



Bachelorarbeit
Albert Ludwigs Universität Freiburg

Kerndichte Schätzfunktionen

Maxim Sessler

21.08.17

Betreuung: Frau Prof. Dr. Angelika Rohde

Inhaltsverzeichnis

1	Einführung	1
2	Kerndichteschätzer	1
2.1	Mittlerer quadratischer Fehler	3
2.1.1	Varianz Schranke	4
2.1.2	Bias Schranke	6
2.2	Integrierter mittlerer quadratischer Fehler	9
2.2.1	Eine Schranke für die Varianz	10
2.2.2	Eine Schranke für den Bias	11
3	Untere Schranke des Minimax Risikos	13
3.1	generelles reduktions Schema	14
3.1.1	Die Reduktion auf Grenzen in Wahrscheinlichkeit	14
3.1.2	Die Reduzierung auf eine endliche Anzahl an Hypothesen	15
3.1.3	Die Wahl von 2s-getrennten Hypothesen	15
3.2	Untere Schranke basierend auf 2 Hypothesen.	16
3.3	Schranken auf Grundlage von Abständen	19
3.4	Untere Schranken des Risikos von Regressionsschätzern an einem Punkt	20
3.4.1	Untere Schranken in L_2	25
3.4.2	Untere Schranken bezüglich der Sup-Norm	27
3.4.3	Konvergenzraten in Supremums- und L_2 -Norm	29
3.4.4	Die Methode der zwei Fuzzy-Hypothesen	30
3.5	Fazit	31
4	Verwendete Sätze und Definitionen	31
5	Eigenständigkeitserklärung	34

1 Einführung

Diese Arbeit entstand im Rahmen des Stochastik- Bachelorseminars an der Universität Freiburg im Sommersemester 2017 unter der Betreuung von Prof. Dr. Angelika Rohde. Das Thema ist der nichtparametrischen Statistik zuzuordnen, diese hat sich in den letzten Jahrzehnten stark weiterentwickelt und lässt sich etwas verallgemeinert in drei fundamentale Themengebiete aufteilen.

- (i) Die Methoden der Konstruktion von Schätzfunktionen.
- (ii) Das Untersuchen der Schätzfunktionen auf ihre Eigenschaften.
- (iii) Die adaptive Schätzung.

Das erste Kapitel beschäftigt sich mit einer bestimmten Art von Schätzfunktionen, dem Kerndichteschätzer und in diesem Zusammenhang dessen Konstruktion. Die Kerndichteschätzung ist neben dem Histogramm eines der ältesten statistischen Verfahren zur Schätzung der Wahrscheinlichkeitsverteilung einer Zufallsvariable und hat dem Histogramm gegenüber den Vorteil stetig zu sein.

Im zweiten Kapitel widmen wir uns primär (iii). Das heißt, es wird herausgearbeitet, weswegen die durch die Kerndichteschätzer zu erreichende Konvergenzrate tatsächlich bereits unter allen Schätzern optimal ist. Dies wird erreicht, indem wir das maximale Risiko betrachten.

Schätzfunktionen sind spezielle Stichprobenfunktionen, um Parameter oder Verteilungen der Grundgesamtheit zu bestimmen. Parameterfreie Modelle unterscheiden sich von parametrischen Modellen dadurch, dass die Modellstruktur nicht von vornherein festgelegt wird, sondern aus den Daten bestimmt wird. Der Begriff parameterfrei bedeutet nicht, dass solche Modelle überhaupt keine Parameter besitzen. Vielmehr ist die Art und Anzahl der Parameter flexibel und nicht von vornherein festgelegt.

Wir benötigen für eine nichtparametrische Schätzung folgende 3 Bedingungen.

- (i) Eine nichtparametrische Klasse von Funktionen Θ (zum Beispiel $\Sigma(\beta, L)$ die Hölder-Klasse), welche die zu schätzende Funktion θ beinhaltet.
- (ii) Eine Metrik $d : \Theta \times \Theta \rightarrow [0, \infty)$ für die Risikodefinition.
- (iii) Eine Familie $\{P_\theta, \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen, indiziert von Θ auf einem messbaren Raum $(\mathcal{X}, \mathcal{A})$ bezüglich der Daten.

Sei angenommen, wir hätten eine Stichprobe X_1, \dots, X_n identisch verteilter Zufallsvariablen mit Wahrscheinlichkeit \mathbb{P}_p bezüglich der gemeinsamen und unbekannten Dichte

$$p : \mathbb{R} \rightarrow [0, \infty)$$

Das zentrale Thema dieser Arbeit wird das Schätzen der Dichte p sein.

Ein Schätzer für p ist eine Funktion

$$x \mapsto p_n(x) = p_n(x, X_1, \dots, X_n),$$

messbar bezüglich gegebenem $X = (X_1, \dots, X_n)$.

2 Kerndichteschätzer

In diesem Kapitel werden wir den Begriff des Kerndichteschätzers herleiten und seine Eigenschaften überprüfen. Die Schätzung der Kerndichte ist ein Verfahren, das eine stetige Schätzung einer unbekannten Verteilung ermöglicht. Ein Kerndichteschätzer ist ein stetiger Schätzer der Dichte eines unbekannten Wahrscheinlichkeitsmaßes. Der erste intuitive Ansatz für eine Schätzung mag sein, dass man das Verhältnis der Empirischen Verteilungsfunktion F zu der Dichtefunktion f benutzt:

$$F' = f$$

Jedoch ist die Empirische Verteilungsfunktion unstetig und stückweise konstant. Das bedeutet, dass man sich das oben genannte Verhältnis von Verteilungsfunktion zu Dichtefunktion nur approximativ zu nutze machen kann, wie wir im folgenden sehen werden.

Definition 1. [Tsy09](Kern)

Ein Kern ist eine messbare Funktion $K : \mathbb{R} \rightarrow \mathbb{R}$ so dass gilt,

$$\int_{\mathbb{R}} K(x) dx = 1. \quad (1)$$

Da es sich um eine positive, integrierbare, reelle Funktion mit (1) handelt, erfüllt K alle notwendigen Eigenschaften einer Wahrscheinlichkeitsdichtefunktion [Kle13]. Seien X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen, mit einer Wahrscheinlichkeitsdichte p bezüglich des Lebesgue-Maßes auf \mathbb{R} . Die dazugehörige Verteilungsfunktion ist

$$F(x) = \int_{-\infty}^x p(t) dt.$$

Betrachten wir die empirische Verteilungsfunktion $\tilde{\mathbb{P}}_n(x)$ einer Stichprobe (X_1, \dots, X_n) , wobei $\mathbb{1}_{\{A\}}$ die Indikatorfunktion bezüglich einer Menge A bezeichnet. Dann können wir schreiben,

$$\tilde{\mathbb{P}}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} = \frac{1}{n} \#\{i \in \{1, \dots, n\} : X_i \leq x\}, x \in \mathbb{R} \quad (2)$$

oder mit Ordnungsstatistik ausgedrückt:

$$\tilde{\mathbb{P}}_n(x) = \begin{cases} 0, & \text{für } x < X_{1:n} \\ \frac{1}{n}, & \text{für } X_{1:n} \leq x < X_{2:n} \\ \frac{2}{n}, & \text{für } X_{2:n} \leq x < X_{3:n} \\ \dots & \\ 1, & \text{für } X_{n:n} \leq x \end{cases} \quad (3)$$

Die Empirische Verteilungsfunktion erfüllt alle Eigenschaften einer Verteilungsfunktion. Sie ist monoton steigend. Es gilt $\lim_{n \rightarrow \infty} \tilde{\mathbb{P}}_n(x) = 1$, $\lim_{n \rightarrow -\infty} \tilde{\mathbb{P}}_n(x) = 0$ und Rechtsstetigkeit. Nach dem starken Gesetz der großen Zahlen gilt, dass die empirische Verteilungsfunktion fast sicher gegen ihre entsprechende Verteilungsfunktion F konvergiert.

Lemma 1. [Tsy09] Es gilt für alle $x \in \mathbb{R}$ und für $n \rightarrow \infty$

$$\tilde{\mathbb{P}}_n(x) \xrightarrow{f.s.} F(x).$$

Für einen Beweis betrachten wir folgende Zufallsvariablen

$$Y_i := \mathbb{1}_{\{X_i \leq x\}},$$

da unsere X_i i.i.d sind, sind es auch die Y_i und das starke Gesetz großer Zahlen greift mit endlichem Erwartungswert

$$E(Y_i) = \int_{\mathbb{R}} \mathbb{1}_{\{X_i \leq x\}} d\mathbb{P} = \int_{-\infty}^x p(t) dt.$$

Beweis.

$$\begin{aligned}\mathbb{P}\left(\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n Y_i - \mathbb{E}(Y_1) \right| = 0\right) &= \mathbb{P}\left(\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}(x) - \mathbb{E}(\mathbb{1}_{\{X_i \leq x\}}) \right| = 0\right) \\ &= \mathbb{P}\left(\lim_{n \rightarrow \infty} |\tilde{\mathbb{P}}_n(x) - F(x)| = 0\right)\end{aligned}\quad (4)$$

□

$\tilde{\mathbb{P}}_n(x)$ ist also ein konsistenter Schätzer von $F(x)$, was bedeutet, dass mit größer werdender Stichprobe der zu schätzende Wert immer genauer geschätzt wird.

Wie können wir die Dichte von p schätzen? Die Beweisidee basiert auf dem folgenden Argument. Aufgrund des Fundamentalsatzes der Analysis können wir eine Approximation über den Differenzenquotienten schreiben als,

$$\lim_{h \rightarrow \infty} \frac{F(x+h) - F(x)}{2h} = \frac{F(x)'}{2}.$$

Für ausreichend kleine $h > 0$ erhalten wir dann folgende Näherung

$$p(x) \approx \frac{F(x+h) - F(x-h)}{2h}.$$

Nach Lemma 1 können wir F durch $\tilde{\mathbb{P}}_n$ für genügend große n ersetzen

$$\tilde{p}_n(x) \approx \frac{\tilde{\mathbb{P}}_n(x+h) - \tilde{\mathbb{P}}_n(x-h)}{2h}.$$

Die Funktion $\tilde{p}_n^R(x)$ ist ein Schätzer von p und wird Rosenblatt Schätzer genannt. Wir können den Term umschreiben in die Form

$$\tilde{p}_n^R(x) = \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}_{\{-1 < \frac{X_i - x}{h} \leq 1\}} = \frac{1}{nh} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h}\right),$$

wobei

$$K_0(u) = \frac{1}{2} \mathbb{1}_{\{-1 < u \leq 1\}}(u).$$

Eine simple Verallgemeinerung des Kerndichteschätzers ist gegeben, indem wir K_0 ganz allgemein durch einen Kern ersetzen.

Definition 2. [Tsy09][LD85] (Kerndichteschätzer)

$$\tilde{p}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (5)$$

Der Parameter h wird Bandbreite des Schätzers (5) genannt. Hier im ersten Kapitel wird dieser eine zentrale Rolle spielen. Der Kerndichteschätzer kann auf den mehrdimensionalen Fall verallgemeinert werden. Zum Beispiel können wir einen Kerndichteschätzer wie folgt im zweidimensionalen definieren.

Beispiel 1. [Tsy09] (Im zweidimensionalen Fall)

Seien $(X_1, Y_1), \dots, (X_n, Y_n)$ n Tupel von Zufallsvariablen, sodass die (X_i, Y_i) je identisch und unabhängig verteilt sind mit einer Dichtefunktion $p(x, y)$ mit Werten in \mathbb{R}^2 . Ein Kernschätzer von $p(x, y)$ ist dann gegeben durch

$$\tilde{p}_n(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right).$$

In der exemplarischen Abbildung 2 finden wir Beispiel 2, Beispiel 3 und Beispiel 4 wieder.

Beispiel 2. [Rei12](Rechtskern)

der Rechtskern aus obiger Herleitung ist definiert durch:

$K_0(x) = \frac{1}{2}\mathbb{1}_{x \in [-1,1]}$, der mit dem Rechtskern assoziierte Kerndichteschätzer ist somit gegeben durch:

$$\tilde{p}_n(x) = \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}_{X_i \in [x-h, x+h]}$$

Beispiel 3. [Rei12] (Dirichletkern)

$$D_n(x) = \sum_{-n}^n e^{itx} = 1 + 2 \sum_{k=1}^n \cos(kx) = \frac{\sin((n+1/2)x)}{\sin(x/2)}$$

Beispiel 4. [Rei12] (Gaußkern)

Der Gaußkern ist auch bekannt als die Standardnormalverteilung.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

Es gilt dann $\frac{1}{h}K\left(\frac{x-X_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-X_i)^2}{2h^2}\right)$, was der Dichte der Normalverteilung mit $N(X_i, h^2)$ entspricht, also entspricht hier die i -te Zufalls Variable dem Erwartungswert und h^2 der Standardabweichung. Der Kerndichteschätzer $\tilde{p}_n(x)$ ist dann wieder das arithmetische Mittel solcher Dichten. (Siehe Abbildung 2)

$$\tilde{p}_n(x) = \frac{1}{2nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$$

Beispiel 5. Wenn wir zur Anschauung die Stichprobe aus Abbildung 3 betrachten mit dem Gaußkern, also mit $\{-2, 0, 0.1, 1, 0.4, .5, 1.3, 2, 2.1\} \subset \mathcal{X}$. Dann ist die jeweilige Schätzfunktion eine Abbildung

$$\mathcal{X} \xrightarrow{\tilde{\theta}_n} P_{\theta_j}$$

unsere einzelnen Schätzfunktionen sind dann für alle $X_i \in \{-2, 0, 0.1, 1, 0.4, .5, 1.3, 2, 2.1\}$, $i = 1, \dots, 9$ gegeben durch:

$$\tilde{\theta}_n = N(X_i, h^2).$$

2.1 Mittlerer quadratischer Fehler

Der mean squared error oder auch der mittlere quadratische Fehler, beschreibt die mittlere quadratische Abweichung und wird im folgenden $MSE(X_0)$ notiert. Der $MSE(X_0)$ gibt an, wie sehr ein Punktschätzer um den zu schätzenden Wert streut. In den meisten Anwendungen ist die interessierende Größe ein Parameter der Wahrscheinlichkeitsverteilung der Beobachtungen (wie z.B. der Mittelwert μ einer Normalverteilung $N(\mu, \sigma^2)$). Damit ist der $MSE(X_0)$ ein zentrales Qualitätskriterium für Schätzer (siehe Abbildung 1). Eine geringe mittlere quadratische Abweichung bedeutet im klassischen Fall, dass gleichzeitig Bias und Varianz des Schätzers klein sind.

