

Asymptotische Äquivalenz eines Poisson-Modells und eines Gauß-Modells

Dissertation

zur

Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Rostock

Vorgelegt von

Christian Warnken,

geb. am 17.09.1982 in Saarbrücken

Erstgutachter:

Prof. Dr. A. Meister

Zweitgutachter:

Prof. Dr. M. Reiß (HU Berlin)

Tag der Abgabe:

02. Februar 2022

Tag der mündlichen Prüfung:

06. Mai 2022

Danksagung

Ich möchte mich an dieser Stelle bei Herrn Prof. Dr. Meister für die gute Betreuung bedanken. Seine Unterstützung und Ideengabe halfen mir, diese Arbeit zu vollenden. Auch möchte ich mich bei meiner Freundin Michelle und meiner Mutter bedanken, die immer für mich da waren, mir Mut zusprachen und mir in allen Lebenslagen beistanden.

gewidmet Max

Inhaltsverzeichnis

0 Aufbau	4
1 Der Le Cam-Abstand	5
1.1 Einleitung	5
1.2 Defizienz und der Le Cam-Abstand	6
1.3 Eigenschaften des Le Cam-Abstands	12
1.4 Likelihoodprozesse	14
1.5 Suffizienz	15
1.6 Einige Resultate	18
2 Dekonvolution	21
2.1 Einleitung	21
2.2 Wiederholte Messungen	26
3 Hauptresultat	31
3.1 Einführung	31
3.2 Abstand Poisson und Diskretnormal	33
3.3 Diskret zu stetig	42
Kurzer Ausblick	53
Literaturverzeichnis	54

0 Aufbau

Im ersten Kapitel werden wir die Begriffe der Defizienz, Suffizienz und den Le Cam-Abstand einführen. Des Weiteren werden wir die asymptotische Äquivalenz betrachten und einige Resultate verschiedener Paper vorstellen. Dieses Kapitel basiert hauptsächlich auf dem Paper von *Mariucci* (2016), in dem sich auch noch weitere Resultate finden lassen.

Das zweite Kapitel beschäftigt sich mit Dekonvolutionsproblemen und insbesondere mit Modellen, in denen wiederholbare Beobachtungen möglich sind. Es orientiert sich an dem Buch von *Meister* (2009).

In Kapitel drei wird das Resultat dieser Dissertation, die asymptotische Äquivalenz eines Poisson-Modells und eines Gauß-Modells vorgestellt.

Abschließend geben wir einen kurzen Ausblick für künftige Studien.

1 Der Le Cam-Abstand

1.1 Einleitung

Die mathematische Statistik beschäftigt sich mit statistischen Modellen, die auch statistische Experimente oder kurz Experimente genannt werden. Wir beginnen zunächst mit der Definition und werden anschließend an einem Beispiel erläutern, wie Experimente praktisch aussehen.

Definition 1.1 Sei Ω eine beliebige Menge und \mathcal{A} eine σ -Algebra auf dieser Menge. Das Tupel (Ω, \mathcal{A}) heißt *Messraum* oder *Ereignisraum*. Sei weiter $(P_\theta : \theta \in \Theta)$ eine Familie von Wahrscheinlichkeitsmaßen auf (Ω, \mathcal{A}) , wobei Θ der *Parameterraum* ist, also die Menge aller Parameter θ von P_θ . Ein mathematisches Experiment definieren wir als Tripel $\mathcal{P} = (\Omega, \mathcal{A}, (P_\theta : \theta \in \Theta))$.

Diese Definition wurde ursprünglich in *Blackwell* (1951) formuliert und mit ihr können wir konkrete Experimente beschreiben, wie in dem folgenden Beispiel aus dem Buch *Le Cam and Yang* (2000):

Ein Physiker möchte die Halbwertszeit von Carbon 14, C^{14} , bestimmen. Er nimmt an, dass die Überlebenswahrscheinlichkeit eines C^{14} -Atoms exponentialverteilt ist mit Parameter θ und betrachtet eine Probe von n C^{14} -Atomen. Er setzt eine feste Zeit fest, sagen wir zwei Stunden, und zählt die Anzahl der Zerfälle. Dies führt uns nun zur formalen Definition des statistischen Modells $\mathcal{P}_1 = (\mathbb{N}, \mathfrak{P}(\mathbb{N}), (P_\theta : \theta \in (0, \infty)))$, wobei P_θ die Verteilung der Zufallsvariablen X ist, die die Anzahl der Zerfälle innerhalb der zwei Stunden zählt. Aber das ist nicht der einzige Weg, die Halbwertszeit zu bestimmen. Der Physiker könnte auch die Zeit Y messen, bis zum Beispiel 10^6 Zerfälle beobachtet wurden. In diesem Fall lautet das statistische Modell $\mathcal{P}_2 = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), (Q_\theta : \theta \in (0, \infty)))$, wobei Q_θ die Verteilung von Y ist.

1.2 Defizienz und der Le Cam-Abstand

Wir können uns nun die Frage stellen, wie viel statistische Information die beiden Experimente aus unserem Beispiel enthalten oder, ob das Experiment \mathcal{P}_1 mehr Information als \mathcal{P}_2 oder \mathcal{P}_2 mehr Information als \mathcal{P}_1 enthält. Diese Frage nach der Vergleichbarkeit von Experimenten stellten erstmals *Bohnenblust, Shapley and Sherman* (1946) in ihrem Paper vor, gefolgt von den Papers von *Blackwell* (1951) und (1953). Zur Beantwortung dieser Frage verglichen sie das mathematische Risiko beider Experimente, zu dessen Formulierung wir noch die folgenden Definitionen benötigen:

Gegeben sei ein mathematisches Experiment $\mathcal{P} = (\Omega, \mathcal{A}, (P_\theta : \theta \in \Theta))$.

Definition 1.2 Ein Entscheidungsraum (D, \mathcal{D}) ist ein Messraum, dessen σ -Algebra \mathcal{D} alle einelementigen Mengen enthält. Für jedes $z \in D$ ist also die Menge $\{z\} \in \mathcal{D}$. Häufig wird hierbei $D = \Theta$ gewählt.

Definition 1.3 Eine Verlustfunktion $L : \Theta \times D \rightarrow (-\infty, \infty]$ ist eine Funktion, die uns den Verlust $L(\theta, z)$ angibt, nachdem wir Entscheidung $z \in D$ getroffen haben, wenn θ der wahre Parameter ist.

Definition 1.4 Seien $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$ zwei Messräume. Ein Markovkern K ist eine Abbildung $K : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ mit den Eigenschaften:

- Die Abbildung $x \mapsto K(x, A)$ ist \mathcal{A}_1 -messbar für jedes $A \in \mathcal{A}_2$.
- Die Abbildung $A \mapsto K(x, A)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{A}_2)$ für jedes $x \in \Omega_1$.

Wir bezeichnen $K : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ auch als Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$.

Definition 1.5 Eine randomisierte Entscheidungsfunktion p in \mathcal{P} ist eine Funktion, die durch einen Markovkern $p : (\Omega, \mathcal{A}) \rightarrow (D, \mathcal{D})$ gegeben ist.

Mit diesen Definitionen können wir nun das mathematische Risiko formulieren:

Definition 1.6 Das mathematische Risiko ist definiert als

$$R_\theta(\mathcal{P}, p, L) = \int_{\Omega} \left(\int_D L(\theta, z) p(y, dz) \right) P_\theta(dy).$$

Wir können uns das Risiko so vorstellen: Ein Statistiker beobachtet den Wert $x \in \Omega$, den er von einem Wahrscheinlichkeitsmaß P_θ erhält. Er kennt den wahren Wert von θ nicht, muss aber eine Entscheidung $z \in D$ treffen. Dies tut er, indem er ein Wahrscheinlichkeitsmaß $p(x, \cdot)$ auf D wählt und sich dann einen zufälligen Punkt bzgl. $p(x, \cdot)$ aus D aussucht. Wenn er sich nun für z entschieden hat, die wahre Verteilung von x aber P_θ ist, erleidet er einen Verlust $L(\theta, z)$. Der durchschnittliche Verlust, wenn er x beobachtet hat, ist also $\int L(\theta, z)p(x, dz)$. Wenn er x nun gemäß P_θ gewählt hat, ist der durchschnittliche Gesamtverlust $\int (\int L(\theta, z)p(x, dz)) P_\theta(dx)$.

Um zwei Experimente miteinander vergleichen zu können, sagen wir \mathcal{P}_1 ist informativer als \mathcal{P}_2 , falls für jede beschränkte Verlustfunktion L , $\|L\|_\infty \leq 1$ und falls für jede Entscheidungsfunktion p_2 in dem Experiment \mathcal{P}_2 eine Entscheidungsfunktion p_1 in dem Experiment \mathcal{P}_1 existiert, so dass

$$R_\theta(\mathcal{P}_1, p_1, L) \leq R_\theta(\mathcal{P}_2, p_2, L), \quad \forall \theta \in \Theta$$

gilt.

Diese klassische Definition im Sinne der statistischen Entscheidungstheorie kann allerdings dazu führen, dass zwei Modelle nicht miteinander vergleichbar sind. Le Cam löste dieses Problem, indem er den Begriff der Defizienz $\delta(\mathcal{P}_1, \mathcal{P}_2)$ einführt. Sie ist eine wohldefinierte nichtnegative reellwertige Zahl für zwei Experimente \mathcal{P}_1 und \mathcal{P}_2 , die sich einen gemeinsamen Parameterraum teilen. Zu der genauen Definition kommen wir später.

Wie wir bereits gesehen haben, können wir also zwei gegebene statistische Modelle mit gleichem Parameterraum miteinander vergleichen, indem wir ihre entsprechenden Risikofunktionen miteinander vergleichen oder uns fragen, wie viel Information verloren geht, wenn wir von einem Modell in das andere wechseln. Wir sagen, dass keine Information verloren geht, wenn wir Beobachtungen von $P_{1,\theta}$ in Beobachtungen von $P_{2,\theta}$ umwandeln können. Dies geschieht mittels Markovkernen.

Nehmen wir nun einen Markovkern $K : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ und ein Wahrscheinlichkeitsmaß P_1 auf $(\Omega_1, \mathcal{A}_1)$, dann können wir ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{A}_2)$ wie folgt konstruieren:

$$KP_1(A) = \int K(x, A)P_1(dx), \quad \forall A \in \mathcal{A}_2.$$

Grob gesprochen können wir sagen, dass zwei Modelle \mathcal{P}_1 und \mathcal{P}_2 dieselbe Information bezüglich θ enthalten, wenn zwei Markovkerne K_1 und K_2 existieren, die unabhängig von θ sind, so dass gilt $K_1 P_{1,\theta} = P_{2,\theta}$ und $K_2 P_{2,\theta} = P_{1,\theta}$. Diese Idee wurde in den Sechzigern von Lucien Le Cam formalisiert und führte zu dem Begriff der Defizienz und damit zur Einführung einer Pseudometrik auf der Klasse aller statistischen Experimente mit demselben Parameterraum.

Die Definition der Defizienz in der allgemeinsten Form beinhaltet den Begriff der Transitionen, die eine Verallgemeinerung von Markovkernen darstellen. In dieser Arbeit beschränken wir uns allerdings auf dominierte statistische Modelle mit polnischen Ergebnisräumen. Der Vorteil besteht darin, dass sich die Defizienz einfacher definieren lässt und das abstrakte Konzept der Transitionen mit dem der Markovkerne übereinstimmt (siehe Nussbaum (1996), Proposition 9.2). Bevor wir fortfahren, benötigen wir einige Definitionen.

Definition 1.7 Ein statistisches Modell $\mathcal{P} = (\Omega, \mathcal{A}, (P_\theta : \theta \in \Theta))$ heißt *polnisch*, wenn sein Ergebnisraum (Ω, \mathcal{A}) ein separabler und vollständig metrisierbarer topologischer Raum ist.

\mathcal{P} heißt *dominiert*, falls ein σ -endliches Maß μ auf (Ω, \mathcal{A}) existiert, so dass für alle $\theta \in \Theta$, P_θ absolut stetig bezüglich μ ist. Das Maß μ heißt dann *dominiertes Maß*.

Beispiel 1.8 Typische Beispiele von polnischen Räumen sind endliche oder abzählbar unendliche diskrete Räume, \mathbb{R}^n ($n \in \mathbb{N}$) mit seiner natürlichen Topologie oder auch der Raum C_T der stetigen Funktionen auf $[0, T]$, versehen mit der Supremumsnorm $d(x, y) = \sup_{0 \leq t \leq T} |x_t - y_t|$.

Definition 1.9 Seien P_1 und P_2 zwei Wahrscheinlichkeitsmaße auf einem Messraum Ω , die von einem Maß ξ dominiert werden, mit dazugehörigen Dichten $g_i = \frac{dP_i}{d\xi}$, $i = 1, 2$. Der L_1 -Abstand ist definiert als

$$L_1(P_1, P_2) = \int_{\Omega} |g_1(x) - g_2(x)| \xi(dx).$$

Definition 1.10 Seien P_1 und P_2 zwei Wahrscheinlichkeitsmaße auf einem Messraum Ω . Der Totalvariationsabstand zwischen P_1 und P_2 lässt sich definieren als

$$\begin{aligned} \|P_1 - P_2\|_{TV} &= \sup_{A \subseteq \Omega} |P_1(A) - P_2(A)| \\ &= \frac{1}{2} \sup_{\varphi} |\mathbb{E}\varphi(P_1) - \mathbb{E}\varphi(P_2)| \quad \text{mit } |\varphi| \leq 1 \text{ messbar} \\ &= \frac{1}{2} L_1(P_1, P_2). \end{aligned}$$

Definition 1.11 Seien $\mathcal{P}_i = (\Omega_i, \mathcal{A}_i, (P_{i,\theta} : \theta \in \Theta))$, $i = 1, 2$ zwei Experimente. Die Defizienz $\delta(\mathcal{P}_1, \mathcal{P}_2)$ von \mathcal{P}_1 bezüglich \mathcal{P}_2 ist definiert als

$$\delta(\mathcal{P}_1, \mathcal{P}_2) = \inf_K \sup_{\theta \in \Theta} \|K P_{1,\theta} - P_{2,\theta}\|_{TV},$$

wobei das Infimum über alle Markovkerne $K : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ genommen wird.

Definition 1.12 Der Le Cam-Abstand oder Δ -Abstand zwischen \mathcal{P}_1 und \mathcal{P}_2 ist definiert als

$$\Delta(\mathcal{P}_1, \mathcal{P}_2) = \max(\delta(\mathcal{P}_1, \mathcal{P}_2), \delta(\mathcal{P}_2, \mathcal{P}_1)).$$

Der Δ -Abstand ist eine Pseudometrik auf dem Raum aller statistischen Modelle, er erfüllt die Dreiecksungleichung $\Delta(\mathcal{P}_1, \mathcal{P}_3) \leq \Delta(\mathcal{P}_1, \mathcal{P}_2) + \Delta(\mathcal{P}_2, \mathcal{P}_3)$, aber aus $\Delta(\mathcal{P}_1, \mathcal{P}_2) = 0$ folgt nicht, dass \mathcal{P}_1 und \mathcal{P}_2 übereinstimmen.

Wenn $\delta(\mathcal{P}_1, \mathcal{P}_2) = 0$ gilt, d.h. wenn sich das Experiment \mathcal{P}_2 aus Experiment \mathcal{P}_1 mittels eines Markovkerns rekonstruieren lässt, nennen wir das Experiment \mathcal{P}_2 *weniger informativ* als \mathcal{P}_1 , bzw. nennen wir das Experiment \mathcal{P}_1 *informativer* als \mathcal{P}_2 . Wenn $\Delta(\mathcal{P}_1, \mathcal{P}_2) = 0$ gilt, nennen wir \mathcal{P}_1 und \mathcal{P}_2 *alent* und zwei Folgen von statistischen Modellen $(\mathcal{P}_{1,n})_{n \in \mathbb{N}}$ und $(\mathcal{P}_{2,n})_{n \in \mathbb{N}}$ heißen *asymptotisch äquivalent*, wenn $\Delta(\mathcal{P}_{1,n}, \mathcal{P}_{2,n}) \rightarrow 0$ für $n \rightarrow \infty$.

