_0$. Zudem erfüllen die Gewichte des Schätzers Annahme 3.4.

Dann gilt für alle $(x_1, \dots, x_D), (y_1, \dots, y_D), (x'_1, \dots, x'_D), (y'_1, \dots, y'_D) \in [0, 1]^D$:

$$\begin{aligned} a) \quad &\left(\bigotimes_{k=1}^D \vec{W}(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) \right) = \mathcal{O} \left(\frac{1}{p^{D-d} h^D} \right) \\ b) \quad &\left(\bigotimes_{k=1}^D \vec{W}(x_k) - \bigotimes_{k=1}^D \vec{W}(x'_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) - \bigotimes_{k=1}^D \vec{W}(y'_k) \right) = \mathcal{O} \left(\frac{1}{p^{D-d} h^D} \right) \end{aligned}$$

wobei $\vec{W}(x) := (W_1(x), W_2(x), \dots, W_p(x))^T$ und $\bigotimes_{k=1}^D$ für das D -fache Kronecker-Produkt steht.

Beweis.

- a) Die Aussage wird wieder über die betragsmäßige Differenz zu 0 untersucht. Sei $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt auf dem \mathbb{R}^{p^D} . Dann gilt mit der *Cauchy-Schwarz-Ungleichung*:

$$\begin{aligned} \left| \left(\bigotimes_{k=1}^D \vec{W}(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) \right) \right| &= \left| \left\langle \bigotimes_{k=1}^D \vec{W}(x_k), A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) \right) \right\rangle \right| \\ &\leq \left\| \bigotimes_{k=1}^D \vec{W}(x_k) \right\|_2 \cdot \left\| A \cdot \bigotimes_{k=1}^D \vec{W}(y_k) \right\|_2 \\ &\leq \|A\|_M \cdot \left\| \bigotimes_{k=1}^D \vec{W}(x_k) \right\|_2 \cdot \left\| \bigotimes_{k=1}^D \vec{W}(y_k) \right\|_2 \end{aligned}$$

$\|\cdot\|_M$ ist eine geeignete Matrixnorm, $\|\cdot\|_2$ ist die *euklidische Norm*, welche mit dem *Kronecker-Produkt* multiplikativ vertauschbar ist. Damit gilt für $\left\| \bigotimes_{k=1}^D \vec{W}(x_k) \right\|_2$ mit Proposition 3.6:

$$\begin{aligned} \left\| \bigotimes_{k=1}^D \vec{W}(x_k) \right\|_2 &= \prod_{k=1}^D \|\vec{W}(x_k)\|_2 = \prod_{k=1}^D \left(\sum_{j=1}^p |W_j(x_k)|^2 \right)^{\frac{1}{2}} \\ &\leq \prod_{k=1}^D \left(\frac{C_{W_2}}{ph} \right)^{\frac{1}{2}} = \left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} \end{aligned} \quad (3.8)$$

Analoge Abschätzung natürlich für $\left\| \bigotimes_{k=1}^D \vec{W}(y_k) \right\|_2$.

$$\Rightarrow \left| \left(\bigotimes_{k=1}^D \vec{W}(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) \right) \right| \leq \|A\|_M \cdot \left(\left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} \right)^2 = \|A\|_M \cdot \left(\frac{C_{W_2}}{ph} \right)^D$$

Für die Matrixnorm verwenden wir die mit der euklidischen Norm verträgliche *Spektralnorm*. $\|A\|_M$ entspricht dann wegen der Symmetrie und positiven Semi-definitheit von A deren größtem Eigenwert, welcher nach Voraussetzung für alle $p \geq p_0 \in \mathbb{N}$ durch Cp^d beschränkt wurde.

$$\Rightarrow \left| \left(\bigotimes_{k=1}^D \vec{W}(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) \right) \right| \leq Cp^d \left(\frac{C_{W_2}}{ph} \right)^D = \mathcal{O} \left(\frac{1}{p^{D-d}h^D} \right)$$

da $C, C_{W_2} \in \mathbb{R}^+$ konstant.

- b) Wir betrachten auch hier die betragsmäßige Differenz und wenden die *Dreiecksungleichung*, sowie die Abschätzung aus Gleichung (3.8) an.

$$\begin{aligned} & \left| \left(\bigotimes_{k=1}^D \vec{W}(x_k) - \bigotimes_{k=1}^D \vec{W}(x'_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) - \bigotimes_{k=1}^D \vec{W}(y'_k) \right) \right| \\ &= \left| \left\langle \bigotimes_{k=1}^D \vec{W}(x_k) - \bigotimes_{k=1}^D \vec{W}(x'_k), A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) - \bigotimes_{k=1}^D \vec{W}(y'_k) \right) \right\rangle \right| \\ &\leq \left\| \bigotimes_{k=1}^D \vec{W}(x_k) - \bigotimes_{k=1}^D \vec{W}(x'_k) \right\|_2 \cdot \left\| A \cdot \left(\bigotimes_{k=1}^D \vec{W}(y_k) - \bigotimes_{k=1}^D \vec{W}(y'_k) \right) \right\|_2 \\ &\leq \|A\|_M \cdot \left\| \bigotimes_{k=1}^D \vec{W}(x_k) - \bigotimes_{k=1}^D \vec{W}(x'_k) \right\|_2 \cdot \left\| \bigotimes_{k=1}^D \vec{W}(y_k) - \bigotimes_{k=1}^D \vec{W}(y'_k) \right\|_2 \\ &\leq \|A\|_M \cdot \left(\left\| \bigotimes_{k=1}^D \vec{W}(x_k) \right\|_2 + \left\| \bigotimes_{k=1}^D \vec{W}(x'_k) \right\|_2 \right) \cdot \left(\left\| \bigotimes_{k=1}^D \vec{W}(y_k) \right\|_2 + \left\| \bigotimes_{k=1}^D \vec{W}(y'_k) \right\|_2 \right) \\ &\leq \|A\|_M \cdot \left(\left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} + \left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} \right) \cdot \left(\left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} + \left(\frac{C_{W_2}}{ph} \right)^{\frac{D}{2}} \right) \\ &= \|A\|_M \cdot 4 \left(\frac{C_{W_2}}{ph} \right)^D \end{aligned}$$

Ab hier ist das Vorgehen identisch zu a), abgesehen vom Faktor 4, was aber innerhalb der O-Notation keine Rolle spielt. □

Bemerkung 3.10.

Proposition 3.7 und 3.9 werden wir in zahlreichen Beweisen für μ, R und M aus (A.1) und (A.3) bzw. $\text{Cov}(\vec{\varepsilon}_1)$ und $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$ aus (A.4) verwenden. Da hier jeweils die geforderten Voraussetzungen gelten, werden wir dies nicht jedes mal explizit in den Beweisen erwähnen, sondern nur kurz auf die Propositionen selbst verweisen.

3.4. Konkrete Beispiele

Bevor wir nun die Eigenschaften der Schätzer erörtern, wollen wir zunächst noch illustrieren, wie die Bedingungen aus Annahme 3.1 an den Fehler nachgerechnet werden können und werfen danach noch einen Blick auf einen speziellen Schätzer, welcher Annahme 3.4 erfüllt und somit natürlich auch alle weiteren Aussagen aus dieser Arbeit.

3.4.1. Fehlerterme

Hat man nicht genug Informationen zu den Eigenwerten von $\text{Cov}(\vec{\varepsilon}_1)$ und $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$, ist es hilfreich, die einzelnen Einträge davon zu betrachten und gewisse Gesetzmäßigkeiten festzustellen. Sie sind als Kovarianzmatrizen beide symmetrisch und positiv semidefinit, der größte Eigenwert entspricht damit gerade der jeweiligen Spektralnorm. Diese kann wiederum durch $\|\cdot\|_2 \leq \sqrt{\|\cdot\|_1 \cdot \|\cdot\|_\infty}$ abgeschätzt werden, wobei $\|\cdot\|_1$ die Spaltensummennorm und $\|\cdot\|_\infty$ die Zeilensummennorm darstellt. Aufgrund der Symmetrie der Matrizen, gilt in diesem Fall $\|\cdot\|_1 = \|\cdot\|_\infty$, der größte Eigenwert ist also jeweils durch die Spalten- und Zeilensummennorm gleichermaßen beschränkt. Wir werden zur Untersuchung die Zeilensummennorm heranziehen.

Beispiel 3.11. (Unabhängigkeit bzgl. der Messstellen)

Falls der Fehler nicht nur bzgl. der Stichprobeneinheiten i , sondern auch bzgl. der Messpunkte x_j unabhängig ist, also falls ε_{ij} für alle $i = 1, \dots, n$ und $j = 1, \dots, p$, unabhängig ist, so gilt für die Kovarianzmatrix von $\vec{\varepsilon}_1$:

$$\text{Cov}(\vec{\varepsilon}_1) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix}, \quad \sigma_j^2 = \mathbb{E}[\varepsilon_{1j}^2]$$

da $\mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}] = \mathbb{E}[\varepsilon_{1j_1}]\mathbb{E}[\varepsilon_{1j_2}] = 0$ für $j_1 \neq j_2$. Die Eigenwerte dieser Kovarianzmatrix sind nun gerade $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$, also gilt $\lambda_{\max,1} = \max_{j \in \{1, \dots, p\}} \sigma_j^2$. (A.4) ist demnach erfüllt, wenn das maximale zweite Moment von ε_{1j} bzgl. j von der Ordnung $\mathcal{O}(p^{d_1})$, $d_1 \in [0, 1)$ ist.

Die Kovarianzmatrix von $\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1$ wird zeilenweise durch j_1, j_2 und spaltenweise durch j_3, j_4 indiziert. Demnach gilt für die Zeilensummennorm:

$$\|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)\|_\infty = \max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p |\text{Cov}(\varepsilon_{1j_1}\varepsilon_{1j_2}, \varepsilon_{1j_3}\varepsilon_{1j_4})|$$

Wir betrachten nun einen Beispieleintrag mit dem *Verschiebungssatz*:

$$\text{Cov}(\varepsilon_{1j_1}\varepsilon_{1j_2}, \varepsilon_{1j_3}\varepsilon_{1j_4}) = \mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}\varepsilon_{1j_3}\varepsilon_{1j_4}] - \mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}]\mathbb{E}[\varepsilon_{1j_3}\varepsilon_{1j_4}]$$

mit $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$ beliebig. Durch die Unabhängigkeit von ε_{1*} ist dieser Term 0, falls ein einziges j_* von allen anderen verschieden ist, denn dann lässt sich der Erwartungswert auf diese, von den anderen unabhängige Variable einzeln anwenden und eliminiert dadurch den ganzen Term. Genauso ist er 0 für $j_1 = j_2 \neq j_3 = j_4$, denn dann gilt:

$$\begin{aligned} \mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}\varepsilon_{1j_3}\varepsilon_{1j_4}] - \mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}]\mathbb{E}[\varepsilon_{1j_3}\varepsilon_{1j_4}] &= \mathbb{E}[\varepsilon_{1j_1}^2\varepsilon_{1j_3}^2] - \mathbb{E}[\varepsilon_{1j_1}^2]\mathbb{E}[\varepsilon_{1j_3}^2] \\ &= \mathbb{E}[\varepsilon_{1j_1}^2]\mathbb{E}[\varepsilon_{1j_3}^2] - \mathbb{E}[\varepsilon_{1j_1}^2]\mathbb{E}[\varepsilon_{1j_3}^2] = 0 \end{aligned}$$

Es verbleiben noch drei weitere Fälle.

1. $j_1 = j_2 = j_3 = j_4$
2. $j_1 = j_3 \neq j_2 = j_4$
3. $j_1 = j_4 \neq j_2 = j_3$

Sehen wir uns nun einmal eine $p^2 \times p^2$ -Matrix, welche quasi die Indizes von $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$ enthält, also eine Matrix mit Einträgen $j_1j_2j_3j_4$, wobei zeilenweise die ersten beiden Einträge, und spaltenweise die letzten beiden Einträge jeweils bis pp ansteigen.

$$\begin{pmatrix} 1111 & 1112 & 1113 & \dots & 111p & 1121 & \dots & 11pp \\ 1211 & 1212 & 1213 & \dots & 121p & 1221 & \dots & 12pp \\ 1311 & 1312 & 1313 & \dots & 131p & 1321 & \dots & 13pp \\ \vdots & \vdots & \vdots & \ddots & & & & \vdots \\ 1p11 & 1p12 & 1p13 & & & & & 1ppp \\ 2111 & 2112 & 2113 & & & & & 21pp \\ \vdots & \vdots & \vdots & & & & & \vdots \\ pp11 & pp12 & pp13 & \dots & pp1p & pp21 & \dots & pppp \end{pmatrix}$$

Für jede Zeile dieser Matrix können nur zwei verschiedene Fälle auftreten:

- Die Zeile enthält einen Eintrag mit $j_1 = j_2$. Betrachte o. B. d. A. die erste Zeile, hier gibt es den Eintrag $j_1 = j_2 = 1$. In dieser Zeile tritt der Fall 1. genau einmal auf, und zwar bei $j_1 = j_2 = j_3 = j_4 = 1$. In allen anderen Einträgen ist entweder ein j_* von allen anderen verschieden, oder es gilt $j_3 = j_4 \neq 1 \Rightarrow j_1 = j_2 \neq j_3 = j_4$. $\text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon})$ hat also in jeder Zeile mit $j_1 = j_2$ höchstens einen von 0 verschiedenen Eintrag.
- Die Zeile enthält einen Eintrag mit $j_1 \neq j_2$. Betrachte o. B. d. A. die zweite Zeile, hier gibt es den Eintrag $1 = j_1 \neq j_2 = 2$. Der Fall 1. kann in dieser Zeile nicht auftreten. Der Fall 2. kann nur in den Spalten mit $j_3 = 1$ auftreten, also in den ersten p Spalten. Hier gilt aber wiederum nur ein einziges mal $j_2 = j_4$ (= 2 in Spalte 2). Der Fall 3. kann nur in den Spalten mit $j_4 = 1$ auftreten, also in Spalte $0 \cdot p + 1/1 \cdot p + 1/2 \cdot p + 1/\dots/(p-1) \cdot p + 1$. Hier gilt aber wiederum nur ein einziges mal $j_2 = j_3$ (= 2 in Spalte $p+1$). $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$ hat also in jeder Zeile mit $j_1 \neq j_2$ höchstens zwei von 0 verschiedene Einträge.

Das bedeutet, die Summe jeder Zeile der Matrix hat entweder die Form $\mathbb{E}[\varepsilon_{1j_1}^4] - (\mathbb{E}[\varepsilon_{1j_1}^2])^2$ oder $2\mathbb{E}[\varepsilon_{1j_1}^2 \varepsilon_{1j_2}^2]$ mit $j_1, j_2 \in \{1, \dots, p\}$ und $j_1 \neq j_2$. Sei nun das maximale vierte Moment von ε_{1j} bzgl. j von der Ordnung $\mathcal{O}(p^{d_2})$, $d_2 \in [0, 2)$. Dann folgt mit der *Cauchy-Schwarz-Ungleichung* für alle $j_1, j_2 \in \{1, \dots, p\}$ und $j_1 \neq j_2$:

$$\begin{aligned} \mathbb{E}[\varepsilon_{1j_1}^4] - (\mathbb{E}[\varepsilon_{1j_1}^2])^2 &\leq \mathbb{E}[\varepsilon_{1j_1}^4] \leq \max_{j_1 \in \{1, \dots, p\}} E[\varepsilon_{1j_1}^4] = \mathcal{O}(p^{d_2}) \\ 2\mathbb{E}[\varepsilon_{1j_1}^2 \varepsilon_{1j_2}^2] &\leq 2(\mathbb{E}[\varepsilon_{1j_1}^4])^{\frac{1}{2}} (\mathbb{E}[\varepsilon_{1j_2}^4])^{\frac{1}{2}} \leq 2 \max_{j_1 \in \{1, \dots, p\}} \mathbb{E}[\varepsilon_{1j_1}^4] = \mathcal{O}(p^{d_2}) \end{aligned}$$

Damit folgt für den größten Eigenwert $\lambda_{\max,2}$ von $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$:

$$\lambda_{\max,2} = \|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)\|_2 \leq \|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)\|_\infty = \mathcal{O}(p^{d_2})$$

wonach (A.4) gilt.

Dieses Beispiel ist insofern sehr hilfreich, weil für die Fehler gar keine konkrete Verteilung festgelegt werden muss. Bspw. könnte es ja durchaus sein, dass an Messpunkten mit niedrigem Wert, also nahe der 0, Messfehler mit anderer Verteilung auftreten, als an Messpunkten mit hohem Wert, also nahe der 1. Sofern die Messfehler untereinander unabhängig sind, gelten nach wie vor alle in dieser Arbeit untersuchten Eigenschaften.

Beispiel 3.12. (White Noise)

Ein allgemeiner diskreter White Noise Prozess ist (in unserem Fall) definiert durch $\{X(t) \mid t \in T\}$ mit T abzählbar, $\mathbb{E}[X(t)] = 0$ für alle $t \in T$ und $\mathbb{E}[X(t)X(t')] = \sigma^2 \delta_{t,t'} < \infty$ für alle $t, t' \in T$, wobei $\delta_{t,t'}$ das *Kronecker-Delta* darstellt. Im Falle eines *allgemeinen White Noise Fehlers* $\vec{\varepsilon}_1$ gilt also $T = \{1, \dots, p\}$ und damit:

$$\text{Cov}(\vec{\varepsilon}_1) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

Diese Kovarianzmatrix besitzt nur den Eigenwert σ^2 . Insofern σ^2 von der Ordnung $\mathcal{O}(p^{d_1})$, $d_1 \in [0, 1)$ ist, gilt damit sofort (A.4).

Für die Kovarianzmatrix von $\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1$ müssen allerdings weitere Annahmen getroffen werden, da im Falle eines allgemeinen White Noise Fehlers die Komponenten von $\vec{\varepsilon}_1$ nicht zwangsläufig unabhängig sind. Das bedeutet, $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$ ist im Allgemeinen vollbesetzt und man hat auch keine konkrete Kenntnis über die einzelnen Einträge. Eine Möglichkeit wäre es, die Unabhängigkeit der Komponenten einfach direkt wieder vorauszusetzen \Rightarrow *Independent White Noise*. Des Weiteren könnte man die Verteilung des Fehlers durch $\vec{\varepsilon}_1 \sim \mathcal{N}_p(0, \text{Cov}(\vec{\varepsilon}_1))$ festlegen \Rightarrow *Standard White Noise*. In diesem Fall folgt, dass die einzelnen Komponenten von $\vec{\varepsilon}_1$ ebenfalls (eindimensional) normalverteilt sind. Aus der zugehörigen Kovarianzmatrix folgt außerdem, dass die Komponenten unkorreliert sind.

Daraus und der Normalverteilung folgt wiederum Unabhängigkeit. (A.4) gilt unter einer dieser beiden Annahmen dann nach der gleichen Argumentation wie bei Beispiel 3.11.

Beispiel 3.13. (Moving Average)

Ein *Moving-Average Fehler* $\vec{\varepsilon}_1$ der Ordnung $q \in \mathbb{N}_0$ ist (in unserem Fall) definiert durch:

$$\varepsilon_{1j} =: b_0 \varepsilon'_j + b_1 \varepsilon'_{j-1} + \dots + b_q \varepsilon'_{j-q}, \quad j = 1, \dots, p$$

wobei $b_0, \dots, b_q \in \mathbb{R}$ und $(\varepsilon'_j)_{j \in \mathbb{Z}}$ ein Independent White Noise Prozess oder ein Standard White Noise Prozess ist. O. B. d. A. sei $b_0 = \dots = b_q = 1$, $(\varepsilon'_j)_{j \in \mathbb{Z}}$ ein Standard White Noise Prozess und p hinreichend groß. Die Einträge der Kovarianzmatrix sehen dann wie folgt aus:

$$\sigma_{j_1 j_2} = \mathbb{E} \left[\left(\sum_{i=0}^q \varepsilon'_{j_1-i} \right) \left(\sum_{i=0}^q \varepsilon'_{j_2-i} \right) \right] = \sum_{i_1=0}^q \sum_{i_2=0}^q \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] \quad (3.9)$$

mit $j_1, j_2 \in \{1, \dots, p\}$ beliebig. Für den Erwartungswert gilt:

$$\begin{aligned} \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] &= \begin{cases} \sigma^2, & \text{für } j_1 - i_1 = j_2 - i_2 \\ 0, & \text{sonst} \end{cases} \\ \Rightarrow \sigma_{j_1 j_2} &= \sigma^2 \sum_{i_1=0}^q \sum_{i_2=0}^q \mathbb{1}\{j_1 - i_1 = j_2 - i_2\} = \sigma^2 \cdot \left(q + 1 - (\max\{j_1, j_2\} - \min\{j_1, j_2\}) \right)^+ \\ \Rightarrow \text{Cov}(\vec{\varepsilon}_1) &= \sigma^2 \begin{pmatrix} q+1 & q & q-1 & q-2 & \dots & 1 & 0 & 0 & 0 & \dots \\ q & q+1 & q & q-1 & \dots & 2 & 1 & 0 & 0 & \dots \\ q-1 & q & q+1 & q & \dots & 3 & 2 & 1 & 0 & \dots \\ q-2 & q-1 & q & q+1 & \dots & 4 & 3 & 2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & & & & & \\ 1 & 2 & 3 & 4 & & & & & & \\ 0 & 1 & 2 & 3 & & & & & & \\ 0 & 0 & 1 & 2 & & & & & & \\ 0 & 0 & 0 & 1 & & & & & & \\ \vdots & \vdots & \vdots & \vdots & & & & & & \end{pmatrix} \quad (3.10) \end{aligned}$$

$\text{Cov}(\vec{\varepsilon}_1)$ ist also eine symmetrische Bandmatrix der Bandbreite $2q + 1$. Für diese Matrix gilt:

$$\|\text{Cov}(\vec{\varepsilon}_1)\|_\infty = \sigma^2 (q + 1 + 2 \sum_{i=1}^q i) = \sigma^2 (q + 1 + q(q + 1)) = \sigma^2 (q + 1)^2$$

was sich auch durch die konkrete Definition der Normen über $\sigma_{j_1 j_2}$ bestätigt:

$$\|\text{Cov}(\vec{\varepsilon}_1)\|_\infty = \max_{j_1 \in \{1, \dots, p\}} \sum_{j_2=1}^p |\sigma_{j_1 j_2}| = \sigma^2 \cdot \max_{j_1 \in \{1, \dots, p\}} \sum_{j_2=1}^p \left(q + 1 - (\max\{j_1, j_2\} - \min\{j_1, j_2\}) \right)^+ \quad (3.11)$$

$\max\{j_1, j_2\} - \min\{j_1, j_2\}$ ist dabei nichts anderes als der Abstand von j_1 und j_2 . Wegen dem Positivteil der Summanden, spielen nur Abstände von 0 bis q eine Rolle. Wenn man j_1 passend wählt (möglichst mittig zwischen 1 und p), kann jeder dieser Abstände von j_2 aus genau 2 mal erreicht werden (mit Ausnahme von Abstand 0, hier nur 1 mal). Damit gilt insgesamt für die Norm:

$$\begin{aligned} \|\text{Cov}(\vec{\varepsilon}_1)\|_\infty &= \sigma^2 \left(\underbrace{q+1-0}_{\text{Abstand 0}} + \underbrace{2(q+1-1)}_{2 \text{ mal Abstand 1}} + \underbrace{2(q+1-2)}_{2 \text{ mal Abstand 2}} + \dots + \underbrace{2(q+1-q)}_{2 \text{ mal Abstand } q} \right) \\ &= \sigma^2 (q+1 + 2q + 2(q-1) + \dots + 2) = \sigma^2 (q+1 + 2 \sum_{i=1}^q i) = \sigma^2 (q+1)^2 \end{aligned}$$

Wenn σ^2 wieder von der Ordnung $\mathcal{O}(p^{d_1})$, $d_1 \in [0, 1)$ ist, gilt:

$$\lambda_{1,\max} = \|\text{Cov}(\vec{\varepsilon}_1)\|_2 \leq \|\text{Cov}(\vec{\varepsilon}_1)\|_\infty = \sigma^2 \underbrace{(q+1)^2}_{\in \mathbb{N}} = \mathcal{O}(p^{d_1})$$

und damit (A.4).

Für die Kovarianzmatrix von $\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1$ betrachten wir wieder einen Beispielintrag. Mit Gleichung (3.9) gilt:

$$\begin{aligned} \text{Cov}(\varepsilon_{1j_1} \varepsilon_{1j_2}, \varepsilon_{1j_3} \varepsilon_{1j_4}) &= \mathbb{E}[\varepsilon_{1j_1} \varepsilon_{1j_2} \varepsilon_{1j_3} \varepsilon_{1j_4}] - \mathbb{E}[\varepsilon_{1j_1} \varepsilon_{1j_2}] \mathbb{E}[\varepsilon_{1j_3} \varepsilon_{1j_4}] \\ &= \mathbb{E} \left[\left(\sum_{i=0}^q \varepsilon'_{j_1-i} \right) \left(\sum_{i=0}^q \varepsilon'_{j_2-i} \right) \left(\sum_{i=0}^q \varepsilon'_{j_3-i} \right) \left(\sum_{i=0}^q \varepsilon'_{j_4-i} \right) \right] \\ &\quad - \left(\sum_{i_1=0}^q \sum_{i_2=0}^q \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] \right) \left(\sum_{i_1=0}^q \sum_{i_2=0}^q \mathbb{E}[\varepsilon'_{j_3-i_1} \varepsilon'_{j_4-i_2}] \right) \\ &= \mathbb{E} \left[\sum_{i_1=0}^q \sum_{i_2=0}^q \sum_{i_3=0}^q \sum_{i_4=0}^q \varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2} \varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4} \right] \\ &\quad - \sum_{i_1=0}^q \sum_{i_2=0}^q \sum_{i_3=0}^q \sum_{i_4=0}^q \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] \mathbb{E}[\varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4}] \\ &= \sum_{i_1=0}^q \sum_{i_2=0}^q \sum_{i_3=0}^q \sum_{i_4=0}^q \left(\mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2} \varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4}] - \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] \mathbb{E}[\varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4}] \right) \end{aligned}$$

mit $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$ beliebig. Für $\mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2} \varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4}] - \mathbb{E}[\varepsilon'_{j_1-i_1} \varepsilon'_{j_2-i_2}] \mathbb{E}[\varepsilon'_{j_3-i_3} \varepsilon'_{j_4-i_4}]$ gilt nun:

$$= \begin{cases} 2\sigma^4, & \text{für } j_1 - i_1 = j_2 - i_2 = j_3 - i_3 = j_4 - i_4 \\ \sigma^4, & \text{für } j_1 - i_1 = j_3 - i_3 \neq j_2 - i_2 = j_4 - i_4 \\ \sigma^4, & \text{für } j_1 - i_1 = j_4 - i_4 \neq j_2 - i_2 = j_3 - i_3 \\ 0, & \text{sonst} \end{cases}$$

Dies gilt wegen der Unabhängigkeit und den Momenteigenschaften der Normalverteilung. Daraus folgt:

$$\begin{aligned}
 \text{Cov}(\varepsilon_{1j_1}\varepsilon_{1j_2}, \varepsilon_{1j_3}\varepsilon_{1j_4}) &= \sigma^4 \sum_{i_1=0}^q \sum_{i_2=0}^q \sum_{i_3=0}^q \sum_{i_4=0}^q \left(2 \cdot \mathbb{1}\{j_1 - i_1 = j_2 - i_2 = j_3 - i_3 = j_4 - i_4\} \right. \\
 &\quad \left. + \mathbb{1}\{j_1 - i_1 = j_3 - i_3 \neq j_2 - i_2 = j_4 - i_4\} \right. \\
 &\quad \left. + \mathbb{1}\{j_1 - i_1 = j_4 - i_4 \neq j_2 - i_2 = j_3 - i_3\} \right) \\
 &= \sigma^4 \left(2 \left(q + 1 - (\max\{j_1, j_2, j_3, j_4\} - \min\{j_1, j_2, j_3, j_4\}) \right)^+ \right) \\
 &\quad + \left(q + 1 - (\max\{j_1, j_3\} - \min\{j_1, j_3\}) \right)^+ \left(q - (\max\{j_2, j_4\} - \min\{j_2, j_4\}) \right)^+ \\
 &\quad + \left(q + 1 - (\max\{j_1, j_4\} - \min\{j_1, j_4\}) \right)^+ \left(q - (\max\{j_2, j_3\} - \min\{j_2, j_3\}) \right)^+ \\
 &\hspace{15em} (3.12)
 \end{aligned}$$

Nun gilt es wieder

$$\|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)\|_\infty = \max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p |\text{Cov}(\varepsilon_{1j_1}\varepsilon_{1j_2}, \varepsilon_{1j_3}\varepsilon_{1j_4})|$$

zu ermitteln. Wir untersuchen dafür jeden Summanden einzeln, was über die, für den Betrag und das Maximum gültige, Δ -Ungleichung legitimiert wird.

- $\max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p \left(q + 1 - (\max\{j_1, j_2, j_3, j_4\} - \min\{j_1, j_2, j_3, j_4\}) \right)^+$

$\max\{j_1, j_2, j_3, j_4\} - \min\{j_1, j_2, j_3, j_4\}$ kann wieder als Abstand interpretiert werden, dieses mal zwischen den gekoppelten Werten j_1, j_2 und j_3, j_4 . Wegen dem Positivteil des Summanden, spielen wieder nur die Abstände 0 bis q eine Rolle. 0 kann wieder nur 1 mal erreicht werden, und zwar genau für $j_1 = j_2 = j_3 = j_4$. Daraus folgt direkt, dass $j_1 = j_2$ sein muss, sonst würde der (größte) Summand $q + 1$ niemals auftauchen. Um wieder alle weiteren möglichen Abstände zu erreichen, muss j_1 bzw. j_2 wieder passend (möglichst mittig zwischen 1 und p) gewählt werden. Die Abstände sind nun genau zu den folgenden Anzahlen erreichbar: 6 mal Abstand 1, 12 mal Abstand 2, 18 mal Abstand 3, ..., $6(q - 1)$ mal Abstand $q - 1$, $6q$ mal Abstand q .

$$\begin{aligned}
 &\Rightarrow \max_{j_3, j_4 \in \{1, \dots, p\}} \sum_{j_1, j_2=1}^p \left(q+1 - (\max\{j_1, j_2, j_3, j_4\} - \min\{j_1, j_2, j_3, j_4\}) \right)^+ \\
 &= q+1 - 0 + 6(q+1-1) + 12(q+1-2) + \dots + 6(q-1)(q+1-q+1) + 6q(q+1-q) \\
 &= q+1 + 6(q+2(q-1) + 3(q-2) + \dots + (q-1)2 + q) = q+1 + 6 \sum_{i=1}^q i(q+1-i) \\
 &= q+1 + 6(q+1) \left(\sum_{i=1}^q i - \sum_{i=1}^q i^2 \right) = q+1 + 6(q+1) \left(\frac{q(q+1)}{2} - \frac{q(q+1)(2q+1)}{6} \right) \\
 &= (q+1)(1 + 3q(q+1) - q(2q+1)) = (q+1)(1 + 3q^2 + 3q - 2q^2 - q) \\
 &= (q+1)(q^2 + 2q + 1) = (q+1)^3
 \end{aligned}$$

$$\bullet \max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p \left(q+1 - (\max\{j_1, j_3\} - \min\{j_1, j_3\}) \right)^+ \left(q - (\max\{j_2, j_4\} - \min\{j_2, j_4\}) \right)^+$$

Dieser Summand kann ohne Probleme in 2 Faktoren zerlegt werden. Diese werden wiederum nach dem gleichen Schema wie in Gleichung (3.11) berechnet.

$$\begin{aligned}
 &\max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p \left(q+1 - (\max\{j_1, j_3\} - \min\{j_1, j_3\}) \right)^+ \left(q - (\max\{j_2, j_4\} - \min\{j_2, j_4\}) \right)^+ \\
 &= \left(\max_{j_1 \in \{1, \dots, p\}} \sum_{j_3=1}^p \left(q+1 - (\max\{j_1, j_3\} - \min\{j_1, j_3\}) \right)^+ \right) \\
 &\quad \cdot \left(\max_{j_2 \in \{1, \dots, p\}} \sum_{j_4=1}^p \left(q - (\max\{j_2, j_4\} - \min\{j_2, j_4\}) \right)^+ \right) = (q+1)^2 q^2
 \end{aligned}$$

Nach selbiger Argumentation folgt auch direkt:

$$\max_{j_1, j_2 \in \{1, \dots, p\}} \sum_{j_3, j_4=1}^p \left(q+1 - (\max\{j_1, j_4\} - \min\{j_1, j_4\}) \right)^+ \left(q - (\max\{j_2, j_3\} - \min\{j_2, j_3\}) \right)^+ = (q+1)^2 q^2$$

Schlussendlich gilt nun mit Gleichung (3.12) und folgende:

$$\lambda_{2, \max} = \left\| \text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) \right\|_2 \leq \left\| \text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) \right\|_\infty \leq 2\sigma^4((q+1)^3 + (q+1)^2 q^2) = \mathcal{O}((p^{d_1})^2) = \mathcal{O}(p^{2d_1})$$

und damit (A.4) mit $d_2 = 2d_1 \in [0, 2)$.

3.4.2. Der lokal lineare Schätzer

Ein Schätzer, welcher durch Gleichung (3.2) ausgedrückt werden kann und Annahme 3.4 erfüllt, zumindest unter ein paar wenigen weiteren Annahmen, ist der sogenannte *lokal lineare Schätzer*. Dieser wurde bereits detailliert untersucht, unter anderem von [Tsy09], [Deg11] und [Her17]. Dieser basiert auf Kleinste-Quadrate- und Kern-Regression Techniken und entspricht im Grunde dem bekannten *Nadaraya-Watson Schätzer*, siehe z. B. [WJ94].

Der Schätzer kann zunächst für alle x als konstante Lösung eines Minimierungsproblems ausgedrückt werden:

$$\hat{\mu}(x) := \arg \min_{\mu(x) \in \mathbb{R}} \sum_{j=1}^p \left(Y_j - \mu(x) \right)^2 K \left(\frac{x - x_j}{h} \right)$$

wobei K eine Kernfunktion ist - z. B. ein Gauß-Kern oder ein Epanechnikov-Kern - und $Y_j - \mu(x)$ die Residuen darstellt. Im nächsten Schritt wird $\mu(x)$ mit einem Taylorpolynom der Ordnung m angenähert. Dieses Polynom kann als Produkt zweier Vektoren ausgedrückt und danach mit den sogenannten *Normalgleichungen* gelöst werden, wodurch eine Darstellung als Gleichung (3.2), also eine Ausdrucksform für die Gewichtsfunktionen erreicht wird, wenn auch implizit. Hier sehen wir nun die Bedeutung des Grads der Ableitungen bzw. Polynome aus Bemerkung 3.2 und 3.5: Wenn ein Taylorpolynom der Ordnung m verwendet werden soll, muss offensichtlich auch die m -te Ableitung von μ existieren und gewisse Eigenschaften erfüllen; je höher die Ordnung, desto akkurater wird der Schätzer. Der Schätzer, der durch diese Technik erreicht wird, wird auch als *lokal polynomialer Schätzer* bezeichnet.

Der *lokal lineare Schätzer* nutzt dagegen eine Taylorapproximation von μ der Ordnung $m = 1$, was einer Annäherung mittels einer Gerade/linearen Funktion entspricht. Für diesen Fall haben [Deg11] und [Her17] gezeigt, dass die Gewichtsfunktionen explizit darstellbar sind und mit ein paar weiteren Anforderungen an den Kern, wie z. B. die kompakte Definitionsmenge $[-1, 1]$ und Beschränktheit, sind auch alle weiteren Annahmen erfüllt. Siehe z. B. [Fan92] für mehr Informationen zum lokal linearen Schätzer.

Eine weitere Kategorie von Schätzern, welche die Annahmen zumindest größtenteils erfüllen, wären *Splineschätzer*, bei welchen die Gewichtsfunktionen auf Splines basieren, siehe hierfür z. B. [Was06]. Diese sind ebenfalls lineare Schätzer und [Sil84] zeigt, dass sie näherungsweise Kernschätzern entsprechen, wobei die Kernfunktion, wie auch die Bandbreite auf eine ganz bestimmte Art und Weise berechnet wird. Auch wenn zum Teil geforderte Voraussetzungen verletzt sind, legen Simulationen nahe, dass die im Folgenden formulierten Aussagen genauso für Splineschätzer gelten.

4. Eigenschaften der Schätzer

Wir werden nun jeweils drei Gütekriterien für die Schätzer formulieren. Zum einen eine Abschätzung für den *Bias* und die *Varianz*, und damit automatisch einhergehend *asymptotische Erwartungstreue* und eine Abschätzung für den *Mean Squared Error*. Weiterhin wird die *asymptotische Normalität* nachgewiesen und schlussendlich *Konsistenz*. Da es sich um Schätzer funktionaler Daten handelt, können diese Eigenschaften, vor allem wenn Grenzwerte involviert sind, entweder *punktweise* oder *gleichmäßig* durchgeführt werden, wobei natürlich die gleichmäßige Abschätzung eine deutlich stärkere Aussage ist, die wir auch oft, jedoch nicht immer, erreichen werden. Meist werden wir die jeweilige Aussage erst punktweise beweisen, und dann zur Gleichmäßigkeit erweitern.

4.1. Bias und Varianz

Wir stellen zunächst Schranken für den jeweiligen Bias und die Varianz auf. Die asymptotische Erwartungstreue lässt sich über den Bias überprüfen, denn dafür muss er gegen 0 konvergieren, die Abschätzung für den Mean Squared Error wiederum durch die *Bias-Varianz-Zerlegung* $MSE = \text{Bias}^2 + \text{Var}$.

4.1.1. Trendschätzer

Satz 4.1.

Sei $\hat{\mu}$ der Schätzer aus Gleichung (3.2) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt. Dann gilt für alle $x \in [0, 1]$:

$$\begin{aligned} \text{Bias}(\hat{\mu}(x)) &= \mathcal{O}(h^{1+\alpha_\mu}), \quad \text{Var}(\hat{\mu}(x)) = \mathcal{O}\left(\frac{1+h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \\ \Rightarrow \text{MSE}(\hat{\mu}(x)) &= \mathcal{O}\left(h^{2(1+\alpha_\mu)} + \frac{1+h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \end{aligned}$$

Der Bias entsteht durch die Glättung, wie wir gleich im Beweis sehen werden. Aus $h \rightarrow 0$ folgt $\text{Bias}(\hat{\mu}(x)) \rightarrow 0 \Leftrightarrow \mathbb{E}[\hat{\mu}(x)] \rightarrow \mu(x)$, also die punktweise asymptotische Erwartungstreue von $\hat{\mu}$. Da die Schranke des Bias für alle $x \in [0, 1]$ existiert, gilt die Abschätzung auch bzgl. des Supremums über x , was nichts anderes bedeutet, als dass $\mathbb{E}[\hat{\mu}(x)]$ gleichmäßig gegen $\mu(x)$ konvergiert. Der Schätzer ist also auch gleichmäßig asymptotisch erwartungstreu.

Beweis.

Wir bilden für den Bias den Erwartungswert des Schätzers. Für alle $x \in [0, 1]$ gilt über Proposition 3.7:

$$\begin{aligned}
 \mathbb{E}[\hat{\mu}(x)] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij}\right] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \mathbb{E}[Y_{ij}] \quad (4.1) \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \mathbb{E}[\mu(x_j) + Z_i(x_j) + \varepsilon_{ij}] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + \underbrace{\mathbb{E}[Z_i(x_j)]}_{=0} + \underbrace{\mathbb{E}[\varepsilon_{ij}]}_{=0}) \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \mu(x_j) = \sum_{j=1}^p W_j(x) \mu(x_j) = \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) \\
 &\Rightarrow \text{Bias}(\hat{\mu}(x)) = \mathbb{E}[\hat{\mu}(x)] - \mu(x) = \mathcal{O}(h^{1+\alpha_\mu})
 \end{aligned}$$

Die Varianz berechnen wir über den *Verschiebungssatz* und bestimmen dafür zunächst $\mathbb{E}[(\hat{\mu}(x))^2]$.

$$\begin{aligned}
 \mathbb{E}[(\hat{\mu}(x))^2] &= \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij}\right)^2\right] = \mathbb{E}\left[\frac{1}{n^2} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) Y_{i_1 j_1} W_{j_2}(x) Y_{i_2 j_2}\right] \\
 &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \mathbb{E}[Y_{i_1 j_1} Y_{i_2 j_2}] \\
 &\stackrel{(*)}{=} \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \left(\sum_{i=1}^n \mathbb{E}[Y_{ij_1} Y_{ij_2}] + \sum_{i_1 \neq i_2=1}^n \mathbb{E}[Y_{i_1 j_1}] \mathbb{E}[Y_{i_2 j_2}] \right) \\
 &= \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \left(\sum_{i=1}^n \mathbb{E}[Y_{ij_1} Y_{ij_2}] + \sum_{i_1 \neq i_2=1}^n \mu(x_{j_1}) \mu(x_{j_2}) \right) \\
 &= \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \left(\sum_{i=1}^n \mathbb{E}[Y_{ij_1} Y_{ij_2}] + n(n-1) \mu(x_{j_1}) \mu(x_{j_2}) \right)
 \end{aligned}$$

(*) gilt, da sich die Unabhängigkeit von Z und ε bzgl. i auf Y überträgt. Weiterhin gilt:

$$\begin{aligned}
 \mathbb{E}[Y_{i j_1} Y_{i j_2}] &= \mathbb{E}[(\mu(x_{j_1}) + Z_i(x_{j_1}) + \varepsilon_{i j_1})(\mu(x_{j_2}) + Z_i(x_{j_2}) + \varepsilon_{i j_2})] \\
 &= \mathbb{E}[\mu(x_{j_1}) \mu(x_{j_2}) + \mu(x_{j_1}) Z_i(x_{j_2}) + \mu(x_{j_1}) \varepsilon_{i j_2} + Z_i(x_{j_1}) \mu(x_{j_2}) + Z_i(x_{j_1}) Z_i(x_{j_2}) \\
 &\quad + Z_i(x_{j_1}) \varepsilon_{i j_2} + \varepsilon_{i j_1} \mu(x_{j_2}) + \varepsilon_{i j_1} Z_i(x_{j_2}) + \varepsilon_{i j_1} \varepsilon_{i j_2}] \\
 &= \mu(x_{j_1}) \mu(x_{j_2}) + \mu(x_{j_1}) \underbrace{\mathbb{E}[Z_i(x_{j_2})]}_{=0} + \mu(x_{j_1}) \underbrace{\mathbb{E}[\varepsilon_{i j_2}]}_{=0} + \underbrace{\mathbb{E}[Z_i(x_{j_1})]}_{=0} \mu(x_{j_2}) + \underbrace{\mathbb{E}[Z_i(x_{j_1}) Z_i(x_{j_2})]}_{=R(x_{j_1}, x_{j_2})} \\
 &\quad + \underbrace{\mathbb{E}[Z_i(x_{j_1})]}_{=0} \underbrace{\mathbb{E}[\varepsilon_{i j_2}]}_{=0} + \underbrace{\mathbb{E}[\varepsilon_{i j_1}]}_{=0} \mu(x_{j_2}) + \underbrace{\mathbb{E}[\varepsilon_{i j_1}]}_{=0} \underbrace{\mathbb{E}[Z_i(x_{j_2})]}_{=0} + \underbrace{\mathbb{E}[\varepsilon_{i j_1} \varepsilon_{i j_2}]}_{=\sigma_{j_1 j_2}} \\
 &= \mu(x_{j_1}) \mu(x_{j_2}) + R(x_{j_1}, x_{j_2}) + \sigma_{j_1 j_2}
 \end{aligned}$$

Daraus folgt mit Proposition 3.7 und 3.9:

$$\begin{aligned}
\Rightarrow \mathbb{E}[(\hat{\mu}(x))^2] &= \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x)W_{j_2}(x) \left(\sum_{i=1}^n (\mu(x_{j_1})\mu(x_{j_2}) + R(x_{j_1}, x_{j_2}) + \sigma_{j_1 j_2}) + n(n-1)\mu(x_{j_1})\mu(x_{j_2}) \right) \\
&= \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x)W_{j_2}(x) (n(\mu(x_{j_1})\mu(x_{j_2}) + R(x_{j_1}, x_{j_2}) + \sigma_{j_1 j_2}) + n(n-1)\mu(x_{j_1})\mu(x_{j_2})) \\
&= \frac{1}{n^2} \sum_{j_1, j_2=1}^p W_{j_1}(x)W_{j_2}(x) (n^2\mu(x_{j_1})\mu(x_{j_2}) + nR(x_{j_1}, x_{j_2}) + n\sigma_{j_1 j_2}) \\
&= \left(\sum_{j=1}^p W_j(x)\mu(x_j) \right)^2 + \frac{1}{n} \sum_{j_1, j_2=1}^p W_{j_1}(x)W_{j_2}(x)R(x_{j_1}, x_{j_2}) + \frac{1}{n} \sum_{j_1, j_2=1}^p W_{j_1}(x)W_{j_2}(x)\sigma_{j_1 j_2} \\
&= \left(\sum_{j=1}^p W_j(x)\mu(x_j) \right)^2 + \frac{R(x, x) + \mathcal{O}(h^{1+\alpha_R})}{n} + \frac{\vec{W}(x)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(x)}{n} \\
&\stackrel{\substack{|R(\cdot, \cdot)| \leq E_{Z_2} \\ \text{Proposition 3.3}}}{=} \left(\sum_{j=1}^p W_j(x)\mu(x_j) \right)^2 + \frac{\mathcal{O}(1 + h^{1+\alpha_R})}{n} + \frac{\mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right)}{n} \\
&= \left(\sum_{j=1}^p W_j(x)\mu(x_j) \right)^2 + \mathcal{O}\left(\frac{1 + h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right)
\end{aligned}$$

$$\stackrel{\text{Gleichung (4.1)}}{\Rightarrow} \text{Var}(\hat{\mu}(x)) = \mathbb{E}[(\hat{\mu}(x))^2] - (\mathbb{E}[\hat{\mu}(x)])^2 = \mathcal{O}\left(\frac{1 + h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right)$$

□

Bemerkung 4.2.

Unter Annahme einer Hölderstetigkeit der (partiellen) Ableitungen bis zur Ordnung m von μ und R , und Reproduzierbarkeit von Polynomen bis zum Grad m des Schätzers - siehe Bemerkung 3.2 und 3.5 - ergeben sich für Satz 4.1 die Abschätzungen

$$\begin{aligned}
\text{Bias}(\hat{\mu}(x)) &= \mathcal{O}(h^{m+\alpha_\mu}), \quad \text{Var}(\hat{\mu}(x)) = \mathcal{O}\left(\frac{1 + h^{m+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \\
\Rightarrow \text{MSE}(\hat{\mu}(x)) &= \mathcal{O}\left(h^{2(m+\alpha_\mu)} + \frac{1 + h^{m+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right)
\end{aligned}$$

Zur Begründung, siehe Bemerkung 3.8.

4.1.2. Kovarianzschätzer

Satz 4.3.

Sei \hat{R} der Schätzer aus Gleichung (3.4) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt. Dann gilt für alle $(s, t) \in [0, 1]^2$:

$$\begin{aligned} \text{Bias}(\hat{R}(s, t)) &= \mathcal{O}\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h} + \frac{1}{n}\right), \quad \text{Var}(\hat{R}(s, t)) = o(1) \\ \Rightarrow \text{MSE}(\hat{R}(s, t)) &= \mathcal{O}\left(\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h} + \frac{1}{n}\right)^2\right) + o(1) \end{aligned}$$

Bei der korrigierten Version des Schätzers aus Gleichung (3.5), würde der Faktor $\frac{1}{n}$ im Bias entfallen, die anderen beiden Summanden entstehen durch die Glättung und den Fehlerterm, wie wir gleich im Beweis sehen werden. Aus $h \rightarrow 0$ und unter der zusätzlichen Annahme, dass $p^{1-d_1}h \rightarrow \infty$ - welche wir im Übrigen in Abschnitt 4.2 so und so brauchen werden - folgt $\text{Bias}(\hat{R}(s, t)) \rightarrow 0$, und damit auch unmittelbar die punktweise asymptotische Erwartungstreue von \hat{R} . Da die Schranke des Bias für alle $(s, t) \in [0, 1]^2$ existiert, gilt die Abschätzung auch bzgl. des Supremums über (s, t) ; der Schätzer ist also auch gleichmäßig asymptotisch erwartungstreu. Auch hier wäre unter den erweiterten Annahmen aus Bemerkung 3.2 und 3.5 analog zu Bemerkung 4.2 eine schnellere Konvergenzrate erzielbar, wieder durch den größeren Exponenten $m + \alpha_R$ von h .

Beweis.

Zunächst überlegen wir uns, dass für alle $x \in [0, 1]$ und alle $i = 1, \dots, n$ gilt:

$$\begin{aligned} \hat{\mu}_i(x) - \hat{\mu}(x) &= \sum_{j=1}^p W_j(x) Y_{ij} - \sum_{j=1}^p W_j(x) \frac{1}{n} \sum_{l=1}^n Y_{lj} = \sum_{j=1}^p W_j(x) Y_{ij} - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_j(x) Y_{lj} \\ &= \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i(x_j) + \varepsilon_{ij}) - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_l(x_j) + \varepsilon_{lj}) \\ &= \sum_{j=1}^p W_j(x) \underbrace{\left(\mu(x_j) - \frac{1}{n} \sum_{l=1}^n \mu(x_j) \right)}_{= 0} \\ &\quad + \underbrace{\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(x) Z_i(x_j) - W_j(x) Z_l(x_j))}_{=: WZ_i(x)} + \underbrace{\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(x) \varepsilon_{ij} - W_j(x) \varepsilon_{lj})}_{=: W\varepsilon_i(x)} \\ &= WZ_i(x) + W\varepsilon_i(x) \end{aligned} \tag{4.2}$$

Die Unabhängigkeit von Z und ε überträgt sich auf WZ und $W\varepsilon$. Damit kann der Schätzer \hat{R} für alle $(s, t) \in [0, 1]^2$ wie folgt zerlegt werden.

$$\begin{aligned}
 \hat{R}(s, t) &= \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) = \frac{1}{n} \sum_{i=1}^n (W Z_i(s) + W \varepsilon_i(s)) (W Z_i(t) + W \varepsilon_i(t)) \\
 &= \frac{1}{n} \sum_{i=1}^n (W Z_i(s) W Z_i(t) + W Z_i(s) W \varepsilon_i(t) + W Z_i(t) W \varepsilon_i(s) + W \varepsilon_i(s) W \varepsilon_i(t)) \\
 \Rightarrow \mathbb{E}[\hat{R}(s, t)] &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[W Z_i(s) W Z_i(t)] + \mathbb{E}[W Z_i(s)] \mathbb{E}[W \varepsilon_i(t)] + \mathbb{E}[W Z_i(t)] \mathbb{E}[W \varepsilon_i(s)] \\
 &\quad + \mathbb{E}[W \varepsilon_i(s) W \varepsilon_i(t)]) \tag{4.3}
 \end{aligned}$$

Wir untersuchen nun die einzelnen Summanden. Mit Proposition 3.7 gilt:

$$\begin{aligned}
 &\mathbb{E} \left[\left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s) Z_i(x_j) - W_j(s) Z_l(x_j)) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t) Z_i(x_k) - W_k(t) Z_l(x_k)) \right) \right] \\
 &= \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) \underbrace{\mathbb{E}[Z_i(x_j) Z_i(x_k)]}_{= R(x_j, x_k)} - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) \underbrace{\mathbb{E}[Z_i(x_j) Z_{l_2}(x_k)]}_{= 0 \text{ für } l_2 \neq i} \\
 &\quad - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) \underbrace{\mathbb{E}[Z_{l_1}(x_j) Z_i(x_k)]}_{= 0 \text{ für } l_1 \neq i} + \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) \underbrace{\mathbb{E}[Z_{l_1}(x_j) Z_{l_2}(x_k)]}_{= 0 \text{ für } l_1 \neq l_2} \\
 &= \sum_{j, k=1}^p W_j(s) W_k(t) R(x_j, x_k) - \frac{2}{n} \sum_{j, k=1}^p W_j(s) W_k(t) R(x_j, x_k) + \frac{1}{n^2} \sum_{l=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) R(x_j, x_k) \\
 &= R(s, t) + \mathcal{O}(h^{1+\alpha_R}) - \frac{2}{n} (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) + \frac{1}{n^2} \sum_{l=1}^n (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) \\
 &= \frac{n-1}{n} (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) \tag{4.4}
 \end{aligned}$$

Auf analogem Rechenweg folgt für den vierten Summanden über Proposition 3.9:

$$\begin{aligned}
 &\mathbb{E} \left[\left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s) \varepsilon_{ij} - W_j(s) \varepsilon_{lj}) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t) \varepsilon_{ik} - W_k(t) \varepsilon_{lk}) \right) \right] \\
 &= \sum_{j, k=1}^p W_j(s) W_k(t) \sigma_{jk} - \frac{2}{n} \sum_{j, k=1}^p W_j(s) W_k(t) \sigma_{jk} + \frac{1}{n^2} \sum_{l=1}^n \sum_{j, k=1}^p W_j(s) W_k(t) \sigma_{jk} \\
 &= \vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t) - \frac{2}{n} \vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t) + \frac{1}{n^2} \sum_{l=1}^n \vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t) \\
 &= \frac{n-1}{n} \cdot \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \tag{4.5}
 \end{aligned}$$

Für den zweiten Summanden gilt:

$$\begin{aligned}
 & \mathbb{E} \left[\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s)Z_i(x_j) - W_j(s)Z_l(x_j)) \right] \mathbb{E} \left[\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t)\varepsilon_{ik} - W_k(t)\varepsilon_{lj}) \right] \\
 &= \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s)\mathbb{E}[Z_i(x_j)] - W_j(s)\mathbb{E}[Z_l(x_j)]) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t)\mathbb{E}[\varepsilon_{ik}] - W_k(t)\mathbb{E}[\varepsilon_{lj}]) \right) \\
 &= \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s) \cdot 0 - W_j(s) \cdot 0) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t) \cdot 0 - W_k(t) \cdot 0) \right) = 0 \quad (4.6)
 \end{aligned}$$

Selbiges gilt für den dritten Summanden. Da keiner der Summanden mehr von i abhängig ist, folgt insgesamt für den Bias des Schätzers über Gleichung (4.3):

$$\begin{aligned}
 \mathbb{E}[\hat{R}(s, t)] - R(s, t) &= \frac{n-1}{n} \left(R(s, t) + \mathcal{O}(h^{1+\alpha_R}) + \mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) \right) \cdot \underbrace{\frac{1}{n} \sum_{i=1}^n 1 - R(s, t)}_{=1} \\
 &= \frac{n-1}{n} \left(\mathcal{O}(h^{1+\alpha_R}) + \mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) \right) - \frac{1}{n} R(s, t) \\
 &\stackrel{|R(\cdot, \cdot)| \leq E_{Z,2}}{\text{Proposition 3.3}} \underbrace{\frac{n-1}{n} \left(\mathcal{O}(h^{1+\alpha_R}) + \mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) \right)}_{= \mathcal{O}(1)} + \mathcal{O}\left(\frac{1}{n}\right) \\
 &= \mathcal{O}\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h} + \frac{1}{n}\right) \quad (4.7)
 \end{aligned}$$

Der Schätzer ist also zum einen über den „falschen“ Vorfaktor $\frac{1}{n}$ durch $\mathcal{O}\left(\frac{1}{n}\right)$ verzerrt, was sich aber leicht über den „richtigen“ Vorfaktor $\frac{1}{n-1}$ korrigieren lässt. Zum anderen über die Glättung durch $\mathcal{O}(h^{1+\alpha_R})$ - was logisch ist, da man nur an endlich vielen Stellen beobachten/approximieren kann - und über den Fehler durch $\mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right)$.

Die Berechnung der Varianz gestaltet sich hier deutlich komplexer als bei $\hat{\mu}$. Wir zeigen daher nur relativ allgemein, dass auch hier die Varianz im Unendlichen verschwindet, ohne eine genauere Rate anzugeben und werden auch an einigen Stellen, mit entsprechender Begründung, abkürzen.

Nach Gleichung (4.3) und den darauffolgenden Rechenschritten gilt:

$$\begin{aligned}
 \left(\mathbb{E}[\hat{R}(s, t)]\right)^2 &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \left(\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)]\mathbb{E}[WZ_{i_2}(s)WZ_{i_2}(t)] \right. \\
 &\quad + 2\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)]\mathbb{E}[W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] \\
 &\quad \left. + \mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_1}(t)]\mathbb{E}[W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] \right) \\
 (\hat{R}(s, t))^2 &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \left(WZ_{i_1}(s)WZ_{i_1}(t) + WZ_{i_1}(s)W\varepsilon_{i_1}(t) + WZ_{i_1}(t)W\varepsilon_{i_1}(s) + W\varepsilon_{i_1}(s)W\varepsilon_{i_1}(t) \right) \\
 &\quad \cdot \left(WZ_{i_2}(s)WZ_{i_2}(t) + WZ_{i_2}(s)W\varepsilon_{i_2}(t) + WZ_{i_2}(t)W\varepsilon_{i_2}(s) + W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t) \right) \tag{4.8}
 \end{aligned}$$

Bildet man davon den Erwartungswert, können wegen der Unabhängigkeit von Z und ε und wegen $\mathbb{E}[WZ_i(\cdot)] = \mathbb{E}[W\varepsilon_i(\cdot)] = 0$ für alle $i = 1, \dots, n$ die Summanden, bei denen WZ und $W\varepsilon$ einzeln auftritt, ausgeblendet werden. Es verbleibt:

$$\begin{aligned}
 \mathbb{E}[(\hat{R}(s, t))^2] &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \left(\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)WZ_{i_2}(s)WZ_{i_2}(t)] + \mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_1}(t)W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] \right. \\
 &\quad + \mathbb{E}[WZ_{i_1}(s)WZ_{i_2}(s)]\mathbb{E}[W\varepsilon_{i_1}(t)W\varepsilon_{i_2}(t)] + \mathbb{E}[WZ_{i_1}(t)WZ_{i_2}(t)]\mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_2}(s)] \\
 &\quad \left. + 2\mathbb{E}[WZ_{i_1}(s)WZ_{i_2}(t)]\mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_2}(t)] + 2\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)]\mathbb{E}[W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] \right)
 \end{aligned}$$

Damit gilt für die Varianz über den *Verschiebungssatz*:

$$\begin{aligned}
 \text{Var}(\hat{R}(s, t)) &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \left(\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)WZ_{i_2}(s)WZ_{i_2}(t)] - \mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)]\mathbb{E}[WZ_{i_2}(s)WZ_{i_2}(t)] \right. \\
 &\quad + \mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_1}(t)W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] - \mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_1}(t)]\mathbb{E}[W\varepsilon_{i_2}(s)W\varepsilon_{i_2}(t)] \\
 &\quad + \mathbb{E}[WZ_{i_1}(s)WZ_{i_2}(s)]\mathbb{E}[W\varepsilon_{i_1}(t)W\varepsilon_{i_2}(t)] + \mathbb{E}[WZ_{i_1}(t)WZ_{i_2}(t)]\mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_2}(s)] \\
 &\quad \left. + 2\mathbb{E}[WZ_{i_1}(s)WZ_{i_2}(t)]\mathbb{E}[W\varepsilon_{i_1}(s)W\varepsilon_{i_2}(t)] \right) \tag{4.9}
 \end{aligned}$$

Nun wäre es immens von Vorteil, wenn $WZ_{i_1}(\cdot)$ und $WZ_{i_2}(\cdot)$ bzw. $W\varepsilon_{i_1}(\cdot)$ und $W\varepsilon_{i_2}(\cdot)$ für $i_1 \neq i_2$ stochastisch unabhängig wären, weil man dann die Erwartungswerte auseinanderziehen kann. Leider ist das nicht der Fall, weil in diesen Termen jeweils die Summe über alle Z_i bzw. ε_i auftritt, somit sind tatsächlich alle $WZ_{i_1}(\cdot)$ und $WZ_{i_2}(\cdot)$ bzw. $W\varepsilon_{i_1}(\cdot)$ und $W\varepsilon_{i_2}(\cdot)$ für alle $i_1, i_2 = 1, \dots, n$ abhängig. Wir dürfen die Erwartungswerte trotzdem für $i_1 \neq i_2$ auseinanderziehen und mit einem $+o(1)$ versehen, weil sich der dadurch entstehende Ausdruck zum tatsächlichen Ausdruck nur durch einen additiven Term unterscheidet, der wiederum asymptotisch verschwindet.

Wir betrachten dazu o. B. d. A. den Term $\mathbb{E}[WZ_{i_1}(s)WZ_{i_2}(t)]$. Für alle $(s, t) \in [0, 1]^2$ und $i_1 \neq i_2$ gilt:

$$\begin{aligned}
 & \mathbb{E} \left[\left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s)Z_{i_1}(x_j) - W_j(s)Z_l(x_j)) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t)Z_{i_2}(x_k) - W_k(t)Z_l(x_k)) \right) \right] \\
 &= \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t) \underbrace{\mathbb{E}[Z_{i_1}(x_j)Z_{i_2}(x_k)]}_{= 0 \text{ da } i_1 \neq i_2} - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t) \underbrace{\mathbb{E}[Z_{i_1}(x_j)Z_{l_2}(x_k)]}_{= 0 \text{ für } l_2 \neq i} \\
 &\quad - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t) \underbrace{\mathbb{E}[Z_{l_1}(x_j)Z_{i_2}(x_k)]}_{= 0 \text{ für } l_1 \neq i} + \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t) \underbrace{\mathbb{E}[Z_{l_1}(x_j)Z_{l_2}(x_k)]}_{= 0 \text{ für } l_1 \neq l_2} \\
 &= -\frac{2}{n} \sum_{j, k=1}^p W_j(s)W_k(t)R(x_j, x_k) + \frac{1}{n^2} \sum_{l=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)R(x_j, x_k) = \underbrace{-\frac{1}{n}}_{\rightarrow 0} \left(\underbrace{R(s, t)}_{|\cdot| < E_{Z,2} < \infty} + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} \right) \\
 &= o(1) \stackrel{\text{Gleichung (4.6)}}{=} \mathbb{E}[WZ_{i_1}(s)]\mathbb{E}[WZ_{i_2}(t)] + o(1)
 \end{aligned}$$

Nach dem gleichen Prinzip lassen sich alle anderen Erwartungswerte der zweifachen Produkte diskutieren. Für die vierfachen Produkte wäre es ein zu großer Aufwand, es explizit zu beweisen, da man hier insgesamt 16 Summanden überprüfen müsste, es würde aber genauso asymptotisch nur noch der erste Summand eine Rolle spielen. Im Fall von $\mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)WZ_{i_2}(s)WZ_{i_2}(t)]$ wäre das:

$$\begin{aligned}
 & \frac{1}{n^4} \sum_{\substack{l_1, l_2=1 \\ l_3, l_4=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t) \mathbb{E}[Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})] \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t) \mathbb{E}[Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})] \mathbb{E}[Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})] \\
 &= \left(\sum_{j_1, k_1=1}^p W_{j_1}(s)W_{k_1}(t)R(x_{j_1}, x_{k_1}) \right) \left(\sum_{j_2, k_2=1}^p W_{j_2}(s)W_{k_2}(t)R(x_{j_2}, x_{k_2}) \right) \\
 &= (R(s, t) + \mathcal{O}(h^{1+\alpha_R}))^2 \stackrel{\text{Gleichung (4.4)}}{=} \underbrace{\frac{n^2}{(n-1)^2}}_{\rightarrow 1} \mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)] \mathbb{E}[WZ_{i_2}(s)WZ_{i_2}(t)] \\
 &= \mathbb{E}[WZ_{i_1}(s)WZ_{i_1}(t)] \mathbb{E}[WZ_{i_2}(s)WZ_{i_2}(t)] + o(1) \tag{4.10}
 \end{aligned}$$

Alle weiteren Summanden haben mindestens den Vorfaktor $\frac{1}{n}$ und wären damit alle, wegen der Endlichkeit der 2. bzw. 4. Verbundmomente von Z und ε (siehe Proposition 3.3), auch $o(1)$.

D. h. es gilt für alle $(s, t) \in [0, 1]^2$ und $i_1 \neq i_2$:

$$\begin{aligned}
 \mathbb{E}[W Z_{i_1}(s)W Z_{i_1}(t)W Z_{i_2}(s)W Z_{i_2}(t)] &= \mathbb{E}[W Z_{i_1}(s)W Z_{i_1}(t)]\mathbb{E}[W Z_{i_2}(s)W Z_{i_2}(t)] + o(1) \\
 \mathbb{E}[W \varepsilon_{i_1}(s)W \varepsilon_{i_1}(t)W \varepsilon_{i_2}(s)W \varepsilon_{i_2}(t)] &= \mathbb{E}[W \varepsilon_{i_1}(s)W \varepsilon_{i_1}(t)]\mathbb{E}[W \varepsilon_{i_2}(s)W \varepsilon_{i_2}(t)] + o(1) \\
 \mathbb{E}[W Z_{i_1}(s)W Z_{i_2}(t)] &= \underbrace{\mathbb{E}[W Z_{i_1}(s)]}_{=0} \underbrace{\mathbb{E}[W Z_{i_2}(t)]}_{=0} + o(1) = o(1) \\
 \mathbb{E}[W \varepsilon_{i_1}(s)W \varepsilon_{i_2}(t)] &= \underbrace{\mathbb{E}[W \varepsilon_{i_1}(s)]}_{=0} \underbrace{\mathbb{E}[W \varepsilon_{i_2}(t)]}_{=0} + o(1) = o(1) \quad (4.11)
 \end{aligned}$$

und somit für die Varianz nach Gleichung (4.9):

$$\begin{aligned}
 \text{Var}(\hat{R}(s, t)) &= \frac{1}{n^2} \sum_{i=1}^n \left(\mathbb{E}[W Z_i(s)W Z_i(t)W Z_i(s)W Z_i(t)] - (\mathbb{E}[W Z_i(s)W Z_i(t)])^2 \right. \\
 &\quad \left. + \mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)W \varepsilon_i(s)W \varepsilon_i(t)] - (\mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)])^2 \right. \\
 &\quad \left. + \mathbb{E}[W Z_i(s)W Z_i(s)]\mathbb{E}[W \varepsilon_i(t)W \varepsilon_i(t)] + \mathbb{E}[W Z_i(t)W Z_i(t)]\mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(s)] \right. \\
 &\quad \left. + 2\mathbb{E}[W Z_i(s)W Z_i(t)]\mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)] \right) + o(1)
 \end{aligned}$$

Wir führen nun eine, zugegebenermaßen sehr großzügige Abschätzung durch, die zumindest darlegt, dass der Term im Unendlichen verschwindet. Dies ist genau dann der Fall, wenn die komplette Summe für $n \rightarrow \infty$ mit einer geringeren Rate als n^2 gegen Unendlich konvergiert. Wir schätzen die innere Summe der Erwartungswerte nach oben ab. Die Erwartungswerte der zweifachen Produkte haben wir in Gleichung (4.4) und (4.5) schon berechnet, damit folgt für alle $(s, t) \in [0, 1]^2$ und $i = 1, \dots, n$ mit Proposition 3.3 und 3.6 und der Δ -Ungleichung:

$$\begin{aligned}
 \mathbb{E}[W Z_i(s)W Z_i(t)] &= \underbrace{\frac{n-1}{n} \sum_{j,k=1}^p W_j(s)W_k(t)R(x_j, x_k)}_{\leq 1} \stackrel{|\cdot|}{\leq} \sum_{j,k=1}^p |W_j(s)||W_k(t)||R(x_j, x_k)| \\
 &\leq E_{Z,2} \left(\sum_{j=1}^p |W_j(s)| \right) \left(\sum_{k=1}^p |W_k(t)| \right) \leq E_{Z,2} C_{W_1}^2 < \infty \\
 \mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)] &= \underbrace{\frac{n-1}{n} \sum_{j,k=1}^p W_j(s)W_k(t)\sigma_{jk}}_{\leq 1} \stackrel{|\cdot|}{\leq} \dots \leq E_{\varepsilon,2}(p) C_{W_1}^2 < \infty
 \end{aligned}$$

Für die vierfachen Produkte müssten wir nun eigentlich wieder 16 Summanden auswerten. Das ist aber nicht nötig, da nach gleicher Argumentation wie bei Gleichung (4.10) nur der erste Summand eine Rolle spielt, weil wieder alle weiteren Summanden mindestens den Vorfaktor $\frac{1}{n}$ besitzen und somit asymptotisch verschwinden. Für die jeweils ersten Summanden von $\mathbb{E}[W Z_i(s)W Z_i(t)W Z_i(s)W Z_i(t)]$ und $\mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)W \varepsilon_i(s)W \varepsilon_i(t)]$ gilt nach den gleichen Voraussetzungen wie oben mit Proposition 3.3:

$$\begin{aligned}
 & \frac{1}{n^4} \sum_{\substack{l_1, l_2=1 \\ l_3, l_4=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t)\mathbb{E}[Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})Z_i(x_{k_2})] \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t)M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 &\stackrel{|\cdot|}{\leq} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p |W_{j_1}(s)||W_{k_1}(t)||W_{j_2}(s)||W_{k_2}(t)||M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2})| \\
 &\leq E_{Z,4} \left(\sum_{j_1=1}^p |W_{j_1}(s)| \right) \left(\sum_{k_1=1}^p |W_{k_1}(t)| \right) \left(\sum_{j_2=1}^p |W_{j_2}(s)| \right) \left(\sum_{k_2=1}^p |W_{k_2}(t)| \right) \leq E_{Z,4} C_{W_1}^4 < \infty \\
 & \frac{1}{n^4} \sum_{\substack{l_1, l_2=1 \\ l_3, l_4=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t)\mathbb{E}[\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2}] \stackrel{|\cdot|}{\leq} \dots \leq E_{\varepsilon,4}(p)C_{W_1}^4 < \infty \\
 & \Rightarrow \mathbb{E}[W Z_i(s)W Z_i(t)W Z_i(s)W Z_i(t)] \stackrel{|\cdot|}{\leq} E_{Z,4}C_{W_1}^4 + o(1) < \infty \\
 & \mathbb{E}[W \varepsilon_i(s)W \varepsilon_i(t)W \varepsilon_i(s)W \varepsilon_i(t)] \stackrel{|\cdot|}{\leq} E_{\varepsilon,4}(p)C_{W_1}^4 + o(1) < \infty
 \end{aligned}$$

Insgesamt folgt jetzt für die Varianz:

$$\begin{aligned}
 \text{Var}(\hat{R}(s, t)) &\stackrel{|\cdot|}{\leq} \frac{1}{n^2} \sum_{i=1}^n \left(E_{Z,4}C_{W_1}^4 + o(1) + E_{Z,2}^2C_{W_1}^4 + E_{\varepsilon,4}(p)C_{W_1}^4 + o(1) + E_{\varepsilon,2}(p)^2C_{W_1}^4 + 4E_{Z,2}E_{\varepsilon,2}(p)C_{W_1}^4 \right) + o(1) \\
 &= \frac{E_{Z,4}C_{W_1}^4 + E_{Z,2}^2C_{W_1}^4 + E_{\varepsilon,4}(p)C_{W_1}^4 + E_{\varepsilon,2}(p)^2C_{W_1}^4 + 4E_{Z,2}E_{\varepsilon,2}(p)C_{W_1}^4}{\underbrace{\phantom{E_{Z,4}C_{W_1}^4 + E_{Z,2}^2C_{W_1}^4 + E_{\varepsilon,4}(p)C_{W_1}^4 + E_{\varepsilon,2}(p)^2C_{W_1}^4 + 4E_{Z,2}E_{\varepsilon,2}(p)C_{W_1}^4}}_n} + o(1) = o(1)
 \end{aligned} \tag{4.12}$$

Der letzte Gleichungsschritt folgt aus der Endlichkeit des Zählers. \square

4.2. Asymptotische Normalität

Wir weisen nun für beide Schätzer die asymptotische Normalität nach, d. h. die Konvergenz in Verteilung der standardisierten Schätzer gegen eine Normalverteilung bzw., da es sich um funktionale Schätzer handelt, gegen einen Gauß-Prozess (Trend) und ein Gauß-Feld (Kovarianz). Wegen der Funktionalität, handelt es sich auch nicht um Verteilungskonvergenz von Zufallsvariablen im klassischen Sinne, sondern um die allgemeinere schwache Konvergenz innerhalb des Banachraums $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$, siehe Definition 2.11 d).