Definition 3. [ELL98](MSE)

der mittlere quadratische Fehler ist definiert als

$$MSE = MSE(x_0) := \mathbb{E}_p[(\tilde{p}_n(x_0) - p(x_0))^2] = \int \dots \int (\tilde{p}_n(x_0, x_1, \dots, x_n) - p(x_0))^2 \prod_{i=1}^n [p(x_i) dx_i]$$

wobei \mathbb{E}_p die Erwartung beschreibt, bezüglich der Verteilung von (X_1, \dots, X_n) .

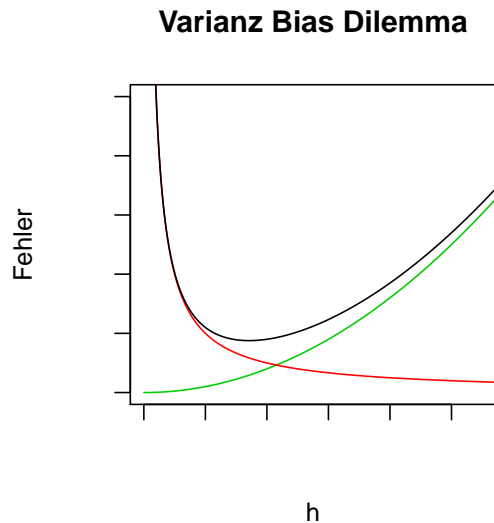


Abbildung 1: Kumulierte Summe der oberen Schranken von Varianz und Bias $\frac{C}{nh} + C_2^2 h^{2\beta}$

Wir können den mittleren quadratischen Fehler ausdrücken mittels Varianz und Bias. Wir erinnern an dieser Stelle an die Definitionen $\sigma^2(x_0) = \mathbb{E}_p[(\tilde{p}_n(x_0) - \mathbb{E}_p[\tilde{p}(x_0)])^2]$, beziehungsweise $b(x_0) = \mathbb{E}_p[\tilde{p}_n(x_0)] - p(x_0) = \mathbb{E}_p[\tilde{p}_n(x_0) - p(x_0)]$.

Satz 1. Für den Mittleren quadratischer Fehler gilt

$$MSE(X_0) = b^2(x_0) + \sigma^2(x_0).$$

Beweis. Gleichheit wird mit einer Nulladdition des Schätzers gezeigt. und der Tatsache, dass $\mathbb{E}(\tilde{p}_n(x_0)) - p(x_0)$ so wie $\tilde{p}_n(x_0)$ konstant sind.

$$\begin{aligned} MSE(X_0) &= \mathbb{E}_p\left[\left(\tilde{p}_n(x_0) - \left(\mathbb{E}_p[(\tilde{p}_n(x_0))] + \mathbb{E}_p[\tilde{p}_n(x_0)]\right) - p(x_0)\right)^2\right] \\ &= \mathbb{E}_p\left[\left(\left(\tilde{p}_n(x_0) - \mathbb{E}_p[(\tilde{p}_n(x_0))]\right) + \mathbb{E}_p[\tilde{p}_n(x_0)] - p(x_0)\right)^2\right] \\ &= \sigma^2(x_0) + 2\mathbb{E}_p\left[\left(\tilde{p}_n(x_0) - \mathbb{E}_p[\tilde{p}_n(x_0)]\right) \cdot \left(\mathbb{E}_p[\tilde{p}_n(x_0)] - p(x_0)\right)\right] + b^2(x_0) \quad (6) \\ &= \sigma^2(x_0) + 2\underbrace{\left(\mathbb{E}_p[\tilde{p}_n(x_0)] - \mathbb{E}_p[\tilde{p}_n(x_0)]\right)}_{=0} \cdot \left(\mathbb{E}_p[\tilde{p}_n(x_0)] - p(x_0)\right) + b^2(x_0) \end{aligned}$$

Der mittlere Term verschwindet aufgrund der Linearität des Erwartungswertes und der Tatsache, dass wir die X_0 fixiert betrachten. \square

2.1.1 Varianz Schranke

Um den mittleren quadratischen Fehler von \tilde{p}_n zu bestimmen, analysieren wir Varianz und den Bias getrennt. Beginnend mit der Varianz des Schätzers \tilde{p}_n nehmen wir an, dass die Dichte p der Ungleichung $p(x) \leq p_{max} < \infty$ für alle $x \in \mathbb{R}$ genügt.

(Einen Beweis wird es später geben, in Theorem 1 bei der Verwendung dieser Definition)

Proposition 1. [Tsy09](Abschätzung der Varianz)

Sei $K : \mathbb{R} \rightarrow \mathbb{R}$ ein Kern, sodass für $u = \frac{X_i - x}{h}$ gilt,

$$\int K^2(u) du < \infty, \quad \forall x \in \mathbb{R}. \quad (7)$$

Kernbeispiele

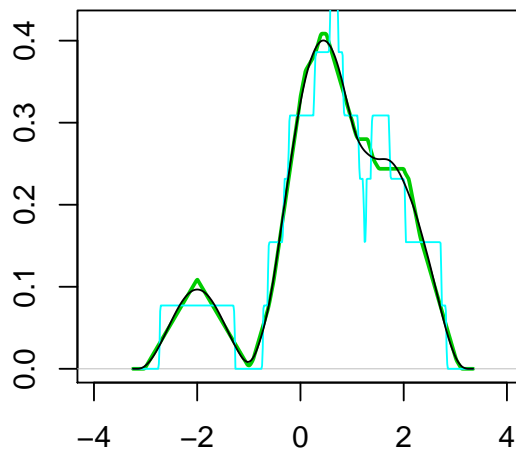


Abbildung 2: Mit der Stichprobe $\{-2, 0, 0.1, 1, 0.4, .5, 1.3, 2, 2.1\}$ dargestellt sind: Grün Dreieckskern, Türkis Rechtskern, Schwarz Dirichletkern, jeweils mit Bandbreite $h = 0.8$

Dann gilt für alle $x_0 \in \mathbb{R}$, $h > 0$ und $n \geq 1$

$$\sigma^2(x_0) \leq \frac{C_1}{nh}, \quad (8)$$

wobei $C_1 := p_{\max} \int K^2(u) du$ und $u := \frac{X_i - x_0}{h}$.

Beweis. Wir betrachten für unsere Zwecke geschickt gewählte Zufallsvariablen

$$\eta_i(x_0) = K\left(\frac{X_i - x_0}{h}\right) - E_p\left[K\left(\frac{X_i - x_0}{h}\right)\right]$$

Die Zufallsvariablen $\eta_i(x_0)$, $i = 1, \dots, n$ sind identisch und unabhängig Verteilt, denn die X_i sind es, mit $E_p[\eta_i(x_0)] = 0$ und Varianz

$$E_p[\eta_i^2(x_0)] \leq E_p\left[K\left(\frac{X_i - x_0}{h}\right)^2\right] = \int K\left(\frac{z - x_0}{h}\right)^2 p(z) dz \leq p_{\max} h \int K^2(u) du,$$

dann können wir folgern,

$$\sigma^2(x_0) = E_p\left[\left(\tilde{p}_n(x_0) - E_p(\tilde{p}_n(x_0))\right)^2\right] = E_p\left[\left(\frac{1}{nh} \sum_{i=1}^n \eta_i(x_0)\right)^2\right] \stackrel{\text{N}}{=} \frac{1}{nh^2} E_p[\eta_1^2(x_0)] \leq \frac{C_1}{nh}.$$

Bei N) wurde ausgenutzt, dass $(\sum_{i=1}^n \eta_i)^2 = \sum_{i=1}^n \sum_{j=1}^n \eta_i \eta_j = \sum_{1 \leq i, j \leq n} \eta_i(x_0) \eta_j(x_0)$. Durch den Erwartungswert verschwinden alle η mit $i \neq j$ in der Gleichung, da $E_p[\eta_i(x_0)] = 0$ und somit nur die η^2 übrigbleiben. \square

Wir kommen zu dem Schluss, dass die Varianz verschwindet, wenn $h = h_n$ mit $hn \rightarrow \infty$ für $n \rightarrow \infty$.

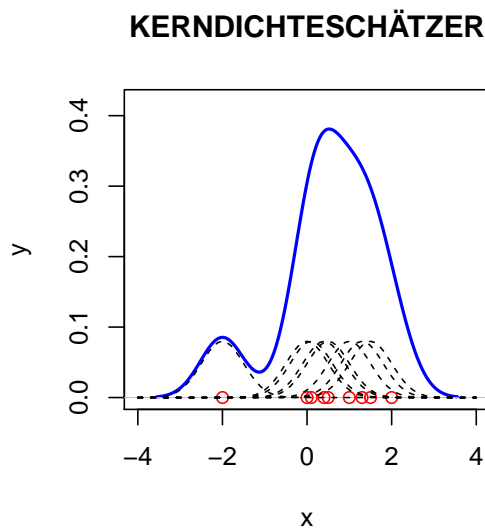


Abbildung 3: Kerndichteschätzer mit Gaußkern. Die Stichprobe $\{-2, 0, 0.1, 1, 0.4, .5, 1.3, 2, 2.1\}$ ist die gleiche wie in Abbildung 1

2.1.2 Bias Schranke

Der Bias bildet das Gegenstück zur Erwartungstreue und formalisiert, dass ein Schätzer im Mittel von dem zu schätzenden Wert abweicht. Der Bias des Kerndichteschätzers ist von der Form

$$b(x_0) = \mathbb{E}_p[\tilde{p}_n(x_0)] - p(x_0) = \frac{1}{h} \int K\left(\frac{z - x_0}{h}\right) p(z) dz - p(x_0).$$

Ein Schätzer heißt erwartungstreu, wenn die Verzerrung gleich null ist. Im folgenden analysieren wir das Verhalten von $b(x_0)$ als Funktion in Abhängigkeit von h unter gleichmäßigen Bedingungen der Dichte p und des Kerns K . Es folgen nun drei wichtige Definitionen, die wir für die Abschätzung des Bias benötigen werden. Die Hölder-Stetigkeit ist eine recht starke Glattheitsannahme und wird im folgenden Beweis über die Abschätzung des Bias eine zentrale Rolle spielen.

Definition 4. [Tsy09] (Hölderstetigkeit)

Sei T ein Intervall in \mathbb{R} und seien $\beta > 0$ und $L > 0$. Die Hölder-Klasse $\Sigma(\beta, L)$ auf T ist definiert als die Menge aller $l = \lfloor \beta \rfloor$ mal differenzierbaren Funktionen $f : T \rightarrow \mathbb{R}$, deren Ableitung $f^{(l)}$ der Gleichung

$$|f^{(l)}(x) - f^{(l)}(x')| \leq L|x - x'|^{(\beta-l)}, \forall x, x' \in \mathbb{R}$$

genügt.

Zur Erinnerung sei angemerkt:
gleichgradig differenzierbar \Rightarrow Lipschitzstetig \Rightarrow Hölderstetig \Rightarrow gleichmäßig stetig \Rightarrow stetig.

Definition 5. [Tsy09] (Ordnung eines Kerns)

Sei $l \in \mathbb{N}$, wir nennen $K : \mathbb{R} \rightarrow \mathbb{R}$ Kern der Ordnung l , falls die $u \mapsto u^j K(u)$, $\forall j = 0, 1, \dots, l$ integrierbar sind und

$$\int K(u)du = 1, \quad \int u^j K(u)du = 0, \quad j = 1, \dots, l$$

genügen.

Man bemerke, dass man in der Literatur oft eine andere Notation für Kerne der Ordnung l verwendet. Man sagt, dass K ein Kern der Ordnung $l + 1$ (mit $l \geq 0$) ist, wenn er Definition 5 und $\int u^{l+1} K(u)du \neq 0$ erfüllt. Definition 5 ist weniger einschränkend und erscheint natürlicher, denn es gibt keinen Bedarf anzunehmen, dass $\int u^{l+1} K(u)du \neq 0$ für $\beta \in \mathbb{R} \setminus \mathbb{Z}$. Zum Beispiel gilt Proposition 1 auch für $\int u^{l+1} K(u)du = 0$, unabhängig von der Existenz des Integrals. Nehmen wir nun an p gehört zur Familie von Dichten $P = P(\beta, L)$ der Hölderklasse (Die Definition folgt im nächsten Absatz).

Beispiel 6. (Kern der Ordnung l)
setze:

$$K = N(0, 1)$$

Die Standardnormalverteilung ist achsensymmetrisch und normiert. Es gilt $\int K(u)du = 1$ und außerdem $\int uK(u)du = 0$, da die zu integrierende Funktion aufgrund des Faktors u punktsymmetrisch ist.

Ansonsten kann man sich nach Belieben mithilfe der normierten Basiselemente der Legendre Polynome Kerne der Ordnung l konstruieren. [Tsy09] [ELL98] Der Sinc-Kern K ist außerdem prädestiniert, als Kern beliebiger Ordnung (sogenannter Superkern) zu dienen. [Rei12]

Definition 6. [Tsy09] (Hölderstetige Dichte)
Wir definieren die Menge der hölderstetigen Dichtefunktionen, als

$$P(\beta, L) = \left\{ p \mid p \geq 0, \int p(x)dx = 1; p \in \Sigma(\beta, L) \text{ auf } \mathbb{R} \right\}.$$

$P(\beta, L)$ ist also die Menge, aller positiven, hölderstetigen, integrierbaren und normierten Funktionen.