Wir können den Le Cam-Abstand als eine Art nummerischen Indikator verstehen, der angibt, wie viel es kostet, ein Modell mit Hilfe von Markovkernen mit einem anderen zu rekonstruieren. Das folgende Theorem (siehe Le Cam (1986), Theorem 2) erlaubt es uns, den Begriff der Defizienz in die Sprache der Entscheidungstheorie zu übertragen.

Theorem 1.13 Sei $\epsilon > 0$ fest. Es gilt $\delta(\mathcal{P}_1, \mathcal{P}_2) < \epsilon$ genau dann, wenn für alle Entscheidungsfunktionen p_2 in \mathcal{P}_2 und für alle beschränkten Verlustfunktionen L , $\|L\|_\infty \leq 1$ eine Entscheidungsfunktion p_1 in \mathcal{P}_1 existiert, so dass

$$R_\theta(\mathcal{P}_1, p_1, L) < R_\theta(\mathcal{P}_2, p_2, L) + \epsilon, \quad \forall \theta \in \Theta.$$

In anderen Worten gilt

$$\delta(\mathcal{P}_1, \mathcal{P}_2) = \inf_{p_1} \sup_{p_2} \sup_{\theta} \sup_L |R(\mathcal{P}_1, p_1, L, \theta) - R(\mathcal{P}_2, p_2, L, \theta)|,$$

wobei das letzte Supremum über der Menge aller Verlustfunktionen L mit $0 \leq L(\theta, z) \leq 1$, $\forall z \in D$, $\forall \theta \in \Theta$ gebildet wird und die p_i zu der Menge aller randomisierten Entscheidungsfunktionen in den Experimenten \mathcal{P}_i , $i = 1, 2$ gehören.

Eine entscheidende Konsequenz aus diesem Theorem ist, dass, wenn zwei Folgen von Experimenten $(\mathcal{P}_{1,n})_{n \in \mathbb{N}}$ und $(\mathcal{P}_{2,n})_{n \in \mathbb{N}}$ asymptotisch äquivalent im Sinne von Le Cam sind, die asymptotischen Eigenschaften die gleichen sind. Wenn wir also gezeigt haben, dass zwei Experimente asymptotisch äquivalent sind, können wir uns das einfacher aussuchen und dort unser Problem untersuchen. Das Ergebnis unserer Untersuchung können wir dann mittels Markovkernen auf das schwerere Problem übertragen.

Durch die Einführung der Defizienz wird auch das oben beschriebene Problem der Vergleichbarkeit gelöst. Es kann passieren, dass beide $\delta(\mathcal{P}_1, \mathcal{P}_2)$ und $\delta(\mathcal{P}_2, \mathcal{P}_1)$ echt positiv und nach der ursprünglichen Definition nicht vergleichbar sind. Nichtsdestotrotz können

wir immer noch sagen, wie viel Information wir verlieren, wenn wir von einem Modell zum anderen wechseln. Die Le Cam-Theorie hat Anwendung in vielen Problemen der Entscheidungstheorie gefunden, wie zum Beispiel in der nichtparametrischen Regression, den nichtparametrischen Dichteschätzproblemen, verallgemeinerten linearen Modellen, Lévy-Modellen und Spektraldichteschätzproblemen. Die ersten Ergebnisse der asymptotischen Äquivalenz im nichtparametrischen Zusammenhang stammen aus dem Jahr 1996 von *Brown and Low* (1996) und *Nussbaum* (1996). In dem ersten Paper wurde die asymptotische Äquivalenz der nichtparametrischen Regression und einem Gaußschen Weißes-Rauschen-Modell gezeigt und in dem zweiten die eines Dichteschätzungsproblems und eines Gaußschen Weißes-Rauschen-Modells. Über die Jahre wurden viele Verallgemeinerungen dieser Ergebnisse veröffentlicht, zum Beispiel *Brown et al.* (2002), *Carter* (2006), (2007), (2009), *Grama and Nussbaum* (2002), *Meister and Reiß* (2013), *Reiß* (2008) und im Bereich der nichtparametrischen Regression *Brown et al.* (2004), *Carter* (2002) und *Jähnisch and Nussbaum* (2003).

Wie in *Mariucci* (2016) gezeigt wird, können wir Entscheidungsfunktionen mit Hilfe von Randomisierung transferieren. Dazu seien $\mathcal{P}_{i,n} = (\Omega_{i,n}, \mathcal{A}_{i,n}, (P_{i,n,\theta} : \theta \in \Theta))$, $i = 1, 2$ zwei Folgen statistischer Modelle mit selbem Parameterraum Θ und polnischen Stichprobenräumen $(\Omega_{i,n}, \mathcal{A}_{i,n})$. Nehmen wir an, es existieren Markovkerne K_n , so dass $\|K_n P_{1,n,\theta} - P_{2,n,\theta}\|_{TV} \rightarrow 0$ gleichmäßig auf unserem Parameterraum. Mit einer gegebenen Entscheidungsfunktion (oder einem Schätzer) $\pi_{2,n}$ auf $\mathcal{P}_{2,n}$ können wir eine Entscheidungsfunktion $\pi_{1,n}$ auf $\mathcal{P}_{1,n}$ definieren, die asymptotisch das gleiche statistische Risiko wie $\pi_{2,n}$ besitzt. Um dies zu zeigen, betrachten wir zunächst den einfachen Fall, in dem sowohl K_n als auch $\pi_{2,n}$ deterministisch sind. Das bedeutet, K_n hat die Gestalt $K_n(A) = \mathbf{1}_A S_n(x)$ für alle $A \in \mathcal{A}_{2,n}$ und Funktionen S_n . Es gilt

$$\begin{aligned} & \left| \int_{\Omega_{1,n}} L(\theta, \pi_{1,n}(y)) P_{1,n,\theta}(dy) - \int_{\Omega_{2,n}} L(\theta, \pi_{2,n}(y)) P_{2,n,\theta}(dy) \right| \\ & \leq \left| \int_{\Omega_{1,n}} L(\theta, \pi_{1,n}(y)) P_{1,n,\theta}(dy) - \int_{\Omega_{2,n}} L(\theta, \pi_{2,n}(y)) K P_{1,n,\theta}(dy) \right| \\ & \quad + \left| \int_{\Omega_{2,n}} L(\theta, \pi_{2,n}(y)) [K P_{1,n,\theta}(dy) - P_{2,n,\theta}(dy)] \right| \\ & \leq \left| \int_{\Omega_{1,n}} L(\theta, \pi_{1,n}(y)) P_{1,n,\theta}(dy) - \int_{\Omega_{1,n}} L(\theta, \pi_{2,n}(S(y))) P_{1,n,\theta}(dy) \right| \\ & \quad + \|L\|_\infty \|K P_{1,n,\theta} - P_{2,n,\theta}\|_{TV}. \end{aligned}$$

Unsere Verlustfunktion L ist durch eins beschränkt und setzen wir

$$\pi_{1,n}(y) := \pi_{2,n}(S(y)),$$

erhalten wir

$$\left| \int_{\Omega_{1,n}} L(\theta, \pi_{1,n}(y)) P_{1,n,\theta}(dy) - \int_{\Omega_{2,n}} L(\theta, \pi_{2,n}(y)) P_{2,n,\theta}(dy) \right| \leq \|KP_{1,n,\theta} - P_{2,n,\theta}\|_{TV} \rightarrow 0,$$

das bedeutet, die Entscheidungsfunktion $\pi_{1,n}(y) = \pi_{2,n}(S(y))$ besitzt asymptotisch das gleiche Risiko wie $\pi_{2,n}$. Die gleiche Rechnung funktioniert auch, wenn die K_n nicht deterministisch sind und $(\pi_{2,n})$ eine Folge von Entscheidungsfunktionen mit Entscheidungsräumen (D_n, \mathcal{D}_n) sind. In diesem Fall können wir zeigen, dass die randomisierte Folge von Entscheidungsfunktionen

$$\pi_{1,n}(y, D) := \int_{\Omega_{2,n}} \pi_{2,n}(x, D) K(y, dx), \quad \forall y \in \Omega_{1,n}, \quad \forall D \in \mathcal{D}_n$$

das gleiche asymptotische Risiko wie $\pi_{2,n}$ besitzt. □

1.3 Eigenschaften des Le Cam-Abstands

Auch wenn die Definition der Defizienz eine genau umschriebene statistische Bedeutung hat, lässt sie sich schwer berechnen. Explizite Berechnungen sind selten, man findet sie z.B. in *Hansen and Torgersen* (1974), *Torgersen* (1972, 1974) und in Sektion 1.9 in *Shiryayev and Spokoiny* (2000). Wir versuchen deswegen, einfachere obere Schranken des Δ -Abstandes zu finden. Eine Auswahl wird im Folgenden vorgestellt.

Lemma 1.14 Seien $\mathcal{P}_i = (\Omega, \mathcal{A}, (P_{i,\theta} : \theta \in \Theta))$, $i = 1, 2$ zwei statistische Modelle mit demselben Stichprobenraum und wir definieren $\Delta_0(\mathcal{P}_1, \mathcal{P}_2) := \sup_{\theta \in \Theta} \|P_{1,\theta} - P_{2,\theta}\|_{TV}$. Dann gilt

$$\Delta(\mathcal{P}_1, \mathcal{P}_2) \leq \Delta_0(\mathcal{P}_1, \mathcal{P}_2).$$

Beweis: Klar, nach Definition des Δ -Abstands. \square

Definition 1.15 Seien P_1 und P_2 zwei Wahrscheinlichkeitsmaße auf Ω , die von einem Maß ξ dominiert werden, mit dazugehörigen Dichten $g_i = \frac{dP_i}{d\xi}$, $i = 1, 2$. Der Hellingerabstand zwischen P_1 und P_2 ist definiert als

$$H(P_1, P_2) = \left(\frac{1}{2} \int_{\Omega} \left(\sqrt{g_1(x)} - \sqrt{g_2(x)} \right)^2 \xi(dx) \right)^{\frac{1}{2}}.$$

In der Literatur wird der Faktor $\frac{1}{\sqrt{2}}$ gelegentlich auch weggelassen, da man häufig, wie auch in dieser Arbeit, an einer Konvergenz gegen null interessiert ist. Der Faktor ist so gewählt, dass gilt

$$0 \leq H(P, Q) \leq 1 \text{ und } 0 \leq \|P - Q\|_{TV} \leq 1$$

mit

$$H(P, Q) = 0 \Leftrightarrow P = Q \Leftrightarrow \|P - Q\|_{TV} = 0$$

und

$$H(P, Q) = 1 \Leftrightarrow P \perp Q \Leftrightarrow \|P - Q\|_{TV} = 1.$$

Weiter gilt

$$H^2(P_1, P_2) \leq \|P_1 - P_2\|_{TV} = \frac{1}{2} L_1(P_1, P_2) \leq \sqrt{2} H(P_1, P_2)$$

sowie

$$H^2(P_1, P_2) = \frac{1}{2} \int_{\Omega} (g_1 + g_2 - 2\sqrt{g_1 g_2}) \xi(dx) = 1 - \int_{\Omega} \sqrt{g_1 g_2} \xi(dx). \quad (1.1)$$

Für den quadrierten Hellingerabstand gilt dann folgende Abschätzung:

Lemma 1.16 Seien P und Q zwei Produktmaße, die auf demselben messbaren Raum definiert sind, $P = \otimes_{j=1}^m P_j$ und $Q = \otimes_{j=1}^m Q_j$ mit dazugehörigen Dichten p_j und q_j , dann gilt

$$H^2(P, Q) = 1 - \prod_{j=1}^m (1 - H^2(P_j, Q_j)) \leq \sum_{j=1}^m H^2(P_j, Q_j).$$

Beweis: Nach (1.1) und dem Satz von Fubini gilt

$$\begin{aligned} H^2(\mu, \nu) &= 1 - \int_{\Omega} \prod_{j=1}^m \sqrt{p_j(x)} \prod_{j=1}^m \sqrt{q_j(x)} \xi(dx) \\ &= 1 - \prod_{j=1}^m \int_{\Omega} \sqrt{p_j(x)q_j(x)} \xi(dx) \\ &= 1 - \prod_{j=1}^m (1 - H^2(P_j, Q_j)) \\ &\leq 1 - \left(1 - \sum_{j=1}^m H^2(P_j, Q_j)\right) = \sum_{j=1}^m H^2(P_j, Q_j). \end{aligned} \quad \square$$

Wir können den quadrierten Hellingerabstand auch für konkrete Verteilungen direkt ausrechnen.

Beispiel 1.17 Seien P und Q normalverteilt mit $P \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $Q \sim \mathcal{N}(\mu_2, \sigma_2^2)$, dann gilt

$$H^2(P, Q) = 1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right).$$

Beispiel 1.18 Seien P und Q poissonverteilt mit Parametern α bzw. β , dann gilt

$$H^2(P, Q) = 1 - \exp\left(-\frac{1}{2}(\sqrt{\alpha} - \sqrt{\beta})^2\right).$$

1.4 Likelihoodprozesse

Einen anderen Weg, den Le Cam-Abstand in den Griff zu bekommen, bietet die enge Beziehung zwischen der Äquivalenz zweier Experimente und der Verteilung der zugehörigen Likelihoodquotienten.

Sei $\mathcal{P} = (\Omega, \mathcal{A}, (P_\theta : \theta \in \Theta))$ ein statistisches Modell, das von P_{θ_0} , $\theta_0 \in \Theta$ dominiert wird und sei $\Lambda(\theta) = \frac{dP_\theta}{dP_{\theta_0}}$ die Dichte von P_θ bezüglich P_{θ_0} . Speziell können wir $\Lambda(\theta)$ als eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, (P_{\theta_0} : \theta_0 \in \Theta))$ betrachten, d.h. wir können $(\Lambda(\theta))_{\theta \in \Theta}$ als einen stochastischen Prozess auffassen. Aus diesem Grund führen wir die Notation $\Lambda_{\mathcal{P}} := (\Lambda(\theta), \theta \in \Theta)$ ein und nennen $\Lambda_{\mathcal{P}}$ den Likelihoodprozess. Ein zentrales Ergebnis aus der Le Cam-Theorie ist das folgende Lemma:

Lemma 1.19 *Seien $\mathcal{P}_j = (\Omega_j, \mathcal{A}_j, (P_{j,\theta} : \theta \in \Theta))$, $j = 1, 2$ zwei Experimente. Wenn die Familie $(P_{j,\theta} : \theta \in \Theta)$ von P_{j,θ_0} dominiert wird, dann sind \mathcal{P}_1 und \mathcal{P}_2 genau dann äquivalent, wenn ihre Likelihoodprozesse unter den dominierenden Maßen P_{1,θ_0} und P_{2,θ_0} übereinstimmen.*

Beweis: siehe Strasser (1985), Korollar 25.9. □

Seien $(\Lambda_j^{n,*}(\theta))_{\theta \in \Theta}$, $j = 1, 2$ zwei Prozesse, die auf demselben Wahrscheinlichkeitsraum $(\Omega^*, \mathcal{A}^*, \Pi^*)$ definiert sind und die Verteilung von $(\Lambda_j^n(\theta))_{\theta \in \Theta}$ unter P_{j,θ_0} äquivalent zu der Verteilung von $(\Lambda_j^{n,*}(\theta))_{\theta \in \Theta}$ unter Π^* , $j = 1, 2$ ist. Dann gilt nach Le Cam and Yang (2000), Lemma 6:

Lemma 1.20 *Wenn $\Lambda_{\mathcal{P}_1}$ und $\Lambda_{\mathcal{P}_2}$ Likelihoodprozesse der Experimente \mathcal{P}_1 und \mathcal{P}_2 sind, gilt*

$$\Delta(\mathcal{P}_1^n, \mathcal{P}_2^n) \leq \sup_{\theta \in \Theta} \mathbb{E}_{\Pi^*} |\Lambda_1^{n,*}(\theta) - \Lambda_2^{n,*}(\theta)|.$$

1.5 Suffizienz

Wenn wir statistische Modelle mit unterschiedlichem Stichprobenraum miteinander vergleichen möchten, kann es hilfreich sein, sich suffiziente Statistiken anzuschauen. Die Einführung des Begriffs der suffizienten Statistik wird gewöhnlich R.A. Fisher zugeschrieben, der verschiedene Definitionen des Konzeptes vorstellt. Le Cam schreibt in *Le Cam* (1964), Fishers wichtigste Aussage sei, dass eine gewählte Statistik alle wichtigen Informationen zusammenfassen soll, die die Stichprobe liefert. Diese Anforderung können wir in unterschiedlichen Weisen präzisieren. Die folgenden drei Interpretationen sind die geläufigsten.