4.2.1. Trendschätzer

Satz 4.4.

Sei $\hat{\mu}$ der Schätzer aus Gleichung (3.2) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelte $\sqrt{nh}^{1+\alpha_\mu} \rightarrow 0$ sowie $\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty$.

Dann gilt:

$$\sqrt{n}(\hat{\mu}(x) - \mu(x)) \xrightarrow{w} \mathcal{G}(0, R) \text{ in } (\mathcal{C}([0, 1]), \|\cdot\|_\infty)$$

$\mathcal{G}(0, R)$ ist ein eindeutig bestimmter Gauß-Prozess mit Erwartungswert 0 und Kovarianzfunktion R , wobei R der Kovarianzfunktion von Z aus (A.3) entspricht.

Es sei hier noch einmal erwähnt, dass p implizit von n , und h implizit von n und p abhängt, sonst wären die geforderten Voraussetzungen gar nicht möglich. Ein Beleg dafür, dass die Konvergenzen hiervon erfüllbar sind, findet sich in Anhang A.3.

Beweis.

Den Beweis haben im Wesentlichen [Deg11] sowie [Her17] durchgeführt, allerdings wurde dort von einer mehrdimensionalen Normalverteilung des Fehler ε ausgegangen, was bei uns nicht gegeben ist. Daher werden nur nachweisen, dass unter den gegebenen Voraussetzungen, die Verteilung von ε keine Rolle für die Prozesskonvergenz spielt.

Mit der letzten Zeile von Gleichung (4.1) lässt sich der empirische Prozess wie folgt zerlegen:

$$\sqrt{n}(\hat{\mu}(x) - \mu(x)) = \underbrace{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Z_i(x_j) \right)}_{=: S_n^1(x)} + \underbrace{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \right)}_{=: S_n^2(x)} + \underbrace{\sqrt{n} \cdot \mathcal{O}(h^{1+\alpha_\mu})}_{=: S_n^3(x)}$$

Die schwache Konvergenz kann nun nach Lemma 2.15 über die Konvergenz der einzelnen Summanden bestimmt werden. Für den ersten Summanden gilt $S_n^1(x) \xrightarrow{w} \mathcal{G}(0, R)$ in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$, siehe dafür [Her17]. Für den dritten Summanden gilt nach den Voraussetzungen aus Satz 4.4:

$$S_n^3(x) = \mathcal{O}(\sqrt{nh}^{1+\alpha_\mu}), \forall x \in [0, 1] \Rightarrow \sup_{x \in [0, 1]} |S_n^3(x)| = \mathcal{O}(\sqrt{nh}^{1+\alpha_\mu}) \rightarrow 0$$

Für $S_n^2(x)$ muss die Konvergenz in \mathbb{P} gegen 0 natürlich auch bzgl. der Supremumsnorm, also gleichmäßig gezeigt werden. Wir kombinieren daher nun die punktweise Konvergenz in \mathbb{P} mit der gleichmäßig gleichgradigen Stetigkeit in \mathbb{P} , wie in Lemma 2.18 erwähnt.

Punktweise Konvergenz in \mathbb{P}

Wir wenden hier Proposition 3.9 an. Sei $x \in [0, 1]$.

$$\begin{aligned}
 \mathbb{E}\left[\left(S_n^2(x)\right)^2\right] &= \mathbb{E}\left[\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^pW_j(x)\varepsilon_{ij}\right)\right)^2\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i_1,i_2=1}^n\sum_{j_1,j_2=1}^pW_{j_1}(x)W_{j_2}(x)\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\right] \\
 &= \frac{1}{n}\sum_{i_1,i_2=1}^n\sum_{j_1,j_2=1}^pW_{j_1}(x)W_{j_2}(x)\mathbb{E}[\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}] \stackrel{\text{iid}}{=} \frac{1}{n}\sum_{i=1}^n\sum_{j_1,j_2=1}^pW_{j_1}(x)W_{j_2}(x)\mathbb{E}[\varepsilon_{1j_1}\varepsilon_{1j_2}] \\
 &= \sum_{j_1,j_2=1}^pW_{j_1}(x)W_{j_2}(x)\sigma_{j_1j_2} = \vec{W}(x)^T \cdot \text{Cov}(\varepsilon) \cdot \vec{W}(x) = \mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) \rightarrow 0
 \end{aligned} \tag{4.13}$$

Dies folgt aus der Voraussetzung $\sum_{n=1}^{\infty}\frac{1}{p^{1-d_1}h} < \infty$, denn dies ist nur möglich, falls $\frac{1}{p^{1-d_1}h}$ eine Nullfolge ist.

Gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei $\Phi(x) = \|S_n^2(x)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty}\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^pW_j(x)\varepsilon_{ij}\right)\right)^2}$. Dann gilt offensichtlich für alle $n \in \mathbb{N}$ und $x \in [0, 1]$: $|S_n^2(x)| \leq \Phi(x) \Leftrightarrow |S_n^2| \leq \Phi$. Mit Gleichung (4.13) folgt weiterhin:

$$\begin{aligned}
 \mathbb{E}[|\Phi(x)|^2] &= \mathbb{E}[|S_n^2(x)|_{\ell^2}^2] = \mathbb{E}\left[\sum_{n=1}^{\infty}\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^pW_j(x)\varepsilon_{ij}\right)\right)^2\right] \\
 &= \sum_{n=1}^{\infty}\mathbb{E}\left[\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^pW_j(x)\varepsilon_{ij}\right)\right)^2\right] = \sum_{n=1}^{\infty}\mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) \\
 &\Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}[|\Phi|^2] \leq C \sum_{n=1}^{\infty}\frac{1}{p^{1-d_1}h} < \infty \Rightarrow (\mathbb{E}[|\Phi|^2])^{\frac{1}{2}} < \infty
 \end{aligned}$$

zu ii):

Wir schätzen zunächst einmal die Überdeckungszahl bzgl. der kanonischen Metrik ab. Auf gleichem Rechenweg wie in Gleichung (4.13) folgt mit Proposition 3.9 für alle $x, x' \in [0, 1]$:

$$\begin{aligned}
 d_n^2(x, x') &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \right) - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x') \varepsilon_{ij} \right) \right)^2 \right] \\
 &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_j(x) - W_j(x')) \varepsilon_{ij} \right) \right)^2 \right] \\
 &= (\vec{W}(x) - \vec{W}(x'))^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot (\vec{W}(x) - \vec{W}(x')) = \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\
 \Rightarrow \exists C \in \mathbb{R}^+ : d_n^2(x, x') &\leq \frac{C}{p^{1-d_1} h} \Rightarrow d_n(x, x') \leq \left(\frac{C}{p^{1-d_1} h} \right)^{\frac{1}{2}} =: C_n \tag{4.14}
 \end{aligned}$$

Die Bezeichnung C_n ist bewusst so gewählt, weil sowohl p , als auch h implizit von n abhängen. Da alle Punkte $x, x' \in [0, 1]$ einen maximalen Abstand von C_n bzgl. d_n haben, benötigt man für $u < C_n$ maximal $\frac{C_n}{u}$ Kugeln mit Radius u , um das Intervall $[0, 1]$ zu überdecken. Ist $u \geq C_n$, so genügt eine einzige Kugel. Für die Überdeckungszahl $N_{\text{cov}}(u, d_n, [0, 1]) =: N_{\text{cov}}$ bedeutet das nun:

$$\begin{cases} N_{\text{cov}} = 1, & \text{für } u \geq C_n \\ 1 \leq N_{\text{cov}} \leq \frac{C_n}{u}, & \text{für } u < C_n \end{cases}$$

Daraus lässt sich herleiten:

$$N_{\text{cov}} \leq \frac{C_n}{u} \Leftrightarrow \frac{N_{\text{cov}}}{\sqrt{u}} \leq \frac{C_n}{u^{\frac{3}{2}}} \stackrel{\text{m.w.}}{\Leftrightarrow} \sqrt{\cdot} \stackrel{\text{m.w.}}{\Leftrightarrow} \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} \leq \left(\log \left(\frac{C_n}{u^{\frac{3}{2}}} \right) \right)^{\frac{1}{2}}$$

Außerdem gilt für C_n wegen $\frac{1}{p^{1-d_1} h} \rightarrow 0$ und der Stetigkeit der Wurzelfunktion:

$$C_n = \left(\frac{C}{p^{1-d_1} h} \right)^{\frac{1}{2}} \rightarrow 0 \tag{4.15}$$

Es gibt also ein hinreichend großes n , ab welchem $C_n < \delta$, für alle $\delta > 0$. Um im weiteren Verlauf die Wohldefiniertheit diverser Ausdrücke nicht zu verletzen, setzen wir von vornherein $\delta < 1$.

Mit der Monotonie des Integrals folgt für das Überdeckungsintegral:

$$\begin{aligned}
 \int_0^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du &= \int_0^{C_n} \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du + \int_{C_n}^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \\
 &\leq \int_0^{C_n} \left(\log \left(\frac{C_n}{u^{\frac{2}{3}}} \right) \right)^{\frac{1}{2}} du + \int_{C_n}^\delta \left(\log \left(\frac{1}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \\
 &= \sqrt{\frac{3}{2}} \int_0^{C_n} \sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)} du + \frac{1}{\sqrt{2}} \int_{C_n}^\delta \sqrt{-\log(u)} du \quad (4.16)
 \end{aligned}$$

Die beiden Integrale werden nun separat untersucht.

Beim ersten Integral gilt für die Integrationsvariable durch die obere Grenze

$$u \leq C_n \Rightarrow C_n^{-\frac{2}{3}} u \leq C_n^{-\frac{2}{3}} C_n = C_n^{\frac{1}{3}} \rightarrow 0 \quad (4.17)$$

wegen Gleichung (4.15) und der Stetigkeit von $f(x) = x^{\frac{1}{3}}$. Ab einem hinreichend großen n ist daher $C_n^{-\frac{2}{3}} u \leq 1$ und $\sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)}$ wohldefiniert. Das Integral wird zunächst umgeformt zu:

$$\int_0^{C_n} \sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)} du = \int_0^{C_n} \frac{2C_n^{-\frac{2}{3}} u}{2C_n^{-\frac{2}{3}} u} \cdot \frac{-\log \left(C_n^{-\frac{2}{3}} u \right)}{\sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)}} du$$

Wir substituieren nun mit den Funktionen

$$f(x) = -\frac{2x^2 \exp(-x^2)}{C_n^{-\frac{2}{3}}}, \quad \varphi(u) = \sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)} \Rightarrow \varphi'(u) = -\frac{1}{2u \sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)}}$$

$$\begin{aligned}
 \Rightarrow \int_0^{C_n} \frac{2C_n^{-\frac{2}{3}} u}{2C_n^{-\frac{2}{3}} u} \cdot \frac{-\log \left(C_n^{-\frac{2}{3}} u \right)}{\sqrt{-\log \left(C_n^{-\frac{2}{3}} u \right)}} du &= -\frac{2}{C_n^{-\frac{2}{3}}} \int_{\infty}^{\sqrt{-\log(C_n^{\frac{1}{3}})}} u^2 \exp(-u^2) du \\
 &= \frac{2}{C_n^{-\frac{2}{3}}} \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} u^2 \exp(-u^2) du
 \end{aligned}$$

Nach Gleichung (4.17) ist $\sqrt{-\log(C_n^{\frac{1}{3}})}$ für ein hinreichend großes n wohldefiniert.

Mit partieller Integration folgt weiterhin:

$$\begin{aligned}
 \frac{2}{C_n^{-\frac{2}{3}}} \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} u \cdot u \exp(-u^2) du &= \frac{2}{C_n^{-\frac{2}{3}}} \left(\left[u \cdot \frac{-\exp(-u^2)}{2} \right]_{u=\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} - \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \frac{-\exp(-u^2)}{2} du \right) \\
 &= \frac{2}{C_n^{-\frac{2}{3}}} \left(\frac{C_n^{\frac{1}{3}} \sqrt{-\log(C_n^{\frac{1}{3}})}}{2} + \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \frac{\exp(-u^2)}{2} du \right) \\
 &= \frac{C_n \sqrt{-\log(C_n)}}{\sqrt{3}} + \frac{1}{C_n^{-\frac{2}{3}}} \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \exp(-u^2) du \quad (4.18)
 \end{aligned}$$

Für das nun verbleibende Integral gilt wegen $\sqrt{-\log(C_n^{\frac{1}{3}})} \leq u \Leftrightarrow 1 \leq \frac{u}{\sqrt{-\log(C_n^{\frac{1}{3}})}}$ und der Monotonie des Integrals:

$$\begin{aligned}
 \frac{1}{C_n^{-\frac{2}{3}}} \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \exp(-u^2) du &\leq \frac{1}{C_n^{-\frac{2}{3}}} \int_{\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \frac{u \exp(-u^2)}{\sqrt{-\log(C_n^{\frac{1}{3}})}} du = \frac{1}{C_n^{-\frac{2}{3}}} \left[-\frac{\exp(-u^2)}{2\sqrt{-\log(C_n^{\frac{1}{3}})}} \right]_{u=\sqrt{-\log(C_n^{\frac{1}{3}})}}^{\infty} \\
 &= \frac{1}{C_n^{-\frac{2}{3}}} \cdot \frac{C_n^{\frac{1}{3}}}{2\sqrt{-\log(C_n^{\frac{1}{3}})}} = \frac{C_n \sqrt{3}}{2\sqrt{-\log(C_n)}}
 \end{aligned}$$

Mit Gleichung (4.18) folgt schließlich:

$$\int_0^{C_n} \sqrt{-\log\left(C_n^{-\frac{2}{3}} u\right)} du \leq \frac{C_n \sqrt{-\log(C_n)}}{\sqrt{3}} + \frac{C_n \sqrt{3}}{2\sqrt{-\log(C_n)}} \quad (4.19)$$

Nun zum zweiten Integral aus Gleichung (4.16). Das Vorgehen ist in weiten Teilen identisch, wir substituieren mit den Funktionen

$$f(x) = -2x^2 \exp(-x^2), \quad \varphi(u) = \sqrt{-\log(u)} \Rightarrow \varphi'(u) = -\frac{1}{2u\sqrt{-\log(u)}}$$

$$\begin{aligned}
 \Rightarrow \int_{C_n}^{\delta} \sqrt{-\log(u)} du &= \int_{C_n}^{\delta} \frac{2u}{2u} \cdot \frac{-\log(u)}{\sqrt{-\log(u)}} du = -2 \int_{\sqrt{-\log(C_n)}}^{\sqrt{-\log(\delta)}} u^2 \exp(-u^2) du \\
 &= 2 \int_{\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} u^2 \exp(-u^2) du \\
 &= 2 \left(\left[u \cdot \frac{-\exp(-u^2)}{2} \right]_{u=\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} - \int_{\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} \frac{-\exp(-u^2)}{2} du \right) \\
 &= \delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \int_{\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} \exp(-u^2) du \\
 &\stackrel{\sqrt{-\log(\delta)} \leq u}{\leq} \delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \int_{\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} \frac{u \exp(-u^2)}{\sqrt{-\log(\delta)}} du \\
 &= \delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \left[-\frac{\exp(-u^2)}{2\sqrt{-\log(\delta)}} \right]_{u=\sqrt{-\log(\delta)}}^{\sqrt{-\log(C_n)}} \\
 &= \delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \frac{\delta}{2\sqrt{-\log(\delta)}} - \frac{C_n}{2\sqrt{-\log(\delta)}}
 \end{aligned} \tag{4.20}$$

Dabei sind die Wurzelfunktionen wegen $C_n < \delta < 1$ alle wohldefiniert. Mit Gleichung (4.19) und Gleichung (4.20), lässt sich Gleichung (4.16) nun endgültig abschätzen.

$$\begin{aligned}
 \int_0^{\delta} \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du &\leq \sqrt{\frac{3}{2}} \left(\frac{C_n \sqrt{-\log(C_n)}}{\sqrt{3}} + \frac{C_n \sqrt{3}}{2\sqrt{-\log(C_n)}} \right) \\
 &\quad + \frac{1}{\sqrt{2}} \left(\delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \frac{\delta}{2\sqrt{-\log(\delta)}} - \frac{C_n}{2\sqrt{-\log(\delta)}} \right)
 \end{aligned} \tag{4.21}$$

Wir untersuchen nun für jeden Summanden einzeln, was passiert, wenn $n \rightarrow \infty$. Die konstanten Vorfaktoren spielen dabei keine Rolle, weswegen diese teilweise ausgeblendet werden. Die Summanden, welche weder p noch h beinhalten, sind diesbezüglich ebenfalls konstant, diese werden als letztes diskutiert.

Als kurze Vorüberlegung, betrachten wir die Funktion

$$f(x) = x \sqrt{-\log(x)} \tag{4.22}$$

Diese mündet für $x \rightarrow 0$ im unbestimmten Ausdruck $0 \cdot \infty$. Nach der *Regel von de L'Hospital* ergibt sich aber:

$$\begin{aligned} \lim_{x \rightarrow 0} x \sqrt{-\log(x)} &= \lim_{x \rightarrow 0} \frac{\sqrt{-\log(x)}}{\frac{1}{x}} = \lim_{x \rightarrow 0} \frac{(\sqrt{-\log(x)})'}{(\frac{1}{x})'} = \lim_{x \rightarrow 0} \frac{-\frac{1}{2x\sqrt{-\log(x)}}}{-\frac{1}{x^2}} \\ &= \lim_{x \rightarrow 0} \frac{\overbrace{x}^{\rightarrow 0}}{\underbrace{2\sqrt{-\log(x)}}_{\rightarrow \infty}} = 0 \end{aligned}$$

Daraus folgt, mit $C_n \rightarrow 0$, für den ersten und vierten Summanden direkt die Konvergenz gegen 0. Für den zweiten und sechsten Summanden folgt ebenfalls mit Stetigkeitsargumenten schnell:

$$\frac{\overbrace{C_n}^{\rightarrow 0} \sqrt{3}}{\underbrace{2\sqrt{-\log(C_n)}}_{\rightarrow \infty}} \rightarrow 0, \quad \frac{\overbrace{C_n}^{\rightarrow 0}}{2\sqrt{-\log(\delta)}} \rightarrow 0$$

Aus diesen Konvergenzen können wir für Gleichung (4.21) bereits folgern:

$$\lim_{n \rightarrow \infty} \int_0^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \leq \frac{1}{\sqrt{2}} \left(\delta \sqrt{-\log(\delta)} + \frac{\delta}{2\sqrt{-\log(\delta)}} \right) \quad (4.23)$$

Die beiden verbleibenden Summanden beinhalten ausschließlich δ als Variable. Wir überprüfen nun, was passiert, wenn δ immer kleiner wird, also letztlich $\delta \rightarrow 0$.

Der vordere Summand konvergiert nach der Argumentation aus Gleichung (4.22) gegen 0. Und auch für den hinteren Summanden folgt mit Stetigkeitsargumenten schnell:

$$\frac{\overbrace{\delta}^{\rightarrow 0}}{\underbrace{2\sqrt{-\log(\delta)}}_{\rightarrow \infty}} \rightarrow 0$$

Daraus können wir letztendlich die folgende Aussage treffen: Für jedes $\gamma > 0$ gibt es, bei einem hinreichend großen n , ein $\delta > 0$, sodass $\int_0^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \leq \gamma$ fast sicher, oder etwas mathematischer formuliert: Für alle $\gamma, \eta > 0$ existiert ein $\delta > 0$, sodass

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\int_0^\delta \left(\log \left(\frac{N_{\text{cov}}(u, d_n, [0, 1])}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du > \gamma \right) = 0 < \eta$$

Damit ist $S_n^2(x)$ nach Lemma 2.19 gleichmäßig gleichgradig stetig in Wahrscheinlichkeit. \square

4.2.2. Kovarianzschätzer

Satz 4.5.

Sei \hat{R} der Schätzer aus Gleichung (3.4) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelte $\sqrt{nh}^{1+\alpha_R} \rightarrow 0$, $\frac{\sqrt{n}}{p^{1-d_1}h} \rightarrow 0$ sowie

$$\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty \text{ und } \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} < \infty. \text{ Dann gilt:}$$

$$\sqrt{n}(\hat{R}(s, t) - R(s, t)) \xrightarrow{w} \mathcal{G}(0, \mathcal{R}) \text{ in } (\mathcal{C}([0, 1]^2), \|\cdot\|_{\infty})$$

$\mathcal{G}(0, \mathcal{R})$ ist ein eindeutig bestimmtes Gauß-Feld mit Erwartungswert 0 und Kovarianzfunktion

$$\mathcal{R} : \begin{cases} [0, 1]^4 \rightarrow \mathbb{R} \\ (s, t, s', t') \mapsto M(s, t, s', t') - R(s, t)R(s', t') \end{cases}$$

wobei M bzw. R dem vierten Verbundmoment bzw. der Kovarianzfunktion von Z aus (A.3) entspricht.

Ein Beleg für die Erfüllbarkeit der Konvergenzen von n , p und h , findet sich ebenfalls in Anhang A.3.

Bemerkung 4.6.

Die vierdimensionale Kovarianzfunktion des Gauß-Felds entspricht dabei gerade der Kovarianz des Produktes zweier Z an verschiedenen Stellen. Denn für dieses gilt für $(s, t, s', t') \in [0, 1]^4$ mit dem Verschiebungssatz:

$$\begin{aligned} \text{Cov}(Z(s)Z(t), Z(s')Z(t')) &= \mathbb{E}[Z(s)Z(t)Z(s')Z(t')] - \mathbb{E}[Z(s)Z(t)]\mathbb{E}[Z(s')Z(t')] \\ &= M(s, t, s', t') - R(s, t)R(s', t') \\ &= \mathcal{R}(s, t, s', t') \end{aligned}$$

Dies erscheint insofern auch intuitiv sinnvoll, wenn man bedenkt, dass die zweidimensionale Kovarianzfunktion im Grenzprozess des lokal polynomialen Schätzers $\hat{\mu}$ für μ der Kovarianzfunktion von Z entspricht. Im Kovarianzschätzer \hat{R} treten wiederum Produkte des Schätzers $\hat{\mu}$ auf, wodurch sich eben auch die Kovarianzfunktion in dessen Grenzprozess bzgl. des Produktes ändert.

In \mathcal{R} treten nur noch Komponenten von Z auf, der Fehlerterm ε kommt nicht mehr vor. Das liegt daran, dass \hat{R} als Input für die Kovarianzschätzung die bereits geglätteten Werte aus $\hat{\mu}$ verwendet. Würde man die Kovarianz bzgl. reinen Beobachtungen Y berechnen und danach erst ausglätten, hätte man zwar immer noch ein Gauß-Feld, jedoch wäre in der Kovarianz des Grenzprozesses \mathcal{G} die Kovarianz von ε in irgendeiner Weise mit enthalten, wie [Rei19] zeigt. Das ist genau das bei Abschnitt 3.2.2 erwähnte Problem von zweidimensionalen Gewichtsfunktionen.

Beweis.

Für den Beweis des Satzes muss zunächst der Schätzer in handhabbarere Komponenten zerlegt werden. Die hierfür nötigen Umformungen finden detailliert in Abschnitt 4.3.2 noch einmal statt, daher werden wir hier darauf verzichten und nur die resultierende Zerlegung angeben. Mit Gleichung (4.3) erhalten wir für alle $(s, t) \in [0, 1]^2$:

$$\begin{aligned}
 & \sqrt{n}(\hat{R}(s, t) - R(s, t)) \\
 &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (WZ_i(s)WZ_i(t) + WZ_i(s)W\varepsilon_i(t) + WZ_i(t)W\varepsilon_i(s) + W\varepsilon_i(s)W\varepsilon_i(t)) - R(s, t) \right) \\
 &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)(Z_i(x_j)Z_i(x_k) - R(x_j, x_k)) \right) \Bigg\} =: S_n^1(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)(\varepsilon_{ij}\varepsilon_{ik} - \sigma_{jk}) \right) \Bigg\} =: S_n^2(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)\varepsilon_{ik} \right) \Bigg\} =: S_n^3(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_k)\varepsilon_{ij} \right) \Bigg\} =: S_n^4(s, t) \\
 &- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)Z_i(x_k) \right) \Bigg\} =: S_n^5(s, t) \\
 &- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)\varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)\varepsilon_{ik} \right) \Bigg\} =: S_n^6(s, t) \\
 &- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)\varepsilon_{ik} \right) \Bigg\} =: S_n^7(s, t) \\
 &- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)\varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)Z_i(x_k) \right) \Bigg\} =: S_n^8(s, t) \\
 &+ \sqrt{n} \left(\sum_{j,k=1}^p W_j(s)W_k(t)R(x_j, x_k) - R(s, t) \right) \Bigg\} =: S_n^9(s, t) \\
 &+ \sqrt{n} \left(\sum_{j,k=1}^p W_j(s)W_k(t)\sigma_{jk} \right) \Bigg\} =: S_n^{10}(s, t) \tag{4.24}
 \end{aligned}$$

S_n^1 und S_n^5 entsteht aus $WZ_i(s)WZ_i(t)$, S_n^2 und S_n^6 aus $W\varepsilon_i(s)W\varepsilon_i(t)$, S_n^3 und S_n^7 aus $WZ_i(s)W\varepsilon_i(t)$ und S_n^4 und S_n^8 aus $W\varepsilon_i(s)WZ_i(t)$. In S_n^1 und S_n^2 werden die Zufallsvariablen noch durch eine Nullergänzung zentriert, was für S_n^2 zwar gar nicht nötig wäre, uns aber an späterer Stelle etwas Arbeit ersparen wird.

Es wird nun zeilenweise gezeigt, dass im Grenzfall unter den gegebenen Voraussetzungen der erste Summand die postulierte Verteilung festlegt und alle andere Summanden gleichmäßig in \mathbb{P} gegen 0 konvergieren. Mit Lemma 2.15 folgt danach insgesamt die Grenzverteilung. Die Konvergenzen müssen aber, wegen des funktionalen Charakters des Schätzers, wieder bzgl. der Supremumsnorm gezeigt werden.

$$S_n^1(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (Z_i(x_j) Z_i(x_k) - R(x_j, x_k)) \right)$$

Für diesen Summanden wird der funktionale zentrale Grenzwertsatz aus Abschnitt 2.3.2 verwendet.

Für $n \in \mathbb{N}$, $i = 1, \dots, n$ und $T = [0, 1]^2$ sei:

$$\begin{aligned} X_{n,i}(s, t) &= \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) \\ S_n^1(s, t) &= \sum_{i=1}^n \left(X_{n,i}(s, t) - \mathbb{E}[X_{n,i}(s, t)] \right) \end{aligned} \quad (4.25)$$

Diese stochastischen Prozesse $X_{n,i}$ bilden ein Dreiecksschema mit $k_n = n$ und sind in jeder Zeile unabhängig, weil für $i = 1, \dots, n$ alle Z_i 's unabhängig sind und damit auch alle Produkte der beiden.

Mit dieser Festlegung gilt:

$$\begin{aligned} S_n^1(s, t) &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) - \mathbb{E} \left[\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) \right] \right) \\ &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) \underbrace{\mathbb{E}[Z_i(x_j) Z_i(x_k)]}_{= R(x_j, x_k)} \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (Z_i(x_j) Z_i(x_k) - R(x_j, x_k)) \right) \quad \checkmark \end{aligned} \quad (4.26)$$

Es müssen nun alle fünf Bedingungen des funktionalen zentralen Grenzwertsatzes nachgewiesen werden.

4. Eigenschaften der Schätzer

Wir beginnen mit der Existenz einer Kovarianzfunktion für den Grenzprozess. Für $(s, t), (s', t') \in [0, 1]^2$ gilt nach Gleichung (4.26):

$$\begin{aligned}
& \mathbb{E}[S_n^1(s, t)S_n^1(s', t')] \\
&= \mathbb{E}\left[\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j,k=1}^p W_j(s)W_k(t)(Z_i(x_j)Z_i(x_k) - R(x_j, x_k))\right)\right)\right. \\
&\quad \cdot \left.\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j,k=1}^p W_j(s')W_k(t')(Z_i(x_j)Z_i(x_k) - R(x_j, x_k))\right)\right)\right) \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{\substack{i_1, i_2=1 \\ j_2, k_2=1}}^n\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')(Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})\right. \\
&\quad \left.- Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})R(x_{j_2}, x_{k_2}) - R(x_{j_1}, x_{k_1})Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2}) + R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2})\right) \\
&= \frac{1}{n}\sum_{\substack{i_1, i_2=1 \\ j_2, k_2=1}}^n\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')\left(\mathbb{E}[Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})] \right. \\
&\quad \left. - \underbrace{\mathbb{E}[Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})]}_{= R(x_{j_1}, x_{k_1})} R(x_{j_2}, x_{k_2}) - R(x_{j_1}, x_{k_1}) \underbrace{\mathbb{E}[Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})]}_{= R(x_{j_2}, x_{k_2})} + R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2})\right) \\
&= \frac{1}{n}\sum_{\substack{i_1, i_2=1 \\ j_2, k_2=1}}^n\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')\left(\underbrace{\mathbb{E}[Z_{i_1}(x_{j_1})Z_{i_1}(x_{k_1})Z_{i_2}(x_{j_2})Z_{i_2}(x_{k_2})]}_{= R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2}), \text{ für } i_1 \neq i_2} - R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2})\right) \\
&= \frac{1}{n}\sum_{i=1}^n\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')\left(\mathbb{E}[Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})Z_i(x_{k_2})] - R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2})\right) \\
&\stackrel{\text{iid}}{=} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')\left(M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) - R(x_{j_1}, x_{k_1})R(x_{j_2}, x_{k_2})\right) \\
&= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
&\quad - \left(\sum_{j_1, k_1=1}^p W_{j_1}(s)W_{k_1}(t)R(x_{j_1}, x_{k_1})\right) \left(\sum_{j_2, k_2=1}^p W_{j_2}(s')W_{k_2}(t')R(x_{j_2}, x_{k_2})\right) \\
&\stackrel{\text{Proposition 3.7}}{=} M(s, t, s', t') + \mathcal{O}(h^{1+\alpha_M}) - (R(s, t) + \mathcal{O}(h^{1+\alpha_R}))(R(s', t') + \mathcal{O}(h^{1+\alpha_R})) \\
&\stackrel{\text{Proposition 3.3}}{=} \underbrace{|R(\cdot, \cdot)| \leq E_{Z,2}}_{\rightarrow 0} M(s, t, s', t') - R(s, t)R(s', t') + \underbrace{\mathcal{O}(h^{1+\alpha_M})}_{\rightarrow 0} + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} \\
&\rightarrow M(s, t, s', t') - R(s, t)R(s', t') =: \mathcal{R}(s, t, s', t')
\end{aligned}$$

Bedingung i) ist also erfüllt.

Sei weiterhin $(s, t), (s', t') \in [0, 1]^2$.

$$\begin{aligned}
 d_{\Sigma n}((s, t), (s', t')) &= \left(\sum_{i=1}^n \mathbb{E} \left[(X_{n,i}(s, t) - X_{n,i}(s', t'))^2 \right] \right)^{\frac{1}{2}} \\
 &= \left(\sum_{i=1}^n \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s') W_k(t') Z_i(x_j) Z_i(x_k) \right)^2 \right] \right)^{\frac{1}{2}} \\
 &= \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) Z_i(x_j) Z_i(x_k) \right)^2 \right] \right)^{\frac{1}{2}} \\
 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) Z_i(x_{j_1}) Z_i(x_{k_1}) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) Z_i(x_{j_2}) Z_i(x_{k_2}) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) \mathbb{E} [Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) Z_i(x_{k_2})] \\
 &\stackrel{\text{iid}}{=} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 &= \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \right)^{\frac{1}{2}} \\
 \Rightarrow d_{\Sigma n}^2((s, t), (s', t')) &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 &\quad - 2 \cdot \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 &\quad + \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s') W_{k_1}(t') W_{j_2}(s) W_{k_2}(t) M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 &= M(s, t, s, t) + \mathcal{O}(h^{1+\alpha_M}) + 2(M(s, t, s', t') + \mathcal{O}(h^{1+\alpha_M})) \\
 &\quad + M(s', t', s', t') + \mathcal{O}(h^{1+\alpha_M}) \\
 &= M(s, t, s, t) + 2M(s, t, s', t') + M(s', t', s', t') + \underbrace{\mathcal{O}(h^{1+\alpha_M})}_{\rightarrow 0} \\
 &\rightarrow M(s, t, s, t) + 2M(s, t, s', t') + M(s', t', s', t') =: d((s, t), (s', t'))
 \end{aligned}$$

Weil $d_{\Sigma n}((s, t), (s', t'))$ für alle $(s, t), (s', t') \in [0, 1]^2$ nach Definition nicht-negativ ist, gilt dies auch für $d((s, t), (s', t'))$. Der Grenzwert ist also wohldefiniert. Weil zudem $h^{1+\alpha_M}$ nicht von $(s, t), (s', t')$ abhängt, ist diese Konvergenz gleichmäßig.

Demnach gilt für alle Folgen $((x_1, y_1)_n)_{n \in \mathbb{N}}, ((x_2, y_2)_n)_{n \in \mathbb{N}} \subseteq [0, 1]^2$, welche $\lim_{n \rightarrow \infty} d((x_1, y_1)_n, (x_2, y_2)_n) = 0$ erfüllen:

$$\begin{aligned}
 0 &\leq d_{\sum n}((x_1, y_1)_n, (x_2, y_2)_n) = d_{\sum n}((x_1, y_1)_n, (x_2, y_2)_n) - d((x_1, y_1)_n, (x_2, y_2)_n) + d((x_1, y_1)_n, (x_2, y_2)_n) \\
 &\leq |d_{\sum n}((x_1, y_1)_n, (x_2, y_2)_n) - d((x_1, y_1)_n, (x_2, y_2)_n) + d((x_1, y_1)_n, (x_2, y_2)_n)| \\
 &\leq |d_{\sum n}((x_1, y_1)_n, (x_2, y_2)_n) - d((x_1, y_1)_n, (x_2, y_2)_n)| + |d((x_1, y_1)_n, (x_2, y_2)_n)| \\
 &\leq \sup_{(s,t), (s',t') \in [0,1]^2} |d_{\sum n}((s, t), (s', t')) - d((s, t), (s', t'))| + |d((x_1, y_1)_n, (x_2, y_2)_n)| \\
 &\rightarrow \sup_{(s,t), (s',t') \in [0,1]^2} |d((s, t), (s', t')) - d((s, t), (s', t'))| = 0 \\
 \Rightarrow \lim_{n \rightarrow \infty} d_{\sum n}((x_1, y_1)_n, (x_2, y_2)_n) &= 0
 \end{aligned}$$

Also gilt auch Bedingung ii).

Für die beiden nächsten Bedingungen wird für die $X_{n,i}$'s eine geeignete Umhüllende Φ_n für $n \in \mathbb{N}$ benötigt. Aus (A.2) kann für $x_j, x_k \in [0, 1]$ folgende Ungleichung ermittelt werden:

$$\begin{aligned}
 |Z_i(x_j)Z_i(x_k)| &= |Z_i(x_j)||Z_i(x_k)| = |Z_i(0) + Z_i(x_j) - Z_i(0)||Z_i(0) + Z_i(x_k) - Z_i(0)| \\
 &\leq (|Z_i(0)| + |Z_i(x_j) - Z_i(0)|)(|Z_i(0)| + |Z_i(x_k) - Z_i(0)|) \\
 &\leq (|Z_i(0)| + M_i|x_j - 0|^\beta)(|Z_i(0)| + M_i|x_k - 0|^\beta) \\
 &\stackrel{x_j, x_k \in [0,1]}{\leq} (|Z_i(0)| + M_i)^2
 \end{aligned} \tag{4.27}$$

Weil es sich bei Z_i um jeweils identische Kopien von Z handelt, gilt die Ungleichung auch für alle unabhängig identischen Kopien K_i von K . Damit und mit Proposition 3.6 kann die Umhüllende für $n \in \mathbb{N}, i = 1, \dots, n$ und alle $s, t \in [0, 1]$ bestimmt werden.

$$\begin{aligned}
 |X_{n,i}(s, t)| &= \frac{1}{\sqrt{n}} \left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) \right| \leq \frac{1}{\sqrt{n}} \sum_{j,k=1}^p |W_j(s)||W_k(t)||Z_i(x_j)Z_i(x_k)| \\
 &\leq \frac{1}{\sqrt{n}} (|Z_i(0)| + K_i)^2 \sum_{j,k=1}^p |W_j(s)||W_k(t)| \\
 &= \frac{1}{\sqrt{n}} (|Z_i(0)| + K_i)^2 \left(\sum_{j=1}^p |W_j(s)| \right) \left(\sum_{k=1}^p |W_k(t)| \right) \leq \frac{C_{W_1}^2}{\sqrt{n}} (|Z_i(0)| + K_i)^2 =: \Phi_{n,i}
 \end{aligned} \tag{4.28}$$

$\Phi_n = (\Phi_{n,1}, \dots, \Phi_{n,n})$ ist also für alle $n \in \mathbb{N}$ eine Umhüllende.

Nach den binomischen Formeln gilt für alle $a, b \in \mathbb{R}$

$$\begin{aligned}
 & (a - b)^2 = a^2 - 2ab + b^2 \\
 \Leftrightarrow & (a - b)^2 + 2ab = a^2 + b^2 \\
 & \Rightarrow 2ab \leq a^2 + b^2 \\
 & (a + b)^2 = a^2 + 2ab + b^2 \leq a^2 + a^2 + b^2 + b^2 \\
 \Rightarrow & (a + b)^2 \leq 2(a^2 + b^2) \\
 \Rightarrow & (a + b)^4 = (a + b)^2(a + b)^2 \leq 2(a^2 + b^2)2(a^2 + b^2) \\
 & = 4(a^2 + b^2)^2 \leq 8(a^4 + b^4)
 \end{aligned} \tag{4.29}$$

und damit für die Summe der Erwartungswerte der quadrierten Umhüllenden

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E}[\Phi_{n,i}^2] &= \sum_{i=1}^n \mathbb{E} \left[\left(\frac{C_{W_1}^2}{\sqrt{n}} (|Z_i(0)| + K_i)^2 \right)^2 \right] = \frac{C_{W_1}^4}{n} \sum_{i=1}^n \mathbb{E} \left[(|Z_i(0)| + K_i)^4 \right] \\
 &\stackrel{\text{iid}}{=} C_{W_1}^4 \mathbb{E} \left[(|Z(0)| + K)^4 \right] \leq C_{W_1}^4 \mathbb{E} \left[8(|Z(0)|^4 + K^4) \right] \\
 &= 8C_{W_1}^4 (\mathbb{E}[Z(0)^4] + \mathbb{E}[K^4]) = 8C_{W_1}^4 (M(0, 0, 0, 0) + \mathbb{E}[K^4]) < \infty
 \end{aligned}$$

weil $\mathbb{E}[K^4]$ und $M(0, 0, 0, 0)$ für alle $n \in \mathbb{N}$ nach (A.2) und Proposition 3.3 allesamt endlich sind, was sich auch durch den \limsup nicht verändert.

Da durch die Existenz der vierten Momente von $Z(0)$ und K auch die niedrigeren Momente existieren, gilt zudem für alle $\delta > 0$:

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E}[\Phi_{n,i}^2 \mathbf{1}_{\{\Phi_{n,i} > \delta\}}] &= \sum_{i=1}^n \mathbb{E} \left[\left(\frac{C_{W_1}^2}{\sqrt{n}} (|Z_i(0)| + K_i)^2 \right)^2 \mathbf{1}_{\left\{ \frac{C_{W_1}^2}{\sqrt{n}} (|Z_i(0)| + K_i)^2 > \delta \right\}} \right] \\
 &\stackrel{\text{iid}}{=} C_{W_1}^4 \mathbb{E} \left[(|Z(0)| + K)^4 \mathbf{1}_{\left\{ C_{W_1}^2 (|Z(0)| + K)^2 > \sqrt{n}\delta \right\}} \right] \rightarrow 0
 \end{aligned}$$

Bedingung iii) und iv) sind somit ebenfalls erfüllt.

Als letztes muss noch die Implikation aus Bedingung v) nachgewiesen werden. Diese ergibt sich für alle $n \in \mathbb{N}$, $i = 1, \dots, n$, $(s, t), (s', t') \in [0, 1]^2$ und $\delta \in (0, 1]$ mit Gleichung (4.25) und Gleichung (4.28) durch:

$$\begin{aligned}
 & |s - s'| + |t - t'| \leq \delta^\gamma \\
 \Rightarrow & \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \leq C\delta\Phi_{n,i} \\
 \Leftrightarrow & \left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \leq C\delta C_{W_1}^2 (|Z_i(0)| + K_i)^2
 \end{aligned} \tag{4.30}$$

wobei $\gamma, C \in \mathbb{R}^+$ Konstanten sein müssen. Hier werden nun zwei Fälle bzgl. h^β unterschieden, wobei $\beta \in (0, 1]$ die positive Konstante aus (A.2) ist. Wir berechnen also jeweils zwei separate Konstanten $\gamma_1, C_1, \gamma_2, C_2 \in \mathbb{R}^+$ für die beiden verschiedenen Fälle. Danach kann schlichtweg $\gamma := \max\{\gamma_1, \gamma_2\}$ und $C := \max\{C_1, C_2\}$ gewählt werden, weil so die Implikation definitiv gilt.

1. Fall: $h^\beta \leq \delta$

Zunächst einmal wird Gleichung (4.30) wie folgt umgeformt:

$$\begin{aligned}
 & \left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \\
 = & \left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \sum_{j,k=1}^p W_j(s)Z_i(x_j)W_k(t')Z_i(x_k) \right. \\
 & \left. + \sum_{j,k=1}^p W_j(s)Z_i(x_j)W_k(t')Z_i(x_k) - \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \\
 = & \left| \sum_{j=1}^p W_j(s)Z_i(x_j) \sum_{k=1}^p (W_k(t)Z_i(x_k) - W_k(t')Z_i(x_k)) \right. \\
 & \left. + \sum_{k=1}^p W_k(t')Z_i(x_k) \sum_{j=1}^p (W_j(s)Z_i(x_j) - W_j(s')Z_i(x_j)) \right| \\
 \leq & \sum_{j=1}^p |W_j(s)| |Z_i(x_j)| \left| \sum_{k=1}^p (W_k(t)Z_i(x_k) - W_k(t')Z_i(x_k)) \right| \\
 & + \sum_{k=1}^p |W_k(t')| |Z_i(x_k)| \left| \sum_{j=1}^p (W_j(s)Z_i(x_j) - W_j(s')Z_i(x_j)) \right|
 \end{aligned} \tag{4.31}$$

Nun gilt mit Proposition 3.6 und Gleichung (4.27) für alle $x, x' \in [0, 1]$:

$$\sum_{j=1}^p |W_j(x)| |Z_i(x_j)| \leq (|Z_i(0)| + K_i) \sum_{j=1}^p |W_j(x)| \leq C_{W_1} (|Z_i(0)| + K_i)$$

sowie für $|x - x'| \leq \delta^{\frac{1}{\beta}}$ mit (A.9):

$$\begin{aligned} & \left| \sum_{j=1}^p \left(W_j(x) Z_i(x_j) - W_j(x') Z_i(x_j) \right) \right| \\ &= \left| \sum_{j=1}^p \left(W_j(x) Z_i(x_j) - W_j(x') Z_i(x_j) \right) + Z_i(x) - Z_i(x) + Z_i(x') - Z_i(x') \right| \\ &\leq \left| \sum_{j=1}^p W_j(x) Z_i(x_j) - Z_i(x) \right| + |Z_i(x) - Z_i(x')| + \left| Z_i(x') - \sum_{j=1}^p W_j(x') Z_i(x_j) \right| \\ &\leq \left| \sum_{j=1}^p W_j(x) Z_i(x_j) - Z_i(x) \underbrace{\sum_{j=1}^p W_j(x)}_{=1} \right| + K_i |x - x'|^\beta + \left| Z_i(x') \underbrace{\sum_{j=1}^p W_j(x')}_{=1} - \sum_{j=1}^p W_j(x') Z_i(x_j) \right| \\ &\leq \sum_{j=1}^p |W_j(x)| |Z_i(x_j) - Z_i(x)| + K_i \delta + \sum_{j=1}^p |W_j(x')| |Z_i(x') - Z_i(x_j)| \\ &\leq K_i |x_j - x|^\beta \sum_{j=1}^p |W_j(x)| + K_i \delta + K_i |x' - x_j|^\beta \sum_{j=1}^p |W_j(x')| \\ &\stackrel{(*)}{\leq} K_i \underbrace{h^\beta}_{\leq \delta} C_{W_1} + K_i \delta + K_i \underbrace{h^\beta}_{\leq \delta} C_{W_1} \leq 2K_i \delta C_{W_1} + K_i \delta \\ &\stackrel{(**)}{\leq} 3K_i \delta C_{W_1} \leq 3\delta C_{W_1} (|Z_i(0)| + K_i) \end{aligned}$$

(*): Da $W_j(x) = 0$ für $|x - x_j| > h$.

(**): Wegen $1 = \sum_{j=1}^p W_j(x) \leq \sum_{j=1}^p |W_j(x)| \leq C_{W_1} \Rightarrow C_{W_1} \geq 1$.

Wenn nun $|s - s'| + |t - t'| \leq \delta^{\frac{1}{\beta}}$ gilt, so folgt sofort $|s - s'| \leq \delta^{\frac{1}{\beta}}$ sowie $|t - t'| \leq \delta^{\frac{1}{\beta}}$, also lässt Gleichung (4.31) mit insgesamt abschätzen durch:

$$\left| \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) - \sum_{j,k=1}^p W_j(s') W_k(t') Z_i(x_j) Z_i(x_k) \right| \leq 6\delta C_{W_1}^2 (|Z_i(0)| + K_i)^2$$

Mit den zugehörigen Konstanten $\gamma_1 = \frac{1}{\beta}$ und $C_1 = 6$ ist die Bedingung somit erfüllt.

2. Fall: $h^\beta > \delta$

Die Umformung von Gleichung (4.30) erfolgt hier etwas anders.

$$\begin{aligned}
& \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \\
&= \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p (W_j(s)W_k(t) - W_j(s')W_k(t'))Z_i(x_j)Z_i(x_k) \right| \\
&\leq \sum_{j,k=1}^p |W_j(s)W_k(t) - W_j(s')W_k(t')| \frac{1}{\sqrt{n}} |Z_i(x_j)Z_i(x_k)| \\
&\leq \sum_{j,k=1}^p |W_j(s)W_k(t) - W_j(s')W_k(t')| \frac{C_{W_1}}{\sqrt{n}} (|Z_i(0)| + K_i)^2 \\
&= \sum_{j,k=1}^p |W_j(s)W_k(t) - W_j(s')W_k(t')| \Phi_{n,i} \\
&= \sum_{j,k=1}^p |W_j(s)W_k(t) - W_j(s)W_k(t') + W_j(s)W_k(t') - W_j(s')W_k(t')| \Phi_{n,i} \\
&\leq \left(\sum_{j=1}^p |W_j(s)| \sum_{k=1}^p |W_k(t) - W_k(t')| + \sum_{k=1}^p |W_k(t')| \sum_{j=1}^p |W_j(s) - W_j(s')| \right) \Phi_{n,i} \\
&= \left(\sum_{k=1}^p |W_k(t) - W_k(t')| + \sum_{j=1}^p |W_j(s) - W_j(s')| \right) C_{W_1} \Phi_{n,i}
\end{aligned}$$

Für alle $x, x' \in [0, 1]$ kann $\sum_{j=1}^p |W_j(x) - W_j(x')|$ mit Proposition 3.6 insgesamt höchstens $2C_{W_0}ph$ ungleich 0 besitzen. In Kombination mit (A.7) folgt so:

$$\sum_{j=1}^p |W_j(x) - W_j(x')| \leq 2C_{W_0}ph \frac{C_{W_-}|x - x'|}{ph^2} = \frac{2C_{W_0}C_{W_-}|x - x'|}{h}$$

Wenn nun $|s - s'| + |t - t'| \leq \delta^{1+\frac{1}{\beta}}$ gilt, so folgt sofort $|s - s'| \leq \delta^{1+\frac{1}{\beta}}$ sowie $|t - t'| \leq \delta^{1+\frac{1}{\beta}}$.

Damit folgt insgesamt:

$$\begin{aligned}
 & \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s')W_k(t')Z_i(x_j)Z_i(x_k) \right| \\
 & \leq \left(\frac{2C_{W_0}C_{W_-}|s-s'|}{h} + \frac{2C_{W_0}C_{W_-}|t-t'|}{h} \right) C_{W_1}\Phi_{n,i} \leq \frac{4C_{W_0}C_{W_-}\delta^{1+\frac{1}{\beta}}}{\underbrace{h}_{> \delta^{\frac{1}{\beta}}}} C_{W_1}\Phi_{n,i} \\
 & < 4C_{W_0}C_{W_-}C_{W_1}\delta\Phi_{n,i}
 \end{aligned}$$

Mit den zugehörigen Konstanten $\gamma_2 = 1 + \frac{1}{\beta}$ und $C_2 = 4C_{W_0}C_{W_-}C_{W_1}$ ist die Bedingung somit ebenfalls erfüllt. Damit haben wir Bedingung v) komplett nachgewiesen.

Nach dem funktionalen zentralen Grenzwertsatz gilt also schlussendlich

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) (Z_i(x_j)Z_i(x_k) - R(x_j, x_k)) \right) \xrightarrow{w} \mathcal{G}(0, \mathcal{R}) \text{ in } (\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$$

mit

$$\mathcal{R} : \begin{cases} [0, 1]^4 \rightarrow \mathbb{R} \\ (s, t, s', t') \mapsto M(s, t, s', t') - R(s, t)R(s', t') \end{cases}$$

wobei M bzw. R dem vierten Verbundmoment bzw. der Kovarianzfunktion von Z aus (A.3) entspricht.

$$S_n^2(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right)$$

Wir kombinieren wieder die punktweise Konvergenz mit der gleichmäßigen gleichgradigen Stetigkeit.

Punktweise Konvergenz in \mathbb{P} :

Sei $(s, t) \in [0, 1]^2$.

$$\begin{aligned} \mathbb{E} \left[(S_n^2(s, t))^2 \right] &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2 \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} - \sigma_{j_1 k_1}) (\varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2} - \sigma_{j_2 k_2}) \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} \varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2} \right. \\ &\quad \left. - \varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} \sigma_{j_2 k_2} - \varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2} \sigma_{j_1 k_1} + \sigma_{j_1 k_1} \sigma_{j_2 k_2} \right] \\ &= \frac{1}{n} \sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\mathbb{E} [\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} \varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2}] \\ &\quad - \mathbb{E} [\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1}] \sigma_{j_2 k_2} - \mathbb{E} [\varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2}] \sigma_{j_1 k_1} + \sigma_{j_1 k_1} \sigma_{j_2 k_2}) \\ &= \frac{1}{n} \sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\mathbb{E} [\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} \varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2}] - \sigma_{j_1 k_1} \sigma_{j_2 k_2}) \\ &\stackrel{(*)}{=} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\mathbb{E} [\varepsilon_{i j_1} \varepsilon_{i k_1} \varepsilon_{i j_2} \varepsilon_{i k_2}] - \sigma_{j_1 k_1} \sigma_{j_2 k_2}) \\ &\stackrel{\text{iid}}{=} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) (\mathbb{E} [\varepsilon_{1 j_1} \varepsilon_{1 k_1} \varepsilon_{1 j_2} \varepsilon_{1 k_2}] - \mathbb{E} [\varepsilon_{1 j_1} \varepsilon_{1 k_1}] \mathbb{E} [\varepsilon_{1 j_2} \varepsilon_{1 k_2}]) \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \text{Cov}(\varepsilon_{1 j_1} \varepsilon_{1 k_1}, \varepsilon_{1 j_2} \varepsilon_{1 k_2}) \\ &= (\vec{W}(s) \otimes \vec{W}(t))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}(s) \otimes \vec{W}(t)) \\ &\stackrel{\text{Proposition 3.9}}{=} \mathcal{O} \left(\frac{1}{p^2 - d_2 h^2} \right) \rightarrow 0 \end{aligned} \tag{4.32}$$

(*): da $\mathbb{E}[\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1} \varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2}] \stackrel{i_1 \neq i_2}{=} \mathbb{E}[\varepsilon_{i_1 j_1} \varepsilon_{i_1 k_1}] \mathbb{E}[\varepsilon_{i_2 j_2} \varepsilon_{i_2 k_2}] = \mathbb{E}[\varepsilon_{1 j_1} \varepsilon_{1 k_1}] \mathbb{E}[\varepsilon_{1 j_2} \varepsilon_{1 k_2}] = \sigma_{j_1 k_1} \sigma_{j_2 k_2}$

Die Konvergenz folgt wieder, weil $\frac{1}{p^{2-d_2} h^2}$ wegen $\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2} < \infty$ eine Nullfolge sein muss.

Gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei $\Phi(s, t) = \|S_n^2(s, t)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2}$. Dann

gilt offensichtlich für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^2(s, t)| \leq \Phi(s, t) \Leftrightarrow |S_n^2| \leq \Phi$. Mit Gleichung (4.32) folgt weiterhin:

$$\begin{aligned} \mathbb{E}[|\Phi(s, t)|^2] &= \mathbb{E}[\|S_n^2(s, t)\|_{\ell^2}^2] = \mathbb{E} \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2 \right] = \sum_{n=1}^{\infty} \mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right) \\ &\Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}[|\Phi|^2] \leq C \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2} < \infty \Rightarrow (\mathbb{E}[|\Phi|^2])^{\frac{1}{2}} < \infty \end{aligned}$$

zu ii):

Wir schätzen die Überdeckungsanzahl mit der kanonischen Metrik ab. Auf gleichem Rechenweg und den Konventionen wie in Gleichung (4.32), folgt mit Proposition 3.9 für alle $(s, t), (s', t') \in [0, 1]^2$

$$\begin{aligned}
d_n^2((s, t), (s', t')) &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right. \right. \\
&\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s') W_k(t') (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2 \right] \\
&= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) \right)^2 \right] \\
&= (\vec{W}(s) \otimes \vec{W}(t) - \vec{W}(s') \otimes \vec{W}(t'))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}(s) \otimes \vec{W}(t) - \vec{W}(s') \otimes \vec{W}(t')) \\
&= \mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right) \\
\Rightarrow \exists C \in \mathbb{R}^+ : d_n^2((s, t), (s', t')) &\leq \frac{C}{p^{2-d_2} h^2} \Rightarrow d_n((s, t), (s', t')) \leq \left(\frac{C}{p^{2-d_2} h^2} \right)^{\frac{1}{2}} =: C_n
\end{aligned} \tag{4.33}$$

Da alle Punkte $(s, t), (s', t') \in [0, 1]^2$ einen maximalen Abstand von C_n bzgl. d_n haben, benötigt man für $u < C_n$ maximal $\left(\frac{C_n}{u}\right)^2$ Kugeln mit Radius u , um das Rechteck $[0, 1]^2$ zu überdecken. Ist $u \geq C_n$, so genügt eine einzige Kugel. Für die Überdeckungsanzahl $N_{\text{cov}}(u, d_n, [0, 1]^2) =: N_{\text{cov}}$ bedeutet das nun:

$$\begin{cases} N_{\text{cov}} = 1, & \text{für } u \geq C_n \\ 1 \leq N_{\text{cov}} \leq \left(\frac{C_n}{u}\right)^2, & \text{für } u < C_n \end{cases}$$

Wir können wieder herleiten:

$$N_{\text{cov}} \leq \left(\frac{C_n}{u}\right)^2 \stackrel{\log \text{ m.w.}}{\underset{\sqrt{\cdot} \text{ m.w.}}{\Leftrightarrow}} \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}}\right)\right)^{\frac{1}{2}} \leq \left(\log \left(\frac{C_n^2}{u^{\frac{5}{2}}}\right)\right)^{\frac{1}{2}}$$

Außerdem gilt für C_n wegen $\frac{1}{p^{2-d_2} h^2} \rightarrow 0$ und der Stetigkeit der Wurzelfunktion:

$$C_n = 2 \left(\frac{C}{p^{2-d_2} h^2}\right)^{\frac{1}{2}} \rightarrow 0 \tag{4.34}$$

also gibt es ein hinreichend großes n , ab welchem $C_n < \delta$, für alle $\delta > 0$.

Wir setzen wieder $\delta < 1$. Mit der Monotonie des Integrals folgt für das Überdeckungsintegral:

$$\begin{aligned}
 \int_0^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du &= \int_0^{C_n} \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du + \int_{C_n}^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \\
 &\leq \int_0^{C_n} \left(\log \left(\frac{C_n^2}{u^{\frac{5}{2}}} \right) \right)^{\frac{1}{2}} du + \int_{C_n}^\delta \left(\log \left(\frac{1}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du \\
 &= \sqrt{\frac{5}{2}} \int_0^{C_n} \sqrt{-\log \left(C_n^{-\frac{4}{5}} u \right)} du + \frac{1}{\sqrt{2}} \int_{C_n}^\delta \sqrt{-\log(u)} du \quad (4.35)
 \end{aligned}$$

Die Abschätzung dieser beiden Integrale erfolgt nun exakt auf gleiche Weise wie bei Gleichung (4.16), nur mit dem Argument $C_n^{-\frac{4}{5}}u$ statt $C_n^{-\frac{2}{3}}u$ innerhalb des Logarithmus, wodurch z. B. die Integration durch Substitution leicht angepasst werden muss. Die Wohldefiniertheit aller dabei entstehenden Ausdrücke garantiert uns Gleichung (4.34) und die Ungleichungskette

$$u \leq C_n \Rightarrow C_n^{-\frac{4}{5}}u \leq C_n^{-\frac{4}{5}}C_n = C_n^{\frac{1}{5}} \rightarrow 0$$

Es folgt also für Gleichung (4.35)

$$\begin{aligned}
 \int_0^\delta \left(\log \left(\frac{N_{\text{cov}}}{\sqrt{u}} \right) \right)^{\frac{1}{2}} du &\leq \sqrt{\frac{5}{2}} \left(\frac{C_n \sqrt{-\log(C_n)}}{\sqrt{5}} + \frac{C_n \sqrt{5}}{2\sqrt{-\log(C_n)}} \right) \\
 &\quad + \frac{1}{\sqrt{2}} \left(\delta \sqrt{-\log(\delta)} - C_n \sqrt{-\log(C_n)} + \frac{\delta}{2\sqrt{-\log(\delta)}} - \frac{C_n}{2\sqrt{-\log(\delta)}} \right)
 \end{aligned}$$

Ab hier sind alle Schritte vollkommen identisch zu denen von $S_n^2(x)$, der Fehlerterm bei $\hat{\mu}$ (siehe Gleichung (4.21) ff.). Der einzige Unterschied in den beiden Gleichungen ist der Faktor $\sqrt{5}$, dieser entsteht gerade durch den anderen Exponenten von C_n . Da $\sqrt{5}$ aber eine Konstante ist, spielt diese für die weiteren Argumentationen überhaupt keine Rolle. Folglich gilt für $S_n^2(s, t)$ die gleichmäßige gleichgradige Stetigkeit.

$$S_n^3(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right)$$

Punktweise Konvergenz in \mathbb{P} :

Sei $(s, t) \in [0, 1]^2$.

$$\begin{aligned} \mathbb{E} \left[(S_n^3(s, t))^2 \right] &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right) \right)^2 \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) Z_{i_1}(x_{j_1}) \varepsilon_{i_1 k_1} Z_{i_2}(x_{j_2}) \varepsilon_{i_2 k_2} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \mathbb{E} [Z_i(x_{j_1}) Z_i(x_{j_2})] \mathbb{E} [\varepsilon_{i k_1} \varepsilon_{i k_2}] \\ &\quad + \frac{1}{n} \sum_{\substack{i_1 \neq i_2 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{j_1, k_1=1}^p \sum_{j_2, k_2=1}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \underbrace{\mathbb{E} [Z_{i_1}(x_{j_1})] \mathbb{E} [\varepsilon_{i_1 k_1}] \mathbb{E} [Z_{i_2}(x_{j_2})] \mathbb{E} [\varepsilon_{i_2 k_2}]}_{= 0} \\ &\stackrel{\text{iid}}{=} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \mathbb{E} [Z_1(x_{j_1}) Z_1(x_{j_2})] \mathbb{E} [\varepsilon_{1 k_1} \varepsilon_{1 k_2}] \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) R(x_{j_1}, x_{j_2}) \sigma_{k_1 k_2} \\ &= \left(\sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s) R(x_{j_1}, x_{j_2}) \right) \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \sigma_{k_1 k_2} \right) \\ &= \left(\sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s) R(x_{j_1}, x_{j_2}) \right) \cdot \vec{W}(t)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t) \\ &\stackrel{\text{Proposition 3.7 und 3.9}}{=} (R(s, s) + \mathcal{O}(h^{1+\alpha_R})) \cdot \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\ &\stackrel{\substack{|R(\cdot, \cdot)| \leq E_{Z,2} \\ \text{Proposition 3.3}}}{=} \mathcal{O} \left(\frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \right) \rightarrow 0 \end{aligned} \tag{4.36}$$

Diese Konvergenz folgt aus der Nullfolgeneigenschaft von $\frac{1}{p^{1-d_1} h}$ und $h \rightarrow 0$.

Gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei $\Phi(s, t) = \|S_n^2(s, t)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right) \right)^2}$. Dann gilt offensichtlich für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^2(s, t)| \leq \Phi(s, t) \Leftrightarrow |S_n^2| \leq \Phi$. Mit Gleichung (4.36) folgt weiterhin:

$$\begin{aligned} \mathbb{E}[|\Phi(s, t)|^2] &= \mathbb{E}[\|S_n^2(s, t)\|_{\ell^2}^2] = \mathbb{E} \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right) \right)^2 \right] = \sum_{n=1}^{\infty} \mathcal{O} \left(\frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}[|\Phi|^2] &\leq C \sum_{n=1}^{\infty} \frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \stackrel{h < 1}{\leq} 2C \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1} h} < \infty \Rightarrow (\mathbb{E}[|\Phi|^2])^{\frac{1}{2}} < \infty \end{aligned}$$

zu ii):

Wir schätzen die Überdeckungszahl mit der kanonischen Metrik ab. Auf gleichem Rechenweg und den Konventionen wie bei Gleichung (4.36), folgt für alle $(s, t), (s', t') \in [0, 1]^2$

$$\begin{aligned}
 d_n^2((s, t), (s', t')) &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) \varepsilon_{ik} \right) - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s') W_k(t') Z_i(x_j) \varepsilon_{ik} \right) \right)^2 \right] \\
 &= \mathbb{E} \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) Z_i(x_j) \varepsilon_{ik} \right) \right)^2 \right] \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) R(x_{j_1}, x_{j_2}) \sigma_{k_1 k_2} \\
 &\stackrel{(*)}{=} \left(\sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s) R(x_{j_1}, x_{j_2}) \right) \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \sigma_{k_1 k_2} \right) \\
 &\quad - 2 \left(\sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s') R(x_{j_1}, x_{j_2}) \right) \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t') \sigma_{k_1 k_2} \right) \\
 &\quad + \left(\sum_{j_1, j_2=1}^p W_{j_1}(s') W_{j_2}(s') R(x_{j_1}, x_{j_2}) \right) \left(\sum_{k_1, k_2=1}^p W_{k_1}(t') W_{k_2}(t') \sigma_{k_1 k_2} \right) \\
 &\stackrel{|\cdot|}{=} (R(s, s) + \mathcal{O}(h^{1+\alpha_R})) \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) + 2(R(s, s') + \mathcal{O}(h^{1+\alpha_R})) \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\
 &\quad + (R(s', s') + \mathcal{O}(h^{1+\alpha_R})) \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\
 &\stackrel{\text{Proposition 3.3}}{=} \mathcal{O} \left(\frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \right) \tag{4.37}
 \end{aligned}$$

$$\Rightarrow \exists C \in \mathbb{R}^+ : d_n^2((s, t), (s', t')) \leq \frac{C(1 + h^{1+\alpha_R})}{p^{1-d_1} h} \Rightarrow d_n((s, t), (s', t')) \leq \left(\frac{C(1 + h^{1+\alpha_R})}{p^{1-d_1} h} \right)^{\frac{1}{2}} =: C_n \tag{4.38}$$

Praktischerweise sind ab hier alle Schritte identisch zu denen von $S_n^2(s, t)$, sie unterscheiden sich lediglich innerhalb von C_n . Für das jetzige C_n gilt aber wegen $\frac{1+h^{1+\alpha_R}}{p^{1-d_1} h} \rightarrow 0$ Voraussetzung einfügen und der Stetigkeit der Wurzelfunktion genauso:

$$2 \left(\frac{C(1 + h^{1+\alpha_R})}{p^{1-d_1} h} \right)^{\frac{1}{2}} \rightarrow 0$$

Das ist das Pendant zu Gleichung (4.34), auf welcher alle restlichen Argumentationen aufgebaut haben.

$$S_n^4(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_k) \varepsilon_{ij} \right)$$

Dieser Summand funktioniert vollkommen analog zu $S_n^3(s, t)$.

$$S_n^5(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) Z_i(x_k) \right)$$

Im Beweis zu Satz 4.4 wurde gezeigt, dass

$$S_n^1(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Z_i(x_j) \right) \xrightarrow{w} \mathcal{G}(0, R) \text{ in } (\mathcal{C}([0, 1]), \|\cdot\|_\infty). \text{ Daraus folgt}$$

$\sup_{x \in [0, 1]} |S_n^1(x)| \xrightarrow{D} \|\mathcal{G}(0, R)\|_\infty$ und demnach mit $\frac{1}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0$ und Lemma 2.15:

$$\frac{1}{\sqrt{n}} \cdot \left(\sup_{s \in [0, 1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \right| \right) \xrightarrow{D} 0$$

Da 0 eine Konstante ist, folgt aus dieser Konvergenz in Verteilung mit Proposition 2.13 direkt die Konvergenz in \mathbb{P} . Ein weiteres Anwenden von Lemma 2.15 führt zu:

$$\begin{aligned} \sup_{(s,t) \in [0,1]^2} |S_n^5(s, t)| &= \sup_{(s,t) \in [0,1]^2} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) Z_i(x_k) \right) \right| \\ &= \underbrace{\frac{1}{\sqrt{n}} \cdot \left(\sup_{s \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \right| \right)}_{\xrightarrow{\mathbb{P}} 0} \cdot \underbrace{\left(\sup_{t \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) Z_i(x_k) \right) \right| \right)}_{\xrightarrow{D} \|\mathcal{G}(0,R)\|_\infty} \xrightarrow{D} 0 \end{aligned}$$

Proposition 2.13 impliziert wieder die Konvergenz in \mathbb{P} , also konvergiert $S_n^5(s, t)$ gleichmäßig in \mathbb{P} gegen 0.

$$S_n^6(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) \varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right)$$

Im Beweis zu Satz 4.4 haben wir zudem gezeigt, dass $\sup_{x \in [0,1]} |S_n^2(x)| \xrightarrow{\mathbb{P}} 0$ mit

$$S_n^2(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \right). \text{ Daraus folgt mit Proposition 2.14 direkt:}$$

$$\begin{aligned} \sup_{(s,t) \in [0,1]^2} |S_n^5(s, t)| &= \sup_{(s,t) \in [0,1]^2} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) \varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right) \right| \\ &= \underbrace{\frac{1}{\sqrt{n}}}_{\xrightarrow{\mathbb{P}} 0} \cdot \left(\underbrace{\sup_{s \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) \varepsilon_{ij} \right) \right|}_{\xrightarrow{\mathbb{P}} 0} \right) \cdot \left(\underbrace{\sup_{t \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right) \right|}_{\xrightarrow{\mathbb{P}} 0} \right) \xrightarrow{\mathbb{P}} 0 \end{aligned}$$

$$S_n^7(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right)$$

Hier folgt schnell über die Argumentationen der beiden vorherigen Summanden und Lemma 2.15 und Proposition 2.14:

$$\begin{aligned} \sup_{(s,t) \in [0,1]^2} |S_n^5(s, t)| &= \sup_{(s,t) \in [0,1]^2} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right) \right| \\ &= \underbrace{\frac{1}{\sqrt{n}}}_{\xrightarrow{\mathbb{P}} 0} \cdot \left(\underbrace{\sup_{s \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) Z_i(x_j) \right) \right|}_{\xrightarrow{\mathbb{D}} \|\mathcal{G}(0, R)\|_\infty} \right) \cdot \left(\underbrace{\sup_{t \in [0,1]} \left| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) \varepsilon_{ik} \right) \right|}_{\xrightarrow{\mathbb{P}} 0} \right) \xrightarrow{\mathbb{D}} 0 \end{aligned}$$

und aus der Konvergenz in Verteilung gegen 0 wieder mit Proposition 2.13 die gleichmäßige Konvergenz gegen 0 in \mathbb{P} .

$$S_n^8(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s) \varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t) Z_i(x_k) \right)$$

Dieser Summand funktioniert vollkommen analog zu $S_n^7(s, t)$.

$$S_n^9(s, t) = \sqrt{n} \left(\sum_{j,k=1}^p W_j(s) W_k(t) R(x_j, x_k) - R(s, t) \right)$$

Aus Proposition 3.7 folgt für alle $(s, t) \in [0, 1]^2$:

$$\begin{aligned} \sqrt{n} \left(\sum_{j,k=1}^p W_j(s) W_k(t) R(x_j, x_k) - R(s, t) \right) &= \sqrt{n} (R(s, t) + \mathcal{O}(h^{1+\alpha_R}) - R(s, t)) = \mathcal{O}(\sqrt{n} h^{1+\alpha_R}) \\ \Rightarrow \sup_{(s,t) \in [0,1]^2} |S_n^9(s, t)| &= \mathcal{O}(\sqrt{n} h^{1+\alpha_R}) \rightarrow 0 \end{aligned}$$

$$S_n^{10}(s, t) = \sqrt{n} \left(\sum_{j,k=1}^p W_j(s) W_k(t) \sigma_{jk} \right)$$

Mit Proposition 3.9 gilt für alle $(s, t) \in [0, 1]^2$:

$$\begin{aligned} \sqrt{n} \left(\sum_{j,k=1}^p W_j(s) W_k(t) \sigma_{jk} \right) &= \sqrt{n} (\vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t)) = \mathcal{O} \left(\frac{\sqrt{n}}{p^{1-d_1} h} \right) \rightarrow 0 \\ \Rightarrow \sup_{(s,t) \in [0,1]^2} |S_n^{10}(s, t)| &= \mathcal{O} \left(\frac{\sqrt{n}}{p^{1-d_1} h} \right) \rightarrow 0 \end{aligned}$$

□

4.3. Konsistenz

Die Konsistenz eines Schätzers kann auf mehrere Arten definiert werden. Wir unterscheiden die drei folgenden:

- schwache Konsistenz \Leftrightarrow der Schätzer konvergiert in \mathbb{P} gegen den wahren Wert.
- Konsistenz im quadratischen Mittel \Leftrightarrow der Schätzer konvergiert im 2-ten bzw. quadratischen Mittel gegen den wahren Wert.
- starke Konsistenz \Leftrightarrow der Schätzer konvergiert f. s. gegen den wahren Wert.