Proposition 2. (Abschätzung des Bias)

Nehmen wir an, dass $p \in P(\beta, L)$ ist und es sei K ein Kern der Ordnung $l = \lfloor \beta \rfloor$ wobei K

$$\int |u|^\beta |K(u)|du < \infty$$

genügt. Dann gilt für alle $x_0 \in \mathbb{R}$ und $h > 0, n \in \mathbb{N}$,

$$|b(x_0)| \leq C_2 h^\beta$$

wobei

$$C_2 := \frac{L}{l!} \int |u|^\beta |K(u)|du.$$

Beweis. Wir haben

$$b(x_0) = \frac{1}{h} \int K\left(\frac{z - x_0}{h}\right) p(z) dz - p(x_0) = \int K(u) [p(x_0 + uh) - p(x_0)] du,$$

weiter ist

$$p(x_0 + uh) = p(x_0) + p'(x_0)uh + \dots + \frac{(uh)^l}{l!} p^{(l)}(x_0 + \tau uh)$$

die Taylor-Formel [For11] mit Legendre-Restglied, wobei für τ gilt $0 \leq \tau \leq 1$ (siehe Definition 19 Kapitel 3). Alternativ schreibt man $p^{(l)}(\xi)$ für $\xi \in [x_0, x_0 + uh]$. Da K von Ordnung $l = \lfloor \beta \rfloor$ ist, erhalten wir

$$\begin{aligned} b(x_0) &= f(x) - T_{n-1}p(x_0 + uh; x_0) + \frac{f^{(n)}(\xi)}{n!}(uh)^n \\ &= \int (K(u) \sum_{i=1}^l \frac{p(x_0)^{(i)}}{i!} (uh)^i + \frac{(uh)^l}{l!} p^{(l)}(x_0 + \tau uh)) du \end{aligned} \quad (9)$$

Nach der Definition eines Kerns von Ordnung l verschwinden alle Terme der Taylorentwicklung bis auf das Restglied.

$$\int K(u) \frac{(uh)^l}{l!} p^{(l)}(x_0 + \tau uh) du = \int K(u) \frac{(uh)^l}{l!} (p^{(l)}(x_0 + \tau uh) - p^{(l)}(x_0)) du$$

und

$$\begin{aligned} |b(x_0)| &\leq \int |K(u)| \frac{|uh|^l}{l!} |p^{(l)}(x_0 + \tau uh) - p^{(l)}(x_0)| du \\ &\leq L \int |K(u)| \frac{|uh|^l}{l!} |\tau uh|^{(\beta-l)} du \leq C_2 h^\beta. \end{aligned} \quad (10)$$

□

Obergrenze des mittleren quadratischen Fehlers

Nach Definition 4 und Proposition 1, stellen wir fest, dass die oberen Grenzen von Bias und Varianz sich für eine Änderung an h gegensätzlich verhalten:

$$\begin{aligned} \sigma^2(x_0) &\leq \frac{C_1}{nh}, \\ b(x_0) &\leq C_2 h^\beta. \end{aligned}$$

Die Varianz verringert sich wenn h wächst, wohingegen b^2 monoton anwächst. Die Wahl eines kleinen h , welches einer großen Varianz entspricht, wird undersmoothing genannt. Alternativ für großen Bias spricht man vom oversmoothing (siehe Abbildung 4). Der Optimale Wert von h liegt in der Mitte von $b^2 + \sigma^2$ (siehe Abbildung 1). Um Einblick in die optimale Wahl von h zu bekommen, minimieren wir die obere Grenze des MSE. Wenn p und K der Annahmen von Proposition 1 und 2 genügen, gilt:

$$MSE \leq C_2^2 h^{2\beta} + \frac{C_1}{nh} \quad (11)$$

Das Minimum in Bezug auf h auf der rechten Seite ist erreicht bei:

$$h_n^* = \left(\frac{C_1}{2\beta C_2^2} \right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$$

h_n^* resultiert aus dem Nullsetzen der Ableitung von (11). Es folgt für $h = h_n^*$

$$MSE(x_0)_{h_n^*} = O(n^{-\frac{2\beta}{2\beta+1}}).$$

Das heißt es gilt

$$\limsup_{n \rightarrow \infty} \left| \frac{MSE(x_0)}{n^{-\frac{1}{2\beta+1}}} \right| < \infty.$$

Theorem 1. [Tsy09]

Sei K quadratintegrierbar mit $\int K(u)du = 1$. Die Eigenschaften wie in Proposition 1 sind erfüllt für fixes $h = \alpha n^{-\frac{1}{2\beta+1}}$, $\alpha > 0$. Dann gilt für den Kerndichteschätzer \tilde{p}_n , $n \geq 1$

$$\sup_{x_0 \in \mathbb{R}} \sup_{p \in P(\beta, L)} MSE(x_0) = \sup_{x_0 \in \mathbb{R}} \sup_{p \in P(\beta, L)} \mathbb{E}_p[(\tilde{p}_n(x_0) - p(x_0))^2] \leq C n^{-\frac{2\beta}{2\beta+1}}. \quad (12)$$

Wobei $C > 0$ eine Konstante ist, die nur von dem Kern K sowie α, β, L abhängt.

Beweis. Wir benutzen (11) wie oben gezeigt. Um die Anwendung von Proposition 1 zu rechtfertigen, fehlt noch zu beweisen, dass eine Konstante $p_{max} < \infty$ existiert mit,

$$\sup_{x \in \mathbb{R}} \sup_{p \in P(\beta, L)} p(x) \leq p_{max}. \quad (13)$$

Um (13) zu zeigen, nehmen wir an K^* ist ein beschränkter Kern der Ordnung l (Nicht notwendigerweise gleich K). Unter Verwendung von Proposition 2 mit $h = 1$ erhalten wir für $x \in \mathbb{R}$ beliebig und $p \in P(\beta, L)$,

$$b_{h=1}(x_0) = \left| \int K^*(z - x_0)p(z)dz - p(x_0) \right| \leq C_2^* := \frac{L}{l!} \int |u|^\beta |K^*(u)|du$$

nach Definition. Folglich ergibt sich wegen: $p(x_0) = \int K(z - x_0)p(z)dz - b(x_0)$, für $x \in \mathbb{R}$ beliebig und $p \in P(\beta, L)$

$$p(x) \leq C_2^* + \int |K^*(z - x)|p(z)dz \leq C_2^* + K_{max}^*,$$

wobei $K_{max}^* = \sup_{u \in \mathbb{R}} |K^*(u)|$ gilt, da K^* beschränkt ist nach der Annahme. \square

Unter der Annahme von Theorem 1, wird die Konvergenzrate des Schätzers $\tilde{p}_n(x_0)$ beschrieben, durch $\psi_n = n^{-\frac{\beta}{2\beta+1}}$. Das bedeutet, dass für $C < \infty$ und alle $n \geq 1$ gilt

$$\sup_{p \in P(\beta, L)} \mathbb{E}_p[(\tilde{p}_n(x_0) - p(x_0))^2] \leq C \psi_n^2$$

Finden wir Schätzer T_n , welche ein besseres Konvergenzverhalten ψ_n als die bisher betrachteten Kerndichteschätzer aufweisen? Was ist die bestmögliche Konvergenzrate die es zu bestimmen gibt? Um den Fragen nachzugehen, betrachten wir das Minimum Maximum Risiko von $P(\beta, L)$

$$R_n^*(P(\beta, L)) := \inf_{T_n} \sup_{p \in P(\beta, L)} \mathbb{E}_p[(T_n(x_0) - p(x_0))^2]$$

wobei das Infimum über alle Schätzer geht. Wenn wir eine Konstante $C' > 0$ finden, sodass C' zusammen mit der Konvergenzrate ψ_n^2 eine untere Schranke bildet, $R_n^*(P(\beta, L)) \geq C' \psi_n^2 = C' n^{-\frac{2\beta}{2\beta+1}}$. Dann wäre bewiesen, dass der Kerndichteschätzer bezüglich des MSE in Theorem 1 bereits die optimale Konvergenzrate $n^{-\frac{2\beta}{2\beta+1}}$ erreicht, bezüglich $P(\beta, L)$. Dies wird im zweiten Kapitel betrachtet.

2.2 Integrierter mittlerer quadratischer Fehler

Einführend haben wir uns mit dem Verhalten von Dichteschätzern \tilde{p}_n an einem festen Punkt x_0 befasst. Im Folgenden werden wir uns mit dem MSE, dem mean integrated squared error, also dem globalen Risiko von \tilde{p}_n befassen.

Definition 7. [Tsy09] Der integrierte mittlere quadratische Fehler ist definiert als

$$MISE := \mathbb{E}_p \left[\int (\tilde{p}_n(x) - p(x))^2 dx \right].$$

Mit Tonelli und unserer Identifikation $MSE = \sigma^2 + b^2$ und der Linearität des Integrals folgt:

$$MISE = \int MSE(x)dx = \int b^2(x)dx + \int \sigma^2(x)dx. \quad (14)$$

(Bei Tonelli wird die Integrierbarkeit bezüglich des Produktmaßes als Voraussetzung nicht benötigt) folglich ist der MISE auch durch die Summe von Varianz und Bias repräsentiert. Wir werden in diesem Abschnitt wieder mit der Aufteilung arbeiten und zwei Abschätzungen suchen. Allerdings wird sich herausstellen, dass dies für den Bias diesmal schwieriger wird. Intuitiv kann man sich dies schnell bewusst machen indem man Abbildung 1 betrachtet. Der Bias wächst monoton.

2.2.1 Eine Schranke für die Varianz

Proposition 3. [Tsy09] (*Abschätzung der Integrierten Varianz*)

Sei $K : \mathbb{R} \rightarrow \mathbb{R}$ ist eine quadratintegrierbare Funktion, dann gilt:

$$\int \sigma^2(x)dx \leq \frac{1}{nh} \int K^2(u)du$$

Beweis. Wie in Proposition 1 nutzen wir die Zufallsvariablen $\eta_i(x_0) = K(\frac{X_i - x_0}{h}) - E_p[K(\frac{X_i - x_0}{h})]$ und erinnern uns an die folgende Abschätzung:

$$\begin{aligned} \sigma^2(x_0) &= \mathbb{E}_p[(\tilde{p}_n(x_0) - \mathbb{E}_p(\tilde{p}_n(x_0)))^2] = \mathbb{E}_p[(\frac{1}{nh} \sum_{i=1}^n \eta_i(x_0))^2] \\ &= \frac{1}{nh^2} \mathbb{E}_p[\eta_1^2(x_0)] \leq \frac{1}{nh^2} \mathbb{E}_p[K^2(\frac{X_1 - x}{h})] \end{aligned} \quad (15)$$

mit Fubini gilt $\forall x \in \mathbb{R}$

$$\begin{aligned} \int \sigma^2(x)dx &\leq \frac{1}{nh} \int [\int K^2(\frac{z-x}{h})p(z)dz]dx \\ &= \frac{1}{nh} \int p(z) [\int K^2(\frac{z-x}{h})dx]dz \\ &= \frac{1}{nh} \int K^2(u)du \end{aligned} \quad (16)$$

□

Die obere Schranke für den Varianzterm in Proposition 2 stellt keine Bedingungen an p . Das Ergebnis gilt für beliebige Dichten. Für den Biasterm in (10) verhält es sich anders. Wir können ihn nur auf einer beschränkten Teilmenge von Dichten kontrollieren. Anschaulich wird das Problem bei dem Betrachten von Abbildung 3 klar, der Bias steigt im Gegensatz zur Varianz streng monoton. Wie oben stellen wir die gleichen Glattheitsannahmen an p . Da der MISE ein Risiko bezüglich der $L_2(\mathbb{R})$ -Norm ist, erscheint es natürlich anzunehmen, dass p glatt ist bezüglich dieser Norm.

Zum Beispiel können wir annehmen p gehört zur Nikol'ski Klasse von Funktionen.

Definition 8. [Tsy09] (*Nikol'ski Klasse*)

Sei $\beta > 0$ und $L > 0$. Die Nikol'ski Klasse $H(\beta, L)$ ist definiert als die Menge der Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ deren Ableitungen $f^{(l)}$ der Ordnung $l = [\beta]$ existieren und für die weiterhin gilt:

$$[\int (f^{(l)}(x+t) - f^{(l)}(x))^2 dx]^{\frac{1}{2}} \leq L|t|^{b-l}, \forall t \in \mathbb{R}$$

Sobolevklassen liefern einen weiteren Weg Glattheit in $L_2(\mathbb{R})$ zu beschreiben.

Definition 9. [Tsy09] (Sobolev Klasse)

Sei $\beta \geq 1$ ganzzahlig und $L > 0$. Die Sobolev Klasse $S(\beta, L)$ ist definiert, als die Menge aller $\beta - 1$ mal differenzierbaren Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ mit absolut stetigen Ableitungen $f^{(\beta-1)}$, so dass

$$\int (f^{(\beta)}(x))^2 dx \leq L^2. \quad (17)$$

Mit der Minkowski Ungleichung folgt, dass $S(\beta, L) \subset H(\beta, L)$

Wir geben nun eine obere Schranke an, für den Biasterm $\int b^2 dx$ wobei p zu der Klasse von Wahrscheinlichkeitsdichten gehört, die glatt sind im Sinne von Nikol'ski,

$$P_H(\beta, L) = \left\{ p \in H(\beta, L) \mid p \geq 0; \int p(x) dx = 1 \right\}.$$

Die Schranke wird aufgrund der Implikation auch für Sobolevklassen gelten.

2.2.2 Eine Schranke für den Bias

Proposition 4. (Abschätzung des Integrierten Bias)

Sei $p \in P_H(\beta, L)$ und sei K wieder Kern der Ordnung $l = \lfloor \beta \rfloor$ sodass K

$$\int |u|^\beta |K(u)| du < \infty$$

genügt. Dann ist für $h > 0, n \geq 1$

$$\int b^2(x) dx \leq C_2^2 h^2 \beta$$

wobei wieder

$$C_2 := \frac{L}{l!} \int |u|^\beta |K(u)| du.$$

Beweis. Wähle $x, u \in \mathbb{R}, h > 0$ beliebig und betrachte die Taylor Entwicklung mit Integralrestglied [For11].

$$p(x + uh) = p(x) + p'(x)uh + \dots + \frac{(uh)^2}{(l-1)!} \int_0^1 (1-\tau)^{l-1} p^{(l)}(x + \tau uh) d\tau$$

Da K von der Ordnung $l = \lfloor \beta \rfloor$ ist folgt:

$$\begin{aligned} b(x) &\stackrel{Ord.l}{=} \int K(u) \frac{(uh)^l}{(l-1)!} \left[\int_0^1 (1-\tau)^{l-1} p^{(l)}(x + \tau uh) d\tau \right] du \\ &= \int K(u) \frac{(uh)^l}{(l-1)!} \left[\int_0^1 (1-\tau)^{l-1} p^{(l)}(x + \tau uh) - p^{(l)}(x) d\tau \right] du \end{aligned}$$

Für die folgende Abschätzung benötigen wir zwei mal die Minkowski Ungleichung und die Tatsache, dass $p \in P_H(\beta, L)$ ist.

$$\begin{aligned} \int b^2(x) dx &\leq \int \left(\int |K(u)| \frac{|uh|^l}{(l-1)!} \int_0^1 (1-\tau)^{l-1} |p^{(l)}(x + \tau uh) - p^{(l)}(x)| d\tau du \right)^2 dx \\ &\stackrel{Minowski}{\leq} \left(\int |K(u)| \frac{|uh|^l}{(l-1)!} \left(\int \left(\int_0^1 (1-\tau)^{l-1} |p^{(l)}(x + \tau uh) - p^{(l)}(x)| d\tau \right)^2 dx \right)^{\frac{1}{2}} du \right)^2 \end{aligned}$$

$$\begin{aligned}
& \stackrel{Minowski}{\leq} \left(\int |K(u)| \frac{|uh|^l}{(l-1)!} \left(\int \left(\int_0^1 (1-\tau)^{l-1} \left(p^{(l)}(x + \tau uh) - p^{(l)}(x) \right)^2 dx \right)^{\frac{1}{2}} d\tau \right) du \right)^2 \\
& \stackrel{p \in P_H(\beta, L)}{\leq} \left(\int |K(u)| \frac{|uh|^l}{(l-1)!} \underbrace{\left[\int_0^1 (1-\tau)^{(l-1)} L |uh|^{\beta-l} d\tau \right]}_{=\frac{1}{l}} du \right)^2 = C_2^2 h^{2\beta}.
\end{aligned}$$

□

Es folgt also mit Proposition 3 und 4, dass der *MISE* ebenfalls nach oben abgeschätzt werden kann:

$$MISE \leq C_2^2 h^{2\beta} + \frac{1}{nh} \int K^2(u) du$$

Der Minimierer h^* folgt direkt mit unserem Ergebnis aus dem MSE

$$h_n^* = \left(\frac{\int K^2(u) du}{2\beta C_2^2} \right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$$

$MSE(x_0) = O(n^{-\frac{2\beta}{2\beta+1}})$, $n \rightarrow \infty$ Das Verhalten ist Analog zum MSE.