- i) Die klassische Definition der Suffizienz:

Eine Statistik S ist suffizient, wenn wir durch Randomisation Zufallsvariablen reproduzieren können, die die gleiche Verteilung wie die ursprünglich beobachteten Variablen haben.

- ii) Die Bayessche Interpretation:

Eine Statistik S ist suffizient, wenn für jede a priori Verteilung des Parameters die a posteriori Verteilung des Parameters gegeben S dieselbe ist, wie wenn das gesamte Resultat gegeben wäre.

- iii) Das entscheidungstheoretische Konzept:

Eine Statistik S ist suffizient, wenn für jedes Entscheidungsproblem und jede Entscheidungsfunktion, die durch das Experiment gegeben ist, eine Entscheidungsfunktion existiert, die nur auf S beruht und dieselben Eigenschaften besitzt.

Eine abstrakte Betrachtung der Suffizienz wird in *Halmos and Savage* (1949) beschrieben. Dort wird auch das Problem diskutiert, dass nicht immer klar sei, was mit ‚allen wichtigen Informationen‘ gemeint sei. Eine Fortführung dieser Arbeit findet sich in *Bahadur* (1954). Ein Hauptresultat dort ist das Theorem 7.1, dass die Äquivalenz des entscheidungstheoretischen Konzepts der Suffizienz mit dem des Konzepts der bedingten Wahrscheinlichkeiten betrachtet. Wir erwähnen dies, da es ähnlich dem Resultat von Le Cam (hier in Theorem 1.13 vorgestellt) ist, welches eine Kernaussage in der Theorie des Vergleichens von statistischen Experimenten ist.

Sei $\mathcal{P} = (\Omega, \mathcal{A}, (P_\theta : \theta \in \Theta))$ ein statistisches Modell. Eine Statistik S ist eine messbare Abbildung von einem messbaren Raum (Ω, \mathcal{A}) in einen anderen messbaren Raum (Ψ, \mathcal{B}) . Wir bezeichnen mit $S_\# P_\theta$ die Bildverteilung von S unter P_θ , d.h. $S_\# P_\theta(B) = P_\theta(S^{-1}(B))$ für alle $B \in \mathcal{B}$.

Definition 1.21 S ist eine suffiziente Statistik für $(P_\theta : \theta \in \Theta)$, wenn für jedes $A \in \mathcal{A}$

eine Funktion ϕ_A existiert, wobei $\phi_A \circ S$ \mathcal{A} -messbar ist, so dass

$$P_\theta(A \cap S^{-1}(B)) = \int_B \phi_A(y) S_\# P_\theta(dy), \quad \forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \forall \theta \in \Theta.$$

Eine beliebige Subalgebra \mathcal{A}_0 von \mathcal{A} heißt suffizient für $(P_\theta : \theta \in \Theta)$, wenn für alle $A \in \mathcal{A}$ eine \mathcal{A}_0 -messbare Funktion ϕ_A existiert, so dass

$$P_\theta(A \cap A_0) = \int_{A_0} \phi_A(y) P_\theta(dy), \quad \forall A_0 \in \mathcal{A}_0, \forall \theta \in \Theta.$$

Die Menge $\{S^{-1}(B) : B \in \mathcal{B}\}$ wird als Subalgebra bezeichnet, die von der Statistik S induziert wird.

Lemma 1.22 Eine Statistik S ist suffizient für $(P_\theta : \theta \in \Theta)$, wenn die Subalgebra, die von S induziert wird, suffizient für $(P_\theta : \theta \in \Theta)$ ist.

Beweis: siehe z.B. Bahadur (1954) □

In Übereinstimmung mit der Notation aus Abschnitt 1.2 stellen wir nun Theorem 7.1 aus Bahadur (1954) vor, wobei (D, \mathcal{D}) den Entscheidungsraum bezeichnet.

Theorem 1.23 Wenn die Subalgebra \mathcal{A}_0 von \mathcal{A} für $(P_\theta : \theta \in \Theta)$ suffizient ist, dann existiert für jede Entscheidungsfunktion $\rho : (\Omega, \mathcal{A}) \rightarrow (D, \mathcal{D})$ eine Entscheidungsfunktion $\pi : (\Omega, \mathcal{A}_0) \rightarrow (D, \mathcal{D})$, so dass

$$P_\theta \rho(C) = P_\theta \pi(C), \quad \forall C \in \mathcal{D}, \forall \theta \in \Theta.$$

Bevor wir uns dem Zusammenhang zwischen suffizienten Statistiken und der Äquivalenz von statistischen Modellen zuwenden, führen wir die Neyman-Fisher-Faktorisation ein. Sie stellt ein mächtiges Werkzeug zur Identifizierung von suffizienten Statistiken in dominierten Wahrscheinlichkeitsfamilien dar.

Theorem 1.24 Sei $(P_\theta : \theta \in \Theta)$ eine Familie von Wahrscheinlichkeiten auf (Ω, \mathcal{A}) , die absolut stetig bezüglich des σ -endlichen Maßes μ ist und bezeichnen mit $p_\theta := \frac{dP_\theta}{d\mu}$ die Dichte. Eine Statistik $S : (\Omega, \mathcal{A}) \rightarrow (\Psi, \mathcal{B})$ ist suffizient für $(P_\theta : \theta \in \Theta)$ genau dann, wenn eine \mathcal{B} -messbare Funktion $g_\theta \forall \theta \in \Theta$ und eine \mathcal{A} -messbare Funktion $h \neq 0$ existieren, so dass

$$p_\theta(x) = g_\theta(S(x))h(x), \quad \mu\text{-fast sicher } \forall x \in \Omega$$

gilt.

Ein wichtiges Resultat, um den Le Cam-Abstand mit der Existenz von suffizienten Statistiken zu verknüpfen, ist folgendes Lemma:

Lemma 1.25 *Seien $\mathcal{P}_i = (\Omega_i, \mathcal{A}_i, (P_{i,\theta} : \theta \in \Theta))$, $i = 1, 2$ zwei statistische Modelle. Sei $S : (\Omega_1 \rightarrow \Omega_2)$ eine suffiziente Statistik, so dass die Verteilung von S unter $P_{1,\theta}$ die gleiche wie $P_{2,\theta}$ ist. Dann gilt $\Delta(\mathcal{P}_1, \mathcal{P}_2) = 0$.*

Beweis: (siehe z.B. Mariucci (2015)) Um zu zeigen, dass $\delta(\mathcal{P}_1, \mathcal{P}_2) = 0$ gilt, reicht es, sich den Markovkern $M : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ mit $M(x, B) := \mathbb{1}_B(S(x)) \forall x \in \Omega_1$ und $\forall B \in \mathcal{A}_2$ anzuschauen.

Andersherum, um $\delta(\mathcal{P}_2, \mathcal{P}_1) = 0$ zu zeigen, betrachten wir den Markovkern $K : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_1, \mathcal{A}_1)$ mit $K(y, A) = \mathbb{E}_{P_{2,\theta}}(\mathbb{1}_A | S = y) \forall y \in \Omega_2, \forall A \in \mathcal{A}_1$. Da S eine suffiziente Statistik ist, ist der Markovkern K unabhängig von θ . Bezeichnen wir nun mit $S_\# P_{1,\theta}$ die Verteilung von S unter $P_{1,\theta}$, dann gilt

$$K P_{2,\theta}(A) = \int K(y, A) P_{2,\theta}(dy) = \int \mathbb{E}_{P_{2,\theta}}(\mathbb{1}_A | S = y) S_\# P_{1,\theta}(dy) = P_{1,\theta}(A)$$

und somit $\delta(\mathcal{P}_2, \mathcal{P}_1) = 0$. □

Für asymptotische Betrachtungsweisen benötigen wir eine entsprechende Version der Definition der Suffizienz.

Definition 1.26 *Sei $\mathcal{P}_n = (\Omega_n, \mathcal{A}_n, (P_{n,\theta} : \theta \in \Theta))$ eine Folge von statistischen Modellen. Die Folge der Subalgebren $\tilde{\mathcal{A}}_n$ von \mathcal{A}_n heißt asymptotisch suffizient für $(P_{n,\theta} : \theta \in \Theta)$, wenn $\Delta(\mathcal{P}_n, \mathcal{P}_{n|\tilde{\mathcal{A}}}) \rightarrow 0$, wobei $\mathcal{P}_{n|\tilde{\mathcal{A}}}$ die Beschränkung des Experiments \mathcal{P}_n auf $\tilde{\mathcal{A}}_n$ bezeichnet, d.h. $\mathcal{P}_{n|\tilde{\mathcal{A}}} = (\Omega_n, \tilde{\mathcal{A}}_n, (\tilde{P}_{n,\theta} : \theta \in \Theta))$ mit $\tilde{P}_{n,\theta}(A) = P_{n,\theta}(A)$ für alle $A \in \tilde{\mathcal{A}}_n$.*

Seien $\mathcal{P}_{1,n}$ und $\mathcal{P}_{2,n}$ zwei Folgen von Experimenten mit demselben Parameterraum. Wenn es nun zwei Folgen $S_{1,n}$ und $S_{2,n}$ von asymptotisch suffizienten Statistiken in $\mathcal{P}_{1,n}$ bzw. in $\mathcal{P}_{2,n}$ gibt, die Werte im selben messbaren Raum annehmen, so dass

$$\sup_{\theta \in \Theta} \|S_{1,n\#} P_{1,\theta} - S_{2,n\#} P_{2,\theta}\|_{TV} \rightarrow 0, \text{ für } n \rightarrow \infty$$

gilt, dann sind die Folgen $\mathcal{P}_{1,n}$ und $\mathcal{P}_{2,n}$ asymptotisch äquivalent.

Darüber hinaus gibt es neben dem Begriff der Suffizienz auch noch den allgemeineren Begriff der Insuffizienz. Dessen Betrachtung geht uns in dieser Arbeit jedoch zu weit und wir verweisen auf Kapitel 5 in Le Cam (1986).

1.6 Einige Resultate

Um die Resultate im Bereich der asymptotischen Äquivalenz besser verstehen zu können, werden wir einige Beispiele präsentieren. Zu Beginn betrachten wir einen einfachen parametrischen Fall.

Beispiel 1.27 Sei $\mathcal{P}_{1,n}$ ein statistisches Modell und der zugehörige Vektor X enthalte die Beobachtungen von n unabhängigen normalverteilten Zufallsvariablen X_i mit $X_i \sim \mathcal{N}(\theta, 1)$, $i = 1, \dots, n$. Der Parameterraum Ω sei ein Intervall auf \mathbb{R} und es gilt

$$\mathcal{P}_{1,n} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{1,\theta} : \theta \in \Theta)),$$

wobei $P_{1,\theta}$ die Verteilung von X ist.

Sei $\mathcal{P}_{2,n}$ das Experiment, das den empirischen Erwartungswert der Beobachtungen aus $\mathcal{P}_{1,n}$ betrachtet, d.h.

$$\mathcal{P}_{2,n} = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (P_{2,\theta} : \theta \in \Theta)),$$

wobei $P_{2,\theta}$ die Verteilung der normalverteilten Zufallvariablen Y mit $Y \sim \mathcal{N}(\theta, \frac{1}{n})$ ist.

Mit der Neyman-Fisher-Faktorisation sieht man leicht, dass $S : \mathbb{R}^n \rightarrow \mathbb{R}$,

$S(x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i}{n}$ eine suffiziente Statistik ist und mit Definition 1.26 folgt, dass $\Delta(\mathcal{P}_{1,n}, \mathcal{P}_{2,n}) = 0$ für alle $n \in \mathbb{N}$.

Bevor wir einige nichtparametrische Arbeiten vorstellen, wenden wir uns Carters Arbeit zu, die ein ähnliches Problem wie diese Arbeit behandelt. Er zeigt die asymptotische Äquivalenz eines multinomialen und eines Gaußschen multivariaten Experiments. Sie stellt ein nützliches Werkzeug dar, wenn wir asymptotische Äquivalenz in Dichteschätzungsproblemen betrachten.

Beispiel 1.28 Sei $X = (X_1, \dots, X_m)$ ein Zufallsvektor mit multinomialverteilten Einträgen mit Parameter n und (p_1, \dots, p_m) mit $p_i \geq 0$ für $i = 1, \dots, m$ und $\sum_{i=1}^m p_i = 1$. \mathcal{P} sei nun das statistische Modell mit multinomialer Verteilung $P_\theta = \mathcal{M}(n; (\theta_1, \dots, \theta_m))$ mit Parametern $\theta = (\theta_1, \dots, \theta_m)$, die zu $\Theta_R \subset \mathbb{R}^m$ gehören, einer Menge, die alle Zufallsvektoren enthält, für die

$$\frac{\max_i \theta_i}{\min_i \theta_i} \leq R$$

gilt.

Das Hauptergebnis in Carter (2002) ist eine Schranke des Le Cam-Abstands zwischen multinomialverteilten Experimenten und multivariat normalverteilten Experimenten mit denselben Erwartungswerten und Kovarianzen. Genauer, sei \mathcal{Q} ein statistisches Modell mit dazugehörigen multivariaten Normalverteilungen $Q_\theta = \mathcal{N}(\mu, \Sigma)$, $\theta \in \Theta_R$ mit

$$\mu = (n\theta_1, \dots, n\theta_m), \quad \Sigma = (\sigma_{i,j})_{i,j=1, \dots, m}, \quad \sigma_{i,j} = n\theta_i(1 - \theta_i)\delta_{i=j} - n\theta_i\theta_j\delta_{i \neq j}.$$

Mit obiger Notation gilt dann folgendes Theorem:

Theorem 1.29 (siehe Carter (2002), Seite 709) Es gilt

$$\Delta(\mathcal{P}, \mathcal{Q}) \leq C_R \frac{m \ln m}{\sqrt{n}}$$

mit einer Konstanten C_R , die nur von R abhängt.

Ein weiteres interessantes Resultat in Carter (2002) beschreibt die Approximation von \mathcal{Q} durch ein Gaußexperiment mit unabhängigen Koordinaten. Sei $\tilde{\mathcal{Q}}$ ein statistisches Modell, das m unabhängige normalverteilte Zufallsvariablen X_i mit $X_i \sim \mathcal{N}(\sqrt{\theta_i}, 1/(4n))$, $i = 1, \dots, m$ enthält.

Theorem 1.30 (siehe Carter (2002), Seite 717-719) Es gilt

$$\Delta(\mathcal{Q}, \tilde{\mathcal{Q}}) \leq C_R \frac{m}{\sqrt{n}}$$

mit einer Konstante C_R , die nur von R abhängt.

Zum Schluss dieses Kapitels stellen wir noch zwei Beispiele im nichtparametrischen Rahmen vor. Das erste ist ein Ergebnis aus Brown and Low (1996) und das zweite aus Nussbaum (1996), welche die ersten Resultate asymptotischer Äquivalenz für nichtparametrische Experimente sind.

Beispiel 1.31 In Brown and Low (1996) wird das Problem behandelt, eine Funktion f aus Beobachtungen eines stetigen Gaußschen Prozesses $y(t)$, $t \in [0, 1]$ zu schätzen, für den die stochastische Differenzialgleichung

$$dy_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dW_t, \quad t \in [0, 1]$$

gilt, wobei dW_t ein Gaußsches Weißes-Rauschen ist. Das Resultat zeigt, dass das statistische Modell mit den stetigen Beobachtungen (y_t) asymptotisch äquivalent zu demjenigen mit diskreten Beobachtungen ist, d.h.

$$y_i = f(t_i) + \sigma(t_i)\xi_i, \quad i = 1, \dots, n.$$

Es gilt $t_i = \frac{i-1}{n}$ und die ξ_i sind standardnormalverteilte Zufallsvariablen. Es wird angenommen, dass f in einer nichtparametrischen Teilmenge \mathcal{F} von $L_2[0, 1]$ lebt. Zusätzlich

werden einige Glättungseigenschaften gefordert und, dass n nicht zu langsam gegen unendlich geht. Die Driftfunktion f ist unbekannt und für eine positive Konstante B gilt:

$$\sup\{|f(t)| : t \in [0, 1], f \in \mathcal{F}\} = B < \infty.$$

Weiter wird gefordert, dass

$$\bar{f}_n(t) = \begin{cases} f(\frac{i}{n}), & \text{falls } \frac{i-1}{n} \leq t < \frac{i}{n}, \quad i = 1, \dots, n \\ f(1), & \text{falls } t = 1 \end{cases}$$

und

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} n \int_0^1 \frac{(f(t) - \bar{f}_n(t))^2}{\sigma^2(t)} dt = 0.$$

Der Diffusionskoeffizient $\sigma^2(\cdot) > 0$ ist eine bekannte absolut stetige Funktion auf $[0, 1]$ mit

$$\left| \frac{d}{dt} \ln \sigma(t) \right| \leq C, \quad t \in [0, 1]$$

mit positiver Konstante C .