Da es sich um funktionale Schätzer handelt, können diese Konsistenzen zudem entweder nur punktweise, oder gleichmäßig gelten - d. h. entweder bzgl. der Betragsnorm, oder der Supremumsnorm, wobei letzteres natürlich eine deutlich stärkere Eigenschaft ist. Wir werden nun verschiedene Konsistenzarten der beiden Schätzer eruieren.

4.3.1. Trendschätzer

Satz 4.7.

Sei $\hat{\mu}$ der Schätzer aus Gleichung (3.2) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelten alle Voraussetzungen aus Satz 4.4. Dann gilt:

- $\hat{\mu}$ ist punktweise konsistent im quadratischen Mittel.
- $\hat{\mu}$ ist punktweise stark konsistent.
- $\hat{\mu}$ ist gleichmäßig schwach konsistent.

Beweis.

- Diese Eigenschaft folgt relativ schnell aus Satz 4.1. Für alle $x \in [0, 1]$ gilt wegen $n \rightarrow \infty$, $h \rightarrow 0$ und $p^{1-d_1}h \rightarrow \infty$:

$$\mathbb{E}[|\hat{\mu}(x) - \mu(x)|^2] = \text{MSE}(\hat{\mu}(x)) = \mathcal{O}\left(h^{2(1+\alpha_\mu)} + \frac{1 + h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \rightarrow 0$$

Das entspricht der punktweisen Konsistenz im quadratischen Mittel.

- b) Die punktweise starke Konsistenz zeigen wir direkt. Für alle $x \in [0, 1]$ gilt für $\hat{\mu}$ ausgeschrieben mit Proposition 3.7:

$$\begin{aligned}
 \hat{\mu}(x) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i(x_j) + \varepsilon_{ij}) \\
 &= \sum_{j=1}^p W_j(x) \mu(x_j) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Z_i(x_j) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \\
 &= \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Z_i(x_j) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \quad (4.39)
 \end{aligned}$$

Für feste $x_j \in [0, 1]$ ist $Z_i(x_j)$ eine reellwertige Zufallsvariable. Genauso ist die deterministische Funktion $W_j(x)$ für alle $j = 1, \dots, p$ und $x \in [0, 1]$ eine reelle Zahl. Wir können daher $\sum_{j=1}^p W_j(x) Z_i(x_j) =: X_{1,i}$ und $\sum_{j=1}^p W_j(x) \varepsilon_{ij} =: X_{2,i}$ jeweils als einzelne Zufallsvariablen betrachten (welche wiederum aus einer Summe von p Zufallsvariablen inkl. reellen Vorfaktoren bestehen). Da die Z_i 's nach den Voraussetzungen aus Abschnitt 3.1 unabhängige Kopien von Z sind, sind die $Z_i(x_j)$'s unabhängig identisch verteilt, genauso wie die ε_{ij} 's nach (A.4), jeweils bzgl. i . Das überträgt sich wiederum auf $X_{1,i}$ und $X_{2,i}$. Außerdem gilt mit der Dreiecksungleichung, Proposition 3.3 und 3.6 für alle $i, p \in \mathbb{N}$:

$$\begin{aligned}
 \mathbb{E}[|X_{1,i}|] &= \mathbb{E}\left[\left|\sum_{j=1}^p W_j(x) Z_i(x_j)\right|\right] \leq \sum_{j=1}^p |W_j(x)| \mathbb{E}[|Z_i(x_j)|] \leq E_{Z,1} \sum_{j=1}^p |W_j(x)| \\
 &\leq E_{Z,1} C_{W_1} < \infty \\
 \mathbb{E}[|X_{2,i}|] &= \mathbb{E}\left[\left|\sum_{j=1}^p W_j(x) \varepsilon_{ij}\right|\right] \leq \sum_{j=1}^p |W_j(x)| \mathbb{E}[|\varepsilon_{ij}|] \leq E_{\varepsilon,1}(p) \sum_{j=1}^p |W_j(x)| \\
 &\leq E_{\varepsilon,1}(p) C_{W_1} < \infty
 \end{aligned}$$

Demnach ist (2. SLLN) anwendbar und die folgende Darstellung ist legitim:

$$\begin{aligned}
 \bar{X}_1 &= \frac{1}{n} \sum_{i=1}^n X_{1,i} = \mathbb{E}[X_{1,1}] + o_{\text{f.s.}}(1) = \sum_{j=1}^p W_j(x) \mathbb{E}[Z_1(x_j)] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1) \\
 \bar{X}_2 &= \frac{1}{n} \sum_{i=1}^n X_{2,i} = \mathbb{E}[X_{2,1}] + o_{\text{f.s.}}(1) = \sum_{j=1}^p W_j(x) \mathbb{E}[\varepsilon_{1j}] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1)
 \end{aligned}$$

Damit folgt für Gleichung (4.39):

$$\hat{\mu}(x) = \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) + \bar{X}_1 + \bar{X}_2 = \mu(x) + \underbrace{\mathcal{O}(h^{1+\alpha_\mu})}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \mu(x)$$

- c) Für die gleichmäßig schwache Konsistenz machen wir uns die asymptotische Normalität zunutze. Aus Satz 4.4 folgt:

$$\sqrt{n} \|\hat{\mu} - \mu\|_{\infty} \xrightarrow{D} \|\mathcal{G}(0, R)\|_{\infty}$$

Wegen $\frac{1}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0$ gilt nach Lemma 2.15:

$$\|\hat{\mu} - \mu\|_{\infty} = \frac{1}{\sqrt{n}} \cdot \sqrt{n} \|\hat{\mu} - \mu\|_{\infty} \xrightarrow{D} 0$$

Da 0 eine Konstante ist, folgt in diesem Fall nach Proposition 2.13 direkt die Konvergenz in Wahrscheinlichkeit:

$$\|\hat{\mu} - \mu\|_{\infty} \xrightarrow{D} 0 \Leftrightarrow \|\hat{\mu} - \mu\|_{\infty} \xrightarrow{\mathbb{P}} 0$$

Das entspricht gerade der gleichmäßig schwachen Konsistenz.

□

4.3.2. Kovarianzschätzer

Satz 4.8.

Sei \hat{R} der Schätzer aus Gleichung (3.4) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelten alle Voraussetzungen aus Satz 4.5. Dann gilt:

- a) \hat{R} ist punktweise konsistent im quadratischen Mittel.
- b) \hat{R} ist punktweise stark konsistent.
- c) \hat{R} ist gleichmäßig schwach konsistent.

Beweis.

- a) Diese Eigenschaft folgt schnell aus Satz 4.3. Für alle $(s, t) \in [0, 1]^2$ gilt wegen $h \rightarrow 0$ und $p^{1-d_1}h \rightarrow \infty$:

$$\mathbb{E}[|\hat{R}(s, t) - R(s, t)|^2] = \text{MSE}(\hat{R}(s, t)) = \mathcal{O}\left(\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h}\right)^2\right) + o(1) \rightarrow 0$$

- b) Die punktweise starke Konsistenz wird wieder direkt gezeigt. Gleichung (4.2) und (4.3) lieferte uns für alle $(s, t) \in [0, 1]^2$:

$$\hat{R}(s, t) = \frac{1}{n} \sum_{i=1}^n \left(\underbrace{WZ_i(s)WZ_i(t)}_{=: I} + \underbrace{WZ_i(s)W\varepsilon_i(t)}_{=: II} + \underbrace{WZ_i(t)W\varepsilon_i(s)}_{=: III} + \underbrace{W\varepsilon_i(s)W\varepsilon_i(t)}_{=: IV} \right) \quad (4.40)$$

Wir untersuchen nun die einzelnen Summanden inklusive der äußeren Summe auf ihre Konvergenz.

$$\begin{aligned} I &= \frac{1}{n} \sum_{i=1}^n WZ_i(s)WZ_i(t) \\ I &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s)Z_i(x_j) - W_j(s)Z_l(x_j)) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t)Z_i(x_k) - W_k(t)Z_l(x_k)) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_{l_2}(x_k) \right. \\ &\quad \left. - \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_{l_1}(x_j)Z_i(x_k) + \frac{1}{n^2} \sum_{l_1, l_2=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_{l_1}(x_j)Z_{l_2}(x_k) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)Z_i(x_j) \right) \left(\frac{1}{n} \sum_{l_2=1}^n \sum_{k=1}^p W_k(t)Z_{l_2}(x_k) \right) \\ &\quad - \left(\frac{1}{n} \sum_{l_1=1}^n \sum_{j=1}^p W_j(s)Z_{l_1}(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)Z_i(x_k) \right) \\ &\quad + \left(\frac{1}{n} \sum_{l_1=1}^n \sum_{j=1}^p W_j(s)Z_{l_1}(x_j) \right) \left(\frac{1}{n} \sum_{l_2=1}^n \sum_{k=1}^p W_k(t)Z_{l_2}(x_k) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j, k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)Z_i(x_k) \right) \end{aligned}$$

Für feste $x_j, x_k \in [0, 1]$ sind $Z_i(x_j)$ und $Z_i(x_k)$ wieder reelle Zufallsvariablen, die deterministischen Funktionen $W_j(s)$ und $W_k(t)$ für alle $j, k = 1, \dots, p$ und $s, t \in [0, 1]$ reelle Zahlen.

Wir setzen $\sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k) =: X_{1,i}$, $\sum_{j=1}^p W_j(s)Z_i(x_j) =: X_{2,i}$ und $\sum_{k=1}^p W_k(t)Z_i(x_k) =: X_{3,i}$, was wieder jeweils als einzelne Zufallsvariable betrachtet werden kann (bestehend aus einer Summe von p^2 bzw. p Zufallsvariablen inkl. reellen Vorfaktoren). Die unabhängig identische Verteilung der $Z_i(x_j)$'s bzw. $Z_i(x_k)$'s überträgt sich wieder auf $X_{1,i}$, $X_{2,i}$ und $X_{3,i}$. Mit der Δ -Ungleichung, Proposition 3.3 und 3.6 gilt für alle $i, p \in \mathbb{N}$:

$$\begin{aligned} \mathbb{E}[|X_{1,i}|] &= \mathbb{E}\left[\left|\sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k)\right|\right] \leq \sum_{j,k=1}^p |W_j(s)||W_k(t)|\mathbb{E}[|Z_i(x_j)Z_i(x_k)|] \\ &\leq E_{Z,2} \left(\sum_{j=1}^p |W_j(s)|\right) \left(\sum_{k=1}^p |W_k(t)|\right) \leq E_{Z,2} C_{W_1}^2 < \infty \end{aligned}$$

$$\begin{aligned} \mathbb{E}[|X_{2,i}|] &= \mathbb{E}\left[\left|\sum_{j=1}^p W_j(s)Z_i(x_j)\right|\right] \leq \sum_{j=1}^p |W_j(s)|\mathbb{E}[|Z_i(x_j)|] \\ &\leq E_{Z,1} \sum_{j=1}^p |W_j(s)| \leq E_{Z,1} C_{W_1} < \infty \end{aligned}$$

$$\begin{aligned} \mathbb{E}[|X_{3,i}|] &= \mathbb{E}\left[\left|\sum_{k=1}^p W_k(t)Z_i(x_k)\right|\right] \leq \sum_{k=1}^p |W_k(t)|\mathbb{E}[|Z_i(x_k)|] \\ &\leq E_{Z,1} \sum_{k=1}^p |W_k(t)| \leq E_{Z,1} C_{W_1} < \infty \end{aligned}$$

Nach (2. SLLN) folgt die Darstellung:

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n} \sum_{i=1}^n X_{1,i} = \mathbb{E}[X_{1,1}] + o_{f.s.}(1) = \sum_{j,k=1}^p W_j(s)W_k(t)\mathbb{E}[Z_1(x_j)Z_1(x_k)] + o_{f.s.}(1) \\ &= \sum_{j,k=1}^p W_j(s)W_k(t)R(x_j, x_k) + o_{f.s.}(1) \\ \bar{X}_2 &= \frac{1}{n} \sum_{i=1}^n X_{2,i} = \mathbb{E}[X_{2,1}] + o_{f.s.}(1) = \sum_{j=1}^p W_j(s)\mathbb{E}[Z_1(x_j)] + o_{f.s.}(1) = o_{f.s.}(1) \\ \bar{X}_3 &= \frac{1}{n} \sum_{i=1}^n X_{3,i} = \mathbb{E}[X_{3,1}] + o_{f.s.}(1) = \sum_{k=1}^p W_k(t)\mathbb{E}[Z_1(x_k)] + o_{f.s.}(1) = o_{f.s.}(1) \end{aligned}$$

Damit und mit Proposition 3.7 gilt:

$$\begin{aligned} I &= \bar{X}_1 - \bar{X}_2 \cdot \bar{X}_3 = \sum_{j,k=1}^p W_j(s)W_k(t)R(x_j, x_k) + o_{\text{f.s.}}(1) - o_{\text{f.s.}}(1) \cdot o_{\text{f.s.}}(1) \\ &= R(s, t) + \mathcal{O}(h^{1+\alpha_R}) + o_{\text{f.s.}}(1) \end{aligned}$$

$$II = \frac{1}{n} \sum_{i=1}^n W Z_i(s) W \varepsilon_i(t)$$

Über die gleichen Umformungen wie bei I folgt:

$$II = \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)\varepsilon_{ik} - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)Z_i(x_j) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)\varepsilon_{ik} \right)$$

Seien hier nun $X_{1,i} := \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)\varepsilon_{ik}$, $X_{2,i} := \sum_{j=1}^p W_j(s)Z_i(x_j)$ und

$X_{3,i} := \sum_{k=1}^p W_k(t)\varepsilon_{ik}$ wieder für feste $x_j, s, t \in [0, 1]$ reelle Zufallsvariablen. Auch hier gilt nach (A.4) die unabhängig identische Verteilung und mit der Unabhängigkeit von Z und ε für alle $i, p \in \mathbb{N}$:

$$\begin{aligned} \mathbb{E}[|X_{1,i}|] &= \mathbb{E} \left[\left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)\varepsilon_{ik} \right| \right] = \sum_{j,k=1}^p |W_j(s)||W_k(t)|\mathbb{E}[|Z_i(x_j)|]\mathbb{E}[|\varepsilon_{ik}|] \\ &\leq E_{Z,1}E_{\varepsilon,1}(p) \left(\sum_{j=1}^p |W_j(s)| \right) \left(\sum_{k=1}^p |W_k(t)| \right) \leq E_{Z,1}E_{\varepsilon,1}(p)C_{W_1}^2 < \infty \end{aligned}$$

$$\mathbb{E}[|X_{2,i}|] \leq E_{Z,1}C_{W_1} < \infty, \quad \mathbb{E}[|X_{3,i}|] \leq E_{\varepsilon,1}(p)C_{W_1} < \infty$$

Es ist also wieder (2. SLLN) anwendbar und die folgende Darstellung legitim:

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n} \sum_{i=1}^n X_{1,i} = \mathbb{E}[X_{1,1}] + o_{\text{f.s.}}(1) = \sum_{j,k=1}^p W_j(s)W_k(t)\mathbb{E}[Z_1(x_j)]\mathbb{E}[\varepsilon_{1k}] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1) \\ \bar{X}_2 &= \frac{1}{n} \sum_{i=1}^n X_{2,i} = \mathbb{E}[X_{2,1}] + o_{\text{f.s.}}(1) = \sum_{j=1}^p W_j(s)\mathbb{E}[Z_1(x_j)] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1) \\ \bar{X}_3 &= \frac{1}{n} \sum_{i=1}^n X_{3,i} = \mathbb{E}[X_{3,1}] + o_{\text{f.s.}}(1) = \sum_{k=1}^p W_k(t)\mathbb{E}[\varepsilon_{1k}] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1) \end{aligned}$$

Damit folgt schnell:

$$II = \bar{X}_1 - \bar{X}_2 \cdot \bar{X}_3 = o_{f.s.}(1) - o_{f.s.}(1) \cdot o_{f.s.}(1) = o_{f.s.}(1)$$

$$III = \frac{1}{n} \sum_{i=1}^n WZ_i(t)W\varepsilon_i(s)$$

Dieser Summand lässt sich exakt so wie Summand II diskutieren.

$$IV := \frac{1}{n} \sum_{i=1}^n W\varepsilon_i(s)W\varepsilon_i(t)$$

Über die gleichen Umformungen wie bei I und II folgt:

$$IV = \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)\varepsilon_{ij}\varepsilon_{ik} - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(s)\varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_k(t)\varepsilon_{ik} \right)$$

Seien hier nun $X_{1,i} := \sum_{j,k=1}^p W_j(s)W_k(t)\varepsilon_{ij}\varepsilon_{ik}$, $X_{2,i} := \sum_{j=1}^p W_j(s)\varepsilon_{ij}$ und $X_{3,i} :=$

$\sum_{k=1}^p W_k(t)\varepsilon_{ik}$ wieder für feste $s, t \in [0, 1]$ reelle Zufallsvariablen. Auch hier gilt wegen der unabhängig identischen Verteilung für alle $i, p \in \mathbb{N}$ analog zu I :

$$\mathbb{E}[|X_{1,i}|] \leq E_{\varepsilon,2}(p)C_{W_1}^2 < \infty, \quad \mathbb{E}[|X_{2,i}|] \leq E_{\varepsilon,1}(p)C_{W_1} < \infty, \quad \mathbb{E}[|X_{3,i}|] \leq E_{\varepsilon,1}(p)C_{W_1} < \infty,$$

Es ist also wieder (2. SLLN) anwendbar und die folgende Darstellung legitim:

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n} \sum_{i=1}^n X_{1,i} = \mathbb{E}[X_{1,1}] + o_{f.s.}(1) = \sum_{j,k=1}^p W_j(s)W_k(t)\mathbb{E}[\varepsilon_{1j}\varepsilon_{1k}] + o_{f.s.}(1) \\ &= \sum_{j,k=1}^p W_j(s)W_k(t)\sigma_{jk} + o_{f.s.}(1) \end{aligned}$$

$$\bar{X}_2 = \frac{1}{n} \sum_{i=1}^n X_{2,i} = \mathbb{E}[X_{2,1}] + o_{f.s.}(1) = \sum_{j=1}^p W_j(s)\mathbb{E}[\varepsilon_{1j}] + o_{f.s.}(1) = o_{f.s.}(1)$$

$$\bar{X}_3 = \frac{1}{n} \sum_{i=1}^n X_{3,i} = \mathbb{E}[X_{3,1}] + o_{f.s.}(1) = \sum_{k=1}^p W_k(t)\mathbb{E}[\varepsilon_{1k}] + o_{f.s.}(1) = o_{f.s.}(1)$$

Mit Proposition 3.9 folgt:

$$\begin{aligned} IV &= \bar{X}_1 - \bar{X}_2 \cdot \bar{X}_3 = \sum_{j,k=1}^p W_j(s)W_k(t)\sigma_{jk} + o_{f.s.}(1) \cdot o_{f.s.}(1) \\ &= \vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t) + o_{f.s.}(1) = \mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right) + o_{f.s.}(1) \end{aligned}$$

Schlussendlich folgt für Gleichung (4.40) für alle $(s, t) \in [0, 1]^2$:

$$\hat{R}(s, t) = R(s, t) + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} + \underbrace{\mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right)}_{\rightarrow 0} + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0} \xrightarrow{f.s.} R(s, t)$$

- c) Der Beweis der gleichmäßig schwachen Konsistenz ist vollkommen analog zum Beweis von Satz 4.7 c), legitimiert durch Satz 4.5.

□

5. Konfidenzbänder

Satz 4.4 und 4.5 können nun zur Konstruktion von asymptotischen Konfidenzbändern benutzt werden, also oberen und unteren Grenzfunktionen, welche für $n \rightarrow \infty$ die tatsächlichen Funktionen μ bzw. R mit einer Wahrscheinlichkeit von mindestens $(1 - \gamma) \cdot 100\%$, $\gamma \in (0, 1)$ einschließen. Da die asymptotische Normalität jeweils auf funktionaler Ebene berechnet wurde, ist es möglich, die Konfidenzbänder nicht nur *punktweise*, sondern *simultan* zu formulieren. D. h. für die Bänder $B_{\gamma,n,\mu}(x)$ und $B_{\gamma,n,R}(s, t)$ gilt:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\mu(x) \in B_{\gamma,n,\mu}(x), x \in [0, 1]) &\geq 1 - \gamma \\ \lim_{n \rightarrow \infty} \mathbb{P}(R(s, t) \in B_{\gamma,n,R}(s, t), (s, t) \in [0, 1]^2) &\geq 1 - \gamma \end{aligned}$$

Das ist eine deutlich stärkere Aussage als eine punktweise Überdeckungswahrscheinlichkeit von $1 - \gamma$, da derartige Bänder im Allgemeinen eine simultane Überdeckungswahrscheinlichkeit aufweisen, welche kleiner als $1 - \gamma$ ist, siehe z. B. [FFdR16].

5.1. Trendschätzer

Satz 5.1.

Sei $\hat{\mu}$ der Schätzer aus Gleichung (3.2) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelten alle Voraussetzungen aus Satz 4.4. Dann ist ein asymptotisches simultanes $(1 - \gamma) \cdot 100\%$ Konfidenzband für die tatsächliche Trendfunktion μ gegeben durch

$$B_{\gamma,n,\mu}(x) := \left[\hat{\mu}(x) - \frac{\hat{c}_{\gamma,\mu}}{\sqrt{n}}, \hat{\mu}(x) + \frac{\hat{c}_{\gamma,\mu}}{\sqrt{n}} \right], \mathbb{P}^* \left(\|\mathcal{G}(0, \hat{R})\|_{\infty} > \hat{c}_{\gamma,\mu} \right) = \gamma$$

wobei \hat{R} der Kovarianzschätzer von Z nach Gleichung (3.4) ist und in \mathbb{P}^* auf die Stichprobe Y bedingt wird.

$\hat{c}_{\gamma,\mu}$ kann in gewisser Weise als das $(1 - \frac{\gamma}{2})$ -Quantil des Gauß-Prozesses $\mathcal{G}(0, \hat{R})$ bzw. das $(1 - \gamma)$ -Quantil seiner Supremumsnorm aufgefasst werden. Da es sich um einen stochastischen Prozess und keine reellwertige Zufallsvariable handelt, muss die Supremumsnorm zu dessen Bestimmung herangezogen werden.

Bemerkung 5.2.

Es sei hier erwähnt, dass dieses Konfidenzband auch schon in [Deg11] angegeben wird. Dort wird der empirische Prozess $\sqrt{n}(\hat{\mu} - \mu)$ jedoch noch standardisiert - insofern man bei einer funktionalen Zufallsvariable davon sprechen kann - indem zusätzlich durch die geschätzte Standardabweichung von Z , also $\hat{\sigma}_Z(x) := \sqrt{\hat{R}(x, x)}$ dividiert wird.

Da \hat{R} nach Satz 4.8 ein gleichmäßig schwach konsistenter Schätzer ist, führt dies über Lemma 2.15 zu einer schwachen Konvergenz gegen $\mathcal{G}(0, \rho_Z)$, wobei ρ_Z die Korrelationsfunktion von Z , also die Normierung der Kovarianzfunktion R ist, und zu den Bändern:

$$B'_{\gamma, n, \mu}(x) := \left[\hat{\mu}(x) - \hat{c}_{\gamma, \mu} \frac{\hat{\sigma}_Z(x)}{\sqrt{n}}, \hat{\mu}(x) + \hat{c}_{\gamma, \mu} \frac{\hat{\sigma}_Z(x)}{\sqrt{n}} \right]$$

$$\mathbb{P}^* \left(\|\mathcal{G}(0, \hat{\rho}_Z)\|_\infty > \hat{c}_{\gamma, \mu} \right) = \gamma, \quad \hat{\rho}_Z(s, t) = \frac{\hat{R}(s, t)}{\hat{\sigma}_Z(s)\hat{\sigma}_Z(t)}$$

Beide Versionen funktionieren, zumindest ab einer hinreichend großen Stichprobe n und Anzahl von Beobachtungspunkten p . Bei einer sehr kleinen Stichprobengröße führen allerdings die Bänder $B'_{\gamma, n, \mu}(x)$, in Kombination mit einem gebootstrapteten Quantil $c_{\gamma, \mu}^*$, zu einer marginalen Verbesserung der Überdeckungswahrscheinlichkeit. Demgegenüber muss aber auch erwähnt werden, dass bei der normalen Berechnung die Bänder $B'_{\gamma, n, \mu}(x)$ etwas schlechter abschneiden als $B_{\gamma, n, \mu}(x)$. Beides werden wir später in Kapitel 11 anhand einer Simulation belegen und diskutieren.

Die wahre Kovarianzfunktion R für den Gauß-Prozess ist nicht bekannt, deshalb wird die geschätzte Version \hat{R} verwendet, deshalb muss in der Wahrscheinlichkeit zusätzlich noch auf die Daten Y bedingt werden. Wir beweisen nun die Gültigkeit der Konfidenzbänder $B_{\gamma, n, \mu}(x)$ in aller Ausführlichkeit. Der Beweis für die Bänder $B'_{\gamma, n, \mu}(x)$ aus Bemerkung 5.2 funktioniert bis auf geringfügige Abweichungen genau so, daher werden wir diesen nicht explizit durchführen, sondern jeweils nur kurz auf die Unterschiede eingehen, sofern nötig.

Beweis.

Sei $c_{\gamma, \mu}$ so definiert, dass $\mathbb{P}(\|\mathcal{G}(0, R)\|_\infty > c_{\gamma, \mu}) = \gamma$, wobei R die (tatsächliche) Kovarianzfunktion von Z ist. Dann folgt wegen der in Satz 4.4 ermittelten schwachen Konvergenz:

$$\begin{aligned} \gamma &= \mathbb{P}(\|\mathcal{G}(0, R)\|_\infty > c_{\gamma, \mu}) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{x \in [0, 1]} |\sqrt{n}(\hat{\mu}(x) - \mu(x))| > c_{\gamma, \mu}\right) \\ &\Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}\left(\underbrace{\sup_{x \in [0, 1]} |\sqrt{n}(\hat{\mu}(x) - \mu(x))| \leq c_{\gamma, \mu}}_{\subseteq \{|\sqrt{n}(\hat{\mu}(x) - \mu(x))| \leq c_{\gamma, \mu}, x \in [0, 1]\}}\right) = 1 - \gamma \\ &\Rightarrow \lim_{n \rightarrow \infty} \mathbb{P}\left(|\sqrt{n}(\hat{\mu}(x) - \mu(x))| \leq c_{\gamma, \mu}, x \in [0, 1]\right) \geq 1 - \gamma \\ &\Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}\left(\hat{\mu}(x) - \frac{c_{\gamma, \mu}}{\sqrt{n}} \leq \mu(x) \leq \hat{\mu}(x) + \frac{c_{\gamma, \mu}}{\sqrt{n}}, x \in [0, 1]\right) \geq 1 - \gamma \quad (5.1) \end{aligned}$$

was genau der Definition eines asymptotischen Konfidenzbandes zum Niveau $1 - \gamma$ entspricht. Da allerdings, wie bereits erwähnt, \hat{R} statt R verwendet wurde, was in dem von $c_{\gamma, \mu}$ verschiedenen Quantil $\hat{c}_{\gamma, \mu}$ mündet, muss noch gezeigt werden, dass:

$$\lim_{n \rightarrow \infty} \mathbb{P}^* \left(\|\mathcal{G}(0, \hat{R})\|_\infty > \hat{c}_{\gamma, \mu} \right) \stackrel{\text{f.s.}}{=} \mathbb{P} \left(\|\mathcal{G}(0, R)\|_\infty > \hat{c}_{\gamma, \mu} \right)$$

Denn daraus folgt $\gamma \stackrel{\text{f.s.}}{=} \mathbb{P}(\|\mathcal{G}(0, R)\|_\infty > \hat{c}_{\gamma, \mu})$ und Gleichung (5.1) ist auch für $\hat{c}_{\gamma, \mu}$ erfüllt. Es ist also zu zeigen, dass $\mathcal{G}(0, \hat{R}) \mid Y$ in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ schwach, also gleichmäßig schwach, f. s. gegen $\mathcal{G}(0, R)$ konvergiert. Für die Bänder $B'_{\gamma, n, \mu}(x)$ müsste man $c_{\gamma, \mu}$ durch $\mathbb{P}(\|\mathcal{G}(0, \rho_Z)\|_\infty > c_{\gamma, \mu}) = \gamma$ definieren, was zu einer erforderlichen gleichmäßig schwachen Konvergenz von $\mathcal{G}(0, \hat{\rho}_Z) \mid Y$ gegen $\mathcal{G}(0, \rho_Z)$ führen würde.

Zur gleichmäßig schwachen Konvergenz folgendes Lemma:

Lemma 5.3.

Seien $(\mathbb{P}_n)_{n \in \mathbb{N}}$ und \mathbb{P} W-Maße auf $\mathcal{C}([0, 1])$. Dann sind äquivalent:

- a) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$
- b) $\mathbb{P}_n \xrightarrow{\text{fdd}} \mathbb{P}$ und $(\mathbb{P}_n)_{n \in \mathbb{N}}$ ist straff.

Beweis.

Siehe [Kle13], Seite 495. Anzumerken ist hierbei noch, dass die Aussage dort für W-Maße auf $\mathcal{C}(\mathbb{R}_0^+)$ formuliert ist, was aber wegen $\mathcal{C}([0, 1]) \subset \mathcal{C}(\mathbb{R}_0^+)$ keine Einschränkung darstellt. \square

Es muss also f. s. $\mathcal{G}_n \xrightarrow{\text{fdd}} \mathcal{G}$ und die Straffheit von $(\mathcal{G}_n)_{n \in \mathbb{N}}$ nachgewiesen werden, wobei $\mathcal{G} := \mathcal{G}(0, R)$ und $\mathcal{G}_n := \mathcal{G}(0, \hat{R}) \mid Y$, da \hat{R} natürlich implizit vom Stichprobenumfang n abhängt.

fdd-Konvergenz:

\mathcal{G}_n ist für alle $n \in \mathbb{N}$ ein Gauß-Prozess, was bedeutet, dass all seine endlichdimensionalen Verteilungen multivariaten Normalverteilungen entsprechen. Genauer: Falls $\{X_n(t) \mid t \in [0, 1]\} \sim \mathcal{G}_n$, so ist für alle $t_1, \dots, t_j \in [0, 1]$, $j \in \mathbb{N}$ die Verteilung des Zufallsvektors $(X_n(t_1), \dots, X_n(t_j))^T$ durch eine j -dimensionale Normalverteilung gegeben. Wir betrachten also Zufallsvektoren $\vec{X}_n, \vec{X} \in \mathbb{R}^j$ als endlichdimensionale Teilvektoren von $\{X_n(t) \mid t \in [0, 1]\} \sim \mathcal{G}_n$ bzw. $\{X(t) \mid t \in [0, 1]\} \sim \mathcal{G}$. Diese Vektoren sind verteilt nach $\mathcal{N}_j(0, \hat{\Sigma}) \mid Y$ bzw. $\mathcal{N}_j(0, \Sigma)$. Da sich die Einträge von $\hat{\Sigma}$ bzw. Σ aus den Funktionen \hat{R} bzw. R bilden, wäre z. B. die Komponente $\hat{\Sigma}_{12}$ bzw. Σ_{12} die Kovarianz von $X_n(t_1)$ und $X_n(t_2)$ bzw. $X(t_1)$ und $X(t_2)$, also

$$\text{Cov}^*(X_n(t_1), X_n(t_2)) = \hat{R}(t_1, t_2) \text{ (geschätzt)}, \quad \text{Cov}(X(t_1), X(t_2)) = R(t_1, t_2) \text{ (tatsächlich)} \quad (5.2)$$

Nach dem *Stetigkeitssatz von Lévy* gilt nun $\vec{X}_n \xrightarrow{D} \vec{X}$ genau dann, wenn $\lim_{n \rightarrow \infty} \phi_n(\vec{x}) = \phi(\vec{x})$ für alle $\vec{x} \in \mathbb{R}^j$, wobei ϕ_n bzw. ϕ die (bedingte) charakteristische Funktion von \vec{X}_n bzw. \vec{X} ist. Für diesen Beweis der punktweisen Konvergenz verwenden wir noch ein kurzes Lemma.

Lemma 5.4.

Seien $f_n : \mathbb{R}^j \rightarrow \mathbb{R}$ bzw. $g_n : \mathbb{R}^j \rightarrow \mathbb{R}$ punktweise gegen f bzw. g konvergente Funktionenfolgen und $h : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion. Dann gilt:

a) $f_n + g_n \rightarrow f + g$ punktweise.

b) $h \circ f_n \rightarrow h \circ f$ punktweise.

Beweis.

Für feste $\vec{x} \in \mathbb{R}^j$ sind die Funktionen $f_n(\vec{x})$ und $g_n(\vec{x})$ nichts anderes als Folgen reeller Zahlen. Damit folgt a) direkt aus der Linearität des Grenzwertes. Da h eine stetige Funktion ist, gilt $\lim_{n \rightarrow \infty} h(a_n) = h(a)$ für jede reellwertige Folge $(a_n)_{n \in \mathbb{N}}$ mit $\lim_{n \rightarrow \infty} a_n = a$, also insbesondere auch $\lim_{n \rightarrow \infty} h(f_n(\vec{x})) = h(f(\vec{x}))$ für alle (festen) $\vec{x} \in \mathbb{R}^j$. \square

Die (bedingte) charakteristische Funktion bei den hier vorliegenden j -dimensionalen Normalverteilung ist gegeben durch:

$$\phi_n(\vec{x}) = \exp\left(-\frac{1}{2}\vec{x}^T \hat{\Sigma} \vec{x}\right) \quad \text{bzw.} \quad \phi(\vec{x}) = \exp\left(-\frac{1}{2}\vec{x}^T \Sigma \vec{x}\right)$$

Nun folgt mit Lemma 5.4 und der punktweise starken Konsistenz von \hat{R} aus Satz 4.8 für alle $\vec{x} \in \mathbb{R}^j$ und $j \in \mathbb{N}$:

$$\begin{aligned} \lim_{n \rightarrow \infty} \phi_n(\vec{x}) &= \lim_{n \rightarrow \infty} \exp\left(-\frac{1}{2}\vec{x}^T \hat{\Sigma} \vec{x}\right) = \lim_{n \rightarrow \infty} \exp\left(-\frac{1}{2}\left(\sum_{j_1, j_2=1}^j x_{j_1} x_{j_2} \hat{R}(t_{j_1}, t_{j_2})\right)\right) \\ &= \exp\left(-\frac{1}{2}\left(\sum_{j_1, j_2=1}^j x_{j_1} x_{j_2} \lim_{n \rightarrow \infty} \hat{R}(t_{j_1}, t_{j_2})\right)\right) \stackrel{\text{f.s.}}{=} \exp\left(-\frac{1}{2}\left(\sum_{j_1, j_2=1}^j x_{j_1} x_{j_2} R(t_{j_1}, t_{j_2})\right)\right) \\ &= \exp\left(-\frac{1}{2}\vec{x}^T \Sigma \vec{x}\right) = \phi(\vec{x}) \end{aligned}$$

Insgesamt also:

$$\vec{X}_n \xrightarrow{D} \vec{X} \Leftrightarrow \mathcal{G}_n \xrightarrow{\text{fdd}} \mathcal{G} \text{ f. s.}$$

Für $B'_{\gamma, n, \mu}(x)$ funktioniert die fdd-Konvergenz genauso, da die punktweise starke Konsistenz von \hat{R} eine punktweise starke Konsistenz von $\hat{\rho}_Z$ impliziert.

Straffheit:

Zur Straffheit ein weiteres Lemma:

Lemma 5.5.

Sei $(\mathbb{P}_n)_{n \in \mathbb{N}}$ eine Folge von W -Maßen auf $\mathcal{C}([0, 1])$. Dann ist die Folge genau dann straff, wenn gilt:

i) Für alle $\eta > 0$ existiert ein $\gamma > 0$ und ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$:

$$\mathbb{P}_n(x : |x(0)| > \gamma) < \eta$$

ii) Für alle $\gamma, \eta > 0$ existiert ein $\delta > 0$ und ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$:

$$\mathbb{P}_n(x : \sup_{d(t,t') \leq \delta} |x(t) - x(t')| > \gamma) < \eta$$

wobei x eine beliebige messbare Funktion auf $[0, 1]$ und d eine Metrik auf $[0, 1]$ ist.

Beweis.

Siehe [Bil99], Seite 82. □

Diese Schreibweise ist z. B. für i) eine äquivalente Darstellung zu $\mathbb{P}(|X_n(0)| > \gamma)$ mit $X_n \sim \mathbb{P}_n$, da \mathbb{P}_n ja gerade die Verteilung der messbaren Funktion x festlegt. Da es sich bei uns um eine bedingte Verteilung handelt, müssen die Wahrscheinlichkeiten ebenfalls bedingt betrachtet werden.

zu i):

Sei wieder $X_n \sim \mathcal{G}_n$. Mit der *bedingten Markov-Ungleichung* folgt:

$$\begin{aligned} \mathbb{P}^*(|X_n(0)| > \gamma) &\leq \frac{\mathbb{E}^*[(X_n(0))^2]}{\gamma^2} = \frac{\text{Var}^*(X_n(0)) + (\mathbb{E}^*[X_n(0)])^2}{\gamma^2} \\ &= \frac{\text{Cov}^*(X_n(0), X_n(0)) + (\mathbb{E}^*[X_n(0)])^2}{\gamma^2} = \frac{\hat{R}(0, 0)}{\gamma^2} \stackrel{\text{f.s.}}{=} \frac{R(0, 0) + \hat{a}_n}{\gamma^2} \end{aligned}$$

wobei $(\hat{a}_n)_{n \in \mathbb{N}}$ eine f. s. Nullfolge ist, d. h. $\hat{a}_n \xrightarrow{\text{f.s.}} 0$. Dies folgt wieder aus der punktweise starken Konsistenz von \hat{R} . Unter diesen Voraussetzungen gibt es ein $n_0 \in \mathbb{N}$, so dass

$\mathbb{P}(\hat{a}_n \leq \hat{a}_{n_0}) = 1$ für alle $n \geq n_0$. Wähle $\gamma > \sqrt{\frac{R(0, 0) + \hat{a}_{n_0}}{\eta}}$, dann folgt i) f. s. aus der

Ungleichung:

$$\mathbb{P}^*(|X_n(0)| > \gamma) \stackrel{\text{f.s.}}{<} \eta \cdot \underbrace{\frac{R(0, 0) + \hat{a}_n}{R(0, 0) + \hat{a}_{n_0}}}_{\leq 1, \text{ f.s.}} \leq \eta$$

zu ii):

Bei genauerer Betrachtung wird klar, dass die zweite Bedingung nichts anderes als die gleichmäßige gleichgradige Stetigkeit in Wahrscheinlichkeit aus Definition 2.17 darstellt. Weil es in diesem Fall sehr schwierig bzw. unmöglich ist, den Gauß-Prozess zu beschränken, verwenden wir Lemma 2.20. Wir benötigen zudem eine Eigenschaft der multivariaten Normalverteilung, welche aus diversen Teilbereichen der Statistik wohl bekannt sein sollte.

Proposition 5.6.

Sei $X \sim \mathcal{N}_j(\mu, \Sigma)$, $A \in \mathbb{R}^{k \times j}$ und $b \in \mathbb{R}^k$, dann gilt:

$$Y = AX + b \sim \mathcal{N}_k(A\mu + b, A\Sigma A^T)$$

Das ist eine fundamentale Eigenschaft der multivariaten Normalverteilung: Sie bleibt nach einer linearen Transformation, ggf. in anderer Dimension, erhalten.

Diese Aussage lässt sich leicht mittels der Transformationsformel für Dichten und den Eigenschaften der mehrdimensionalen Normalverteilung beweisen.

Zunächst einmal gilt nach den Argumentationen bei Gleichung (5.2) $(X_n(t), X_n(t'))^T \sim \mathcal{N}_2(0, \hat{\Sigma}) \mid Y$ für alle $t, t' \in [0, 1]$, wobei $\hat{\Sigma} = \begin{pmatrix} \hat{R}(t, t) & \hat{R}(t, t') \\ \hat{R}(t', t) & \hat{R}(t', t') \end{pmatrix}$. Nun gilt mit Proposition 5.6 und der Symmetrie von \hat{R} als Kovarianzfunktion:

$$\begin{aligned} X_n(t) - X_n(t') &= (1, -1) \begin{pmatrix} X_n(t) \\ X_n(t') \end{pmatrix} \\ \Rightarrow X_n(t) - X_n(t') &\sim \mathcal{N} \left((1, -1) \cdot 0, (1, -1) \hat{\Sigma} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) \mid Y = \mathcal{N}(0, \hat{R}(t, t) - 2\hat{R}(t, t') + \hat{R}(t', t')) \mid Y \end{aligned}$$

Exakt die gleiche Verteilung hat f. s. nach den Eigenschaften der *Standardisierung* eine Zufallsvariable der Form

$$(\hat{R}(t, t) - 2\hat{R}(t, t') + \hat{R}(t', t'))^{\frac{1}{2}} \cdot N, \text{ mit } N \sim \mathcal{N}(0, 1)$$

Der Ausdruck unter der Wurzel ist f. s. größer oder gleich 0, da für $\hat{\Sigma}$ als positiv semidefinite Kovarianzmatrix $\vec{v}^T \hat{\Sigma} \vec{v} \geq 0$ für alle $\vec{v} \in \mathbb{R}^2$ f. s. gilt, damit ist die Darstellung der Zufallsvariable in dieser Form wohldefiniert.

Zudem ist er messbar bzgl. \hat{R} . Unter diesen Voraussetzungen gilt nun:

$$\begin{aligned} \mathbb{E}^* \left[|X_n(t) - X_n(t')|^4 \right] &= \mathbb{E}^* \left[\left| (\hat{R}(t, t) - 2\hat{R}(t, t') + \hat{R}(t', t'))^{\frac{1}{2}} N \right|^4 \right] \\ &\stackrel{\text{f.s.}}{=} (\hat{R}(t, t) - 2\hat{R}(t, t') + \hat{R}(t', t'))^2 \cdot \mathbb{E}^*[N^4] \\ &\stackrel{\substack{N, \hat{R} \\ \text{unabh.}}}{=} (\hat{R}(t, t) - 2\hat{R}(t, t') + \hat{R}(t', t'))^2 \underbrace{\mathbb{E}[N^4]}_{= 3 \cdot 1^2} \\ &\stackrel{\text{f.s.}}{=} 3(R(t, t) - R(t, t') + R(t', t') - R(t, t'))^2 + \hat{a}_n \end{aligned}$$

wobei $(\hat{a}_n)_{n \in \mathbb{N}}$ wieder eine f. s. Nullfolge ist. (A.3) setzt $R \in C^{1, \alpha_R}([0, 1]^2)$ voraus, R ist also auf dem kompakten Gebiet $[0, 1]^2$ stetig partiell differenzierbar. Nach dem *Schranke*satz impliziert diese Voraussetzung für R die Lipschitzstetigkeit. Damit folgt:

$$\begin{aligned} \mathbb{E}^* \left[|X_n(t) - X_n(t')|^4 \right] &\stackrel{\text{f.s.}}{\leq} 3(|R(t, t) - R(t, t')| + |R(t', t') - R(t, t')|)^2 + \hat{a}_n \\ &\leq 3L_R^2(|t - t| + |t - t'| + |t' - t| + |t' - t'|)^2 + \hat{a}_n \\ &= 12L_R^2|t - t'|^2 + \hat{a}_n \end{aligned}$$

mit L_R als Lipschitzkonstante von R . Wähle nun n_0 so, dass $\mathbb{P}(\hat{a}_{n_0} \leq 12L_R^2|t - t'|^2) = 1$, was durch die Nullfolgen-Eigenschaft für alle $t, t' \in [0, 1]$ möglich ist. Dann gilt auch insbesondere für alle $n \geq n_0$:

$$\mathbb{E}^* \left[|X_n(t) - X_n(t')|^4 \right] \stackrel{\text{f.s.}}{\leq} 24L_R^2|t - t'|^2$$

Damit ist Lemma 2.20 mit $C = 24L_R^2$, $a = 4$ und $b = 2 > 1 = D$ für alle $n \geq n_0$ f. s. erfüllt, wonach auch ii) f. s. gilt. Aus i) und ii) folgt f. s. die Straffheit und schlussendlich, in Kombination mit der fdd-Konvergenz, f. s. die gleichmäßig schwache Konvergenz von $\mathcal{G}(0, \hat{R}) \mid Y$ gegen $\mathcal{G}(0, R)$.

Für $B'_{\gamma, n, \mu}(x)$ kann die Straffheit von $\mathcal{G}(0, \hat{\rho}_Z) \mid Y$ z. B. durch Entropie-Berechnungen mittels Kapazitäts- und Überdeckungszahlen aus Definition 2.8 nachgewiesen werden. \square

5.2. Kovarianzschätzer

Satz 5.7.

Sei \hat{R} der Schätzer aus Gleichung (3.4) innerhalb des Modells aus Abschnitt 3.1. Weiter seien Annahme 3.1 und 3.4 erfüllt und es gelten alle Voraussetzungen aus Satz 4.5. Dann ist ein asymptotisches simultanes $(1-\gamma) \cdot 100\%$ Konfidenzband für die tatsächliche Kovarianzfunktion R von Z gegeben durch

$$B_{\gamma,n,R}(s,t) := \left[\hat{R}(s,t) - \frac{\hat{c}_{\gamma,R}}{\sqrt{n}}, \hat{R}(s,t) + \frac{\hat{c}_{\gamma,R}}{\sqrt{n}} \right], \quad \mathbb{P}^* \left(\|\mathcal{G}(0, \hat{\mathcal{R}})\|_{\infty} > \hat{c}_{\gamma,R} \right) = \gamma$$

wobei $\hat{\mathcal{R}} : \begin{cases} [0,1]^4 \rightarrow \mathbb{R} \\ (s,t,s',t') \mapsto \hat{M}(s,t,s',t') - \hat{R}(s,t)\hat{R}(s',t') \end{cases}$ und in \mathbb{P}^* auf die Stichprobe Y bedingt wird.

\hat{M} sei dabei der **4. Verbundmomentschätzer** von Z .

$$\hat{M}(s,t,s',t') = \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) (\hat{\mu}_i(s') - \hat{\mu}(s')) (\hat{\mu}_i(t') - \hat{\mu}(t'))$$

Der Begriff „Momentschätzer“ ist hier nicht zu verwechseln mit der *Momentenmethode*, ein spezielles Verfahren zur Gewinnung von Schätzfunktionen. Es handelt sich hier tatsächlich um einen Schätzer für das 4. Verbundmoment von Z , basierend auf der empirischen Verteilung der Stichprobe, ähnlich zum Stichprobenmittel oder der Stichprobenvarianz.

Bei numerischen Berechnungen der Bänder ist es sinnvoller, $\hat{\mathcal{R}}$ vor der Anwendung des *Verschiebungssatzes* zu berechnen, da sonst Auslöschungseffekte auftreten können. Nach Bemerkung 4.6 ist $\mathcal{R}(s,t,s',t') = \text{Cov}(Z(s)Z(t), Z(s')Z(t'))$, der dazu korrespondierende Schätzer ist also:

$$\frac{1}{n} \sum_{i=1}^n \left((\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) - \hat{R}(s,t) \right) \left((\hat{\mu}_i(s') - \hat{\mu}(s')) (\hat{\mu}_i(t') - \hat{\mu}(t')) - \hat{R}(s',t') \right) \quad (5.3)$$

bzw. in korrigierter Version bzgl. der Erwartungstreue mit Vorfaktor $\frac{1}{n-1}$.

Bemerkung 5.8.

Prinzipiell könnte man hier nun auch wie in Bemerkung 5.2 eine Standardisierung von $\sqrt{n}(R - \hat{R})$ vornehmen, indem man durch die geschätzte Standardabweichung von $Z \cdot Z$, also $\hat{\sigma}_{ZZ}(s, t) := \sqrt{\hat{\mathcal{R}}(s, t, s, t)}$ teilt, was wiederum zu den Bändern

$$B'_{\gamma, n, R}(s, t) := \left[\hat{R}(s, t) - \hat{c}_{\gamma, R} \frac{\hat{\sigma}_{ZZ}(s, t)}{\sqrt{n}}, \hat{R}(s, t) + \hat{c}_{\gamma, R} \frac{\hat{\sigma}_{ZZ}(s, t)}{\sqrt{n}} \right]$$

$$\mathbb{P}^* (\|\mathcal{G}(0, \hat{\rho}_{ZZ})\|_{\infty} > \hat{c}_{\gamma, R}) = \gamma, \quad \hat{\rho}_{ZZ}(s, t, s', t') = \frac{\hat{\mathcal{R}}(s, t, s', t')}{\hat{\sigma}_{ZZ}(s, t) \hat{\sigma}_{ZZ}(s', t')}$$

führen würde. Auch hier schneiden diese Bänder für sehr kleine Stichprobengrößen mit einem gebootstrapteten Quantil $c_{\gamma, R}^*$ leicht besser ab. Allerdings nur, wenn man von vornherein davon ausgeht, dass die gesuchte Kovarianzfunktion R einen verhältnismäßig moderaten Charakter hat. Bei zu großen Schwingungen oder Änderungsraten von R auf kurze Distanzen, performen die Bänder $B'_{\gamma, n, R}(s, t)$ innerhalb des Bootstraps schlechter und versagen bei normaler Berechnung komplett, wie wir in den Simulationen in Kapitel 11 noch sehen werden. Eine mögliche Erklärung hierfür wäre, dass $\hat{\mathcal{R}}$ zwar punktweise stark konsistent ist - das werden wir im folgenden Beweis zeigen - jedoch nicht gleichmäßig schwach konsistent sein könnte, wodurch Lemma 2.15 nicht mehr anwendbar ist und der empirische Prozess ggf. nicht gegen ein Gauß-Feld konvergiert. Wir zeigen daher wieder nur den Beweis für die normalen Bänder $B_{\gamma, n, R}(s, t)$ und gehen nur an wenigen Stellen näher auf $B'_{\gamma, n, R}(s, t)$ ein.

Beweis.

Sei $c_{\gamma, R}$ so definiert, dass $\mathbb{P} (\|\mathcal{G}(0, \mathcal{R})\|_{\infty} > c_{\gamma, R}) = \gamma$, wobei

$$\mathcal{R} : \begin{cases} [0, 1]^4 \rightarrow \mathbb{R} \\ (s, t, s', t') \mapsto M(s, t, s', t') - R(s, t)R(s', t') \end{cases}$$

Für das so definierte $c_{\gamma, R}$ folgt die Validität der Konfidenzbänder mit Satz 4.5 analog zu Gleichung (5.1). Das Problem ist, dass in Satz 5.7 auch wieder $\hat{\mathcal{R}}$ statt \mathcal{R} verwendet wurde. Es muss also, wieder über Lemma 5.3, f. s. nachgewiesen werden, dass $(\mathcal{G}_n)_{n \in \mathbb{N}}$ mit $\mathcal{G}_n := \mathcal{G}(0, \hat{\mathcal{R}}) \mid Y$, gleichmäßig schwach gegen $\mathcal{G} := \mathcal{G}(0, \mathcal{R})$ konvergiert.

fdd-Konvergenz:

\mathcal{G}_n ist für alle $n \in \mathbb{N}$ ein zweidimensionales Gauß-Feld, was bedeutet, dass all seine endlichdimensionalen Verteilungen matrixvariater Normalverteilungen entsprechen. Genauer: Falls $\{X_n(s, t) \mid (s, t) \in [0, 1]^2\} \sim \mathcal{G}_n$, so ist $\forall (s_j, t_k) \in [0, 1]^2, j, k \in \mathbb{N}$ die Verteilung der Zufallsmatrix

$$MX_n := \begin{pmatrix} X_n(s_1, t_1) & X_n(s_1, t_2) & \dots & X_n(s_1, t_k) \\ X_n(s_2, t_1) & X_n(s_2, t_2) & \dots & X_n(s_2, t_k) \\ \vdots & \vdots & \ddots & \vdots \\ X_n(s_j, t_1) & X_n(s_j, t_2) & \dots & X_n(s_j, t_k) \end{pmatrix} \in \mathbb{R}^{j \times k} \quad (5.4)$$

durch eine $j \times k$ -matrixvariante Normalverteilung gegeben. Diese Eigenschaft lässt sich jedoch auch Vektorwertig formulieren. Und zwar ist die matrixvariante Normalverteilung von MX_n äquivalent dazu, dass der Zufallsvektor

$$\vec{X}_n := (X_n(s_1, t_1), \dots, X_n(s_1, t_k), X_n(s_2, t_1), \dots, X_n(s_2, t_k), \dots, X_n(s_j, t_1), \dots, X_n(s_j, t_k))^T \in \mathbb{R}^{j \cdot k}$$

eine $(j \cdot k)$ -dimensionale Normalverteilung besitzt. Das Wort „Feld“ bezieht sich dabei nur auf die Menge, auf welcher der Prozess definiert ist, deswegen kann die gaußsche Eigenschaft nach wie vor durch einen endlichdimensionalen multivariat normalverteilten Vektor definiert werden, siehe für diese Korrespondenz z. B. [GN99].

Wir betrachten also Zufallsvektoren $\vec{X}_n, \vec{X} \in \mathbb{R}^{j \cdot k}$ als endlichdimensionale Teilvektoren von $\{X_n(s, t) \mid (s, t) \in [0, 1]^2\} \sim \mathcal{G}_n$ bzw. $\{X(s, t) \mid (s, t) \in [0, 1]^2\} \sim \mathcal{G}$. Diese Vektoren sind verteilt nach $\mathcal{N}_{j \cdot k}(0, \hat{\Sigma}) \mid Y$ bzw. $\mathcal{N}_{j \cdot k}(0, \Sigma)$, wobei sich die Einträge von $\hat{\Sigma}$ bzw. Σ aus den Funktionen $\hat{\mathcal{R}}$ bzw. \mathcal{R} bilden. Z. B. wäre die Komponente $\hat{\Sigma}_{12}$ bzw. Σ_{12} die Kovarianz von $X_n(s_1, t_1)$ und $X_n(s_1, t_2)$ bzw. $X(s_1, t_1)$ und $X(s_1, t_2)$:

$$\begin{aligned} \text{Cov}^*(X_n(s_1, t_1), X_n(s_1, t_2)) &= \hat{\mathcal{R}}(s_1, t_1, s_1, t_2) & (5.5) \\ &= \hat{M}(s_1, t_1, s_1, t_2) - \hat{R}(s_1, t_1)\hat{R}(s_1, t_2) \text{ (geschätzt)} \\ \text{Cov}(X(s_1, t_1), X(s_1, t_2)) &= \mathcal{R}(s_1, t_1, s_1, t_2) \\ &= M(s_1, t_1, s_1, t_2) - R(s_1, t_1)R(s_1, t_2) \text{ (tatsächlich)} \end{aligned}$$

Ab dieser Stelle sind die Argumentationen analog zu Gleichung (5.2) ff., mit dem einzigen Unterschied, dass die $\hat{\Sigma}$ bzw. Σ höhere Dimensionen haben. Letztendlich muss also für alle $(s, t), (s', t') \in [0, 1]^2$ die punktweise f. s. Konvergenz von

$$\hat{\mathcal{R}}(s, t, s', t') = \hat{M}(s, t, s', t') - \hat{R}(s, t)\hat{R}(s', t')$$

bzw. wegen Proposition 2.14

$$\hat{M}(s, t, s', t') \quad \text{und} \quad \hat{R}(s, t)$$

überprüft werden. Nach Satz 4.8 gilt f. s. die punktweise Konvergenz für \hat{R} .

\hat{M} ist ein deutlich größerer Aufwand, denn hier muss Konvergenz erneut explizit gezeigt werden. Mit der Darstellung von $\hat{\mu}_i(x) - \hat{\mu}(x)$ über Gleichung (4.2), spalten wir den Momentschätzer auf.

$$\begin{aligned} \hat{M}(s, t, s', t') &= \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) (\hat{\mu}_i(s') - \hat{\mu}(s')) (\hat{\mu}_i(t') - \hat{\mu}(t')) \\ &= \frac{1}{n} \sum_{i=1}^n (WZ_i(s) + W\varepsilon_i(s)) (WZ_i(t) + W\varepsilon_i(t)) (WZ_i(s') + W\varepsilon_i(s')) (WZ_i(t') + W\varepsilon_i(t')) \end{aligned} \tag{5.6}$$

Um genaueres über das Konvergenzverhalten zu erfahren, bleibt jetzt nichts anderes übrig, als das Produkt auszumultiplizieren. Einen ähnlichen Fall hatten wir bereits im Beweis zu Satz 4.3 für $(\hat{R}(s, t))^2$, jedoch war dort ein Erwartungswert im Spiel, welcher einige der Summanden direkt eliminiert hat, was hier leider nicht der Fall ist. Nach dem Ausmultiplizieren ergibt sich:

$$\begin{aligned}
 & \underbrace{\frac{1}{n} \sum_{i=1}^n W Z_i(s) W Z_i(t) W Z_i(s') W Z_i(t')}_{=: I} + \underbrace{\frac{1}{n} \sum_{i=1}^n W Z_i(s) W Z_i(t) W Z_i(s') W \varepsilon_i(t')}_{=: II} \\
 & + \frac{1}{n} \sum_{i=1}^n W Z_i(s) W Z_i(t) W \varepsilon_i(s') W Z_i(t') + \underbrace{\frac{1}{n} \sum_{i=1}^n W Z_i(s) W Z_i(t) W \varepsilon_i(s') W \varepsilon_i(t')}_{=: III} \\
 & + \frac{1}{n} \sum_{i=1}^n W Z_i(s) W \varepsilon_i(t') W Z_i(s') W Z_i(t') + \frac{1}{n} \sum_{i=1}^n W Z_i(s) W \varepsilon_i(t') W Z_i(s') W \varepsilon_i(t') \\
 & + \frac{1}{n} \sum_{i=1}^n W Z_i(s) W \varepsilon_i(t') W \varepsilon_i(s') W Z_i(t') + \underbrace{\frac{1}{n} \sum_{i=1}^n W Z_i(s) W \varepsilon_i(t') W \varepsilon_i(s') W \varepsilon_i(t')}_{=: IV} \\
 & + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W Z_i(t) W Z_i(s') W Z_i(t') + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W Z_i(t) W Z_i(s') W \varepsilon_i(t') \\
 & + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W Z_i(t) W \varepsilon_i(s') W Z_i(t') + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W Z_i(t) W \varepsilon_i(t') \\
 & + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W \varepsilon_i(t) W Z_i(s') W Z_i(t') + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W \varepsilon_i(t) W Z_i(s') W \varepsilon_i(t') \\
 & + \frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W \varepsilon_i(t) W \varepsilon_i(s') W Z_i(t') + \underbrace{\frac{1}{n} \sum_{i=1}^n W \varepsilon_i(s) W \varepsilon_i(t) W \varepsilon_i(s') W \varepsilon_i(t')}_{=: V} \quad (5.7)
 \end{aligned}$$

Es genügt, die mit $I - V$ gekennzeichneten Summanden zu diskutieren, da sich die restlichen nur bzgl. der Auswertungsstellen (s, t, s', t') unterscheiden und die Argumentationen äquivalent durchgeführt werden können. Nach der Festlegung von WZ und $W\varepsilon$ aus Gleichung (4.2), müssten wir nun erneut für jeden dieser fünf Summanden 16 neue Summanden, also insgesamt 80 Summanden untersuchen. Wir zeigen daher o. B. d. A. an Summand I , dass sich diese 80 Summanden insgesamt auf nur 5 relevante reduzieren, da alle anderen für alle $(s, t, s', t') \in [0, 1]^4$ für $n \rightarrow \infty$ verschwinden.

$$\begin{aligned}
 I &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s)Z_i(x_j) - W_j(s)Z_l(x_j)) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t)Z_i(x_k) - W_k(t)Z_l(x_k)) \right) \\
 &\quad \cdot \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_j(s')Z_i(x_j) - W_j(s')Z_l(x_j)) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_k(t')Z_i(x_k) - W_k(t')Z_l(x_k)) \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^4} \sum_{\substack{l_1, l_2=1 \\ l_3, l_4=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})Z_i(x_{k_2}) \right) \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^4} \sum_{\substack{l_1, l_2=1 \\ l_3, l_4=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})Z_{l_4}(x_{k_2}) \right) + \dots \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t')Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})Z_i(x_{k_2}) \\
 &\quad - \left(\frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k_2=1}^p W_{k_2}(t')Z_i(x_{k_2}) \right) + \dots
 \end{aligned} \tag{5.8}$$

In diesem Term betrachten wir nun das Konvergenzverhalten des zweiten Summanden.

Sei hier $X_{1,i} := \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')Z_i(x_{j_1})Z_i(x_{k_1})Z_i(x_{j_2})$ und $X_{2,i} := \sum_{k_2=1}^p W_{k_2}(t')Z_i(x_{k_2})$,

nach der gleichen Argumentation wie im Beweis zu Satz 4.7 und 4.8 reellwertige, bzgl. i unabhängig identisch verteilte Zufallsvariablen. Mit der Δ -Ungleichung, Proposition 3.3 und 3.6 gilt außerdem für alle $i, p \in \mathbb{N}$:

$$\begin{aligned}
 \mathbb{E}[|X_{1,i}|] &\leq E_{Z,3}C_{W_1}^3 < \infty, \quad \mathbb{E}[|X_{2,i}|] \leq E_{Z,1}C_{W_1} < \infty \\
 \stackrel{(2. \text{ SLLN})}{\Rightarrow} \bar{X}_1 &= \frac{1}{n} \sum_{i=1}^n X_{1,i} = \mathbb{E}[X_{1,1}] + o_{\text{f.s.}}(1) \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')\mathbb{E}[Z_1(x_{j_1})Z_1(x_{k_1})Z_1(x_{j_2})] + o_{\text{f.s.}}(1) \\
 \bar{X}_2 &= \frac{1}{n} \sum_{i=1}^n X_{2,i} = \mathbb{E}[X_{2,1}] + o_{\text{f.s.}}(1) = \sum_{k_2=1}^p W_{k_2}(t')\mathbb{E}[Z_1(x_{k_2})] + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1)
 \end{aligned}$$

Damit folgt schließlich:

$$\begin{aligned}
& \left(\frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k_2=1}^p W_{k_2}(t') Z_i(x_{k_2}) \right) \\
&= \bar{X}_1 \cdot \bar{X}_2 = \left(\sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') \mathbb{E}[Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2})] + o_{f.s.}(1) \right) o_{f.s.}(1) \\
&\stackrel{|\cdot|}{\leq} \left(\sum_{\substack{j_1, k_1=1 \\ j_2}}^p |W_{j_1}(s)| |W_{k_1}(t)| |W_{j_2}(s')| |\mathbb{E}[Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2})]| + o_{f.s.}(1) \right) o_{f.s.}(1) \\
&\leq \left(E_{Z,3} \left(\sum_{j_1=1}^p |W_{j_1}(s)| \right) \left(\sum_{k_1=1}^p |W_{k_1}(t)| \right) \left(\sum_{j_2=1}^p |W_{j_2}(s')| \right) + o_{f.s.}(1) \right) o_{f.s.}(1) \\
&\leq \underbrace{\left(E_{Z,3} C_{W_1}^3 + o_{f.s.}(1) \right)}_{< \infty} o_{f.s.}(1) = o_{f.s.}(1)
\end{aligned}$$

Der Summand konvergiert also f. s. gegen 0. Bei allen weiteren Summanden aus Gleichung (5.8), abgesehen vom ersten, kann genau auf die gleiche Weise argumentiert werden, da sich der Summand immer in ein Produkt aufspalten lässt, wobei immer mindestens ein Faktor auftritt, der f. s. gegen 0 konvergiert, und alle anderen Faktoren kleiner Unendlich sind. Asymptotisch verschwinden aus Gleichung (5.8) also alle Summanden, außer dem ersten. Das bedeutet, wir können $I - V$ aus Gleichung (5.7) umschreiben, indem wir einfach $WZ_i(x)$ bzw. $W\varepsilon_i(x)$ durch $\sum_{j=1}^p W_j(x) Z_i(x_j)$ bzw. $\sum_{j=1}^p W_j(x) \varepsilon_{ij}$ ersetzen, für alle $x \in [0, 1]$, inklusive einem additiven, f. s. gegen 0 konvergierenden Term. Dieser Restterm wird künftig ignoriert und die asymptotisch fast sichere Gleichheit mit $\stackrel{f.s.}{\simeq}$ symbolisiert. Wir wenden im Folgenden mehrfach die Δ -Ungleichung, Proposition 3.3 und 3.6 und (2. SLLN) und erwähnen dies nicht jedes mal erneut.

$$I \stackrel{\text{f.s.}}{\simeq} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) Z_i(x_{k_2})$$

Sei hier $X_i := \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) Z_i(x_{k_2})$ mit $\mathbb{E}[|X_i|] \leq E_{Z,4} C_{W_1}^4 < \infty$ für alle $i, p \in \mathbb{N}$. Mit Proposition 3.7 gilt für alle $(s, t, s', t') \in [0, 1]^4$:

$$\begin{aligned} \bar{X} &= \mathbb{E}[X_1] + o_{\text{f.s.}}(1) = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \mathbb{E}[Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) Z_i(x_{k_2})] + o_{\text{f.s.}}(1) \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) + o_{\text{f.s.}}(1) \\ &= M(s, t, s', t') + \underbrace{\mathcal{O}(h^{1+\alpha_M})}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \xrightarrow{\text{f.s.}} M(s, t, s', t') \end{aligned}$$

$$II \stackrel{\text{f.s.}}{\simeq} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) \varepsilon_{ik_2}$$

Sei hier $X_i := \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2}) \varepsilon_{ik_2}$ mit $\mathbb{E}[|X_i|] \leq E_{Z,3} E_{\varepsilon,1}(p) C_{W_1}^4 < \infty$ für alle $i, p \in \mathbb{N}$ (wegen Z und ε unabh.). Es folgt für alle $(s, t, s', t') \in [0, 1]^4$:

$$\begin{aligned} \bar{X} &= \mathbb{E}[X_1] + o_{\text{f.s.}}(1) = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \mathbb{E}[Z_i(x_{j_1}) Z_i(x_{k_1}) Z_i(x_{j_2})] \underbrace{\mathbb{E}[\varepsilon_{ik_2}]}_{=0} + o_{\text{f.s.}}(1) \\ &= o_{\text{f.s.}}(1) \xrightarrow{\text{f.s.}} 0 \end{aligned}$$

Dieser Summand verschwindet also asymptotisch ohne größere weitere Annahmen.

$$III \stackrel{\text{f.s.}}{\simeq} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) \varepsilon_{ij_2} \varepsilon_{ik_2}$$

Sei hier $X_i := \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) Z_i(x_{k_1}) \varepsilon_{ij_2} \varepsilon_{ik_2}$ mit $\mathbb{E}[|X_i|] \leq$

$E_{Z,2} E_{\varepsilon,2}(p) C_{W_1}^4 < \infty$ für alle $i, p \in \mathbb{N}$ (wegen Z und ε unabh.). Mit Proposition 3.7 und 3.9 folgt für alle $(s, t, s', t') \in [0, 1]^4$:

$$\begin{aligned} \bar{X} &= \mathbb{E}[X_1] + o_{\text{f.s.}}(1) = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \mathbb{E}[Z_i(x_{j_1}) Z_i(x_{k_1})] \mathbb{E}[\varepsilon_{ij_2} \varepsilon_{ik_2}] + o_{\text{f.s.}}(1) \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') R(x_{j_1}, x_{k_1}) \sigma_{j_2 k_2} + o_{\text{f.s.}}(1) \\ &= \left(\sum_{j_1, k_1=1}^p W_{j_1}(s) W_{k_1}(t) R(x_{j_1}, x_{k_1}) \right) \cdot \left(\vec{W}^T(s') \text{Cov}(\vec{\varepsilon}) \vec{W}(t') \right) + o_{\text{f.s.}}(1) \\ &= \underbrace{(R(s, t) + \mathcal{O}(h^{1+\alpha_R}))}_{\rightarrow 0} \cdot \underbrace{\left(\frac{1}{p^{1-d_1} h} \right)}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \end{aligned}$$

Dieser Summand verschwindet also mit unseren Annahmen asymptotisch ebenfalls.

$$IV \stackrel{\text{f.s.}}{\simeq} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') Z_i(x_{j_1}) \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}$$

Dieser Summand lässt sich vollkommen analog zu II behandeln, nur dass in diesem Fall $\mathbb{E}[\cdot] = 0$ aus dem Faktor Z und nicht aus ε kommt. Ansonsten ist die Argumentation absolut identisch.

$$V \stackrel{\text{f.s.}}{\simeq} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}$$

Sei hier $X_i := \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}$ mit $\mathbb{E}[|X_i|] \leq E_{\varepsilon,4}(p) C_{W_1}^4 < \infty$ für alle $i, p \in \mathbb{N}$. Mit Proposition 3.9 folgt für alle $(s, t, s', t') \in [0, 1]^4$:

$$\begin{aligned} \bar{X} &= \mathbb{E}[X_1] + o_{\text{f.s.}}(1) = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \mathbb{E}[\varepsilon_{1j_1} \varepsilon_{1k_1} \varepsilon_{1j_2} \varepsilon_{1k_2}] + o_{\text{f.s.}}(1) \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') (\text{Cov}(\varepsilon_{1j_1} \varepsilon_{1k_1}, \varepsilon_{1j_2} \varepsilon_{1k_2}) + \sigma_{j_1 k_1} \sigma_{j_2 k_2}) \\ &= (\vec{W}(s) \otimes \vec{W}(t))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}(s) \otimes \vec{W}(t)) + (\vec{W}(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(t))^2 + o_{\text{f.s.}}(1) \\ &= \underbrace{\mathcal{O}\left(\frac{1}{p^{2-d_2} h^2}\right)}_{\rightarrow 0} + \underbrace{\left(\mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right)\right)^2}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \end{aligned}$$

Dieser Summand verschwindet also mit unseren Annahmen asymptotisch ebenfalls.

Insgesamt gilt also mit Proposition 2.14 und der Argumentation eines durch die verschiedenen Schritte entstehenden, zwar sehr großen und komplexen, dennoch f. s. im Unendlichen verschwindenden Terms:

$$\hat{M} \xrightarrow{\text{f.s.}} M \tag{5.9}$$

Demnach konvergiert \hat{M} bzw. \hat{R} punktweise f. s. gegen M bzw. R womit die fdd-Konvergenz endgültig nachgewiesen ist.

Straffheit:

Die Straffheit wird wieder mit Lemma 5.5 nachgewiesen.

zu i):

Sei wieder $X_n \sim \mathcal{G}_n$. Da nun X_n eine zweidimensionale Funktion ist, verwenden wir für 0 den Nullvektor $(0, 0)$. Es gilt wieder mit der *bedingten Markov-Ungleichung*:

$$\begin{aligned} \mathbb{P}^*(|X_n(0, 0)| > \gamma) &\leq \frac{\mathbb{E}^*[(X_n(0, 0))^2]}{\gamma^2} = \frac{\text{Var}^*(X_n(0, 0)) + (\mathbb{E}^*[X_n(0, 0)])^2}{\gamma^2} \\ &= \frac{\text{Cov}^*(X_n(0, 0), X_n(0, 0)) + (\mathbb{E}^*[\hat{X}(0, 0)])^2}{\gamma^2} \\ &\stackrel{\text{f.s.}}{=} \frac{\hat{\mathcal{R}}(0, 0, 0, 0)}{\gamma^2} \stackrel{\text{f.s.}}{=} \frac{\mathcal{R}(0, 0, 0, 0) + \hat{a}_n}{\gamma^2} \end{aligned} \quad (5.10)$$

mit $(\hat{a}_n)_{n \in \mathbb{N}}$ Nullfolge f. s.. Diese folgt aus der punktweise f. s. Konvergenz von $\hat{\mathcal{R}}$ gegen \mathcal{R} , was gerade ausführlich erläutert wurde. Wähle nun wieder $n_0 \in \mathbb{N}$ so, dass $\mathbb{P}(\hat{a}_n \leq \hat{a}_{n_0}) = 1$ für alle $n \geq n_0$ und $\gamma > \sqrt{\frac{\mathcal{R}(0, 0, 0, 0) + \hat{a}_{n_0}}{\eta}}$. Dann folgt:

$$\mathbb{P}^*(|X_n(0, 0)| > \gamma) \stackrel{\text{f.s.}}{<} \eta \cdot \underbrace{\frac{\mathcal{R}(0, 0, 0, 0) + \hat{a}_n}{\mathcal{R}(0, 0, 0, 0) + \hat{a}_{n_0}}}_{\leq 1} \leq \eta$$

wonach i) f. s. gilt.

zu ii):

Wir weisen die gleichmäßige gleichgradige Stetigkeit wieder mit Lemma 2.20 nach. $X(s, t) - X(s', t')$ hat nach analoger Argumentation wie im eindimensionalen Fall für alle $(s, t), (s', t') \in [0, 1]^2$ f. s. die gleiche Verteilung wie

$$(\hat{\mathcal{R}}(s, t, s, t) - 2\hat{\mathcal{R}}(s, t, s', t') + \hat{\mathcal{R}}(s', t', s', t'))^{\frac{1}{2}} \cdot N, \text{ mit } N \sim \mathcal{N}(0, 1)$$

was wieder durch die positive Semidefinitheit von $\hat{\Sigma}$ wohldefiniert ist.