Theorem 2. [Tsy09]

Sei $\alpha > 0$ fix, $h := \alpha n^{-\frac{1}{2\beta+1}}$ mit $n \geq 1$, dann genügt \tilde{p}_n der Ungleichung

$$\sup_{p \in P_H(\beta, L)} E_p \left[\int (\tilde{p}_n(x) - p(x))^2 dx \right] \leq C n^{-\frac{2\beta}{2\beta+1}}.$$

C ist eine Konstante in Abhängigkeit von α, β, L, K .

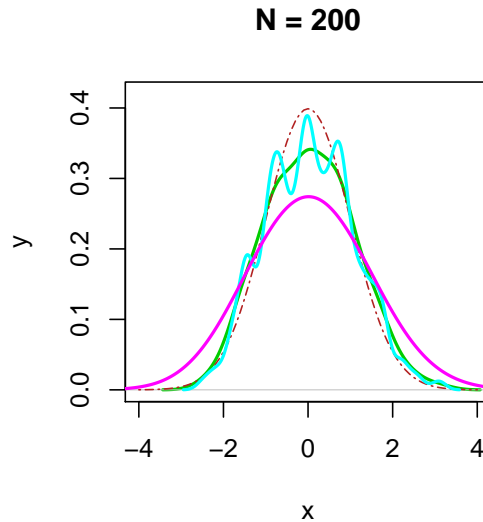


Abbildung 4: drei Schätzfunktionen mit Gauß-Kern. einer Standardnormalverteilten Stichprobe (200). Als Bandbreite wurde zwischen: 1, 0.5, und 3 gewechselt. Die Glockenkurve im Hintergrund (rot gestrichelt) stellt die Standardnormalverteilung dar.

3 Untere Schranke des Minimax Risikos

Es sei zuerst bemerkt, dass es keinen Sinn ergibt, das Risiko für eine feste Dichte f von unten abzuschätzen, da ansonsten der konstante Schätzer das Risiko Null besitzen würde. Ein Standardansatz ist daher das maximale Risiko über eine feste Parametermenge zu betrachten. Die Strategie für den Beweis unterer Schranken beruht wie wir sehen werden stets auf denselben Konzepten. Damit wir behaupten können, dass unter den Voraussetzungen von Theorem 1, die optimale Konvergenzrate $n^{\frac{2\beta}{2\beta+1}}$ tatsächlich von den Kerndichteschätzern bezüglich des L_2 Abstandes erreicht wird, fehlt noch der Beweis der Existenz einer unteren Schranke des Minimax Risikos. Das heißt, es bleibt noch zu zeigen:

$$R_n^*(P(\beta, L)) = \inf_{T_n} \sup_{p \in P(\beta, L)} \mathbb{E}_p[(T_n(x_0) - p(x_0))^2] \geq C' \psi_n^2 = C' n^{-\frac{2\beta}{2\beta+1}}, \quad C' > 0.$$

Bildlich gesprochen wird eine untere Schranke für die größte mögliche Abweichung des besten Schätzers gesucht. Bezüglich einer Metrik d ist die Ausführung eines Schätzers $\tilde{\theta}_n$ von θ gemessen durch das *Maximale Risiko* des Schätzers auf Θ

$$r(\tilde{\theta}_n) := \sup_{\theta \in \Theta} E_\theta[d^2(\tilde{\theta}_n, \theta)]$$

wobei E_θ die Erwartung bezüglich P_θ beschreibt. Im ersten Kapitel haben wir durch Theorem 1 obere Schranken für das Maximale Risiko gefunden, in Abhängigkeit der Konvergenzrate ψ_n

$$\sup_{\theta \in \Theta} E_\theta[d^2(\tilde{\theta}_n, \theta)] \leq C \psi_n^2$$

für Kerndichteschätzer $\tilde{\theta}_n = \tilde{p}_n$ und bestimmte positive Folgen $\lim_{n \rightarrow \infty} \psi_n = 0$. Um jetzt untere Schranken für das maximale Risiko zu finden, ist es sinnvoll, das minimum maximum Risiko zu betrachten.

Definition 10. [Tsy09](*minimax Risiko*)

Wir definieren das *minimax Risiko* verbunden mit einem Statistischen Modell $P_\theta, \theta \in \Theta$ und einer Pseudo-Metrik $d(\cdot, \cdot)$ als,

$$\mathcal{R}_n^* := \inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} E_\theta[d^2(\tilde{\theta}_n, \theta)].$$

Das Infimum geht also über alle Schätzer $\tilde{\theta}_n$. Durch die oberen Schranken, die wir im ersten Kapitel über Kerndichteschätzer eingeführt haben wird impliziert, dass eine Konstante $C < \infty$ existiert, so dass

$$\limsup_{n \rightarrow \infty} \psi_n^2 \mathcal{R}_n^* \leq C \quad (18)$$

für eine Nullfolge ψ_n . Die entsprechenden unteren Schranken fordern, dass es außerdem eine Schranke c gibt, mit:

$$\liminf_{n \rightarrow \infty} \psi_n^2 \mathcal{R}_n^* \geq c \quad (19)$$

Definition 11. [Tsy09](*optimale Konvergenzrate*)

Eine positive Folge $(\psi_n)_{n \in \mathbb{N}}$ wird *optimale Konvergenzrate* von Schätzern auf (Θ, d) genannt, falls (18), (19) gelten. Ein Schätzer θ_n^* der

$$\sup_{\theta \in \Theta} E_\theta[d^2(\theta_n^*, \theta)] \leq C' \psi_n^2$$

erfüllt, mit der Einschränkung, dass $(\psi_n)_{n \in \mathbb{N}}$ die optimale Konvergenzrate darstellt und $C' < \infty$ als Konstante existiert, wird *Schätzer der optimalen Rate* auf (Θ, d) genannt.

Definition 12. Ein Schätzer θ_n^* wird asymptotisch effizient auf (Θ, d) genannt, falls:

$$\lim_{n \rightarrow \infty} \frac{r(\theta_n^*)}{\mathcal{R}_n^*} = 1$$

also wenn für genügend große Stichproben, das Maximale Risiko nicht schneller wächst, als das Minimum-Maximum Risiko.

Optimale Konvergenzraten sind definiert bezüglich einer multiplikativen Konstanten (oder bis zu einem beschränkten Faktor abhängig von n). Wenn also ψ_n eine optimale Konvergenzrate ist, dann ist eine beliebige Folge ψ'_n welche

$$0 < \liminf_{n \rightarrow \infty} (\psi_n / \psi'_n) \leq \limsup_{n \rightarrow \infty} (\psi_n / \psi'_n) < \infty \quad (20)$$

genügt ebenfalls eine optimale Konvergenzrate. Folgen ψ, ψ' die (20) erfüllen, haben äquivalente Größenordnungen. Eine beliebige Folge aus der Klasse äquivalenter Folgen kann als optimale Konvergenzrate angesehen werden.

Bevor wir zum generellen Reduktions Schema kommen, einigen wir uns auf etwas allgemeinere Rahmenbedingungen. Dazu benötigen wir die Verlustfunktion ω (Definition 16).

Das maximum Risiko ist dann definiert, als:

$$r_\omega(\tilde{\theta}_n) = \sup_{\theta \in \Theta} E_\theta[\omega(\psi_n^{-1} d(\tilde{\theta}_n, \theta))]$$

Das heißt, dass wir im Folgenden untere Schranken betrachten werden, die über Ungleichungen dieser Art

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\omega(\psi^{-1} d(\tilde{\theta}_n, \theta))] \geq c > 0$$

formuliert sind.

Beispiel 7. [Tsy09] Verlustfunktionen

(i) $\omega(u) = u^p, p > 0$;

(ii) $\omega(u) = \mathbb{1}_{u \geq A} \quad A > 0$

ii) beschreibt das Risiko die Wahrscheinlichkeit über A hinauszuragen.

3.1 generelles reduktions Schema

[Tsy09] Wir betrachten ein generelles Schema, welches in drei Schritten beim Finden unterer Schranken hilft. Denn es stellt sich heraus, dass das Arbeiten mit der Minimax Risk nicht immer zielführend ist. Gesucht wird ein Alternatives Werkzeug, welches man stattdessen nach unten abschätzen kann, um somit untere Schranken zu bestimmen.

3.1.1 Die Reduktion auf Grenzen in Wahrscheinlichkeit

Einführend sei bemerkt, dass es ausreichend ist die Verlustfunktion $\omega_0(u) = \mathbb{1}_{\{u \geq A\}}$ zu betrachten. Mit der Markov-Ungleichung (Satz 3) und der Bedingung $A > 0 \Rightarrow \omega(A) > 0$ folgt,

$$\mathbb{E}_\theta[\omega(\psi^{-1}(d(\tilde{\theta}_n, \theta)))] \geq \omega(A) P_\theta(\psi^{-1} d(\tilde{\theta}_n, \theta) \geq A) = \omega(A) P_\omega(d(\tilde{\theta}_n, \theta) \geq s)$$

mit $s = s_n = A\psi_n$. Das bedeutet, anstelle eine untere Schranke für das minimax Risiko \mathcal{R}_n^* zu suchen, es ausreicht eine untere Schranke der minimax Wahrscheinlichkeiten zu finden:

$$\inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} P_\theta(d(\tilde{\theta}_n, \theta) \geq s)$$

Dies ist die erste Vereinfachung.

3.1.2 Die Reduzierung auf eine endliche Anzahl an Hypothesen

Wir schätzen nach unten ab, indem wir anstelle des Supremums über eine endliche Anzahl von Hypothesen iterieren.

$$\inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \geq \inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \quad (21)$$

für eine endliche Menge $\{\theta_0, \dots, \theta_M\} \in \Theta$ und $s = A\psi_n$. In den Beispielen werden wir $M \geq 1$ und $\theta_0, \dots, \theta_M$ angemessen wählen. Wir werden die $M+1$ Elemente $\theta_0, \theta_1, \dots, \theta_M$ aus Θ Hypothesen nennen, die wir auswählen, um untere Schranken für das Minimax Risiko zu finden.

3.1.3 Die Wahl von 2s-getrennten Hypothesen

Wir wählen unsere $\theta \in \Theta$ so, dass für alle j, k , mit $k \neq j$ gilt

$$d(\theta_j, \theta_k) \geq 2s. \quad (22)$$

Lemma 2. Mindestabstands Test

Sei d eine Pseudometrik, und sei außerdem angenommen, dass wir $M+1$ Hypothesen gefunden haben, so dass (22) für alle $j \neq k$ erfüllt ist. Dann definieren wir für einen beliebigen Schätzer $\tilde{\theta}_n$ den Mindestabstandstest $\Psi^*(\tilde{\theta}_n) : \mathcal{X} \rightarrow \{0, 1, \dots, M\}$

$$\Psi^*(\tilde{\theta}_n) := \arg \min_{0 \leq k \leq M} d(\tilde{\theta}_n, \theta_k)$$

Dann impliziert $\Psi^*(\tilde{\theta}_n) \neq j$, dass $d(\tilde{\theta}_n, \theta_j) \geq s_n$

Beweis. Sei dazu angenommen $\Psi^*(\tilde{\theta}_n) \neq j \Leftrightarrow \exists k \neq j : d(\tilde{\theta}_n, \theta_k) \leq d(\tilde{\theta}_n, \theta_j)$ Dann ist:

$$2s_n \leq d(\theta_j, \theta_k) \leq d(\tilde{\theta}_n, \theta_j) + d(\tilde{\theta}_n, \theta_k) \leq 2d(\tilde{\theta}_n, \theta_j)$$

□

Lemma 2 impliziert

$$P_{\theta_j}(d(\tilde{\theta}_n, \theta_j) \geq s) \geq P_{\theta_j}(\Psi^*(\tilde{\theta}_n) \neq j), \quad j = 1, \dots, M \quad (23)$$

$$\begin{aligned} \inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} P_{\theta_j}(d(\tilde{\theta}_n, \theta_j) \geq s) &\geq \inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \\ &\geq \inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(\Psi^*(\tilde{\theta}_n) \neq j) \\ &\geq \inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(\Psi \neq j)) \\ &:= p_{e,M} \end{aligned} \quad (24)$$

$$p_{e,M} := \inf_{\Psi} \max_{0 \leq j \leq M} P(\Psi \neq j)$$

Wir schreiben für $P_j := P_{\theta_j}$. Hier beschreibt \inf_{Ψ} das Infimum aller Tests. Man kann sich vielleicht grafisch die Minimax Wahrscheinlichkeit $p_{e,M}$ erläutern, indem folgendes Bild betrachtet wird, wobei jede Fläche einer Wahrscheinlichkeit entspricht.

$P(\Psi = 1)$	$P(\Psi = 2)$	\dots	$P(\Psi = M)$
---------------	---------------	---------	---------------

Zuerst wählt man die Summe der M Wahrscheinlichkeiten, deren Summe Maximal ist, und wählt dann Ψ so, dass sich dieses Verhältnis minimiert. Wenn wir hier keine weiteren Einschränkung an unseren Hypothesen vornehmen, bedeutet dies alle Flächen $P(\Psi = j)$ für $j = 1, \dots, M$ müssen von gleicher Größe sein, da wäre dies nicht der Fall, bereits $\max_{1 \leq j \leq M} P(\Psi \neq j)$ falsch gewählt worden.