Beispiel 1.32 In Nussbaum (1996) wird eine globale asymptotische Äquivalenz eines Dichteschätzungsproblems mit unabhängig identisch verteilten Zufallsgrößen und die eines Gaußschen Weißes-Rauschen-Modells betrachtet. Genauer formuliert, seien $(Y_i)_{i=1}^n$ unabhängig identisch verteilte Zufallsvariablen mit Dichte f auf $[0, 1]$ bezüglich des Lebesgue-Maßes. Die Dichten f seien die unbekannten Parameter und gehören zu einer nichtparametrischen Klasse \mathcal{F} , für die die Hölderbedingung

$$|f(x) - f(y)| \leq C|x - y|^\alpha, \quad \alpha > \frac{1}{2}$$

gilt sowie

$$f(x) \geq \epsilon > 0.$$

Wir bezeichnen mit $\mathcal{P}_{1,n}$ das statistische Modell bezüglich der Beobachtungen der Y_i und mit $\mathcal{P}_{2,n}$ das Experiment, in dem wir den stochastischen Prozess $(y_t)_{t \in [0,1]}$ beobachten mit

$$d_{y_t} = \sqrt{f(t)} dt + \frac{1}{2\sqrt{n}} dW_t, \quad t \in [0, 1],$$

wobei $(W_t)_{t \in [0,1]}$ eine Standard-Brownsche-Bewegung ist. Das Hauptresultat ist, dass $\Delta(\mathcal{P}_{1,n}, \mathcal{P}_{2,n}) \rightarrow 0$ für $n \rightarrow \infty$. Im Beweis, ohne tiefer ins Detail zu gehen, wird gezeigt, dass das Resultat für gewisse Teilmengen $\mathcal{F}_n(f_0)$ der Klasse \mathcal{F} gilt und man diese dann passend zusammensetzen kann. Bei den wichtigen Schritten werden Poissonisationstechniken und KMT-Ungleichungen verwendet.

2 Dekonvolution

2.1 Einleitung

In diesem Kapitel wollen wir uns mit Dekonvolutionsproblemen beschäftigen. Diese tauchen in vielen Feldern der nichtparametrischen Statistik auf, zum Beispiel bei Dichteschätzungsproblemen mit kontaminierten Daten, nichtparametrischer Regression mit errors-in-variables und bei Bild- und Signalentzerrung. Während der letzten drei Jahrzehnte haben diese Themen mehr und mehr Aufmerksamkeit gewonnen. Als Anwendungsbereiche zählen Probleme in der Ökonometrie, Biometrie, medizinischen Statistik oder der Bildwiederherstellung.

Unser Ziel ist es, eine Funktion f zu schätzen, zu der wir nur über

$$h = f * G = \int f(x - y)dG(y)$$

Zugang haben, sprich über die Faltung von f mit einem Wahrscheinlichkeitsmaß G . Wir können f also durch Beobachtungen nur indirekt schätzen.

Häufig gehen wir davon aus, dass wir die Verteilung von G kennen, die die Lebesgue-Dichte g , die sogenannte Fehlerdichte, besitzt. Unser entsprechendes neues Modell lautet dann

$$h = f * g = \int f(x - y)g(y)dy,$$

wobei f und g reellwertige Funktionen sind, die auf \mathbb{R} abbilden.

Ein nützliches Hilfsmittel für unsere Dekonvolutionsprobleme ist die Fouriertransformation. Sie erlaubt es uns, Faltungen als einfache Multiplikationen zu interpretieren, wie wir in Lemma 2.2 ii) sehen werden. Dazu benötigen wir zunächst:

Definition 2.1 Sei $f \in L_1(\mathbb{R})$. Die Fouriertransformierte f^{ft} ist definiert als

$$f^{ft}(t) = \int \exp(itx)f(x)dx, \quad t \in \mathbb{R}.$$

Das folgende Lemma zeigt einige wichtige Eigenschaften:

Lemma 2.2 Seien $f, g \in L_1(\mathbb{R})$, $\lambda, \mu \in \mathbb{C}$, dann gilt

- i) Linearität: $(\lambda f + \mu g)^{ft} = \lambda f^{ft} + \mu g^{ft}$,
- ii) Faltung: $(f * g)^{ft} = f^{ft}g^{ft}$,
- iii) Beschränktheit: $\sup_{t \in \mathbb{R}} |f^{ft}(t)| \leq \|f\|_1$,
- iv) Symmetrie: Falls f reellwertig ist, gilt $f^{ft}(-t) = \overline{f^{ft}(t)}$ $\forall t \in \mathbb{R}$ und falls f zusätzlich symmetrisch ist, gilt $f^{ft}(-t) = f^{ft}(t)$ $\forall t \in \mathbb{R}$.

Beweis:

- i) Ergibt sich direkt aus der Linearität des Integrals.
- ii) Mit Hilfe von Substitution und dem Satz von Fubini erhalten wir

$$\begin{aligned} (f * g)^{ft}(t) &= \int \exp(itx) \int f(y)g(x-y)dydx \\ &= \int \exp(ity)f(y) \int \exp(it(x-y))g(x-y)dxdy \\ &= \int \exp(ity)f(y)dy \int \exp(ity)g(y)dy = f^{ft}(t)g^{ft}(t). \end{aligned}$$

iii) Indem wir den Betrag in das Integral ziehen, erhalten wir die Behauptung.

iv) Es gilt

$$f^{ft}(-t) = \int \exp(-itx)f(x)dx = \int \overline{\exp(itx)}f(x)dx = \overline{f^{ft}(t)}$$

und falls f symmetrisch ist, gilt

$$\begin{aligned} f^{ft}(t) &= \int_{-\infty}^0 \exp(itx)f(x)dx + \int_0^\infty \exp(itx)f(x)dx \\ &= \int_0^\infty \exp(-itx)f(x)dx + \int_0^\infty \exp(itx)f(x)dx \\ &= 2 \int_0^\infty \cos(tx)f(x)dx \in \mathbb{R}, \quad \forall t \in \mathbb{R}. \end{aligned}$$

Da der Kosinus und f symmetrisch sind, folgt die Behauptung. □

Um unsere ursprüngliche Funktion f zurückzuerhalten, benötigen wir die inverse Fouriertransformation.

Satz 2.3 Sei $f \in L_1(\mathbb{R})$ beschränkt und stetig für fast alle $x \in \mathbb{R}$ und $f^{ft} \in L_1(\mathbb{R})$, dann gilt

$$f(x) = \frac{1}{2\pi} \int \exp(-itx) f^{ft}(t) dt.$$

Beweis: siehe z.B. Meister (2009)

□

Wir sehen, dass sich die inverse Fouriertransformation und die Fouriertransformation ähneln und nur im Vorzeichen des Arguments und im Skalierungsfaktor $1/(2\pi)$ unterscheiden. In der Literatur gibt es auch alternative Definitionen. Dort wird bei der Fouriertransformierten der Skalierungsfaktor $1/(2\pi)$ oder $1/\sqrt{2\pi}$ eingefügt und das Vorzeichen des Arguments umgekehrt, womit sich auch die Inverse entsprechend ändert.

Wir können, unter Zuhilfenahme der Fouriertransformation, unser ursprüngliches Problem wie folgt umformulieren:

$$h = f * G \Leftrightarrow h^{ft} = f^{ft} \cdot G^{ft}.$$

Unsere Strategie besteht nun darin, in einem ersten Schritt h^{ft} über empirische Daten zu schätzen. Sei \hat{h}^{ft} die empirische Version von h^{ft} . Der zweite Schritt besteht darin, $\hat{h}^{ft}(t)$ zu berechnen und durch $G^{ft}(t)$ zu teilen, was uns den Schätzer $\hat{f}^{ft}(t)$ liefert. Wenn G bekannt ist, können wir G^{ft} berechnen, ansonsten müssen wir einen anderen Zugang zu G finden. Im letzten Schritt wenden wir auf \hat{f}^{ft} die inverse Fouriertransformation an und erhalten \hat{f} . Dieses \hat{f} wählen wir als unseren Dekonvolutionsschätzer von f .

All diese Schritte müssen gegebenenfalls noch angepasst werden, zum Beispiel muss gewährleistet sein, dass die inverse Fouriertransformation von \hat{f}^{ft} überhaupt existiert.

In der Praxis wollen wir häufig die Dichtefunktion f mit entsprechender Verteilung F schätzen, indem wir reellwertige Zufallsvariablen X_1, \dots, X_n beobachten, wobei jedes X_j die Verteilung F besitzt. Dazu lassen sich verschiedene Problemstellungen finden, zum Beispiel:

- i) In der Ökonometrie:

Wir sind an der Verteilung F des Einkommens von Angestellten interessiert. Die Zufallsvariablen X_1, \dots, X_n entsprechen dem Monatsgehalt von n zufällig ausgewählten Angestellten.

ii) In der Biometrik:

Wir sind an der Verteilung F der Größe einer Tierart interessiert. Wir messen dazu n Individuen, die wir aus einer Gruppe einzeln entnommen und wieder freigelassen haben. Die Zufallsvariablen X_1, \dots, X_n beschreiben unsere Messungen.

iii) In statistischen Erhebungen:

Wir sind an der Verteilung F des Verkehrsaufkommens einer bestimmten Straße interessiert. Hier entsprechen unsere Zufallsvariablen X_1, \dots, X_n der Anzahl der vorbeikommenden Autos an verschiedenen Tagen.

In all diesen Beispielen gehen wir davon aus, dass unsere Zufallsvariablen X_1, \dots, X_n unabhängig und identisch verteilt sind. Stochastische Unabhängigkeit der Zufallsvariablen kann dahingehend interpretiert werden, dass sich die Variablen untereinander nicht beeinflussen. Im zweiten Beispiel könnte es passieren, dass wir ein Tier doppelt messen, womit wir gegen die Unabhängigkeit verstößen würden. Gegen die identische Verteilung könnte im dritten Beispiel verstößen werden, wenn wir ein Mal die Anzahl der Autos an Werktagen beobachten und ein anderes Mal an Feiertagen. Es gibt Gebiete der Stochastik, die sich auch mit diesen Problemen beschäftigen. In dieser Arbeit werden wir es jedoch nur mit unabhängigen und identisch verteilten Zufallsvariablen zu tun haben.

Die obigen Beispiele sind natürlich recht simpel und lassen den Fakt außer Betracht, dass jede Messung oder Beobachtung fehlerbehaftet ist. Es ist uns nur möglich, fehlerbehaftete Daten Y_1, \dots, Y_n zu beobachten, statt der wahren Daten X_1, \dots, X_n . Unser sogenanntes additives Fehlermodell sieht dann wie folgt aus:

$$Y_j = X_j + \epsilon_j \quad j \in \{1, \dots, n\}, \quad (2.1)$$

wobei X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen sind und wir wollen die Dichte f von X_1 schätzen. Die ebenfalls unabhängigen und identisch verteilten Zufallsvariablen $\epsilon_1, \dots, \epsilon_n$ beschreiben den Fehler in unseren Beobachtungen und besitzen die Fehlerdichte g . Darüber hinaus seien X_j und ϵ_j unabhängig und reellwertig. Häufig gehen wir davon aus, g zu kennen.

In unserem hypothetischen Beispiel mit den Tiergrößen könnte es zum Beispiel passieren, dass wir das Maßband nicht exakt ablesen können, wenn sich das Tier etwas windet. Hier könnten wir annehmen, dass g normalverteilt ist und wir hätten ein recht einfaches additives Fehlermodell.

Ein praktisches Beispiel liefert das National Health and Nutrition Examination Survey (NHANES I). In dieser Langzeitstudie wollte man wissen, wie die Verteilung der Aufnahme von gesättigten Fettsäuren aussieht. *Stefanski and Carroll* (1990) stellen einen Dekonvolutionsansatz zu dieser Studie zur Verfügung. NHANES I war eine der ersten Studien, in der nichtparametrische Dekonvolutionstechniken angewendet wurden.

In *Dunn* (2004) und *Turner* (1986) werden Daten von 50 Psychiatriepatienten verwendet, deren Köpfe mit einem Computertomographen untersucht wurden. Dabei wurde das Hirnventrikelverhältnis (engl. VBR) bei jedem Patienten doppelt mit einem Planimeter, einem mechanischen Messgerät zur Flächeninhaltsbestimmung, gemessen. Diese Messmethode ist fehlerbehaftet und wenn die Fehlerdichte g bekannt wäre, könnten wir die Verteilung des VBR eines zufällig ausgewählten Patienten mit unseren Dekonvolutionstechniken bestimmen. In diesem Beispiel erhalten wir allerdings einen empirischen Zugang zu g , da die Messungen mehrfach durchgeführt wurden.

Ein weiteres Beispiel zu wiederholten Messungen liefert uns *Bissantz et al.* (2007). In diesem Paper wird das Problem, das Verhältnis von Eisen zu Wasserstoff eines Sterns zu schätzen, um Aufschluss über den Anteil an schweren Elementen zu erhalten, behandelt. Das Verhältnis wird durch die Helligkeit bestimmter Spektralbänder bestimmt, wobei schwierige Kalibrierungen nötig sind. Der Fehler, der hierbei entsteht, kann als additives Rauschen interpretiert werden und ist durch wiederholte Beobachtungen desselben Sterns empirisch zugänglich.

Wie wir an den letzten beiden Beispielen gesehen haben, kann es von Vorteil sein, wenn man ein Phänomen wiederholt messen kann. Dies bringt uns zu unserem nächsten Kapitel.

2.2 Wiederholte Messungen

Anstatt die Zufallsvariable X_j , wie anfangs beschrieben, nur ein Mal zu beobachten, können wir sie auch mehrfach beobachten, wobei jede Beobachtung auch wieder fehlerbehaftet ist. Wir ändern unser additives Modell dahingehend ab, dass wir nun mehrere Beobachtungen $Y_{j,k}$, $j = 1, \dots, n$, $k = 1, \dots, m_j$, $m_j \geq 2$ der Zufallsvariablen X_j erhalten. Es ergibt sich dann

$$Y_{j,k} = X_j + \epsilon_{j,k}.$$

Unser Ziel ist es noch immer, die Dichte f von X_1 zu schätzen, dabei seien alle zugehörigen Zufallsvariablen $X_{\bullet}, \epsilon_{\bullet,\bullet}$ unabhängig und jedes $\epsilon_{j,k}$ habe die Fehlerdichte g .