Durch die Messbarkeit bzgl. $\hat{\mathcal{R}}$ gilt:

$$\begin{aligned}
 \mathbb{E}^* \left[|X_n(s, t) - X_n(s', t')|^6 \right] &= \mathbb{E}^* \left[\left| (\hat{\mathcal{R}}(s, t, s, t) - 2\hat{\mathcal{R}}(s, t, s', t') + \hat{\mathcal{R}}(s', t', s', t')) \right|^{\frac{1}{2}} N \right]^6 \\
 &\stackrel{\text{f.s.}}{=} \left(\hat{\mathcal{R}}(s, t, s, t) - 2\hat{\mathcal{R}}(s, t, s', t') + \hat{\mathcal{R}}(s', t', s', t') \right)^3 \cdot \mathbb{E}^*[N^6] \\
 &\stackrel{\substack{N, \hat{\mathcal{R}} \\ \text{unabh.}}}{=} 15 \left(\hat{\mathcal{R}}(s, t, s, t) - 2\hat{\mathcal{R}}(s, t, s', t') + \hat{\mathcal{R}}(s', t', s', t') \right)^3 \underbrace{\mathbb{E}[N^6]}_{= 15 \cdot 16} \\
 &\stackrel{\text{f.s.}}{=} 15 \left(\mathcal{R}(s, t, s, t) - 2\mathcal{R}(s, t, s', t') + \mathcal{R}(s', t', s', t') \right)^3 + \hat{a}_n
 \end{aligned}$$

mit $(\hat{a}_n)_{n \in \mathbb{N}}$ Nullfolge f. s.. Mit $R \in C^{1, \alpha_R}([0, 1]^2)$, $M \in C^{1, \alpha_M}([0, 1]^4)$ aus (A.3) und der daraus resultierenden *Lipschitzstetigkeit* gilt über Proposition 3.3:

$$\begin{aligned}
 \mathcal{R}(s, t, s, t) - \mathcal{R}(s, t, s', t') &= M(s, t, s, t) - (R(s, t))^2 - M(s, t, s', t') + R(s, t)R(s', t') \\
 &= M(s, t, s, t) - M(s, t, s', t') + R(s, t)(R(s', t') - R(s, t)) \\
 &\leq |M(s, t, s, t) - M(s, t, s', t')| + |R(s, t)| |R(s', t') - R(s, t)| \\
 &\leq L_M(|s - s'| + |t - t'| + |s - s'| + |t - t'|) + E_{Z,2} L_R(|s' - s| + |t' - t|) \\
 &= (L_M + E_{Z,2} L_R)(|s - s'| + |t - t'|)
 \end{aligned}$$

mit L_M und L_R als *Lipschitzkonstanten* von M und R . Aus der analogen Herleitung für $\mathcal{R}(s', t', s', t') - \mathcal{R}(s, t, s', t')$ folgt damit:

$$\mathbb{E}^* \left[|X_n(s, t) - X_n(s', t')|^6 \right] \stackrel{\text{f.s.}}{\leq} 120(L_M + E_{Z,2} L_R)^3 (|s - s'| + |t - t'|)^3 + \hat{a}_n$$

Wähle nun n_0 so, dass $\mathbb{P}(\hat{a}_{n_0} \leq 120(L_M + E_{Z,2} L_R)^3 (|s - s'| + |t - t'|)^3) = 1$, was durch die Nullfolgen-Eigenschaft für alle $(s, t), (s', t') \in [0, 1]^2$ möglich ist. Dann gilt auch insbesondere für alle $n \geq n_0$:

$$\mathbb{E}^* \left[|X_n(s, t) - X_n(s', t')|^6 \right] \stackrel{\text{f.s.}}{\leq} 240(L_M + E_{Z,2} L_R)^3 (|s - s'| + |t - t'|)^3$$

Damit ist Lemma 2.20 mit $C = 240(L_M + E_{Z,2} L_R)^3$, $a = 6$ und $b = 3 > 2 = D$ für alle $n \geq n_0$ f. s. erfüllt, wonach auch ii) f. s. gilt. Aus i) und ii) folgt f. s. die Straffheit und schlussendlich, in Kombination mit der fdd-Konvergenz, f. s. die gleichmäßig schwache Konvergenz von $\mathcal{G}(0, \hat{\mathcal{R}}) | Y$ gegen $\mathcal{G}(0, \mathcal{R})$. □

6. Bootstrap-Methode der Konfidenzbänder

In Kapitel 5 wurde die asymptotische Normalität aus Satz 4.4 und 4.5 benutzt, um simultane Konfidenzbänder zu konstruieren. Die Beweise der Sätze basierten wiederum auf dem funktionalen zentralen Grenzwertsatz aus Abschnitt 2.3.2, also letztlich auf einer Art Normal-Approximation des empirischen Prozesses. Diese Approximation ist natürlich erst ab einer hinreichend großen Stichprobe legitim, was bereits seit einem der ersten zentralen Grenzwertsätze überhaupt - dem *Satz von Moivre-Laplace*, die Annäherung der Binomialverteilung an die Normalverteilung - klar ist, siehe z. B. [Sac13], Seite 129. Bei einem zu kleinen n ist der empirische Prozess bzw. dessen Verteilung schlichtweg zu weit entfernt von einem Gauß-Prozess, wodurch die Quantile $\hat{c}_{\gamma,\mu}$ und $\hat{c}_{\gamma,R}$ in Satz 5.1 und 5.7 falsch geschätzt werden. Für diesen Fall formulieren wir nun *naive* Bootstrap-Varianten zur Bestimmung eines validen Quantils $c_{\gamma,\mu}^*$ und $c_{\gamma,R}^*$, ebenfalls basierend auf einem Vorschlag aus [Deg11].

6.1. Trendschätzer

Satz 6.1.

Die in Satz 5.1 angeführten Konfidenzbänder $B_{\gamma,n,\mu}(x)$ behalten unter sonst gleichen Bedingungen f. s. ihre Gültigkeit, wenn $\hat{c}_{\gamma,\mu}$ durch $c_{\gamma,\mu}^*$, ermittelt über folgendes Bootstrap-Verfahren, ersetzt wird:

- (B.1) Berechne $\hat{\mu}_i := \hat{\mu}_i(x)$ für alle $i = 1, \dots, n$ und daraus $\hat{\mu}$ nach Gleichung (3.2).
- (B.2) Ermittle eine Bootstrap-Stichprobe μ_1^*, \dots, μ_n^* , wobei μ_i^* für alle $i = 1, \dots, n$ durch Ziehen mit Zurücklegen aus $\{\hat{\mu}_1, \dots, \hat{\mu}_n\}$ bestimmt wird und berechne $\mu^* = \frac{1}{n} \sum_{i=1}^n \mu_i^*$.
- (B.3) Berechne $X_{\mu^*} = \sqrt{n} \|\mu^* - \hat{\mu}\|_\infty$.
- (B.4) Wiederhole Schritt (B.2) und (B.3) N mal und verwende die bedingte empirische Verteilung von $X_{\mu^*1}, \dots, X_{\mu^*N} \mid Y$ zur Bestimmung des $(1-\gamma) \cdot 100\%$ -Quantils $c_{\gamma,\mu}^*$, wobei X_{μ^*i} das X_{μ^*} aus dem i -ten Durchlauf und $Y := (Y_{ij})_{i \in \{1, \dots, n\}, j \in \{1, \dots, p\}}$ die Stichprobe ist.

Bemerkung 6.2.

Werden die Bänder $B'_{\gamma,n,\mu}(x)$ aus Bemerkung 5.2 verwendet, so muss in (B.3) $X_{\mu^*} = \sqrt{n} \left\| \frac{\mu^* - \hat{\mu}}{\sigma_{\mu^*}} \right\|_{\infty}$ berechnet werden, wobei σ_{μ^*} die empirische funktionale Standardabweichung der μ_i^* 's ist.

Es passiert im Prinzip hierbei nichts anderes, als dass der empirische Prozess $\sqrt{n}(\hat{\mu} - \mu)$ (mit zufälligem $\hat{\mu}$ und deterministischem μ) durch seine Bootstrap-Version $\sqrt{n}(\mu^* - \hat{\mu})$ (mit zufälligem μ^* und deterministischem $\hat{\mu}$) ersetzt wird. Das ist ein sehr gängiges und auch - bezüglich der Mallows' Distanz aus Definition 2.27 - stark konsistentes Verfahren, siehe z. B. [Das08], Seite 465. Mittels mehrmaliger Berechnung der Supremumsnorm, wird schließlich das Quantil $\hat{c}_{\gamma,\mu}$ durch $c_{\gamma,\mu}^*$ numerisch angenähert. Auch wenn die Konsistenz für endlichdimensionale Zufallsvariablen wie bereits erwähnt schon häufig untersucht wurde, kommen wir nicht wirklich umhin, die Validität des Bootstraps explizit zu beweisen, da es sich bei uns um die Supremumsnorm einer funktionalen Zufallsvariable handelt.

Beweis.

Der Beweis läuft darauf hinaus, dass $\sqrt{n}(\mu^* - \hat{\mu}) \mid Y \xrightarrow{w} \mathcal{G}(0, R)$ in $(\mathcal{C}([0, 1]), \|\cdot\|_{\infty})$ f. s. gezeigt werden muss. Dies lässt sich zum Einen auf eine mehr oder weniger anschauliche Art, zum Anderen auf eine etwas technischere mathematische Art erklären.

Sei wieder $\mathbb{P}(\|\mathcal{G}(0, R)\|_{\infty} > c_{\gamma,\mu}) = \gamma$. Dann wissen wir aus Gleichung (5.1), dass:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\hat{\mu}(x) - \frac{c_{\gamma,\mu}}{\sqrt{n}} \leq \mu(x) \leq \hat{\mu}(x) + \frac{c_{\gamma,\mu}}{\sqrt{n}}, x \in [0, 1] \right) \geq 1 - \gamma$$

Unter Annahme der oben erwähnten Konvergenz gilt nun:

$$\lim_{n \rightarrow \infty} \mathbb{P}^*(\sqrt{n}\|\mu^* - \hat{\mu}\|_{\infty} > c_{\gamma,\mu}^*) \stackrel{\text{f.s.}}{=} \mathbb{P}(\|\mathcal{G}(0, R)\|_{\infty} > c_{\gamma,\mu}^*)$$

also $\gamma \stackrel{\text{f.s.}}{=} \mathbb{P}(\|\mathcal{G}(0, R)\|_{\infty} > c_{\gamma,\mu}^*)$, wonach Gleichung (5.1) auch für $c_{\gamma,\mu}^*$ erfüllt ist. Konkret bedeutet das, dass das Konfidenzband zum Niveau $1 - \gamma$ f. s. seine Gültigkeit behält, oder anders formuliert, dass die Schranke $c_{\gamma,\mu}^*$ über die Bootstrap-Variante f. s. korrekt berechnet wird.

Die andere Betrachtungsweise stützt sich auf die Überprüfung der Konsistenz von X_{μ^*} im Sinne von Definition 2.28. Die Schranke $c_{\gamma,\mu}^*$ muss das $(1 - \gamma) \cdot 100\%$ -Quantil für $\|\mathcal{G}(0, R)\|_{\infty}$ sein. Das ist gerade dann f. s. erfüllt, wenn X_{μ^*} ein stark konsistenter Bootstrapschätzer für $\|\mathcal{G}(0, R)\|_{\infty}$ ist, also wenn $\rho_2^*(\|\mathcal{G}(0, R)\|_{\infty}, X_{\mu^*}) \xrightarrow{\text{f.s.}} 0$. Wenn nun $\sqrt{n}(\mu^* - \hat{\mu}) \mid Y \xrightarrow{w} \mathcal{G}(0, R)$ in $(\mathcal{C}([0, 1]), \|\cdot\|_{\infty})$ f. s., dann impliziert das $X_{\mu^*} \mid Y \xrightarrow{D} \|\mathcal{G}(0, R)\|_{\infty}$ f. s.. Außerdem gilt für die stetige Funktion $f(x) = x^2$ mit dem *Satz von der stetigen Abbildung* $(X_{\mu^*})^2 \mid Y \xrightarrow{D} (\|\mathcal{G}(0, R)\|_{\infty})^2$, und damit logischerweise

$$\mathbb{E}^* [|X_{\mu^*}|^2] = \mathbb{E}^* [(X_{\mu^*})^2] \xrightarrow{\text{f.s.}} \mathbb{E} [(\|\mathcal{G}(0, R)\|_{\infty})^2] = \mathbb{E} [|\|\mathcal{G}(0, R)\|_{\infty}|^2]$$

Damit folgt schließlich die starke Konsistenz von X_{μ^*} aus Lemma 2.29 e).

Es ist also $\sqrt{n}(\mu^* - \hat{\mu}) \mid Y \xrightarrow{w} \mathcal{G}(0, R)$ in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ f. s. zu zeigen bzw. im Fall von Bemerkung 6.2 $\sqrt{n}(\frac{\mu^* - \hat{\mu}}{\sigma_{\mu^*}}) \mid Y \xrightarrow{w} \mathcal{G}(0, \rho_Z)$, was aber aus der f. s. schwach gleichmäßigen Konsistenz von σ_{μ^*} - denn σ_{μ^*} ist äquivalent zu $\hat{\sigma}_Z$ definiert und dieser Schätzer ist schwach gleichmäßig konsistent, siehe Bemerkung 5.2 - mit Lemma 2.15 sofort folgt. Wir benötigen dafür erst wieder eine geeignete Zerlegung von $\sqrt{n}(\mu^* - \hat{\mu})$.

$$\begin{aligned}
 \sqrt{n}(\mu^*(x) - \hat{\mu}(x)) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p W_j(x) Y_{ij} \right)^* - \hat{\mu}(x) \right) \\
 &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij}^* - \hat{\mu}(x) \right) \\
 &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i(x_j) + \varepsilon_{ij})^* - \hat{\mu}(x) \right) \\
 &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + (Z_i(x_j) + \varepsilon_{ij})^*) - \hat{\mu}(x) \right) \quad (6.1)
 \end{aligned}$$

Hier entsteht nun ein Problem: Da nur die Stichprobe gebootstrapt werden kann, ist innerhalb jeder Bootstrap-Stichprobeneinheit der Prozess Z sowie der Fehler ε gleichermaßen mit enthalten. Unser Ziel ist es, dass Z die Grenzverteilung festlegt und ε ausgeglättet wird. Dieser Nachweis gelingt, wenn man von einem separaten Bootstrap für Z und ε ausgeht, sprich $(Z_i(x_j) + \varepsilon_{ij})^*$ durch $Z_i^*(x_j) + \varepsilon_{ij}^*$ ersetzt. Im Allgemeinen sind das aber nicht die gleichen Zufallsvariablen, denn beim gemeinsamen Bootstrap entsteht eine Abhängigkeitsstruktur zwischen Z und ε . Wir wollen daher vorher noch begründen, dass beide Varianten, zumindest asymptotisch f. s., die Verteilung von $Z_i(x_j) + \varepsilon_{ij}$ für alle $x_j \in [0, 1]$ und $j = 1, \dots, p$ korrekt „simulieren“.

Aufgrund des Bootstrap-Charakters (Ziehen mit Zurücklegen), gelten bzgl. der Zufallsstichprobe Y die folgenden empirischen Verteilungen:

$$\begin{aligned}
 Z_1^*(x_j), \dots, Z_n^*(x_j) \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{Z_i(x_j)}, \quad j = 1, \dots, p \\
 \varepsilon_{1j}^*, \dots, \varepsilon_{nj}^* \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\varepsilon_{ij}}, \quad j = 1, \dots, p \\
 (Z_1(x_j) + \varepsilon_{1j})^*, \dots, (Z_n(x_j) + \varepsilon_{nj})^* \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{Z_i(x_j) + \varepsilon_{ij}}, \quad j = 1, \dots, p \quad (6.2)
 \end{aligned}$$

Die Unabhängigkeit gilt dabei im Sinne von Definition 2.24, Z^* und ε^* sind also bedingt auf die Stichprobe Y unabhängig bzgl. i .

Die so definierte empirische Verteilung ist ein zufälliges Maß, womit im Allgemeinen äußerst schwer umzugehen ist. Da wir hauptsächlich nur Momente empirisch verteilter Zufallsvariablen benötigen, formulieren wir nun eine Proposition, welches uns den Umgang mit ebendiesen etwas vereinfacht.

Proposition 6.3.

Sei $X := (X_i)_{i \in \{1, \dots, n\}} \stackrel{\text{iid}}{\sim} F$, $n \in \mathbb{N}$ eine Folge reellwertiger Zufallsvariablen und $Y := (Y_i)_{i \in \{1, \dots, n\}}$ eine, von X linear abhängige Stichprobe, das bedeutet Y_i ist abhängig von $f(X_i)$ mit f linear für alle $i \in \mathbb{N}$. Sei weiter $X_1^*, \dots, X_n^* | Y \stackrel{\text{iid}}{\sim} F_n$ für alle $n \in \mathbb{N}$, wobei F_n die empirische Verteilung von X ist. Falls das k -te absolute Moment von X_1 existiert, dann gilt für alle $i = 1, \dots, n$:

$$\mathbb{E}^*[(X_i^*)^k] = \mathbb{E}[(X_1)^k] + o_{\text{f.s.}}(1)$$

Analoge Aussage gilt für Verbundmomente verschiedener Folgen von Zufallsvariablen. Dabei ist zu beachten, dass auf der rechten Seite nach wie vor eine Zufallsvariable steht, die jedoch für $n \rightarrow \infty$ fast sicher gegen einen deterministischen Wert konvergiert, ganz im Sinne des starken Gesetz der großen Zahlen, welches auch im Beweis verwendet wird.

Beweis.

Mit F_n als empirischer Verteilung von X gilt für alle $k \in \mathbb{N}$:

$$X_1^*, \dots, X_n^* | Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \Rightarrow (X_1^*)^k, \dots, (X_n^*)^k | Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{(X_i)^k}$$

Wegen der Linearität gibt es eine messbare Funktion g , für welche $X(\omega) = x = g(y)$ für feste Realisierungen $Y(\omega) = y$. Es folgt also:

$$X_1^*, \dots, X_n^* | Y(\omega) = y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{g(y_i)} \Rightarrow (X_1^*)^k, \dots, (X_n^*)^k | Y(\omega) = y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{(g(y_i))^k}$$

Daraus folgt für den bedingten Erwartungswert für alle $i = 1, \dots, n$:

$$\mathbb{E}[(X_i^*)^k | Y(\omega) = y] = \frac{1}{n} \sum_{i=1}^n (g(y_i))^k = \frac{1}{n} \sum_{i=1}^n (x_i)^k$$

Das *Faktorisierungslemma* (siehe z. B. [Bau11], Seite 71) rechtfertigt damit f. s. die Darstellung:

$$\mathbb{E}[(X_i^*)^k | Y] = \frac{1}{n} \sum_{i=1}^n (g(Y_i))^k = \frac{1}{n} \sum_{i=1}^n (X_i)^k$$

Für X^k kann nach den gegebenen Voraussetzungen (2. SLLN) angewendet werden:

$$\frac{1}{n} \sum_{i=1}^n (X_i)^k \xrightarrow{\text{f.s.}} \mathbb{E}[(X_1)^k] \Leftrightarrow \frac{1}{n} \sum_{i=1}^n (X_i)^k = \mathbb{E}[(X_1)^k] + o_{\text{f.s.}}(1)$$

Für Verbundmomente funktioniert der Beweis auf die exakt gleiche Weise. Weil eine präzise allgemeine Formulierung aber sehr unübersichtlich sein würde, verzichten wir in diesem Fall darauf und weisen nur an den relevanten Stellen darauf hin. \square

Wir halten noch in einer weiteren kurzen Proposition zwei weitere Eigenschaften von Z^* und ε^* , welche später von Relevanz sein werden.

Proposition 6.4.

Für alle $n, D \in \mathbb{N}$ und $j_1, \dots, j_D = 1, \dots, p$ gilt:

$$\begin{aligned} \left(\prod_{k=1}^D Z_i^*(x_{j_k}) \right)_{i \in \{1, \dots, n\}} \Big| Y \stackrel{D}{=} \left(\left(\prod_{k=1}^D Z_i(x_{j_k}) \right)^* \right)_{i \in \{1, \dots, n\}} \Big| Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\prod_{k=1}^D Z_i(x_{j_k})} \\ \left(\prod_{k=1}^D \varepsilon_{ij_k}^* \right)_{i \in \{1, \dots, n\}} \Big| Y \stackrel{D}{=} \left(\left(\prod_{k=1}^D \varepsilon_{ij_k} \right)^* \right)_{i \in \{1, \dots, n\}} \Big| Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\prod_{k=1}^D \varepsilon_{ij_k}} \end{aligned}$$

Diese Proposition besagt letztlich nichts anderes, als dass sich die empirische Verteilung auf beliebige Produkte von Z^* bzw. ε^* bzgl. der Messpunkte überträgt und es zudem egal ist, ob man zuerst die Variablen bootstrapt und dann multipliziert, oder umgekehrt. Diese Schlussfolgerung rührt daher, weil der Bootstrap bzgl. i , nicht aber bzgl. j stattfindet: In jeder Bootstrap-Stichprobeneinheit wird effektiv der komplette stochastische Prozess Z bzw. Vektor $\vec{\varepsilon}$ simultan für alle x_1, \dots, x_p bzw. $j = 1, \dots, p$ gezogen wird. Man befindet sich also für jede Bootstrap-Stichprobeneinheit in einer bestimmten ursprünglichen Stichprobeneinheit, die sich jedoch für alle x_1, \dots, x_p bzw. $j = 1, \dots, p$ nicht ändert. Je nach Beschränkung der ursprünglichen Verbundmomente dieser Bootstrap-Variablen, kann für Verbundmomente von Z^* bzw. ε^* Proposition 6.3 angewendet werden.

Seien nun die Folgen $(\vec{X}_n)_{n \in \mathbb{N}}$, $(\vec{X}_n^{1*})_{n \in \mathbb{N}}$ und $(\vec{X}_n^{2*})_{n \in \mathbb{N}}$ definiert durch:

$$\begin{aligned} \vec{X}_i &:= \vec{Z}_i + \vec{\varepsilon}_i = (Z_i(x_1) + \varepsilon_{i1}, \dots, Z_i(x_p) + \varepsilon_{ip})^T \\ \vec{X}_i^{1*} &:= (\vec{Z}_i + \vec{\varepsilon}_i)^* = ((Z_i(x_1) + \varepsilon_{i1})^*, \dots, (Z_i(x_p) + \varepsilon_{ip})^*)^T \\ \vec{X}_i^{2*} &:= \vec{Z}_i^* + \vec{\varepsilon}_i^* = (Z_i^*(x_1) + \varepsilon_{i1}^*, \dots, Z_i^*(x_p) + \varepsilon_{ip}^*)^T \end{aligned}$$

Für feste $x_1, \dots, x_p \in [0, 1]$ entspricht das p -dimensionalen Zufallsvariablen. Da Z und ε bzgl. i unabhängig identisch verteilt sind, gilt $(\vec{X}_n)_{n \in \mathbb{N}} \stackrel{\text{iid}}{\sim} F$ und außerdem nach Gleichung (6.2) $\vec{X}_n^{1*} | Y \sim F_n$ also nach Lemma 2.29 d) und e):

$$\rho_2^*(\vec{X}_1, \vec{X}_n^{1*}) \rightarrow 0 \Rightarrow \vec{X}_n^{1*} | Y \xrightarrow{D} \vec{X}_1 \text{ f. s.}$$

$(\vec{X}_n^{2*})_{n \in \mathbb{N}}$ kann wiederum durch $\left((I_p, I_p) \begin{pmatrix} \vec{Z}_n^* \\ \vec{\varepsilon}_n^* \end{pmatrix} \right)_{n \in \mathbb{N}}$ dargestellt werden, wobei I_p die

$p \times p$ Einheitsmatrix ist - selbiges gilt für $(\vec{X}_n)_{n \in \mathbb{N}}$. (I_p, I_p) ist eine reellwertige $p \times 2p$ Matrix, wonach die Abbildung $\vec{v} \mapsto (I_p, I_p) \vec{v}$ für alle $\vec{v} \in \mathbb{R}^{2p}$ ein linearer Operator auf \mathbb{R}^{2p} ist. Für die Operatornorm $\|\cdot\|$ bietet sich hier die Spaltensummennorm als natürliche

Matrixnorm an, was $\|(I_p, I_p)\| = 1$ liefert. Schließlich gilt erneut mit der empirischen Verteilung von $Z^* | Y$ und $\varepsilon^* | Y$ und Lemma 2.29:

$$\begin{aligned} \rho_2^*(\vec{X}_1, \vec{X}_n^{2*}) &= \rho_2^*(\vec{Z}_1 + \vec{\varepsilon}_1, \vec{Z}_n^* + \vec{\varepsilon}_n^*) = \rho_2^*\left(\left(I_p, I_p\right) \begin{pmatrix} \vec{Z}_1 \\ \vec{\varepsilon}_1 \end{pmatrix}, \left(I_p, I_p\right) \begin{pmatrix} \vec{Z}_n^* \\ \vec{\varepsilon}_n^* \end{pmatrix}\right) \\ &\stackrel{a)}{\leq} \|(I_p, I_p)\| \cdot \rho_2^*\left(\begin{pmatrix} \vec{Z}_1 \\ \vec{\varepsilon}_1 \end{pmatrix}, \begin{pmatrix} \vec{Z}_n^* \\ \vec{\varepsilon}_n^* \end{pmatrix}\right) = \rho_2^*\left(\begin{pmatrix} \vec{Z}_1 \\ \vec{\varepsilon}_1 \end{pmatrix}, \begin{pmatrix} \vec{Z}_n^* \\ \vec{\varepsilon}_n^* \end{pmatrix}\right) \stackrel{d)}{\rightarrow} 0 \text{ f. s.} \\ &\stackrel{e)}{\Rightarrow} \vec{X}_n^{2*} | Y \stackrel{D}{\rightarrow} \vec{X}_1 \text{ f. s.} \end{aligned}$$

Nach dieser Argumentation spielt es schließlich keine Rolle, ob wir im empirischen Prozess $(\vec{X}_n^{1*})_{n \in \mathbb{N}}$ oder $(\vec{X}_n^{2*})_{n \in \mathbb{N}}$ verwenden, weil im Grenzfall $n \rightarrow \infty$ - und nur dafür interessieren wir uns - beide f. s. die tatsächliche Verteilung von $Z + \varepsilon$ gleich gut wiedergeben.

Der empirische Prozess kann also über Gleichung (6.1) geschrieben werden als:

$$\begin{aligned} \sqrt{n}(\mu^*(x) - \hat{\mu}(x)) &\stackrel{\text{f. s.}}{\stackrel{D}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i^*(x_j) + \varepsilon_{ij}^*) - \hat{\mu}(x) \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i^*(x_j) + \varepsilon_{ij}^*) - \mathbb{E}^*[\mu^*(x)] + \mathbb{E}^*[\mu^*(x)] - \hat{\mu}(x) \right) \end{aligned}$$

f. s.

$\stackrel{D}{\simeq}$ steht hier für „asymptotisch in Verteilung f. s. gleich“, f. s. bzgl. der zufälligen Stichprobe Y . Zur weiteren Vereinfachung betrachten wir nun noch den bedingten Erwartungswert des Schätzers μ^* . Für alle $x \in [0, 1]$ gilt:

$$\begin{aligned} \mathbb{E}^*[\mu^*(x)] &= \mathbb{E}^* \left[\left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij} \right)^* \right] = \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) Y_{ij}^* \right] \\ &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i(x_j) + \varepsilon_{ij})^* \right] \\ &\stackrel{\text{f. s.}}{\simeq} \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + Z_i^*(x_j) + \varepsilon_{ij}^*) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\mu(x_j) + \mathbb{E}^*[Z_i^*(x_j)] + \mathbb{E}^*[\varepsilon_{ij}^*]) \end{aligned} \quad (6.3)$$

μ^* ist eine, auf Y bedingte Schätzfunktion für $\hat{\mu}$, also

$$\sum_{j=1}^p W_j(x) (\mu(x_j) + \mathbb{E}^*[Z_i^*(x_j)] + \mathbb{E}^*[\varepsilon_{ij}^*]) - \hat{\mu}(x) = \mathbb{E}^*[\mu^*(x)] - \hat{\mu}(x) =: \text{Bias}^*(\mu^*(x))$$

Da es sich um einen naiven Bootstrap handelt, ist μ_i^* wie in (B.2) definiert empirisch verteilt, wobei sich ähnlich zu Gleichung (6.2) und Proposition 6.4 die empirische Verteilung wieder auf beliebige Produkte bzgl. des Funktionsarguments überträgt:

$$\begin{aligned} \mu_1^*(x_1), \dots, \mu_n^*(x_1) \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\hat{\mu}_i(x_1)}, \quad x_1 \in [0, 1] \\ \mu_1^*(x_1)\mu_1^*(x_2), \dots, \mu_n^*(x_1)\mu_n^*(x_2) \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\hat{\mu}_i(x_1)\hat{\mu}_i(x_2)}, \quad x_1, x_2 \in [0, 1] \\ &\vdots \end{aligned} \tag{6.4}$$

Die Unabhängigkeit gilt wieder im Sinne von Definition 2.24, also bedingt auf die Stichprobe Y . Dadurch folgt, wegen der Festlegung von $\hat{\mu}$ aus (B.1), für den Erwartungswert des Schätzers für alle $x \in [0, 1]$:

$$\mathbb{E}^*[\mu^*(x)] = \mathbb{E}^*\left[\frac{1}{n} \sum_{i=1}^n \mu_i^*(x)\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mu_i^*(x)] \stackrel{\text{iid}}{=} \mathbb{E}^*[\mu_1^*(x)] = \frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(x) = \hat{\mu}(x) \tag{6.5}$$

Der Bootstrap-Schätzer ist also ein bedingt erwartungstreuer Schätzer für $\hat{\mu}$ und mit Gleichung (6.3) folgt für den empirischen Prozess:

$$\begin{aligned} \sqrt{n}(\mu^*(x) - \hat{\mu}(x)) &\stackrel{\text{f.s.}}{\underset{\text{D}}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (Z_i^*(x_j) - \mathbb{E}^*[Z_i^*(x_j)]) \right) \Bigg\} =: S_n^{1*}(x) \\ &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \Bigg\} =: S_n^{2*}(x) \end{aligned}$$

Die Grenzverteilung kann nun wieder wegen Lemma 2.15 durch Diskussion der einzelnen Summanden ermittelt werden.

$$S_n^{1*}(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \left(Z_i^*(x_j) - \mathbb{E}^* [Z_i^*(x_j)] \right) \right)$$

Dieser Term wird wieder mit dem funktionalen zentralen Grenzwertsatz aus Abschnitt 2.3.2 bearbeitet. Für $n \in \mathbb{N}$, $i = 1, \dots, n$ und $T = [0, 1]$ sei:

$$\begin{aligned} X_{n,i}^*(x) &= \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) \\ S_n^{1*}(x) &= \sum_{i=1}^n \left(X_{n,i}^*(x) - \mathbb{E}^* [X_{n,i}^*(x)] \right) \end{aligned} \quad (6.6)$$

Diese stochastischen Prozesse $X_{n,i}^*$ bilden bzgl. i ein Dreiecksschema mit $k_n = n$ und sind in jeder Zeile bedingt unabhängig, weil für $i = 1, \dots, n$ alle Z_i^* 's bedingt unabhängig sind. Mit dieser Festlegung gilt:

$$\begin{aligned} S_n^{1*}(x) &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) - \mathbb{E}^* \left[\frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) \right] \right) \\ &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) \mathbb{E}^* [Z_i^*(x_j)] \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \left(Z_i^*(x_j) - \mathbb{E}^* [Z_i^*(x_j)] \right) \right) \quad \checkmark \end{aligned} \quad (6.7)$$

Es müssen nun alle 5 Bedingungen des funktionalen zentralen Grenzwertsatzes f. s. nachgewiesen werden. Wir beginnen mit der Existenz einer Kovarianzfunktion für den Grenzprozess. Für $x, x' \in [0, 1]$ gilt nach Gleichung (6.7):

$$\begin{aligned}
 \mathbb{E}^* [S_n^{1*}(x)S_n^{1*}(x')] &= \mathbb{E}^* \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (Z_i^*(x_j) - \mathbb{E}[Z_i^*(x_j)]) \right) \right. \\
 &\quad \cdot \left. \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^p W_j(x') (Z_i^*(x_j) - \mathbb{E}^*[Z_i^*(x_j)]) \right) \right] \\
 &= \frac{1}{n} \mathbb{E}^* \left[\sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x') (Z_{i_1}^*(x_{j_1}) - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})]) \cdot (Z_{i_2}^*(x_{j_2}) - \mathbb{E}^*[Z_{i_2}^*(x_{j_2})]) \right] \\
 &= \frac{1}{n} \mathbb{E}^* \left[\sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x') (Z_{i_1}^*(x_{j_1}) Z_{i_2}^*(x_{j_2}) - Z_{i_1}^*(x_{j_1}) \mathbb{E}^*[Z_{i_2}^*(x_{j_2})] \right. \\
 &\quad \left. - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})] Z_{i_2}^*(x_{j_2}) + \mathbb{E}^*[Z_{i_1}^*(x_{j_1})] \mathbb{E}^*[Z_{i_2}^*(x_{j_2})]) \right] \\
 &= \frac{1}{n} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x') (\mathbb{E}^*[Z_{i_1}^*(x_{j_1}) Z_{i_2}^*(x_{j_2})] - 2\mathbb{E}^*[Z_{i_1}^*(x_{j_1})] \mathbb{E}^*[Z_{i_2}^*(x_{j_2})] \\
 &\quad + \mathbb{E}^*[Z_{i_1}^*(x_{j_1})] \mathbb{E}^*[Z_{i_2}^*(x_{j_2})]) \\
 &\stackrel{(*)}{=} \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x') (\mathbb{E}^*[Z_i^*(x_{j_1}) Z_i^*(x_{j_2})] \\
 &\quad - \mathbb{E}^*[Z_i^*(x_{j_1})] \mathbb{E}^*[Z_i^*(x_{j_2})]) \\
 &\stackrel{\text{iid}}{=} \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(x) Z_1(x_{j_1}) \right)^* \left(\sum_{j_2=1}^p W_{j_2}(x') Z_1(x_{j_2}) \right)^* \right] \\
 &\quad - \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(x) Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(x') Z_1(x_{j_2}) \right)^* \right] \\
 &\stackrel{(**)}{=} \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(x) Z_1(x_{j_1}) \right) \left(\sum_{j_2=1}^p W_{j_2}(x') Z_1(x_{j_2}) \right) \right] + o_{\text{f.s.}}(1) \\
 &\quad - \left(\mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(x) Z_1(x_{j_1}) \right) \right] + o_{\text{f.s.}}(1) \right) \left(\mathbb{E} \left[\left(\sum_{j_2=1}^p W_{j_2}(x') Z_1(x_{j_2}) \right) \right] + o_{\text{f.s.}}(1) \right) \\
 &= \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x') \mathbb{E} [Z_1(x_{j_1}) Z_1(x_{j_2})] + o_{\text{f.s.}}(1) \\
 &\quad - \left(\sum_{j_1=1}^p W_{j_1}(x) \underbrace{\mathbb{E} [Z_1(x_{j_1})]}_{=0} + o_{\text{f.s.}}(1) \right) \left(\sum_{j_2=1}^p W_{j_2}(x') \underbrace{\mathbb{E} [Z_1(x_{j_2})]}_{=0} + o_{\text{f.s.}}(1) \right) \\
 &= \sum_{j_1, j_2=1}^p W_{i_{j_1}}(x) W_{i_{j_2}}(x') R(x_{i_{j_1}}, x_{i_{j_2}}) + o_{\text{f.s.}}(1) \\
 &\stackrel{\text{Proposition 3.7}}{=} R(x, x') + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} R(x, x') \tag{6.8}
 \end{aligned}$$

(*): Wegen der bedingten Unabhängigkeit von Z_i^* und Lemma 2.25.

(**): Nach Proposition 6.3, was nach Proposition 3.3 und Proposition 3.6 anwendbar ist.

Bedingung i) ist also f. s. erfüllt.

Sei weiterhin $x, x' \in [0, 1]$.

$$\begin{aligned}
(d_{\sum n}^*(x, x'))^2 &= \sum_{i=1}^n \mathbb{E}^* \left[(X_{n,i}^*(x) - X_{n,i}^*(x'))^2 \right] \\
&= \sum_{i=1}^n \mathbb{E}^* \left[\left(\frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x') Z_i^*(x_j) \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* \left[\left(\sum_{j=1}^p (W_j(x) - W_j(x')) Z_i^*(x_j) \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* \left[\sum_{j_1, j_2=1}^p (W_{j_1}(x) - W_{j_1}(x')) (W_{j_2}(x) - W_{j_2}(x')) Z_i^*(x_{j_1}) Z_i^*(x_{j_2}) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p (W_{j_1}(x) - W_{j_1}(x')) (W_{j_2}(x) - W_{j_2}(x')) \mathbb{E}^* [Z_i^*(x_{j_1}) Z_i^*(x_{j_2})] \\
&\stackrel{\text{iid}}{=} \mathbb{E}^* \left[\left(\sum_{j_1=1}^p (W_{j_1}(x) - W_{j_1}(x')) Z_1(x_{j_1}) \right)^* \left(\sum_{j_2=1}^p (W_{j_2}(x) - W_{j_2}(x')) Z_1(x_{j_2}) \right)^* \right] \\
&\stackrel{(*)}{=} \mathbb{E} \left[\left(\sum_{j_1=1}^p (W_{j_1}(x) - W_{j_1}(x')) Z_1(x_{j_1}) \right) \left(\sum_{j_2=1}^p (W_{j_2}(x) - W_{j_2}(x')) Z_1(x_{j_2}) \right) \right] + o_{\text{f.s.}}(1) \\
&= \sum_{j_1, j_2=1}^p (W_{j_1}(x) - W_{j_1}(x')) (W_{j_2}(x) - W_{j_2}(x')) \mathbb{E} [Z_1(x_{j_1}) Z_1(x_{j_2})] + o_{\text{f.s.}}(1) \\
&= \sum_{j_1, j_2=1}^p (W_{j_1}(x) - W_{j_1}(x')) (W_{j_2}(x) - W_{j_2}(x')) R(x_{j_1}, x_{j_2}) + o_{\text{f.s.}}(1) \quad (6.9)
\end{aligned}$$

(*): Nach Proposition 6.3.

Ab hier ist das Vorgehen identisch zu dem aus dem Beweis von Satz 4.4. Bedingung ii) ist demnach ebenfalls f. s. erfüllt.

Für Bedingung iii) - v) benötigen wir wieder eine Umhüllende. Für alle $x, x' \in [0, 1]$ und $i = 1, \dots, n$ gilt nach (A.2) die Ungleichung:

$$|Z_i(x) - Z_i(x')| \stackrel{\text{f.s.}}{\leq} K_i |x - x'|^\beta \stackrel{x, x' \in [0, 1]}{\leq} K_i \quad (6.10)$$

Damit folgt nun für alle $x \in [0, 1]$ und $i = 1, \dots, n$:

$$|Z_i^*(x)| = |Z_i(x)|^* = |Z_i(0) + Z_i(x) - Z_i(0)|^* \leq (|Z_i(0)| + |Z_i(x) - Z_i(0)|)^* \stackrel{\text{f.s.}}{\leq} (|Z_i(0)| + K_i)^* \quad (6.11)$$

und daraus wiederum

$$\begin{aligned} |X_{n,i}^*(x)| &= \frac{1}{\sqrt{n}} \left| \sum_{j=1}^p W_j(x) Z_i^*(x_j) \right| \leq \frac{1}{\sqrt{n}} \sum_{j=1}^p |W_j(x)| |Z_i(x_j)|^* \leq \frac{1}{\sqrt{n}} (|Z_i(0)| + K_i)^* \sum_{j=1}^p |W_j(x)| \\ &\leq \frac{C_{W_1}}{\sqrt{n}} (|Z_i(0)| + K_i)^* =: \Phi_{n,i}^* \end{aligned}$$

$\Phi_n^* = (\Phi_{n,1}^*, \dots, \Phi_{n,n}^*)$ ist also eine Umhüllende für $X_{n,i}^*(x)$ und entspricht gerade der gebootstrapteten Version von Φ_n aus dem Beweis von Satz 4.4. Nun gilt für alle $k \in \mathbb{N}$ wegen der empirischen Verteilung die folgende Implikation:

$$\begin{aligned} (|Z_1(0)| + K_1)^*, \dots, (|Z_n(0)| + K_n)^* \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{|Z_i(0)| + K_i} \\ \Rightarrow (\Phi_{n,1}^*)^k, \dots, (\Phi_{n,n}^*)^k \mid Y &\stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{(\Phi_{n,i}^*)^k} \text{ mit } \Phi_{n,i} = \frac{C_{W_1}}{\sqrt{n}} (|Z_i(0)| + K_i) \quad (6.12) \end{aligned}$$

und außerdem mit (A.2) und Gleichung (4.29):

$$\mathbb{E}[|\Phi_{n,1}^*|^2] = \frac{C_{W_1}^2}{n} \mathbb{E}[(|Z_1(0)| + K_1)^2] \leq \underbrace{\frac{2C_{W_1}^2}{n}}_{< \infty} \cdot \left(\underbrace{\mathbb{E}[(Z_1(0))^2]}_{=R(0,0) \leq E_{Z,2} < \infty} + \underbrace{\mathbb{E}[(K_1)^2]}_{< \infty} \right) < \infty \quad (6.13)$$

da die K_i 's durch die Festlegung der Z_i 's unabhängige Kopien von K sind und dessen Eigenschaften erben. Somit ist für das zweite Moment der Einträge von Φ_n^* für alle $n \in \mathbb{N}$ Proposition 6.3 anwendbar.

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}^*[\Phi_{n,i}^2] &= \sum_{i=1}^n \mathbb{E}[\Phi_{n,1}^2] + o_{\text{f.s.}}(1) = \frac{C_{W_1}^2}{n} \sum_{i=1}^n \mathbb{E}[(|Z_1(0)| + K_1)^2] + o_{\text{f.s.}}(1) \\ &= C_{W_1}^2 \mathbb{E}[(|Z_1(0)| + K_1)^2] + o_{\text{f.s.}}(1) \leq 2C_{W_1}^2 \left(\underbrace{\mathbb{E}[(Z_1(0))^2]}_{=R(0,0) \leq E_{Z,2} < \infty} + \underbrace{\mathbb{E}[(K_1)^2]}_{< \infty} \right) + o_{\text{f.s.}}(1) \\ &\stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

für alle $n \in \mathbb{N}$ und damit auch für den \limsup .

Daraus folgt weiterhin für alle $\delta > 0$:

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E}^* [(\Phi_{n,i}^*)^2 \mathbf{1}_{\{\Phi_{n,i}^* > \delta\}}] &= \sum_{i=1}^n \mathbb{E} [\Phi_{n,1}^2 \mathbf{1}_{\{\Phi_{n,1} > \delta\}}] + o_{f.s.}(1) \\
 &= \frac{C_{W_1}^2}{n} \sum_{i=1}^n \mathbb{E} \left[(|Z_1(0)| + K_1)^2 \mathbf{1}_{\left\{ \frac{C_{W_1} (|Z_1(0)| + K_1)}{\sqrt{n}} > \delta \right\}} \right] + o_{f.s.}(1) \\
 &= C_{W_1}^2 \mathbb{E} \left[(|Z_1(0)| + K_1)^2 \underbrace{\mathbf{1}_{\{C_{W_1} (|Z_1(0)| + K_1) > \sqrt{n}\delta\}}}_{\rightarrow 0} \right] + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0}
 \end{aligned}$$

denn Proposition 6.3 ist auch für $(\Phi_{n,i}^*)^2 \mathbf{1}_{\{\Phi_{n,i}^* > \delta\}}$ anwendbar, da die bedingt unabhängige identische Verteilung vererbt wird und die Existenz der Momente durch Gleichung (6.13) für alle $\Phi_{n,i} \geq \delta > 0$ gesichert ist. Also gelten auch die Bedingungen iii) und iv) f. s.

Für Bedingung v) ist zu zeigen, dass für alle $n \in \mathbb{N}$, $i = 1, \dots, n$ und $x, x' \in [0, 1]$ eine Konstante $C \in \mathbb{R}$ existiert, so dass f. s.:

$$|x - x'| \leq \delta^\gamma \Rightarrow \left| \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x') Z_i^*(x_j) \right| \leq C \delta \Phi_{n,i}^*$$

Aufgrund der denkbar einfachen, jedoch erlaubten und äußerst effektiven Umformung:

$$\left| \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i^*(x_j) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x') Z_i^*(x_j) \right| = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x) Z_i(x_j) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_j(x') Z_i(x_j) \right|^*$$

sind im Inneren alle Schritte äquivalent Beweis von Satz 4.4 durchführbar, welche schließlich auch zur geforderten Implikation führen.

Also auch Bedingung v) f. s. erfüllt und nach dem funktionalen zentralen Grenzwertsatz gilt dadurch:

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \left(Z_i^*(x_j) - \mathbb{E}^* [Z_i^*(x_j)] \right) \right) | Y \xrightarrow{w} \mathcal{G}(0, R)$$

f. s. in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$, wobei R die Kovarianzfunktion der Zufallsfunktion Z darstellt.

$$S_n^{2*}(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right)$$

Die asymptotische Vernachlässigbarkeit wird im Sinne von Lemma 2.18 wieder durch Kombination von gleichmäßiger gleichgradiger Stetigkeit und punktwiser Konvergenz gezeigt, wobei natürlich auch hier nun, wegen der empirischen Verteilung von ε^* , mit bedingten Wahrscheinlichkeiten bzw. Erwartungswerten bzgl. Y gearbeitet werden muss. Die insgesamt gleichmäßige Konvergenz gilt also wieder nur f. s. bedingt auf Y .

f. s. Punktweise Konvergenz in \mathbb{P} :

Die f. s. punktweise Konvergenz gegen 0 in \mathbb{P} kann im Bootstrap sogar auf zwei verschiedene Arten gezeigt werden. Entweder wie im Rest der Arbeit als Folgerung aus der f. s. Konvergenz im 2-ten Mittel, oder über die Mallows' Distanz. Da beides durch die bereits erbrachte Vorarbeit relativ schnell vonstattengeht, werden wir für diesen einen Fall auch einmal beide Varianten durchführen.

Mittels Konvergenz im 2-ten Mittel:

Für alle $x \in [0, 1]$ gilt:

$$\begin{aligned} \mathbb{E}^*[(S_n^{2*}(x))^2] &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2 \right] \\ &= \frac{1}{n} \mathbb{E}^* \left[\sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) (\varepsilon_{i_1 j_1}^* - \mathbb{E}^*[\varepsilon_{i_1 j_1}^*]) (\varepsilon_{i_2 j_2}^* - \mathbb{E}^*[\varepsilon_{i_2 j_2}^*]) \right] \\ &= \frac{1}{n} \mathbb{E}^* \left[\sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) (\varepsilon_{i_1 j_1}^* \varepsilon_{i_2 j_2}^* - \varepsilon_{i_1 j_1}^* \mathbb{E}^*[\varepsilon_{i_2 j_2}^*] \right. \\ &\quad \left. - \mathbb{E}^*[\varepsilon_{i_1 j_1}^*] \varepsilon_{i_2 j_2}^* + \mathbb{E}^*[\varepsilon_{i_1 j_1}^*] \mathbb{E}^*[\varepsilon_{i_2 j_2}^*]) \right] \\ &= \frac{1}{n} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) (\mathbb{E}^*[\varepsilon_{i_1 j_1}^* \varepsilon_{i_2 j_2}^*] - \mathbb{E}^*[\varepsilon_{i_1 j_1}^*] \mathbb{E}^*[\varepsilon_{i_2 j_2}^*] \\ &\quad - \mathbb{E}^*[\varepsilon_{i_1 j_1}^*] \mathbb{E}^*[\varepsilon_{i_2 j_2}^*] + \mathbb{E}^*[\varepsilon_{i_1 j_1}^*] \mathbb{E}^*[\varepsilon_{i_2 j_2}^*]) \\ &\stackrel{(*)}{=} \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) (\mathbb{E}^*[\varepsilon_{i j_1}^* \varepsilon_{i j_2}^*] - \mathbb{E}^*[\varepsilon_{i j_1}^*] \mathbb{E}^*[\varepsilon_{i j_2}^*]) \\ &\stackrel{\text{iid}}{=} \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \underbrace{(\mathbb{E}^*[\varepsilon_{1 j_1}^* \varepsilon_{1 j_2}^*] - \mathbb{E}^*[\varepsilon_{1 j_1}^*] \mathbb{E}^*[\varepsilon_{1 j_2}^*])}_{=: \sigma_{j_1 j_2}^*} \end{aligned} \tag{6.14}$$

(*)): Wegen der bedingten Unabhängigkeit von ε_i^* und Lemma 2.25.

Dieses $\sigma_{j_1 j_2}^*$ wollen wir nun ein wenig genauer inspizieren. Nach Gleichung (6.2) gilt:

$$\sigma_{j_1 j_2}^* = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_1} \varepsilon_{ij_2} - \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_1} \right) \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_2} \right)$$

Es handelt sich dabei um die bedingte, über gebootstrappte Werte von ε , geschätzte Kovarianz von ε an den Stellen j_1 und j_2 , also letztlich f.s. die Stichprobenkovarianz von ε an den Stellen j_1 und j_2 . Die Stichprobenkovarianz ist unter unseren Annahmen stark konsistent, was sich leicht mit Lemma 2.15 und 2.16 zeigen lässt. Das bedeutet, für alle $j_1, j_2 = 1, \dots, p$ gilt $\sigma_{j_1 j_2}^* = \sigma_{j_1 j_2} + \hat{a}_n$ mit $\hat{a}_n \xrightarrow{\text{f.s.}} 0$, oder in Matrixform

$$\text{Cov}^*(\vec{\varepsilon}_1^*) = \text{Cov}(\vec{\varepsilon}_1) + \hat{A}_n \in \mathbb{R}^{p \times p} \text{ mit } \hat{A}_n \xrightarrow{\text{f.s.}} 0_p$$

$0_p \hat{=} p \times p$ Nullmatrix. \hat{A}_n ist für alle $n \in \mathbb{N}$ ebenfalls symmetrisch, denn durch die Symmetrie von $\text{Cov}^*(\vec{\varepsilon}_1^*)$ und $\text{Cov}(\vec{\varepsilon}_1)$ gilt:

$$\hat{A}_n^T = (\text{Cov}^*(\vec{\varepsilon}_1^*) - \text{Cov}(\vec{\varepsilon}_1))^T = \text{Cov}(\vec{\varepsilon}_1)^T - \text{Cov}^*(\vec{\varepsilon}_1^*)^T = \text{Cov}^*(\vec{\varepsilon}_1^*) - \text{Cov}(\vec{\varepsilon}_1) = \hat{A}_n$$

$\Rightarrow \hat{A}_n^T = \hat{A}_n$. Aus der Nullfolgen-Eigenschaft folgt außerdem $\|\hat{A}_n\|_2 \xrightarrow{\text{f.s.}} 0$, es muss also ein $n_0 \in \mathbb{N}$ geben für welches f.s. $\mu_n = \|\hat{A}_n\|_2 \leq \|\hat{A}_{n_0}\|_2 = \mu_{n_0} < \infty$ für alle $n \geq n_0$, dabei bezeichne μ_n den betragsmäßig größten Eigenwert von \hat{A}_n . Damit gilt für den maximalen Eigenwert $\lambda_{\max,1}^*$ von $\text{Cov}^*(\vec{\varepsilon}_1^*)$ als f.s. symmetrischer, positiv semidefiniter Matrix für alle $n \geq n_0$:

$$\lambda_{\max,1}^* = \|\text{Cov}^*(\vec{\varepsilon}_1^*)\|_2 = \|\text{Cov}(\vec{\varepsilon}_1) + \hat{A}_n\|_2 \leq \|\text{Cov}(\vec{\varepsilon}_1)\|_2 + \|\hat{A}_n\|_2 \stackrel{\text{f.s.}}{\leq} \lambda_{\max,1} + \mu_{n_0}$$

Nach (A.4) gilt $\lambda_{\max,1} = \mathcal{O}(p^{d_1})$ und da $p \rightarrow \infty$, $d_1 \in [0, 1)$ und $\mu_0 < \infty$, folgt daraus $\lambda_{\max,1}^* = \mathcal{O}(p^{d_1})$, wodurch Proposition 3.9 für $\text{Cov}^*(\vec{\varepsilon}_1^*)$ anwendbar ist. Es folgt schließlich für Gleichung (6.14):

$$\mathbb{E}^*[(S_n^{2*}(x))^2] = \sum_{j_1, j_2=1}^p W_{j_1}(x) W_{j_2}(x) \sigma_{j_1 j_2}^* = \vec{W}^T(x) \cdot \text{Cov}(\vec{\varepsilon}_1^*) \cdot \vec{W}(x) = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \xrightarrow{\text{f.s.}} 0 \quad (6.15)$$

$S_n^{2*}(x)$ konvergiert somit für alle $x \in [0, 1]$ bedingt auf Y f.s. gegen 0 in \mathbb{P} .

Mittels Mallows' Distanz:

In diesem Fall wird nicht ε^* einzeln betrachtet, sondern eine Verbindung zum ursprünglichen Fehler ε hergestellt, was mit der Mallows' Distanz gut gelingt.

Sei im Folgenden wieder $\vec{\varepsilon}_i := (\varepsilon_{i1}, \dots, \varepsilon_{ip})^T$ und $\vec{\varepsilon}_i^* := (\varepsilon_{i1}^*, \dots, \varepsilon_{ip}^*)^T$, also beides p -dimensionale Zufallsvariablen. Für ein festes $x \in [0, 1]$ ist $\vec{W}(x)^T$ eine reellwertige $1 \times p$ Matrix, wonach die Abbildung $\vec{v} \mapsto \vec{W}(x)^T \vec{v}$ für alle $\vec{v} \in \mathbb{R}^p$ ein linearer Operator auf \mathbb{R}^p ist. Für die Operatornorm $\|\cdot\|$ bietet sich hier die Zeilensummennorm als natürliche Matrixnorm an. Wir betrachten nun die Mallows' Distanz zwischen $S_n^{2*}(x)$ und $S_n^2(x)$ aus der Zerlegung des ursprünglichen Prozesses (siehe Beweis zu Satz 4.4) für ein festes $x \in [0, 1]$ und wenden die verschiedenen Eigenschaften aus Lemma 2.29 für die bedingte Mallows' Distanz an.

$$\begin{aligned}
 \rho_2^*(S_n^2(x), S_n^{2*}(x)) &= \rho_2^* \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \varepsilon_{ij} \right), \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right) \\
 &\stackrel{a)}{=} \frac{1}{\sqrt{n}} \rho_2^* \left(\sum_{i=1}^n \vec{W}(x)^T \vec{\varepsilon}_i, \sum_{i=1}^n \vec{W}(x)^T (\vec{\varepsilon}_i^* - \mathbb{E}^*[\vec{\varepsilon}_i^*]) \right) \\
 &\stackrel{c)}{=} \left(\frac{1}{n} \sum_{i=1}^n \rho_2^{*2}(\vec{W}(x)^T \vec{\varepsilon}_i, \vec{W}(x)^T (\vec{\varepsilon}_i^* - \mathbb{E}^*[\vec{\varepsilon}_i^*])) \right)^{\frac{1}{2}} \\
 &\stackrel{iid}{=} \left(\frac{1}{n} \sum_{i=1}^n \rho_2^{*2}(\vec{W}(x)^T \vec{\varepsilon}_1, \vec{W}(x)^T (\vec{\varepsilon}_1^* - \mathbb{E}^*[\vec{\varepsilon}_1^*])) \right)^{\frac{1}{2}} \\
 &= \rho_2^*(\vec{W}(x)^T \vec{\varepsilon}_1, \vec{W}(x)^T (\vec{\varepsilon}_1^* - \mathbb{E}^*[\vec{\varepsilon}_1^*])) \\
 &\stackrel{a)}{=} \|\vec{W}(x)^T\| \cdot \rho_2^*(\vec{\varepsilon}_1, \vec{\varepsilon}_1^* - \mathbb{E}^*[\vec{\varepsilon}_1^*]) = \sum_{i=1}^p |W_j(x)| \cdot \rho_2^*(\varepsilon_{1i}, \varepsilon_{1i}^* - \mathbb{E}^*[\varepsilon_{1i}^*]) \\
 &\leq C_{W_1} \cdot \rho_2^*(\vec{\varepsilon}_1, \vec{\varepsilon}_1^* - \mathbb{E}^*[\vec{\varepsilon}_1^*]) \stackrel{b)}{=} C_{W_1} \cdot (\rho_2^{*2}(\vec{\varepsilon}_1, \vec{\varepsilon}_1^*) - \|\mathbb{E}[\vec{\varepsilon}_1] - \mathbb{E}^*[\vec{\varepsilon}_1^*]\|^2)^{\frac{1}{2}} \\
 &= C_{W_1} \cdot \left(\underbrace{\rho_2^{*2}(\vec{\varepsilon}_1, \vec{\varepsilon}_1^*)}_{\stackrel{d)}{\xrightarrow{\text{f.s.}} 0}} - \underbrace{\left\| \mathbb{E}[\vec{\varepsilon}_1] - \frac{1}{n} \sum_{i=1}^n \vec{\varepsilon}_i \right\|^2}_{\stackrel{\text{f.s.}}{\xrightarrow{0}}} \right)^{\frac{1}{2}} \xrightarrow{\text{f.s.}} 0
 \end{aligned}$$

$\frac{1}{n} \sum_{i=1}^n \vec{\varepsilon}_i$ konvergiert als konsistenter empirischer Stichprobenmittelwert f. s. gegen $\mathbb{E}[\vec{\varepsilon}_1]$.

Wie wir aus dem Beweis zu Satz 4.4 wissen, konvergiert $S_n^2(x)$ in Wahrscheinlichkeit, also auch in Verteilung, gegen 0. Nach Lemma 2.29 f) konvergiert damit $S_n^{2*}(x)$ f. s. in Verteilung gegen 0 und da 0 eine Konstante ist, nach Proposition 2.13 auch f. s. in \mathbb{P} .

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei $\Phi^*(x) = \|S_n^{2*}(x)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2}$. Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $x \in [0, 1]$: $|S_n^{2*}(x)| \leq \Phi^*(x) \Leftrightarrow |S_n^{2*}| \leq \Phi^*$. Mit Gleichung (6.14) und (6.15) folgt weiterhin:

$$\begin{aligned} \mathbb{E}^*[|\Phi^*(x)|^2] &= \mathbb{E}^*[\|S_n^{2*}(x)\|_{\ell^2}^2] = \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2 \right] = \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^*[|\Phi^*|^2] &\stackrel{\text{f.s.}}{\leq} C \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1} h} < \infty \Rightarrow (\mathbb{E}^*[|\Phi^*|^2])^{\frac{1}{2}} \stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

zu ii):

Wir schätzen wieder die Überdeckungsanzahl bzgl. der kanonischen Metrik ab. Auf gleicher Argumentation wie bei Gleichung (6.14) und (6.15) folgt mit Proposition 3.9 für alle $x, x' \in [0, 1]$:

$$\begin{aligned} d_n^{*2}(x, x') &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x') (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2 \right] \\ &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_j(x) - W_j(x')) (\varepsilon_{ij}^* - \mathbb{E}^*[\varepsilon_{ij}^*]) \right) \right)^2 \right] \\ &= \sum_{j_1, j_2=1}^p (W_{j_1}(x) - W_{j_1}(x')) (W_{j_2}(x) - W_{j_2}(x')) \sigma_{j_1 j_2}^* \\ &= (\vec{W}(x) - \vec{W}(x'))^T \text{Cov}^*(\vec{\varepsilon}_1^*) (\vec{W}(x) - \vec{W}(x')) = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}(x, x') &\stackrel{\text{f.s.}}{\leq} \frac{C}{p^{1-d_1} h} \Rightarrow d_n^*(x, x') \stackrel{\text{f.s.}}{\leq} \left(\frac{C}{p^{1-d_1} h} \right)^{\frac{1}{2}} =: C_n \end{aligned}$$

Ab hier sind alle weiteren Schritte identisch zu denen von Gleichung (4.14) und folgende, also ist $S_n^{2*}(x)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} . \square

6.2. Kovarianzschätzer

Satz 6.5.

Die in Satz 5.7 angeführten Konfidenzbänder $B_{\gamma,n,R}(s,t)$ behalten unter sonst gleichen Bedingungen f. s. ihre Gültigkeit, wenn $\hat{c}_{\gamma,R}$ durch $c_{\gamma,R}^*$, ermittelt durch folgendes Bootstrap-Verfahren, ersetzt wird:

(B.1) Berechne $\hat{\mu}_i := (\hat{\mu}_i(s), \hat{\mu}_i(t))$ für alle $i = 1, \dots, n$ und daraus $\hat{\mu}$ und \hat{R} nach Gleichung (3.2) und (3.4).

(B.2) Ermittle eine Bootstrap-Stichprobe μ_1^*, \dots, μ_n^* , wobei μ_i^* für alle $i = 1, \dots, n$ durch Ziehen mit Zurücklegen aus $\{\hat{\mu}_1, \dots, \hat{\mu}_n\}$ bestimmt wird.

(B.3) Berechne für alle $i = 1, \dots, n$

$$a) R_i^*(s,t) = (\mu_i^*(s) - \hat{\mu}(s))(\mu_i^*(t) - \hat{\mu}(t)) \text{ und daraus } R^*(s,t) = \frac{1}{n} \sum_{i=1}^n R_i^*(s,t)$$

oder

$$b) R_i^*(s,t) = (\mu_i^*(s) - \mu^*(s))(\mu_i^*(t) - \mu^*(t)) \text{ mit } \mu^* = \frac{1}{n} \sum_{i=1}^n \mu_i^* \text{ und daraus}$$

$$R^*(s,t) = \frac{1}{n-1} \sum_{i=1}^n R_i^*(s,t).$$

(B.4) Berechne $X_{R^*} = \sqrt{n} \|R^* - \hat{R}\|_{\infty}$.

(B.5) Wiederhole (B.2), (B.3) und (B.4) N mal und verwende die bedingte empirische Verteilung von $(X_{R^*1}, \dots, X_{R^*N}) \mid Y$ zur Bestimmung des $(1 - \gamma) \cdot 100\%$ -Quantils $c_{\gamma,R}^*$, wobei X_{R^*i} das X_{R^*} aus dem i -ten Durchlauf und $Y := (Y_{ij})_{i \in \{1, \dots, n\}, j \in \{1, \dots, p\}}$ die Stichprobe ist.

Man sollte hier natürlich innerhalb der N Durchläufe nicht zwischen (B.3) a) und b) wechseln sondern sich von vornherein für eine der beiden Methoden entscheiden.

Bemerkung 6.6.

Werden die Bänder $B'_{\gamma,n,R}(s,t)$ aus Bemerkung 5.8 verwendet, so muss (B.4) $X_{R^*} = \sqrt{n} \left\| \frac{R^* - \hat{R}}{\sigma_{R^*}} \right\|_{\infty}$ berechnet werden, wobei σ_{R^*} die funktionale empirische Standardabweichung der R_i^* 's ist. Wird zudem der Schätzer \hat{R} aus Gleichung (3.5) betrachtet, so muss die Berechnung von R^* in (B.3) a) durch $R^*(s,t) = \frac{1}{n-1} \sum_{i=1}^n R_i^*(s,t)$ und in b) durch

$$R^*(s,t) = \frac{1}{n} \sum_{i=1}^n R_i^*(s,t) \text{ angepasst werden.}$$

Die verschiedenen Berechnungen rühren daher, dass R^* wieder möglichst ein bedingt erwartungstreuer Schätzer für \hat{R} sein soll. Für μ_i^* und dessen Produkte bzgl. verschiedener Argumente, kann die Verteilung wieder aus Gleichung (6.4) entnommen werden. Es gilt für R^* nach (B.3) a) bedingt auf die Stichprobe Y mit Gleichung (6.5) und der Y -Messbarkeit von $\hat{\mu}$:

$$\begin{aligned} \mathbb{E}^*[R^*(s,t)] &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \hat{\mu}(s)) (\mu_i^*(t) - \hat{\mu}(t)) \right] \\ &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \mu_i^*(s) \mu_i^*(t) - \hat{\mu}(t) \frac{1}{n} \sum_{i=1}^n \mu_i^*(s) - \hat{\mu}(s) \frac{1}{n} \sum_{i=1}^n \mu_i^*(t) + \hat{\mu}(s) \hat{\mu}(t) \right] \\ &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \mu_i^*(s) \mu_i^*(t) - \hat{\mu}(t) \mu^*(s) - \hat{\mu}(s) \mu^*(t) + \hat{\mu}(s) \hat{\mu}(t) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mu_i^*(s) \mu_i^*(t)] - \hat{\mu}(t) \mathbb{E}^*[\mu^*(s)] - \hat{\mu}(s) \mathbb{E}^*[\mu^*(t)] + \hat{\mu}(s) \hat{\mu}(t) \\ &= \mathbb{E}^*[\mu_1^*(s) \mu_1^*(t)] - \hat{\mu}(s) \hat{\mu}(t) - \hat{\mu}(s) \hat{\mu}(t) + \hat{\mu}(s) \hat{\mu}(t) \\ &= \frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) = \hat{R}(s,t) \end{aligned} \tag{6.16}$$

Nach (B.3) b) gilt für R^* wiederum:

$$\begin{aligned}
 \mathbb{E}^* [R^*(s, t)] &= \mathbb{E}^* \left[\frac{1}{n-1} \sum_{i=1}^n (\mu_i^*(s) - \mu^*(s)) (\mu_i^*(t) - \mu^*(t)) \right] \\
 &= \mathbb{E}^* \left[\frac{1}{n-1} \sum_{i=1}^n \mu_i^*(s) \mu_i^*(t) - \mu^*(s) \frac{1}{n-1} \sum_{i=1}^n \mu_i^*(s) - \mu^*(s) \frac{1}{n-1} \sum_{i=1}^n \mu_i^*(t) \right. \\
 &\quad \left. + \frac{n}{n-1} \mu^*(s) \mu^*(t) \right] \\
 &= \mathbb{E}^* \left[\frac{1}{n-1} \sum_{i=1}^n \mu_i^*(s) \mu_i^*(t) - \frac{n}{n-1} \mu^*(s) \mu^*(t) - \frac{n}{n-1} \mu^*(s) \mu^*(t) + \frac{n}{n-1} \mu^*(s) \mu^*(t) \right] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}^* [\mu_i^*(s) \mu_i^*(t)] - \frac{n}{n-1} \mathbb{E}^* [\mu^*(s) \mu^*(t)] \\
 &= \frac{n}{n-1} \left(\mathbb{E}^* [\mu_1^*(s) \mu_1^*(t)] - \mathbb{E}^* \left[\left(\frac{1}{n} \sum_{i=1}^n \mu_i^*(s) \right) \left(\frac{1}{n} \sum_{i=1}^n \mu_i^*(t) \right) \right] \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \mathbb{E}^* \left[\frac{1}{n^2} \sum_{i_1, i_2=1}^n \mu_{i_1}^*(s) \mu_{i_2}^*(t) \right] \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \frac{1}{n^2} \sum_{i_1, i_2=1}^n \mathbb{E}^* [\mu_{i_1}^*(s) \mu_{i_2}^*(t)] \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{E}^* [\mu_i^*(s) \mu_i^*(t)] + \sum_{i_1 \neq i_2}^n \mathbb{E}^* [\mu_{i_1}^*(s)] \mathbb{E}^* [\mu_{i_2}^*(t)] \right) \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \frac{1}{n^2} \left(n \cdot \mathbb{E}^* [\mu_1^*(s) \mu_1^*(t)] + \sum_{i_1 \neq i_2}^n \hat{\mu}(s) \hat{\mu}(t) \right) \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \frac{1}{n^2} \left(\sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) + n \cdot (n-1) \cdot \hat{\mu}(s) \hat{\mu}(t) \right) \right) \\
 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) - \frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) \right) \right) \\
 &= \frac{n}{n-1} \cdot \frac{n-1}{n} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) \right) = \hat{R}(s, t) \tag{6.17}
 \end{aligned}$$

Verwendet man für \hat{R} die korrigierte Version aus Gleichung (3.5), so ergibt sich:

$$\begin{aligned}
a) : R^*(s, t) &= \frac{1}{n-1} \sum_{i=1}^n (\mu_i^*(s) - \hat{\mu}(s)) (\mu_i^*(t) - \hat{\mu}(t)) = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \hat{\mu}(s)) (\mu_i^*(t) - \hat{\mu}(t)) \\
\Rightarrow \mathbb{E}^* [R^*(s, t)] &= \frac{n}{n-1} \cdot \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \hat{\mu}(s)) (\mu_i^*(t) - \hat{\mu}(t)) \right] \\
&\stackrel{\text{Gleichung (6.16)}}{=} \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i(s) - \hat{\mu}(s)) (\hat{\mu}_i(t) - \hat{\mu}(t)) = \hat{R}(s, t) \\
b) : R^*(s, t) &= \frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \mu^*(s)) (\mu_i^*(t) - \mu^*(t)) = \frac{n-1}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\mu_i^*(s) - \mu^*(s)) (\mu_i^*(t) - \mu^*(t)) \\
\Rightarrow \mathbb{E}^* [R^*(s, t)] &= \frac{n-1}{n} \cdot \mathbb{E}^* \left[\frac{1}{n-1} \sum_{i=1}^n (\mu_i^*(s) - \mu^*(s)) (\mu_i^*(t) - \mu^*(t)) \right] \\
&\stackrel{\text{Gleichung (6.17)}}{=} \frac{n}{n-1} \cdot \frac{n}{n-1} \cdot \frac{n-1}{n} \left(\frac{1}{n} \sum_{i=1}^n \hat{\mu}_i(s) \hat{\mu}_i(t) - \hat{\mu}(s) \hat{\mu}(t) \right) = \hat{R}(s, t)
\end{aligned}$$

Bei einem falschen Vorfaktor würde wieder der klassische Verzerrungsfaktor $\frac{1}{n}$ auftauchen, was bei einer größeren Stichprobe wegen $\frac{1}{n} \rightarrow 0$ zwar keine große Rolle spielt, sich bei einer kleineren Stichprobe jedoch durchaus bemerkbar machen kann. Gerade weil der Bootstrap für kleine Stichproben gedacht ist, sollte man in diesem Fall die unverzerrte Variante heranziehen.