Korollar 1. *Untere Schranke des minimax Risikos*

Anstelle untere Schranken für das minimax Risiko $\liminf_{n \rightarrow \infty} \inf_{\tilde{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\omega(\psi^{-1}(d(\tilde{\theta}_n, \theta)))] \geq c$ zu finden, ist es folglich ausreichend $p_{e,M}$ wie oben definiert durch eine Konstante c' abzuschätzen. in Formeln:

$$p_{e,M} := \inf_{\Psi} \max_{0 \leq j \leq M} P(\Psi \neq j) \geq c' \quad (25)$$

Wobei unsere die Hypothesen θ_j dem $2s$ Abstand von (22) entsprechen, mit $s = A\psi_n$ und das c' unabhängig von n ist. Die Wahrscheinlichkeit $p_{e,M}$ wird minimax Wahrscheinlichkeit des Fehlers genannt, bezüglich des Tests von $M + 1$ Hypothesen $\theta_0, \theta_1, \dots, \theta_M$.

3.2 Untere Schranke basierend auf 2 Hypothesen.

[Tsy09] Betrachten wir zunächst den einfachsten Fall $M = 1$. Dann haben wir nur zwei Hypothesen θ_0, θ_1 aus Θ . Zuerst suchen wir Schranken für die Minimax Wahrscheinlichkeit $p_{e,1}$ und dann für das Minimax Risiko

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} P_{\tilde{\theta}}(d(\tilde{\theta}, \theta) \geq s).$$

Sei $s > 0$, betrachten wir weiter die Zerlegung des signierten Maßes $P_{\theta_0} = P_0 = P_0^a + P_0^s$ in die absolut stetigen und singulären Komponenten P_0^a, P_0^s des Maßes P_0 bezüglich P_1 . Dann gilt nach dem Zerlegungssatz von Lebesgue $P_0^a \ll dP_1$. Wenn Mehrdeutigkeit ausgeschlossen ist, nutzen wir die Notation $\frac{dP_0^a}{dP_1}$ für die Radon-Nikodym Ableitung $\frac{dP_0^a}{dP_1}(X)$.

Proposition 5. [Tsy09]

$$p_{e,1} \geq \sup_{\tau > 0} \left\{ \frac{\tau}{1 + \tau} P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \right\}$$

Beweis. Sei τ fixiert. Dann gilt für einen beliebigen Test $\Psi : \mathcal{X} \rightarrow \{0, 1\}$

$$\begin{aligned} P_0(\Psi \neq 0) &= P_0(\Psi = 1) \geq P_0^a(\Psi = 1) \\ &= \int \mathbb{1}_{\{\Psi=1\}} \frac{dP_0^a}{dP_1} dP_1 \\ &\geq \tau \int \mathbb{1}_{\{\Psi=1\} \cap \{\frac{dP_0^a}{dP_1} \geq \tau\}} dP_1 \geq \tau(p - \alpha_1) \end{aligned} \quad (26)$$

wobei $p = P_1(\Psi = 1)$ und $\alpha_1 = P_1(\frac{dP_0^a}{dP_1} < \tau)$

Dann ist $p_{e,1} = \inf_{\Psi} \max_{j=0,1} P_j(\Psi \neq j) \geq \min_{0 \leq p \leq 1} \max\{\tau(p - \alpha_1), 1 - p\} = \frac{\tau(1 - \alpha_1)}{1 + \tau}$ \square

Wir sehen, dass es anstelle untere Schranken für das Minimax Risiko zu suchen, sinnvoller ist konstante τ und $0 < \alpha < 1$ zu finden, welche der Gleichung

$$P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \geq 1 - \alpha \quad (27)$$

genügen.

Theorem 3. [Tsy09] Wir nehmen an, dass Θ zwei Elemente θ_0, θ_1 enthält, die $d(\theta_0, \theta_1) \geq 2s > 0$ genügen, dann gilt

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \geq \sup_{\tau > 0} \left\{ \frac{\tau}{1 + \tau} P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \right\}$$

Denn es ist,

Beweis.

$$\begin{aligned} \inf_{\tilde{\theta}} \sup_{\theta \in \Theta} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) &\stackrel{(21)}{\geq} \inf_{\Psi} \max_{j=0,1} P_j(\Psi \neq j) \\ &\stackrel{Prop.5}{\geq} \sup_{\tau > 0} \left\{ \frac{\tau}{1 + \tau} P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \right\} \\ &\stackrel{(27)}{\geq} \frac{\tau M}{1 + \tau M} (1 - \alpha) \end{aligned} \tag{28}$$

□

(27) bedeutet die beiden Wahrscheinlichkeiten sind nicht sehr weit von einander entfernt. Beziehungsweise: Je näher P_0 an P_1 ist, desto größer ist die untere Schranke aus Theorem 3. Gilt Gleichheit $P_0 = P_1$, dann behält (27) Gültigkeit für $\tau = 1, \alpha = 0$ und die beste untere Schranke welche wir nach Proposition 5 finden, ist

$$p_{e,1} \geq \frac{1(1-0)}{1+1} = 1/2.$$

Hier ist zu beachten, dass diese Schranke nicht immer scharf ist. Sei beispielsweise $P_0 = P_1$ dann erhalten wir

$$p_{e,1} = \inf_{\Psi} \max_{j=0,1} \{P_0(\Psi = 1), P_0(\Psi = 0)\}.$$

Und wir können die rechte Seite beliebig an die Eins nähern, indem wir P_0 als passende Bernoulliverteilung wählen. In einem anderen Extremfall sind die Maße P_1, P_0 gegenseitig singulär und Theorem 3 ist trivial, da die Schranken null sind. Außerdem haben wir in diesem Fall die Wahrscheinlichkeit $p_{e,1} = 0$ und das Minimum bezüglich Ψ der Minimax Wahrscheinlichkeit des Fehlers wird erreicht, von dem Test mit Wert Eins bezüglich P_1 und Wert Null bezüglich P_0 . Man bemerke außerdem, selbst wenn $P_0 = P_1$ was auf den ersten Blick als günstigster Fall für das Bestimmen von unteren Schranken erscheint, können die Hypothesen θ_1, θ_0 so liegen, dass Theorem 3 keine guten Lösungen liefern würde. Die Wahl der Hypothesen ist also sehr wichtig.

Beispiel 8. [Tsy09] *Eine schlechte Wahl von zwei Hypothesen.*

Betrachten wir die Regressionsanalyse

$$Y_i = f\left(\frac{i}{n}\right) + \xi_i \quad i = 1, \dots, n.$$

Wobei für f gilt: $f \in \Sigma(1, 1)$. Das heißt f ist Lipschitz-stetig mit $L = 1$. Nehmen wir folgende Hypothesen an

$$\theta_0 = f_0(\cdot) \equiv 0,$$

also konstant null und

$$\theta_1 = f_1(\cdot).$$

Für geeignete f_0, f_1 wobei wir $f_1(x) = (2\pi n)^{-1} \sin(2\pi n x)$ wählen, dann ist wegen $f_1(i/n) = (2\pi n)^{-1} \sin(2\pi i) = 0$ $f_0(i/n) = f_1(i/n)$ für alle i . Es folgt außerdem, dass die Beobachtungen (Y_1, \dots, Y_n) die gleichen sind für $f = f_0$ und $f = f_1$. Dann ist $P_0 = P_1$, und nach Proposition 4 haben wir $p_{e,1} \geq 1/2$ für beliebige Fehler ξ_i . Nehmen wir als Abstand $d(f, g) = \|f - g\|_{\infty} = \sup_{0 < x < 1} |f - g|$, dann ist $d(f_0, f_1) = (2\pi n)^{-1}$ und da $f_0, f_1 \in \Sigma(1, 1)$, können wir mit Theorem

3 und unserer Ungleichung von oben mit $s = (4\pi n)^{-1}$ folgende Abschätzung vornehmen. Wir zeigen mit der Markov Ungleichung

$$\mathbb{E}_\theta[\omega(\psi^{-1}(d(\tilde{\theta}_n, \theta)))] \geq \omega(A)P_\omega(d(\tilde{\theta}_n, \theta) \geq (4\pi n)^{-1}),$$

um die Ungleichung (19) für die Klasse $\Theta = \Sigma(1, 1)$ mit der Rate

$$0 < \liminf_{n \rightarrow \infty}(\psi_n/\frac{1}{n}) \leq \limsup_{n \rightarrow \infty}(\psi_n/\frac{1}{n}) < \infty$$

zu erhalten. Also es gilt $\psi_n \asymp \frac{1}{n}$. Dieses Ergebnis ist nicht zufriedenstellend, wie sich im weiteren Verlauf noch zeigen wird, ist diese Rate zu schnell.

Folgende zwei Annahmen werden im Folgenden häufiger verwendet und finden daher hier ihren Platz.

Annahme 1. *Wir betrachten das nichtparametrische Regressionsmodell*

$$Y_i = f(X_i) + \xi_i \quad \forall i = 1, \dots, n$$

wobei f eine Funktion von $[0, 1]$ nach \mathbb{R} beschreibt. Die Zufallsvariablen X_i sind unabhängig mit $\mu = 0$ und $\mathbb{E}(\xi_i^2) = \sigma_\xi^2 < \infty$ und $X_i = i/n$ für $i = 1, \dots, n$

$\{\varphi_j\}_{j=1}^\infty$ ist die trigonometrische Basis.

iii) Die Fourier Koeffizienten $\theta_j = \int_0^1 f \varphi_j$ von f genügen:

$$\sum_{j=1}^\infty |\theta_j| < \infty$$

Es folgt aus ii) und iii), dass die Reihe $\sum_{j=1}^\infty \theta_j \psi_j(x)$ absolut konvergent ist, für alle $x \in [0, 1]$.

Annahme 2. *Das statistische Modell ist das von nichtparametrischer Regression.*

1. $Y_i = f(X_i) + \xi_i$, $\forall i = 1, \dots, n$ wobei $f : [0, 1] \rightarrow \mathbb{R}$.
2. Die Zufallsvariablen ξ_i sind i.i.d. mit einer Wahrscheinlichkeitsdichte bezüglich des Lebesguemaßes auf den reellen Zahlen, so dass

$$\exists p_* > 0, v_0 > 0 : \int p_\xi(u) \log \frac{p_x^i(u)}{p_\xi(u+v)} du \leq p_* v^2 \quad (29)$$

für alle $|v| \leq v_0$.

3. Die Variablen $X_i \in [0, 1]$ sind deterministisch.

Sei $(\mathcal{X}, \mathcal{A})$ ein Messraum und seinen P, Q zwei Wahrscheinlichkeitsmaße auf \mathcal{X} . Nehmen wir an ν ist ein σ -endliches Maß auf $(\mathcal{X}, \mathcal{A})$ wobei hier P, Q jeweils absolut stetig sind bezüglich ν ($P, Q \ll \nu$).

Definiere $p = \frac{dP}{d\nu}$, $q = \frac{dQ}{d\nu}$ und beachte, dass ein solches Maß ν immer existiert, da wir beispielsweise $\nu = P + Q$ betrachten können.

Definition 13. [Tsy09] (Hellinger Abstand)

Seien P, Q zwei Wahrscheinlichkeitsmaße die absolut stetig sind, bezüglich eines dritten Wahrscheinlichkeitsmaßes ν . Der Hellinger-Abstand zwischen P und Q ist definiert als

$$H(P, Q) = \left(\int (\sqrt{p} - \sqrt{q})^2 d\nu \right)^{\frac{1}{2}} = \left(\int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \right)^{\frac{1}{2}} := \left(\int (\sqrt{dP} - \sqrt{dQ})^2 \right)^{\frac{1}{2}} \quad (30)$$

Es ist leicht zu sehen, dass $H(P, Q)$ nicht abhängig ist von der Wahl des dominierenden Maßes ν . Dies erklärt die symbolische Notation auf der rechten Seite von (23).

Es folgen vier praktische Eigenschaften des Hellinger Abstandes

- (i) Der Hellinger Abstand definiert eine Metrik.
- (ii) $0 \leq H^2(P, Q) = \left(\int dP - \int 2\sqrt{dP \cdot dQ} + \int dQ \right) \leq 2$.
- (iii) $H^2(P, Q) = 2 \left(1 - \int \sqrt{pq} d\nu \right) = 2 \left(1 - \int \sqrt{dP dQ} \right)$.
- (iv) Wenn P, Q Produktmaße sind, $P = \otimes_{i=1}^n P_i$, $Q = \otimes_{i=1}^n Q_i$ dann ist

$$H^2(P, Q) = 2 \left(1 - \prod_{i=1}^n \left(1 - \frac{H^2(P_i, Q_i)}{2} \right) \right).$$

Es folgt eine weitere Abstandsdefinition.

Definition 14. [Tsy09] (*Gesamtabweichung*)

Die Gesamtabweichung zwischen P, Q ist definiert, als

$$V(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \sup_{A \in \mathcal{A}} \left| \int_A (p - q) d\nu \right|.$$

Wir haben zwei Eigenschaften

- (i) $V(P, Q)$ ist eine Metrik
- (ii) $0 \leq V(P, Q) \leq 1$

Beide Eigenschaften werden aus dem nächsten Lemma folgen. Definiere

$$\int \min(dP, dQ) := \int \min(p, q) d\nu.$$

Lemma 3. [Tsy09] (*Scheffè's Theorem*)

$$V(P, Q) = \frac{1}{2} \int |p - q| d\nu = 1 - \int \min(dP, dQ)$$

Beweis. Betrachte die Menge $A := \{x \in \mathcal{X} : q(x) \geq p(x)\}$. Dann ist

$$\int |p - q| d\nu = 2 \int_{A_0} (q - p) d\nu$$

und

$$V(P, Q) \geq Q(A_0) - P(A_0) = \frac{1}{2} \int |p - q| d\nu = 1 - \int \min(p, q) d\nu$$

auf der anderen Seite. □

3.3 Schranken auf Grundlage von Abständen

Damit wir Theorem 3 und Proposition 5 anwenden können, benötigen wir die Bedingung (27)

$$P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \geq 1 - \alpha$$

Diese Bedingung ist zwar sehr allgemein, aber nicht immer leicht zu überprüfen. Deswegen werden oft andere Schranken der Minimax Wahrscheinlichkeit des Fehlers für zwei Hypothesen genutzt und zwar auf Grundlage der Distanz oder der Divergenz von P_0 und P_1 . Drei Schranken sind in folgendem Theorem beschrieben.

Theorem 4. [Tsy09] Seien P_0, P_1 zwei Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$, dann erhalten wir drei Varianten um $p_{e,1}$ abschätzen zu können.