Die Fehlerdichte g ist, im Gegensatz zu der in unserem Standardmodell 2.1, in diesem Modell identifizierbar. Wieso dies so ist, erläutern wir im Folgenden. Wir betrachten zur Vereinfachung den kleinstmöglichen Fall, d.h. wir wählen $m_j \equiv 2$ für alle j . Nun bilden wir die Differenz unserer Beobachtungen

$$\Delta Y_j = Y_{j,1} - Y_{j,2} = \epsilon_{j,1} - \epsilon_{j,2}, \quad j = 1, \dots, n \quad \text{fast sicher.}$$

Wie wir sehen, wurde die Zufallsvariable X_j eliminiert und ΔY_j ist messbar bezüglich der σ -Algebra, die von $\epsilon_{j,1}$ und $\epsilon_{j,2}$ erzeugt wird. Die charakteristische Funktion von ΔY_j ist

$$\psi_{\Delta Y_j}(t) = \mathbb{E} \exp(it\Delta Y_j) = \mathbb{E} \exp(it\epsilon_{j,1}) \cdot \mathbb{E} \exp(-it\epsilon_{j,2}) = g^{ft}(t)g^{ft}(-t) = |g^{ft}(t)|^2.$$

Wenn wir annehmen, dass g symmetrisch bezüglich der y -Achse ist, ist g^{ft} reellwertig und nach Lemma 2.2 iv) gilt

$$g^{ft}(-t) = g^{ft}(t), \quad \text{für alle } t \in \mathbb{R}.$$

Wenn wir zusätzlich die Standardannahme treffen, dass g^{ft} nicht verschwindet, kann man leicht sehen, dass

$$g^{ft}(t) > 0, \quad \text{für alle } t \in \mathbb{R}$$

gilt und daher ist es sinnvoll

$$\hat{g}^{ft}(t) = \left| \frac{1}{n} \sum_{j=1}^n \exp(it\Delta Y_j) \right|^{\frac{1}{2}}$$

als Schätzer für $g^{ft}(t)$ zu wählen, wobei die empirische charakteristische Funktion von ΔY_j verwendet wird. Wenn wir davon ausgehen, dass g^{ft} reellwertig ist, können wir

uns auf den reellen Teil beschränken, also den Term $\exp(it\Delta Y_j)$ durch $\cos(t\Delta Y_j)$ ersetzen. Wir sollten nun auch mehr als zwei Beobachtungen von X_j berücksichtigen und verallgemeinern unseren Fall auf $m_j > 2$. Dafür definieren wir

$$\hat{g}^{ft}(t) = \left| \frac{1}{N} \sum_{j=1}^n \sum_{(k_1, k_2) \in K_j} \cos(t\Delta Y_j) \right|^{\frac{1}{2}}, \quad (2.2)$$

wobei wir mit K_j die Menge aller Paare $(k_1, k_2) \in \mathbb{Z}^2$ mit $1 \leq k_1 < k_2 \leq m_j$ bezeichnen. K_j enthält $m_j(m_j - 1)/2$ Elemente, wie wir durch leichte Berechnung sehen können. Wir setzen $N = \sum_{j=1}^n m_j(m_j - 1)/2$, was der Anzahl der Summanden in der Definition des Schätzers (2.2) entspricht. Es kann passieren, dass einige der m_j gleich eins sind, dann wäre nur eine Beobachtung von X_j gegeben. Diese Beobachtungen werden jedoch bei der Konstruktion von (2.2) einfach ignoriert.

Während der Gebrauch wiederholter Messungen in der Dichtedekonvolution als erstes in *Horowitz and Markatou* (1996) in einem ökonometrischen Zusammenhang erwähnt wurde, haben viele neuere Papers diesen Ansatz übernommen, um die Fehlerdichte identifizierbar zu machen. Einige Beispiele hierzu sind *Delaigle, Hall and Meister* (2008), *Schennach* (2004a), *Schennach* (2004b), wobei der Schätzer (2.2) von der letzteren Arbeit übernommen wurde. In *Li and Vuong* (1998) wurde die Forderung der Symmetrie von g gelockert und eine eher kompliziertere Rekonstruktion von f^{ft} vorgestellt. Eine weitere Rekonstruktion von f wird in *Hall and Yao* (2003) behandelt. *Neumann* (2007) zeigt, dass f nicht identifizierbar ist, wenn für eine Dichte g , sowohl g als auch die verschobene Funktion $g(\cdot - a)$, $a > 0$ als wahre Fehlerdichten zugelassen sind. In Bezug auf unseren Schätzer stellen wir fest, dass ΔY_j die gleiche Verteilung für beide konkurrierenden Fehlerdichten besitzt, so dass g nicht wirklich auf Basis der Differenz zweier Messungen identifiziert werden kann. Auf der anderen Seite zeigt *Neumann* (2007) Konsistenz für einen allgemeineren Fall, in dem $\epsilon_{j,k}$ zentriert ist.

Ein weiterer großer Vorteil der wiederholten Messungen ist die Handhabung von Fehlerdichten g , deren Fouriertransformierte Oszillationen aufweisen und Nullstellen besitzen. Dies wurde in *Meister and Neumann* (2008) untersucht. In dem Modell (2.1) ohne wiederholte Messungen führen isolierte Nullstellen von g^{ft} in vielen Fällen unweigerlich zu langsameren Konvergenzraten, außer die Ordnung der Nullstellen ist klein verglichen mit der Glattheit β unter restriktiveren Glattheitsannahmen oder es werden restriktive Annahmen an die Momente von f getroffen. Im Modell mit wiederholten Messungen ändert sich die Situation drastisch. Zum Beispiel bei der Dekonvolution von gleichverteilten Fehlerdichten und ihren Faltungen sind wir in der Lage, die gleichen Konvergenzraten zu erreichen wie für gewöhnlich glatte g mit entsprechenden Fourier-Tails, ohne stärkere Bedingungen für die Dichte zu fordern; eine gewöhnliche Sobolevannahme

$$\int |g^{ft}(t)|^2 |t|^{2\beta} dt \leq C, \quad \text{mit } C > 0, \beta > 0$$

kann verwendet werden. Gewöhnlich glatt bedeutet in diesem Zusammenhang, dass für g gilt

$$C_1(1+|t|)^{-\alpha} \leq |g^{ft}(t)| \leq C_2(1+|t|)^{-\alpha}, \quad \text{für alle } t \in \mathbb{R} \quad (2.3)$$

mit Konstanten $C_2 > C_1 > 0$ und $\alpha > 0$.

Der Hauptgrund für dieses Phänomen ist die charakteristische Funktion der bivariaten Beobachtungen $(Y_{j,1}, Y_{j,2})$, die gleich $f^{ft}(t_1 + t_2)g^{ft}(t_1)g^{ft}(t_2)$ ist oder äquivalent $f^{ft}(t)g^{ft}(t_1)g^{ft}(t-t_1)$, so dass wir im Fourierbereich durch $g^{ft}(t_1)g^{ft}(t-t_1)$ teilen müssen und immer noch die Möglichkeit haben, den Parameter $t_1 \in \mathbb{R}$ frei zu wählen. Bei Fehlerdichten, die nur isolierte Nullstellen besitzen, kann man t_1 dann für jedes t so wählen, dass der Nenner hinreichend weit von null entfernt ist. Wenn wir nur eine Beobachtung Y_j haben, ist ihre charakteristische Funktion $f^{ft}(t)g^{ft}(t)$ und somit kein Parameter wählbar, um $f^{ft}(t)$ zu schätzen. Es handelt sich hierbei natürlich nur um eine intuitive Erklärung, wo die Unterschiede der beiden Modelle liegen.

In Modellen mit mehrfacher Beobachtung versucht man häufig, die Anzahl der Beobachtungen durch Mittelwertbildung zu reduzieren. Das bedeutet, dass wir unserem Schätzer die Beobachtungen $(\bar{Y}_1, \dots, \bar{Y}_n)$ zugrunde legen mit

$$\bar{Y}_j = \frac{1}{m_j} \sum_{k=1}^{m_j} Y_{j,k}, \quad k = 1, \dots, m_j, \quad j = 1, \dots, n \quad (2.4)$$

anstatt der Originaldaten $Y_{\bullet,\bullet}$. Es ergibt sich also

$$\bar{Y}_j = X_j + \frac{1}{m_j} \sum_{k=1}^{m_j} \epsilon_{j,k}.$$

Nicht nur, dass diese Mittelwertbildung die Varianz der Fehlerkomponenten reduziert, sie macht die Fehlerdichten glatter, wenn diese gewöhnlich glatt, wie in (2.3), sind. Wenn zum Beispiel $m_j \equiv 2$ ist, fällt die charakteristische Funktion der Zufallsvariablen $\frac{1}{2} \sum_{k=1}^2 \epsilon_{j,k}$ genau wie $|t|^{-2\alpha}$ ab. Dabei wird, solange die m_j bezüglich des Stichprobenumfangs n beschränkt sind, der Glattheitsgrad mit der Anzahl der Beobachtungen, über die gemittelt wurde, multipliziert. Das bedeutet allerdings, dass die Mittelwertbildung die Konvergenzraten schlechter macht. Daher ist es günstiger nur eine Reihe der Daten, zum Beispiel $Y_{\cdot,1}$, für den Dekonvolutionsschritt zu verwenden.

Wenn auf der anderen Seite die Fehlerdichte g normalverteilt ist, ändert sich die Situation. Normalverteilte Dichten sind abgeschlossen bezüglich der Faltung, da die Faltung zweier normalverteilter Dichten wieder normalverteilt ist. Hier erhöht sich der Glättungsgrad von g durch Mittelwertbildung nicht, aber die Varianz wird reduziert. In *Delaigle and Meister (2008)* wird unter anderem gezeigt, dass $(\bar{Y}_1, \dots, \bar{Y}_n)$ statistisch suffizient für die Dichte f ist, was die Mittelwertbildung rechtfertigt. Wir werden den Beweis hier genauer vorstellen.

Wir befinden uns in der Situation, in der die Fehlerterme homoskedastisch sind, was bedeutet, dass sich ihre Varianzen nicht signifikant unterscheiden. Darüber hinaus sind mehrfache Beobachtungen verfügbar. Unser Modell sieht also wie anfangs beschrieben aus:

$$Y_{j,k} = X_j + \epsilon_{j,k}, \quad j = 1, \dots, n, k = 1, \dots, r_{j,n}, \quad (2.5)$$

wobei $\epsilon_{j,k} \sim f_\epsilon$.

Wir werden wie in (2.4) mit gemittelten Beobachtungen $\bar{Y}_j = r_{j,n}^{-1} \sum_{k=1}^{r_{j,n}} Y_{j,k}$ arbeiten. Unser neues Modell hat also die Form

$$\bar{Y}_j = X_j + \bar{\epsilon}_j$$

mit $\bar{\epsilon}_j = r_{j,n}^{-1} \sum_{k=1}^{r_{j,n}} \epsilon_{j,k}$.

Sind die Fehlerdichten normalverteilt, d.h. $f_\epsilon = \mathcal{N}(\mu, \sigma^2)$ und $f_{\epsilon_j} = \mathcal{N}(\mu, \sigma_{j,n}^2)$ mit $\sigma_{j,n}^2 = \sigma^2/r_{j,n}$ und $f_{\epsilon_j}^{ft}(t) = f_{\epsilon_j}^{ft}(t/r_{j,n})$, dann geht keine Information verloren, wenn wir die gemittelte Stichprobe verwenden, um f_X zu schätzen.

Theorem 2.4 Sei $f_\epsilon = \mathcal{N}(\mu, \sigma^2)$ in dem Modell (2.5), dann ist die Stichprobe $\bar{Y}_1, \dots, \bar{Y}_n$ suffizient für f_X .

Beweis: Wir führen zunächst eine orthonormale $(r_{j,n} \times r_{j,n})$ -Matrix $A_{j,n}$ ein, in deren erster Zeile $r_{j,n}^{-\frac{1}{2}} \cdot (1, \dots, 1)$ steht und setzen $W_{j,\bullet} := A_{j,n} Y_{j,\bullet}$ mit $Y_{j,\bullet} := (Y_{j,1}, \dots, Y_{j,r_{j,n}})^t$. Wir sehen, dass $W_{j,1} = r_{j,n}^{\frac{1}{2}} \bar{Y}_j$ gilt. Die anderen Komponenten von $W_{j,\bullet}$ sind messbar bzgl. der σ -Algebra, die von $\epsilon_{j,1}, \dots, \epsilon_{j,r_{j,n}}$ erzeugt wird, da sich jede Zeile von $A_{j,n}$ zu null summiert, außer der ersten. Dies ergibt sich aus der orthonormalen Struktur von $A_{j,n}$. Für die Dichte $f_{Y_{j,\bullet}}$ von $Y_{j,\bullet}$ gilt

$$\begin{aligned} f_{Y_{j,\bullet}}(y_{j,\bullet}) &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{r_{j,n}} \int f_X(x) \exp \left(-\frac{\|y_{j,\bullet} - (x + \mu) \cdot (1, \dots, 1)^t\|^2}{2\sigma^2} \right) dx \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{r_{j,n}} \int f_X(x) \exp \left(-\frac{\|A_{j,n}y_{j,\bullet} - r_{j,n}^{\frac{1}{2}}(x + \mu) \cdot (1, 0, \dots, 0)^t\|^2}{2\sigma^2} \right) dx \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{r_{j,n}} \exp \left(-\frac{\sum_{k=2}^{r_{j,n}} |w_{j,k}|^2}{2\sigma^2} \right) \\ &\quad \cdot \int f_X(x) \exp \left(-\frac{|w_{j,1} - r_{j,n}^{\frac{1}{2}}(x + \mu)|^2}{2\sigma^2} \right) dx, \end{aligned}$$

wobei $\|\cdot\|$ die euklidische Norm und $w_{j,\bullet} = A_{j,n}y_{j,\bullet}$ bezeichnet. Wir sehen, dass die bedingte Verteilung von $Y_{j,\bullet}$ gegeben $W_{j,1}$ und damit die Verteilung all unserer Daten

$$dP(Y_{\bullet,\bullet} = y_{\bullet,\bullet}|W_{\bullet,1}) = \prod_{j=1}^n dP(Y_{j,\bullet} = y_{j,\bullet}|W_{j,1})$$

nicht von f_X abhängt. Somit haben wir die Suffizienz gezeigt. \square

3 Hauptresultat

3.1 Einführung

Wir werden in diesem Kapitel das Hauptresultat dieser Dissertation vorstellen. Betrachtet werden Folgen von Experimenten, deren Wahrscheinlichkeitsverteilungen gegeben der zu schätzenden Zufallsvariablen poissonverteilt sind, und Folgen von Experimenten mit normalverteilten Wahrscheinlichkeitsverteilungen. Das Ziel ist es, bei wiederholten Beobachtungen die asymptotische Äquivalenz der beiden Folgen zu zeigen. Der Beweis ist in mehrere Schritte gegliedert, und wir beginnen damit, die diskrete Normalverteilung einzuführen.

Definition 3.1 Wir nennen N eine diskret-normalverteilte Zufallsvariable mit Parametern μ und σ^2 , falls sie die Zähldichte

$$p(k) = \alpha^{-1} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right), \quad k \in \mathbb{Z} \quad (3.1)$$

besitzt mit

$$\alpha = \sum_{k \in \mathbb{Z}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right), \quad (3.2)$$

damit $\sum_{k \in \mathbb{Z}} p(k) = 1$ gilt.

Für $\alpha = 1$ stimmen die Werte der Dichte (3.1) auf den ganzen Zahlen mit denen der Normalverteilung mit Erwartungswert μ und Varianz σ^2 überein. Von dieser Eigenschaft werden wir später Gebrauch machen. Wieso wir $\alpha = 1$ verwenden können, zeigt uns das folgende Lemma:

Lemma 3.2 Seien $p_n(k)$ Zähldichten wie in (3.1) mit $\alpha = 1$. Sei weiter X_n^* eine Folge diskret-normalverteilter Zufallsvariablen mit Zähldichte $p_n^*(k)$ und α wie in (3.2) und Y_n eine Folge von beliebigen diskreten Zufallsvariablen, die eine Zähldichte $q_n(k)$ besitzen, $n \in \mathbb{N}$.

Es gilt die Abschätzung

$$\sum_{k \in \mathbb{Z}} |q_n(k) - p_n^*(k)| \leq 2 \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)|.$$

Das bedeutet, wenn

$$\sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| \rightarrow 0 \text{ für } n \rightarrow \infty$$

gilt, gilt also auch

$$\sum_{k \in \mathbb{Z}} |q_n(k) - p_n^*(k)| \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Beweis: Da $p_n^*(k) = \frac{p_n(k)}{\sum_{j \in \mathbb{Z}} p_n(j)}$ gilt, folgt

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |q_n(k) - p_n^*(k)| &\leq \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \sum_{k \in \mathbb{Z}} |p_n(k) - p_n^*(k)| \\ &= \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \sum_{k \in \mathbb{Z}} p_n(k) \cdot \left| 1 - \frac{1}{\sum_{j \in \mathbb{Z}} p_n(j)} \right| \\ &= \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \sum_{k \in \mathbb{Z}} p_n(k) \cdot \frac{|\sum_{j \in \mathbb{Z}} p_n(j) - 1|}{\sum_{j \in \mathbb{Z}} p_n(j)} \\ &= \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \left| \sum_{j \in \mathbb{Z}} p_n(j) - 1 \right| \\ &= \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \left| \sum_{j \in \mathbb{Z}} p_n(j) - \underbrace{\sum_{j \in \mathbb{Z}} q_n(j)}_{=1} \right| \\ &\leq \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| + \sum_{j \in \mathbb{Z}} |p_n(j) - q_n(j)| \\ &= 2 \sum_{k \in \mathbb{Z}} |q_n(k) - p_n(k)| \rightarrow 0 \text{ für } n \rightarrow \infty. \end{aligned}$$

□

3.2 Abstand Poisson und Diskretnormal

Wir beobachten Zufallsvariablen $Y_{j,k}$, $j = 1, \dots, n$, $k = 1, \dots, r_n$ die gegeben der Zufallsvariable X_j poissonverteilt sind, d.h. $Y_{j,k}|X_j \sim \mathcal{P}(X_j)$ und seien $N_j|X_j$ diskret-normalverteilte Zufallsvariablen mit $\mu = r_n X_j$ und $\sigma^2 = r_n X_j$ wie in Definition 3.1.