Beide Bootstrap-Methoden funktionieren, es stellt sich jedoch nun die Frage, welche Methode die bessere ist. Bezüglich der Berechnungsdauer ist (B.3) a) besser, weil in (B.3) b) in jedem Bootstrap-Durchlauf μ^* neu berechnet werden muss. Bezüglich der Überdeckung ist wiederum (B.3) b) zum Teil leicht im Vorteil, weil die so berechneten Bänder eine etwas größere Spannweite aufweisen. Darauf werden wir in Kapitel 11 noch genauer eingehen.

Beweis.

Wir validieren aus Gründen des Umfangs o. B. d. A. nur den Schätzer nach (B.3) a).

Nach gleicher Argumentation wie im Beweis zu Satz 6.1, muss $\sqrt{n}(R^* - \hat{R}) \mid Y \xrightarrow{w} \mathcal{G}(0, \mathcal{R})$ in $(\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$ f. s. gezeigt werden, bzw. für Bemerkung 6.6 $\sqrt{n} \left(\frac{R^* - \hat{R}}{\sigma_{R^*}} \right) \mid Y \xrightarrow{w} \mathcal{G}(0, \rho_{ZZ})$. Letzteres müsste nun explizit gezeigt werden, da wir uns, wie in Bemerkung 5.8 bereits erwähnt, einer gleichmäßig schwachen Konsistenz von $\hat{\mathcal{R}}$ nicht sicher sein können und damit Lemma 2.15 evtl. nicht anwendbar ist. Da die Simulationen aus Kapitel 11 suggerieren, dass die Bänder nach $B'_{\gamma, n, R}(s, t)$ für gewisse Settings nicht wirklich einen Vorteil bieten, verzichten wir auch an dieser Stelle auf den Beweis.

Zunächst überlegen wir uns, dass für alle $x \in [0, 1]$ und alle $i = 1, \dots, n$, zumindest asymptotisch f. s. nach der Argumentation wie bei Gleichung (6.1) und den Verteilungseigenschaften nach Gleichung (6.2), gilt:

$$\begin{aligned}
 \mu_i^*(x) - \hat{\mu}(x) &= \left(\sum_{j=1}^p W_j(x) Y_{ij} \right)^* - \sum_{j=1}^p W_j(x) \frac{1}{n} \sum_{i=1}^n Y_{ij} = \sum_{j=1}^p W_j(x) \left(Y_{ij}^* - \frac{1}{n} \sum_{i=1}^n Y_{ij} \right) \\
 &\stackrel{\text{f.s.}}{\approx} \sum_{j=1}^p W_j(x) \left(\mu(x_j) + Z_i^*(x_j) + \varepsilon_{ij}^* - \frac{1}{n} \sum_{i=1}^n (\mu(x_j) + Z_i(x_j) + \varepsilon_{ij}) \right) \\
 &= \sum_{j=1}^p W_j(x) \left(\underbrace{\mu(x_j) - \frac{1}{n} \sum_{i=1}^n \mu(x_j)}_{=0} + Z_i^*(x_j) - \frac{1}{n} \sum_{i=1}^n Z_i(x_j) + \varepsilon_{ij}^* - \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij} \right) \\
 &= \sum_{j=1}^p \underbrace{\left(W_j(x) Z_i^*(x_j) - \mathbb{E}^* [W_j(x) Z_i^*(x_j)] \right)}_{=: WZ_i^*(x)} + \sum_{j=1}^p \underbrace{\left(W_j(x) \varepsilon_{ij}^* - \mathbb{E}^* [W_j(x) \varepsilon_{ij}^*] \right)}_{=: W\varepsilon_i^*(x)}
 \end{aligned}$$

Damit gilt für den Bootstrap-Schätzer:

$$\begin{aligned}
 R^*(s, t) &= \frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \hat{\mu}(s)) (\mu_i^*(t) - \hat{\mu}(t)) = \frac{1}{n} \sum_{i=1}^n (WZ_i^*(s) + W\varepsilon_i^*(s)) (WZ_i^*(t) + W\varepsilon_i^*(t)) \\
 &= \frac{1}{n} \sum_{i=1}^n (WZ_i^*(s)WZ_i^*(t) + WZ_i^*(s)W\varepsilon_i^*(t) + WZ_i^*(t)W\varepsilon_i^*(s) + W\varepsilon_i^*(s)W\varepsilon_i^*(t)) \\
 \Rightarrow \mathbb{E}^* [R^*(s, t)] &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}^* [WZ_i^*(s)WZ_i^*(t)] + \mathbb{E}^* [WZ_i^*(s)] \mathbb{E}^* [W\varepsilon_i^*(t)] + \mathbb{E}^* [WZ_i^*(t)] \mathbb{E}^* [W\varepsilon_i^*(s)] \\
 &\quad + \mathbb{E}^* [W\varepsilon_i^*(s)W\varepsilon_i^*(t)])
 \end{aligned}$$

da sich die Unabhängigkeit von Z und ε auf WZ^* und $W\varepsilon^*$ bedingt auf Y überträgt.

Nun kann der empirische Prozess mit der bedingten Erwartungstreue von R^* nach Gleichung (6.16) wie folgt zerlegt werden:

$$\begin{aligned}
\sqrt{n}(R^*(s, t) - \hat{R}(s, t)) &= \sqrt{n}(R^*(s, t) - \mathbb{E}^*[R^*(s, t)]) \\
&\stackrel{\text{f.s.}}{\underset{D}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(W Z_i^*(s) W Z_i^*(t) - \mathbb{E}^*[W Z_i^*(s) W Z_i^*(t)] \right) \right) \\
&\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(W \varepsilon_i^*(s) W \varepsilon_i^*(t) - \mathbb{E}^*[W \varepsilon_i^*(s) W \varepsilon_i^*(t)] \right) \right) \\
&\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(W Z_i^*(s) W \varepsilon_i^*(t) - \mathbb{E}^*[W Z_i^*(s)] \mathbb{E}^*[W \varepsilon_i^*(t)] \right) \right) \\
&\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(W Z_i^*(t) W \varepsilon_i^*(s) - \mathbb{E}^*[W Z_i^*(t)] \mathbb{E}^*[W \varepsilon_i^*(s)] \right) \right)
\end{aligned}$$

Wir vereinfachen o. B. d. A. den ersten Summanden:

$$\begin{aligned}
W Z_i^*(s) W Z_i^*(t) &= \left(\sum_{j=1}^p \left(W_j(s) Z_i^*(x_j) - \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \right) \right) \left(\sum_{j=1}^p \left(W_j(t) Z_i^*(x_j) - \mathbb{E}^*[W_j(t) Z_i^*(x_j)] \right) \right) \\
&= \sum_{j,k=1}^p \left(W_j(s) Z_i^*(x_j) W_k(t) Z_i^*(x_k) - W_j(s) Z_i^*(x_j) \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \right. \\
&\quad \left. - W_k(t) Z_i^*(x_k) \mathbb{E}^*[W_j(s) Z_i^*(x_j)] + \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \right) \\
\Rightarrow \frac{1}{n} \sum_{i=1}^n W Z_i^*(s) W Z_i^*(t) &\stackrel{\text{iid}}{=} \sum_{j,k=1}^p \left(\frac{1}{n} \sum_{i=1}^n W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \underbrace{\frac{1}{n} \sum_{i=1}^n W_j(s) Z_i^*(x_j)}_{\mathbb{E}^*[W_j(s) Z_i^*(x_j)]} \right. \\
&\quad \left. - \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \underbrace{\frac{1}{n} \sum_{i=1}^n W_k(t) Z_i^*(x_k)}_{\mathbb{E}^*[W_k(t) Z_i^*(x_k)]} + \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \right) \\
&= \sum_{j,k=1}^p \left(\frac{1}{n} \sum_{i=1}^n W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \right) \\
\Rightarrow \frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[W Z_i^*(s) W Z_i^*(t)] &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n W Z_i^*(s) W Z_i^*(t) \right] \\
&= \sum_{j,k=1}^p \left(\frac{1}{n} \sum_{i=1}^n W_j(s) W_k(t) \mathbb{E}^*[Z_i^*(x_j) Z_i^*(x_k)] - \mathbb{E}^*[W_j(s) Z_i^*(x_j)] \mathbb{E}^*[W_k(t) Z_i^*(x_k)] \right) \\
\Rightarrow \frac{1}{n} \sum_{i=1}^n \left(W Z_i^*(s) W Z_i^*(t) - \mathbb{E}^*[W Z_i^*(s) W Z_i^*(t)] \right) &= \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^*[Z_i^*(x_j) Z_i^*(x_k)] \right)
\end{aligned} \tag{6.18}$$

Genau auf die gleiche Weise können die anderen 3 Summanden ebenfalls umgeformt werden, wodurch sich die Zerlegung vereinfacht zu:

$$\begin{aligned}
 \sqrt{n}(R^*(s, t) - \hat{R}(s, t)) &\stackrel{\text{f.s.}}{\cong} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^* [Z_i^*(x_j) Z_i^*(x_k)] \right) \right) \Bigg\} =: S_n^{1*}(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*] \right) \right) \Bigg\} =: S_n^{2*}(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^* [Z_i^*(x_j)] \mathbb{E}^* [\varepsilon_{ik}^*] \right) \right) \Bigg\} =: S_n^{3*}(s, t) \\
 &+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_k) \varepsilon_{ij}^* - \mathbb{E}^* [Z_i^*(x_k)] \mathbb{E}^* [\varepsilon_{ij}^*] \right) \right) \Bigg\} =: S_n^{4*}(s, t)
 \end{aligned}$$

Praktischerweise muss hier überhaupt nichts ergänzt werden, was bei der Diskussion der asymptotischen Verteilung von \hat{R} so nicht funktioniert hat, eben weil \hat{R} kein erwartungstreuer Schätzer für R ist. Wir erhalten die Grenzverteilung wieder durch Diskussion der einzelnen Summanden, wobei die Verteilungen der verschiedenen Verbundmomente von Z^* und ε^* aus Gleichung (6.2) und Proposition 6.4 entnommen werden können:

$$S_n^{1*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^* [Z_i^*(x_j) Z_i^*(x_k)] \right) \right)$$

Für diesen Summanden wird der funktionale zentrale Grenzwertsatz verwendet.

Für $n \in \mathbb{N}$, $i = 1, \dots, n$ und $T = [0, 1]^2$ sei:

$$\begin{aligned}
 X_{n,i}^*(s, t) &= \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) \tag{6.19} \\
 S_n^{1*}(s, t) &= \sum_{i=1}^n \left(X_{n,i}^*(s, t) - \mathbb{E}^* [X_{n,i}^*(s, t)] \right)
 \end{aligned}$$

Diese stochastischen Prozesse $X_{n,i}^*$ bilden ein Dreiecksschema mit $k_n = n$ und sind nach Proposition 6.4 in jeder Zeile bedingt unabhängig.

Mit dieser Festlegung folgt:

$$\begin{aligned}
S_n^{1*}(s, t) &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i^*(x_j)Z_i^*(x_k) - \mathbb{E}^* \left[\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i^*(x_j)Z_i^*(x_k) \right] \right) \\
&= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)Z_i^*(x_j)Z_i^*(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s)W_k(t)\mathbb{E}^*[Z_i^*(x_j)Z_i^*(x_k)] \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)Z_i^*(x_j)Z_i^*(x_k) - \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t)\mathbb{E}^*[Z_i^*(x_j)Z_i^*(x_k)] \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) \left(Z_i^*(x_j)Z_i^*(x_k) - \mathbb{E}^*[Z_i^*(x_j)Z_i^*(x_k)] \right) \right) \quad \checkmark \quad (6.20)
\end{aligned}$$

Es müssen nun alle 5 Bedingungen des funktionalen zentralen Grenzwertsatzes f. s. nachgewiesen werden. Wir beginnen mit der Existenz einer Kovarianzfunktion für den Grenzprozess. Für $(s, t), (s', t') \in [0, 1]^2$ gilt:

$$\begin{aligned}
&\mathbb{E}^*[S_n^{1*}(s, t)S_n^{1*}(s', t')] \\
&= \mathbb{E}^* \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) \left(Z_i^*(x_j)Z_i^*(x_k) - \mathbb{E}^*[Z_i^*(x_j)Z_i^*(x_k)] \right) \right) \right. \\
&\quad \cdot \left. \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s')W_k(t') \left(Z_i^*(x_j)Z_i^*(x_k) - \mathbb{E}^*[Z_i^*(x_j)Z_i^*(x_k)] \right) \right) \right] \\
&= \frac{1}{n} \mathbb{E}^* \left[\sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t') \left(Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2}) \right. \right. \\
&\quad \left. \left. - Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2}) \right. \right. \\
&\quad \left. \left. + \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right) \right] \\
&= \frac{1}{n} \sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t') \left(\mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right. \\
&\quad \left. - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right. \\
&\quad \left. + \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right) \\
&= \frac{1}{n} \sum_{\substack{i_1, i_2=1 \\ j_1, k_1=1 \\ j_2, k_2=1}}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s')W_{k_2}(t') \left(\mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right. \\
&\quad \left. - \mathbb{E}^*[Z_{i_1}^*(x_{j_1})Z_{i_1}^*(x_{k_1})]\mathbb{E}^*[Z_{i_2}^*(x_{j_2})Z_{i_2}^*(x_{k_2})] \right)
\end{aligned}$$

$$\begin{aligned}
 & \stackrel{(*)}{=} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \left(\mathbb{E}^* [Z_i^*(x_{j_1}) Z_i^*(x_{k_1}) Z_i^*(x_{j_2}) Z_i^*(x_{k_2})] \right. \\
 & \quad \left. - \mathbb{E}^* [Z_i^*(x_{j_1}) Z_i^*(x_{k_1})] \mathbb{E}^* [Z_i^*(x_{j_2}) Z_i^*(x_{k_2})] \right) \\
 & \stackrel{\text{iid}}{=} \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right)^* \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right)^* \right] \\
 & \quad - \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right)^* \right] \\
 & \stackrel{(**)}{=} \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right) \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right) \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right) \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right) \right] + o_{\text{f.s.}}(1) \\
 & \quad - \left(\mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right) \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right) \right] + o_{\text{f.s.}}(1) \right) \\
 & \quad \cdot \left(\mathbb{E} \left[\left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right) \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right) \right] + o_{\text{f.s.}}(1) \right) \\
 & = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \mathbb{E} [Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2}) Z_1(x_{k_2})] + o_{\text{f.s.}}(1) \\
 & \quad - \left(\sum_{j_1, k_1=1}^p W_{j_1}(s) W_{k_1}(t) \mathbb{E} [Z_1(x_{j_1}) Z_1(x_{k_1})] + o_{\text{f.s.}}(1) \right) \left(\sum_{j_2, k_2=1}^p W_{j_2}(s') W_{k_2}(t') \mathbb{E} [Z_1(x_{j_2}) Z_1(x_{k_2})] + o_{\text{f.s.}}(1) \right) \\
 & = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2}) \\
 & \quad - \left(\sum_{j_1, k_1=1}^p W_{j_1}(s) W_{k_1}(t) R(x_{j_1}, x_{k_1}) \right) \left(\sum_{j_2, k_2=1}^p W_{j_2}(s') W_{k_2}(t') R(x_{j_2}, x_{k_2}) \right) + o_{\text{f.s.}}(1) \\
 & \stackrel{\text{Proposition 3.7}}{=} M(s, t, s', t') + \mathcal{O}(h^{1+\alpha_M}) - (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) (R(s', t') + \mathcal{O}(h^{1+\alpha_R})) + o_{\text{f.s.}}(1) \\
 & \stackrel{|\cdot| \leq E_{Z,2}}{=} M(s, t, s', t') - R(s, t) R(s', t') + \underbrace{\mathcal{O}(h^{1+\alpha_M})}_{\rightarrow 0} + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \\
 & \stackrel{\text{Proposition 3.3}}{\xrightarrow{\text{f.s.}}} M(s, t, s', t') - R(s, t) R(s', t') \tag{6.21}
 \end{aligned}$$

(*): da $\mathbb{E}^* [Z_{i_1}^*(x_{j_1}) Z_{i_1}^*(x_{k_1}) Z_{i_2}^*(x_{j_2}) Z_{i_2}^*(x_{k_2})] \stackrel{i_1 \neq i_2}{=} \mathbb{E}^* [Z_{i_1}^*(x_{j_1}) Z_{i_1}^*(x_{k_1})] \mathbb{E}^* [Z_{i_2}^*(x_{j_2}) Z_{i_2}^*(x_{k_2})]$, legitimiert durch Lemma 2.25.

(**): Nach Proposition 6.3, was nach Proposition 3.3 und Proposition 3.6 anwendbar ist.

Bedingung i) ist also f. s. erfüllt. Sei weiterhin $(s, t), (s', t') \in [0, 1]^2$.

$$\begin{aligned}
& \left(d_{\Sigma n}^*((s, t), (s', t')) \right)^2 = \sum_{i=1}^n \mathbb{E}^* \left[\left(X_{n,i}^*(s, t) - X_{n,i}^*(s', t') \right)^2 \right] \\
&= \sum_{i=1}^n \mathbb{E}^* \left[\left(\frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s') W_k(t') Z_i^*(x_j) Z_i^*(x_k) \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* \left[\left(\sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) Z_i^*(x_j) Z_i^*(x_k) \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* \left[\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) Z_i^*(x_{j_1}) Z_i^*(x_{k_1}) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) Z_i^*(x_{j_2}) Z_i^*(x_{k_2}) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) \mathbb{E}^* [Z_i^*(x_{j_1}) Z_i^*(x_{k_1}) Z_i^*(x_{j_2}) Z_i^*(x_{k_2})] \\
&\stackrel{\text{iid}}{=} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) \mathbb{E}^* [Z_1^*(x_{j_1}) Z_1^*(x_{k_1}) Z_1^*(x_{j_2}) Z_1^*(x_{k_2})] \\
&= \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right)^* \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \left(\sum_{k_2=1}^p W_{k_2}(t) Z_1(x_{k_2}) \right)^* \right] \\
&\quad - 2 \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right)^* \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right)^* \right] \\
&\quad + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s') Z_1(x_{j_1}) \right)^* \left(\sum_{k_1=1}^p W_{k_1}(t') Z_1(x_{k_1}) \right)^* \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \left(\sum_{k_2=1}^p W_{k_2}(t) Z_1(x_{k_2}) \right)^* \right] \\
&\stackrel{(*)}{=} \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right) \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right) \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right) \left(\sum_{k_2=1}^p W_{k_2}(t) Z_1(x_{k_2}) \right) \right] + o_{\text{f.s.}}(1) \\
&\quad - 2 \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right) \left(\sum_{k_1=1}^p W_{k_1}(t) Z_1(x_{k_1}) \right) \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right) \left(\sum_{k_2=1}^p W_{k_2}(t') Z_1(x_{k_2}) \right) \right] + o_{\text{f.s.}}(1) \\
&\quad + \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s') Z_1(x_{j_1}) \right) \left(\sum_{k_1=1}^p W_{k_1}(t') Z_1(x_{k_1}) \right) \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right) \left(\sum_{k_2=1}^p W_{k_2}(t) Z_1(x_{k_2}) \right) \right] + o_{\text{f.s.}}(1) \\
&= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \underbrace{\mathbb{E} [Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2}) Z_1(x_{k_2})]}_{=M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2})} \\
&\quad - 2 \cdot \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s') W_{k_2}(t') \underbrace{\mathbb{E} [Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2}) Z_1(x_{k_2})]}_{=M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2})} \\
&\quad + \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s') W_{k_1}(t') W_{j_2}(s) W_{k_2}(t) \underbrace{\mathbb{E} [Z_1(x_{j_1}) Z_1(x_{k_1}) Z_1(x_{j_2}) Z_1(x_{k_2})]}_{=M(x_{j_1}, x_{k_1}, x_{j_2}, x_{k_2})} + o_{\text{f.s.}}(1) \tag{6.22}
\end{aligned}$$

(*): Mit Proposition 6.3.

Ab hier ist das Vorgehen identisch zu dem aus dem Beweis von Satz 4.5, nur mit zusätzlichem $o_{f.s.}(1)$, was aber f. s. im Unendlichen verschwindet. Bedingung ii) ist demnach ebenfalls f. s. erfüllt.

Wir brauchen nun wieder eine Umhüllende. Mit Gleichung (6.11) folgt für alle $(s, t) \in [0, 1]^2$ schnell:

$$|Z_i^*(s)Z_i^*(t)| = |Z_i^*(s)||Z_i^*(t)| \leq \left((|Z_i(0)| + K_i)^*\right)^2 \quad (6.23)$$

und daraus wiederum:

$$\begin{aligned} |X_{n,i}^*(s, t)| &= \frac{1}{\sqrt{n}} \left| \sum_{j,k=1}^p W_j(s)W_k(t)Z_i^*(x_j)Z_i^*(x_k) \right| \leq \frac{1}{\sqrt{n}} \sum_{j,k=1}^p |W_j(s)||W_k(t)||Z_i^*(x_j)Z_i^*(x_k)| \\ &\leq \frac{1}{\sqrt{n}} \left((|Z_i(0)| + K_i)^*\right)^2 \sum_{j,k=1}^p |W_j(s)||W_k(t)| \leq \frac{C_{W_1}^2}{\sqrt{n}} \left((|Z_i(0)| + K_i)^*\right)^2 = \Phi_{n,i}^* \end{aligned}$$

$\Phi_n^* = (\Phi_{n,1}^*, \dots, \Phi_{n,n}^*)$ ist also eine Umhüllende für $X_{n,i}^*(s, t)$ und entspricht wieder der gebootstrapteten Version von Φ_n aus dem Beweis von Satz 4.5. Wegen der empirischen Verteilung von $(Z + K)^*$, siehe Gleichung (6.12) gilt für alle $k \in \mathbb{N}$ die folgende Implikation:

$$\begin{aligned} &\Rightarrow \left((|Z_1(0)| + K_1)^*\right)^2, \dots, \left((|Z_n(0)| + K_n)^*\right)^2 \mid Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{|Z_i(0)| + K_i} \\ &\Rightarrow \left(\Phi_{n,1}^*\right)^k, \dots, \left(\Phi_{n,n}^*\right)^k \mid Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{(\Phi_{n,i}^*)^k} \text{ mit } \Phi_{n,i}^* = \frac{C_{W_1}^2}{\sqrt{n}} \left(|Z_i(0)| + K_i\right)^2 \quad (6.24) \end{aligned}$$

und außerdem mit Gleichung (4.29), (A.2) und Proposition 3.3:

$$\mathbb{E}[|\Phi_{n,1}^*|^2] = \frac{C_{W_1}^4}{n} \mathbb{E} \left[\left(|Z_1(0)| + K_1\right)^4 \right] \leq \underbrace{\frac{8C_{W_1}^4}{n}}_{< \infty} \left(\underbrace{\mathbb{E}[(Z_1(0))^4]}_{=M(0,0,0,0) < \infty} + \underbrace{\mathbb{E}[(K_1)^4]}_{< \infty} \right) < \infty \quad (6.25)$$

Somit ist für das zweite Moment der Einträge von Φ_n^* für alle $n \in \mathbb{N}$ Proposition 6.3 anwendbar.

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}^*[\Phi_{n,i}^*] &= \sum_{i=1}^n \mathbb{E}[\Phi_{n,1}^*] + o_{f.s.}(1) = \frac{C_{W_1}^4}{n} \sum_{i=1}^n \mathbb{E} \left[\left(|Z_1(0)| + K_1\right)^4 \right] + o_{f.s.}(1) \\ &= C_{W_1}^4 \mathbb{E} \left[\left(|Z_1(0)| + K_1\right)^4 \right] + o_{f.s.}(1) \\ &\leq \underbrace{8C_{W_1}^4}_{< \infty} \left(\underbrace{\mathbb{E}[(Z_1(0))^4]}_{=M(0,0,0,0) < \infty} + \underbrace{\mathbb{E}[(K_1)^4]}_{< \infty} \right) + o_{f.s.}(1) \stackrel{f.s.}{<} \infty \end{aligned}$$

für alle $n \in \mathbb{N}$ und damit auch für den \limsup . Daraus folgt weiterhin für alle $\delta > 0$:

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E}^* [\Phi_{n,i}^2 \mathbf{1}_{\{\Phi_{n,i} > \delta\}}] &= \sum_{i=1}^n \mathbb{E} [\Phi_{n,1}^2 \mathbf{1}_{\{\Phi_{n,1} > \delta\}}] + o_{f.s.}(1) \\
 &= \frac{C_{W_1}^4}{n} \sum_{i=1}^n \mathbb{E} \left[\left(|Z_1(0)| + K_1 \right)^4 \mathbf{1}_{\left\{ \frac{C_{W_1}^2}{\sqrt{n}} \left(|Z_1(0)| + K_1 \right)^2 > \delta \right\}} \right] + o_{f.s.}(1) \\
 &= C_{W_1}^4 \mathbb{E} \left[\left(|Z_1(0)| + K_1 \right)^4 \underbrace{\mathbf{1}_{\left\{ C_{W_1}^2 \left(|Z_1(0)| + K_1 \right)^2 > \sqrt{n} \delta \right\}}}_{\rightarrow 0} \right] + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0}
 \end{aligned}$$

Also gelten auch die Bedingungen iii) und iv) f. s.

Für Bedingung v) ist zu zeigen, dass für alle $n \in \mathbb{N}$, $i = 1, \dots, n$ und $(s, t), (s', t') \in [0, 1]^2$ eine Konstante $C \in \mathbb{R}$ existiert, so dass f. s.:

$$\begin{aligned}
 &|s - s'| + |t - t'| \leq \delta^\gamma \\
 \Rightarrow &\left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s') W_k(t') Z_i^*(x_j) Z_i^*(x_k) \right| \leq C \delta \Phi_{n,i}^*
 \end{aligned}$$

Proposition 6.4 erlaubt die Umformung

$$\begin{aligned}
 &\left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i^*(x_j) Z_i^*(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s') W_k(t') Z_i^*(x_j) Z_i^*(x_k) \right| \\
 &= \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s) W_k(t) Z_i(x_j) Z_i(x_k) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_j(s') W_k(t') Z_i(x_j) Z_i(x_k) \right|^*
 \end{aligned}$$

also sind auch hier im Inneren alle Schritte äquivalent Beweis von Satz 4.5 durchführbar, welche schließlich auch zur geforderten Implikation führen.

Somit ist auch Bedingung v) f. s. erfüllt und nach dem funktionalen zentralen Grenzwertsatz gilt dadurch:

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) Z_i^*(x_k) - \mathbb{E}^* [Z_i^*(x_j) Z_i^*(x_k)] \right) \right) \mid Y \xrightarrow{w} \mathcal{G}(0, \mathcal{R})$$

f. s. in $(\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$, wobei \mathcal{R} durch Satz 4.5 gegeben ist.

$$S_n^{2*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*] \right) \right)$$

Wir kombinieren wieder f. s. die gleichmäßige gleichgrade Stetigkeit und punktweise Konvergenz in \mathbb{P} .

f. s. Punktweise Konvergenz in \mathbb{P} :

Wir haben es hier mit einer Doppelsumme eines Produkts zweier Zufallsvariablen zu tun, was sich noch durch ein Vektor-(Zufalls)Matrix-Vektor-Produkt ausdrücken ließe. Allerdings haben wir für beide Darstellungen keine Aussagen bzgl. der Mallows' Distanz, da diese maximal mit gewöhnlichen Vektoren von Zufallsvariablen arbeitet, nicht jedoch mit Produkten oder Matrizen. Möglicherweise gelten die Aussagen auch dafür, was aber definitiv weiterer Nachforschung bedürfte, also beschränken wir uns dieses mal auf den Beweis mittels der Konvergenz im 2-ten Mittel. Durch die bedingte unabhängig identische Verteilung von $\varepsilon^* \cdot \varepsilon^*$ bzgl. der Stichprobe i und Lemma 2.25, entsprechen die Rechenschritte hier exakt denen aus Gleichung (4.32). Für alle $(s, t) \in [0, 1]^2$ gilt also:

$$\begin{aligned} \mathbb{E}^* [(S_n^{2*}(s, t))^2] &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right)^2 \right] \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \underbrace{(\mathbb{E}^* [\varepsilon_{1j_1}^* \varepsilon_{1k_1}^* \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*] - \mathbb{E}^* [\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*] \mathbb{E}^* [\varepsilon_{1j_2}^* \varepsilon_{1k_2}^*])}_{= \text{Cov}^*(\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*, \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*)} \end{aligned} \quad (6.26)$$

$\text{Cov}^*(\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*, \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*)$ entspricht nun gerade wieder der bedingten, über gebootstrapte Werte von ε , geschätzten Kovarianz von $\varepsilon \cdot \varepsilon$ an den Stellen (j_1, k_1) und (j_2, k_2) , also letztlich f. s. der Stichprobenkovarianz:

$$\text{Cov}^*(\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*, \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*) = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} - \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_1} \varepsilon_{ik_1} \right) \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_{ij_2} \varepsilon_{ik_2} \right)$$

Auch hier gilt für alle $j_1, k_1, j_2, k_2 = 1, \dots, p$ die Konsistenz $\text{Cov}^*(\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*, \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*) = \text{Cov}(\varepsilon_{1j_1} \varepsilon_{1k_1}, \varepsilon_{1j_2} \varepsilon_{1k_2}) + \hat{a}_n$ mit $\hat{a}_n \xrightarrow{\text{f.s.}} 0$, oder in Matrixform

$$\text{Cov}^*(\vec{\varepsilon}^* \otimes \vec{\varepsilon}^*) = \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) + \hat{A}_n \in \mathbb{R}^{p^2 \times p^2} \text{ mit } \hat{A}_n \xrightarrow{\text{f.s.}} 0_{p^2}$$

Mit der gleichen Argumentation wie bei der Herleitung von Gleichung (6.15) existiert nun wieder ein $n_0 \in \mathbb{N}$, so dass für den maximalen Eigenwert $\lambda_{\max,2}^*$ von $\text{Cov}^*(\vec{\varepsilon}_1^* \otimes \vec{\varepsilon}_1^*)$ für alle $n \geq n_0$:

$$\lambda_{\max,2}^* = \|\text{Cov}^*(\vec{\varepsilon}_1^* \otimes \vec{\varepsilon}_1^*)\|_2 = \|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) + \hat{A}_n\|_2 \leq \|\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)\|_2 + \|\hat{A}_n\|_2 \stackrel{\text{f.s.}}{\leq} \lambda_{\max,2} + \mu_{n_0}$$

mit $\lambda_{\max,2}$ als beschränkten maximalen Eigenwert von $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1)$ und μ_{n_0} als betragsmäßig größten, ebenfalls beschränkten Eigenwert von \hat{A}_{n_0} . Nach (A.4) gilt $\lambda_{\max,2} = \mathcal{O}(p^{d_2})$ und da $p \rightarrow \infty$, $d_2 \in [0, 2)$ und $\mu_0 < \infty$, folgt wieder $\lambda_{\max,2}^* = \mathcal{O}(p^{d_2})$, wodurch auch wieder Proposition 3.9 für $\text{Cov}^*(\vec{\varepsilon}_1^* \otimes \vec{\varepsilon}_1^*)$ anwendbar ist. Es folgt schließlich für Gleichung (6.26):

$$\begin{aligned} \mathbb{E}^*[(S_n^{2*}(s, t))^2] &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s)W_{k_1}(t)W_{j_2}(s)W_{k_2}(t) \text{Cov}^*(\varepsilon_{1j_1}^* \varepsilon_{1k_1}^*, \varepsilon_{1j_2}^* \varepsilon_{1k_2}^*) \\ &= (\vec{W}(s) \otimes \vec{W}(t))^T \cdot \text{Cov}(\vec{\varepsilon}_1^* \otimes \vec{\varepsilon}_1^*) \cdot (\vec{W}(s) \otimes \vec{W}(t))^T = \mathcal{O}_{\text{f.s.}}\left(\frac{1}{p^{2-d_2}h^2}\right) \xrightarrow{\text{f.s.}} 0 \end{aligned} \quad (6.27)$$

$S_n^{2*}(s, t)$ konvergiert somit für alle $(s, t) \in [0, 1]^2$ bedingt auf Y f. s. gegen 0 in \mathbb{P} .

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

$$\text{Sei } \Phi^*(s, t) = \|S_n^{2*}(s, t)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^*[\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right)^2}.$$

Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^{2*}(s, t)| \leq \Phi^*(s, t) \Leftrightarrow |S_n^{2*}| \leq \Phi^*$. Mit Gleichung (6.26) und (6.27) folgt weiterhin:

$$\begin{aligned} \mathbb{E}^*[|\Phi^*(s, t)|^2] &= \mathbb{E}^*[\|S_n^{2*}(s, t)\|_{\ell^2}^2] = \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^*[\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^*[\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right)^2 \right] = \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2}h^2} \right) \\ &\Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^*[|\Phi^*|^2] \stackrel{\text{f.s.}}{\leq} C \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} < \infty \Rightarrow (\mathbb{E}^*[|\Phi^*|^2])^{\frac{1}{2}} \stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

zu ii):

Wir schätzen die Überdeckungszahl mit der kanonischen Metrik ab. Auf gleichem Rechenweg und den Konventionen bzw. Argumentationen wie in Gleichung (6.26) und (6.27), folgt mit Proposition 3.9 für alle $(s, t), (s', t') \in [0, 1]^2$

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right. \right. \\
 &\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s') W_k(t') (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right) \right]^2 \\
 &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) (\varepsilon_{ij}^* \varepsilon_{ik}^* - \mathbb{E}^* [\varepsilon_{ij}^* \varepsilon_{ik}^*]) \right) \right) \right]^2 \\
 &= (\vec{W}(s) \otimes \vec{W}(t) - \vec{W}(s') \otimes \vec{W}(t'))^T \cdot \text{Cov}(\varepsilon_1^* \otimes \varepsilon_1^*) \cdot (\vec{W}(s) \otimes \vec{W}(t) - \vec{W}(s') \otimes \vec{W}(t')) \\
 &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} \right)
 \end{aligned}$$

$$\Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \frac{C}{p^{2-d_2} h^2} \Rightarrow d_n^*((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \left(\frac{C}{p^{2-d_2} h^2} \right)^{\frac{1}{2}} =: C_n$$

Ab hier sind alle weiteren Schritte identisch zu denen von Gleichung (4.33) und folgende, also ist $S_n^{2*}(s, t)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} .

$$S_n^{3*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^* [Z_i^*(x_j)] \mathbb{E}^* [\varepsilon_{ik}^*] \right) \right)$$

Wir kombinieren wieder f. s. die gleichmäßige gleichgrade Stetigkeit und punktweise Konvergenz in \mathbb{P} .

f. s. Punktweise Konvergenz in \mathbb{P} :

Wegen den bereits erwähnten potentiellen Problemen der Mallows' Distanz bei Produkten von Zufallsvariablen, verwenden wir wieder die Konvergenz im 2-ten Mittel. Für alle $(s, t) \in [0, 1]^2$ erhalten wir mit der Unabhängigkeit von Z^* und ε^* und Lemma 2.25 über die gleichen Rechenschritte bzw. der Argumentation bei Gleichung (6.26):

$$\begin{aligned} \mathbb{E}^* [(S_n^{3*}(s, t))^2] &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^* [Z_i^*(x_j)] \mathbb{E}^* [\varepsilon_{ik}^*] \right) \right) \right)^2 \right] \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \\ &\quad \cdot \left(\mathbb{E}^* [Z_1^*(x_{j_1}) Z_1^*(x_{j_2})] \mathbb{E}^* [\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] - \mathbb{E}^* [Z_1^*(x_{j_1})] \mathbb{E}^* [Z_1^*(x_{j_2})] \mathbb{E}^* [\varepsilon_{1k_1}^*] \mathbb{E}^* [\varepsilon_{1k_2}^*] \right) \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{j_1}(s) W_{k_1}(t) W_{j_2}(s) W_{k_2}(t) \left(\mathbb{E}^* [\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \underbrace{\left(\mathbb{E}^* [Z_1^*(x_{j_1}) Z_1^*(x_{j_2})] - \mathbb{E}^* [Z_1^*(x_{j_1})] \mathbb{E}^* [Z_1^*(x_{j_2})] \right)}_{= \text{Cov}^* (Z_1^*(x_{j_1}), Z_1^*(x_{j_2}))} \right. \\ &\quad \left. + \mathbb{E}^* [Z_1^*(x_{j_1})] \cdot \mathbb{E}^* [Z_1^*(x_{j_2})] \cdot \underbrace{\left(\mathbb{E}^* [\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] - \mathbb{E}^* [\varepsilon_{1k_1}^*] \mathbb{E}^* [\varepsilon_{1k_2}^*] \right)}_{= \sigma_{k_1 k_2}^*} \right) \\ &= \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \mathbb{E}^* [\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \right) \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right) \\ &\quad + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right] \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \sigma_{k_1 k_2}^* \right) \end{aligned} \tag{6.28}$$

Wir betrachten jeden dieser Faktoren einzeln, angefangen beim Letzten.

Nach den Argumentationen bei Gleichung (6.15) gilt:

$$\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \sigma_{k_1 k_2}^* = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right)$$

Nach Proposition 6.3 gilt:

$$\begin{aligned}
 & \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right] \\
 &= \left(\mathbb{E} \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right) \right] + o_{\text{f.s.}}(1) \right) \left(\mathbb{E} \left[\left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right) \right] + o_{\text{f.s.}}(1) \right) \\
 &= \left(\sum_{j_1=1}^p W_{j_1}(s) \underbrace{\mathbb{E}[Z_1(x_{j_1})]}_{=0} + o_{\text{f.s.}}(1) \right) \left(\sum_{j_2=1}^p W_{j_2}(s) \underbrace{\mathbb{E}[Z_1(x_{j_2})]}_{=0} + o_{\text{f.s.}}(1) \right) = o_{\text{f.s.}}(1)
 \end{aligned}$$

Wegen der starken Konsistenz der Stichproben Kovarianz gilt nach Proposition 3.7 und 6.3, legitimiert durch Proposition 3.3 und 3.6:

$$\begin{aligned}
 & \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right) \\
 &= \text{Cov} \left(\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right), \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right) \right) + o_{\text{f.s.}}(1) \\
 &= \sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s) \text{Cov} (Z_1(x_{j_1}), Z_1(x_{j_2})) + o_{\text{f.s.}}(1) \\
 &= \sum_{j_1, j_2=1}^p W_{j_1}(s) W_{j_2}(s) R(x_{j_1}, x_{j_2}) + o_{\text{f.s.}}(1) = R(s, s) + \mathcal{O}(h^{1+\alpha_R}) + o_{\text{f.s.}}(1)
 \end{aligned}$$

Für den verbleibenden Faktor muss ein wenig ausgeholt werden. Nach Proposition 6.3, wieder legitimiert durch Proposition 3.3, gilt $\mathbb{E}^* [\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] = \mathbb{E} [\varepsilon_{1k_1} \varepsilon_{1k_2}] + o_{\text{f.s.}}(1)$ für alle $k_1, k_2 = 1, \dots, p$. Schreibt man dieses bedingte zweite Verbundmoment wieder als Matrix bzgl. der Indizes k_1 und k_2 , so erhält man:

$$\mathbb{E}^* [\vec{\varepsilon}_1^* \cdot \vec{\varepsilon}_1^{*T}] := \mathbb{E} [\vec{\varepsilon}_1 \cdot \vec{\varepsilon}_1^T] + \hat{A}_n = \underbrace{\text{Cov}(\vec{\varepsilon}_1)}_{=\vec{0}} + \underbrace{\mathbb{E}[\vec{\varepsilon}_1] \cdot \mathbb{E}[\vec{\varepsilon}_1]^T}_{=\vec{0}} + \hat{A}_n = \text{Cov}(\vec{\varepsilon}_1) + \hat{A}_n \in \mathbb{R}^{p \times p}$$

mit $\hat{A}_n \xrightarrow{\text{f.s.}} 0_p$. Durch die Symmetrie von $\text{Cov}(\vec{\varepsilon}_1)$ und $\mathbb{E}^* [\vec{\varepsilon}_1^* \cdot \vec{\varepsilon}_1^{*T}]$ (denn das Verbundmoment ist symmetrisch), ist \hat{A}_n auch wieder symmetrisch für alle $n \in \mathbb{N}$. Durch die Nullfolgen-Eigenschaft, gibt es wieder ein $n_0 \in \mathbb{N}$, für welches für alle $n \geq n_0$ f. s. $\mu_n = \|\hat{A}_n\|_2 \leq \|\hat{A}_{n_0}\|_2 = \mu_{n_0} < \infty$.

Es gilt also auch hier für den maximalen Eigenwert λ_{\max}^* von $\mathbb{E}^*[\vec{\varepsilon}_1^* \cdot \vec{\varepsilon}_1^{*T}]$ als f. s. symmetrischer, positiv semidefiniter Matrix für alle $n \geq n_0$:

$$\lambda_{\max}^* = \|\mathbb{E}^*[\vec{\varepsilon}_1^* \cdot \vec{\varepsilon}_1^{*T}]\|_2 = \|\text{Cov}(\vec{\varepsilon}_1) + \hat{A}_n\|_2 \leq \|\text{Cov}(\vec{\varepsilon}_1)\|_2 + \|\hat{A}_n\|_2 \stackrel{\text{f.s.}}{\leq} \lambda_{\max,1} + \mu_{n_0}$$

und somit ist erneut Proposition 3.9 anwendbar:

$$\sum_{k_1, k_2=1}^p W_{k_1}(t)W_{k_2}(t)\mathbb{E}^*[\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] = \vec{W}^T(t) \cdot \mathbb{E}^*[\vec{\varepsilon}_1^* \cdot \vec{\varepsilon}_1^{*T}] \cdot \vec{W}^T(t) = \mathcal{O}_{\text{f.s.}}\left(\frac{1}{p^{1-d_1}h}\right)$$

Insgesamt folgt damit für Gleichung (6.28) durch die Nullfolgeneigenschaft von $\frac{1}{p^{1-d_1}h}$ und $h \rightarrow 0$:

$$\begin{aligned} \mathbb{E}^*[(S_n^{3*}(s, t))^2] &= \mathcal{O}_{\text{f.s.}}\left(\frac{1}{p^{1-d_1}h}\right) \left(R(s, s) + \mathcal{O}(h^{1+\alpha_R}) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{= \mathcal{O}_{\text{f.s.}}(1)}\right) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{= \mathcal{O}_{\text{f.s.}}(1)} \cdot \mathcal{O}_{\text{f.s.}}\left(\frac{1}{p^{1-d_1}h}\right) \\ &\stackrel{\substack{|R(\cdot, \cdot)| \leq E_{Z,2} \\ \text{Proposition 3.3}}}{\mathcal{O}_{\text{f.s.}}\left(\frac{1+h^{1+\alpha_R}}{p^{1-d_1}h}\right)} \stackrel{\text{f.s.}}{\rightarrow} 0 \end{aligned} \quad (6.29)$$

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei

$$\Phi^*(s, t) = \|S_n^{3*}(s, t)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) \left(Z_i^*(x_j)\varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)]\mathbb{E}^*[\varepsilon_{ik}^*] \right) \right) \right)^2}$$

Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^{3*}(s, t)| \leq \Phi^*(s, t) \Leftrightarrow |S_n^{3*}| \leq \Phi^*$. Mit Gleichung (6.28) und (6.29) folgt weiterhin:

$$\begin{aligned} \mathbb{E}^*[|\Phi^*(s, t)|^2] &= \mathbb{E}^*[\|S_n^{3*}(s, t)\|_{\ell^2}^2] \\ &= \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) \left(Z_i^*(x_j)\varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)]\mathbb{E}^*[\varepsilon_{ik}^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s)W_k(t) \left(Z_i^*(x_j)\varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)]\mathbb{E}^*[\varepsilon_{ik}^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}}\left(\frac{1+h^{1+\alpha_R}}{p^{1-d_1}h}\right) \\ &\Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^*[|\Phi^*|^2] \stackrel{\text{f.s.}}{\leq} C \sum_{n=1}^{\infty} \frac{1+h^{1+\alpha_R}}{p^{1-d_1}h} \stackrel{h < 1}{<} 2C \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty \Rightarrow (\mathbb{E}^*[|\Phi^*|^2])^{\frac{1}{2}} \stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

zu ii):

Wir schätzen die Überdeckungszahl mit der kanonischen Metrik ab. Auf gleichem Rechenweg und den Konventionen wie in Gleichung (6.28) folgt mit Proposition 6.3 für alle $(s, t), (s', t') \in [0, 1]^2$:

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) (Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)] \mathbb{E}^*[\varepsilon_{ik}^*]) \right) \right. \right. \\
 &\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s') W_k(t') (Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)] \mathbb{E}^*[\varepsilon_{ik}^*]) \right) \right) \right]^2 \\
 &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p (W_j(s) W_k(t) - W_j(s') W_k(t')) (Z_i^*(x_j) \varepsilon_{ik}^* - \mathbb{E}^*[Z_i^*(x_j)] \mathbb{E}^*[\varepsilon_{ik}^*]) \right) \right) \right]^2 \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p (W_{j_1}(s) W_{k_1}(t) - W_{j_1}(s') W_{k_1}(t')) (W_{j_2}(s) W_{k_2}(t) - W_{j_2}(s') W_{k_2}(t')) \\
 &\quad \cdot (\mathbb{E}^*[\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \text{Cov}^*(Z_1^*(x_{j_1}), Z_1^*(x_{j_2})) + \mathbb{E}^*[Z_1^*(x_{j_1})] \mathbb{E}^*[Z_1^*(x_{j_2})] \sigma_{k_1 k_2}^*) \\
 &\stackrel{(*)}{=} \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \mathbb{E}^*[\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \right) \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right) \\
 &\quad + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s) Z_1(x_{j_2}) \right)^* \right] \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t) \sigma_{k_1 k_2}^* \right) \\
 &\quad - 2 \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t') \mathbb{E}^*[\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \right) \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \right) \\
 &\quad - 2 \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s) Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \right] \left(\sum_{k_1, k_2=1}^p W_{k_1}(t) W_{k_2}(t') \sigma_{k_1 k_2}^* \right) \\
 &\quad + \left(\sum_{k_1, k_2=1}^p W_{k_1}(t') W_{k_2}(t') \mathbb{E}^*[\varepsilon_{1k_1}^* \varepsilon_{1k_2}^*] \right) \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{j_1}(s') Z_1(x_{j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \right) \\
 &\quad + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{j_1}(s') Z_1(x_{j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{j_2}(s') Z_1(x_{j_2}) \right)^* \right] \left(\sum_{k_1, k_2=1}^p W_{k_1}(t') W_{k_2}(t') \sigma_{k_1 k_2}^* \right) \\
 &\stackrel{|\cdot|}{\leq} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \left(R(s, s) + \mathcal{O}(h^{1+\alpha_R}) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \right) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \cdot \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \\
 &\quad + 2 \cdot \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \left(R(s, s') + \mathcal{O}(h^{1+\alpha_R}) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \right) + 2 \cdot \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \cdot \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \\
 &\quad + \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \left(R(s', s') + \mathcal{O}(h^{1+\alpha_R}) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \right) + \underbrace{\mathcal{O}_{\text{f.s.}}(1)}_{=\mathcal{O}_{\text{f.s.}}(1)} \cdot \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right) \\
 &\stackrel{|\text{R}(\cdot, \cdot)| \leq E_{Z,2}}{\leq} \underbrace{\mathcal{O}_{\text{f.s.}}}_{\text{Proposition 3.3}} \left(\frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \right)
 \end{aligned}$$

$$\Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \frac{C(1 + h^{1+\alpha_R})}{p^{1-d_1}h} \Rightarrow d_n^*((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \left(\frac{C(1 + h^{1+\alpha_R})}{p^{1-d_1}h} \right)^{\frac{1}{2}} =: C_n$$

Nun sind wir wieder bei der Aussage von Gleichung (4.38) angelangt, nur dass die Aussagen alle nur f. s. gelten, alle restlichen Argumentationen sind identisch. Also ist $S_n^{3*}(s, t)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} .

$$S_n^{4*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \left(Z_i^*(x_k) \varepsilon_{ij}^* - \mathbb{E}^* [Z_i^*(x_k)] \mathbb{E}^* [\varepsilon_{ij}^*] \right) \right)$$

Dieser Summand lässt sich vollkommen analog zu $S_n^{3*}(s, t)$ diskutieren. □

7. Erweiterung auf asynchrone Messpunkte

In Kapitel 3, 4, 5 und 6 wurde ausschließlich der Fall synchroner Messpunkte behandelt, dass die zu Y_{ij} gehörigen x_j für alle $i = 1, \dots, n$ gleich sind. Das kann man zu einer allgemeineren Situation erweitern, indem für jedes Y_{ij} ein separater Messpunkt x_{ij} herangezogen wird, für welche - im Gegensatz zum synchronen Fall - für mindestens ein $j = 1, \dots, p$ und $i_1 \neq i_2$ gilt, dass $x_{i_1 j} \neq x_{i_2 j}$. Daraus resultieren $p \cdot n$ Messpaare (Y_{ij}, x_{ij}) mit $1 \leq i \leq n$ und $1 \leq j \leq p$, wobei wieder $x_{ij} := x_{ij}(n, p)$ und o. B. d. A. $x_{ij} \in [0, 1]$ gelten soll. Für die Modellgleichung gilt damit:

$$Y_{ij} = \mu(x_{ij}) + Z_i(x_{ij}) + \varepsilon_{ij} \quad (7.1)$$

μ , Z und ε sind die gleichen mathematischen Elemente wie in Kapitel 3. Da jeder Messpunkt quasi „seine eigene“ Gewichtsfunktion braucht, verändern sich die Schätzer aus Abschnitt 3.2 durch diese neue Festlegung zu:

$$\begin{aligned} \tilde{\mu}_i(x) &= \sum_{j=1}^p W_{ij}(x) Y_{ij} & (7.2) \\ \Rightarrow \tilde{\mu}(x) &= \frac{1}{n} \sum_{i=1}^n \tilde{\mu}_i(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) Y_{ij} \\ \tilde{R}(s, t) &= \frac{1}{n} \sum_{i=1}^n (\tilde{\mu}_i(s) - \tilde{\mu}(s)) (\tilde{\mu}_i(t) - \tilde{\mu}(t)) = \frac{1}{n} \sum_{i=1}^n \tilde{\mu}_i(s) \tilde{\mu}_i(t) - \tilde{\mu}(s) \tilde{\mu}(t) \end{aligned}$$

bzw. als unverzerrte Version

$$\tilde{R}(s, t) = \frac{1}{n-1} \sum_{i=1}^n (\tilde{\mu}_i(s) - \tilde{\mu}(s)) (\tilde{\mu}_i(t) - \tilde{\mu}(t))$$

Eine Umformung von $\tilde{\mu}$ im Sinne von Gleichung (3.3) ist somit nicht mehr möglich.

Der wesentliche Unterschied von $\hat{\mu}$ und $\tilde{\mu}$ ist, dass bei Ersterem zuerst die Daten Y_{ij} gemittelt werden (können) und dann geglättet wird, wohingegen bei Letzterem unabdingbar zuerst geglättet werden muss, und danach gemittelt wird, gerade weil die Gewichte nun auch von i abhängen. Für den Spezialfall $x_{1j} = x_{2j} = \dots = x_{nj}$, für alle $j = 1, \dots, p$, sind die Schätzer $\hat{\mu}$ und $\tilde{\mu}$ identisch.

Wir müssen nun die geforderten Annahmen etwas verändern bzw. erweitern, damit die bereits eruierten Eigenschaften und Rechenmethoden der Schätzer erhalten bleiben.

7.1. Annahmen (asynchrones Modell)

7.1.1. Annahmen an das asynchrone Modell

Hier bleiben die Annahmen größtenteils gleich, bis auf die Erzeugung der Messpunkte x_{ij} und eine Forderung der Existenz höherer Momente von Z und ε .

Annahme 7.1.

($\tilde{A}.1$) Identisch zu (A.1)

($\tilde{A}.2$) Identisch zu (A.2)

($\tilde{A}.3$) Identisch zu (A.3). Zudem soll das achte Verbundmoment \mathcal{M} von Z mit der Abbildungsvorschrift

$$\mathcal{M} : \begin{cases} [0, 1]^8 \rightarrow \mathbb{R} \\ (s, t, s', t', s'', t'', s''', t''') \mapsto \mathbb{E}[Z(s)Z(t)Z(s')Z(t')Z(s'')Z(t'')Z(s''')Z(t''')] \end{cases}$$

existieren und hölderstetige partielle Ableitungen zum Exponenten $\alpha_{\mathcal{M}} \in (0, 1]$ mit Konstanten $H_{\mathcal{M}} \in \mathbb{R}^+$ haben, schreibe $\mathcal{M} \in \mathcal{C}^{1, \alpha_{\mathcal{M}}}([0, 1]^8)$.

($\tilde{A}.4$) Identisch zu (A.4). Sei zudem $\lambda_{\max, 4}$ der maximale Eigenwert der Kovarianzmatrix $\text{Cov}(\vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1 \otimes \vec{\varepsilon}_1) \in \mathbb{R}^{p^4 \times p^4}$. Dann soll $\lambda_{\max, 4} = \mathcal{O}(p^{d_4})$, $d_4 \in [0, 4)$ gelten, d. h. es soll eine endliche Konstante $C_4 \in \mathbb{R}^+$ existieren, so dass $\lambda_{\max, 4} \leq C_4 p^{d_4}$. Die Beschränkung soll ebenfalls gleichmäßig in n und p gelten.

($\tilde{A}.5$) Für alle $i = 1, \dots, n$ werden die Punkte x_{ij} , $j = 1, \dots, p$ durch die Vorschrift $\int_0^{x_{ij}} f_i(t) dt = \frac{j - 0.5}{p}$ erzeugt, wobei $f_i : [0, 1] \rightarrow \mathbb{R}^+$ lipschitzstetige positive Lebesgue-Dichten sind. Zudem existieren Konstanten $c_f, C_f \in \mathbb{R}^+$ sodass für alle $i = 1, \dots, n$ und $x \in [0, 1]$ die Bedingung $c_f \leq f_i(x) \leq C_f$ gilt.

Wir brauchen diese neue Annahme hauptsächlich deswegen, weil die, durch die asynchronen Messpunkte, bzgl. i ebenfalls asynchronen Gewichtsfunktionen nun oftmals die identische Verteilung zerstören, wodurch einige Aussagen anders als im synchronen Fall bewiesen werden müssen. Es ist nun eine Erweiterung von Proposition 3.3 möglich.

Proposition 7.2.

Sei Annahme 7.1 erfüllt und $m \in \{1, \dots, 8\}$. Dann existieren für alle $x_1, \dots, x_m \in [0, 1]$ und alle $j_1, \dots, j_m = 1, \dots, p$, $p \in \mathbb{N}$ Konstanten $E_{Z,1}, \dots, E_{Z,m}, E_{\varepsilon,1}(p), \dots, E_{\varepsilon,m}(p) \in \mathbb{R}^+$, für welche folgende Ungleichungen gelten:

$$\mathbb{E} \left[\left| \prod_{k=1}^m Z(x_k) \right| \right] \leq E_{Z,m}, \quad \mathbb{E} \left[\left| \prod_{k=1}^m \varepsilon_{1j_k} \right| \right] \leq E_{\varepsilon,m}(p)$$

Durch die Jensensche Ungleichung gelten diese Schranken automatisch auch, falls der Betrag außerhalb des Erwartungswertes steht.

Beweis.

Äquivalent zum Beweis von Proposition 3.3, nur in höherer Dimension, legitimiert durch (Ä.3) und (Ä.4). \square

7.1.2. Annahmen an die asynchronen Schätzer

Da wir es nun auch mit asynchronen Gewichten bzgl. i zu tun haben, müssen die Annahmen daran vollständig angepasst werden. Sie sollen nun für alle $x, x' \in [0, 1]$, $i = 1, \dots, n$ und $j = 1, \dots, p$, $n, p \in \mathbb{N}$ gelten.

Annahme 7.3.

$$(\tilde{A}.6) \quad |W_{ij}(x)| \leq \frac{C'_W}{ph} \text{ mit } C'_W \in \mathbb{R}^+.$$

$$(\tilde{A}.7) \quad |W_{ij}(x) - W_{ij}(x')| \leq \frac{C'_{W_-} |x - x'|}{ph^2} \text{ mit } C'_{W_-} \in \mathbb{R}^+.$$

$$(\tilde{A}.8) \quad \text{Falls } |x - x_{ij}| > h, \text{ dann soll } W_{ij}(x) = 0 \text{ gelten.}$$

$$(\tilde{A}.9) \quad \sum_{j=1}^p W_{ij}(x) = 1 \text{ und } \sum_{j=1}^p W_{ij}(x)(x - x_{ij}) = 0.$$

Unter diesen Voraussetzungen folgt wieder eine äquivalente Proposition zu Proposition 3.6.

Proposition 7.4.

Für die Gewichte und Messpunkte seien Annahme 7.1 und 7.3 erfüllt. Dann gilt:

a) Für alle $i = 1, \dots, n$ und $x \in [0, 1]$ existiert eine Konstante $C'_{W_0} \in \mathbb{R}^+$, sodass die maximale Anzahl der von 0 verschiedenen Gewichte $W_{i1}(x), \dots, W_{ip}(x)$ ab einem hinreichend großen p höchstens $C'_{W_0} ph$ beträgt.

b) Für alle $i = 1, \dots, n$, $x \in [0, 1]$ und $r \in \mathbb{R}^+$ gilt:

$$\sum_{j=1}^p |W_{ij}(x)|^r \leq \frac{C'_{W_r}}{(ph)^{r-1}}, \quad C'_{W_r} < \infty$$

Beweis.

a) Siehe [Her17] Seite 32 und 76/77.

b) Aus Teil a) und (Ã.6) folgt:

$$\sum_{j=1}^p |W_{ij}(x)|^r \leq C'_{W_0} p h \cdot \left(\frac{C'_W}{ph} \right)^r = \overbrace{\frac{C'_{W_0} C'^r_W}{(ph)^{r-1}}}^{=: C'_{W_r}}, \text{ mit } C'_{W_0}, C'_W \in \mathbb{R}^+ \Rightarrow C'_{W_r} \in \mathbb{R}^+$$

□

Nun lassen sich auch Proposition 3.7, 3.9 und 6.3 wie folgt neu formulieren, in deren Beweisen wir nur die marginal anderen Stellen anpassen.

Proposition 7.5.

Sei $f \in \mathcal{C}^{1,\alpha}(\mathbb{R}^D)$, mit $D \in \mathbb{N}$ und $\alpha \in (0, 1]$, eine deterministische Funktion und die Gewichte des Schätzers erfüllen Annahme 7.3. Dann gilt für alle $i = 1, \dots, n$ und $(x_{ij_1}, \dots, x_{ij_D}), (x_1, \dots, x_D) \in [0, 1]^D$:

$$\begin{aligned} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(x_{ij_1}, \dots, x_{ij_D}) &= \sum_{j_1, \dots, j_D=1}^p W_{ij_1}(x_1) \cdot \dots \cdot W_{ij_D}(x_D) f(x_{ij_1}, \dots, x_{ij_D}) \\ &= f(x_1, \dots, x_D) + \mathcal{O}(h^{1+\alpha}) \end{aligned}$$

Beweis.

Hier ist die Betragsmäßige Differenz

$$\left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}_{ij}) - f(\vec{x}) \right| \quad (7.3)$$

mit $\vec{x}_{ij} := (x_{ij_1}, \dots, x_{ij_D})$ und $\vec{x} := (x_1, \dots, x_D)$ zu untersuchen. Auf exakt gleiche Weise wie im vorherigen Beweis gilt für alle $i = 1, \dots, n$ die Gleichung:

$$\begin{aligned} \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}_{ij}) &= \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}) \\ &+ \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) \left(\sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} (x_{ij_l} - x_l) \right) \\ &+ \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) \left(\sum_{l=1}^D \left(\frac{\partial f(\vec{x} + \tau_{ij}(\vec{x}_{ij} - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right) (x_{ij_l} - x_l) \right) \end{aligned}$$

$f(\vec{x})$ beinhaltet nach wie vor keinen Index, der für die Summe bzw. das Produkt bzgl. der Gewichte relevant wäre. Daher folgt für den ersten Summanden mit (Ä.9) für alle $i = 1, \dots, n$:

$$\sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}) = f(\vec{x}) \cdot \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) = f(\vec{x}) \cdot \prod_{k=1}^D \underbrace{\left(\sum_{j=1}^p W_{ij}(x_k) \right)}_{=1} = f(\vec{x})$$

Der zweite Summand lässt sich wieder umformen zu:

$$\sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) \left(\sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} (x_{ij_l} - x_l) \right) = \sum_{l=1}^D \frac{\partial f(\vec{x})}{\partial x_l} \left(\left(\prod_{k=1}^D \left(\sum_{j=1}^p W_{ij}(x_k) \right) \right) (x_{ij_l} - x_l) \right)$$

Für den Fall $l = k$ gilt für den jeweiligen Teilsummanden nach (Ä.9) für alle $i = 1, \dots, n$:

$$\frac{\partial f(\vec{x})}{\partial x_k} \left(\sum_{j=1}^p W_{ij}(x_k) (x_{ij_k} - x_k) \right) = 0$$

also ist auch hier dieser komplette Summand 0. Das Mittel über n spielt daher für den ersten und den zweiten Summanden keine Rolle mehr und es folgt für Gleichung (7.3) mit der Δ -Ungleichung:

$$\left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}_{ij}) - f(\vec{x}) \right| \leq \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{ij_k}(x_k)| \right) \cdot \left(\sum_{l=1}^D \left| \frac{\partial f(\vec{x} + \tau_{ij}(\vec{x}_{ij} - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right| |x_{ij_l} - x_l| \right) \quad (7.4)$$

Da $f \in C^{1,\alpha}(\mathbb{R}^D)$, gilt wieder für alle $l = 1, \dots, D$ und $i = 1, \dots, n$:

$$\left| \frac{\partial f(\vec{x} + \tau_{ij}(\vec{x}_{ij} - \vec{x}))}{\partial x_l} - \frac{\partial f(\vec{x})}{\partial x_l} \right| \leq H_f \|\vec{x}_{ij} - \vec{x}\|^\alpha$$

mit H_f als Hölderkonstante von f . Daraus folgt über Gleichung (7.4):

$$\left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}_{ij}) - f(\vec{x}) \right| \leq \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{ij_k}(x_k)| \right) \cdot \left(\sum_{l=1}^D H_f \left(\sum_{k=1}^D |x_{ij_k} - x_k| \right)^\alpha |x_{ij_l} - x_l| \right)$$

Nach (Ä.8) gilt für alle $i = 1, \dots, n$, dass $|W_{ij_k}(x_k)| = 0 \Rightarrow \prod_{k=1}^D |W_{ij_k}(x_k)| = 0$, falls $|x_{ij_k} - x_j| > h$. Dass heißt, $|x_{ij_k} - x_j|$ muss für alle $k = 1, \dots, D$ kleiner oder gleich h sein. Damit folgt schlussendlich mit Proposition 7.4 annahmen gewichte über obige Gleichung:

$$\begin{aligned} \left| \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D W_{ij_k}(x_k) \right) f(\vec{x}_{ij}) - f(\vec{x}) \right| &\leq \sum_{j_1, \dots, j_D=1}^p \left(\prod_{k=1}^D |W_{ij_k}(x_k)| \right) \left(\sum_{l=1}^D H_f \left(\sum_{k=1}^D h \right)^\alpha h \right) \\ &= D^{1+\alpha} H_f h^{1+\alpha} \left(\prod_{k=1}^D \left(\sum_{j=1}^p |W_{ij}(x_k)| \right) \right) \\ &\leq D^{1+\alpha} H_f h^{1+\alpha} \left(\prod_{k=1}^D C'_{W_1} \right) = D^{1+\alpha} H_f C'^D_{W_1} h^{1+\alpha} \\ &= \mathcal{O}(h^{1+\alpha}) \end{aligned}$$

□

Proposition 7.6.

Sei $A \in \mathbb{R}^{p^D \times p^D}$, mit $D \in \mathbb{N}$, eine symmetrische, positiv semidefinite Matrix, für die, für $p \rightarrow \infty$, die folgende Eigenschaft gilt: Der maximale Eigenwert ist von der Ordnung $\mathcal{O}(p^d)$ mit $d \in \mathbb{R}$. Anders formuliert: Es existiert eine endliche Konstante $C \in \mathbb{R}^+$ und ein $p_0 \in \mathbb{N}$, so dass für den maximalen Eigenwert λ_{\max} von A gilt $\lambda_{\max} \leq Cp^d$ für alle $p \geq p_0$. Zudem erfüllen die Gewichte des Schätzers Annahme 7.3.

Dann gilt für alle $i = 1, \dots, n$ und $(x_1, \dots, x_D), (y_1, \dots, y_D), (x'_1, \dots, x'_D), (y'_1, \dots, y'_D) \in [0, 1]^D$:

$$\begin{aligned} a) \quad &\left(\bigotimes_{k=1}^D \vec{W}_i(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}_i(y_k) \right) = \mathcal{O} \left(\frac{1}{p^{D-d} h^D} \right) \\ b) \quad &\left(\bigotimes_{k=1}^D \vec{W}_i(x_k) - \bigotimes_{k=1}^D \vec{W}_i(x'_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}_i(y_k) - \bigotimes_{k=1}^D \vec{W}_i(y'_k) \right) = \mathcal{O} \left(\frac{1}{p^{D-d} h^D} \right) \end{aligned}$$

wobei $\vec{W}_i(x) := (W_{i1}(x), W_{i2}(x), \dots, W_{ip}(x))^T$ und $\bigotimes_{k=1}^D$ für das D -fache Kronecker-Produkt steht.

Beweis.

Über die *Cauchy-Schwarz-Ungleichung* folgt wie im Beweis zu Proposition 3.9:

$$\left| \left(\bigotimes_{k=1}^D \vec{W}_i(x_k) \right)^T \cdot A \cdot \left(\bigotimes_{k=1}^D \vec{W}_i(y_k) \right) \right| \leq \|A\|_M \cdot \left\| \bigotimes_{k=1}^D \vec{W}_i(x_k) \right\|_2 \cdot \left\| \bigotimes_{k=1}^D \vec{W}_i(y_k) \right\|_2$$

Nun gilt nach Proposition 7.4:

$$\begin{aligned} \left\| \bigotimes_{k=1}^D \vec{W}_i(x_k) \right\|_2 &= \prod_{k=1}^D \|\vec{W}_i(x_k)\|_2 = \prod_{k=1}^D \left(\sum_{j=1}^p |W_{ij}(x_k)|^2 \right)^{\frac{1}{2}} \\ &\leq \prod_{k=1}^D \left(\frac{C'_{W_2}}{ph} \right)^{\frac{1}{2}} = \left(\frac{C'_{W_2}}{ph} \right)^{\frac{D}{2}} \end{aligned}$$

Analog für $\left\| \bigotimes_{k=1}^D \vec{W}_i(y_k) \right\|_2$. Der Rest ist absolut identisch zum Beweis von Proposition 3.9. □

Proposition 7.7.

Sei $X := (X_i)_{i \in \{1, \dots, n\}} \sim F$ eine Folge unabhängiger reellwertiger Zufallsvariablen und $Y := (Y_i)_{i \in \{1, \dots, n\}}$ eine, von X linear abhängige Stichprobe, das bedeutet Y_i ist abhängig von $f(X_i)$ mit f linear für alle $i \in \mathbb{N}$. Sei weiter $X_1^*, \dots, X_n^* \mid Y \stackrel{\text{iid}}{\sim} F_n$ für alle $n \in \mathbb{N}$, wobei F_n die empirische Verteilung von X ist. Falls für alle $n \in \mathbb{N}$ das $2 \cdot k$ -te Moment von X_n existiert und $\text{Var}(X_n^k)$ durch eine Konstante $C < \infty$ beschränkt ist, dann gilt für alle $i = 1, \dots, n$:

$$\mathbb{E}^*[(X_i^*)^k] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[(X_i)^k] + o_{\text{f.s.}}(1)$$

Analoge Aussage gilt für Verbundmomente verschiedener Folgen von Zufallsvariablen.

Beweis.

Der Beweis ist identisch zum Beweis von Proposition 6.3, allerdings kann für $(X_n^k)_{n \in \mathbb{N}}$ nicht mehr (2. SLLN) angewendet werden, da dieses eine identische Verteilung voraussetzt, was hier im Allgemeinen nicht mehr der Fall ist. Wir müssen daher auf (1. SLLN) zurückgreifen. Wenn nun $\text{Var}(X_n)$ für alle $n \in \mathbb{N}$ durch eine Konstante $C \in \mathbb{R}^+$ beschränkt ist, dann gilt nach dem *Basler Problem*:

$$\sum_{n \in \mathbb{N}} \frac{\text{Var}(X_n)}{n^2} \leq C \sum_{n \in \mathbb{N}} \frac{1}{n^2} = \frac{C\pi^2}{6} < \infty \quad (7.5)$$

Gilt also $\text{Var}(X_n^k) \leq C < \infty$ für alle $n \in \mathbb{N}$, so folgt mit (1. SLLN):

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n ((X_i)^k - \mathbb{E}[(X_i)^k]) &\xrightarrow{\text{f.s.}} 0 \Leftrightarrow \frac{1}{n} \sum_{i=1}^n (X_i)^k - \frac{1}{n} \sum_{i=1}^n \mathbb{E}[(X_i)^k] = o_{\text{f.s.}}(1) \\ &\Leftrightarrow \mathbb{E}^*[(X_i^*)^k] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[(X_i)^k] + o_{\text{f.s.}}(1) \end{aligned}$$

□

8. Eigenschaften der asynchronen Schätzer

Wir werden nun die Gültigkeit der in Kapitel 4, 5 und 6 ausgearbeiteten Sätze für den asynchronen Fall verifizieren. Manches davon lässt sich mehr oder weniger eins zu eins übernehmen, manchmal müssen die Sätze und allen voran die Voraussetzungen aber leicht umformuliert werden. Da die Beweise teilweise völlig analog zum synchronen Fall durchführbar sind, werden wir an diversen Stellen abkürzen. Der größte Unterschied wird sein, dass nun deterministische Funktionen und Fehlerterme zuerst geglättet und dann gemittelt werden. Ein exemplarischer Vergleich für μ , R und σ :

Synchroner Fall: Verwende Proposition 3.7 und 3.9.

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_j(x) \mu(x_j) &= \sum_{j=1}^p W_j(x) \mu(x_j) = \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) \\ \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) R(x_j, x_k) &= \sum_{j,k=1}^p W_j(s) W_k(t) R(x_j, x_k) = R(s, t) + \mathcal{O}(h^{1+\alpha_R}) \\ \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_j(s) W_k(t) \sigma_{jk} &= \sum_{j,k=1}^p W_j(s) W_k(t) \sigma_{jk} = \vec{W}(x)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}(x) = \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right)\end{aligned}$$

Asynchroner Fall: Verwende Proposition 7.5 und 7.6.

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \mu(x_{ij}) &= \frac{1}{n} \sum_{i=1}^n (\mu(x) + \mathcal{O}(h^{1+\alpha_\mu})) = \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) \\ \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) R(x_{ij}, x_{ik}) &= \frac{1}{n} \sum_{i=1}^n (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) = R(s, t) + \mathcal{O}(h^{1+\alpha_R}) \\ \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) \sigma_{jk} &= \frac{1}{n} \sum_{i=1}^n (\vec{W}_i(x)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}_i(x)) = \frac{1}{n} \sum_{i=1}^n \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right) \\ &= \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right)\end{aligned}$$

8.1. Bias und Varianz

8.1.1. Trendschätzer

Satz 8.1.