Gesamtvarianten Version

i) Sei $V(P_1, P_0) \leq \alpha < 1$, dann gilt

$$p_{e,1} \geq \frac{1 - \alpha}{2}.$$

Hellinger Version

ii) Sei $H^2(P_1, P_0) \leq \alpha < 2$, dann gilt

$$p_{e,1} \geq \frac{1}{2} \left(1 - \sqrt{\alpha(1 - \alpha/4)} \right).$$

Kullback Version

iii) Sei $K(P_1, P_0) \leq \alpha < \infty$ oder $(\chi^2(P_1, P_0) \leq \alpha < \infty)$, dann gilt

$$p_{e,1} \geq \max \left(\frac{1}{4} \exp(-\alpha), \frac{1 - \sqrt{\alpha/2}}{2} \right).$$

Beweis. Es ist

$$p_{e,1} = \inf_{\Psi} \max_{j=0,1} P_j(\Psi \neq j) \geq \frac{1}{2} \inf_{\Psi} (P_0(\Psi^* \neq 0) + P_1(\Psi^* \neq 1))$$

wobei Ψ^* hier den Maximum-Likelihood Test beschreibt, das heißt hier $\Psi^* = 0$, falls $p_0 > p_1$ gilt, ansonsten ist $\Psi^* = 1$. Es sind p_j die Dichtefunktionen von P_j bezüglich ν

$$P_j(E) = \int_E p_j d\nu \quad \forall E \in \mathcal{A}, \quad j = 0, 1.$$

Mit Scheffè's Theorem (Lemma 3) folgt

$$\frac{1}{2} (P_0(\Psi^* \neq 0) + P_1(\Psi^* \neq 1)) = \frac{1}{2} \int \min(dP_0, dP_1) = (1 - V(P_0, P_1))/2.$$

□

3.4 Untere Schranken des Risikos von Regressionsschätzern an einem Punkt

Wir nehmen in diesem Kapitel an, dass Annahme 1 gilt. Unser Ziel ist, eine untere Schranke zu finden für das Minimum Maximum Risiko auf (Θ, d) , wobei Θ wie oben definiert ist:

$$\Theta = \Sigma(\beta, L), \quad \beta > 0, \quad L > 0.$$

wobei d einen Abstand zu einem fixen Punkt $x_0 \in [0, 1]$ beschreibt,

$$d(f, g) = |f(x_0) - g(x_0)|.$$

Die Rate, die es zu zeigen gilt, wird wie im ersten Kapitel beschrieben durch

$$\psi_n = n^{-\frac{\beta}{2\beta+1}}. \quad (31)$$

Nach (3.1) wissen wir, dass es ausreicht, sich nach einer Schranke für die Wahrscheinlichkeit umzusehen,

$$\inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \geq c' > 0.$$

Wobei $s = A\psi_n$ mit einer Konstanten $A > 0$. Wenn wir nur 2 Hypothesen betrachten, erhalten wir mit gleicher Notation:

$$\inf_{T_n} \max_{f \in \{f_{0n}, f_{1n}\}} P_f(|T_n(x_0) - f(x_0)| \geq A\psi_n) \geq c' > 0. \quad (32)$$

Wobei $f_{0n}(\cdot) = \theta_0, f_{1n}(\cdot) = \theta_1$ zwei Hypothesen beschreiben. $A > 0$, und T_n beschreibt alle Schätzer. Anstatt (32) zu erhalten, werden wir die Kullback Version von Theorem 4 ($p_{e,1} \geq \frac{1}{2}(1 - \sqrt{\alpha(1 - \alpha/4)})$) und die Ungleichung (21) benutzen. Wir wählen die Hypothesen $\theta_0 = f_{0n}(\cdot)$ und $\theta_1 = f_{1n}(\cdot)$ folgendermaßen,

$$f_{0n}(x) \equiv 0, f_{1n}(x) = Lh^\beta K\left(\frac{x - x_0}{h_n}\right), x \in [0, 1]$$

wobei

$$h_n = c_0 n^{-\frac{1}{2\beta+1}}, c_0 > 0$$

und wobei die Funktion $K : \mathbb{R} \rightarrow [0, \infty)$ der Bedingung

$$K \in \Sigma(\beta, 1/2) \cap C^\infty(\mathbb{R}) \text{ und } K(u) > 0 \Leftrightarrow u \in (-1/2, 1/2) \quad (33)$$

genügt. Das heißt, wir fordern neben der Hölderstetigkeitsbedingung auch, dass K unendlich oft differenzierbar ist.

Beispiel 9. [Tsy09] Für genügend kleine $a > 0$, um die Hölderbedingung nicht zu verletzen, betrachten wir

$$K(u) = aK_0(2u) \text{ mit } K_0(u) = e^{\left(-\frac{1}{1-u^2}\right)} \mathbb{1}_{\{|u| \leq 1\}}. \quad (34)$$

Um Theorem 4 und (21) benutzen zu können, müssen wir 3 Bedingungen überprüfen:

- (i) $f_{jn} \in \Sigma(\beta, L)$ $j = 0, 1$
- (ii) $d(f_{1n}, f_{0n}) \geq 2s$
- (iii) $K(P_0, P_1) \leq \alpha < \infty$

Im Folgenden wird gezeigt, dass die drei Bedingungen gelten für ausreichend kleine c_0 , beziehungsweise große n .

i) für $l = \lfloor \beta \rfloor$, die l -te Ableitung von f_{1n} ist:

$$f_{1n}^{(l)}(x) = Lh_n^{\beta-l} K^{(l)}\left(\frac{x - x_0}{h_n}\right)$$

dann folgt mit unserer Bedingung (33) an den Kern:

$$|f_{1n}^{(l)}(x) - f_{1n}^{(l)}(x')| = Lh_n^{\beta-l} |K^{(l)}(u) - K^{(l)}(u')| \leq Lh_n^{\beta-l} |u - u'|/2 = L|x - x'|^{\beta-l}/2 \quad (35)$$

Das bedeutet, dass f_{1n} zur Hölderklasse gehört, woraus folgt, dass f_{1n} beschränkt auf $[0, 1]$ zur Klasse $\Sigma(\beta, L/2)$ auf $[0, 1]$ gehört.

ii) Wir haben $d(f_{1n}, f_{0n}) = |f_{1n}(x_0)| = Lh_n^\beta K(0) = Lc_0^\beta K(0)n^{-\frac{\beta}{2\beta+1}}$, das bedeutet, die Bedingung $d(f_{1n}, f_{0n}) \geq 2s$ gilt für:

$$s = s_n = \frac{1}{2} Lc_0^\beta K(0)n^{-\frac{\beta}{2\beta+1}} := An^{-\frac{\beta}{2\beta+1}} = A\psi_n.$$

iii) Beachte dazu, dass P_j (die Verteilung von Y_1, \dots, Y_n für $f = f_{jn}$) folgende Verteilung bezüglich des Lebesgue Maßes liefert:

$$p_j(u_1, \dots, u_n) = \prod_{i=1}^n p_\xi(u_i - f_{jn}(X_i)), j = 0, 1.$$

dann existiert ein $n \in \mathbb{N}$ in Abhängigkeit von $c_0, L, \beta, K_{max}, v_0$, sodass für alle $n > n_0$ gilt: $nh_n \geq 1$ und $Lh_n^\beta K_{max} \leq v_0$ mit $K_{max} = \max_u K(u)$. Dann haben wir mit Annahme 3 und Annahme 1 für $n > n_0$:

$$\begin{aligned}
K(P_0, P_1) &= \int \log \frac{dP_0}{dP_1} dP_0 \\
&= \int \dots \int \log \prod_{i=1}^n \frac{p_\xi(u_i)}{p_\xi(u_i - f_{1n}(X_i))} \prod_{i=1}^n [p_\xi(u_i) du_i] \\
&= \sum_{i=1}^n \int \log \frac{p_\xi(y)}{p_\xi(y - f_{1n}(X_i))} p_\xi(y) dy \\
&\leq p_* \sum_{i=1}^n f_{1n}^2(X_i) p_* L^2 h_n^{2\beta} \sum_{i=1}^n K^2 \left(\frac{X_i - x_0}{h_n} \right) \\
&\leq p_* L^2 h_n^{2\beta} K_{max}^2 \sum_{i=1}^n \mathbb{1}_{\{|\frac{X_i - x_0}{h_n}| \leq 1\}} \\
&\leq p_* a_0 L^2 h_n^{2\beta} K_{max}^2 \max(nh_n, 1) \\
&= p_* a_0 L^2 K_{max}^2 n h_n^{2\beta+1}.
\end{aligned} \tag{36}$$

Es beschreibt a_0 die Konstante aus Annahme 1, wenn wir c_0 folgendermaßen definieren:

$$c_0 = \left(\frac{\alpha}{p_* a_0 L^2 K_{max}^2} \right)^{\frac{1}{2\beta+1}}$$

Dann folgt mit $h_n = c_0 n^{-\frac{1}{2\beta+1}}$, dass $K(P_0, P_1) \leq \alpha$ ist. Nach der Kullback-Version von Theorem 4 impliziert das obige Argument, dass für $n > n_0$ beliebig und T_n beliebig:

$$\begin{aligned}
\sup_{f \in \Sigma(\beta, L)} P_f(|T_n(x_0) - f(x_0)| \geq s_n) &\geq \max_{j=0,1} P_j(|T_n(x_0) - f_j(x_0)| \geq s_n) \\
&\geq \max \left(\frac{1}{4} e^{-\alpha}, \frac{1 - \sqrt{\alpha/2}}{2} \right) \\
&= V_0(\alpha)
\end{aligned} \tag{37}$$

Dies liefert uns das folgende Ergebnis:

Theorem 5. [Tsy09] Wir nehmen $\beta > 0$ und $L > 0$ an. Unter den Annahmen 1 und 3 erhalten wir für alle $x_0 \in [0, 1], t > 0$

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} P_f \left(n^{\frac{\beta}{2\beta+1}} |T_n(x_0) - f(x_0)| \geq t^{\frac{\beta}{2\beta+1}} \right) \geq V_0(ct) \tag{38}$$

wobei über das Infimum aller Schätzer T_n iteriert wird und $c_0 > 0$ nur von β, L, p_* und a abhängt. Außerdem

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \mathbb{E} \left[n^{\frac{\beta}{2\beta+1}} (T_n(x_0) - f(x_0))^2 \right] \geq c_1 \tag{39}$$

wobei c_1 nur von β, L, p_* und a_0 abhängt.

Betrachten wir das nichtparametrische Regressionsmodell unter folgenden zwei Annahmen.

(i) $X_i = i/n$

(ii) Die Zufallsvariablen ξ_i sind i.i.d mit Dichte p_{x_i} genügen Annahme 2 und:

$$\mathbb{E}(\xi_i) = 0, \quad \mathbb{E}(\xi_i^2) < \infty$$

Dann ist für $\beta > 0$ und $L > 0$ die Konvergenzrate $\psi_n = n^{-\frac{\beta}{2\beta+1}}$ optimal auf $(\Sigma(\beta, L), d_0)$ mit d_0 als Abstand eines fest gewählten Punktes $x_0 \in [0, 1]$

Die untere Schranke basierend auf zwei Hypothesen stellt sich als ungünstig heraus, wenn wir Schätzungen in L^p Abständen betrachten. Betrachten wir als Beispiel den L^2 Abstand:

$$d(f, g) = \|f - g\|_2 = \left(\int_0^1 (f(x) - g(x))^2 dx \right)^{\frac{1}{2}}$$

und nehmen wir außerdem noch an, die zwei obigen Annahmen 1 und 2 gelten. Wir versuchen eine Technik für zwei Hypothesen anzuwenden, mit f_{0n} und f_{1n} (mit $x_0 = 1/2$ als Beispiel)

$$f_{0n}(x) \equiv 0,$$

$$f_{1n}(x) = Lh_n^\beta K\left(\frac{x - 1/2}{h_n}\right).$$

für K gilt (33) wie oben, $h_n > 0$. Jetzt verwenden wir die Kullback-Version von Theorem 4, die Bedingung $K(P_0, P_1) \leq \alpha < \infty$ und Ungleichung (36) $K(P_0, P_1) \leq p_* a_0 L^2 K_{max}^2 n h_n^{2\beta+1}$. dies führt uns zu folgender Einschränkung von h_n :

$$\limsup_{n \rightarrow \infty} n h_n^{2\beta+1} < \infty$$

In anderen Worten heißt das, wir erhalten $h_n = \mathcal{O}(n^{-\frac{1}{2\beta+1}})$ wie in Kapitel 1.

$$\begin{aligned} d(f_{0n}, f_{1n}) &= \|f_{0n} - f_{1n}\|_2 = \left(\int_0^1 f_{1n}^2(x) dx \right)^{\frac{1}{2}} \\ &= Lh_n^\beta \left(\int_0^1 K^2\left(\frac{x - 1/2}{h_n}\right) dx \right)^{\frac{1}{2}} \\ &= Lh_n^{\beta+\frac{1}{2}} \left(\int K^2(u) du \right)^{\frac{1}{2}} \end{aligned} \tag{40}$$

für n genügend groß. Deswegen

$$0 < d(f_{0n}, f_{1n}) \asymp h_n^{\beta+1/2} = \liminf_{n \rightarrow \infty} (d(f_{0n}, f_{1n})/h_n^{\beta+1/2}) \leq \limsup_{n \rightarrow \infty} (d(f_{0n}, f_{1n})/h_n^{\beta+1/2}) < \infty$$

ist von der Ordnung $\mathcal{O}(n^{-\frac{1}{2}})$ was impliziert, dass unsere Ungleichung aus Annahme 3 nur benutzt werden kann für:

$$s \leq d(f_{0n}, f_{1n})/2 = \mathcal{O}(n^{-\frac{1}{2}}).$$

Abschließend können wir also sagen, dass die Technik, welche auf Grundlage zweier Hypothesen eine untere Schranke ermittelt, mit einer Rate von $n^{-1/2}$ ein nicht zufriedenstellendes Ergebnis liefert. Die Schranke ist wesentlich kleiner, als $n^{-\frac{\beta}{2\beta+1}}$ welche in der unteren Schranke des L_2 -Abstandes auf $\Sigma(\beta, L)$ vorkommt. Dieses Problem kann gelöst werden, indem man zu M Hypothesen wechselt, wobei M nach unendlich strebt für $n \rightarrow \infty$.

Lemma 4. [Tsy09] Seien P_0, P_1, \dots, P_M Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$, dann ist

$$p_{e,M} \geq \sup_{\tau > 0} \frac{\tau M}{1 + \tau M} \left[\frac{1}{M} \sum_{j=1}^M P_j \left(\frac{dP_{0,j}^a}{P_j} \geq \tau \right) \right]$$

wobei $dP_{0,j}^a$ die absolut stetig Komponente des Maßes P_0 bezüglich P_j ist.