An unsere Zufallsvariablen X_j stellen wir die Bedingungen, dass sie nach oben beschränkt und nach unten von Null weg beschränkt sind, d.h. es existieren Konstanten $C_1 > 0$ und $C_2 > 0$ mit $C_1 \geq X_j > C_2 > 0$ f.s. Weiter seien die X_j kompakt getragen und beliebig oft stetig differenzierbar mit Dichten f_{X_j} . Ähnlich zu Theorem 2.4 wollen wir die gemittelten Beobachtungen Y_1, \dots, Y_n mit $Y_j := \sum_{k=1}^{r_n} Y_{j,k}$, $j = 1, \dots, n$ verwenden. Dazu benötigen wir das folgende Lemma:

Lemma 3.3 Y_1, \dots, Y_n ist suffizient für f_X mit $X = (X_1, \dots, X_n)$.

Beweis: Es gilt

$$\begin{aligned} f_{Y_{j,k}}(y_{j,k}) &= \prod_{j=1}^n \int f_X(x) \exp(-x) \frac{x^{y_{j,1}}}{y_{j,1}!} \cdots \exp(-x) \frac{x^{y_{j,r_n}}}{y_{j,r_n}!} dx \\ &= \prod_{j=1}^n \frac{1}{\prod_{k=1}^{r_n} y_{j,k}!} \int f_X(x) \exp(-r_n x) x^{\sum_{k=1}^{r_n} y_{j,k}} dx. \end{aligned}$$

Mit der Neyman-Fisher-Faktorisierung sehen wir also, dass wir mit den gemittelten Beobachtungen arbeiten können. \square

Theorem 3.4 Mit obigen Bezeichnungen gilt

$$\Delta(\mathcal{P}_{Y_j}, \mathcal{P}_{N_j}) \leq \mathcal{O}\left(h_n a_n^4 r_n + \frac{r_n}{h_n^2} + h_n \exp(-c a_n^2 r_n)\right) \rightarrow 0, \quad c > 0$$

für $h_n \rightarrow \infty$, $r_n \rightarrow \infty$, $a_n \rightarrow 0$ mit $h_n a_n^4 r_n \rightarrow 0$, $\frac{r_n}{h_n^2} \rightarrow 0$ und $a_n^2 r_n \rightarrow \infty$, $n \rightarrow \infty$.

Wir verwenden hierbei die Landau-Notation

$$\mathcal{O}(g(x)) = f(x) \iff \exists c > 0 \ \exists x_0 \geq 0 \ \forall x \geq x_0 : |f(x)| \leq c \cdot |g(x)|.$$

Beweis: Es gilt mit $|\varphi| \leq 1$ messbar

$$\begin{aligned}
\Delta(\mathcal{P}_{Y_j}, \mathcal{P}_{N_j}) &\leq \sup_{f_{X_j}} \|P_{Y_j} - P_{N_j}\|_{TV} = \frac{1}{2} \sup_{f_{X_j}} \sup_{\varphi} |\mathbb{E}\varphi(Y_j) - \mathbb{E}\varphi(N_j)| \\
&= \frac{1}{2} \sup_{f_{X_j}} \sup_{\varphi} |\mathbb{E}(\mathbb{E}(\varphi(Y_j)|X_j)) - \mathbb{E}(\mathbb{E}(\varphi(N_j)|X_j))| \\
&\leq \frac{1}{2} \sup_{f_{X_j}} \sup_{\varphi} \mathbb{E} \{ |\mathbb{E}(\varphi(Y_j)|X_j) - \mathbb{E}(\varphi(N_j)|X_j)| \} \\
&\leq \frac{1}{2} \sup_{f_{X_j}} \mathbb{E} \left\{ \sup_{\varphi} |\mathbb{E}(\varphi(Y_j)|X_j) - \mathbb{E}(\varphi(N_j)|X_j)| \right\} \\
&= \frac{1}{2} \sup_{f_{X_j}} \mathbb{E} \|P_{Y_j|X_j} - P_{N_j|X_j}\|_{TV}.
\end{aligned}$$

f_{X_j} ist nach oben beschränkt, also betrachten wir

$$\mathbb{E} \|P_{Y_j|X_j} - P_{N_j|X_j}\|_{TV} = \mathbb{E} \left\{ \sum_{l \in \mathbb{Z}} |P(Y_j = l|X_j) - P(N_j = l|X_j)| \right\}.$$

Wir werden uns als erstes den inneren Teil des Erwartungswertes ansehen und uns dann Schritt für Schritt weiter nach außen arbeiten. Beginnen wir mit $P(Y_j = l|X_j)$.

Schritt 1:

Mit Hilfe der Inversionsformel gilt dann

$$\begin{aligned}
P(Y_j = l|X_j) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \\
&= \frac{1}{2\pi} \int_{-\pi}^{-a_n} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \\
&\quad + \frac{1}{2\pi} \int_{-a_n}^{a_n} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \\
&\quad + \frac{1}{2\pi} \int_{a_n}^{\pi} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx.
\end{aligned}$$

Den Betrag des äußeren Integralbereichs von a_n bis π schätzen wir nach oben gegen

einen Term, der gegen null geht, ab

$$\begin{aligned}
& \left| \int_{a_n}^{\pi} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \right| \\
&= \left| \int_{a_n}^{\pi} \exp(-ixl) (\exp(X_j(e^{ix} - 1)))^{r_n} dx \right| \\
&\leq \int_{a_n}^{\pi} |\exp(-ixl)| |\exp(r_n X_j(e^{ix} - 1))| dx \\
&\leq \int_{a_n}^{\pi} |\exp(r_n X_j \cos(x) + ir_n X_j \sin(x) - r_n X_j)| dx \\
&= \int_{a_n}^{\pi} \exp(r_n X_j(\cos(x) - 1)) dx \\
&\leq (\pi - a_n) \exp(r_n X_j(\cos(a_n) - 1)) \leq \mathcal{O}(\exp(-c r_n a_n^2)) \rightarrow 0 \quad (c > 0)
\end{aligned} \tag{3.3}$$

für $r_n \rightarrow \infty$, $a_n \rightarrow 0$ und $r_n a_n^2 \rightarrow \infty$.

Und analog mit der gleichen Abschätzung gilt

$$\left| \int_{-\pi}^{-a_n} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \right| \leq \mathcal{O}(\exp(-c r_n a_n^2)) \rightarrow 0 \quad (c > 0)$$

für $r_n \rightarrow \infty$, $a_n \rightarrow 0$ und $r_n a_n^2 \rightarrow \infty$.

Bemerkung 3.5 Da X_j durch $C_2 \geq X_j > C_1$ mit $C_1 > 0$, $C_2 > 0$ beschränkt ist, lässt sich $c > 0$ entsprechend wählen.

Zur Abschätzung des Integrals über den inneren Bereich zentrieren wir zunächst unsere Zufallsvariablen $Y_{j,k}|X_j$, $j = 1, \dots, n$, $k = 1, \dots, r_n$ und führen anschließend eine Taylorentwicklung durch.

Wir setzen also

$$Y_{j,k}^*|X_j := Y_{j,k}|X_j - \mathbb{E}(Y_{j,k}|X_j) = Y_{j,k}|X_j - X_j, \quad j = 1, \dots, n, \quad k = 1, \dots, r_n,$$

dann gilt

$$\begin{aligned}
& \int_{-a_n}^{a_n} \exp(-ixl) (\psi_{Y_{j,1}|X_j}(x))^{r_n} dx \\
&= \int_{-a_n}^{a_n} \exp(-ixl) (\psi_{Y_{j,1}^*|X_j+X_j}(x))^{r_n} dx \\
&= \int_{-a_n}^{a_n} \exp(-ixl) \left(\exp(ixX_j) \psi_{Y_{j,1}^*|X_j}(x) \right)^{r_n} dx \\
&= \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ixX_j) (\psi_{Y_{j,1}^*|X_j}(x))^{r_n} dx.
\end{aligned} \tag{3.4}$$

Eine Taylorentwicklung von $\psi_{Y_{j,1}^*|X_j}$ um $a = 0$ ergibt

$$\psi_{Y_{j,1}^*|X_j}(x) = 1 - \frac{1}{2}X_j x^2 + R_1(x, X_j) \quad (3.5)$$

mit

$$\begin{aligned} |R_1(x, X_j)| &= \left| \frac{1}{2} \int_0^x \psi'''_{Y_{j,1}^*|X_j}(t)(x-t)^2 dt \right| \\ &\leq \frac{1}{6} |x|^3 \sup_{t \in [0, x]} |\psi'''_{Y_{j,1}^*|X_j}(t)| \\ &= \frac{1}{6} |x|^3 \sup_{t \in [0, x]} |\mathbb{E}((iY_{j,1}^*)^3 \exp(itY_{j,1}^*)|X_j)| \\ &\leq \frac{1}{6} |x|^3 \mathbb{E}(|Y_{j,1}^*|^3|X_j) \\ &\leq \frac{1}{6} |x|^3 (X_j + 3X_j^2 + X_j^3) \\ &\leq \mathcal{O}(x^3). \end{aligned}$$

Mit Hilfe der Potenzreihenentwicklung

$$\exp(x) = 1 + x + x^2 \sum_{j=0}^{\infty} \frac{x^j}{j!} \cdot \frac{1}{(j+1)!(j+2)!} \leq 1 + x + \frac{1}{2} x^2 \exp(|x|)$$

gilt

$$\exp(-\frac{1}{2}X_j x^2) = 1 - \frac{1}{2}X_j x^2 + R_2(x, X_j) \quad (3.6)$$

und da $|x| \leq \pi$, gilt weiter

$$|R_2(x, X_j)| = \left| \frac{1}{8} X_j^2 x^4 \exp(-\frac{1}{2}X_j x^2) \right| \leq \mathcal{O}(x^4).$$

Setzen wir nun (3.6) in (3.5) ein, erhalten wir

$$\begin{aligned} \psi_{Y_{j,1}^*|X_j}(x) &= \exp(-\frac{1}{2}X_j x^2) + R_1(x, X_j) - R_2(x, X_j) \\ &= \exp(-\frac{1}{2}X_j x^2) \underbrace{(1 + \exp(\frac{1}{2}X_j x^2)(R_1(x, X_j) - R_2(x, X_j)))}_{=: \exp(R(x, X_j))} \end{aligned}$$

mit

$$R(x, X_j) \leq \mathcal{O}(x^3).$$

Damit ist das Integral über den inneren Bereich in (3.4)

$$\begin{aligned}
& \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) (\psi_{Y_{j,1}^* | X_j}(x))^{r_n} dx \right| \\
& \leq \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) \exp(r_n R(x, X_j)) dx \right| \\
& = \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) (1 + \exp(r_n R(x, X_j)) - 1) dx \right| \\
& = \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) dx \right| \\
& \quad + \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) (\exp(r_n R(x, X_j)) - 1) dx \right|.
\end{aligned}$$

Der zweite Summand lässt sich wie folgt abschätzen:

$$\begin{aligned}
& \left| \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) (\exp(r_n R(x, X_j)) - 1) dx \right| \quad (3.7) \\
& \leq \int_{-a_n}^{a_n} |\exp(-\frac{1}{2} r_n X_j x^2) (\exp(r_n R(x, X_j)) - 1)| dx \\
& \leq \mathcal{O}(a_n) \cdot \mathcal{O}(r_n a_n^3) = \mathcal{O}(r_n a_n^4) \longrightarrow 0, \text{ falls } r_n \rightarrow \infty, a_n \rightarrow 0 \text{ und } r_n a_n^3 \rightarrow 0.
\end{aligned}$$

Mit den Abschätzungen der Integrale der inneren und äußeren Bereiche wenden wir uns nun dem nächsten Schritt zu.

Schritt 2:

In diesem Schritt betrachten wir die Differenz zwischen den einzelnen Wahrscheinlichkeiten von $Y_j | X_j$ und Zähldichten $p_n(l, X_j)$ wie in Definition 3.1 mit Parametern $\mu = r_n X_j$, $\sigma^2 = r_n X_j$ und $\alpha = 1$. Die entsprechenden Integrale spalten wir wie in Schritt 1 auf.

Für $p_n(l, X_j)$ an der Stelle l gilt

$$\begin{aligned}
& \frac{1}{2\pi} \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) dx \\
& = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) dx \\
& \quad - \frac{1}{2\pi} \int_{\mathbb{R} \setminus [-a_n, a_n]} \exp(-ixl) \exp(r_n ix X_j) \exp(-\frac{1}{2} r_n X_j x^2) dx,
\end{aligned}$$

wobei das Integral über $\mathbb{R} \setminus [-a_n, a_n]$ für $a_n \rightarrow 0$ verschwindet und für die Differenz gilt

dann

$$\begin{aligned}
& |P(Y_j = l|X_j) - p_n(l, X_j)| \\
&= \left| \frac{1}{2\pi} \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp\left(-\frac{1}{2} r_n X_j x^2\right) \exp(r_n R(x, X_j)) \right. \\
&\quad - \exp(-ixl) \exp(r_n ix X_j) \exp\left(-\frac{1}{2} r_n X_j x^2\right) dx \\
&\quad + \frac{1}{2\pi} \int_{a_n}^{\pi} \exp(-ixl) \exp(-ix X_j) \exp(r_n X_j (e^{ix} - 1)) dx \\
&\quad + \frac{1}{2\pi} \int_{-\pi}^{-a_n} \exp(-ixl) \exp(-ix X_j) \exp(r_n X_j (e^{ix} - 1)) dx \\
&\quad \left. - \frac{1}{2\pi} \int_{|x|>a_n} \exp(-ixl) \exp(r_n ix X_j) \exp\left(-\frac{1}{2} r_n X_j x^2\right) dx \right| \\
&\leq \left| \frac{1}{2\pi} \int_{-a_n}^{a_n} \exp(-ixl) \exp(r_n ix X_j) \exp\left(-\frac{1}{2} r_n X_j x^2\right) (\exp(r_n R(x)) - 1) dx \right| \quad (=: A) \\
&\quad + \left| \frac{1}{2\pi} \int_{a_n}^{\pi} \exp(-ixl) \exp(-ix X_j) \exp(r_n X_j (e^{ix} - 1)) dx \right| \quad (=: B_1) \\
&\quad + \left| \frac{1}{2\pi} \int_{-\pi}^{-a_n} \exp(-ixl) \exp(-ix X_j) \exp(r_n X_j (e^{ix} - 1)) dx \right| \quad (=: B_2) \\
&\quad + \left| \frac{1}{2\pi} \int_{|x|>a_n} \exp(-ixl) \exp(r_n ix X_j) \exp\left(-\frac{1}{2} r_n X_j x^2\right) dx \right| \quad (=: C).
\end{aligned}$$

Für A gilt nach (3.7)

$$A \leq \mathcal{O}(r_n a_n^4),$$

für B_1 und B_2 nach (3.3)

$$B_1 \leq \mathcal{O}(\exp(-c r_n a_n^2)) \text{ und } B_2 \leq \mathcal{O}(\exp(-c r_n a_n^2)) \quad (c > 0)$$

und schließlich

$$C \leq \mathcal{O}(\exp(-c r_n a_n^2)) \quad (c > 0)$$

für $r_n \rightarrow \infty$, $a_n \rightarrow 0$ und gleichzeitig $r_n a_n^3 \rightarrow 0$.