Sei $\tilde{\mu}$ der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt. Dann gilt für alle $x \in [0, 1]$:

$$\begin{aligned} \text{Bias}(\tilde{\mu}(x)) &= \mathcal{O}(h^{1+\alpha_\mu}), \quad \text{Var}(\tilde{\mu}(x)) = \mathcal{O}\left(\frac{1+h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \\ \Rightarrow \text{MSE}(\tilde{\mu}(x)) &= \mathcal{O}\left(h^{2(1+\alpha_\mu)} + \frac{1+h^{1+\alpha_R}}{n} + \frac{1}{np^{1-d_1}h}\right) \end{aligned}$$

Aus $h \rightarrow 0$ folgt $\text{Bias}(\tilde{\mu}(x)) \rightarrow 0$, und damit auch unmittelbar die punktweise asymptotische Erwartungstreue von $\hat{\mu}$. Da die Schranke des Bias für alle $x \in [0, 1]$ existiert, gilt die Abschätzung auch bzgl. des Supremums über x , der Schätzer ist also auch gleichmäßig asymptotisch erwartungstreu. Bemerkung 4.2 gilt ebenfalls im asynchronen Fall.

Beweis.

Mit Proposition 7.5 und 7.6 erfolgt der Beweis absolut analog zum Beweis von Satz 4.1. \square

8.1.2. Kovarianzschätzer

Satz 8.2.

Sei \tilde{R} der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt. Dann gilt für alle $(s, t) \in [0, 1]^2$:

$$\begin{aligned} \text{Bias}(\tilde{R}(s, t)) &= \mathcal{O}\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h} + \frac{1}{n} + h^{2(1+\alpha_\mu)}\right), \quad \text{Var}(\tilde{R}(s, t)) = o(1) \\ \Rightarrow \text{MSE}(\tilde{R}(s, t)) &= \mathcal{O}\left(\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1}h} + \frac{1}{n} + h^{2(1+\alpha_\mu)}\right)^2\right) + o(1) \end{aligned}$$

Bei der korrigierten Version des Schätzers, würde der Faktor $\frac{1}{n}$ im Bias entfallen, die anderen drei Summanden entstehen durch die Glättung, den Fehlerterm und die asynchronen Messpunkten, wie wir gleich im Beweis sehen werden. Aus $h \rightarrow 0$ und mit $p^{1-d_1}h \rightarrow \infty$ folgt $\text{Bias}(\hat{R}(s, t)) \rightarrow 0$, und damit wieder die punktweise asymptotische Erwartungstreue von \hat{R} . Da die Schranke des Bias für alle $(s, t) \in [0, 1]^2$ existiert, gilt die Abschätzung auch bzgl. des Supremums über (s, t) , der Schätzer ist also auch gleichmäßig asymptotisch erwartungstreu. Auch hier wäre unter den erweiterten Annahmen aus Bemerkung 3.2 und 3.5 analog zu Bemerkung 4.2 eine schnellere Konvergenzrate erzielbar, wieder durch die größeren Exponenten $m + \alpha_R$ bzw. $2(m + \alpha_\mu)$ von h .

Beweis.

Für alle $x \in [0, 1]$ und alle $i = 1, \dots, n$ gilt mit Proposition 7.5:

$$\begin{aligned}
\tilde{\mu}_i(x) - \tilde{\mu}(x) &= \sum_{j=1}^p W_{ij}(x)Y_{ij} - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_{lj}(x)Y_{lj} \\
&= \sum_{j=1}^p W_{ij}(x)(\mu(x_{ij}) + Z_i(x_{ij}) + \varepsilon_{ij}) - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_{lj}(x)(\mu(x_{lj}) + Z_l(x_{lj}) + \varepsilon_{lj}) \\
&= \frac{1}{n} \sum_{l=1}^n \left(\underbrace{\sum_{j=1}^p W_{ij}(x)\mu(x_{ij})}_{=\mu(x)+\mathcal{O}(h^{1+\alpha_\mu})} - \underbrace{\sum_{j=1}^p W_{lj}(x)\mu(x_{lj})}_{=\mu(x)+\mathcal{O}(h^{1+\alpha_\mu})} \right) \\
&\quad + \underbrace{\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_{ij}(x)Z_i(x_{ij}) - W_{lj}(x)Z_l(x_{lj}))}_{=: W_i Z_i(x)} + \underbrace{\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_{ij}(x)\varepsilon_{lj} - W_{lj}(x)\varepsilon_{lj})}_{=: W_i \varepsilon_i(x)} \\
&= W_i Z_i(x) + W_i \varepsilon_i(x) + \mathcal{O}(h^{1+\alpha_\mu}) \tag{8.1}
\end{aligned}$$

$W_i Z_i(x)$ und $W_i \varepsilon_i(x)$ sind die Äquivalente zu $W Z_i(x)$ und $W \varepsilon_i(x)$ aus Gleichung (4.2), nur dass W nun ebenfalls vom Index der Stichprobeneinheit i abhängt. Damit kann der Schätzer \tilde{R} nach Gleichung (7.2) für alle $(s, t) \in [0, 1]^2$ wie folgt zerlegt werden.

$$\begin{aligned}
\tilde{R}(s, t) &= \frac{1}{n} \sum_{i=1}^n (\tilde{\mu}_i(s) - \tilde{\mu}(s))(\tilde{\mu}_i(t) - \tilde{\mu}(t)) \\
&= \frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) + W_i \varepsilon_i(s) + \mathcal{O}(h^{1+\alpha_\mu})) (W_i Z_i(t) + W_i \varepsilon_i(t) + \mathcal{O}(h^{1+\alpha_\mu}))
\end{aligned}$$

Daraus würden insgesamt 9 neue Summanden entstehen. Jedoch begründen wir nun, dass jeder Summand, welcher genau einen Faktor $\mathcal{O}(h^{1+\alpha_\mu})$ enthält, gleich 0 ist, wodurch sich der Term auf insgesamt 4 Summanden $+(\mathcal{O}(h^{1+\alpha_\mu}))^2 = \mathcal{O}(h^{2(1+\alpha_\mu)})$ reduziert. Wir betrachten o. B. d. A. den Summanden $W_i Z_i(s)\mathcal{O}(h^{1+\alpha_\mu})$.

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n W_i Z_i(s)\mathcal{O}(h^{1+\alpha_\mu}) &= \mathcal{O}(h^{1+\alpha_\mu}) \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_{ij}(s)Z_i(x_{ij}) - W_{lj}(s)Z_l(x_{lj})) \right) \\
&= \mathcal{O}(h^{1+\alpha_\mu}) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(s)Z_i(x_{ij}) - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_{lj}(s)Z_l(x_{lj}) \right) \\
&= 0
\end{aligned}$$

$\mathcal{O}(h^{1+\alpha_\mu})$ darf aus der Summe herausgezogen werden, da dieser Term zwar von n , nicht jedoch von i oder der Variablen s abhängt, denn Proposition 7.5 wurde gleichmäßig für alle $i = 1, \dots, n$ und $x \in [0, 1]$ formuliert.

$$\begin{aligned}
 \Rightarrow \tilde{R}(s, t) &= \frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) W_i Z_i(t) + W_i Z_i(s) W_i \varepsilon_i(t) + W_i Z_i(t) W_i \varepsilon_i(s) + W_i \varepsilon_i(s) W_i \varepsilon_i(t)) + \mathcal{O}(h^{2(1+\alpha_\mu)}) \\
 \Rightarrow \mathbb{E}[\tilde{R}(s, t)] &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[W_i Z_i(s) W_i Z_i(t)] + \mathbb{E}[W_i Z_i(s)] \mathbb{E}[W_i \varepsilon_i(t)] + \mathbb{E}[W_i Z_i(t)] \mathbb{E}[W_i \varepsilon_i(s)] \\
 &\quad + \mathbb{E}[W_i \varepsilon_i(s) W_i \varepsilon_i(t)]) + \mathcal{O}(h^{2(1+\alpha_\mu)}) \tag{8.2}
 \end{aligned}$$

Dies sind wiederum die exakten Äquivalente zu Gleichung (4.3), für welche sich alle Schritte wegen dem angepassten Proposition 7.5 und 7.6 analog durchführen lassen. Es folgt somit direkt durch Gleichung (4.7):

$$\begin{aligned}
 \mathbb{E}[\tilde{R}(s, t)] - R(s, t) &= \frac{n-1}{n} \left(R(s, t) + \mathcal{O}(h^{1+\alpha_R}) + \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right) \right) \cdot \underbrace{\frac{1}{n} \sum_{i=1}^n 1}_{=1} + \mathcal{O}(h^{2(1+\alpha_\mu)}) - R(s, t) \\
 &= \mathcal{O}\left(h^{1+\alpha_R} + \frac{1}{p^{1-d_1} h} + \frac{1}{n} + h^{2(1+\alpha_\mu)}\right) \tag{8.3}
 \end{aligned}$$

Wir haben also hier wieder eine Verzerrung durch den Vorfaktor $\mathcal{O}\left(\frac{1}{n}\right)$, die Glättung $\mathcal{O}(h^{1+\alpha_R})$ und den Fehler $\mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right)$. Zudem tritt hier ein weiterer Term der Ordnung $\mathcal{O}(h^{2(1+\alpha_\mu)})$ auf. Dieser Term ist auf die asynchronen Messpunkte zurückzuführen. Da an verschiedenen Punkten x_{ij} beobachtet wird, ergo auch μ in jeder Stichprobe an verschiedenen Punkten ausgewertet wird, hebt sich dieser Teil in den Faktoren $\tilde{\mu}_i(s) - \tilde{\mu}(s)$ und $\tilde{\mu}_i(t) - \tilde{\mu}(t)$ nicht auf. Erst die Ausglättung durch die Gewichtsfunktionen führt zum konkreten Wert $\mu(s) - \mu(s)$ bzw. $\mu(t) - \mu(t)$, allerdings erneut mit dem Fehler $\mathcal{O}(h^{1+\alpha_\mu})$, innerhalb der Kovarianz quadratisch $\mathcal{O}(h^{2(1+\alpha_\mu)})$.

Für die Varianz benötigen wir wieder den quadrierten Erwartungswert und den Erwartungswert des quadrierten Schätzers. Nach Gleichung (8.2) und den darauffolgenden Rechenschritten, insbesondere $\mathbb{E}[W_i Z_i(\cdot)] = \mathbb{E}[W_i \varepsilon_i(\cdot)] = 0$ gilt:

$$\begin{aligned}
 \tilde{R}(s, t) &= \frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) W_i Z_i(t) + W_i Z_i(s) W_i \varepsilon_i(t) + W_i Z_i(t) W_i \varepsilon_i(s) + W_i \varepsilon_i(s) W_i \varepsilon_i(t)) + \mathcal{O}(h^{2(1+\alpha_\mu)}) \\
 \mathbb{E}[\tilde{R}(s, t)] &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[W_i Z_i(s) W_i Z_i(t)] + \mathbb{E}[W_i \varepsilon_i(s) W_i \varepsilon_i(t)]) + \mathcal{O}(h^{2(1+\alpha_\mu)})
 \end{aligned}$$

Und mit der 1. binomischen Formel:

$$\begin{aligned}
 (\tilde{R}(s, t))^2 &= \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) W_i Z_i(t) + W_i Z_i(s) W_i \varepsilon_i(t) + W_i Z_i(t) W_i \varepsilon_i(s) + W_i \varepsilon_i(s) W_i \varepsilon_i(t)) \right)^2 \\
 &\quad + 2 \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) W_i Z_i(t) + W_i Z_i(s) W_i \varepsilon_i(t) + W_i Z_i(t) W_i \varepsilon_i(s) + W_i \varepsilon_i(s) W_i \varepsilon_i(t)) \right) \\
 &\quad \cdot \mathcal{O}(h^{2(1+\alpha_\mu)}) + (\mathcal{O}(h^{2(1+\alpha_\mu)}))^2 \\
 (\mathbb{E}[\tilde{R}(s, t)])^2 &= \left(\frac{1}{n} \sum_{i=1}^n (\mathbb{E}[W_i Z_i(s) W_i Z_i(t)] + \mathbb{E}[W_i \varepsilon_i(s) W_i \varepsilon_i(t)]) \right)^2 \\
 &\quad + 2 \left(\frac{1}{n} \sum_{i=1}^n (\mathbb{E}[W_i Z_i(s) W_i Z_i(t)] + \mathbb{E}[W_i \varepsilon_i(s) W_i \varepsilon_i(t)]) \right) \mathcal{O}(h^{2(1+\alpha_\mu)}) + (\mathcal{O}(h^{2(1+\alpha_\mu)}))^2
 \end{aligned}$$

Würde man nun alle Ausdrücke für sich untersuchen, hätte man insgesamt 28 Summanden, die man betrachten müsste, wir beschränken uns daher auf einige Beispielsummanden, da sich die restlichen ähnlich diskutieren lassen. Die jeweils ersten entsprechen gerade den Summanden aus Gleichung (4.8), mit dem einzigen Unterschied, dass hier $W_i Z_i$ bzw. $W_i \varepsilon_i$ statt $W Z_i$ bzw. $W \varepsilon_i$ auftaucht. Trotzdem lassen sich wegen Annahme 7.3 alle Rechenschritte analog zu denen von Gleichung (4.8) und folgende durchführen. Wenn nun die vier restlichen Summanden im Unendlichen verschwinden, so verschwindet damit auch die Varianz asymptotisch. Dass der Summand $(\mathcal{O}(h^{2(1+\alpha_\mu)}))^2$ gegen 0 konvergiert, ist offensichtlich. Da für die Varianz auch noch von $(\tilde{R}(s, t))^2$ der Erwartungswert gebildet wird, genügt es für alle restlichen Summanden o. B. d. A. $\mathcal{O}(h^{2(1+\alpha_\mu)}) \frac{1}{n} \sum_{i=1}^n \mathbb{E}[W_i Z_i(s) W_i Z_i(t)]$ zu betrachten. Für diesen gilt aber mit Proposition 7.5, über den gleichen Rechenweg wie bei Gleichung (4.4):

$$\begin{aligned}
 \mathcal{O}(h^{2(1+\alpha_\mu)}) \frac{1}{n} \sum_{i=1}^n \mathbb{E}[W_i Z_i(s) W_i Z_i(t)] &= \mathcal{O}(h^{2(1+\alpha_\mu)}) \frac{1}{n} \sum_{i=1}^n \frac{n-1}{n} (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) \\
 &= \underbrace{\frac{n-1}{n}}_{\rightarrow 1} \left(\underbrace{R(s, t)}_{\leq E_{Z,2} < \infty} + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} \right) \underbrace{\mathcal{O}(h^{2(1+\alpha_\mu)})}_{\rightarrow 0} = o(1)
 \end{aligned}$$

Somit folgt insgesamt:

$$\text{Var}(\tilde{R}(s, t)) = o(1)$$

□

8.2. Asymptotische Normalität

8.2.1. Trendschätzer

Satz 8.3.

Sei $\tilde{\mu}$ der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelte $\sqrt{nh}^{1+\alpha_\mu} \rightarrow 0$ sowie $\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty$. Dann gilt:

$$\sqrt{n}(\tilde{\mu}(x) - \mu(x)) \xrightarrow{w} \mathcal{G}(0, R) \text{ in } (\mathcal{C}([0, 1]), \|\cdot\|_\infty)$$

$\mathcal{G}(0, R)$ ist ein eindeutig bestimmter Gauß-Prozess mit Erwartungswert 0 und Kovarianzfunktion R , wobei R der Kovarianzfunktion von Z aus (A.3) entspricht.

Beweis.

Der empirische Prozess lässt sich über Gleichung (4.39) mit Proposition 7.5 wieder wie folgt zerlegen:

$$\sqrt{n}(\tilde{\mu}(x) - \mu(x)) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) Z_i(x_j) \right) + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \right) + \sqrt{n} \cdot \mathcal{O}(h^{1+\alpha_\mu})$$

Durch die Erweiterungen in Annahme 7.1 und 7.3, lassen sich alle Summanden wieder vollkommen analog zum Fall synchroner Messpunkte diskutieren, da die Annahmen alle simultan für alle $i = 1, \dots, n$ gelten. \square

8.2.2. Kovarianzschätzer

Satz 8.4.

Sei \tilde{R} der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelte $\sqrt{nh}^{1+\alpha_R} \rightarrow 0$, $\frac{\sqrt{n}}{p^{1-d_1}h} \rightarrow 0$ sowie $\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty$ und $\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} < \infty$. Dann gilt:

$$\sqrt{n}(\tilde{R}(s, t) - R(s, t)) \xrightarrow{w} \mathcal{G}(0, \mathcal{R}) \text{ in } (\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$$

$\mathcal{G}(0, \mathcal{R})$ ist ein eindeutig bestimmtes Gauß-Feld mit Erwartungswert 0 und Kovarianzfunktion

$$\mathcal{R} : \begin{cases} [0, 1]^4 \rightarrow \mathbb{R} \\ (s, t, s', t') \mapsto M(s, t, s', t') - R(s, t)R(s', t') \end{cases}$$

wobei M bzw. R dem vierten Verbundmoment bzw. der Kovarianzfunktion von Z aus (A.3) entspricht.

Beweis.

Aus Gleichung (8.2) folgt die, zu Gleichung (4.24) identische Zerlegung:

$$\begin{aligned}
\sqrt{n}(\tilde{R}(s,t) - R(s,t)) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) (Z_i(x_{ij}) Z_i(x_{ik}) - R(x_{ij}, x_{ik})) \right) \\
&+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) (\varepsilon_{ij} \varepsilon_{ik} - \sigma_{jk}) \right) + \sqrt{n} \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) \sigma_{jk} \right) \\
&+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) \varepsilon_{ik} \right) \\
&+ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ik}) \varepsilon_{ij} \right) \\
&- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_{ik}(t) Z_i(x_{ik}) \right) \\
&- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(s) \varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_{ik}(t) \varepsilon_{ik} \right) \\
&- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_{ik}(t) \varepsilon_{ik} \right) \\
&- \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(s) \varepsilon_{ij} \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p W_{ik}(t) Z_i(x_{ik}) \right) \\
&+ \sqrt{n} \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) R(x_{ij}, x_{ik}) - R(s,t) \right) \\
&+ \mathcal{O}(\sqrt{nh}^{2(1+\alpha_\mu)}) \tag{8.4}
\end{aligned}$$

Abgesehen vom letzten Summanden, lassen sich durch Annahme 7.1 und 7.3 alle anderen wieder vollkommen analog zum Fall synchroner Messpunkte diskutieren, da die Annahmen alle simultan für alle $i = 1, \dots, n$ gelten. Zudem gilt für alle $n \in \mathbb{N}$ und mit den Voraussetzungen aus Satz 4.5:

$$\sup_{(s,t) \in [0,1]^2} |\sqrt{nh}^{2(1+\alpha_\mu)}| = \sqrt{nh}^{2(1+\alpha_\mu)} \leq nh^{2(1+\alpha_\mu)} \rightarrow 0$$

wodurch auch der letzte Summand bzgl. der Supremumsnorm asymptotisch verschwindet und damit die asymptotische Normalität erfüllt ist. \square

8.3. Konsistenz

8.3.1. Trendschätzer

Satz 8.5.

Sei $\tilde{\mu}$ der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelten alle Voraussetzungen aus Satz 8.3. Dann gilt:

- a) $\tilde{\mu}$ ist punktweise konsistent im quadratischen Mittel.
- b) $\tilde{\mu}$ ist punktweise stark konsistent.
- c) $\tilde{\mu}$ ist gleichmäßig schwach konsistent.

Beweis.

- a) Diese Eigenschaft folgt wie bei Satz 4.7 aus Satz 8.1.
- b) Analog zu Gleichung (4.39) folgt mit Proposition 7.5:

$$\tilde{\mu}(x) = \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) Z_i(x_{ij}) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \quad (8.5)$$

Nun lässt sich mit Annahme 7.3 im Prinzip alles komplett analog zum Fall synchroner Messpunkte diskutieren, bis auf zwei kleine aber feine Unterschiede.

1. Für die, für alle $p \in \mathbb{N}$ und feste $x, x_{ij} \in [0, 1]$ reellen Zufallsvariablen $\sum_{j=1}^p W_{ij}(x) Z_i(x_{ij})$ und $\sum_{j=1}^p W_{ij}(x) \varepsilon_{ij}$ wurde (2. SLLN) verwendet, welches auf der unabhängig identischen Verteilung bzgl. i beruht. Da wir hier aber nun asynchrone Messpunkte - und damit auch asynchrone Gewichte - haben, sind die gerade erwähnten Zufallsvariablen zwar nach wie vor unabhängig, jedoch nicht mehr identisch verteilt. Wir müssen daher auf (1. SLLN) zurückgreifen und zeigen wieder $\text{Var}(X_i) \leq C < \infty$, was nach Gleichung (7.5) die Anwendung von (1. SLLN) rechtfertigt.

Wir verwenden den *Verschiebungssatz*, womit die Existenz des 2. Moments ebenfalls gleich gesichert ist. Die Ungleichungen folgen für alle $i \in \mathbb{N}$ aus Annahme 7.3 und Proposition 7.2.

$$\begin{aligned}
\text{Var} \left(\sum_{j=1}^p W_{ij}(x) Z_i(x_{ij}) \right) &= \mathbb{E} \left[\left(\sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}) \right)^2 \right] - \left(\mathbb{E} \left[\sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}) \right] \right)^2 \\
&= \sum_{j_1, j_2=1}^p W_{ij_1}(s) W_{ij_2}(s) \mathbb{E}[Z_i(x_{ij_1}) Z_i(x_{ij_2})] - \left(\sum_{j=1}^p W_{ij}(s) \mathbb{E}[Z_i(x_{ij})] \right)^2 \\
&\stackrel{|\cdot|}{\leq} E_{Z,2} \sum_{j_1, j_2=1}^p |W_{ij_1}(s)| |W_{ij_2}(s)| = E_{Z,2} \left(\sum_{j=1}^p |W_{ij}(s)| \right)^2 \\
&\leq E_{Z,2} C_{W_1}^2 < \infty \\
\text{Var} \left(\sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \right) &\leq \dots \leq E_{\varepsilon,2}(p) C_{W_1}^2 < \infty \tag{8.6}
\end{aligned}$$

Statt (2. SLLN) im Fall synchroner Messpunkte, kann also hier immer (1. SLLN) verwendet werden.

2. Durch die nicht identische Verteilung der oben erwähnten Zufallsvariablen, ist logischerweise auch deren Erwartungswert für alle $i \in \mathbb{N}$ nicht gleich. Die Aussagen von (1. SLLN) gelten also jeweils in der Form:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] + o_{\text{f.s.}}(1) \tag{8.7}$$

Das spielt hier jedoch keine Rolle, denn in diesem Fall gilt:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) Z_i(x_{ij}) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\sum_{j=1}^p W_{ij}(x) Z_i(x_{ij}) \right] + o_{\text{f.s.}}(1) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \underbrace{\mathbb{E}[Z_i(x_{ij})]}_{=0} + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1) \\
\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \right] + o_{\text{f.s.}}(1) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \underbrace{\mathbb{E}[\varepsilon_{ij}]}_{=0} + o_{\text{f.s.}}(1) = o_{\text{f.s.}}(1)
\end{aligned}$$

Mit dieser Argumentation sind die Rechenschritte identisch zum Fall synchroner Messpunkte, wodurch für Gleichung (8.5) für alle $x \in [0, 1]$ folgt:

$$\tilde{\mu}(x) = \mu(x) + \underbrace{\mathcal{O}(h^{1+\alpha_\mu})}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \xrightarrow{\text{f.s.}} \mu(x)$$

- c) Der Beweis der gleichmäßig schwachen Konsistenz ist analog zum Beweis von Satz 4.7 c), legitimiert durch Satz 8.3.

□

8.3.2. Kovarianzschätzer

Satz 8.6.

Sei \tilde{R} der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelten alle Voraussetzungen aus Satz 8.4. Dann gilt:

- a) \tilde{R} ist punktweise konsistent im quadratischen Mittel.
- b) \tilde{R} ist punktweise stark konsistent.
- c) \tilde{R} ist gleichmäßig schwach konsistent.

Beweis.

- a) Diese Eigenschaft folgt wie bei Satz 4.8 aus Satz 8.2.
- b) Die punktweise starke Konsistenz wird wieder direkt gezeigt. Gleichung (8.2) liefert uns als Schätzer:

$$\tilde{R}(s, t) = \frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) W_i Z_i(t) + W_i Z_i(s) W_i \varepsilon_i(t) + W_i Z_i(t) W_i \varepsilon_i(s) + W_i \varepsilon_i(s) W_i \varepsilon_i(t)) + \mathcal{O}(h^{2(1+\alpha_\mu)}) \tag{8.8}$$

Auch hier lässt sich nach Annahme 7.3 alles analog zum Fall synchroner Messpunkte, bis auf die beiden Unterschiede, wie bereits schon im Beweis von Satz 8.5 erwähnt.

1. Für die, für alle $p \in \mathbb{N}$ und feste $x_{ij}, x_{ik}, s, t \in [0, 1]$ reellen Zufallsvariablen $\sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}), \sum_{j=1}^p W_{ij}(s) \varepsilon_{ij}, \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) Z_i(x_{ik}), \sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) \varepsilon_{ij} \varepsilon_{ik}$ und $\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) \varepsilon_{ik}$ muss erneut (1. SLLN) verwendet werden.

Wir zeigen wieder $\text{Var}(X_i) \leq C < \infty$, für alle $i \in \mathbb{N}$ mit Annahme 7.3 und Proposition 7.2.

$$\begin{aligned}
& \text{Var} \left(\sum_{j=1}^p W_{ij}(s) Z_i(x_{ij}) \right) \stackrel{\text{Gleichung (8.6)}}{\leq} E_{Z,2} C_{W_1}^{\prime 2} < \infty \\
& \text{Var} \left(\sum_{j=1}^p W_{ij}(s) \varepsilon_{ij} \right) \leq \dots \leq E_{\varepsilon,2}(p) C_{W_1}^{\prime 2} < \infty \\
\\
& \text{Var} \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) Z_i(x_{ik}) \right) \\
& = \mathbb{E} \left[\left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) Z_i(x_{ik}) \right)^2 \right] - \left(\mathbb{E} \left[\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) Z_i(x_{ik}) \right] \right)^2 \\
& = \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s) W_{ik_2}(t) \mathbb{E}[Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) Z_i(x_{ik_2})] \\
& \quad - \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) \mathbb{E}[Z_i(x_{ij}) Z_i(x_{ik})] \right)^2 \\
& \stackrel{|\cdot|}{\leq} E_{Z,4} \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p |W_{ij_1}(s)| |W_{ik_1}(t)| |W_{ij_2}(s)| |W_{ik_2}(t)| + E_{Z,2}^2 \left(\sum_{j,k=1}^p |W_{ij}(s)| |W_{ik}(t)| \right)^2 \\
& = (E_{Z,4} + E_{Z,2}^2) \left(\left(\sum_{j=1}^p |W_{ij}(s)| \right) \left(\sum_{k=1}^p |W_{ik}(t)| \right) \right)^2 \leq (E_{Z,4} + E_{Z,2}^2) C_{W_1}^{\prime 4} < \infty \\
& \text{Var} \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) \varepsilon_{ij} \varepsilon_{ik} \right) \leq \dots \leq (E_{\varepsilon,4}(p) + E_{\varepsilon,2}^2) C_{W_1}^{\prime 4} < \infty \\
& \text{Var} \left(\sum_{j,k=1}^p W_{ij}(s) W_{ik}(t) Z_i(x_{ij}) \varepsilon_{ik} \right) \leq \dots \leq E_{Z,2} E_{\varepsilon,2}(p) C_{W_1}^{\prime 4} < \infty \quad (8.9)
\end{aligned}$$

Statt (2. SLLN) im Fall synchroner Messpunkte, kann also hier immer (1. SLLN) verwendet werden.

2. Auch hier gelten die Aussagen von (1. SLLN) wegen der nicht identischen Verteilung jeweils in der Form von Gleichung (8.7). Am Beispielsummanden

$X_i = \sum_{j,k=1}^p W_j(s)W_k(t)Z_i(x_j)Z_i(x_k)$ sehen wir aber schnell, dass das im Endeffekt keine Rolle spielt, denn nach Proposition 7.5 gilt:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\sum_{j,k=1}^p W_{ij}(s)W_{ik}(t)Z_i(x_{ij})Z_i(x_{ik}) \right] + o_{f.s.}(1) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s)W_{ik}(t) \mathbb{E}[Z_i(x_{ij})Z_i(x_{ik})] + o_{f.s.}(1) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p W_{ij}(s)W_{ik}(t)R(x_{ij}, x_{ik}) + o_{f.s.}(1) = \frac{1}{n} \sum_{i=1}^n (R(s, t) + \mathcal{O}(h^{1+\alpha_R})) + o_{f.s.}(1) \\ &= R(s, t) + \mathcal{O}(h^{1+\alpha_R}) + o_{f.s.}(1) \end{aligned}$$

Nach den gegebenen Anforderungen an die Gewichte, egalisiert die Glättung quasi die Asynchronität der Messpunkte und die übergeordnete Mittelung bzgl. i spielt keine Rolle mehr.

Ab hier sind die Rechenschritte identisch zum Fall synchroner Messpunkte, wodurch für Gleichung (8.8) für alle $(s, t) \in [0, 1]^2$ folgt:

$$\hat{R}(s, t) = R(s, t) + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} + \underbrace{\mathcal{O}\left(\frac{1}{p^{1-d_1}h}\right)}_{\rightarrow 0} + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0} + \underbrace{\mathcal{O}(h^{2(1+\alpha_\mu)})}_{\rightarrow 0} \xrightarrow{f.s.} R(s, t)$$

- c) Der Beweis der gleichmäßig schwachen Konsistenz ist analog zum Beweis von Satz 4.7 c), legitimiert durch Satz 8.4.

□

9. Konfidenzbänder (asynchroner Fall)

9.1. Trendschätzer

Satz 9.1.

Sei $\tilde{\mu}$ der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelten alle Voraussetzungen aus Satz 8.3. Dann ist ein asymptotisches simultanes $(1-\gamma) \cdot 100\%$ Konfidenzband für die tatsächliche Trendfunktion μ gegeben durch

$$\tilde{B}_{\gamma,n,\mu}(x) := \left[\tilde{\mu}(x) - \frac{\tilde{c}_{\gamma,\mu}}{\sqrt{n}}, \tilde{\mu}(x) + \frac{\tilde{c}_{\gamma,\mu}}{\sqrt{n}} \right], \quad \mathbb{P}^* \left(\|\mathcal{G}(0, \tilde{R})\|_{\infty} > \tilde{c}_{\gamma,\mu} \right) = \gamma$$

wobei \tilde{R} der Kovarianzschätzer von Z nach Gleichung (7.2) ist und in \mathbb{P}^* auf die Stichprobe Y bedingt wird.

Bemerkung 9.2.

Auch hier ist es möglich, die Bänder über den standardisierten empirischen Prozess $\sqrt{n}(\mu - \tilde{\mu})$ mit $\tilde{\sigma}_Z(x) := \sqrt{\tilde{R}(x, x)}$ zu berechnen:

$$\tilde{B}'_{\gamma,n,\mu}(x) := \left[\tilde{\mu}(x) - \tilde{c}_{\gamma,\mu} \frac{\tilde{\sigma}_Z(x)}{\sqrt{n}}, \tilde{\mu}(x) + \tilde{c}_{\gamma,\mu} \frac{\tilde{\sigma}_Z(x)}{\sqrt{n}} \right]$$

$$\mathbb{P}^* \left(\|\mathcal{G}(0, \tilde{\rho}_Z)\|_{\infty} > \tilde{c}_{\gamma,\mu} \right) = \gamma, \quad \tilde{\rho}_Z(s, t) = \frac{\tilde{R}(s, t)}{\tilde{\sigma}_Z(s)\tilde{\sigma}_Z(t)}$$

Siehe dazu Bemerkung 5.2.

Beweis.

Im Fall synchroner Messpunkte, baut der Beweis der Konfidenzbänder im Wesentlichen auf der punktweise starken Konsistenz von \hat{R} auf. Diese gilt aber für \tilde{R} nach Satz 8.6 genauso, wodurch alle Beweisschritte identisch zu \hat{R} durchgeführt werden können. \square

9.2. Kovarianzschätzer

Satz 9.3.

Sei \tilde{R} der Schätzer aus Gleichung (7.2) innerhalb des Modells aus Kapitel 7. Weiter seien Annahme 7.1 und 7.3 erfüllt und es gelten alle Voraussetzungen aus Satz 8.4. Dann ist ein asymptotisches simultanes $(1-\gamma)\cdot 100\%$ Konfidenzbänder für die tatsächliche Kovarianzfunktion R von Z gegeben durch

$$\tilde{B}_{\gamma,n,R}(s,t) := \left[\tilde{R}(s,t) - \frac{\tilde{c}_{\gamma,R}}{\sqrt{n}}, \tilde{R}(s,t) + \frac{\tilde{c}_{\gamma,R}}{\sqrt{n}} \right], \quad \mathbb{P}^* \left(\|\mathcal{G}(0, \tilde{\mathcal{R}})\|_{\infty} > \tilde{c}_{\gamma,R} \right) = \gamma$$

wobei $\tilde{\mathcal{R}} : \begin{cases} [0,1]^4 \rightarrow \mathbb{R} \\ (s,t,s',t') \mapsto \tilde{M}(s,t,s',t') - \tilde{R}(s,t)\tilde{R}(s',t') \end{cases}$ und in \mathbb{P}^* auf die Stichprobe Y bedingt wird.

\tilde{M} sei dabei der **4. Verbundmomentenschätzer** von Z .

$$\tilde{M}(s,t,s',t') = \frac{1}{n} \sum_{i=1}^n (\tilde{\mu}_i(s) - \tilde{\mu}(s)) (\tilde{\mu}_i(t) - \tilde{\mu}(t)) (\tilde{\mu}_i(s') - \tilde{\mu}(s')) (\tilde{\mu}_i(t') - \tilde{\mu}(t'))$$

Bemerkung 9.4.

Auch hier sind die Bänder nach einer Standardisierung von $\sqrt{n}(R - \tilde{R})$ mit $\tilde{\sigma}_{ZZ}(s,t) := \sqrt{\tilde{\mathcal{R}}(s,t,s,t)}$ möglich:

$$\begin{aligned} \tilde{B}'_{\gamma,n,R}(s,t) &:= \left[\tilde{R}(s,t) - \tilde{c}_{\gamma,R} \frac{\tilde{\sigma}_{ZZ}(s,t)}{\sqrt{n}}, \tilde{R}(s,t) + \tilde{c}_{\gamma,R} \frac{\tilde{\sigma}_{ZZ}(s,t)}{\sqrt{n}} \right] \\ \mathbb{P}^* \left(\|\mathcal{G}(0, \tilde{\rho}_{ZZ})\|_{\infty} > \tilde{c}_{\gamma,R} \right) &= \gamma, \quad \tilde{\rho}_{ZZ}(s,t,s',t') = \frac{\tilde{\mathcal{R}}(s,t,s',t')}{\tilde{\sigma}_{ZZ}(s,t)\tilde{\sigma}_{ZZ}(s',t')} \end{aligned}$$

Siehe dazu Bemerkung 5.8.

Beweis.

Auch hier baut der Beweis hauptsächlich auf der punktweise starken Konsistenz von \hat{R} und \hat{M} auf. \tilde{R} ist nach Satz 8.6 punktweise stark konsistent. Für \tilde{M} gilt nach Gleichung (8.1):

$$\begin{aligned} \tilde{M}_n(s,t,s',t') &= \frac{1}{n} \sum_{i=1}^n (\tilde{\mu}_i(s) - \tilde{\mu}(s)) (\tilde{\mu}_i(t) - \tilde{\mu}(t)) (\tilde{\mu}_i(s') - \tilde{\mu}(s')) (\tilde{\mu}_i(t') - \tilde{\mu}(t')) \\ &= \frac{1}{n} \sum_{i=1}^n (W_i Z_i(s) + W_i \varepsilon_i(s) + \mathcal{O}(h^{1+\alpha_\mu})) (W_i Z_i(t) + W_i \varepsilon_i(t) + \mathcal{O}(h^{1+\alpha_\mu})) \\ &\quad \cdot (W_i Z_i(s') + W_i \varepsilon_i(s') + \mathcal{O}(h^{1+\alpha_\mu})) (W_i Z_i(t') + W_i \varepsilon_i(t') + \mathcal{O}(h^{1+\alpha_\mu})) \end{aligned} \tag{9.1}$$

Das wären insgesamt 81 zu untersuchende Summanden, die sich jedoch wieder sehr ähnlich diskutieren lassen, also beschränken wir uns wieder auf einige wesentliche Beispiele.

Wir zeigen daher zunächst wieder o. B. d. A. an $\frac{1}{n} \sum_{i=1}^n W_i Z_i(s) W_i Z_i(t) W_i Z_i(s') \mathcal{O}(h^{1+\alpha_\mu})$, dass alle Summanden, die den Faktor $\mathcal{O}(h^{1+\alpha_\mu})$ mindestens einmal beinhalten, vernachlässigbar sind. Durch ähnliche Umformungen wie bei Gleichung (5.8) folgt:

$$\begin{aligned}
 & \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_{ij}(s) Z_i(x_{ij}) - W_{ij}(s) Z_l(x_{ij})) \right) \left(\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^p (W_{ik}(t) Z_i(x_{ik}) - W_{ij}(t) Z_l(x_{ik})) \right) \\
 & \quad \cdot \left(\frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p (W_{ij}(s') Z_i(x_{ij}) - W_{ij}(s') Z_l(x_{ij})) \right) \mathcal{O}(h^{1+\alpha_\mu}) \\
 = & \mathcal{O}(h^{1+\alpha_\mu}) \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^3} \sum_{\substack{l_1, l_2=1 \\ l_3}}^n \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) \right) \\
 & - \mathcal{O}(h^{1+\alpha_\mu}) \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^3} \sum_{\substack{l_1, l_2=1 \\ l_3}}^n \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_{l_3}(x_{ij_2}) \right) + \dots \\
 = & \mathcal{O}(h^{1+\alpha_\mu}) \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) \\
 & - \mathcal{O}(h^{1+\alpha_\mu}) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1, k_1=1}^p W_{ij_1}(s) W_{ik_1}(t) Z_i(x_{ij_1}) Z_i(x_{ik_1}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1=1}^p W_{ij_1}(s') Z_i(x_{ij_1}) \right) + \dots
 \end{aligned}$$

In diesem Term betrachten wir das Konvergenzverhalten des ersten Summanden. Sei

$$X_i := \sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}),$$

eine für alle $p \in \mathbb{N}$ und feste $s, t, s' \in [0, 1]$ reellwertige, bzgl. i unabhängig, aber nicht identisch verteilte Zufallsvariable. Um (1. SLLN) anwenden zu können, muss deren Varianz wieder durch eine Konstante beschränkt werden. Die Rechenschritte und Umformungen funktionieren nach dem gleichen Prinzip wie bei Gleichung (8.9), weswegen wir aus Gründen des Umfangs wieder darauf verzichten werden. Letztlich gilt mit Annahme 7.3 und Proposition 7.2

$$\text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) \right) \leq (E_{Z,6} + E_{Z,3}^2) C_{W_1}^{\prime 6} < \infty$$

ergo für den ersten Summanden aus obiger Gleichung mit (1. SLLN)

$$\begin{aligned}
 \mathcal{O}(h^{1+\alpha_\mu})\bar{X} &= \mathcal{O}(h^{1+\alpha_\mu})\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\sum_{\substack{j_1, k_1=1 \\ j_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s')Z_i(x_{ij_1})Z_i(x_{ik_1})Z_i(x_{ij_2})\right] + o_{f.s.}(1)\right) \\
 &= \mathcal{O}(h^{1+\alpha_\mu})\left(\frac{1}{n}\sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s')\mathbb{E}[Z_i(x_{ij_1})Z_i(x_{ik_1})Z_i(x_{ij_2})] + o_{f.s.}(1)\right) \\
 &\stackrel{|\cdot|}{\leq} \mathcal{O}(h^{1+\alpha_\mu})\left(\frac{1}{n}\sum_{i=1}^n \left(\sum_{j_1=1}^p |W_{ij_1}(s)|\right)\left(\sum_{k_1=1}^p |W_{ik_1}(t)|\right)\left(\sum_{j_2=1}^p |W_{ij_2}(s')|\right)E_{Z,3} + o_{f.s.}(1)\right) \\
 &\leq \mathcal{O}(h^{1+\alpha_\mu})\left(\frac{1}{n}\sum_{i=1}^n E_{Z,3}C_{W_1}^{\prime 3} + o_{f.s.}(1)\right) = \underbrace{\mathcal{O}(h^{1+\alpha_\mu})}_{\rightarrow 0} \underbrace{(E_{Z,3}C_{W_1}^{\prime 3})}_{< \infty} + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0} \xrightarrow{f.s.} 0
 \end{aligned}$$

Alle weiteren Summanden in diesem Term lassen sich genauso mit (1. SLLN) abarbeiten, wobei $\mathcal{O}(h^{1+\alpha_\mu})$ immer für die Konvergenz gegen 0 sorgt und alle anderen Faktoren endlich sind. Für Gleichung (9.1) bedeutet das:

$$\begin{aligned}
 \tilde{M}_n(s, t, s', t') &= \frac{1}{n}\sum_{i=1}^n (W_i Z_i(s) + W_i \varepsilon_i(s))(W_i Z_i(t) + W_i \varepsilon_i(t)) \\
 &\quad \cdot (W_i Z_i(s') + W_i \varepsilon_i(s'))(W_i Z_i(t') + W_i \varepsilon_i(t')) + o_{f.s.}(1) \quad (9.2)
 \end{aligned}$$

Jetzt befinden wir uns im asynchronen Äquivalent zu Gleichung (5.6), woraus die gleiche Zerlegung wie bei Gleichung (5.7) folgt, nur jeweils inklusive $+o_{f.s.}(1)$.

Die Diskussion der einzelnen Terme erfolgt nun wieder analog, nur dass (1. SLLN) statt (2. SLLN) verwendet wird, welches wegen der nicht identischen Verteilung wieder eine beschränkte Varianz statt einem endlichen absoluten Moment fordert. Nach Annahme 7.1 an Z und ε ist diese endliche Varianz aber gesichert und es gilt für die im letzten Schritt relevanten Summanden aus den Termen $I - V$ von Gleichung (5.7):

$$\begin{aligned}
 I : \text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') W_{ik_2}(t') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) Z_i(x_{ik_2}) \right) &\leq (E_{Z,8} + E_{Z,4}^2) C_{W_1}'^8 \\
 II : \text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') W_{ik_2}(t') Z_i(x_{ij_1}) Z_i(x_{ik_1}) Z_i(x_{ij_2}) \varepsilon_{ik_2} \right) &\leq E_{Z,6} E_{\varepsilon,2}(p) C_{W_1}'^8 \\
 III : \text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') W_{ik_2}(t') Z_i(x_{ij_1}) Z_i(x_{ik_1}) \varepsilon_{ij_2} \varepsilon_{ik_2} \right) &\leq (E_{Z,4} E_{\varepsilon,2}(p) + E_{Z,2}^2 E_{\varepsilon,4}^2)(p) C_{W_1}'^8 \\
 IV : \text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') W_{ik_2}(t') Z_i(x_{ij_1}) \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} \right) &\leq E_{Z,2} E_{\varepsilon,6}(p) C_{W_1}'^8 \\
 V : \text{Var} \left(\sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s') W_{ik_2}(t') \varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} \right) &\leq (E_{\varepsilon,8}(p) + E_{\varepsilon,4}(p)^2) C_{W_1}'^8
 \end{aligned}$$

Alle Terme auf der rechten Seite sind letztlich endliche Konstanten, (1. SLLN) ist demnach anwendbar und alle Schritte wie im Fall synchroner Messpunkte durchführbar. Die punktweise starke Konsistenz gilt also auch für \tilde{M} , woraus nach den gleichen Argumenten wie im synchronen Fall die Gültigkeit der Konfidenzbänder resultiert. \square

10. Bootstrap-Methode der Konfidenzbänder (asynchroner Fall)

Dass die Bootstrapverfahren aus Kapitel 6 auch hier funktionieren, ist keineswegs selbstverständlich. Der synchrone Fall hatte den Vorteil, dass Zufallsvariablen der Art

$$\sum_{j=1}^p W_{ij}(x) Z_i(x_{ij}) \text{ bzw. } \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \text{ für alle } x \in [0, 1] \text{ bzgl. } i \text{ identisch verteilt waren,}$$

da $x_{1j} = x_{2j} = \dots = x_{nj}$ und damit auch $W_{1j}(x) = W_{2j}(x) = \dots = W_{nj}(x)$, was nun jedoch nicht mehr der Fall ist. Insbesondere bedeutet das, dass jedem Z_i bzw. ε_{ij} nun ein separater Punkt x_{ij} und damit auch ein separates Gewicht W_{ij} zugeordnet wird, was im Nachweis des Bootstraps berücksichtigt werden muss. Da dadurch einige Beweisschritte erheblich angepasst werden müssen, werden wir beide Beweise im Vergleich zu den vorherigen relativ ausgiebig durchführen.

10.1. Trendschätzer

Satz 10.1.

Die in Satz 9.1 angeführten Konfidenzbänder $\tilde{B}_{\gamma, n, \mu}(x)$ behalten unter sonst gleichen Bedingungen, und der zusätzlichen Voraussetzung $\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2} < \infty$, f. s. ihre Gültigkeit, wenn $\tilde{c}_{\gamma, \mu}$ durch $c_{\gamma, \mu}^*$ ersetzt wird, welches nach dem gleichen Bootstrapverfahren wie aus Satz 6.1 berechnet wird, wobei $\hat{\mu}$ durch $\tilde{\mu}$ ersetzt wird.

Beweis.

Dadurch dass sich die Messpunkte x_{ij} - und dadurch auch die Gewichte W_{ij} - nun in jeder Stichprobeneinheit ändern (können), können innerhalb des Bootstraps nicht nur die Zufallsvariablen an sich, sondern auch die Messpunkte und die Gewichte alle Werte zwischen x_{1j} bis x_{nj} bzw. W_{1j} bis W_{nj} mit der Wahrscheinlichkeit $\frac{1}{n}$ annehmen, und werden somit mehr oder weniger auch zu Zufallsvariablen. Da ihre Werte jedoch im normalen Modell deterministisch sind, verzichten wir auf eine weitere Abänderung der Schreibweise. Wichtig dabei ist jedoch, dass in jeder Bootstrap-Strichprobe die Zufallsvariablen, die dazugehörigen Datenpunkte und die Gewichte gleichzeitig gezogen werden, sie sind also jeweils voneinander abhängig.

$$\begin{aligned}
\sqrt{n}(\mu^*(x) - \tilde{\mu}(x)) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p W_{ij}(x) Y_{ij} \right)^* - \tilde{\mu}(x) \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_{ij}(x) Y_{ij})^* - \tilde{\mu}(x) \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_{ij}(x) \mu(x_{ij}) + W_{ij}(x) Z_i(x_{ij}) + W_{ij}(x) \varepsilon_{ij})^* - \tilde{\mu}(x) \right) \\
&\stackrel{\text{f.s.}}{\underset{\text{D}}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) \mu(x_{ij}))^* + (W_{ij}(x) Z_i(x_{ij}))^* + (W_{ij}(x) \varepsilon_{ij})^*) - \tilde{\mu}(x) \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p W_{ij}(x) \mu(x_{ij}) \right)^* + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) Z_i(x_{ij}))^* + (W_{ij}(x) \varepsilon_{ij})^*) - \tilde{\mu}(x) \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (\mu(x) + \mathcal{O}(h^{1+\alpha_\mu}))^* + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) Z_i(x_{ij}))^* + (W_{ij}(x) \varepsilon_{ij})^*) - \tilde{\mu}(x) \right) \\
&= \sqrt{n} \left(\mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_{ij}(x) Z_i(x_{ij}))^* + (W_{ij}(x) \varepsilon_{ij})^* - \mathbb{E}^*[\mu^*(x)] \right. \\
&\quad \left. + \mathbb{E}^*[\mu^*(x)] - \tilde{\mu}(x) \right) \tag{10.1}
\end{aligned}$$

Der vorletzte Schritt gilt mit Proposition 7.5.

Für den bedingten Erwartungswert gilt nun:

$$\begin{aligned}
 \mathbb{E}^*[\mu^*(x)] &= \mathbb{E}^* \left[\left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) Y_{ij} \right)^* \right] = \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_{ij}(x) Y_{ij})^* \right] \\
 &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (W_{ij}(x) \mu(x_{ij}) + W_{ij}(x) Z_i(x_{ij}) + W_{ij}(x) \varepsilon_{ij})^* \right] \\
 &\stackrel{\text{f.s.}}{\simeq} \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) \mu(x_{ij}))^* + (W_{ij}(x) Z_i(x_{ij}))^* + (W_{ij}(x) \varepsilon_{ij})^*) \right] \\
 &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p W_{ij}(x) \mu(x_{ij}) \right)^* \right] + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\mathbb{E}^* [(W_{ij}(x) Z_i(x_{ij}))^*] + \mathbb{E}^* [(W_{ij}(x) \varepsilon_{ij})^*] \right) \\
 &= \mathbb{E}^* \left[\frac{1}{n} \sum_{i=1}^n (\mu(x) + \mathcal{O}(h^{1+\alpha_\mu}))^* \right] + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\mathbb{E}^* [(W_{ij}(x) Z_i(x_{ij}))^*] + \mathbb{E}^* [(W_{ij}(x) \varepsilon_{ij})^*] \right) \\
 &= \mu(x) + \mathcal{O}(h^{1+\alpha_\mu}) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\mathbb{E}^* [(W_{ij}(x) Z_i(x_{ij}))^*] + \mathbb{E}^* [(W_{ij}(x) \varepsilon_{ij})^*] \right) \quad (10.2)
 \end{aligned}$$

Zudem gilt nach den gleichen Rechenschritten wie bei Gleichung (6.5)

$$\mathbb{E}^*[\mu^*(x)] = \tilde{\mu}(x)$$

und somit für die Zerlegung des Schätzers:

$$\begin{aligned}
 \sqrt{n}(\mu^*(x) - \tilde{\mu}(x)) &\stackrel{\text{f.s.}}{\underset{D}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) Z_i(x_{ij}))^* - \mathbb{E}^* [(W_{ij}(x) Z_i(x_{ij}))^*]) \right) \Bigg\} =: S_n^{1*}(x) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) \varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x) \varepsilon_{ij})^*]) \right) \Bigg\} =: S_n^{2*}(x) \\
 &\quad + \mathcal{O}(h^{1+\alpha_\mu}) \quad (10.3)
 \end{aligned}$$

Hier gelten nun bzgl. der Zufallsstichprobe Y , ähnlich zu Gleichung (6.2) die folgenden empirischen Verteilungen:

$$\begin{aligned}
 (W_{1j}(x) Z_1(x_{1j}))^*, \dots, (W_{nj}(x) Z_n(x_{nj}))^* \mid Y &\stackrel{\text{iid}}{\simeq} \frac{1}{n} \sum_{i=1}^n \delta_{W_{ij}(x) Z_i(x_{ij})}, \quad x \in [0, 1], \quad j = 1, \dots, p \\
 (W_{1j}(x) \varepsilon_{1j})^*, \dots, (W_{nj}(x) \varepsilon_{nj})^* \mid Y &\stackrel{\text{iid}}{\simeq} \frac{1}{n} \sum_{i=1}^n \delta_{W_{ij}(x) \varepsilon_{ij}}, \quad x \in [0, 1], \quad j = 1, \dots, p
 \end{aligned} \quad (10.4)$$

Nach analoger Argumentation zu Gleichung (6.2) und Proposition 6.4, überträgt sich auch hier wieder die empirische Verteilung auf beliebige Produkte bzgl. der Messpunkte, und nun auch Funktionsargumente der Gewichte, was wir wieder in einer kurzen Proposition festhalten.

Proposition 10.2.

Für alle $n, D \in \mathbb{N}$ und $j_1, \dots, j_D = 1, \dots, p$ bzw. $x_1, \dots, x_D \in [0, 1]$ gilt:

$$\begin{aligned} \left(\prod_{k=1}^D (W_{ij_k}(x_k) Z_i(x_{j_k}))^* \right)_{i \in \{1, \dots, n\}} \mid Y \stackrel{D}{=} \left(\left(\prod_{k=1}^D W_{ij_k}(x_k) Z_i(x_{j_k}) \right)^* \right)_{i \in \{1, \dots, n\}} \mid Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\prod_{k=1}^D W_{ij_k}(x_k) Z_i(x_{j_k})} \\ \left(\prod_{k=1}^D (W_{ij_k}(x_k) \varepsilon_{ij_k})^* \right)_{i \in \{1, \dots, n\}} \mid Y \stackrel{D}{=} \left(\left(\prod_{k=1}^D W_{ij_k}(x_k) \varepsilon_{ij_k} \right)^* \right)_{i \in \{1, \dots, n\}} \mid Y \stackrel{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{\prod_{k=1}^D W_{ij_k}(x_k) \varepsilon_{ij_k}} \end{aligned}$$

Die Grenzverteilung wird wieder von den einzelnen Summanden determiniert, die dazugehörige Diskussion ist ähnlich zum Fall synchroner Messpunkte, jedoch mit der ein oder anderen Abweichung wegen der nicht identischen Verteilung der Ausgangsvariablen. Der Summand $\mathcal{O}(h^{1+\alpha_\mu})$ bedarf keiner weiteren Diskussion, da er unabhängig von x , also gleichmäßig gegen 0 konvergiert.

$$S_n^{1*}(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x) Z_i(x_{ij}))^* - \mathbb{E}^* [(W_{ij}(x) Z_i(x_{ij}))^*]) \right)$$

Dieser Term wird wieder mit dem funktionalen zentralen Grenzwertsatz aus Abschnitt 2.3.2 bearbeitet. Für $n \in \mathbb{N}, i = 1, \dots, n$ und $T = [0, 1]$ sei:

$$\begin{aligned} X_{n,i}^*(x) &= \frac{1}{\sqrt{n}} \sum_{j=1}^p (W_{ij}(x) Z_i(x_{ij}))^* \tag{10.5} \\ S_n^{1*}(x) &= \sum_{i=1}^n \left(X_{n,i}^*(x) - \mathbb{E}^* [X_{n,i}^*(x)] \right) \end{aligned}$$

Die Voraussetzungen sind nun wie bei Gleichung (6.6) und (6.7) gegeben. Um etwas Zeit zu sparen, kürzen wir hier einige Schritte ab, die wegen der bzgl. i bedingt unabhängig identischen Verteilung von $(W_{ij}(x) Z_i(x_{ij}))^*$ und deren Produkte bzgl. j äquivalent zu Gleichung (6.8) funktionieren, mit dem einzigen Unterschied, dass die Gewichte innerhalb der Bootstrap-Variablen betrachtet werden müssen.

Der erste gravierende Unterschied entsteht bei der Verwendung von Proposition 7.7 anstatt Proposition 6.3. Für $x, x' \in [0, 1]$ gilt also:

$$\begin{aligned}
 \mathbb{E}^* [S_n^{1*}(x) S_n^{1*}(x')] &= \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(x) Z_1(x_{1j_1}) \right)^* \left(\sum_{j_2=1}^p W_{1j_2}(x') Z_1(x_{1j_2}) \right)^* \right] \\
 &\quad - \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(x) Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(x') Z_1(x_{1j_2}) \right)^* \right] \\
 &\stackrel{(**)}{=} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{ij_1}(x) Z_i(x_{ij_1}) \right) \left(\sum_{j_2=1}^p W_{ij_2}(x') Z_i(x_{ij_2}) \right) \right] + o_{f.s.}(1) \\
 &\quad - \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{ij_1}(x) Z_i(x_{ij_1}) \right) \right] + o_{f.s.}(1) \right) \\
 &\quad \cdot \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_2=1}^p W_{ij_2}(x') Z_i(x_{ij_2}) \right) \right] + o_{f.s.}(1) \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x') \mathbb{E} [Z_i(x_{ij_1}) Z_i(x_{ij_2})] + o_{f.s.}(1) \\
 &\quad - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1=1}^p W_{ij_1}(x) \underbrace{\mathbb{E} [Z_i(x_{ij_1})]}_{=0} + o_{f.s.}(1) \right) \\
 &\quad \cdot \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_2=1}^p W_{ij_2}(x') \underbrace{\mathbb{E} [Z_i(x_{ij_2})]}_{=0} + o_{f.s.}(1) \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x') R(x_{ij_1}, x_{ij_2}) + o_{f.s.}(1) \\
 &\stackrel{\text{Proposition 7.5}}{=} \frac{1}{n} \sum_{i=1}^n (R(x, x') + \mathcal{O}(h^{1+\alpha_R})) + o_{f.s.}(1) = R(x, x') + \underbrace{\mathcal{O}(h^{1+\alpha_R})}_{\rightarrow 0} + \underbrace{o_{f.s.}(1)}_{\xrightarrow{f.s.} 0} \\
 &\stackrel{f.s.}{\rightarrow} R(x, x') \tag{10.6}
 \end{aligned}$$

(**): Nach Proposition 7.7, was nach Proposition 7.2 und Proposition 7.4 anwendbar ist.

Die Kennzeichnung (**) ist hier bewusst so gewählt, um einen transparenten Vergleich zu Gleichung (6.8) zu schaffen. Bedingung i) ist also f. s. erfüllt.

Über ebenfalls nahezu identische Schritte zu Gleichung (6.9), lässt sich auch Bedingung ii) nachweisen, nur dass hier ebenfalls die Gewichte innerhalb der Bootstrap-Variablen betrachtet und die Propositionen für den Fall asynchroner Messpunkte verwendet werden müssen.

Bedingung iii) - v) basiert hauptsächlich auf der Umhüllenden von $X_{n,i}^*(x)$. Diese entspricht aber mit Gleichung (6.10) und (6.11) über

$$\begin{aligned} |X_{n,i}^*(x)| &= \frac{1}{\sqrt{n}} \left| \sum_{j=1}^p (W_{ij}(x)Z_i(x_{ij}))^* \right| = \frac{1}{\sqrt{n}} \left| \sum_{j=1}^p W_{ij}(x)Z_i(x_{ij}) \right|^* \leq \frac{1}{\sqrt{n}} \left(\sum_{j=1}^p |W_{ij}(x)||Z_i(x_{ij})| \right)^* \\ &\leq \frac{1}{\sqrt{n}} \left((|Z_i(0)| + K_i) \sum_{j=1}^p |W_{ij}(x)| \right)^* \leq \frac{C'_{W_1}}{\sqrt{n}} (|Z_i(0)| + K_i)^* =: \Phi_{n,i}^* \end{aligned}$$

gerade der Umhüllenden aus dem Fall synchroner Messpunkte, abgesehen von den beiden unterschiedlichen Konstanten C_{W_1} und C'_{W_1} . Da in $\Phi_n^* = (\Phi_{n,1}^*, \dots, \Phi_{n,n}^*)$ die Messpunkte, sowie die Gewichte keine Rolle mehr spielen, ist für Bedingung iii) und iv) durch die die unabhängig identische Verteilung von $|Z_i(0)| + K_i$, und der daraus resultierenden unabhängig identischen Verteilung von $\Phi_{n,i}^*$ sogar Proposition 6.3 aus dem Fall synchroner Messpunkte anwendbar. Die Beweisschritte sind also in diesem Fall tatsächlich absolut identisch. Und auch Bedingung v) gilt wegen der erneut erlaubten Umformung

$$\left| \frac{1}{\sqrt{n}} \sum_{j=1}^p (W_{ij}(x)Z_i(x_{ij}))^* - \frac{1}{\sqrt{n}} \sum_{j=1}^p (W_{ij}(x')Z_i(x_{ij}))^* \right| = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^p W_{ij}(x)Z_i(x_{ij}) - \frac{1}{\sqrt{n}} \sum_{j=1}^p W_{ij}(x')Z_i(x_{ij}) \right|^*$$

Nach dem funktionalen zentralen Grenzwertsatz gilt also

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)Z_i(x_{ij}))^* - \mathbb{E}^* [(W_{ij}(x)Z_i(x_{ij}))^*]) \right) | Y \xrightarrow{w} \mathcal{G}(0, R)$$

f. s. in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$, wobei R die Kovarianzfunktion der Zufallsfunktion Z darstellt.

$$S_n^{2*}(x) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x)\varepsilon_{ij})^*]) \right)$$

Wir kombinieren wieder f. s. die gleichmäßige gleichgrade Stetigkeit und punktweise Konvergenz in \mathbb{P} .

f. s. Punktweise Konvergenz in \mathbb{P} :

Auch hier sparen wir uns etwas Arbeit, da die Schritte wegen der bedingt unabhängig identischen Verteilung von $(W_{ij}(x)\varepsilon_{ij})^*$ und deren Produkte wieder im Prinzip identisch zu Gleichung (6.14) sind, nur mit ebenfalls gebootstrapteten Gewichten. Für alle $x \in [0, 1]$ gilt:

$$\begin{aligned} \mathbb{E}^* [(S_n^{2*}(x))^2] &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x)\varepsilon_{ij})^*]) \right) \right)^2 \right] \\ &= \sum_{j_1, j_2=1}^p \left(\underbrace{\mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^* (W_{1j_2}(x)\varepsilon_{1j_2})^*]}_{=\text{Cov}^* ((W_{1j_1}(x)\varepsilon_{1j_1})^*, (W_{1j_2}(x)\varepsilon_{1j_2})^*)} - \mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x)\varepsilon_{1j_2})^*] \right) \end{aligned} \quad (10.7)$$

Nach Gleichung (10.4) gilt wiederum:

$$\text{Cov}^* ((W_{1j_1}(x)\varepsilon_{1j_1})^*, (W_{1j_2}(x)\varepsilon_{1j_2})^*) = \frac{1}{n} \sum_{i=1}^n W_{ij_1}(x)\varepsilon_{ij_1} W_{ij_2}(x)\varepsilon_{ij_2} - \left(\frac{1}{n} \sum_{i=1}^n W_{ij_1}(x)\varepsilon_{ij_1} \right) \left(\frac{1}{n} \sum_{i=1}^n W_{ij_2}(x)\varepsilon_{ij_2} \right)$$

Nun brauchen wir jedoch ein anderes Vorgehen als im synchronen Fall, da dieser Ausdruck zwar die Form einer Stichprobenkovarianz besitzt, aber dennoch nicht als solche interpretiert werden kann, da hierfür eine identische Verteilung von $W_{1j_1}(x)\varepsilon_{1j_1}, \dots, W_{nj_1}(x)\varepsilon_{nj_1}$ bzw. $W_{1j_2}(x)\varepsilon_{1j_2}, \dots, W_{nj_2}(x)\varepsilon_{nj_2}$ vorlegen müsste, was aber durch die unterschiedlichen Gewichte nicht der Fall ist.

Nach obiger Gleichung gilt mit Gleichung (10.7):

$$\begin{aligned} \mathbb{E}^*[(S_n^{2*}(x))^2] &= \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x) (\varepsilon_{ij_1} \varepsilon_{ij_2} - \sigma_{j_1 j_2}) - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \right)^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x) \sigma_{j_1 j_2} \end{aligned} \quad (10.8)$$

Wir benötigen eine f. s. Konvergenz gegen 0, dafür überprüfen wir jeden Summanden einzeln.

Der dritte Summand ist kein Problem, da er reellwertig ist und nach Proposition 7.6:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x) \sigma_{j_1 j_2} &= \frac{1}{n} \sum_{i=1}^n \vec{W}_i^T(x) \cdot \text{Cov}(\varepsilon) \cdot \vec{W}_i^T(x) = \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right) \cdot \frac{1}{n} \sum_{i=1}^n 1 \\ &= \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{1-d_1} h}\right) \end{aligned}$$

Für die anderen beiden Summanden können wir uns die Diskussion der Fehlerterme aus der Grenzverteilung der normalen Schätzer zu Nutze machen. Diese entsprechen nämlich gerade $\frac{1}{\sqrt{n}} \cdot S_n^2(x, x)$ aus Gleichung (4.32) und $\left(\frac{1}{\sqrt{n}} \cdot S_n^2(x)\right)^2$ aus Gleichung (4.13), nur bzgl. asynchroner Messpunkte. Da wir aber mit Proposition 7.6 ein Äquivalent zu Proposition 3.9 für die asynchronen Messpunkte haben, folgt schnell:

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \cdot S_n^2(x, x)\right)^2\right] &= \frac{1}{n} \mathbb{E}\left[(S_n^2(x, x))^2\right] = \mathcal{O}\left(\frac{1}{np^{2-d_2} h^2}\right) \stackrel{n \in \mathbb{N}}{=} \mathcal{O}\left(\frac{1}{p^{2-d_2} h^2}\right) \\ \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \cdot S_n^2(x)\right)^2\right] &= \frac{1}{n} \mathbb{E}\left[(S_n^2(x))^2\right] = \mathcal{O}\left(\frac{1}{np^{1-d_1} h}\right) \stackrel{n \in \mathbb{N}}{=} \mathcal{O}\left(\frac{1}{p^{1-d_1} h}\right) \end{aligned}$$

Daraus folgt nun

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^p W_{ij_1}(x) W_{ij_2}(x) (\varepsilon_{ij_1} \varepsilon_{ij_2} - \sigma_{j_1 j_2}) &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{2-d_2} h^2}\right) \\ \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij}\right)^2 &= \left(\mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{1-d_1} h}\right)\right)^2 = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{(p^{1-d_1} h)^2}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{1-d_1} h}\right) \end{aligned}$$

also insgesamt

$$\mathbb{E}^*[(S_n^{2*}(x))^2] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{2-d_2} h^2}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{1-d_1} h}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{1-d_1} h}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h}\right) \xrightarrow{\mathbb{P}} 0$$

Die punktweise Konvergenz in \mathbb{P} gilt also im 2-ten Mittel und dadurch auch stochastisch (daher die O-Notation bzgl. \mathbb{P}).

Damit die Konvergenz in \mathbb{P} auch fast sicher gilt, betrachten wir noch die beiden folgenden, durch die *Markov-Ungleichung* geltenden Ungleichungen, die für alle $n \in \mathbb{N}$ und alle $\gamma > 0$ gelten:

$$\begin{aligned}
 & \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x, x) \right| > \gamma \right) \leq \frac{1}{\gamma^2} \cdot \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \cdot S_n^2(x, x) \right)^2 \right] = \mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right) \\
 \Rightarrow & \sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x, x) \right| > \gamma \right) \leq \sum_{n=1}^{\infty} \mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right) \\
 \Rightarrow \exists C \in \mathbb{R}^+ : & \sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x, x) \right| > \gamma \right) \leq C \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2} < \infty \\
 & \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x) \right| > \gamma \right) \leq \frac{1}{\gamma^2} \cdot \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \cdot S_n^2(x) \right)^2 \right] = \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\
 \Rightarrow & \sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x) \right| > \gamma \right) \leq \sum_{n=1}^{\infty} \mathcal{O} \left(\frac{1}{p^{1-d_1} h} \right) \\
 \Rightarrow \exists C \in \mathbb{R}^+ : & \sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \cdot S_n^2(x) \right| > \gamma \right) \leq C \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1} h} < \infty \tag{10.9}
 \end{aligned}$$

Diese beiden endlichen Reihen gewährleisten nach Proposition 2.13 auch die fast sichere Konvergenz der ersten beiden Summanden von Gleichung (10.8), also gilt für Gleichung (10.7):

$$\mathbb{E}^* \left[(S_n^{2*}(x))^2 \right] = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \xrightarrow{\text{f.s.}} 0 \tag{10.10}$$

$S_n^{2*}(x)$ konvergiert somit für alle $x \in [0, 1]$ bedingt auf Y f. s. gegen 0 in \mathbb{P} .

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

$$\text{Sei } \Phi^*(x) = \|S_n^{2*}(x)\|_{\ell^2} = \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^*[(W_{ij}(x)\varepsilon_{ij})^*]) \right) \right)^2}.$$

Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $x \in [0, 1]$: $|S_n^{2*}(x)| \leq \Phi^*(x) \Leftrightarrow |S_n^{2*}| \leq \Phi^*$.
Mit Gleichung (10.7), (10.8) und (10.10) folgt weiterhin:

$$\begin{aligned} \mathbb{E}^*[\|\Phi^*(x)\|^2] &= \mathbb{E}^*[\|S_n^{2*}(x)\|_{\ell^2}^2] = \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^*[(W_{ij}(x)\varepsilon_{ij})^*]) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^*[(W_{ij}(x)\varepsilon_{ij})^*]) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^*[\|\Phi^*\|^2] &\stackrel{\text{f.s.}}{\leq} C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2}}_{< \infty} + C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1} h}}_{< \infty} < \infty \Rightarrow (\mathbb{E}^*[\|\Phi^*\|^2])^{\frac{1}{2}} \stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

zu ii):

Wir schätzen wieder die Überdeckungsanzahl bzgl. der kanonischen Metrik ab. Auf gleicher Argumentation für die bedingte Unabhängigkeit bzgl. i wie bei Gleichung (10.7) folgt mit Proposition 7.6 für alle $x, x' \in [0, 1]$:

$$\begin{aligned}
 d_n^{*2}(x, x') &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x)\varepsilon_{ij})^*]) \right) \right. \right. \\
 &\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x')\varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x')\varepsilon_{ij})^*]) \right) \right) \right]^2 \\
 &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(x)\varepsilon_{ij})^* - (W_{ij}(x')\varepsilon_{ij})^* - \mathbb{E}^* [(W_{ij}(x)\varepsilon_{ij})^* - (W_{ij}(x')\varepsilon_{ij})^*]) \right) \right) \right]^2 \\
 &= \sum_{j_1, j_2=1}^p \left(\mathbb{E}^* [((W_{1j_1}(x)\varepsilon_{1j_1})^* - (W_{1j_1}(x')\varepsilon_{1j_1})^*)((W_{1j_2}(x)\varepsilon_{1j_2})^* - (W_{1j_2}(x')\varepsilon_{1j_2})^*)) \right. \\
 &\quad \left. - \mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^* - (W_{1j_1}(x')\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x)\varepsilon_{1j_2})^* - (W_{1j_2}(x')\varepsilon_{1j_2})^*] \right) \\
 &= \sum_{j_1, j_2=1}^p \left(\mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^* (W_{1j_2}(x)\varepsilon_{1j_2})^*] - \mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^* (W_{1j_2}(x')\varepsilon_{1j_2})^*] \right. \\
 &\quad - \mathbb{E}^* [(W_{1j_1}(x')\varepsilon_{1j_1})^* (W_{1j_2}(x)\varepsilon_{1j_2})^*] + \mathbb{E}^* [(W_{1j_1}(x')\varepsilon_{1j_1})^* (W_{1j_2}(x')\varepsilon_{1j_2})^*] \\
 &\quad - \mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x)\varepsilon_{1j_2})^*] + \mathbb{E}^* [(W_{1j_1}(x)\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x')\varepsilon_{1j_2})^*] \\
 &\quad \left. + \mathbb{E}^* [(W_{1j_1}(x')\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x)\varepsilon_{1j_2})^*] - \mathbb{E}^* [(W_{1j_1}(x')\varepsilon_{1j_1})^*] \mathbb{E}^* [(W_{1j_2}(x')\varepsilon_{1j_2})^*] \right) \\
 &= \sum_{j_1, j_2=1}^p \text{Cov}^* ((W_{1j_1}(x)\varepsilon_{1j_1})^*, (W_{1j_2}(x)\varepsilon_{1j_2})^*) - 2 \sum_{j_1, j_2=1}^p \text{Cov}^* ((W_{1j_1}(x)\varepsilon_{1j_1})^*, (W_{1j_2}(x')\varepsilon_{1j_2})^*) \\
 &\quad + \sum_{j_1, j_2=1}^p \text{Cov}^* ((W_{1j_1}(x')\varepsilon_{1j_1})^*, (W_{1j_2}(x')\varepsilon_{1j_2})^*) \tag{10.11}
 \end{aligned}$$

Nun kann jeder dieser drei Summanden nach dem gleichen Schema wie bei Gleichung (10.8) und folgende mittels Gleichung (10.10) abgeschätzt werden (es spielt dabei auch keine Rolle, dass nun im Allgemeinen verschiedene x und x' in Cov^* auftreten, da sämtliche Schritte äquivalent ausgeführt werden können). Letztlich ergibt sich also auch wieder, über die Schreibweise der O-Notation:

$$\begin{aligned}
 d_n^{*2}(x, x') &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \\
 \Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}(x, x') &\stackrel{\text{f.s.}}{\leq} C \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \\
 \Rightarrow d_n^*(x, x') &\stackrel{\text{f.s.}}{\leq} \sqrt{C} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right)^{\frac{1}{2}} =: C_n
 \end{aligned}$$

Die Konstante C_n hat zwar nun erneut eine andere komplexere Form, wegen $C_n \rightarrow 0$ sind jedoch ab hier alle weiteren Schritte wieder im Prinzip identisch zu denen von Gleichung (4.14) und folgende, also ist $S_n^{2*}(x)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} . \square

10.2. Kovarianzschätzer

Satz 10.3.