Beweis. Sei hierfür Ψ ein Test mit Werten in $\{0, 1, \dots, M\}$. Dann ist klar

$$\bigcup_{j=1}^M \{\Psi = j\} = \{\Psi \neq 0\},$$

und

$$\{\Psi = j\} \cap \{\Psi = k\} = \emptyset, k \neq j.$$

Mit dem Zufallsereignis $A_j = \left\{ \frac{dP_{0,j}^a}{P_j} \geq \tau \right\}$ können wir schreiben

$$\begin{aligned} P_0(\Psi \neq 0) &= \sum_{j=1}^M P_0(\Psi = j) \geq \sum_{j=1}^M P_{0,j}^a(\Psi = j) \\ &\geq \sum_{j=1}^M \tau P_j^a(\{\Psi = j\} \cap A_j) \\ &\geq \tau M \left(\frac{1}{M} \sum_{j=1}^M P_0(\Psi = j) \right) - \tau \sum_{j=1}^M P_j(A_j^c) \\ &= \tau M(p_0 - \alpha). \end{aligned} \tag{41}$$

Mit A_j^c als Komplement von A_j ist,

$$p_0 = \frac{1}{M} \sum_{j=1}^M P_0(\Psi = j), \quad \frac{1}{M} \sum_{j=1}^M P_j \left(\frac{dP_{0,j}^a}{P_j} < \tau \right).$$

Dann erhalten wir

$$\begin{aligned} \max_{0 \leq j \leq M} P_j(\Psi \neq j) &= \max\{P_0(\Psi = 0), \max_{1 \leq j \leq M} P_j(\Psi \neq j)\} \\ &\geq \max\{\tau M(p_0 - \alpha), \frac{1}{M} \sum_{j=1}^M P_j(\Psi \neq j)\} \\ &= \max\{\tau M(p_0 - \alpha), 1 - p_0\} \\ &\geq \min_{0 \leq p \leq 1} \max\{\tau M(p - \alpha), 1 - p\} \\ &= \frac{\tau M(1 - \alpha)}{1 + \tau M}. \end{aligned} \tag{42}$$

□

Theorem 6. [Tsy09] *Haupttheorem für untere Schranken des Risikos*

Wir nehmen an Θ enthält Elemente $\theta_0, \theta_1, \dots, \theta_M$ derart, dass

(i) $d(\theta_j, \theta_k) \geq 2s > 0, \forall 0 \leq j < k \leq M$

(ii) *es gibt $\tau > 0$ und $0 < \alpha < 1$*

sodass

$$\frac{1}{M} \sum_{j=1}^M P_j \left(\frac{dP_{0,j}^a}{dP_j} \geq \tau \right) \geq 1 - \alpha. \tag{43}$$

Wobei $P_{0,j}^a$ den absolut stetigen Teil des Maßes $P_0 = P_{\theta_0}$ bezüglich $P_j = P_{\theta_j}$ beschreibt. Dann haben wir folgende Abschätzung

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} P_{\tilde{\theta}}(d(\tilde{\theta}, \theta) \geq s) \geq \frac{\tau M}{1 + \tau M} (1 - \alpha). \tag{44}$$

Der Beweis folgt direkt aus Lemma 4 und Annahme 3.

Beweis.

$$\begin{aligned}
\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) &\stackrel{(3.1)}{\geq} \inf_{\psi} \max_{0 \leq j \leq M} P(\Psi \neq j) \\
&\stackrel{\text{Lemma 4}}{\geq} \sup_{\tau > 0} \frac{\tau M}{1 + \tau M} \left[\frac{1}{M} \sum_{j=1}^M P_j \left(\frac{dP_{0,j}^a}{dP_j} \geq \tau \right) \right] \\
&\stackrel{(44)}{\geq} \frac{\tau M}{1 + \tau M} (1 - \alpha)
\end{aligned} \tag{45}$$

□

Für $M = 1$ stimmen Lemma 4 und Theorem 6 mit Proposition 4 und Theorem 3 überein.

Proposition 6. [Tsy09] Seien P_0, P_1, \dots, P_M Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ welche

$$\frac{1}{M} \sum_{j=1}^M K(P_j, P_0) \leq \alpha_* \tag{46}$$

mit $0 < \alpha_* < \infty$ genügen, dann gilt

$$p_{e,M} \geq \sup_{0 < \tau < 1} \left[\frac{\tau M}{1 + \tau M} \left(1 + \frac{\alpha_* + \sqrt{\alpha_*/2}}{\log(\tau)} \right) \right].$$

3.4.1 Untere Schranken in L_2

Theorem 6 erlaubt es untere Schranken für das L_p –*Risiko* mit optimalen Raten zu finden. Zur Veranschaulichung wird das nicht-parametrische Regressionsmodell unter Annahme 2 herangezogen und im Folgenden das L_2 –*Risiko* etwas genauer betrachtet. Es ist

$$d(f, g) = \|f - g\|_2 = \left(\int_0^1 (f(x) - g(x))^2 dx \right)^{\frac{1}{2}}. \tag{47}$$

Das erste Ziel ist ein Beweis für die untere Schranke von (19) des Minimax Risikos für die Hölder Klasse $\Theta = \Sigma(\beta, L)$ und den L_2 –*Abstand* (47), mit der Konvergenzrate $\psi_n = n^{-\frac{\beta}{2\beta+1}}$. Sei außerdem $M \in \mathbb{N}$ die Anzahl der Hypothesen. Wir wählen unsere Hypothesen $(f_{nj} \in \Theta)$

$$\theta_j = f_{jn}(\cdot), \quad j = 0, \dots, M.$$

Wir entscheiden uns nach dem Reduktionsschema (3.1) dafür

$$\inf_{\tilde{\theta}_n} \max_{\theta \in \{\theta_0, \dots, \theta_M\}} P_{\theta}(d(\tilde{\theta}, \theta) \geq s) \geq c'$$

zu betrachten. Wobei wir $A > 0$ fordern und $s = A\psi_n$. Da wir allerdings d als L_2 –*Abstand* wählen und $\Theta = \Sigma(\beta, L)$ setzen, erhalten wir

$$\inf_{T_n} \max_{f \in \{f_{0n}, \dots, f_{Mn}\}} P_f(\|T_n - f\|_2 \geq A\psi_n) \geq c'$$

Bevor wir zum eigentlichen Beweis kommen, werden wir Hypothesen konstruieren, was einige Zeit in Anspruch nehmen wird.

[Tsy09]Konstruktion der Hypothesen

f_{jn} Sei $c_0 \in \mathbb{R}$ und $M \geq 1$ ganzzahlig. $\lceil x \rceil$ ist die kleinste ganzzahlige Zahl x welche echt größer ist als x (es wird aufgerundet). Wir definieren

$$m = \lceil c_0 n^{-\frac{1}{2\beta+1}} \rceil, \quad h_n = \frac{1}{m}, \quad x_k = \frac{k - 1/2}{m},$$

$$\varphi_k(x) = L_n^\beta K\left(\frac{x - x_k}{h_n}\right), \quad k = 1, \dots, m, \quad x \in [0, 1]$$

Die Funktion K erfüllt (33). Wegen (35) sind alle Funktionen $\varphi_k \in \Sigma(\beta, L)$ enthalten. Betrachten wir weiterhin die Menge aller binären Folgen der Menge m

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m), \omega_i \in \{0, 1\}\} = \{0, 1\}^m.$$

Die Hypothesen wählen wir aus

$$\mathcal{E} = \left\{ f_\omega(x) = \sum_{k=1}^m \omega_k \varphi_k(x), \omega \in \Omega \right\}.$$

Es gilt für alle ω, ω' , dass

$$\begin{aligned} d(f_\omega, f_{\omega'}) &= \left[\int_0^1 (f_\omega(x) - f_{\omega'}(x))^2 dx \right]^{\frac{1}{2}} \\ &= \left[\int_0^1 \left(\sum_{k=1}^m \omega_k \varphi_k(x) - \sum_{k=1}^m \omega'_k \varphi_k(x) \right)^2 dx \right]^{\frac{1}{2}} \\ &= \left[\sum_{k=1}^m (\omega_k - \omega'_k)^2 \int_{\Delta_k} \varphi_k^2 dx \right]^{\frac{1}{2}} \\ &= L h_n^{2\beta+1/2} \|K\|_2 \left[\sum_{k=1}^m (\omega_k - \omega'_k)^2 \right]^{1/2} \\ &= L h_n^{2\beta+1/2} \|K\|_2 \sqrt{\rho(\omega, \omega')}. \end{aligned} \tag{48}$$

Das ρ steht für den *Hamming Abstand*, $\rho(\omega, \omega') = \sum_{k=1}^m \mathbb{1}_{\omega_k \neq \omega'_k}$. Das Kürzel Δ_k definiert Intervalle wie folgt,

$$\Delta_1 = [0, 1/m], \quad \Delta_k = [(k-1)/m, k/m], \quad k = 2, \dots, m.$$

Die Menge $\{f_{j_n}, j = 1, \dots, M\}$ wird zusammengesetzt aus bestimmten Funktionen $f_{j_n} \in \mathcal{E}$. Um die Kullback Version des Haupttheorems benutzen zu können, werden wir eine Anforderung an den jeweiligen Abstand der Hypothesen stellen. Das heißt, es soll gelten

$$d(f_\omega, f_{\omega'}) \geq 2s_n \asymp n^{-\frac{2\beta}{2\beta+1}}.$$

Es genügt daher die ω, ω' so auszuwählen, dass $\sqrt{\rho(\omega, \omega')} \asymp h_n^{-1/2} \Leftrightarrow \rho(\omega, \omega') \asymp m$ gilt. Man kann mithilfe der *Varshamov – Gilbert – Schranke* eine untere Schranke für Kardinalität der Menge aller binären Sequenzen bestimmen.

Anwendung der Kullback Version des Haupttheorems

[Tsy09] Wir wählen $\alpha \in (0, 1/8)$ fixiert. Um das Theorem anwenden zu können, gilt es drei Bedingungen zu überprüfen.

- (i) Es ist f_{j_n} Hölder-stetig für alle $j = 1, \dots, M$
- (ii) Die Hypothesen sind $d(\theta_j, \theta_k) = \|f_{j_n} - f_{k_n}\|_2 \geq 2s > 0$
- (iii) $\frac{1}{M} K(P_j, P_0) \leq \alpha \log(M)$

Beginnen wir mit (i). Da $\varphi_k \in \Sigma(\beta, L/2)$ enthalten ist, $|\omega_i| \leq 1$ und die φ_k disjunkte Träger besitzen, folgt $f_\omega \in \Sigma(\beta, L)$ für alle $\omega \in \Omega$.

(ii) Wir werden die Varshamov Gilbert Schranke und (48) benutzen.

$$\begin{aligned}
\|f_{jn} - f_{kn}\|_2 &= \|f_{\omega^{(j)}} - f_{\omega^{(k)}}\|_2 \\
&\stackrel{(51)}{=} Lh_n^{2\beta+1/2} \|K\|_2 \sqrt{\rho(\omega^{(j)}, \omega^{(k)})} \\
&\stackrel{VG}{\geq} Lh_n^{2\beta+1/2} \|K\|_2 \sqrt{\frac{m}{16}} \\
&= \frac{L}{4} h_n^\beta \|K\|_2 = \frac{L}{4} m^{-\beta} \|K\|_2,
\end{aligned} \tag{49}$$

für alle $m \geq 8$. Nehmen wir an, dass $n > n_*$ mit $n_* = (7/c_0)^{2\beta+1}$. Dann ist $m \geq 8$ und $m^\beta \leq (1 + 1/7)^\beta c_0^\beta n^{\frac{\beta}{2\beta+1}}$, dass heißt

$$\|f_{jn} - f_{kn}\|_2 \geq 2s,$$

mit

$$s = An^{-\frac{\beta}{2\beta+1}}, \quad A = \frac{L}{8} \|K\|_2 (2c_0)^{-\beta}.$$

(iii) Wir verfahren wie in (36), wo gezeigt wurde, dass $K(P_0, P_1) \leq p_* a_0 L^2 K_{max}^2 n h_n^{2\beta+1}$. Für $n > n_*$ betrachten wir,

$$\begin{aligned}
K(P_j, P_0) &\leq p_* \sum_{i=1}^n f_{jn}^2(X_i) \leq p_* \sum_{k=1}^m \sum_{X_i \in \Delta_k} \varphi_k^2(X_i) \\
&\leq p_* L^2 K_{max}^2 h_n^{2\beta} \sum_{k=1}^m \text{Card}\{i : X_i \in \Delta_k\} \\
&= P_* L^2 K^2 n h_n^{2\beta} \leq p_* L^2 K_{max}^2 c_0^{-(2\beta+1)} m.
\end{aligned} \tag{50}$$

wegen der Varshamov Gilbert Schranke wissen wir, $M \geq 2^{m/8} \Leftrightarrow m \leq 8 \log(M)/\log(2)$. Wir wählen

$$c_0 = \left(\frac{8p_* L^2 K_{max}^2}{\alpha \log(2)} \right)^{\frac{1}{2\beta+1}}$$

und somit folgt $K(P_j, P_0) < \alpha \log(M)$ für $j = 1, \dots, M$.

Das heißt, die Bedingungen (i) – (iii) sind erfüllt und das wiederum heißt, es gilt für einen beliebigen Schätzer T_n ,

$$\max_{f \in \{f_{0n}, \dots, f_{Mn}\}} P_f(\|T_n - f\|_2 \geq A\psi_n) \geq \frac{\sqrt{M}}{1 + \sqrt{M}} \left(1 - 2\alpha \sqrt{\frac{2\alpha}{\log(M)}} \right).$$

Also haben wir mithilfe der Reduzierung auf eine endliche Anzahl von Hypothesen eine untere Schranke für das maximale Risiko gefunden. Unter Verwendung von Annahme 3, können wir folgendes Theorem formulieren.

Theorem 7. [Tsy09] Sei $\beta > 0$ und $L > 0$, dann gilt

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} E_f[n^{\frac{2\beta}{2\beta+1}} \|T_n - f\|_2^2] \geq c.$$

Wobei c nur abhängig ist von den konstanten β, L, p_* .