Schritt 3:

Wir betrachten nun die Summe über die Differenz:

$$\begin{aligned} & \sum_{l \in \mathbb{Z}} |P(Y_j = l | X_j) - p_n(l, X_j)| \\ & \leq \sum_{|l - r_n X_j| < h_n} |P(Y_j = l | X_j) - p_n(l, X_j)| \end{aligned} \tag{3.8}$$

$$+ P(|Y_j - r_n X_j| \geq h_n | X_j) + \sum_{|l - r_n X_j| \geq h_n} p_n(l, X_j). \tag{3.9}$$

(3.8) ist nach Obigem durch $\mathcal{O}(h_n r_n a_n^4) + \mathcal{O}(h_n \exp(-c r_n a_n^2))$ ($c > 0$) nach oben beschränkt. Mit Tschebyscheff folgt für den ersten Term von (3.9)

$$P(|Y_j - r_n X_j| \geq h_n | X_j) \leq \frac{r_n X_j}{h_n^2} \leq \mathcal{O}\left(\frac{r_n}{h_n^2}\right)$$

mit $h_n \rightarrow \infty$ und $r_n \rightarrow \infty$.

Den zweiten Term zentrieren wir mittels Indextransformation

$$\begin{aligned} \sum_{|l - r_n X_j| \geq h_n} p_n(l, X_j) &= \sum_{|l - r_n X_j| \geq h_n} \frac{1}{\sqrt{2\pi r_n X_j}} \exp\left(-\frac{(l - r_n X_j)^2}{2r_n X_j}\right) \\ &= \sum_{|l| \geq h_n} \frac{1}{\sqrt{2\pi r_n X_j}} \exp\left(-\frac{l^2}{2r_n X_j}\right) \\ &= \frac{2}{\sqrt{2\pi r_n X_j}} \sum_{l \geq h_n} \exp\left(-\frac{l^2}{2r_n X_j}\right) \\ &= \frac{2}{\sqrt{2\pi r_n X_j}} \sum_{l=\lceil h_n \rceil}^{\infty} \exp\left(-\frac{l^2}{2r_n X_j}\right) \end{aligned}$$

und da $\exp\left(-\frac{l^2}{2r_n X_j}\right)$ in l für $l > 0$ monoton fallend ist, können wir das Integralkriterium anwenden. Es folgt

$$\begin{aligned} & \frac{2}{\sqrt{2\pi r_n X_j}} \sum_{l=\lceil h_n \rceil}^{\infty} \exp\left(-\frac{l^2}{2r_n X_j}\right) \\ & \leq \frac{2}{\sqrt{2\pi r_n X_j}} \int_{\lceil h_n \rceil - 1}^{\infty} \exp\left(-\frac{x^2}{2r_n X_j}\right) dx \\ & \leq \mathcal{O}\left(\exp\left(-\frac{(\lceil h_n \rceil - 1)^2}{2r_n X_j}\right)\right) \leq \mathcal{O}\left(\frac{r_n}{h_n^2}\right) \end{aligned}$$

für $h_n \rightarrow \infty$, $r_n \rightarrow \infty$ und $\frac{r_n}{h_n^2} \rightarrow 0$.

(3.9) ist also durch

$$\mathcal{O}\left(\frac{r_n}{h_n^2}\right) + \mathcal{O}\left(\frac{r_n}{h_n^2}\right) = \mathcal{O}\left(\frac{r_n}{h_n^2}\right), \quad h_n \rightarrow \infty, r_n \rightarrow \infty \text{ und } \frac{r_n}{h_n^2} \rightarrow 0$$

nach oben beschränkt.

Schritt 4:

Schließlich wird noch der Erwartungswert hinzugenommen. Dieser trägt jedoch nichts mehr zu der Rechnung bei, da sein Argument gegen null konvergiert. Zusammengefasst gilt also

$$\mathbb{E} \left\{ \sum_{l \in \mathbb{Z}} |P(Y_j = l | X_j) - p_n(l, X_j)| \right\} \leq \mathcal{O} \left(h_n a_n^4 r_n + h_n \exp(-c a_n^2 r_n) + \frac{r_n}{h_n^2} \right).$$

Wenden wir nun Lemma 3.2 an, erhalten wir die Behauptung

$$\Delta(\mathcal{P}_{Y_j}, \mathcal{P}_{N_j}) \leq \mathcal{O} \left(h_n a_n^4 r_n + \frac{r_n}{h_n^2} + h_n \exp(-c a_n^2 r_n) \right) \rightarrow 0 \quad (c > 0)$$

für $h_n \rightarrow \infty, r_n \rightarrow \infty, a_n \rightarrow 0$ mit $a_n^3 r_n \rightarrow 0, \frac{r_n}{h_n^2} \rightarrow 0$ und $a^2 r_n \rightarrow \infty$. □

a_n und h_n werden wir erst im nächsten Theorem wählen, da dort noch ein Faktor n hinzu kommt. Nachdem wir also die asymptotische Äquivalenz der einzelnen Experimente $\Delta(\mathcal{P}_{Y_j}, \mathcal{P}_{N_j}), j = 1, \dots, n, n \in \mathbb{N}$ gezeigt haben, werden wir nun die asymptotische Äquivalenz der gesamten Experimente $\mathcal{P}_{Y_1, \dots, Y_n}$ und $\mathcal{P}_{N_1, \dots, N_n}$ beweisen.

Theorem 3.6 *Mit den Bezeichnungen wie in Theorem 3.4 gilt*

$$\Delta(\mathcal{P}_{Y_1, \dots, Y_n}, \mathcal{P}_{N_1, \dots, N_n}) \leq \mathcal{O} \left(n h_n a_n^4 r_n + \frac{n r_n}{h_n^2} + n h_n \exp(-c a_n^2 r_n) \right) \rightarrow 0$$

für $h_n \rightarrow \infty, r_n \rightarrow \infty, a_n \rightarrow 0$ und $n \rightarrow \infty$ mit

$$\begin{aligned} n h_n a_n^4 r_n &\rightarrow 0, \\ \frac{n r_n}{h_n^2} &\rightarrow 0, \\ a_n^3 r_n &\rightarrow 0, \\ n h_n \exp(-c a^2 r_n) &\rightarrow 0 \text{ für hinreichend großes } c > 0. \end{aligned}$$

Die Wahl

$$a_n = \sqrt{\frac{\ln(n)}{r_n}} \quad \text{und} \quad h_n = \sqrt[3]{r_n^2}$$

erfüllt die Bedingungen, falls

$$\frac{n \ln^2(n)}{\sqrt[3]{r_n}} \longrightarrow 0 \text{ für } n \longrightarrow \infty \quad \text{und} \quad r_n \leq \text{const} \cdot n^d$$

für hinreichend großes d .

Beweis: In Theorem 3.4 haben wir gezeigt, dass die einzelnen Experimente P_{Y_j} und P_{N_j} , $j = 1, \dots, n$, $n \in \mathbb{N}$, asymptotisch äquivalent sind. Da die Experimente $\mathcal{P}_{Y_1}, \dots, \mathcal{P}_{Y_n}$ und $\mathcal{P}_{N_1}, \dots, \mathcal{P}_{N_n}$ jeweils unabhängig von einander sind, können wir auch n -mal den Abstand der ersten Experimente P_{Y_1} und P_{N_1} betrachten, womit die Behauptung folgt.

□

3.3 Diskret zu stetig

Wir haben nun also die asymptotische Äquivalenz unserer beiden Folgen diskretverteilter Experimente gezeigt. Jetzt wollen wir in den stetigen Bereich vordringen. Dazu werden wir unsere diskret-normalverteilten Zufallsvariablen N_j verwischen, d.h. wir werden jeweils eine unabhängige auf $[-\frac{1}{2}, \frac{1}{2}]$ gleichverteilte Zufallsvariable G_j hinzufügen, um so aus unseren diskreten Zufallsvariablen stückweise stetige zu machen. Wir erhalten also eine Treppenfunktion f , die die Werte der Zähldichte von N_j an der Stelle l auf dem Intervall $[-\frac{1}{2}, \frac{1}{2}]$ annimmt. Wenn wir f normieren, erhalten wir eine Dichte.

Definition 3.7 Wir nennen M eine *verwischt diskret-normalverteilte Zufallsvariable mit Parametern μ und σ^2* , falls sie die Dichte

$$f(x) = \sum_{l \in \mathbb{Z}} \mathbb{1}_{[l-\frac{1}{2}, l+\frac{1}{2}]}(x) \cdot \beta^{-1} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(l-\mu)^2}{2\sigma^2}\right) \quad (3.10)$$

mit

$$\beta = \int_{\mathbb{R}} \sum_{l \in \mathbb{Z}} \mathbb{1}_{[l-\frac{1}{2}, l+\frac{1}{2}]}(x) \frac{\beta}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(l-\mu)^2}{2\sigma^2}\right) dx$$

besitzt.

Wieso wir im Weiteren $\beta = 1$ setzen können, zeigt uns das folgende Lemma:

Lemma 3.8 Seien f_n Funktionen, wie in (3.10) und g_n beliebige Dichten. Weiter seien f_n^* Dichten von *verwischt diskret-normalverteilten Zufallsvariablen*, also

$$f_n^*(x) = \frac{f_n(x)}{\int_{\mathbb{R}} f_n(y) dy} \quad n \in \mathbb{N}.$$

Dann gilt die Abschätzung

$$\int_{\mathbb{R}} |g_n(x) - f_n^*(x)| dx \leq 2 \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx.$$

Das bedeutet, wenn gilt

$$\int_{\mathbb{R}} |g_n(x) - f_n(x)| dx \rightarrow 0 \text{ für } n \rightarrow \infty,$$

gilt also auch

$$\int_{\mathbb{R}} |g_n(x) - f_n^*(x)| dx \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Beweis: Da $f_n^*(x) = \frac{f_n(x)}{\int_{\mathbb{R}} f_n(y) dy}$ gilt, folgt

$$\begin{aligned}
\int_{\mathbb{R}} |g_n(x) - f_n^*(x)| dx &\leq \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx + \int_{\mathbb{R}} |f_n(x) - f_n^*(x)| dx \\
&= \int_{\mathbb{R}} |g_n(x) - f_n(x)| + \int_{\mathbb{R}} f_n(x) \cdot \left| 1 - \frac{1}{\int_{\mathbb{R}} f_n(y) dy} \right| dx \\
&= \int_{\mathbb{R}} |g_n(x) - f_n(x)| + \int_{\mathbb{R}} f_n(x) \cdot \frac{|\int_{\mathbb{R}} f_n(y) dy - 1|}{\int_{\mathbb{R}} f_n(y) dy} dx \\
&= \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx + \left| \int_{\mathbb{R}} f_n(y) dy - 1 \right| \\
&= \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx + \left| \int_{\mathbb{R}} f_n(y) dy - \underbrace{\int_{\mathbb{R}} g_n(y) dy}_{=1} \right| \\
&\leq \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx + \int_{\mathbb{R}} |f_n(y) - g_n(y)| dy \\
&= 2 \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx \rightarrow 0 \text{ für } n \rightarrow \infty. \quad \square
\end{aligned}$$

Bemerkung 3.9 Zu unserem Ursprungsexperiment kommen wir durch Runden wieder zurück, indem wir den Markovkern H mit $H(x, \cdot) = \delta_{\{x\}}$ verwenden, dabei wird das Argument zur näheren ganzen Zahl auf- bzw. abgerundet und der Kern ist an dieser Stelle eins.

Betrachten wir also in einem Zwischenexperiment den Abstand zwischen unseren verwischten diskret-normalverteilten Zufallsvariablen M_j und normalverteilten Zufallsvariablen Z_j .

Theorem 3.10 Seien g_n die Dichten der Normalverteilungen Z_j mit $Z_j \sim \mathcal{N}(r_n X_j, r_n X_j)$ und f_n Funktionen der Gestalt (3.10) mit $\beta = 1$, dann gilt

$$\int_{\mathbb{R}} |g_n(x) - f_n(x)| dx \rightarrow 0 \quad \text{für } r_n \rightarrow \infty.$$

Beweis: Es gilt

$$\begin{aligned}
& \int_{\mathbb{R}} |g_n(x) - f_n(x)| dx \\
&= \sum_{l \in \mathbb{Z}} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \frac{1}{\sqrt{2\pi r_n X_j}} \exp\left(-\frac{(x - r_n X_j)^2}{2r_n X_j}\right) - \frac{1}{\sqrt{2\pi r_n X_j}} \exp\left(-\frac{(l - r_n X_j)^2}{2r_n X_j}\right) \right| dx \\
&= \frac{1}{\sqrt{2\pi r_n X_j}} \sum_{l \in \mathbb{Z}} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp\left(-\frac{(x - r_n X_j)^2}{2r_n X_j}\right) - \exp\left(-\frac{(l - r_n X_j)^2}{2r_n X_j}\right) \right| dx.
\end{aligned}$$

Betrachten wir zunächst das Innere des Integrals. Nach dem Mittelwertsatz existiert für $x > l$ ein $\xi(x) \in (l, x)$, so dass

$$\frac{\exp\left(-\frac{(x - r_n X_j)^2}{2r_n X_j}\right) - \exp\left(-\frac{(l - r_n X_j)^2}{2r_n X_j}\right)}{x - l} = -2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp\left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j}\right)$$

gilt. Damit folgt für unser Integral

$$\begin{aligned}
& \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp\left(-\frac{(x - r_n X_j)^2}{2r_n X_j}\right) - \exp\left(-\frac{(l - r_n X_j)^2}{2r_n X_j}\right) \right| dx \\
&= \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp\left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j}\right) (x - l) \right| dx
\end{aligned}$$

für ein $\xi(x) \in (l, x)$.

In den folgenden Schritten werden wir das Integral in die Bereiche von $l - \frac{1}{2}$ bis l und von l bis $l + \frac{1}{2}$ zerlegen und ausnutzen, dass die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

$$f(x) = \left| \left(\frac{x - r_n X_j}{2r_n X_j} \right) \exp\left(-\frac{(x - r_n X_j)^2}{2r_n X_j}\right) \right|$$

zwei Maxima bei $x_1 = r_n X_j - \sqrt{r_n X_j}$ und $x_2 = r_n X_j + \sqrt{r_n X_j}$ sowie ein Minimum bei $x_3 = r_n X_j$ besitzt und zusätzlich symmetrisch zu den Senkrechten durch das Minimum ist.

Wir werden im Folgenden die einzelnen Bereiche, in denen die Funktion monoton wachsend bzw. fallend ist, einzeln abschätzen und die Bereiche nah an den Extremstellen separat betrachten.

Bereich 1:

In dem Bereich $l \in (\lceil r_n X_j \rceil + 1, \lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1)$ ist die Funktion monoton wachsend und es gilt

$$\begin{aligned}
& \int_l^{l+\frac{1}{2}} \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \quad \text{mit } \xi(x) \in (l, x) \\
& \leq \int_l^{l+\frac{1}{2}} \left| \left(\frac{x - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& = -\frac{1}{2} \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) \Big|_l^{l+\frac{1}{2}} \\
& = -\frac{1}{2} \exp \left(-\frac{(l + \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right)
\end{aligned}$$

sowie

$$\begin{aligned}
& \int_{l-\frac{1}{2}}^l \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \quad \text{mit } \xi(x) \in (x, l) \\
& \leq \int_{l-\frac{1}{2}}^l \left| \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& = \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right).
\end{aligned}$$

Für die Summe gilt dann

$$\begin{aligned}
& \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) - \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& \leq \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} -\frac{1}{2} \exp \left(-\frac{(l + \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \quad (3.11) \\
& + \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right). \quad (3.12)
\end{aligned}$$

Um eine Teleskopsumme zu erhalten, schätzen wir (3.11) nach oben ab:

$$\begin{aligned}
& \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} -\frac{1}{2} \exp \left(-\frac{(l + \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} -\frac{1}{2} \exp \left(-\frac{(l + 1 - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{2} \exp \left(-\frac{(\lceil r_n X_j \rceil + 1 - r_n X_j)^2}{2r_n X_j} \right) - \frac{1}{2} \exp \left(-\frac{(\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{2}.
\end{aligned}$$

Ebenso schätzen wir die Summe in (3.12) nach oben ab, indem wir die Anzahl der Summanden mit dem größten Summanden multiplizieren. Für die Anzahl der Summanden gilt außerdem $(\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1) - (\lceil r_n X_j \rceil + 1) + 1 \leq \sqrt{r_n X_j}$ und damit

$$\begin{aligned}
& \sum_{l=\lceil r_n X_j \rceil + 1}^{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1} \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{2} \sqrt{r_n X_j} \left(\frac{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1 - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1 - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{2} \sqrt{r_n X_j} \left(\frac{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1 - r_n X_j}{2r_n X_j} \right) \\
& \leq \frac{1}{2} \sqrt{r_n X_j} \left(\frac{\lfloor \sqrt{r_n X_j} \rfloor}{2r_n X_j} \right) \\
& \leq \frac{1}{4}.
\end{aligned}$$

Bereich 2:

In dem Bereich $l \in (\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1, \infty)$ ist die Funktion monoton fallend und es gilt

$$\begin{aligned} & \int_l^{l+\frac{1}{2}} \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \quad \text{mit } \xi(x) \in (l, x) \\ & \leq \int_l^{l+\frac{1}{2}} \left| \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\ & = \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \end{aligned}$$

und

$$\begin{aligned} & \int_{l-\frac{1}{2}}^l \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \quad \text{mit } \xi(x) \in (x, l) \\ & \leq \int_{l-\frac{1}{2}}^l \left| \left(\frac{x - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\ & = -\frac{1}{2} \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) \Big|_{l-\frac{1}{2}}^l \\ & = -\frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right). \end{aligned}$$

Für die Summe gilt dann

$$\begin{aligned} & \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) - \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\ & \leq \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \end{aligned} \tag{3.13}$$

$$+ \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} -\frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right). \tag{3.14}$$

Für die Abschätzung von (3.13) verwenden wir das Integralkriterium.