Die in Satz 9.3 angeführten Konfidenzbänder $\tilde{B}_{\gamma,n,R}(x)$ behalten unter sonst gleichen Bedingungen, und der zusätzlichen Voraussetzung $\sum_{n=1}^{\infty} \frac{1}{p^{4-d_4} h^4} < \infty$, f. s. ihre Gültigkeit, wenn $\tilde{c}_{\gamma,R}$ durch $c_{\gamma,R}^*$ ersetzt wird, welches nach dem gleichen Bootstrapverfahren wie aus Satz 6.5 berechnet wird, wobei \hat{R} durch \tilde{R} ersetzt wird.

Beweis.

Wir beweisen aus Umfangsgründen wieder o. B. d. A. nur den Schätzer nach (B.3) a). Zunächst wird wieder $\mu_i^*(x) - \tilde{\mu}(x)$, für alle $i = 1, \dots, n$ vereinfacht.

$$\begin{aligned}
\mu_i^*(x) - \tilde{\mu}(x) &= \left(\sum_{j=1}^p W_{ij}(x) Y_{ij} \right)^* - \sum_{j=1}^p W_{ij}(x) \frac{1}{n} \sum_{i=1}^n Y_{ij} = \sum_{j=1}^p (W_{ij}(x) Y_{ij})^* - \sum_{j=1}^p W_{ij}(x) \frac{1}{n} \sum_{i=1}^n Y_{ij} \\
&= \sum_{j=1}^p \left(W_{ij}(x) (\mu(x_{ij}) + Z_i(x_{ij}) + \varepsilon_{ij}) \right)^* - \frac{1}{n} \sum_{l=1}^n \sum_{j=1}^p W_{lj}(x) (\mu(x_{lj}) + Z_l(x_{lj}) + \varepsilon_{lj}) \\
&\stackrel{\text{f.s.}}{\underset{\text{D}}{\simeq}} \underbrace{\left(\sum_{j=1}^p W_{ij}(x) \mu(x_{ij}) \right)^*}_{= \mu(x) + \mathcal{O}(h^{1+\alpha_\mu})} - \frac{1}{n} \sum_{l=1}^n \underbrace{\sum_{j=1}^p W_{lj}(x) \mu(x_{lj})}_{= \mu(x) + \mathcal{O}(h^{1+\alpha_\mu})} \\
&\quad + \sum_{j=1}^p (W_{ij}(x) Z_i(x_{ij}))^* - \underbrace{\sum_{j=1}^p \frac{1}{n} \sum_{l=1}^n W_{lj}(x) Z_l(x_{lj})}_{= \mathbb{E}^*[(W_{ij}(x) Z_i(x_{ij}))^*]} + \sum_{j=1}^p (W_{ij}(x) \varepsilon_{ij})^* - \underbrace{\sum_{j=1}^p \frac{1}{n} \sum_{l=1}^n W_{lj}(x) \varepsilon_{lj}}_{= \mathbb{E}^*[(W_{ij}(x) \varepsilon_{ij})^*]} \\
&= \underbrace{(\mu(x) + \mathcal{O}(h^{1+\alpha_\mu}))^* - \frac{1}{n} \sum_{l=1}^n (\mu(x) + \mathcal{O}(h^{1+\alpha_\mu}))}_{= \mathcal{O}(h^{1+\alpha_\mu})} \\
&\quad + \underbrace{\sum_{j=1}^p ((W_{ij}(x) Z_i(x_{ij}))^* - \mathbb{E}^*[(W_{ij}(x) Z_i(x_{ij}))^*])}_{:= (W_i Z_i)^*(x)} + \underbrace{\sum_{j=1}^p ((W_{ij}(x) \varepsilon_{ij})^* - \mathbb{E}^*[(W_{ij}(x) \varepsilon_{ij})^*])}_{:= (W_i \varepsilon_i)^*(x)} \\
&= (W_i Z_i)^*(x) + (W_i \varepsilon_i)^*(x) + \mathcal{O}(h^{1+\alpha_\mu})
\end{aligned}$$

Daraus folgt für R^* :

$$\begin{aligned}
 R^*(s, t) &= \frac{1}{n} \sum_{i=1}^n (\mu_i^*(s) - \tilde{\mu}(s)) (\mu_i^*(t) - \tilde{\mu}(t)) \\
 &= \frac{1}{n} \sum_{i=1}^n ((W_i Z_i)^*(s) + (W_i \varepsilon_i)^*(s) + \mathcal{O}(h^{1+\alpha_\mu})) ((W_i Z_i)^*(t) + (W_i \varepsilon_i)^*(t) + \mathcal{O}(h^{1+\alpha_\mu})) \\
 &= \frac{1}{n} \sum_{i=1}^n ((W_i Z_i)^*(s)(W_i Z_i)^*(t) + (W_i Z_i)^*(s)(W_i \varepsilon_i)^*(t) + (W_i Z_i)^*(s)\mathcal{O}(h^{1+\alpha_\mu}) \\
 &\quad + (W_i \varepsilon_i)^*(s)(W_i Z_i)^*(t) + (W_i \varepsilon_i)^*(s)(W_i \varepsilon_i)^*(t) + (W_i \varepsilon_i)^*(s)\mathcal{O}(h^{1+\alpha_\mu}) \\
 &\quad + \mathcal{O}(h^{1+\alpha_\mu})(W_i Z_i)^*(t) + \mathcal{O}(h^{1+\alpha_\mu})(W_i \varepsilon_i)^*(t) + (\mathcal{O}(h^{1+\alpha_\mu}))^2) \\
 \Rightarrow \mathbb{E}^*[R^*(s, t)] &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}^*[(W_i Z_i)^*(s)(W_i Z_i)^*(t)] + \mathbb{E}^*[(W_i Z_i)^*(s)(W_i \varepsilon_i)^*(t)] \\
 &\quad + \mathbb{E}^*[(W_i Z_i)^*(t)(W_i \varepsilon_i)^*(s)] + \mathbb{E}^*[(W_i \varepsilon_i)^*(s)(W_i \varepsilon_i)^*(t)]) + \mathcal{O}(h^{2(1+\alpha_\mu)})
 \end{aligned}$$

Die Vereinfachung des bedingten Erwartungswertes folgt aus dessen Linearität und der Tatsache, dass $\mathbb{E}^*[(W_i Z_i)^*(\cdot)] = \mathbb{E}^*[(W_i \varepsilon_i)^*(\cdot)] = 0$. Aufgrund der Erwartungstreue von R^* , folgt nun:

$$\begin{aligned}
 \sqrt{n}(R^*(s, t) - \tilde{R}(s, t)) &= \sqrt{n}(R^*(s, t) - \mathbb{E}^*[R^*(s, t)]) \\
 &\stackrel{\text{f.s.}}{\underset{\text{D}}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n ((W_i Z_i)^*(s)(W_i Z_i)^*(t) - \mathbb{E}^*[(W_i Z_i)^*(s)(W_i Z_i)^*(t)]) \right) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n ((W_i \varepsilon_i)^*(s)(W_i \varepsilon_i)^*(t) - \mathbb{E}^*[(W_i \varepsilon_i)^*(s)(W_i \varepsilon_i)^*(t)]) \right) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n ((W_i Z_i)^*(s)(W_i \varepsilon_i)^*(t) - \mathbb{E}^*[(W_i Z_i)^*(s)]\mathbb{E}^*[(W_i \varepsilon_i)^*(t)]) \right) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n ((W_i Z_i)^*(t)(W_i \varepsilon_i)^*(s) - \mathbb{E}^*[(W_i Z_i)^*(t)]\mathbb{E}^*[(W_i \varepsilon_i)^*(s)]) \right) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i)^*(s) \right) \mathcal{O}(h^{1+\alpha_\mu}) + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i)^*(t) \right) \mathcal{O}(h^{1+\alpha_\mu}) \\
 &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i \varepsilon_i)^*(s) \right) \mathcal{O}(h^{1+\alpha_\mu}) + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i \varepsilon_i)^*(t) \right) \mathcal{O}(h^{1+\alpha_\mu}) \\
 &\quad + \mathcal{O}(h^{2(1+\alpha_\mu)}) \tag{10.12}
 \end{aligned}$$

Wir begründen nun zunächst, dass die unteren fünf Summanden vernachlässigbar sind. $\mathcal{O}(h^{1+\alpha\mu})$, wie auch $\mathcal{O}(h^{2(1+\alpha\mu)})$, konvergieren dabei wieder unabhängig von s und t , ergo gleichmäßig gegen 0.

Nun gilt für die jeweils ersten Faktoren der Summanden:

$$\begin{aligned}\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i)^*(\cdot) \right) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(\cdot) Z_i(x_{ij}))^* - \mathbb{E}^*[(W_{ij}(\cdot) Z_i(x_{ij}))^*]) \right) \\ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i \varepsilon_i)^*(\cdot) \right) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p ((W_{ij}(\cdot) \varepsilon)^* - \mathbb{E}^*[(W_{ij}(\cdot) \varepsilon)^*]) \right)\end{aligned}$$

Das entspricht praktischerweise genau den Summanden $S_n^{1*}(\cdot)$ und $S_n^{2*}(\cdot)$ aus Gleichung (10.3), für welche wiederum $S_n^{1*}(\cdot) | Y \xrightarrow{\mathbb{W}} \mathcal{G}(0, R)$ und $S_n^{2*}(\cdot) | Y \xrightarrow{\mathbb{P}} 0$, jeweils f. s. in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$. Damit und mit Lemma 2.15 gilt schließlich f. s.:

$$\left\| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i Z_i)^* \right) \mathcal{O}(h^{1+\alpha\mu}) \right\|_\infty | Y = \underbrace{\mathcal{O}(h^{1+\alpha\mu})}_{\xrightarrow{\mathbb{P}} 0} \cdot \underbrace{\|S_n^{1*}\|_\infty}_{\xrightarrow{\mathbb{D}} \|\mathcal{G}(0, R)\|_\infty} | Y \xrightarrow{\mathbb{D}} 0$$

und da 0 eine Konstante ist, auch direkt f. s. die gleichmäßige Konvergenz in \mathbb{P} über Proposition 2.13, sowie mit Proposition 2.14 f. s.:

$$\left\| \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (W_i \varepsilon_i)^* \right) \mathcal{O}(h^{1+\alpha\mu}) \right\|_\infty | Y = \underbrace{\mathcal{O}(h^{1+\alpha\mu})}_{\xrightarrow{\mathbb{P}} 0} \cdot \underbrace{\|S_n^{2*}\|_\infty}_{\xrightarrow{\mathbb{P}} 0} | Y \xrightarrow{\mathbb{P}} 0$$

Nun lässt sich, durch die exakt gleiche Vorgehensweise wie bei Gleichung (6.18), o. B. d. A. der erste Summand von Gleichung (10.12) wieder umformen zu:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left((W_i Z_i)^*(s) (W_i Z_i)^*(t) - \mathbb{E}^* [(W_i Z_i)^*(s) (W_i Z_i)^*(t)] \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) Z_i(x_{ik}))^* - \mathbb{E}^* [(W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) Z_i(x_{ik}))^*]) \right) \end{aligned}$$

wodurch sich der empirische Prozess wie folgt vereinfacht:

$$\begin{aligned} & \sqrt{n}(R^*(s, t) - \tilde{R}(s, t)) \\ & \stackrel{\text{f.s.}}{\underset{\text{D}}{\simeq}} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) Z_i(x_{ik}))^* - \mathbb{E}^* [(W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) Z_i(x_{ik}))^*]) \right) \right) \Bigg\} =: S_n^{1*}(s, t) \\ & + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s) \varepsilon_{ij})^* (W_{ik}(t) \varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s) \varepsilon_{ij})^* (W_{ik}(t) \varepsilon_{ik})^*]) \right) \right) \Bigg\} =: S_n^{2*}(s, t) \\ & + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) \varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s) Z_i(x_{ij}))^* (W_{ik}(t) \varepsilon_{ik})^*]) \right) \right) \Bigg\} =: S_n^{3*}(s, t) \\ & + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(t) Z_i(x_{ij}))^* (W_{ik}(s) \varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(t) Z_i(x_{ij}))^* (W_{ik}(s) \varepsilon_{ik})^*]) \right) \right) \Bigg\} =: S_n^{4*}(s, t) \\ & + o_{\mathbb{P}}(1) \tag{10.13} \end{aligned}$$

$o_{\mathbb{P}}(1)$ übernimmt dabei die Rolle der fünf restlichen, asymptotisch verschwindenden Summanden, wie eingangs erklärt. Ein detaillierterer Restterm ist nicht notwendig, da wir uns nur für die funktionierende Prozesskonvergenz, nicht aber für deren Rate interessieren. Es erfolgt nun wieder eine Diskussion der einzelnen Summanden, die jeweils relevanten Verteilungen können aus Gleichung (10.4) und Proposition 10.2 erschlossen werden.

$$S_n^{1*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)Z_i(x_{ik}))^* - \mathbb{E}^*[(W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)Z_i(x_{ik}))^*]) \right) \right)$$

Dieser Term wird wieder mit dem funktionalen zentralen Grenzwertsatz aus Abschnitt 2.3.2 bearbeitet. Für $n \in \mathbb{N}$, $i = 1, \dots, n$ und $T = [0, 1]^2$ sei:

$$X_{n,i}^*(s, t) = \frac{1}{\sqrt{n}} \sum_{j,k=1}^p (W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)Z_i(x_{ik}))^* \quad (10.14)$$

$$S_n^{1*}(s, t) = \sum_{i=1}^n \left(X_{n,i}^*(s, t) - \mathbb{E}^*[X_{n,i}^*(s, t)] \right)$$

Die Voraussetzungen sind nun wie bei Gleichung (6.19) und (6.20) gegeben. Wir kürzen hier wieder die, zu Gleichung (6.21) äquivalenten Schritte ab. Für $(s, t), (s', t') \in [0, 1]^2$ gilt:

$$\begin{aligned} & \mathbb{E}^*[S_n^{1*}(s, t)S_n^{1*}(s', t')] \\ \stackrel{(**)}{=} & \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{ij_1}(s)Z_i(x_{ij_1}) \right) \left(\sum_{k_1=1}^p W_{ik_1}(t)Z_i(x_{ik_1}) \right) \left(\sum_{j_2=1}^p W_{ij_2}(s')Z_i(x_{ij_2}) \right) \left(\sum_{k_2=1}^p W_{ik_2}(t')Z_i(x_{ik_2}) \right) \right] + o_{f.s.}(1) \\ & - \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{ij_1}(s)Z_i(x_{ij_1}) \right) \left(\sum_{k_1=1}^p W_{ik_1}(t)Z_i(x_{ik_1}) \right) \right] + o_{f.s.}(1) \right) \\ & \quad \cdot \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_2=1}^p W_{ij_2}(s')Z_i(x_{ij_2}) \right) \left(\sum_{k_2=1}^p W_{ik_2}(t')Z_i(x_{ik_2}) \right) \right] + o_{f.s.}(1) \right) \\ = & \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s')W_{ik_2}(t') \mathbb{E}[Z_i(x_{ij_1})Z_i(x_{ik_1})Z_i(x_{ij_2})Z_i(x_{ik_2})] + o_{f.s.}(1) \\ & - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1, k_1=1}^p W_{ij_1}(s)W_{ik_1}(t) \mathbb{E}[Z_i(x_{ij_1})Z_i(x_{ik_1})] + o_{f.s.}(1) \right) \\ & \quad \cdot \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_2, k_2=1}^p W_{ij_2}(s')W_{ik_2}(t') \mathbb{E}[Z_i(x_{ij_2})Z_i(x_{ik_2})] + o_{f.s.}(1) \right) \\ = & \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s')W_{ik_2}(t') M(x_{ij_1}, x_{ik_1}, x_{ij_2}, x_{ik_2}) \\ & - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1, k_1=1}^p W_{ij_1}(s)W_{ik_1}(t) R(x_{ij_1}, x_{ik_1}) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_2, k_2=1}^p W_{ij_2}(s')W_{ik_2}(t') R(x_{ij_2}, x_{ik_2}) \right) + o_{f.s.}(1) \\ \stackrel{\text{Proposition 7.5}}{=} & \frac{1}{n} \sum_{i=1}^n (M(s, t, s', t') + \mathcal{O}(h^{1+\alpha})) - \left(\frac{1}{n} \sum_{i=1}^n (R(s, t) + \mathcal{O}(h^{1+\alpha})) \right) \left(\frac{1}{n} \sum_{i=1}^n (R(s', t') + \mathcal{O}(h^{1+\alpha})) \right) + o_{f.s.}(1) \end{aligned}$$

$$\begin{aligned}
 &= M(s, t, s', t') + \underbrace{\mathcal{O}(h^{1+\alpha})}_{\rightarrow 0} - \underbrace{(R(s, t) + \mathcal{O}(h^{1+\alpha}))}_{\rightarrow 0} \underbrace{(R(s', t') + \mathcal{O}(h^{1+\alpha}))}_{\rightarrow 0} + \underbrace{o_{\text{f.s.}}(1)}_{\xrightarrow{\text{f.s.}} 0} \\
 &\xrightarrow{\text{f.s.}} M(s, t, s', t') - R(s, t)R(s', t')
 \end{aligned} \tag{10.15}$$

(**): Nach Proposition 7.7, was nach Proposition 3.3 und Proposition 3.6 anwendbar ist.

Die Kennzeichnung (**) ist hier wieder für den Vergleich zu Gleichung (6.21). Bedingung i) ist also f. s. erfüllt.

Über ebenfalls nahezu identische Schritte zu Gleichung (6.22), lässt sich auch Bedingung ii) nachweisen, nur dass hier ebenfalls die Gewichte innerhalb der Bootstrap-Variablen betrachtet und die Propositionen für den Fall asynchroner Messpunkte verwendet werden müssen.

Bedingung iii) - v) basiert wieder auf der Umhüllenden von $X_{n,i}^*(s, t)$. Diese entspricht aber mit Gleichung (6.10) und (6.11) und Proposition 10.2 über

$$\begin{aligned}
 |X_{n,i}^*(s, t)| &= \frac{1}{\sqrt{n}} \left| \sum_{j,k=1}^p (W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)Z_i(x_{ik}))^* \right| = \frac{1}{\sqrt{n}} \left| \sum_{j,k=1}^p W_{ij}(s)Z_i(x_{ij})W_{ik}(t)Z_i(x_{ik}) \right|^* \\
 &\leq \frac{1}{\sqrt{n}} \left(\sum_{j,k=1}^p |W_{ij}(s)||W_{ik}(t)||Z_i(x_{ij})||Z_i(x_{ik})| \right)^* \\
 &\leq \frac{1}{\sqrt{n}} \left((|Z_i(0)| + K_i)^2 \sum_{j,k=1}^p |W_{ij}(s)||W_{ik}(t)| \right)^* \leq \frac{C_{W_1}^{\prime 2}}{\sqrt{n}} \left((|Z_i(0)| + K_i)^* \right)^2 =: \Phi_{n,i}^*
 \end{aligned}$$

wieder genau der Umhüllenden aus dem Fall synchroner Messpunkte mit der geringfügig anderen Konstanten C'_{W_1} statt C_{W_1} , wodurch der Beweis wieder identisch zu diesem Fall wird. Und auch Bedingung v) gilt wegen der durch Proposition 10.2 legitimierten Umformung:

$$\begin{aligned}
 &\left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p (W_{ij}(s)Z_i(x_{ij}))^* (W_{ij}(t)Z_i(x_{ij}))^* - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p (W_{ij}(s')Z_i(x_{ij}))^* (W_{ij}(t')Z_i(x_{ij}))^* \right| \\
 &= \left| \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_{ij}(s)W_{ij}(t)Z_i(x_{ij})Z_i(x_{ij}) - \frac{1}{\sqrt{n}} \sum_{j,k=1}^p W_{ij}(s')W_{ij}(t')Z_i(x_{ij})Z_i(x_{ij}) \right|^*
 \end{aligned}$$

Nach dem funktionalen zentralen Grenzwertsatz gilt also

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p ((W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)Z_i(x_{ik}))^* - \mathbb{E}^* [(W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)Z_i(x_{ik}))^*]) \right) \Big| Y \xrightarrow{w} \mathcal{G}(0, \mathcal{R})$$

f. s. in $(\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$, wobei \mathcal{R} durch Satz 4.5 gegeben ist.

$$S_n^{2*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^*(W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^*(W_{ik}(t)\varepsilon_{ik})^*] \right) \right)$$

Wir kombinieren wieder f. s. die gleichmäßige gleichgrade Stetigkeit und punktweise Konvergenz in \mathbb{P} .

f. s. Punktweise Konvergenz in \mathbb{P} :

Über Proposition 10.2 können die Rechenschritte wieder aus Gleichung (4.32) entnommen werden, was in Gleichung (6.26) resultiert, nur eben zusätzlich mit gebootstrapteten Gewichten. Für alle $(s, t) \in [0, 1]^2$ gilt also:

$$\begin{aligned} \mathbb{E}^* [(S_n^{2*}(s, t))^2] &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^*(W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^*(W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2 \right] \\ &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p \left(\mathbb{E}^* [(W_{1j_1}(s)\varepsilon_{1j_1})^*(W_{1k_1}(t)\varepsilon_{1k_1})^*(W_{1j_2}(s)\varepsilon_{1j_2})^*(W_{1k_2}(t)\varepsilon_{1k_2})^*] \right. \\ &\quad \left. - \mathbb{E}^* [(W_{1j_1}(s)\varepsilon_{1j_1})^*(W_{1k_1}(t)\varepsilon_{1k_1})^*] \mathbb{E}^* [(W_{1j_2}(s)\varepsilon_{1j_2})^*(W_{1k_2}(t)\varepsilon_{1k_2})^*] \right) \\ &=: \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p \text{Cov}^* \left((W_{1j_1}(s)\varepsilon_{1j_1})^*(W_{1k_1}(t)\varepsilon_{1k_1})^*, (W_{1j_2}(s)\varepsilon_{1j_2})^*(W_{1k_2}(t)\varepsilon_{1k_2})^* \right) \end{aligned} \tag{10.16}$$

und nach Gleichung (10.4) und Proposition 10.2:

$$\begin{aligned} &\text{Cov}^* \left((W_{1j_1}(s)\varepsilon_{1j_1})^*(W_{1k_1}(t)\varepsilon_{1k_1})^*, (W_{1j_2}(s)\varepsilon_{1j_2})^*(W_{1k_2}(t)\varepsilon_{1k_2})^* \right) \\ &= \frac{1}{n} \sum_{i=1}^n W_{ij_1}(s)\varepsilon_{ij_1} W_{ik_1}(t)\varepsilon_{ik_1} W_{ij_2}(s)\varepsilon_{ij_2} W_{ik_2}(t)\varepsilon_{ik_2} - \left(\frac{1}{n} \sum_{i=1}^n W_{ij_1}(s)\varepsilon_{ij_1} W_{ik_1}(t)\varepsilon_{ik_1} \right) \left(\frac{1}{n} \sum_{i=1}^n W_{ij_2}(s)\varepsilon_{ij_2} W_{ik_2}(t)\varepsilon_{ik_2} \right) \end{aligned}$$

Wir bemühen nun das gleiche Vorgehen wie bei Gleichung (10.7), da

$W_{1j_1}(s)\varepsilon_{1j_1}W_{1k_1}(t)\varepsilon_{1k_1}, \dots, W_{nj_1}(s)\varepsilon_{nj_1}W_{nk_1}(t)\varepsilon_{nk_1}$ bzw.

$W_{1j_2}(s)\varepsilon_{1j_2}W_{1k_2}(t)\varepsilon_{1k_2}, \dots, W_{nj_2}(s)\varepsilon_{nj_2}W_{nk_2}(t)\varepsilon_{nk_2}$ durch die verschiedenen Gewichte wieder nicht mehr identisch verteilt sind. Über obige Gleichung folgt für Gleichung (10.16):

$$\begin{aligned}
 \mathbb{E}^*[(S_n^{2*}(s, t))^2] &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s)W_{ik_2}(t)(\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2} - \mathbb{E}[\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2}]) \\
 &\quad - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j, k=1}^p W_{ij}(s)W_{ik}(t)(\varepsilon_{ij}\varepsilon_{ik} - \sigma_{jk}) + \frac{1}{n} \sum_{i=1}^n \sum_{j, k=1}^p W_{ij}(s)W_{ik}(t)\sigma_{jk} \right)^2 \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s)W_{ik_2}(t) \underbrace{(\mathbb{E}[\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2}] - \sigma_{j_1k_1}\sigma_{j_2k_2})}_{\stackrel{\text{iid}}{=} \text{Cov}(\varepsilon_{1j_1}\varepsilon_{1k_1}, \varepsilon_{1j_2}\varepsilon_{1k_2})} \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \left(\sum_{j, k=1}^p W_{ij}(s)W_{ik}(t)\sigma_{jk} \right)^2 \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s)W_{ik_1}(t)W_{ij_2}(s)W_{ik_2}(t)(\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2} - \mathbb{E}[\varepsilon_{ij_1}\varepsilon_{ik_1}\varepsilon_{ij_2}\varepsilon_{ik_2}]) \\
 &\quad - \left(\frac{1}{\sqrt{n}} \cdot S_n^2(s, t) + \frac{1}{n} \sum_{i=1}^n \vec{W}_i(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}_i(t) \right)^2 \\
 &\quad + \frac{1}{n} \sum_{i=1}^n (\vec{W}_i(s) \otimes \vec{W}_i(t))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}_i(s) \otimes \vec{W}_i(t)) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \left(\vec{W}_i(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}_i(t) \right)^2 \tag{10.17}
 \end{aligned}$$

$S_n^2(s, t)$ entspricht wieder dem aus Gleichung (4.32). Die drei hinteren Summanden sind schnell erklärt. Mit (Ä.4), Proposition 7.6 und der gleichen Argumentation wie bei Gleichung (10.8) und folgende gilt:

$$\begin{aligned}
 \left(\frac{1}{\sqrt{n}} \cdot S_n^2(s, t) + \frac{1}{n} \sum_{i=1}^n \vec{W}_i(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}_i(t) \right)^2 &= \left(\mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2}h^2} \right) + \mathcal{O} \left(\frac{1}{p^{1-d_1}h} \right) \right)^2 \\
 &\stackrel{\text{Gleichung (4.29)}}{\leq} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{(p^{2-d_2}h^2)^2} \right) + \mathcal{O} \left(\frac{1}{(p^{1-d_1}h)^2} \right) \\
 \frac{1}{n} \sum_{i=1}^n (\vec{W}_i(s) \otimes \vec{W}_i(t))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}_i(s) \otimes \vec{W}_i(t)) &= \mathcal{O} \left(\frac{1}{p^{2-d_2}h^2} \right) \cdot \frac{1}{n} \sum_{i=1}^n 1 = \mathcal{O} \left(\frac{1}{p^{2-d_2}h^2} \right) \\
 \frac{1}{n} \sum_{i=1}^n \left(\vec{W}_i(s)^T \cdot \text{Cov}(\vec{\varepsilon}) \cdot \vec{W}_i(t) \right)^2 &= \left(\mathcal{O} \left(\frac{1}{p^{1-d_1}h} \right) \cdot \frac{1}{n} \sum_{i=1}^n 1 \right)^2 = \mathcal{O} \left(\frac{1}{(p^{1-d_1}h)^2} \right)
 \end{aligned}$$

$$\begin{aligned} \Rightarrow \mathbb{E}^* \left[(S_n^{2*}(s, t))^2 \right] &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s) W_{ik_2}(t) (\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} - \mathbb{E}[\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}]) \\ &\quad + \mathcal{O}_{\text{f.s.}} \left(\frac{1}{(p^{2-d_2} h^2)^2} \right) + \mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right) + \mathcal{O} \left(\frac{1}{(p^{1-d_1} h)^2} \right) \end{aligned}$$

Der erste Summand wird nun wieder über die Konvergenz im 2-ten Mittel überprüft. Wir sparen uns auch hier wieder einiges an Schreibarbeit und verweisen auf Gleichung (4.32), da die Schritte hier im Prinzip wieder identisch sind, abgesehen von der höheren Dimension und dem Vorfaktor $\frac{1}{n}$ statt $\frac{\sqrt{n}}{n}$.

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s) W_{ik_2}(t) (\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} - \mathbb{E}[\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}]) \right)^2 \right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{\substack{j_1, k_1, j_2, k_2=1 \\ j_3, k_3, j_4, k_4=1}}^p W_{ij_1}(s) W_{ik_1}(t) W_{ij_2}(s) W_{ik_2}(t) W_{ij_3}(s) W_{ik_3}(t) W_{ij_4}(s) W_{ik_4}(t) \\ &\quad \cdot \underbrace{\left(\mathbb{E}[\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2} \varepsilon_{ij_3} \varepsilon_{ik_3} \varepsilon_{ij_4} \varepsilon_{ik_4}] - \mathbb{E}[\varepsilon_{ij_1} \varepsilon_{ik_1} \varepsilon_{ij_2} \varepsilon_{ik_2}] \mathbb{E}[\varepsilon_{ij_3} \varepsilon_{ik_3} \varepsilon_{ij_4} \varepsilon_{ik_4}] \right)}_{\stackrel{\text{iid}}{=} \text{Cov}(\varepsilon_{1j_1} \varepsilon_{1k_1} \varepsilon_{1j_2} \varepsilon_{1k_2}, \varepsilon_{1j_3} \varepsilon_{1k_3} \varepsilon_{1j_4} \varepsilon_{1k_4})} \\ &= \frac{1}{n^2} \sum_{i=1}^n (\vec{W}(s) \otimes \vec{W}(t) \otimes \vec{W}(s) \otimes \vec{W}(t))^T \cdot \text{Cov}(\vec{\varepsilon} \otimes \vec{\varepsilon} \otimes \vec{\varepsilon} \otimes \vec{\varepsilon}) \cdot (\vec{W}(s) \otimes \vec{W}(t) \otimes \vec{W}(s) \otimes \vec{W}(t)) \\ &= \mathcal{O} \left(\frac{1}{p^{4-d_4} h^4} \right) \cdot \frac{1}{n^2} \sum_{i=1}^n 1 = \mathcal{O} \left(\frac{1}{n p^{4-d_4} h^4} \right) \stackrel{n \in \mathbb{N}}{=} \mathcal{O} \left(\frac{1}{p^{4-d_4} h^4} \right) \rightarrow 0 \end{aligned}$$

Das impliziert die stochastische Konvergenz. Durch die Voraussetzung $\sum_{n=1}^{\infty} \frac{1}{p^{4-d_4} h^4} < \infty$ kann aber wieder, äquivalent zu Gleichung (10.9), die f. s. Konvergenz gefolgert werden. Es folgt damit insgesamt für Gleichung (10.18):

$$\begin{aligned} \mathbb{E}^* \left[(S_n^{2*}(s, t))^2 \right] &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{4-d_4} h^4} \right) + \underbrace{\mathcal{O}_{\text{f.s.}} \left(\frac{1}{(p^{2-d_2} h^2)^2} \right)}_{p^{2-d_2} h^2 \rightarrow \infty \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} \right)} + \underbrace{\mathcal{O} \left(\frac{1}{p^{2-d_2} h^2} \right)}_{= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} \right)} + \underbrace{\mathcal{O} \left(\frac{1}{(p^{1-d_1} h)^2} \right)}_{p^{1-d_1} h \rightarrow \infty \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{1-d_1} h} \right)} \\ &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{4-d_4} h^4} + \frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \xrightarrow{\text{f.s.}} 0 \end{aligned} \quad (10.18)$$

$S_n^{2*}(s, t)$ konvergiert somit für alle $(s, t) \in [0, 1]^2$ bedingt auf Y f. s. gegen 0 in \mathbb{P} .

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei

$$\begin{aligned}\Phi^*(s, t) &= \left\| S_n^{2*}(s, t) \right\|_{\ell^2} \\ &= \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2}\end{aligned}$$

Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^{2*}(s, t)| \leq \Phi^*(s, t) \Leftrightarrow |S_n^{2*}| \leq \Phi^*$. Mit Gleichung (10.16), (10.17) und (10.18) folgt weiterhin:

$$\begin{aligned}\mathbb{E}^* [|\Phi^*(x)|^2] &= \mathbb{E}^* \left[\left\| S_n^{2*}(x) \right\|_{\ell^2}^2 \right] \\ &= \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{4-d_4} h^4} + \frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^* [|\Phi^*|^2] &\stackrel{\text{f.s.}}{\leq} C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{4-d_4} h^4}}_{< \infty} + C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2} h^2}}_{< \infty} + C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1} h}}_{< \infty} < \infty \Rightarrow (\mathbb{E}^* [|\Phi^*|^2])^{\frac{1}{2}} \stackrel{\text{f.s.}}{<} \infty\end{aligned}$$

zu ii):

Wir schätzen wieder die Überdeckungszahl bzgl. der kanonischen Metrik ab. Auch hier sparen wir uns Schreibearbeit, denn wegen der bedingte Unabhängigkeit bzgl. i wie bei Gleichung (10.16) folgt mit Proposition 7.6 über die gleiche Vorgehensweise wie bei Gleichung (10.11) für alle $(s, t), (s', t') \in [0, 1]^2$:

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)\varepsilon_{ij})^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right. \right. \\
 &\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s')\varepsilon_{ij})^* (W_{ik}(t')\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s')\varepsilon_{ij})^* (W_{ik}(t')\varepsilon_{ik})^*] \right) \right) \right) \right)^2 \right] \\
 &= \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p \text{Cov}^* \left((W_{1j_1}(s)\varepsilon_{1j_1})^* (W_{1k_1}(t)\varepsilon_{1k_1})^*, (W_{1j_2}(s)\varepsilon_{1j_2})^* (W_{1k_2}(t)\varepsilon_{1k_2})^* \right) \\
 &\quad - 2 \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p \text{Cov}^* \left((W_{1j_1}(s)\varepsilon_{1j_1})^* (W_{1k_1}(t)\varepsilon_{1k_1})^*, (W_{1j_2}(s')\varepsilon_{1j_2})^* (W_{1k_2}(t')\varepsilon_{1k_2})^* \right) \\
 &\quad + \sum_{\substack{j_1, k_1=1 \\ j_2, k_2=1}}^p \text{Cov}^* \left((W_{1j_1}(s')\varepsilon_{1j_1})^* (W_{1k_1}(t')\varepsilon_{1k_1})^*, (W_{1j_2}(s')\varepsilon_{1j_2})^* (W_{1k_2}(t')\varepsilon_{1k_2})^* \right)
 \end{aligned}$$

Die Abschätzung erfolgt nun wieder über Gleichung (10.17) und resultiert in Gleichung (10.18):

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{4-d_4}h^4} + \frac{1}{p^{2-d_2}h^2} + \frac{1}{p^{1-d_1}h} \right) \\
 &\Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} C \left(\frac{1}{p^{4-d_4}h^4} + \frac{1}{p^{2-d_2}h^2} + \frac{1}{p^{1-d_1}h} \right) \\
 &\Rightarrow d_n^*((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \sqrt{C} \left(\frac{1}{p^{4-d_4}h^4} + \frac{1}{p^{2-d_2}h^2} + \frac{1}{p^{1-d_1}h} \right)^{\frac{1}{2}} =: C_n
 \end{aligned}$$

Wegen $C_n \rightarrow 0$ sind wieder alle weiteren Schritte im Grunde identisch zu denen von Gleichung (4.34) und folgende, also ist $S_n^{2*}(s, t)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} .

$$S_n^{3*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right)$$

Wir kombinieren wieder f. s. die gleichmäßige gleichgrade Stetigkeit und punktweise Konvergenz in \mathbb{P} .

f. s. Punktweise Konvergenz in \mathbb{P} :

Mit Proposition 10.2 landen wir wieder über die gleichen Rechenschritte bei Gleichung (6.28), nur mit gebootstrapteten Gewichten. Für alle $(s, t) \in [0, 1]^2$ gilt:

$$\begin{aligned} \mathbb{E}^* [(S_n^{3*}(s, t))^2] &= \mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t)\varepsilon_{1k_2} \right)^* \right] \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right) \\ &\quad + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right] \text{Cov}^* \left(\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^*, \left(\sum_{k_2=1}^p W_{1k_2}(t)\varepsilon_{1k_2} \right)^* \right) \end{aligned} \quad (10.19)$$

Wir betrachten wieder jeden Faktor einzeln, angefangen beim Letzten.

Nach den Argumentationen bei Gleichung (10.7), resultierend in Gleichung (10.10) gilt:

$$\text{Cov}^* \left(\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^*, \left(\sum_{k_2=1}^p W_{1k_2}(t)\varepsilon_{1k_2} \right)^* \right) = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2}h^2} + \frac{1}{p^{1-d_1}h} \right)$$

Nach Proposition 7.7 gilt:

$$\begin{aligned} &\mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right] \\ &= \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_1=1}^p W_{ij_1}(s)Z_i(x_{ij_1}) \right) \right] + o_{\text{f.s.}}(1) \right) \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j_2=1}^p W_{ij_2}(s)Z_i(x_{ij_2}) \right) \right] + o_{\text{f.s.}}(1) \right) \\ &= \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_1=1}^p W_{ij_1}(s) \underbrace{\mathbb{E}[Z_i(x_{ij_1})]}_{=0} + o_{\text{f.s.}}(1) \right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j_2=1}^p W_{ij_2}(s) \underbrace{\mathbb{E}[Z_i(x_{ij_2})]}_{=0} + o_{\text{f.s.}}(1) \right) = o_{\text{f.s.}}(1) \end{aligned}$$

Der zweite Faktor ist nach dem *Verschiebungssatz*:

$$\begin{aligned} &\text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right) \\ &= \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right] - \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right] \end{aligned}$$

Das ist wiederum das Äquivalent zu Gleichung (10.6) nur ausgewertet an den Stellen s und s , statt x und x' . Daher folgt:

$$\text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s) Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s) Z_1(x_{1j_2}) \right)^* \right) = R(s, s) + \mathcal{O}(h^{1+\alpha_R}) + o_{\text{f.s.}}(1)$$

Für den letzten Faktor gilt:

$$\begin{aligned} \mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t) \varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t) \varepsilon_{1k_2} \right)^* \right] &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{k_1=1}^p W_{ik_1}(t) \varepsilon_{ik_1} \right) \left(\sum_{k_2=1}^p W_{ik_2}(t) \varepsilon_{ik_2} \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k_1, k_2=1}^p W_{ik_1}(t) W_{ik_2}(t) (\varepsilon_{ik_1} \varepsilon_{ik_2} - \sigma_{k_1 k_2}) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{k_1, k_2=1}^p W_{ik_1}(t) W_{ik_2}(t) \sigma_{k_1 k_2} \end{aligned}$$

Das entspricht genau Gleichung (10.8), nur ohne den Summanden $\left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p W_{ij}(x) \varepsilon_{ij} \right)^2$,

wir können also wieder über Gleichung (10.10) auf

$$\mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t) \varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t) \varepsilon_{1k_2} \right)^* \right] = \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right)$$

schließen. Insgesamt folgt für Gleichung (10.19):

$$\begin{aligned} \mathbb{E}^* \left[(S_n^{3*}(s, t))^2 \right] &= \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) \cdot (R(s, s) + \mathcal{O}(h^{1+\alpha_R}) + o_{\text{f.s.}}(1)) \\ &\quad + o_{\text{f.s.}}(1) \cdot \mathcal{O}_{\text{f.s.}} \left(\frac{1}{p^{2-d_2} h^2} + \frac{1}{p^{1-d_1} h} \right) = \mathcal{O}_{\text{f.s.}} \left(\frac{1 + h^{1+\alpha_R}}{p^{2-d_2} h^2} + \frac{1 + h^{1+\alpha_R}}{p^{1-d_1} h} \right) \xrightarrow{\text{f.s.}} 0 \end{aligned} \tag{10.20}$$

$S_n^{3*}(s, t)$ konvergiert somit für alle $(s, t) \in [0, 1]^2$ bedingt auf Y f. s. gegen 0 in \mathbb{P} .

f. s. gleichmäßige gleichgradige Stetigkeit in \mathbb{P} (über Lemma 2.19)

zu i):

Sei

$$\begin{aligned} \Phi^*(s, t) &= \|S_n^{3*}(s, t)\|_{\ell^2} \\ &= \sqrt{\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^*[(W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2} \end{aligned}$$

Dann gilt offensichtlich f. s. für alle $n \in \mathbb{N}$ und $(s, t) \in [0, 1]^2$: $|S_n^{3*}(s, t)| \leq \Phi^*(s, t) \Leftrightarrow |S_n^{3*}| \leq \Phi^*$. Mit Gleichung (10.19) und (10.20) folgt weiterhin:

$$\begin{aligned} \mathbb{E}^*[\|\Phi^*(x)\|^2] &= \mathbb{E}^*[\|S_n^{3*}(x)\|_{\ell^2}^2] \\ &= \mathbb{E}^* \left[\sum_{n=1}^{\infty} \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^*[(W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^*[(W_{ij}(s)Z_i(x_{ij}))^*(W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right)^2 \right] \\ &= \sum_{n=1}^{\infty} \mathcal{O}_{\text{f.s.}} \left(\frac{1+h^{1+\alpha_R}}{p^{2-d_2}h^2} + \frac{1+h^{1+\alpha_R}}{p^{1-d_1}h} \right) \\ \Rightarrow \exists C \in \mathbb{R}^+ : \mathbb{E}^*[\|\Phi^*\|^2] &\stackrel{\text{f.s.}}{\leq} C \sum_{n=1}^{\infty} \frac{1+h^{1+\alpha_R}}{p^{2-d_2}h^2} + C \sum_{n=1}^{\infty} \frac{1+h^{1+\alpha_R}}{p^{1-d_1}h} \stackrel{h \leq 1}{\leq} 2C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2}}_{< \infty} + 2C \underbrace{\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h}}_{< \infty} < \infty \\ \Rightarrow (\mathbb{E}^*[\|\Phi^*\|^2])^{\frac{1}{2}} &\stackrel{\text{f.s.}}{<} \infty \end{aligned}$$

zu ii):

Wir schätzen wieder die Überdeckungszahl bzgl. der kanonischen Metrik ab. Mit Proposition 7.6 und Gleichung (10.19) folgt über die gleiche Vorgehensweise wie bei Gleichung (10.11) für alle $(s, t), (s', t') \in [0, 1]^2$:

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathbb{E}^* \left[\left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s)Z_i(x_{ij}))^* (W_{ik}(t)\varepsilon_{ik})^*] \right) \right) \right. \right. \\
 &\quad \left. \left. - \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(s')Z_i(x_{ij}))^* (W_{ik}(t')\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(s')Z_i(x_{ij}))^* (W_{ik}(t')\varepsilon_{ik})^*] \right) \right) \right) \right)^2 \right] \\
 &= \mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t)\varepsilon_{1k_2} \right)^* \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right) \right. \\
 &\quad \left. + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s)Z_1(x_{1j_2}) \right)^* \right] \text{Cov}^* \left(\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^*, \left(\sum_{k_2=1}^p W_{1k_2}(t)\varepsilon_{1k_2} \right)^* \right) \right. \right. \\
 &\quad \left. - 2\mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t')\varepsilon_{1k_2} \right)^* \right] \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s')Z_1(x_{1j_2}) \right)^* \right) \right. \\
 &\quad \left. - 2\mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s)Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s')Z_1(x_{1j_2}) \right)^* \right] \text{Cov}^* \left(\left(\sum_{k_1=1}^p W_{1k_1}(t)\varepsilon_{1k_1} \right)^*, \left(\sum_{k_2=1}^p W_{1k_2}(t')\varepsilon_{1k_2} \right)^* \right) \right. \\
 &\quad \left. + \mathbb{E}^* \left[\left(\sum_{k_1=1}^p W_{1k_1}(t')\varepsilon_{1k_1} \right)^* \left(\sum_{k_2=1}^p W_{1k_2}(t')\varepsilon_{1k_2} \right)^* \right] \text{Cov}^* \left(\left(\sum_{j_1=1}^p W_{1j_1}(s')Z_1(x_{1j_1}) \right)^*, \left(\sum_{j_2=1}^p W_{1j_2}(s')Z_1(x_{1j_2}) \right)^* \right) \right. \\
 &\quad \left. + \mathbb{E}^* \left[\left(\sum_{j_1=1}^p W_{1j_1}(s')Z_1(x_{1j_1}) \right)^* \right] \mathbb{E}^* \left[\left(\sum_{j_2=1}^p W_{1j_2}(s')Z_1(x_{1j_2}) \right)^* \right] \text{Cov}^* \left(\left(\sum_{k_1=1}^p W_{1k_1}(t')\varepsilon_{1k_1} \right)^*, \left(\sum_{k_2=1}^p W_{1k_2}(t')\varepsilon_{1k_2} \right)^* \right) \right]
 \end{aligned}$$

Die Abschätzung erfolgt wieder wie bei Gleichung (10.19) und resultiert in Gleichung (10.20):

$$\begin{aligned}
 d_n^{*2}((s, t), (s', t')) &= \mathcal{O}_{\text{f.s.}} \left(\frac{1 + h^{1+\alpha_R}}{p^{2-d_2}h^2} + \frac{1 + h^{1+\alpha_R}}{p^{1-d_1}h} \right) \\
 &\Rightarrow \exists C \in \mathbb{R}^+ : d_n^{*2}((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} C \left(\frac{1 + h^{1+\alpha_R}}{p^{2-d_2}h^2} + \frac{1 + h^{1+\alpha_R}}{p^{1-d_1}h} \right) \\
 &\Rightarrow d_n^*((s, t), (s', t')) \stackrel{\text{f.s.}}{\leq} \sqrt{C} \left(\frac{1 + h^{1+\alpha_R}}{p^{2-d_2}h^2} + \frac{1 + h^{1+\alpha_R}}{p^{1-d_1}h} \right)^{\frac{1}{2}} =: C_n
 \end{aligned}$$

Wegen $C_n \rightarrow 0$ sind wieder alle weiteren Schritte im Grunde identisch zu denen von Gleichung (4.34) und folgende, also ist $S_n^{3*}(s, t)$ bedingt auf Y f. s. gleichmäßig gleichgradig stetig in \mathbb{P} .

$$S_n^{4*}(s, t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^p \left((W_{ij}(t)Z_i(x_{ij}))^* (W_{ik}(s)\varepsilon_{ik})^* - \mathbb{E}^* [(W_{ij}(t)Z_i(x_{ij}))^* (W_{ik}(s)\varepsilon_{ik})^*] \right) \right)$$

Dieser Summand lässt sich vollkommen analog zu $S_n^{3*}(s, t)$ diskutieren. \square

11. Simulationen

Wir werden nun die Gültigkeit der Konfidenzbänder für R nach Satz 5.7 und 6.5 anhand mehrerer Datensettings überprüfen. Eine Simulationsstudie zu den Bändern für μ wurde bereits in [Deg11] durchgeführt. Der Fall der asynchronen Messpunkte wird hier nicht betrachtet. Nichtsdestotrotz suggerieren weitere, etwas weniger umfangreiche Simulationen (für μ), dass auch hier die Konfidenzbänder korrekt sind, wenn auch mit einer leicht geringeren Überdeckungswahrscheinlichkeit. Darauf werden wir etwas weiter unten noch einmal kurz eingehen. Als konkreter Schätzer für $\hat{\mu}$ wird der lokal lineare Schätzer aus Abschnitt 3.4.2 verwendet.

11.1. Settings und Aufbau der Simulationen

Wir betrachten insgesamt vier verschiedene Datensätze, welche sich wie folgt zusammensetzen:

$$(S.1) \quad \begin{cases} Y_{ij} = \mu(x_j) + Z_i(x_j), & i = 1, \dots, n, \quad j = 1, \dots, p \\ \mu(x) = 6x^5 - 15x^4 + 10x^3 \\ Z \sim \mathcal{G}(0, R), \quad R(s, t) = (0.25)^2 \cdot \exp\{20 \log(0.9)(s - t)^2\} \end{cases}$$

$$(S.2) \quad \begin{cases} Y_{ij} = \mu(x_j) + Z_i(x_j) + \varepsilon_{ij}, & i = 1, \dots, n, \quad j = 1, \dots, p \\ \mu(x) = \sin(8\pi x) \exp\{-3x\} \\ Z(x) = \frac{\sqrt{2}}{6}(\eta_1 - 1) \sin(\pi x) + \frac{2}{3}(\eta_2 - 1)(x - 0.5), & \eta_1 \sim \chi_1^2, \quad \eta_2 \sim \text{Exp}(1) \\ \varepsilon_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.01), & i = 1, \dots, n, \quad j = 1, \dots, p \end{cases}$$

$$(S.3) \quad \begin{cases} Y_{ij} = \mu(x_j) + Z_i(x_j) + \varepsilon_{ij}, & i = 1, \dots, n, \quad j = 1, \dots, p \\ \mu(x) = \sin(8\pi x) \exp\{-3x\} \\ Z(x) = \frac{\sqrt{2}}{6}(\eta_1 - 1) \sin(10\pi x) + \frac{2}{3}(\eta_2 - 1)(x - 1), & \eta_1 \sim \chi_1^2, \quad \eta_2 \sim \text{Exp}(1) \\ \varepsilon_{ij} \stackrel{\text{iid}}{\sim} \mathcal{U}_{[-0.1, 0.1]}, & i = 1, \dots, n, \quad j = 1, \dots, p \end{cases}$$

$$(S.4) \quad \begin{cases} Y_{ij} = \mu(x_j) + Z_i(x_j) + \varepsilon_{ij}, & i = 1, \dots, n, \quad j = 1, \dots, p \\ \mu(x) = 6x^5 - 15x^4 + 10x^3 \\ Z \sim \mathcal{G}(0, R), \quad R(s, t) = (0.25)^2 \cdot \exp\{20 \log(0.9)(s - t)^2\} \\ \vec{\varepsilon}_1, \dots, \vec{\varepsilon}_n \stackrel{\text{iid}}{\sim} \mathcal{N}_p(0, \text{Cov}(\vec{\varepsilon}_1)), \quad \text{Cov}(\vec{\varepsilon}_1) = (\sigma_{jk})_{j,k=1,\dots,p} \text{ mit } \sigma_{jk} = \frac{p-|j-k|}{10p} \end{cases}$$

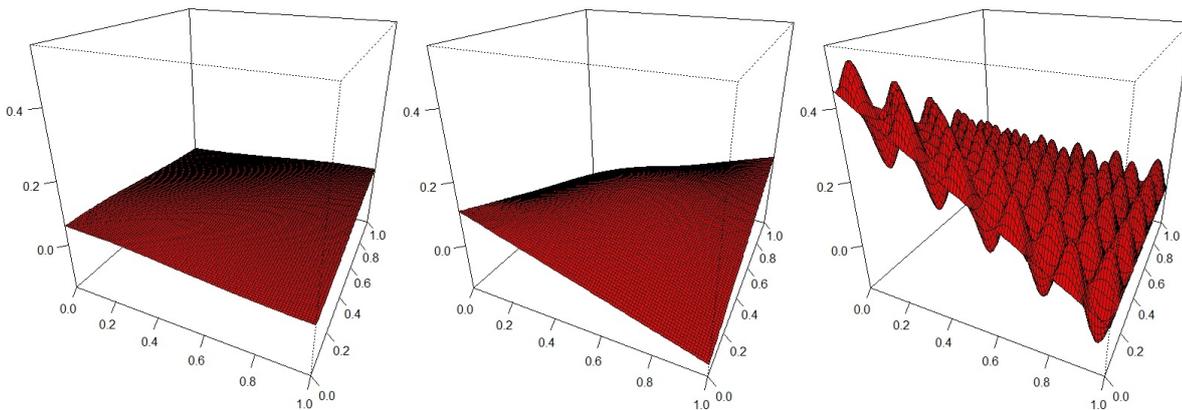
(S.1) ist ein sehr angenehmes Setting, da μ ein Polynom ist, Z einem Gauß-Prozess mit sehr flacher Kovarianzstruktur folgt und das Modell gänzlich ohne Fehlerterm ε auskommt. (S.2) und (S.3) sind deutlich komplexere Settings, denn μ ist kein Polynom und Z ist durch eine spezifische Funktion über zwei weitere Zufallsvariablen definiert. Explizit berechnet, gilt für die Kovarianzfunktion R in diesen beiden Fällen:

$$(S.2) : R(s, t) = \frac{1}{9} \sin(\pi s) \sin(\pi t) + \frac{4}{9}(s - 0.5)(t - 0.5)$$

$$(S.3) : R(s, t) = \frac{1}{9} \sin(10\pi s) \sin(10\pi t) + \frac{4}{9}(s - 1)(t - 1)$$

(S.4) entspricht im Prinzip genau (S.1) nur dass hier zusätzlich ein Fehlerterm existiert. Dieser folgt einer zentrierten Normalverteilung, also eigentlich relativ unproblematisch. Dennoch werden wir sehen, dass mit diesem Fehler eine präzise Bestimmung der Konfidenzbänder nicht möglich ist, da eine Abhängigkeitsstruktur bzgl. der Beobachtungspunkte vorliegt, welche (A.4) nicht erfüllt, siehe Anhang A.2. Abbildung 11.1 zeigt die Graphen der Kovarianzfunktionen, wobei deutlich erkennbar ist, dass der Verlauf der Funktionen von Setting zu Setting vertrackter wird.

Abbildung 11.1.: Plot der Kovarianzfunktionen R aus den Settings (S.1), (S.2) und (S.3).



Die Konfidenzbänder wurden insgesamt über sechs verschiedene Verfahren berechnet: Mit dem nicht standardisierten empirischen Prozess nach Satz 5.7 und den beiden Bootstrapversionen nach Satz 6.5, und dann noch einmal mit einem komplett standardisierten empirischen Prozess nach Bemerkung 5.8 und 6.6. Dabei wurden jeweils verschiedene Kombinationen der Stichprobengröße n , der Anzahl der Beobachtungspunkte p und der Bandbreite h betrachtet, wobei die Beobachtungspunkte x_j , $j = 1, \dots, p$ durch die Vorschrift $x_j = \frac{j - 0.5}{p}$ erzeugt wurden, welche (A.5) per Definition erfüllen.

Bemerkung 11.1.

Eine Möglichkeit asynchrone Messpunkte zu erzeugen, wäre z. B.:

$$x_{ij} = \frac{j - (1 - \frac{i-1}{n})}{p}, \quad i = 1, \dots, \frac{n}{2}, \quad x_{ij} = \frac{j - (1 - \frac{i}{n})}{p}, \quad i = \frac{n}{2} + 1, \dots, n$$

falls n gerade und

$$x_{ij} = \frac{j - (1 - \frac{i-1}{n-1})}{p}, \quad i = 1, \dots, \frac{n+1}{2}, \quad x_{ij} = \frac{j - (1 - \frac{i}{n-1})}{p}, \quad i = \frac{n+1}{2} + 1, \dots, n$$

falls n ungerade. Das sind im Prinzip die obigen Messpunkte, nur in jeder Stichprobeneinheit um einen gewissen Betrag, abhängig von n , nach rechts bzw. links verschoben, wodurch (Ä.5) wieder per Definition erfüllt ist.

Die Schranke $\hat{c}_{\gamma,R}$ wurde mittels einer *Monte Carlo Simulation* ermittelt. Dafür wurde $\hat{\mathcal{R}}$ an insgesamt 10000 Stellen berechnet und in einer Matrix $\hat{\Sigma} \in \mathbb{R}^{100 \times 100}$ abgespeichert. Schließlich wurde 500 mal ein Zufallsvektor $\vec{X}_n \sim \mathcal{N}_{100}(0, \hat{\Sigma}) \mid Y$ erzeugt, der jeweils betragsmäßig maximale Wert abgespeichert und aus diesen 500 Werten das empirische 95%-Quantil ermittelt. Das Berechnungsschema von $\hat{\Sigma}$ und \vec{X}_n entspricht dem von Gleichung (5.4) und folgende, wobei $s_j = \frac{j-1}{9}$ bzw. $t_k = \frac{k-1}{9}$ mit $j, k = 1, \dots, 10$. $c_{\gamma,R}^*$ wurde bzgl. 100 Stellen $(s_j, t_k) \in [0, 1]^2$, $j, k = 1, \dots, 10$ jeweils mit $N = 5000$ Bootstrap-Wiederholungen berechnet, s_j und t_k definiert wie oben.

Die Bänder wurden jeweils insgesamt 1000 mal ermittelt und an 2500 Stellen $(s_j, t_k) \in [0, 1]^2$ ausgewertet, wobei $s_j = \frac{j-1}{49}$ bzw. $t_k = \frac{k-1}{49}$ mit $j, k = 1, \dots, 50$.

Wir werden nun die beiden Fälle einer großen und kleinen Stichprobe separat diskutieren, da genau genommen nur bei letzterem die Bänder nach dem Bootstrapverfahren wirklich relevant sind. Um die Ergebnisse besser interpretieren zu können, betrachten wir zunächst immer einige Plots der wahren Kovarianzfunktion und der Bänder, und gehen danach erst auf die jeweiligen Überdeckungen ein. Da man in einem 3D-Plot mit insgesamt drei verschiedenen Bändern so gut wie nichts mehr erkennen kann, betrachten wir t -Schnitte an verschiedenen Stellen.

11.2. Fall (hinreichend) große Stichprobe

Wir legen eine Stichprobe ab der Größe $n = 50$ als „hinreichend gut“ fest, da sich ab hier auch für die komplexeren Settings relativ passable Ergebnisse abzeichnen. Abbildung 11.2, 11.3, 11.4 und 11.5 zeigen jeweils für $n = p = 100$ und $h = 0.05$ die Verläufe für die verschiedenen Settings, sowohl berechnet nach dem nicht standardisierten, als auch nach dem standardisierten empirischen Prozess.

Abbildung 11.2.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.1).

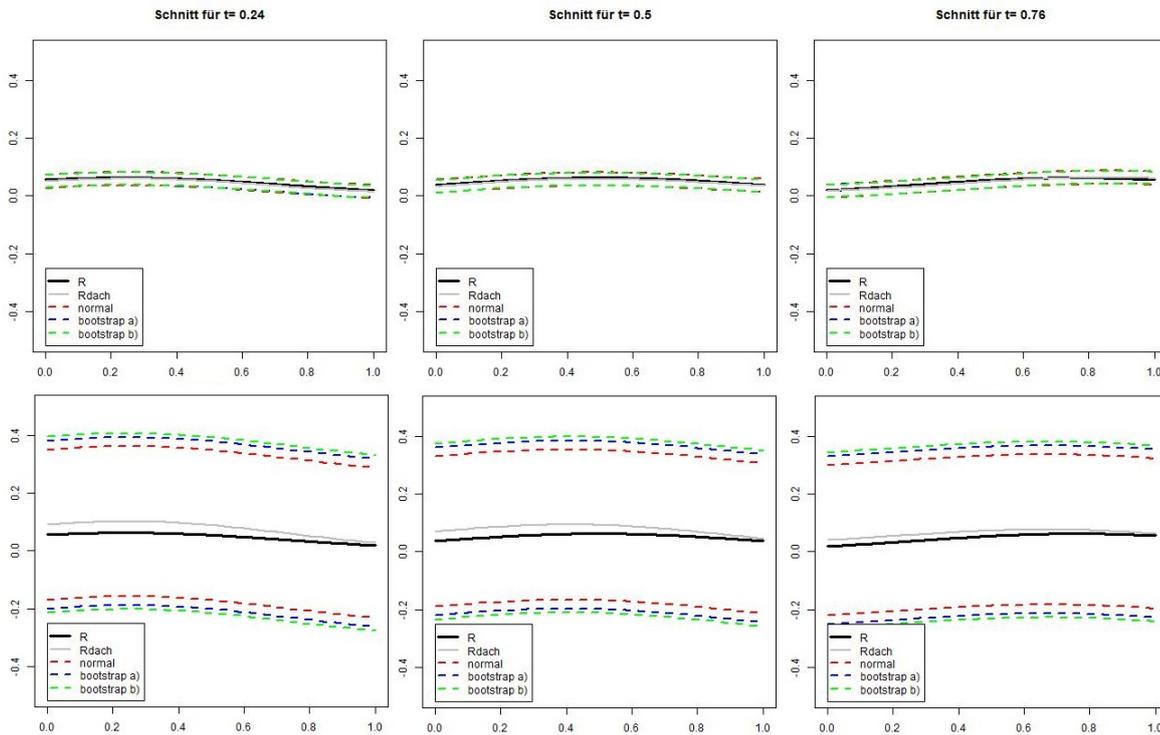


Abbildung 11.3.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.2).

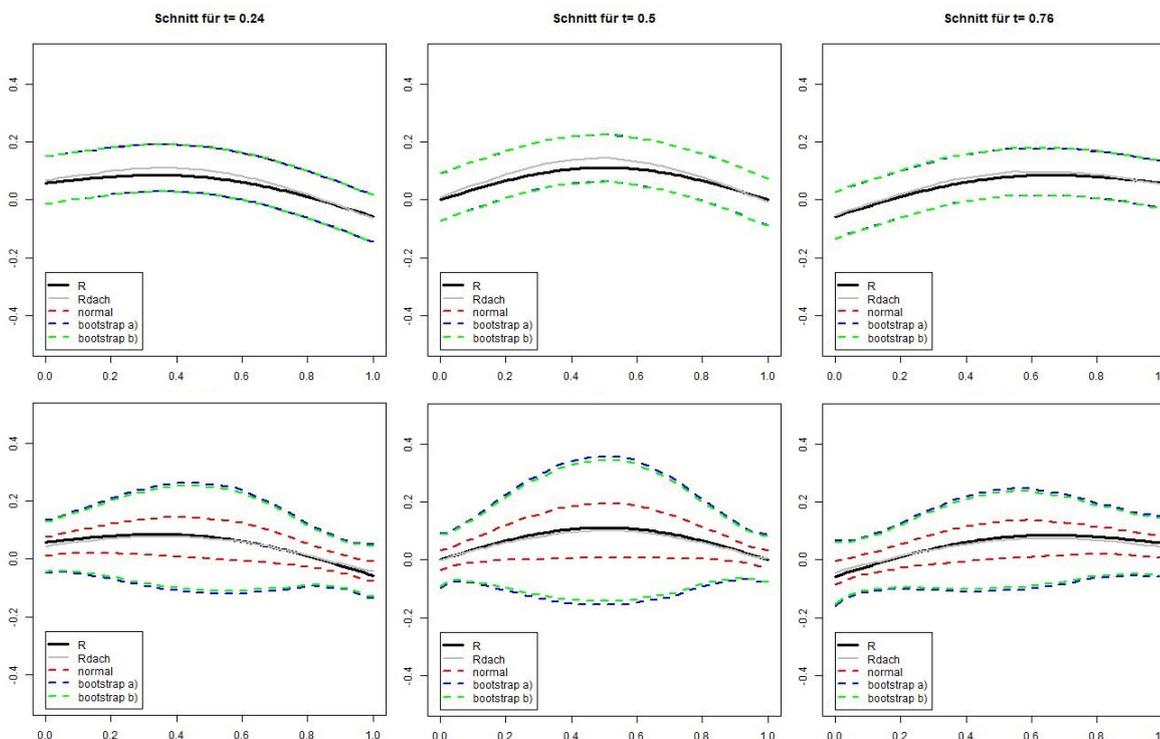


Abbildung 11.4.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.3).

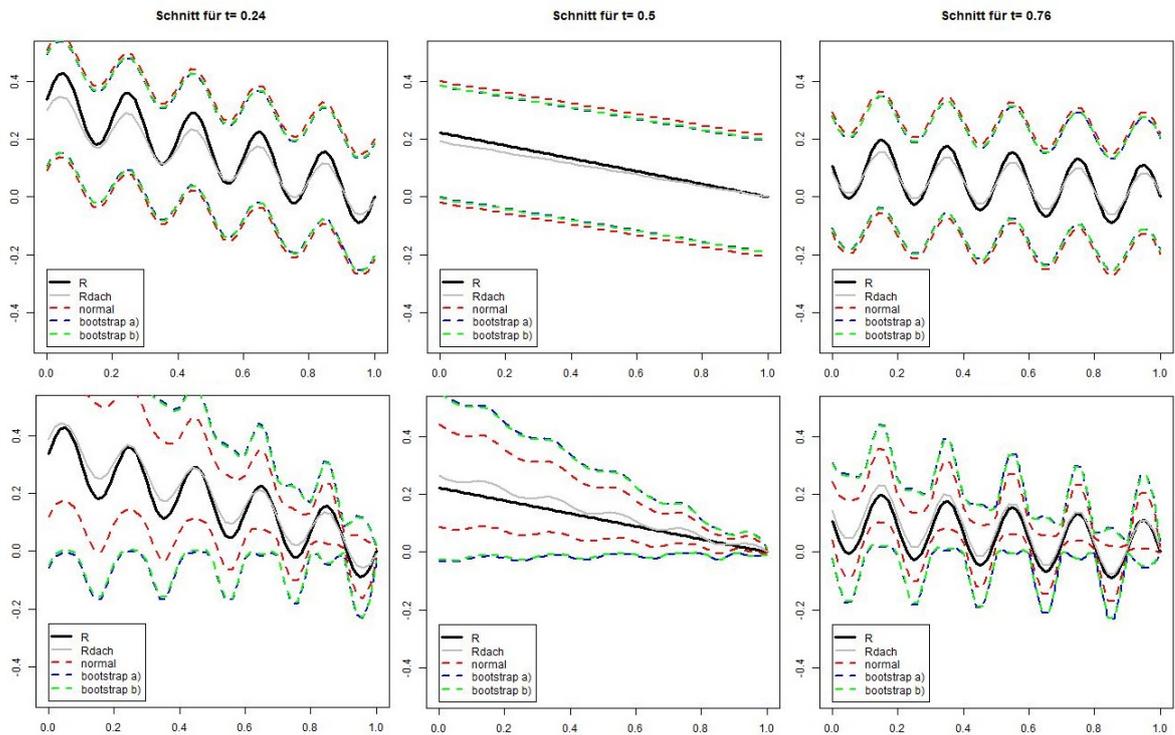
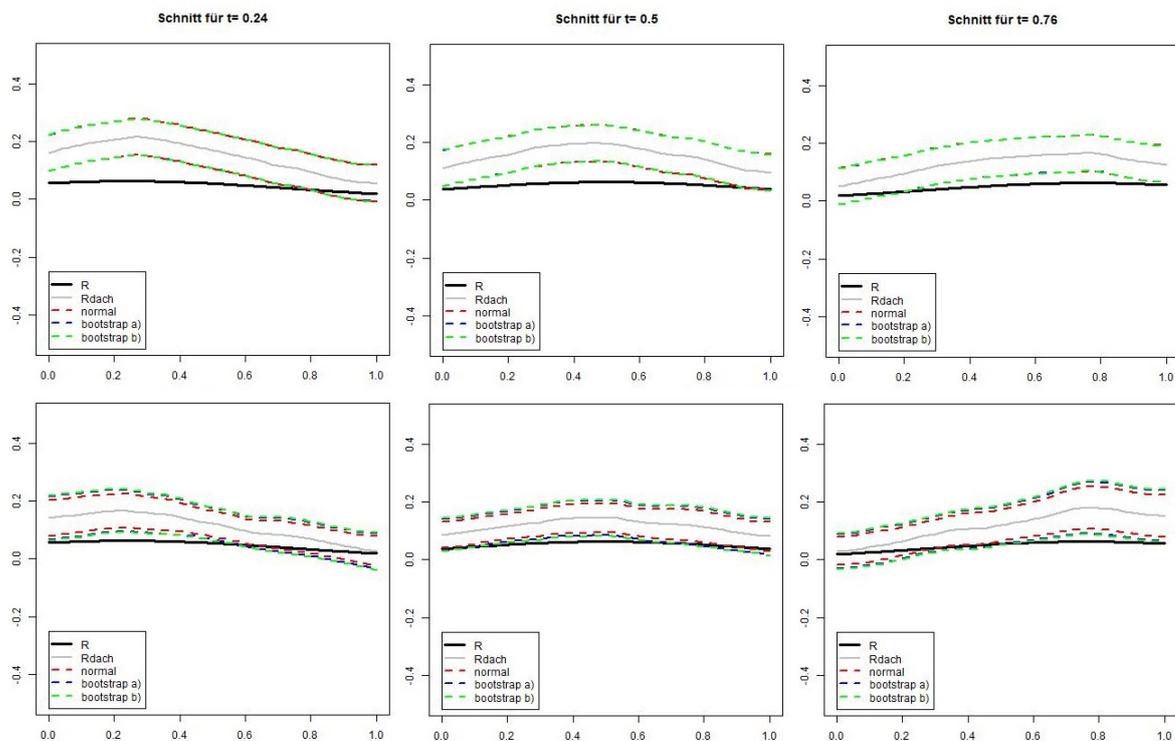


Abbildung 11.5.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.4).



Eines ist in allen Grafiken sofort ersichtlich: Berechnet man die Bänder nicht standardisiert, so spielt es keine Rolle, ob man die Bänder normal oder gebootstrapt bildet, da sie nahezu übereinstimmen. Erst nach der Standardisierung sieht man in den Bändern markante Unterschiede. Um diesen Unterschied zu erklären, werfen wir noch einmal einen Blick auf die Berechnungsweise der Schranken $\hat{c}_{\gamma,R}$ und $c_{\gamma,R}^*$. Die Schranken sind die empirischen 95%-Quantile von $\|\mathcal{G}(0, \hat{\mathcal{R}})\|_\infty$ und $\sqrt{n}\|R^* - \hat{R}\|_\infty$ bzw. $\|\mathcal{G}(0, \hat{\rho}_{ZZ})\|_\infty$ und $\sqrt{n}\|\frac{R^* - \hat{R}}{\sigma_{R^*}}\|_\infty$. Tabelle 11.1 zeigt für Setting (S.1) und verschiedene n die Mittelwerte von 500 erzeugten Zufallsvariablen, welche den Suprema der Gauß-Felder folgen, und die Mittelwerte der Bootstrap Suprema für 5000 Wiederholungen, Supremum wieder jeweils gebildet über 100 Stellen und Schätzer berechnet mit $p = 100$ und $h = 0.05$. Wie man sieht, sind die Mittelwerte für $\|\mathcal{G}(0, \hat{\mathcal{R}})\|_\infty$ und $\sqrt{n}\|R^* - \hat{R}\|_\infty$ durchweg nahezu identisch, wohingegen die für $\sqrt{n}\|\frac{R^* - \hat{R}}{\sigma_{R^*}}\|_\infty$ bei kleinem n größer als die für $\|\mathcal{G}(0, \hat{\rho}_{ZZ})\|_\infty$ sind. Das liegt schlichtweg daran, dass σ_{R^*} in der Regel deutlich unterschätzt und dadurch der Bruch vergrößert wird. Damit gilt natürlich in aller Regel auch $\hat{c}_{\gamma,R} \approx c_{\gamma,R}^*$ im nicht standardisierten Fall und $\hat{c}_{\gamma,R} < c_{\gamma,R}^*$ im standardisierten Fall, was zur Folge hat, dass die Spannweite der standardisiert gebootstrapt Bänder viel größer ist.

Tabelle 11.1.: Mittelwerte 500 Monte Carlo simulierter Zufallsvariablen (Spalte 1 und 4) und Bootstrap simulierter Zufallsvariablen mit 5000 Wiederholungen (Spalte 2, 3, 5 und 6) bei ein und derselben Stichprobe Y , erzeugt nach (S.1).

n	$\ \mathcal{G}(0, \hat{\mathcal{R}})\ _\infty Y$	$\sqrt{n}\ R^* - \hat{R}\ _\infty Y$	$\ \mathcal{G}(0, \hat{\rho}_{ZZ})\ _\infty Y$	$\sqrt{n}\ \frac{R^* - \hat{R}}{\sigma_{R^*}}\ _\infty Y$	
		a)	b)	a)	b)
10	0.0969	0.1012	0.1063	1.4543	6.5253 7.1453
20	0.1112	0.1154	0.1140	1.5811	2.2564 2.7877
50	0.1062	0.1093	0.1093	1.5848	1.8471 1.9481
100	0.1297	0.1281	0.1298	1.4895	1.6664 1.6986

Dieser Umstand bietet allerdings nicht nur Vorteile. Während in Abbildung 11.2 die Überdeckung nahezu perfekt ist, sieht man bereits in Abbildung 11.3 eine deutlich stärkere Schwankung im Verlauf. In Abbildung 11.4 sind die standardisierten Bänder nicht mehr unbedingt präzise und überdecken die wahre Kovarianz immer seltener - vor allem bei den normal berechneten. Je komplexer die wahre Kovarianz wird, desto fehleranfälliger werden die standardisiert berechneten Bänder. Hat man also eine verhältnismäßig große Stichprobe, so sind auf jeden Fall die nicht standardisierten Bänder vorzuziehen. Ob diese normal oder über den Bootstrap berechnet werden spielt durch die immer stärkere Annäherung letztlich nicht wirklich eine Rolle. Tabellen 11.2, 11.3 und 11.4 bestätigen diese Aussagen.

Darüber hinaus zeigt Abbildung 11.5, dass (A.4) wesentlich für die Qualität der Bänder ist, weil der Fehler die Überdeckung des einfachsten Settings vollständig zerstört, sehr schön auch noch einmal zu sehen in Tabelle 11.5.

Tabelle 11.2.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.1).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert		50	50	0.05	0.941	0.2227	0.941	0.2242	0.938	0.2200
		50	100	0.05	0.942	0.2244	0.952	0.2257	0.945	0.2218
		50	100	0.1	0.927	0.2239	0.933	0.2252	0.926	0.2214
		100	10	0.2	0.937	0.2221	0.938	0.2232	0.93	0.2210
		100	20	0.1	0.947	0.2220	0.95	0.2233	0.949	0.2212
		100	20	0.15	0.932	0.2211	0.942	0.2227	0.939	0.2205
		100	20	0.2	0.926	0.2191	0.933	0.2204	0.927	0.2183
		100	50	0.05	0.947	0.2223	0.951	0.2244	0.946	0.2221
		100	100	0.05	0.939	0.2227	0.937	0.2235	0.936	0.2214
		100	100	0.1	0.945	0.2228	0.95	0.2239	0.947	0.2217
		100	100	0.15	0.954	0.2211	0.96	0.2225	0.957	0.2204
	standardisiert		50	50	0.05	0.814	2.6476	0.933	3.7371	0.945
		50	100	0.05	0.824	2.6430	0.945	3.7309	0.963	3.9681
		50	100	0.1	0.831	2.6421	0.941	3.7269	0.959	3.9734
		100	10	0.2	0.883	2.6361	0.946	3.1822	0.949	3.2875
		100	20	0.1	0.887	2.6431	0.951	3.1822	0.957	3.2898
		100	20	0.15	0.886	2.6385	0.936	3.1836	0.948	3.2922
		100	20	0.2	0.855	2.6390	0.923	3.1648	0.93	3.2724
		100	50	0.05	0.897	2.6491	0.948	3.1834	0.957	3.2906
		100	100	0.05	0.897	2.6506	0.949	3.1910	0.955	3.2980
		100	100	0.1	0.887	2.6513	0.94	3.1776	0.946	3.2817
		100	100	0.15	0.875	2.6449	0.933	3.1709	0.939	3.2771

Tabelle 11.3.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.2).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert		50	50	0.05	0.87	0.8097	0.867	0.8047	0.86	0.7778
		50	100	0.05	0.871	0.7791	0.871	0.7768	0.864	0.7512
		50	100	0.1	0.882	0.7783	0.877	0.7710	0.872	0.7466
		100	10	0.2	0.89	0.8496	0.887	0.8397	0.885	0.8263
		100	20	0.1	0.922	0.8468	0.921	0.8392	0.917	0.8249
		100	20	0.15	0.896	0.8167	0.89	0.8089	0.887	0.7948
		100	20	0.2	0.891	0.8087	0.897	0.7996	0.89	0.7867
		100	50	0.05	0.903	0.8418	0.907	0.8315	0.9	0.8183
		100	100	0.05	0.914	0.8221	0.915	0.8114	0.909	0.7984
		100	100	0.1	0.899	0.7957	0.893	0.7862	0.889	0.7732
		100	100	0.15	0.915	0.8158	0.914	0.8057	0.911	0.7934
	standardisiert		50	50	0.05	0.491	2.8093	0.894	9.7150	0.888
		50	100	0.05	0.544	2.7614	0.886	10.0051	0.876	9.3973
		50	100	0.1	0.519	2.7307	0.899	10.0320	0.892	9.3343
		100	10	0.2	0.594	2.7947	0.884	6.1867	0.874	5.9046
		100	20	0.1	0.657	2.8100	0.903	6.4837	0.9	6.1732
		100	20	0.15	0.661	2.7779	0.917	6.3022	0.909	5.9867
		100	20	0.2	0.63	2.7602	0.891	6.2888	0.882	5.9570
		100	50	0.05	0.625	2.8139	0.901	6.5009	0.894	6.1551
		100	100	0.05	0.634	2.7640	0.909	6.4194	0.901	6.0705
		100	100	0.1	0.681	2.7402	0.924	6.3924	0.921	6.0072
		100	100	0.15	0.647	2.7276	0.903	6.1870	0.898	5.8409

Tabelle 11.4.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.3).

		normal		bootstrap a)		bootstrap b)				
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert		50	50	0.05	0.728	2.0294	0.72	1.9926	0.715	1.9529
		50	100	0.05	0.72	2.0443	0.721	2.0148	0.714	1.9745
		50	100	0.025	0.762	2.0165	0.76	1.9842	0.753	1.9442
		100	10	0.2	0.147	4.1124	0.125	4.0706	0.112	4.0180
		100	20	0.1	0.711	2.3585	0.713	2.3170	0.712	2.2983
		100	20	0.08	0.714	2.2783	0.722	2.2471	0.718	2.2300
		100	50	0.05	0.807	2.2346	0.8	2.2018	0.796	2.1873
		100	50	0.035	0.795	2.2107	0.796	2.1780	0.794	2.1595
		100	100	0.05	0.762	2.1456	0.754	2.1054	0.754	2.0916
		100	100	0.035	0.791	2.1770	0.792	2.1372	0.788	2.1253
		100	100	0.025	0.828	2.1990	0.827	2.1629	0.829	2.1445
		100	100	0.02	0.823	2.2341	0.821	2.2018	0.822	2.1802
standardisiert		50	50	0.05	0	2.7715	0.369	7.1665	0.338	7.0063
		50	100	0.05	0.001	2.7372	0.492	7.1852	0.469	6.9760
		50	100	0.025	0	2.7932	0.409	7.4836	0.375	7.2333
		100	10	0.2	0	2.6796	0.009	6.0850	0.004	5.7386
		100	20	0.1	0	2.6554	0.005	5.5729	0.005	5.3678
		100	20	0.08	0	2.6856	0.064	5.7571	0.055	5.5373
		100	50	0.05	0	2.7778	0.029	5.1311	0.021	4.9583
		100	50	0.035	0	2.8059	0.026	5.3287	0.023	5.1491
		100	100	0.05	0	2.7490	0.127	5.3333	0.112	5.1410
		100	100	0.035	0	2.7898	0.018	5.2564	0.015	5.0545
		100	100	0.025	0	2.8019	0.037	5.5260	0.025	5.3094
		100	100	0.02	0	2.8114	0.032	5.4328	0.021	5.2250

Tabelle 11.5.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.4).

		normal		bootstrap a)		bootstrap b)				
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert		50	50	0.05	0.005	0.6176	0.005	0.6232	0.003	0.6106
		50	100	0.05	0.004	0.6137	0.001	0.6190	0.001	0.6062
		50	100	0.1	0.005	0.6025	0.003	0.6073	0.002	0.5951
		100	10	0.2	0	0.6133	0	0.6190	0	0.6127
		100	20	0.1	0	0.6061	0	0.6112	0	0.6045
		100	20	0.15	0	0.5944	0	0.5998	0	0.5933
		100	20	0.2	0	0.5865	0	0.5914	0	0.5852
		100	50	0.05	0	0.6112	0	0.6155	0	0.6091
		100	100	0.05	0	0.6082	0	0.6123	0	0.6055
		100	100	0.1	0	0.6082	0	0.6067	0	0.6003
		100	100	0.15	0	0.5889	0	0.5929	0	0.5864
	standardisiert		50	50	0.05	0.011	2.8170	0.554	4.1620	0.671
		50	100	0.05	0.012	2.8198	0.559	4.1604	0.677	4.4236
		50	100	0.1	0.01	2.7838	0.562	4.0962	0.662	4.3556
		100	10	0.2	0	2.7513	0.001	3.3563	0.001	3.4640
		100	20	0.1	0	2.8027	0.001	3.4367	0.004	3.5505
		100	20	0.15	0	2.7636	0.001	3.3702	0.001	3.4802
		100	20	0.2	0	2.7363	0.003	3.3238	0.007	3.4327
		100	50	0.05	0	2.8198	0.004	3.4799	0.006	3.5918
		100	100	0.05	0	2.8209	0.004	3.4785	0.006	3.5970
		100	100	0.1	0	2.7886	0.002	3.4144	0.004	3.5271
		100	100	0.15	0	2.7546	0.01	3.3746	0.013	3.4867

11.3. Fall kleine Stichprobe

Abbildung 11.6, 11.7, 11.8 und 11.9 zeigen wieder die Verläufe für die verschiedenen Settings, dieses mal für $n = 10$, $p = 100$ und $h = 0.05$.

Auch hier unterscheiden sich im nicht standardisierten Fall die normalen Bänder kaum von den gebootstrapteten und liegen in allen Fällen verhältnismäßig nahe am Schätzer selbst. Sollte nun der Schätzer ein klein wenig unpräzise berechnet worden sein, so hat diese geringe Spannweite zur Folge, dass die wahre Kovarianzfunktion relativ schnell außerhalb aller Bänder liegt. Daher ist bei einer kleinen Stichprobe zunächst einmal die standardisierte Berechnung vorzuziehen, auch wenn bei einer sehr flachen Kovarianz wie bei (S.1) die nicht standardisierten Bänder ebenfalls sehr akzeptable Ergebnisse liefern, wie Tabelle 11.6 zeigt. Abbildung 11.7 und 11.8 und Tabellen 11.7 und 11.8 zeigen zudem eindrucksvoll, dass die standardisierten Bänder zwingend gebootstrapt berechnet werden müssen, weil die normalen Bänder in diesem Fall komplett versagen. Weiterhin ist hier sowohl in den Abbildungen, als auch in den Tabellen der leichte Vorteil von Methode b) durch die größere Spannweite der Bänder erkennbar. Der Hintergrund für beides ist wieder in Tabelle 11.1 abzulesen: $\hat{c}_{\gamma,R}$ ist zwar im standardisierten Fall durch die höheren Supremumswerte des Gauß-Felds größer als im nicht standardisierten Fall, allerdings für alle n mehr oder weniger gleich, und da die Bänder selbst jeweils noch $\hat{\sigma}_{ZZ}$ beinhalten, reicht diese Vergrößerung für eine Optimierung der Bänder einfach nicht aus. Zudem ist in den letzten beiden Spalten ersichtlich, dass Methode b) leicht höhere Werte für die Suprema als Methode a) liefert, woraus eine größere Schranke $c_{\gamma,R}^*$, ergo eine größere Spannweite und damit in aller Regel auch eine bessere Überdeckung resultiert.

Nichtsdestotrotz sind die standardisiert gebootstrapteten Bänder auch nicht perfekt. Obschon sie bei einer kleinen Stichprobe in aller Regel bessere Ergebnisse als die nicht standardisierten liefern, ist deren Verlauf recht unpräzise, wie man gut in Abbildung 11.7 und 11.8 sehen kann und in Abschnitt 11.2 bereits erwähnt wurde. Hat man einen Datensatz mit einer kleinen Stichprobe, so ist vermutlich das beste Vorgehen, sich einmal alle drei verschiedenen Bänder standardisiert plotten zu lassen. Sind sich diese im Verlauf alle sehr ähnlich - wie in Abbildung 11.6 - so kann man sich relativ sicher sein, dass die wahre Kovarianz tatsächlich innerhalb des engsten Bandes verläuft. Lässt sich beim Vergleich der Bänder jedoch nicht wirklich eine Struktur feststellen, wie in Abbildung 11.7 und 11.8, so sollten noch keine Aussagen getroffen und die Stichprobe vergrößert werden.

Zu guter Letzt zeigt Abbildung 11.9 erneut die Relevanz von (A.4). Trotz der einfachen Kovarianz, suggerieren die Bänder einen vollkommen anderen Verlauf. Auch wenn die Überdeckung bei den gebootstrapteten Bändern in Tabelle 11.9 bei 100% liegt, ist das nicht unbedingt ein Qualitätsmerkmal, weil eine Überanpassung stattfindet und die wahre Kovarianz durch die viel zu große Spannweite schlichtweg zwischen den Bändern liegen „muss“. Erhöht man in diesem Fall jedoch die Stichprobe, so haben wir in Abschnitt 11.2 bereits eruiert, dass keines der drei Bänder mehr funktioniert.

Abbildung 11.6.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.1).

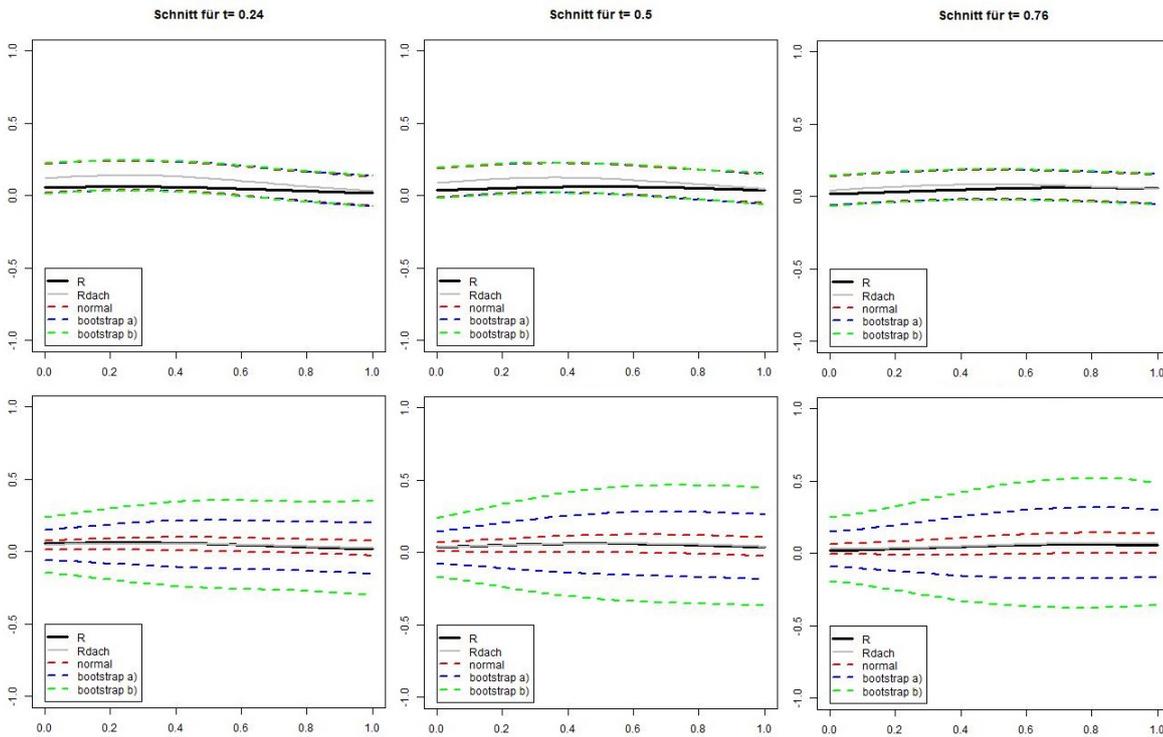


Abbildung 11.7.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.2).

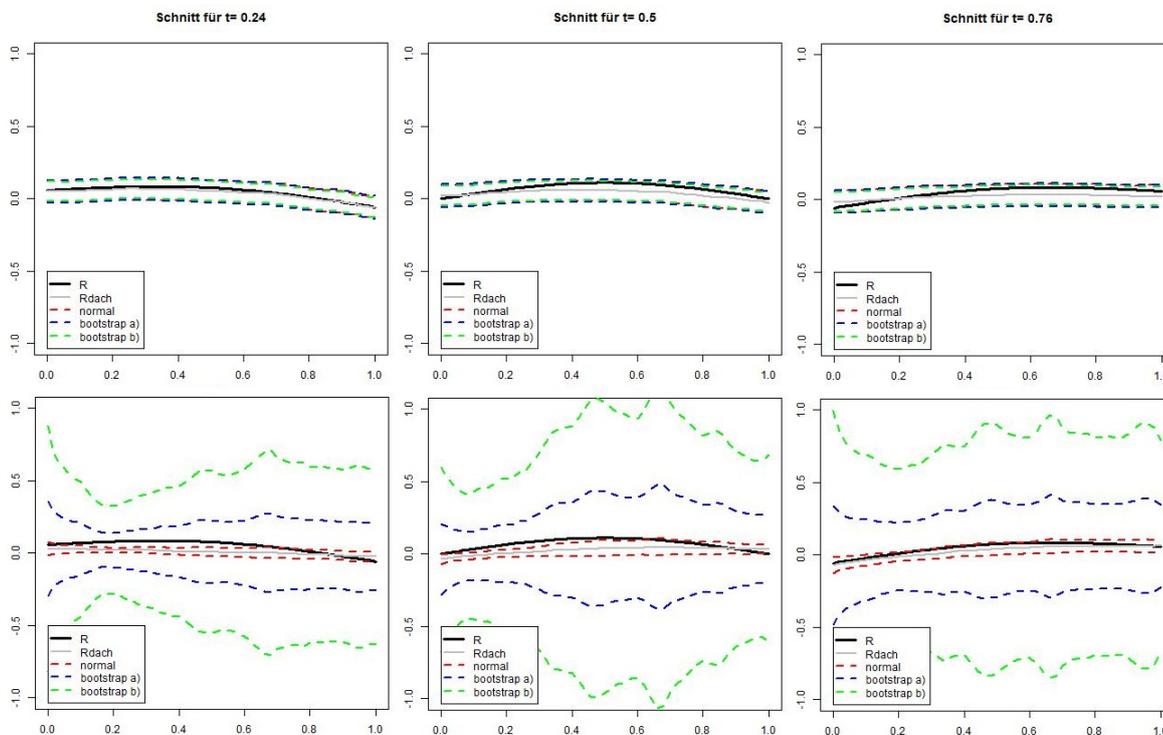


Abbildung 11.8.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.3).

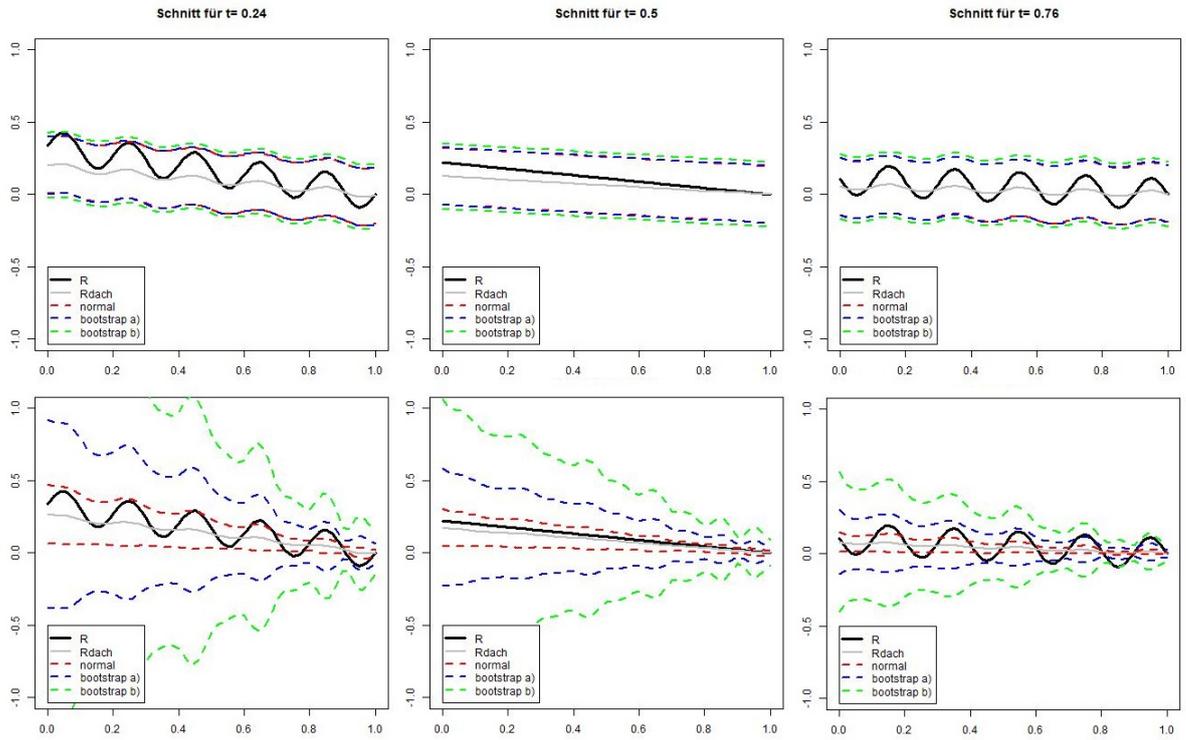


Abbildung 11.9.: R , \hat{R} und die Konfidenzbänder bzgl. des nicht standardisierten (oben) und des standardisierten (unten) empirischen Prozess für Setting (S.4).

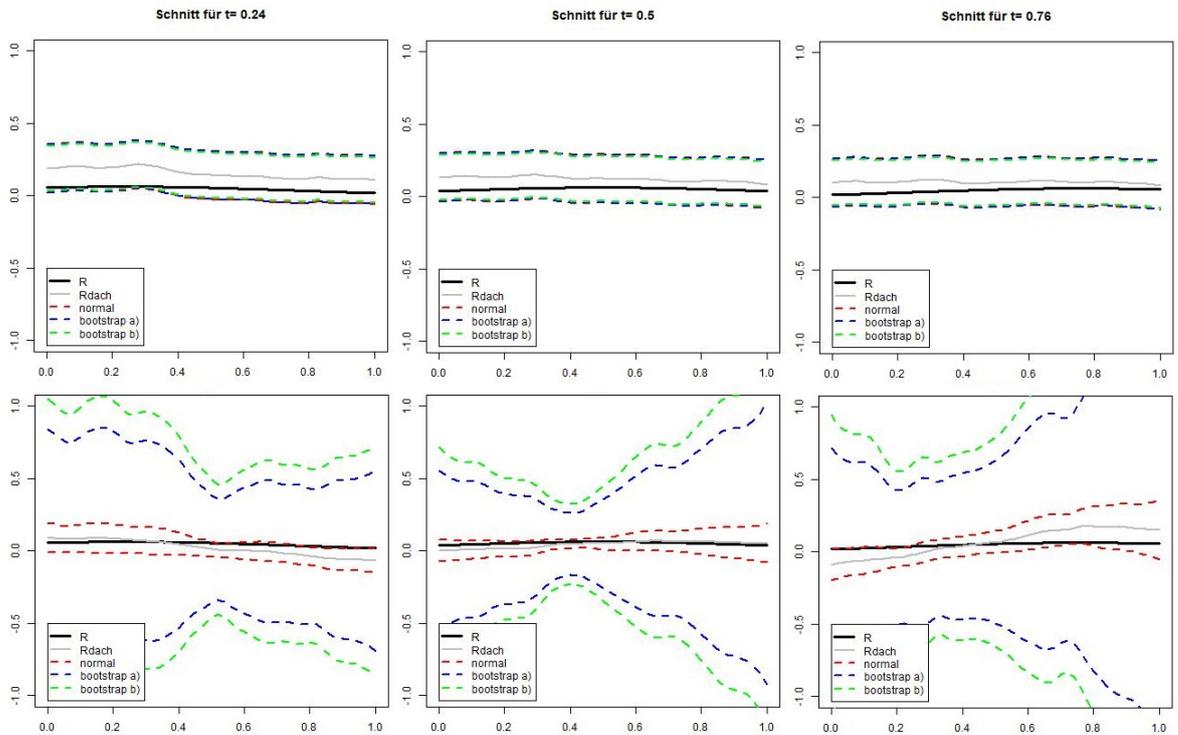


Tabelle 11.6.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.1).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert	10	10	0.2	0.883	0.2104	0.899	0.2167	0.89	0.2041	
	10	20	0.1	0.865	0.2129	0.884	0.2188	0.878	0.2073	
	10	50	0.05	0.882	0.2114	0.896	0.2181	0.889	0.2048	
	10	100	0.05	0.88	0.2135	0.902	0.2198	0.899	0.2075	
	20	20	0.1	0.93	0.2186	0.935	0.2217	0.928	0.2134	
	20	20	0.15	0.921	0.2197	0.923	0.2230	0.921	0.2149	
	20	20	0.2	0.919	0.2216	0.929	0.2246	0.92	0.2166	
	20	50	0.05	0.925	0.2182	0.931	0.2216	0.917	0.2134	
	20	100	0.05	0.913	0.2198	0.926	0.2229	0.918	0.2149	
standardisiert	10	10	0.2	0.558	2.6150	0.938	12.6780	0.975	20.7810	
	10	20	0.1	0.519	2.6211	0.936	13.5528	0.979	21.4942	
	10	50	0.05	0.536	2.6227	0.942	12.3396	0.984	20.4503	
	10	100	0.05	0.497	2.6191	0.917	13.0061	0.972	21.2213	
	20	20	0.1	0.686	2.6374	0.934	5.8820	0.963	7.0058	
	20	20	0.15	0.721	2.6354	0.929	5.8786	0.953	6.9939	
	20	20	0.2	0.675	2.6342	0.932	5.7652	0.95	6.8296	
	20	50	0.05	0.694	2.6424	0.934	5.9058	0.959	6.9863	
	20	100	0.05	0.699	2.6393	0.934	5.9031	0.959	6.9852	

Tabelle 11.7.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.2).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert	10	10	0.2	0.756	0.6204	0.77	0.6612	0.756	0.5541	
	10	20	0.1	0.725	0.6193	0.75	0.6642	0.726	0.5549	
	10	50	0.05	0.763	0.6205	0.777	0.6615	0.754	0.5505	
	10	100	0.05	0.701	0.5978	0.707	0.6444	0.697	0.5313	
	20	20	0.1	0.856	0.7529	0.857	0.7679	0.839	0.7048	
	20	20	0.15	0.814	0.6971	0.81	0.7131	0.797	0.6557	
	20	20	0.2	0.831	0.6730	0.834	0.6860	0.82	0.6303	
	20	50	0.05	0.854	0.7546	0.851	0.7701	0.844	0.7090	
	20	100	0.05	0.831	0.7170	0.823	0.7338	0.813	0.6720	
standardisiert	10	10	0.2	0.171	2.7692	0.766	23.8348	0.897	69.6397	
	10	20	0.1	0.164	2.7938	0.742	24.2065	0.881	72.1401	
	10	50	0.05	0.147	2.7867	0.717	24.5265	0.873	75.0898	
	10	100	0.05	0.18	2.7363	0.759	24.8419	0.887	80.5643	
	20	20	0.1	0.312	2.8056	0.835	16.7677	0.872	23.2269	
	20	20	0.15	0.325	2.7688	0.849	16.2843	0.886	22.0769	
	20	20	0.2	0.341	2.7434	0.846	15.7811	0.887	21.5710	
	20	50	0.05	0.312	2.8038	0.84	17.0747	0.885	23.5616	
	20	100	0.05	0.32	2.7607	0.834	16.6834	0.867	22.6223	

Tabelle 11.8.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.3).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert	10	10	0.2	0.686	2.5511	0.704	2.6580	0.663	2.3354	
	10	20	0.1	0.594	1.5077	0.604	1.5821	0.597	1.3816	
	10	50	0.05	0.555	1.3068	0.567	1.3913	0.541	1.1870	
	10	100	0.025	0.599	1.5255	0.611	1.6280	0.578	1.3828	
	20	20	0.1	0.643	1.8458	0.65	1.8566	0.634	1.7434	
	20	20	0.08	0.66	1.8335	0.658	1.8390	0.651	1.7350	
	20	50	0.05	0.659	1.7330	0.662	1.7468	0.648	1.6412	
	20	100	0.025	0.711	1.8740	0.713	1.9093	0.701	1.7768	
standardisiert	10	10	0.2	0	2.6398	0.227	20.2956	0.516	63.3245	
	10	20	0.1	0.004	2.6372	0.241	19.4353	0.487	49.2937	
	10	50	0.05	0.027	2.7201	0.601	19.1679	0.775	46.3495	
	10	100	0.025	0.027	2.7388	0.659	19.3954	0.812	45.2966	
	20	20	0.1	0.004	2.6385	0.166	12.3321	0.237	16.3785	
	20	20	0.08	0.005	2.6743	0.359	12.5783	0.459	17.1581	
	20	50	0.05	0.004	2.7441	0.606	11.2616	0.679	14.4677	
	20	100	0.025	0	2.7674	0.747	11.9468	0.794	15.5448	

Tabelle 11.9.: Überdeckungen und Schranken der Konfidenzbänder für Setting (S.4).

		normal			bootstrap a)		bootstrap b)			
		n	p	h	Überdeckung	$\hat{c}_{\gamma,R}$	Überdeckung	$c_{\gamma,R}^*$	Überdeckung	$c_{\gamma,R}^*$
nicht standardisiert	10	10	0.2	0.595	0.5935	0.648	0.6121	0.572	0.5766	
	10	20	0.1	0.626	0.5912	0.686	0.6100	0.602	0.5729	
	10	50	0.05	0.573	0.6092	0.624	0.6277	0.538	0.5851	
	10	100	0.05	0.592	0.5980	0.641	0.6164	0.551	0.5782	
	20	20	0.1	0.317	0.6166	0.349	0.6270	0.248	0.6002	
	20	20	0.15	0.317	0.6031	0.32	0.6119	0.239	0.5874	
	20	20	0.2	0.331	0.5955	0.351	0.6029	0.262	0.5787	
	20	50	0.05	0.275	0.6195	0.309	0.6300	0.226	0.6019	
	20	100	0.05	0.318	0.6245	0.332	0.6331	0.255	0.6057	
standardisiert	10	10	0.2	0.544	2.7195	0.995	13.9928	1	22.8367	
	10	20	0.1	0.497	2.7563	0.999	14.1387	1	23.2107	
	10	50	0.05	0.467	2.7611	0.999	15.2914	1	24.3669	
	10	100	0.05	0.466	2.7668	0.995	14.8765	0.998	24.5459	
	20	20	0.1	0.377	2.7767	0.995	6.6696	0.999	7.8898	
	20	20	0.15	0.414	2.7450	0.989	6.4639	0.997	7.6643	
	20	20	0.2	0.409	2.7188	0.99	6.2981	0.999	7.4902	
	20	50	0.05	0.359	2.8027	0.996	6.7391	0.999	7.9653	
	20	100	0.05	0.357	2.8009	0.993	6.7638	1	8.0021	

12. Fazit

12.1. Resultate und Interpretation

Werfen wir noch einmal einen kurzen Blick auf die in dieser Arbeit gewonnenen Erkenntnisse. Wir haben uns mit zwei Schätzern beschäftigt, welche dem Modell nach Gleichung (3.1) zugrunde liegen. Einmal für die Trend- und einmal für die Kovarianzfunktion, wobei für Letzteres der Trendschätzer in gewisser Weise als Argument der Stichprobenkovarianz verwendet wurde. Dieser Ansatz basiert, wie eingangs erwähnt, auf der Idee von [Deg11], welcher in seinem Artikel bereits einige Nachforschungen zum Trendschätzer anstellte, ähnlich zu denen in dieser Arbeit. Es gibt jedoch zwei markante Unterschiede: Zum einen wurde dort der lokal lineare Schätzer verwendet (siehe Abschnitt 3.4.2) und zum anderen wurde der Messfehler als normalverteilt angenommen. Beides wurde in dieser Arbeit auf einen viel allgemeinen Fall erweitert. Der Schätzer kann im Prinzip jeder lineare Schätzer der Form von Gleichung (3.2) sein, dessen Gewichtsfunktionen Annahme 3.4 erfüllen, der lokal lineare Schätzer ist demnach nur ein Spezialfall. Zudem muss der Fehler keineswegs normalverteilt sein, tatsächlich ist keine Annahme einer konkreten Verteilung möglich, solange der Fehler (A.4) erfüllt. Gerade der erste Punkt ist eine deutliche Verbesserung, weil man dadurch für jeden beliebigen linearen Schätzer, nach Überprüfung von (A.4) weiß, dass alle Aussagen aus dieser Arbeit für ihn gelten. Der Kovarianzschätzer hingegen wurde in [Deg11] nicht weiter erörtert und diente nur als Mittel zum Zweck, um die Kovarianzfunktion des Gaußprozesses in der Grenzverteilung zu schätzen. Dies wurde in dieser Arbeit ausführlich ergänzt und gezeigt, dass mehr oder weniger alle Eigenschaften des Trendschätzers auf den Kovarianzschätzer übertragen werden können, natürlich mit dementsprechender Erhöhung der Dimension.

Für die Schätzer selbst wurden zunächst diverse wünschenswerte Eigenschaften nachgewiesen, wobei sich die folgenden Eigenschaften herausstellten: Für beide Schätzer lässt sich der Bias, sowie die Varianz - und damit auch der Mean Squared Error - abschätzen bzw. beschränken, siehe dafür Abschnitt 8.1. Beide konvergieren für $n \rightarrow \infty$ punktweise gegen 0, woraus sich unmittelbar die punktweise Erwartungstreue und Konsistenz im quadratischen Mittel folgern ließ. Der Bias konvergiert sogar bzgl. der Supremumsnorm gegen 0, was bedeutet, dass die Schätzer auch gleichmäßig asymptotisch erwartungstreu sind (tatsächlich konvergiert auch die Varianz, und damit der Mean Squared Error bzgl. der Supremumsnorm gegen 0, was aber keine wirklich nützliche Erkenntnis ist, da man für gleichmäßige Konsistenz im quadratischen Mittel das Supremum innerhalb, und nicht wie in diesem Fall außerhalb des Erwartungswertes bräuchte).

Weiterhin wurde in Abschnitt 4.3 belegt, dass die Schätzer auch punktweise stark, sowie gleichmäßig schwach konsistent sind. Die wichtigste Aussage ist aber vermutlich die asymptotische Normalität aus Abschnitt 4.2. Für den Trendschätzer ist das nur eine bedingt neue Erkenntnis, da es bereits in [Deg11] und [Her17] ausführlich nachgewiesen wurde, dort aber wie bereits erwähnt unter deutlich stärkeren Annahmen. Für den Kovarianzschätzer stellt die Aussage ein komplettes Novum dar.

In Kapitel 5 wurde die asymptotische Normalität der Schätzer dafür genutzt, um Konfidenzbänder für die wahre Funktion zu bestimmen, wobei die Bänder aufgrund der schwachen Konvergenz in $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ bzw. $(\mathcal{C}([0, 1]^2), \|\cdot\|_\infty)$ simultan, also mit einer besseren Überdeckungs-Performance formuliert werden konnten. Die Bänder können dabei auf zwei verschiedene Arten berechnet werden: Entweder nicht standardisiert, also mit einer Art Kovarianzfunktion, die den Gauß-Prozess im Grenzfall für das benötigte Quantil festlegt, oder eben standardisiert, also mit einer Korrelationsfunktion im Gauß-Prozess. Welche Bänder besser sind, hängt vom konkreten Fall ab, wie später ermittelt wurde. Die große Schwierigkeit hierbei war, dass die Kovarianz- bzw. Korrelationsfunktion im Gauß-Prozess nicht bekannt ist, und daher geschätzt werden musste. Dies machte den Beweis deutlich komplexer, weil somit zusätzlich relevant war, dass der Gauß-Prozess mit den geschätzten Kovarianz-/Korrelationsfunktionen schwach gegen den Gauß-Prozess mit den wahren Funktionen konvergiert.

Für die Konfidenzbänder wurde in Kapitel 6 zusätzlich eine (naive) Bootstrap-Methode vorgestellt. Auch diese ist bereits in [Deg11] zu finden, jedoch auch nur für die Bänder des Trends und ohne jeglichen Beweis, was hier, inklusive des Bootstraps für die Bänder der Kovarianz, ausführlich ergänzt wurde. Für Letzteren gibt es sogar zwei Methoden, die sich hinsichtlich der Berechnung der Bootstrap-Stichprobe leicht unterscheiden, wobei eine der beiden eine bessere Laufzeit, dafür aber eine etwas schlechtere Überdeckung aufweist.

Als letzten theoretischen Teil, wurde in Kapitel 7 begründet, dass sämtliche Eigenschaften der Schätzer, wie auch die Konfidenzbänder und die Bootstrap-Methoden erhalten bleiben, falls man in asynchrone Messpunkte, also in jeder Stichprobe potentiell verschiedene Messstellen x_{ij} , $i = 1, \dots, n$, $j = 1, \dots, p$ betrachtet. Allerdings war es dafür auch nötig, die Annahmen an das Modell, wie auch an den Schätzer leicht anzupassen bzw. zu erweitern, weil hier schlichtweg keine so schön simultane Glättung wie im synchronen Fall möglich ist.

In Kapitel 11 wurde schließlich die Gültigkeit der Konfidenzbänder für die Kovarianz (im synchronen Fall) untermauert, indem die Bänder an vier verschiedenen Datensätzen bzgl. ihrer Überdeckung getestet wurden. Für die ersten drei Settings funktionierte die Überdeckung für große Stichproben und viele Messstellen problemlos, auch für eine sehr zerfahrene Kovarianzfunktion.

Hier war die Haupteigentnis, dass bei einer großen Stichprobe in jedem Fall die nicht standardisierten Bänder vorzuziehen sind, weil die standardisierten gerade abseits des Bootstraps eine zu große Schwankung, und damit schlechtere Überdeckung aufweisen. Im Fall einer kleinen Stichprobe und wenig Beobachtungspunkten, sind die standardisierten Bänder in jedem Fall besser, wenn auch nicht durchweg perfekt. Je nach Komplexität der tatsächlichen Kovarianzstruktur hat man gesehen, dass die Bänder mal besser, mal schlechter funktionieren. Daher sollte man die Bänder zunächst einmal sowohl normal, als auch gebootstrapt berechnen und vergleichen, bevor man vorschnell Schlüsse zieht. Ggf. ist dann auch eine Vergrößerung der Stichprobe unumgänglich, falls sich die Bänder zu sehr voneinander unterscheiden.

Das letzte Setting zeigte schließlich noch abrundend, dass die Schätzer keineswegs „alle“ möglichen und willkürlichen Daten gut verarbeiten, da selbst die moderateste Kovarianzfunktion innerhalb des einfachsten Settings nicht erkannt wird, wenn der Fehlerterm nicht die geforderten Eigenschaften aufweist. Dies ist auch noch einmal eine schöne Visualisierung dafür, dass es beim Fehler tatsächlich nicht auf die Verteilung selbst, sondern nur auf die Kovarianzstruktur ankommt: Für das komplexe Setting (S.3) ist der Fehler nicht normalverteilt, für das mutmaßlich einfache (S.4) schon. Trotzdem funktionieren die Bänder für Letzteres nicht, für Ersteres hingegen schon.

12.2. Ausblick für weitere Nachforschungen

Auch wenn diese Arbeit eine möglichst breit gefächerte Zusammenfassung der Eigenschaften und Verwendungszwecke - inkl. Beweise - für die Schätzer $\hat{\mu}$ und \hat{R} liefert, gibt es natürlich noch einige weitere Aspekte, welche sich für weitere Nachforschungen eignen.

Zwar wurde gezeigt, dass die Schätzer sowohl asymptotisch erwartungstreu, als auch konsistent sind, allerdings heißt das noch nicht, dass es unter Umständen nicht auch noch deutlich bessere Schätzer für μ und R gibt. Mit anderen Worten wurde nicht untersucht, ob es sich hierbei um **UMVU-Schätzer** handelt. Ggf. ließen sich die Schätzer über die Sätze von *Rao-Blackwell* bzw. *Lehmann-Scheffé* noch verbessern bzw. zeigen, dass es bereits die gleichmäßig besten sind. Genauso könnte man evtl. über die *Cramer-Rao-Ungleichung* Effizienz, oder, falls dies nicht gilt, eine asymptotische Effizienz überprüfen. Alles natürlich wieder jeweils auf funktionaler Ebene betrachtet.

Weiterhin wurde die asymptotische Normalität nachgewiesen und gezeigt, dass man in der Wahl des Fehlerterms relativ frei ist. Die gleichmäßige asymptotische Vernachlässigbarkeit der Fehlerterme wurde durch Kombination der punktweisen Konvergenz und gleichmäßigen gleichgradigen Stetigkeit - jeweils in \mathbb{P} - gezeigt. Gerade für letzteres wurde in Kauf genommen, dass sich die Stichprobengröße n , die Anzahl der Beobachtungspunkte p und die Bandbreite h nach einem relativ striktem Schema verhalten müssen. Insbesondere die vier Voraussetzungen

$$\sqrt{n}h^{1+\alpha_R} \rightarrow 0, \frac{\sqrt{n}}{p^{1-d_1}h} \rightarrow 0, \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty, \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} < \infty$$

aus Satz 4.5 erfordern, dass p extrem schnell gegen ∞ konvergiert. Eventuell gibt es noch effizientere Möglichkeiten, welche die Vernachlässigbarkeit des Fehlerterms gewährleisten und zugleich einen größeren Spielraum für n , p und h bieten.

Für die Resultate mit asynchronen Messpunkte aus Kapitel 7, wurde in Annahme 7.1 zudem die Existenz und Höldereigenschaft der achten Verbundmomente von Z und ε gefordert. Eine Abschwächung auf die vierten Momente wäre wünschenswert und ist auch nicht völlig aus der Luft gegriffen, da diese im Fall synchroner Messpunkte auch ausreichen. Eine weitere Herangehensweise wäre es, die Messpunkte nicht nur asynchron, sondern sogar zufällig festzulegen. Das ist auch durchaus für Anwendungsfälle interessant: Man stelle sich bspw. eine Messstation vor, welche einen stetigen Temperaturverlauf eines Tages durch diskrete Werte ausmisst, allerdings erst ab einer bestimmten Minimalschwankung der Temperatur einen neuen Messwert sendet. Diese Temperaturschwankungen sind in erster Linie zufällig, also weiß man im Vorfeld nicht, an welchen Punkten (neue) Temperaturwerte für die Daten verfügbar sind. Mutmaßlich wären für die Resultate dann aber vermutlich deutlich mehr Annahmen nötig, da hierdurch eine weitere stochastische Komponente in das Modell einfließt.

Was in dieser Arbeit ebenfalls keine große Rolle spielt, ist die Konvergenzgeschwindigkeit; es wurde nur bestätigt, dass jeweils die asymptotische Erwartungstreue, Normalität, sowie Konsistenz gilt, ohne groß Wert auf die Rate der jeweiligen Terme zu legen. Manch einem Anwender sind diese Raten möglicherweise deutlich zu langsam. [BHH23] überprüfen bspw. den exakt gleichen Schätzer bzgl. der Konvergenzgeschwindigkeit und erzielen ein deutlich besseres Ergebnis als hier, wobei dazu gesagt werden muss, dass dafür auch wieder stärkere Einschränkungen an den Fehler, wie eine Sub-Gaußverteilung, gemacht werden. Es wäre durchaus erstrebenswert, diese Aussagen für einen Fehler wie hier, ohne konkrete Verteilung auch zu erreichen.

Zudem darf man auch nach wie vor nicht vergessen, dass es in dieser Arbeit zwar um funktionale Daten geht, allerdings im Kern der *sparse data* Ansatz verfolgt wird. Die Daten werden also nur an endlich vielen Punkten ausgewertet, und danach wird ein Grenzübergang zu unendlich vielen vollzogen, um die Funktionalität zu erhalten. Das ist legitim und für die Praxis oftmals auch ausreichend, ist in der Realität doch meist auch nur eine Beobachtung bzw. Datensammlung endlicher Größe durchführbar. Nichtsdestotrotz könnte man zumindest theoretisch auch versuchen, das Modell von vornherein funktional in der Form

$$Y_i(t) = \mu(t) + X_i(t) \quad (+ \varepsilon_i(t))$$

mit $X_i : T \rightarrow \mathbb{R}$ auszudrücken und die Aussagen für dieses Modell zu beweisen. Die Stichprobendaten würden also direkt kontinuierlich erfasst, wobei der Fehler hier erst einmal eingeklammert ist, da man sich die Frage stellen sollte, ob die Annahme eines derartigen Fehlers, zusätzlich zu den individuellen Abweichung $X_i(t)$, überhaupt sinnvoll ist, oder ob er das Modell evtl. so stark beeinflusst, dass er bzgl. der Schätzungen untrennbar mit $X_i(t)$ verbunden ist und sich nicht ausglätten lässt.

Eine weitere Möglichkeit als Erweiterung zu dieser Arbeit wäre die Fragestellung: Was passiert, wenn bestimmte Voraussetzungen aus Annahme 3.1 bzw. Annahme 7.1 verletzt sind, der Rest aber erfüllt ist? Funktionieren die Schätzer dann immer noch in jeder Hinsicht? Funktionieren sie gar nicht mehr? Funktionieren sie nur noch für bestimmte n , p und h ? Simulationsstudien legen nahe, dass zumindest $\hat{\mu}$ auch innerhalb des Settings (S.4) aus Kapitel 11, also inklusive eines Fehlers, mit augenscheinlich zu starker Abhängigkeitsstruktur, und sogar für nicht differenzierbare oder gar nicht stetige zugrunde liegende Funktionen μ zu einem gewissen Grad trotzdem funktioniert. Eine solche Erweiterung ist natürlich auch für den Schätzer \hat{R} in diversen Ausführungen denkbar, was ggf. zu einer legitimen Abschwächung der Annahmen führen könnte.

Zu guter Letzt sind die Bootstrap-Versionen aus Kapitel 6 natürlich nur eine von vielen Möglichkeiten. Es ist durchaus möglich, dass es noch Weitere gibt, die entweder eine noch bessere Laufzeit, oder Performance haben. Gerade bei Datensätzen mit einer stark schwankenden Kovarianzfunktion, haben wir in den Simulationen gesehen, dass hier auch die Qualität des Bootstraps stark schwankt. Oder um es etwas plakativer zu formulieren: Qualität lässt nach, wenn sie nicht verbessert wird!

A. Anhang

A.1. Beweiserganzung zu Satz 2.22

Wir definieren zunachst einmal die Handhabbarkeit, wie in [Pol90] formuliert. Sei da-
fur $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ ein Dreiecksschema von stochastischen Prozessen mit $t \in T$. Diese
Prozesse sollen in jeder Zeile unabhangig sein, das heit die Folgen $(X_{n,i})_{1 \leq i \leq k_n}$ sind
unabhangig bezuglich i . Fur ein festes $t \in T$ sei

$$X_n^{k_n}(t) := (X_{n,1}(t), \dots, X_{n,k_n}(t))^T \in \mathbb{R}^{k_n}$$

und

$$Y_n := \{X_n^{k_n}(t) : t \in T\} \subset \mathbb{R}^{k_n}$$

Zudem sei Φ_n eine Umhullende von Y_n , nach Definition 2.21.

Definition A.1. (*Handhabbarkeit*)

Ein Dreiecksschema von stochastischen Prozessen $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ heit **handhabbar
bezuglich der Umhullenden** Φ_n , wenn eine deterministische Funktion $\lambda > 0$ existiert,
fur die gilt

- i) Fur alle $n \in \mathbb{N}$, $\delta \in (0, 1]$ und alle Vektoren $\alpha \in \mathbb{R}^{k_n}$ mit $\alpha \geq 0$ gilt fur die
Kapazitatszahl nach Definition 2.8:

$$N_{\text{cap}}(\delta \|\alpha \odot \Phi_n\|_2, d^2, \alpha \odot Y_n) < \lambda(\delta)$$

ii) $\int_0^1 \sqrt{\log(\lambda(\delta))} d\delta < \infty$.

wobei \odot das komponentenweise Produkt von Vektoren ist und auch der Vergleichsopera-
tor \geq komponentenweise betrachtet wird. d^2 entspricht der euklidischen Metrik, also der
von der euklidischen Norm erzeugten Metrik.

Eine Folge von stochastischen Prozessen $(X_i)_{i \in \mathbb{N}}$ heit handhabbar, wenn das dazugeho-
rige Dreiecksschema, definiert durch $X_{n,i} = X_i$, $i \leq n$ handhabbar ist.

Proposition A.2.

Ein Dreiecksschema von stochastischen Prozessen $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq k_n}$ erfulle Bedingung v) aus Satz 2.22. Dann ist das Schema handhabbar nach Definition A.1.

Beweis.

Wenn wir fur alle $n \in \mathbb{N}$, $\delta \in (0, 1]$ und alle $\alpha \in \mathbb{R}^{k_n}$ mit $\alpha \geq 0$ eine uberdeckungszahl finden, fur die

$$N_{\text{cov}}(\delta \|\alpha \odot \Phi_n\|_2, d^2, \alpha \odot Y_n) < \lambda(\delta)$$

gilt, dann impliziert Lemma 2.9 insbesondere

$$\begin{aligned} N_{\text{cap}}(2\delta \|\alpha \odot \Phi_n\|_2, d^2, \alpha \odot Y_n) &\leq N_{\text{cov}}(\delta \|\alpha \odot \Phi_n\|_2, d^2, \alpha \odot Y_n) \\ \Rightarrow N_{\text{cap}}(2\delta \|\alpha \odot \Phi_n\|_2, d^2, \alpha \odot Y_n) &< \lambda(\delta) \end{aligned} \quad (\text{A.1})$$

Wir konnen also die Kapazitatzahl wiederum uber die uberdeckungszahl abschatzen, wobei dadurch sogar zu viele Werte betrachtet werden. Es wurde genugen, die uberdeckungszahl nur fur alle $\delta \in (0, \frac{1}{2}]$ herzuleiten und $\sqrt{\log(\lambda(\delta))}$ musste nur uber $(0, \frac{1}{2}]$ integrierbar sein, was aber aus Bedingung i) von Definition A.1 unmittelbar folgt.

Die eben genannte uberdeckungszahl versuchen wir nun genauer zu verstehen. Nach Definition 2.8 suchen wir fur alle $n \in \mathbb{N}$ und alle Vektoren $\alpha \in \mathbb{R}^{k_n}$ mit $\alpha \geq 0$ die kleinste Zahl $N(\delta; n, \alpha)$, sodass Punkte $\tau_1, \dots, \tau_{N(\delta; n, \alpha)} \in T$ existieren, welche folgende Bedingung erfullen.

$$\begin{aligned} \forall t \in T, \exists m \in \{1, \dots, N(\delta; n, \alpha)\} : &\|\alpha \odot X_n^{k_n}(t) - \alpha \odot X_n^{k_n}(\tau_m)\|_2 \leq \delta \|\alpha \odot \Phi_n\|_2 \\ \Leftrightarrow \left(\sum_{i=1}^{k_n} \alpha_i^2 (X_{n,i}(t) - X_{n,i}(\tau_m))^2 \right)^{\frac{1}{2}} &\leq \delta \left(\sum_{i=1}^{k_n} \alpha_i^2 \Phi_{n,i}^2 \right)^{\frac{1}{2}} \\ \Leftrightarrow \sum_{i=1}^{k_n} \alpha_i^2 (X_{n,i}(t) - X_{n,i}(\tau_m))^2 &\leq \delta^2 \sum_{i=1}^{k_n} \alpha_i^2 \Phi_{n,i}^2 \end{aligned}$$

Dann soll fur alle $\delta \in (0, 1]$ eine Abschatzung fur $N(\delta; n, \alpha)$ getroffen werden, welche unabhangig von n und α immer gilt. Fur ein festes $\delta \in (0, 1]$ sei $\tilde{\delta} := \frac{\delta}{C}$. Falls fur dieses $\tilde{\delta}$ Punkte $\tau_1, \dots, \tau_{N(\tilde{\delta})} \in T$ existieren, sodass fur alle $t \in T$ ein $m \in \{1, \dots, N(\tilde{\delta})\}$ mit

$$d^1(t, \tau_m) \leq \tilde{\delta}^\gamma$$

existiert, so folgt aus der Implikation von Gleichung (2.3) fur die betrachteten τ 's fur alle $i = 1, \dots, k_n$:

$$|X_{n,i}(t) - X_{n,i}(\tau_m)| \leq C \tilde{\delta} \Phi_{n,i} = C \frac{\delta}{C} \Phi_{n,i} = \delta \Phi_{n,i}$$

Daher gilt für alle nicht-negativen Vektoren $\alpha \in \mathbb{R}^{k_n}$:

$$\begin{aligned} \alpha_i |X_{n,i}(t) - X_{n,i}(\tau_m)| &\leq \delta \alpha_i \Phi_{n,i} \Rightarrow \alpha_i^2 (X_{n,i}(t) - X_{n,i}(\tau_m))^2 \leq \delta^2 \alpha_i^2 \Phi_{n,i}^2 \\ &\Rightarrow \sum_{i=1}^{k_n} \alpha_i^2 (X_{n,i}(t) - X_{n,i}(\tau))^2 \leq \delta^2 \sum_{i=1}^{k_n} \alpha_i^2 \Phi_{n,i}^2 \end{aligned}$$

Nach den vorigen Überlegungen entspricht $N(\tilde{\delta})$ gerade $N(\delta; n, \alpha)$, also auch der hergeleiteten Kapazitätzahl von Gleichung (A.1). Um $N(\tilde{\delta})$ zu bestimmen, muss man sich überlegen, wie viele abgeschlossene Hyperkugeln der Dimension $D := \dim(T)$ mit Radius $\tilde{\delta}^\gamma$ mindestens nötig sind, um T zu überdecken. Mit $T_{\max} := \max_{t \in T} \|t\|_1$ und der bzgl. der Abstände innerhalb T verwendeten Manhattan-Metrik d^1 , erhalten wir über Beispiel 2.10:

$$N(\tilde{\delta}) \leq \left(\frac{2T_{\max} \sqrt{D}}{\tilde{\delta}^\gamma} \right)^D = \left(\frac{2T_{\max} \sqrt{D}}{\frac{\delta^\gamma}{C}} \right)^D = (2T_{\max} \sqrt{D} C \delta^{-\gamma})^D$$

Wir definieren nun

$$\lambda(\delta) := (C^* \delta^{-\gamma})^D \text{ mit } C^* := \max\{2T_{\max} \sqrt{D} C, 1\} < \infty, \text{ da } T_{\max} < \infty$$

dann gilt i) von Definition A.1. Nun muss noch ii), also die Endlichkeit des Integrals überprüft werden. Nach den Rechenregeln des Logarithmus gilt:

$$\log((C^* \delta^{-\gamma})^D) = D \log(C^*) + D\gamma(-\log(\delta))$$

Weil $\delta \in (0, 1]$, $\gamma \in \mathbb{R}^+$ und $C^* \geq 1$, sind alle Summanden nicht-negativ, der Integrand ist wohldefiniert. Außerdem gilt für $x, y \in \mathbb{R}^+$:

$$x + y \leq x + 2\sqrt{x}\sqrt{y} + y = (\sqrt{x} + \sqrt{y})^2 \Leftrightarrow \sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$$

Wegen der Monotonie des natürlichen Logarithmus' und die Wurzelfunktion, kann der Integrand weiter abgeschätzt werden durch:

$$\left(\log((C^* \delta^{-\gamma})^D) \right)^{\frac{1}{2}} \leq \sqrt{D \log(C^*)} + \sqrt{D\gamma} \sqrt{-\log(\delta)}$$

Es folgt insgesamt:

$$\begin{aligned} \int_0^1 \left(\log((C^* \delta^{-\gamma})^D) \right)^{\frac{1}{2}} d\delta &\leq \int_0^1 \sqrt{D \log(C^*)} + \sqrt{D\gamma} \sqrt{-\log(\delta)} d\delta \\ &= \int_0^1 \sqrt{D \log(C^*)} d\delta + \sqrt{D\gamma} \int_0^1 \sqrt{-\log(\delta)} d\delta \\ &= \sqrt{D \log(C^*)} + \frac{\sqrt{D\gamma\pi}}{2} < \infty \end{aligned}$$

letztlich also die Handhabbarkeit. □

A.2. ε aus Setting (S.4) erfüllt Annahme (A.4) nicht

Proposition A.3.

Sei $\text{Cov}(\vec{\varepsilon}_1) = (\sigma_{jk})_{j,k=1,\dots,p}$ mit $\sigma_{jk} = \frac{p - |j - k|}{10p}$. Dann existiert kein $d_1 \in [0, 1)$, für welches der maximale Eigenwert $\lambda_{\max,1}$ von $\text{Cov}(\vec{\varepsilon}_1)$ von der Ordnung $\mathcal{O}(p^{d_1})$ ist, d. h. es gibt keine Konstante $C_1 \in \mathbb{R}^+$, so dass ein $p_1 \in \mathbb{N}$ mit $\lambda_{\max,1} \leq C_1 p^{d_1}$ für alle $p \geq p_1$ existiert.

Beweis.

Nach [WM08] ist der maximale Eigenwert einer symmetrischen Matrix $A = (a_{jk})_{j,k=1,\dots,p}$ mindestens so groß wie $\frac{1}{p} \sum_{j,k=1}^p a_{jk}$, d. h. es gilt $\lambda_{\max,1} \geq \frac{1}{p} \sum_{j,k=1}^p \sigma_{jk}$.

$$\sum_{j,k=1}^p \sigma_{jk} = \sum_{j,k=1}^p \frac{p - |j - k|}{10p} = \frac{1}{10} \left(p^2 - \frac{1}{p} \sum_{j,k=1}^p |j - k| \right)$$

$|j - k|$ ist der Abstand zwischen j und k , deshalb gilt für die Summe:

$$\begin{aligned} \sum_{j,k=1}^p |j - k| &= \underbrace{p \cdot 0}_{p \text{ mal Abstand } 0} + \underbrace{2 \cdot (p-1) \cdot 1}_{2(p-1) \text{ mal Abstand } 1} + \underbrace{2 \cdot (p-2) \cdot 2}_{2(p-2) \text{ mal Abstand } 2} + \dots + \underbrace{2 \cdot 1 \cdot (p-1)}_{2(p-(p-1)) \text{ mal Abstand } p-1} \\ &= \sum_{j=0}^{p-1} \underbrace{2(p-j)j}_{=0 \text{ für } j=p} = \sum_{j=0}^p 2(p-j)j = \sum_{j=0}^p (2pj - 2j^2) = 2p \sum_{j=0}^p j - 2 \sum_{j=0}^p j^2 \\ &= 2p \cdot \frac{p(p+1)}{2} - 2 \cdot \frac{p(p+1)(2p+1)}{6} = p^3 + p^2 - \frac{2p^3 + p^2 + 2p^2 + p}{3} \\ &= \frac{3p^3 + 3p^2 - 2p^3 - p^2 - 2p^2 - p}{3} = \frac{p^3 - p}{3} \end{aligned}$$

$$\Rightarrow \sum_{j,k=1}^p \sigma_{jk} = \frac{1}{10} \left(p^2 - \frac{1}{p} \cdot \frac{p^3 - p}{3} \right) = \frac{1}{10} \cdot \frac{3p^3 - p^3 + p}{3p} = \frac{2p^2 + 1}{30}$$

$$\Rightarrow \frac{1}{p} \sum_{j,k=1}^p \sigma_{jk} = \frac{1}{p} \cdot \frac{2p^2 + 1}{30} = \frac{2}{30}p + \frac{1}{30p} = \mathcal{O}(p) \supset \mathcal{O}(p^{d_1}), \quad \forall d_1 \in [0, 1)$$

$$\Rightarrow \forall C_1 \in \mathbb{R}^+, \exists p_1 \in \mathbb{N}, \forall p \geq p_1 : \lambda_{\max,1} > C_1 p^{d_1}$$

□

A.3. Erfüllbarkeit der Annahmen von Satz 4.4 und Satz 4.5

Satz 4.4 hat $\sqrt{nh}^{1+\alpha_\mu} \rightarrow 0$ sowie $\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty$ als zusätzliche Voraussetzungen. Sei

z. B. $p = \lceil n^{\frac{b}{1-d_1-a}} \rceil$ und $h = \frac{1}{np^a}$, wobei $0 < a < 1 - d_1$ und $b > 2$ gelten soll. Dann gilt nämlich zum Einen:

$$p \geq n^{\frac{b}{1-d_1-a}} \xrightarrow{1-d_1-a > 0} \infty \Rightarrow p \rightarrow \infty \Rightarrow p^a \xrightarrow{a > 0} \infty \Rightarrow h = \frac{1}{np^a} \rightarrow 0$$

Die Parameter sind also sinnvoll definiert. Zum Anderen:

$$\sqrt{nh}^{1+\alpha_\mu} \stackrel{n,p > 0}{\leq} \sqrt{nh} = \frac{1}{\sqrt{np^a}} \rightarrow 0$$

und zudem mit der *allgemeinen harmonischen Reihe*:

$$\frac{1}{p^{1-d_1}h} = \frac{n}{p^{1-d_1-a}} \leq \frac{n}{(n^{\frac{b}{1-d_1-a}})^{1-d_1-a}} = \frac{1}{n^{b-1}} \Rightarrow \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} \leq \sum_{n=1}^{\infty} \frac{1}{n^{b-1}} \stackrel{b-1 > 1}{<} \infty$$

Satz 4.5 hat $\sqrt{nh}^{1+\alpha_R} \rightarrow 0$, $\frac{\sqrt{n}}{p^{1-d_1}h} \rightarrow 0$ sowie $\sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} < \infty$ und $\sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} < \infty$

als zusätzliche Voraussetzungen. Das Beispiel hierfür ähnelt obigem, ist jedoch etwas komplexer. Sei nun $p = \lceil n^{\frac{b}{\min\{1-d_1-a, 2-d_2-2a\}}} \rceil$ und wieder $h = \frac{1}{np^a}$, wobei dieses mal $0 < a < 1 - d_1$, $0 < 2a < 2 - d_2$ und $b > 3$ gelten soll. Dann folgt wieder:

$$p \geq n^{\frac{b}{\min\{1-d_1-a, 2-d_2-2a\}}} \xrightarrow{\min\{1-d_1-a, 2-d_2-2a\} > 0} \infty \Rightarrow p \rightarrow \infty \Rightarrow p^a \xrightarrow{a > 0} \infty \Rightarrow h = \frac{1}{np^a} \rightarrow 0$$

Die Parameter sind also sinnvoll definiert. $\sqrt{nh}^{1+\alpha_R} \rightarrow 0$ folgt damit wie oben. Nach der Festlegung von a und p gilt weiterhin sowohl $p \geq n^{\frac{b}{1-d_1-a}}$, als auch $p \geq n^{\frac{b}{2-d_2-2a}}$; je nachdem, ob $1 - d_1 - a$ oder $2 - d_2 - 2a$ kleiner ist, kann aus dem einen jeweils das andere gefolgert werden. Diese Erkenntnis kann nun, wieder in Kombination mit der *allgemeinen harmonischen Reihe*, gewinnbringend eingesetzt werden.

$$\frac{\sqrt{n}}{p^{1-d_1}h} = \frac{n\sqrt{n}}{p^{1-d_1-a}} \leq \frac{n^{\frac{3}{2}}}{(n^{\frac{b}{1-d_1-a}})^{1-d_1-a}} = \frac{1}{n^{b-\frac{3}{2}}} \stackrel{b > 3}{<} \frac{1}{n^{\frac{3}{2}}} \rightarrow 0$$

$$\frac{1}{p^{1-d_1}h} = \frac{n}{p^{1-d_1-a}} \leq \frac{n}{(n^{\frac{b}{1-d_1-a}})^{1-d_1-a}} = \frac{1}{n^{b-1}} \Rightarrow \sum_{n=1}^{\infty} \frac{1}{p^{1-d_1}h} \leq \sum_{n=1}^{\infty} \frac{1}{n^{b-1}} \stackrel{b-1 > 1}{<} \infty$$

$$\frac{1}{p^{2-d_2}h^2} = \frac{n^2}{p^{2-d_2-2a}} \leq \frac{n^2}{(n^{\frac{b}{2-d_2-2a}})^{2-d_2-2a}} = \frac{1}{n^{b-2}} \Rightarrow \sum_{n=1}^{\infty} \frac{1}{p^{2-d_2}h^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^{b-2}} \stackrel{b-2 > 1}{<} \infty$$

Literaturverzeichnis

- [Adl90] R.J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Ims Lecture Series. Institute of Mathematical Statistics, 1990.
- [Bau11] H. Bauer. *Maß- und Integrationstheorie*. De Gruyter Lehrbuch. De Gruyter, 2011.
- [BF81] Peter J. Bickel and David A. Freedman. Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*, 9(6):1196 – 1217, 1981.
- [BHH23] Max Berger, Philipp Hermann, and Hajo Holzmann. From dense to sparse design: Optimal rates under the supremum norm for estimating the mean function in functional data analysis. 2023.
- [Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [cda] cdatac.de. Climate Data Check - Natürliche Variabilität der Temperaturdaten. <https://cdatac.de/index.php/global-warming/interne-variabilitaet>. (zuletzt aufgerufen am 17.11.2023).
- [CR87] J. Cigler and H.C. Reichel. *Topologie: eine Grundvorlesung*. BI-Hochschultaschenbücher. Bibliographisches Institut, 1987.
- [CT97] Y.S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer Texts in Statistics. Springer New York, 1997.
- [Dam10] Dambeck, H. Ist Fußball ein Glücksspiel? <https://www.spektrum.de/magazin/ist-fussball-ein-gluecksspiel/1030089>, 2010. (zuletzt aufgerufen am 17.11.2023).
- [Das08] A. DasGupta. *Asymptotic theory of statistics and probability*. Springer, 2008.
- [Dav15] François David. *The Formalisms of Quantum Mechanics*, volume 893 of *Lecture Notes in Physics*. Springer, January 2015.
- [Deg11] David A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765, 2011.

- [Don51] M.D. Donsker. *An Invariance Principle for Certain Probability Limit Theorems*. American Mathematical Society. Memoirs. 1951.
- [Dud67] R.M. Dudley. The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis*, 1(3):290–330, 1967.
- [Düm17] L. Dümbgen. *Empirische Prozesse*. Skripten zur mathematischen Statistik. Ges. zur Förderung der Math. Statistik, 2017.
- [Dun99] W. Dunham. *Euler: The Master of Us All*. Number Bd. 22 in Dolciani Mathematical Expositions. Mathematical Association of America, 1999.
- [Efr79] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.
- [Els13] J. Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2013.
- [ET93] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.
- [Fan92] Jianqing Fan. Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420):998–1004, 1992.
- [FFdR16] Mario Francisco-Fernández and Alejandro Quintela del Rio. Comparing Simultaneous and Pointwise Confidence Intervals for Hydrological Processes. *PLOS ONE*, 11(2):1–28, February 2016.
- [FP07] William Ford and James Penline. When does convergence in the mean imply uniform convergence? *American Mathematical Monthly*, 114, 01 2007.
- [Fre81] D. A. Freedman. Bootstrapping Regression Models. *The Annals of Statistics*, 9(6):1218 – 1228, 1981.
- [FV06] Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis: theory and practice*. Springer, 2006.
- [GN99] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Monographs and Surveys in Pure and Applied Mathematics. Taylor & Francis, 1999.
- [Her17] P. Hermann. Prozesskonvergenz bezüglich der supremum norm in der funktionalen datenanalyse. Masterarbeit, Philipps-Universität Marburg, 2017.
- [Kle13] A. Klenke. *Wahrscheinlichkeitstheorie*. Masterclass. Springer Berlin Heidelberg, 2013.

- [Kön03] K. Königsberger. *Analysis 1*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2003.
- [Kön06] K. Königsberger. *Analysis 2*. Number Bd. 2 in Springer-Lehrbuch. Springer Berlin Heidelberg, 2006.
- [Mal72] C. L. Mallows. A Note on Asymptotic Joint Normality. *The Annals of Mathematical Statistics*, 43(2):508 – 515, 1972.
- [MV13] R. Meise and D. Vogt. *Einführung in die Funktionalanalysis*. vieweg studium; Aufbaukurs Mathematik. Vieweg+Teubner Verlag, 2013.
- [Nad65] É. A. Nadaraya. On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, 10(1):186–190, 1965.
- [New91] Whitney Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59:1161–67, 02 1991.
- [Pie07] A. Pietsch. *History of Banach Spaces and Linear Operators*. Birkhäuser Boston, Boston, MA, 2007.
- [Pol84] D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer New York, 1984.
- [Pol90] D. Pollard. *Empirical Processes: Theory and Applications*. Conference Board of the Mathematical Science: NSF-CBMS regional conference series in probability and statistics. Institute of Mathematical Statistics, 1990.
- [Ram82] James Ramsay. When data are functions. *Psychometrika*, 47:379–396, 02 1982.
- [Rei19] C. Reihl. Schwache konvergenz des kovarianzschätzers banachraumwertiger funktionaler daten. Masterarbeit, Universität Bayreuth, 2019.
- [RS07] J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer New York, 2007.
- [Sac13] M. Sachs. *Wahrscheinlichkeitsrechnung und Statistik: für Ingenieurstudenten an Fachhochschulen*. Mathematik-Studienhilfen. Carl Hanser Verlag GmbH & Company KG, 2013.
- [Sch11] Klaus D. Schmidt. *Maß und Wahrscheinlichkeit*. Springer-Verlag, Berlin Heidelberg New York, 2011.

- [Sil84] B. W. Silverman. Spline Smoothing: The Equivalent Variable Kernel Method. *The Annals of Statistics*, 12(3):898 – 916, 1984.
- [SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press, 2014.
- [ST12] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer Series in Statistics. Springer New York, 2012.
- [Tal14] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2014.
- [Tao14] Tao, T. Metric entropy analogues of sum set theory. <https://terrytao.wordpress.com/2014/03/19/metric-entropy-analogues-of-sum-set-theory>, 2014. (Zuletzt aufgerufen am 17.11.2023).
- [Tik93] V. M. Tikhomirov. ϵ -Entropy and ϵ -Capacity of Sets In Functional Spaces, pages 86–170. Springer Netherlands, Dordrecht, 1993.
- [Tsy09] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, 2009.
- [vdVvdVW96] AW van der Vaart, A. van der Vaart, and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.
- [Wai19] M.J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.
- [Was06] L. Wasserman. *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer New York, 2006.
- [WJ94] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.
- [WM08] S. Walker and P. Mieghem. On lower bounds for the largest eigenvalue of a symmetric matrix. *Linear Algebra and Its Applications - LINEAR ALGEBRA APPL*, 429:519–526, 07 2008.
- [You76] A. P. Youschkevitch. The concept of function up to the middle of the 19th century. 16(1):37–85, March 1976.