Es gilt

$$\begin{aligned}
& \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \int_{\lceil r_n X_j + \sqrt{r_n X_j} \rceil}^{\infty} \frac{1}{2} \left(\frac{l - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) dl \\
& = \frac{1}{4} \exp \left(-\frac{(\lceil r_n X_j + \sqrt{r_n X_j} \rceil - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{4}
\end{aligned}$$

und (3.14) lässt sich wieder gegen eine Teleskopsumme abschätzen:

$$\begin{aligned}
& \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} -\frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - \frac{1}{2} - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \sum_{l=\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1}^{\infty} -\frac{1}{2} \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) + \frac{1}{2} \exp \left(-\frac{(l - 1 - r_n X_j)^2}{2r_n X_j} \right) \\
& = \frac{1}{2} \exp \left(-\frac{(\lceil r_n X_j + \sqrt{r_n X_j} \rceil - r_n X_j)^2}{2r_n X_j} \right) \\
& \leq \frac{1}{2}.
\end{aligned}$$

Bereich 3:

Für den Bereich $l \in [\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1, \lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1]$ schätzen wir das Integral gegen das Maximum der Funktion ab, multipliziert mit der Länge des Integrals. Für die Länge gilt $\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1 + \frac{1}{2} - (\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1 - \frac{1}{2}) \leq 4$ und es existiert ein $\xi(x) \in (l, x)$, so dass

$$\begin{aligned}
& \sum_{l=\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1}^{\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \\
& \leq \sum_{l=\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - 1}^{\lceil r_n X_j + \sqrt{r_n X_j} \rceil + 1} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& = \int_{\lfloor r_n X_j + \sqrt{r_n X_j} \rfloor - \frac{3}{2}}^{\lceil r_n X_j + \sqrt{r_n X_j} \rceil + \frac{3}{2}} \left| \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& \leq 4 \left| \left(\frac{r_n X_j + \sqrt{r_n X_j} - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(r_n X_j + \sqrt{r_n X_j} - r_n X_j)^2}{2r_n X_j} \right) \right| \\
& = \frac{2}{\sqrt{r_n X_j}} \exp \left(-\frac{1}{2} \right)
\end{aligned}$$

gilt.

Bereich 4:

Wir betrachten in diesem Schritt den Bereich $l \in [\lfloor r_n X_j \rfloor - 1, \lceil r_n X_j \rceil + 1]$ und schätzen das Integral genau wie im dritten Bereich für ein $\xi(x) \in (l, x)$ ab. Das Maximum des Integrals liegt entweder am linken oder rechten Rand und es gilt

$$\begin{aligned}
& \sum_{l=\lfloor r_n X_j \rfloor - 1}^{\lceil r_n X_j \rceil + 1} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| 2 \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) (x - l) \right| dx \\
& \leq \sum_{l=\lfloor r_n X_j \rfloor - 1}^{\lceil r_n X_j \rceil + 1} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& = \int_{\lfloor r_n X_j \rfloor - \frac{3}{2}}^{\lceil r_n X_j \rceil + \frac{3}{2}} \left| \left(\frac{\xi(x) - r_n X_j}{2r_n X_j} \right) \exp \left(-\frac{(\xi(x) - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& \leq 4 \left| \left(\frac{\max\{\lfloor r_n X_j \rfloor - \frac{3}{2}, \lceil r_n X_j \rceil + \frac{3}{2}\} - r_n X_j}{2r_n X_j} \right) \right. \\
& \quad \cdot \left. \exp \left(-\frac{(\max\{\lfloor r_n X_j \rfloor - \frac{3}{2}, \lceil r_n X_j \rceil + \frac{3}{2}\} - r_n X_j)^2}{2r_n X_j} \right) \right| \\
& \leq \frac{5}{r_n X_j}.
\end{aligned}$$

Wir haben nun den Bereich rechts der Symmetriearchse und einen kleinen positiven Bereich links abgeschätzt. Wir erhalten also aus Symmetriegründen

$$\begin{aligned}
& \sum_{l \in \mathbb{Z}} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) - \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& \leq 2 \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{2} + \frac{2}{\sqrt{r_n X_j}} \exp \left(-\frac{1}{2} \right) + \frac{5}{r_n X_j} \right).
\end{aligned}$$

Zusammengefasst gilt also

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi r_n X_j}} \sum_{l \in \mathbb{Z}} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \left| \exp \left(-\frac{(x - r_n X_j)^2}{2r_n X_j} \right) - \exp \left(-\frac{(l - r_n X_j)^2}{2r_n X_j} \right) \right| dx \\
& \leq \frac{1}{\sqrt{2\pi r_n X_j}} \left(3 + \frac{4}{\sqrt{r_n X_j}} \exp \left(-\frac{1}{2} \right) + \frac{10}{r_n X_j} \right) \rightarrow 0 \quad \text{für } r_n \rightarrow \infty. \quad \square
\end{aligned}$$

Mit diesem Ergebnis können wir unser nächstes Theorem, ähnlich Theorem 3.6 formulieren.

Theorem 3.11 Für den Abstand zwischen unseren gesamten Experimenten P_{M_1, \dots, M_n} und P_{Z_1, \dots, Z_n} gilt

$$\Delta(P_{M_1, \dots, M_n}, P_{Z_1, \dots, Z_n}) \rightarrow 0 \quad \text{für } r_n \rightarrow \infty \text{ und } n \rightarrow \infty$$

mit $\frac{n}{\sqrt{r_n}} \rightarrow 0$, wobei M_j unsere verwischten diskret-normalverteilten Zufallsvariablen und Z_j normalverteilte Zufallsvariablen mit $Z_j \sim \mathcal{N}(r_n X_j, r_n X_j)$, $j = 1, \dots, n$, $n \in \mathbb{N}$ sind.

Beweis: Wie auch im Beweis zu Theorem 3.6 können wir aufgrund der jeweiligen Unabhängigkeit der Einzelexperimente n -mal den Abstand von P_{M_1} und P_{Z_1} betrachten und erhalten die Behauptung. \square

Nun können wir alle Teile des dritten Kapitels zusammenfügen und erhalten unser Hauptresultat, den Abstand zwischen unserem Poissonexperiment und unserem Gaußexperiment.

Theorem 3.12 Mit den Bezeichnungen aus (3.4) und (3.10) gilt

$$\begin{aligned} \Delta(P_{Y_1, \dots, Y_n}, P_{Z_1, \dots, Z_n}) &\leq \Delta(P_{Y_1, \dots, Y_n}, P_{N_1, \dots, N_n}) + \Delta(P_{N_1, \dots, N_n}, P_{Z_1, \dots, Z_n}) \\ &\leq \mathcal{O}\left(h_n a_n^4 r_n + \frac{r_n}{h_n^2} + h_n \exp(-c a_n^2 r_n)\right) + \mathcal{O}\left(\frac{n}{\sqrt{r_n}}\right) \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad (c > 0) \end{aligned}$$

mit $a_n = \sqrt{\frac{\ln(n)}{r_n}}$, $h_n = \sqrt[3]{r_n^2}$, $\frac{n \ln^2(n)}{\sqrt[3]{r_n}} \rightarrow 0$ für $n \rightarrow \infty$ und $r_n \leq \text{const} \cdot n^d$ für hinreichend großes d .

Beweis: Die Behauptung ergibt sich direkt aus den Ergebnissen von Theorem 3.6 und Theorem 3.11. \square

Kurzer Ausblick

Carter (2002) hat den Abstand zwischen multinomialen Experimenten und Gaußexperimenten untersucht und wir in dieser Arbeit den Abstand zwischen Poissonexperimenten und Gaußexperimenten. In weiteren Studien könnten wir nun versuchen, das Problem zu verallgemeinern und den Abstand zwischen Gaußexperimenten und Experimenten, die aus einer Exponentialfamilie stammen, betrachten. Unter Zuhilfenahme von wiederholten Beobachtungen könnten wir also folgende Modelle untersuchen:

$$Y_{j,k} = X_j + \epsilon_{j,k}, \quad j = 1, \dots, n \text{ und } k = 1, \dots, r_n,$$

wobei die Verteilung von $\epsilon_{j,k}$ aus einer Exponentialfamilie stammt und entsprechend

$$Z_{j,k} = X_j + \delta_{j,k}, \quad j = 1, \dots, n \text{ und } k = 1, \dots, r_n,$$

wobei $\delta_{j,k}$ normalverteilt ist.

Literaturverzeichnis

- [1] Bahadur, R.R. (1954). ‘Sufficiency and statistical decision functions’. In: *Ann. Math. Statistics* 25, pp. 423–462.
- [2] Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). ’Nonparametric confidence bands in deconvolution density estimation’. In: *J. Roy. Stat. Soc. Ser. B* 69, pp. 483–506.
- [3] Blackwell, D. (1951). ‘Comparison of experiments’. In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*. Berkeley and Los Angeles: University of California Press, pp. 93–102.
– (1953). ‘Equivalent comparisons of experiments’. In: *Ann. Math. Statistics* 24, pp. 265– 272.
- [4] Bohnenblust, A., Shapley, L. and Sherman, S. (1949). ‘Reconnaissance in game theory’. In: *Rand Research Memorandum* 1949/208.
- [5] Brown, L.D. and Low, M.G. (1996). ‘Asymptotic equivalence of nonparametric regression and white noise’. In: *Ann. Statist.* 24.6, pp. 2384–2398.
- [6] Brown, L.D., Cai, T.T., Low, M.G. and Zhang, C.-H. (2002). ‘Asymptotic equivalence theory for nonparametric regression with random design’. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 688–707.
- [7] Brown, L.D., Carter, A.V., Low, M.G. and Zhang, C.-H. (2004). ‘Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift’. In: *Ann. Statist.* 32.5, pp. 2074–2097.
- [8] Carter, A.V. (2002). ‘Deficiency distance between multinomial and multivariate normal experiments’. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 708–730.
– (2006). ‘A continuous Gaussian approximation to a nonparametric regression in two dimensions’. In: *Bernoulli* 12.1, pp. 143–156.
– (2007). ‘Asymptotic approximation of nonparametric regression experiments with unknown variances’. In: *Ann. Statist.* 35.4, pp. 1644–1673.

- (2009). ‘Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise’. In: *J. Probab. Stat. Art.* ID 275308, 19.
- [9] Delaigle, A., Hall, P. and Meister, A. (2008). ’On deconvolution with repeated measurements’. In: *Ann. Stat.* 36, pp. 665-685.
- [10] Delaigle, A. and Meister, A. (2008). ’Density estimation with heteroscedastic error’. In: *Bernoulli* 14, pp. 562-579.
- [11] Dunn, G. (2004). *Statistical Evaluation of Measurement Errors, Design and Analysis of Reliability Studies*, 2nd Edn. Arnold, London.
- [12] Ester Mariucci. LE CAM THEORY ON THE COMPARISON OF STATISTICAL MODELS. 2016. hal-01314223.
- [13] Grama, I. and Nussbaum, M. (2002). ‘Asymptotic equivalence for nonparametric regression’. In: *Math. Methods Statist.* 11.1, pp. 1–36.
- [14] Hall, P. and Yao, Q. (2003). ’Inference in components of variance models with low replication’. In: *Ann. Stat.* 31, pp. 414-441.
- [15] Halmos, P.R. and Savage, L.J. (1949). ‘Application of the Radon-Nikodym theorem to the theory of sufficient statistics’. In: *Ann. Math. Statistics* 20, pp. 225–241.
- [16] Hansen, O.H. and Torgersen, E.N. (1974). ‘Comparison of linear normal experiments’. In: *The Annals of Statistics*, pp. 367–373.
- [17] Horowitz, J.L. and Markatou, M. (1996). ’Semiparametric estimation of regression models for panel data’. In: *Rev. Econom. Stud.* 63, pp. 145-168.
- [18] Jähnisch, M. and Nussbaum, M. (2003). ‘Asymptotic equivalence for a model of independent non identically distributed observations’. In: *Statist. Decisions* 21.3, pp. 197–218.
- [19] Le Cam, L. (1964). ‘Sufficiency and approximate sufficiency’. In: *Ann. Math. Statist.* 35, pp. 1419–1455.
- [20] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. New York: Springer-Verlag, pp. xxvi+742.
- [21] Le Cam, L. and Yang, G.L. (2000). *Asymptotics in statistics*. Second. Springer Series in Statistics. Some basic concepts. New York: Springer-Verlag, pp. xiv+285.

- [22] Mariucci, E. (2015). 'Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise'. In: *ESAIM Probab. Stat.* 19, pp. 560–577.
- [23] Li, T. and Vuong, Q. (1998). 'Nonparametric estimation of the measurement error model using multiple indicators'. In: *J. Multivar. Anal.* 65, pp. 139-165.
- [24] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Lecture Notes in Statistics, Band 193. Springer-Verlag Berlin Heidelberg, pp. 1-8 + pp. 88-91.
- [25] Meister, A. and Neumann, M.H. (2008). 'Deconvolution from non-standard error densities under replicated measurements'. submitted preprint.
- [26] Meister, A. and Reiß, M. (2013). 'Asymptotic equivalence for nonparametric regression with non-regular errors'. In: *Probab. Theory Related Fields* 155.1-2, pp. 201–229.
- [27] Neumann, M.H. (2007). 'Deconvolution from panel data with unknown error distribution'. In: *J. Multivar. Anal.* 98, pp. 1955-1968.
- [28] Nussbaum, M. (1996). 'Asymptotic equivalence of density estimation and Gaussian white noise'. In: *Ann. Statist.* 24.6, pp. 2399–2430.
- [29] Reiß, M. (2008). 'Asymptotic equivalence for nonparametric regression with multivariate and random design'. In: *Ann. Statist.* 36.4, pp. 1957–1982.
- [30] Schennach, S.M. (2004a). 'Estimation of nonlinear models with measurement error'. In: *Econometrica* 72, pp. 33-75.
- [31] Schennach, S.M. (2004b). 'Nonparametric regression in the presence of measurement error'. In: *Econometric Theo.* 20, pp. 1046-1093.
- [32] Shiryaev, A.N. and Spokoiny, V.G. (2000). *Statistical experiments and decisions*. Vol. 8. Advanced Series on Statistical Science & Applied Probability. Asymptotic theory. River Edge, NJ: World Scientific Publishing Co. Inc., pp. xvi+281.
- [33] Stefanski, L.A. and Carroll, R.J. (1990). 'Deconvoluting kernel density estimators'. In: *Statistics* 21, pp. 169-184.
- [34] Strasser, H. (1985). *Mathematical theory of statistics*. Vol. 7. de Gruyter Studies in Mathematics. Statistical experiments and asymptotic decision theory. Berlin: Walter de Gruyter & Co., pp. xii+492.
- [35] Torgersen, E.N. (1972). 'Comparison of translation experiments'. In: *The Annals*

of Mathematical Statistics 43.5, pp. 1383–1399.
– (1974). ‘Comparison of experiments by factorization’. In: *Stat.Res.Report*, Univ. of Oslo.

- [36] Turner, S.W., Tonne, B.K. and Brett-Jones, J.R. (1986). ‘Computerized tomographic scan changes in early schizophrenia - preliminary findings’. In: *Psychol. Med.* 16, pp. 219-225.