3.4.2 Untere Schranken bezüglich der Sup-Norm

Betrachten wir hierzu eine weitere Definition über die wir unseren Abstand definieren können,

$$d(f, g) = \|f - g\|_\infty = \sup_{x \in [0, 1]} |f(x) - g(x)|.$$

Das Ziel ist die Bestimmung einer Schranke des minimax Risikos bezüglich der Konvergenzrate

$$\psi_n = \left(\frac{\log(n)}{n} \right)^{\frac{\beta}{2\beta+1}},$$

wobei für die nichtparametrische Klasse von Funktionen $\Theta = \Sigma(\beta, L)$ gilt. Wir verwenden wie bei der Abschätzung bezüglich der L_2 -Norm die Kullback Version des Haupttheorems. Wir wählen die Hypothesen

$$\theta_0 = f_{0n}(\cdot) \equiv 0, \quad \theta_j = f_{jn}(\cdot) \quad j = 1, \dots, M,$$

derart, dass

$$f_{jn}(x) = Lh_n^\beta K\left(\frac{x - x_j}{h_n}\right), \quad x_j = \frac{j - 1/2}{M}, \quad h_n = 1/M.$$

Der Kern $K : \mathbb{R} \rightarrow [0, \infty)$ erfüllt dabei (33) und $M > 1$ ist ganzzahlig. Sei $\alpha \in (0, 1/8)$. Um die Kullback-Version des Haupttheorems wieder anwenden zu können, müssen wir wie in 3.8.2 die drei Bedingungen des Theorems überprüfen.

- (i) Es ist $f_{jn} \in \Sigma(\beta, L)$ für alle $j = 1, \dots, M$
 - (ii) Es gilt $d(\theta_j, \theta_k) = \sup_{x \in [0,1]} |f_{jn} - f_{kn}| \geq 2s > 0 \quad k \neq j$
 - (iii) $\frac{1}{M} K(P_j, P_0) \leq \alpha \log(M)$
- (i) folgt mit (35).
(ii) Es ist

$$d(f_{jn}, f_{kn}) = \|f_{jn} - f_{kn}\|_\infty \geq Lh_n^\beta K(0) := 2s,$$

mit $s := \frac{Lh_n^\beta K(0)}{2}$. Unsere Anforderungen an s sind das gleiche Konvergenzverhalten, wie die einführend gewünschte Rate ψ_n . Das heißt wir fordern,

$$s \asymp \psi_n = \left(\frac{\log(n)}{n} \right)^{\frac{\beta}{2\beta+1}}. \quad (51)$$

Dies wird erreicht, indem wir einfach $h \asymp \left(\frac{\log(n)}{n} \right)^{\frac{1}{2\beta+1}}$ wählen. Eine geschickte Wahl ist $h_n = 1/M$, sodass

$$M = \left\lceil c_0 \left(\frac{n}{\log(n)} \right)^{\frac{1}{2\beta+1}} \right\rceil.$$

c_0 ist eine konstante, die im späteren Verlauf gewählt wird.

(iii) Jetzt benötigen wir (39), das heißt die Abschätzung $K(P_0, P_1) \leq \alpha$.

$$\begin{aligned} \frac{1}{M} \sum_{j=1}^M K(P_j, P_0) &\leq \frac{1}{M} \sum_{j=1}^M p_* \sum_{i=1}^n f_{jn}^2(X_i) \\ &\leq p_* L^2 K_{max}^2 h_n^{2\beta} \frac{1}{M} \sum_{j=1}^M \text{Card}\{i : X_i \in \text{supp}(f_{jn})\} \\ &= p_* L^2 K_{max}^2 h_n^{2\beta} \frac{n}{M} \stackrel{h:=1/M}{=} p_* L^2 K_{max}^2 M^{-2\beta+1} n \\ &\leq p_* L^2 K_{max}^2 c_0^{-2\beta+1} \log(n) \end{aligned} \quad (52)$$

Es ist

$$\log(M) \geq \log \left(c_0 \left(\frac{n}{\log(n)} \right)^{\frac{1}{2\beta+1}} \right) = \frac{\log(n)}{2\beta+1} (1 + o(1)) \geq \frac{\log(n)}{2\beta+2}$$

für n genügend groß. Hieraus folgt für c_0 entsprechend groß, die untere Schranke in der Supremumsnorm bezüglich unserer geforderten Rate ψ_n (vgl. (54)).

Theorem 8. [Tsy09] Seien $\beta, L > 0$, unter Annahme 3 wurde gezeigt, dass

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \left(\frac{n}{\log(n)} \right)^{\frac{2\beta}{2\beta+1}} E_f[\|T_n - f\|_\infty^2] \geq c.$$

Das \inf_{T_n} beschreibt das Infimum aller Schätzer. $c > 0$ ist abhängig von β, L, p_* .

Wir benötigen eine Verschärfung der Annahme 3 für folgendes Theorem. Das heißt wir fordern Gültigkeit von 1. und 2. Außerdem seien X_i beliebige Zufallsvariablen auf $[0, 1]$ und zwar derart, dass (X_1, \dots, X_n) unabhängig ist von (ξ_1, \dots, ξ_n) .

3.4.3 Konvergenzraten in Supremums- und L_2 -Norm

Lemma 5. [Tsy09] Seien P_0, P_1, \dots, P_M Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ und $M \geq 2$. Wenn

$$\frac{1}{M+1} \sum_{j=1}^M K(P_j, P_0) \leq \alpha \log(M)$$

mit $0 < \alpha < 1$, dann ist

$$p_{e,M} \geq \bar{p}_{e,M} \frac{\log(M+1) - \log(2)}{\log(M)} - \alpha. \quad (53)$$

Theorem 9. [Tsy09]

Sei $\beta, L > 0$ und Annahme 3 sei erfüllt, dann gilt

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f[\psi_{n,p} \|T_n(x_0) - f(x_0)\|_p^2] \geq c$$

für

$$\psi_{n,2} = n^{\frac{\beta}{2\beta+1}} \quad \text{und} \quad \psi_{n,\infty} = \left(\frac{\log(n)}{n} \right)^{\frac{\beta}{2\beta+1}}.$$

Das heißt, wir erhalten hier zwei Konvergenzraten ψ abhängig davon, ob wir die Differenz von T_n, f im L_2 -Abstand oder bezüglich der Supremumsnorm betrachten. wobei wir über das Infimum aller Schätzer T_n iterieren und wobei unsere Konstante ausschließlich von β, L und p_* abhängt. Wir benötigen für den nächsten Beweis eine nützliche Eigenschaft des bedingten Erwartungswertes (siehe Definition 17).

Beweis. Sei f_{0n}, \dots, f_{Mn} . Nach Konstruktion ist $\|f_{jn} - f_{kn}\|_p \geq 2s, j \neq k$ mit $s = A\psi_{n,p}$ und $A > 0$. Wir bezeichnen mit E_{X_1, \dots, X_n} die Erwartung bezüglich der gemeinsamen Verteilung von X_1, \dots, X_n und setzen $P_j = P_{f_{jn}}$. Dann haben wir für einen beliebigen Schätzer T_n ,

$$\begin{aligned} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f[\psi_{n,p} \|T_n(x_0) - f(x_0)\|_p^2] &\geq A^2 \sup_{f \in \Sigma(\beta, L)} P_f \left(\|T_n - f\|_p \geq A\psi_{n,p} \right) \\ &\geq A^2 \max_{f \in \{f_{0n}, \dots, f_{Mn}\}} P_f \left(\|T_n - f\|_p \geq A\psi_{n,p} \right) \\ &= A^2 \max_{f \in \{f_{0n}, \dots, f_{Mn}\}} \mathbb{E}_f[\mathbb{1}_{(\|T_n - f\|_p \geq s)}] \\ &= A^2 \max_{f \in \{f_{0n}, \dots, f_{Mn}\}} \mathbb{E}_{X_1, \dots, X_n}[\mathbb{1}_{(\|T_n - f\|_p \geq s)} | \sigma(X_1, \dots, X_n)] \\ &\geq A^2 \frac{1}{M+1} \sum_{j=0}^M E_{X_1, \dots, X_n} [P_j \left(\|T_n - f\|_p \geq s | \sigma(X_1, \dots, X_n) \right)] \\ &= A^2 E_{X_1, \dots, X_n} \left[\frac{1}{M+1} \sum_{j=0}^M P_j \left(\|T_n - f\|_p \geq s | \sigma(X_1, \dots, X_n) \right) \right] \\ &\geq A^2 E_{X_1, \dots, X_n} \left[\inf_{\Psi} \frac{1}{M+1} \sum_{j=0}^M P_j \left(\Psi \neq j | X_1, \dots, X_n \right) \right]. \end{aligned} \quad (54)$$

Die erste Ungleichung folgt direkt mit der Markov Ungleichung, der zweite mit (21), also der Reduzierung auf eine endliche Anzahl von Hypothesen. Der letzte Teil folgt mit (24). Seien X_1, \dots, X_n gegeben und

$$\frac{1}{M+1} \sum_{j=1}^M K(P_j, P_0) \leq \alpha \log(M)$$

mit $0 < \alpha < 1/8$ erfüllt, mit Proposition 6 können wir dann folgende Abschätzung machen.

$$\bar{p}_{e,M} = \inf_{\Psi} \frac{1}{M+1} \sum_{j=0}^M P_j(\Psi \neq j | X_1, \dots, X_n) \geq \frac{\log(M+1) - \log(2)}{\log(M)} - \alpha,$$

Wegen der Unabhängigkeit der Rechten Seite von X_1, \dots, X_n sind wir fertig. Also gilt

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f[\psi_{n,p} ||T_n(x_0) - f(x_0)||_p^2] \geq A^2 E_{X_1, \dots, X_n} \left[\frac{\log(M+1) - \log(2)}{\log(M)} - \alpha \right]. \quad (55)$$

□

3.4.4 Die Methode der zwei Fuzzy-Hypothesen

Betrachten wir eine Verallgemeinerung der Zweihypothesentechnik (vergleiche hierzu Theorem 3 und 4). Die Ergebnisse dieses Abschnittes können benutzt werden, um untere Schranken auch für Funktionale und parameterfreie Tests zu bestimmen. Denn anders als sich vielleicht vermuten lässt, lassen sich die Schranken für die Funktionale hierfür auf die gleiche Art wie oben bestimmen. Sei $F(\theta)$ hierzu ein Funktional auf einem Messraum (Θ, \mathcal{U}) , das Werte in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ annimmt. Jetzt geht es darum, $F(\theta)$ auf Grundlage von Beobachtungen \mathbb{X} bezüglich des Statistischen Modells $\{P_\theta, \theta \in \Theta\}$ zu schätzen, wobei die Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ leben. Da in der Regel $\mathbb{X}, P_\theta, \mathcal{X}, \mathcal{A}$ von der Stichprobengröße n abhängen, wird dies im weiteren Verlauf nicht durch Indizierung ausgedrückt. Sei F_n ein Schätzer von $F(\theta)$. Für eine gegebene Verlustfunktion ω und eine Konvergenzrate ψ_n definieren wir das Maximumrisiko von F_n wie in der Einleitung von Kapitel 3, als

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta[\omega(\psi^{-1} |\tilde{F}_n - F(\theta)|)] \quad (56)$$

Das Ziel hier ist das Finden einer nicht trivialen Schranke für (56).

Zuerst wird das Minimax Risiko wieder mit der Markov Ungleichung abgeschätzt.

$$\inf_{\tilde{F}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\omega(\psi^{-1} |\tilde{F}_n - F(\theta)|)] \geq \omega(A) \inf_{\tilde{F}_n} \sup_{\theta \in \Theta} P_\theta(|\tilde{F}_n - F(\theta)| \geq A\psi_n)$$

Es ist bis jetzt das gleiche Vorgehen wie in 3.1. Im folgenden wird allerdings die Wahrscheinlichkeit nicht wieder über eine endliche Anzahl von Hypothesen nach unten abgeschätzt, sondern es werden zwei Wahrscheinlichkeitsmaße eingeführt. Seien μ_0, μ_1 Wahrscheinlichkeitsmaße auf (Θ, \mathcal{U}) , dann können wir folgende Abschätzung betrachten

$$\sup_{\theta \in \Theta} P_\theta(|\tilde{F}_n - F(\theta)| \geq A\psi_n) \geq \max \left\{ \int \sup_{\theta \in \Theta} P_\theta(|\tilde{F}_n - F(\theta)| \geq s) \mu_0(d\theta), \sup_{\theta \in \Theta} P_\theta(|\tilde{F}_n - F(\theta)| \geq s) \mu_1(d\theta) \right\} \quad (57)$$

Die hier eingeführten Maße werden *Fuzzy-Hypothesen* genannt (etwa unscharf oder verschwommen), da sich ihr Volumen oder ihre Masse über der ganzen Menge Θ verteilen kann. Jetzt definieren wir noch zwei »post«-Wahrscheinlichkeitsmaße (in Abhängigkeit von μ_1, μ_2) $\mathbb{P}_0, \mathbb{P}_1$ auf $(\mathcal{X}, \mathcal{A})$ wie folgt

$$\mathbb{P}_j(S) = \int P_\theta(S) \mu_j(d\theta), \quad \forall S \in \mathcal{A}, \quad j = 0, 1$$

Theorem 10. [Tsy09] nehmen wir an

(i) Es gibt $c \in \mathbb{R}, s > 0, 0 \leq \beta_0, \beta_1 < 1$, so dass

$$\begin{aligned}\mu_0(\theta : F(\theta) \leq c) &\geq 1 - \beta_0 \\ \mu_1(\theta : F(\theta) \geq c + 2s) &\geq 1 - \beta_1\end{aligned}\tag{58}$$

(i) Es existiert $\tau > 0$ und $0 < \alpha < 1$, so dass

$$\mathbb{P}_1\left(\frac{d\mathbb{P}_0^\alpha}{d\mathbb{P}_1} \geq \tau\right) \geq 1 - \alpha$$

wobei \mathbb{P}_0^α die absolut stetige Komponente von \mathbb{P}_0 bezüglich \mathbb{P}_1 darstellt. Dann gilt für einen beliebigen funktionalen Schätzer \tilde{F} :

$$\sup_{\theta \in \Theta} P_\theta(|\tilde{F}_n - F(\theta)| \geq s) \geq \frac{\tau(1 - \alpha - \beta_1) - \beta_0}{1 + \tau}$$

3.5 Fazit

Theorem 1 hat unter den Bedingungen, die wir an die Kerndichteschätzer gestellt hatten, das Ergebnis (12)

$$\sup_{x_0 \in \mathbb{R}} \sup_{p \in P(\beta, L)} \mathbb{E}_p[\psi_{n,2}(\tilde{p}_n(x_0) - p(x_0))^2] \leq C(\beta, L, \alpha)$$

geliefert. Das heißt, das maximale Risiko der Kerndichteschätzer bezüglich der L_2 -Norm an einem Punkt übersteigt den Wert der Konstanten C nicht. Unter der Gültigkeit von Annahme 3 wurde in Theorem 9 für $\beta > 0$ und $L > 0$ gezeigt, dass wir für

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f[\psi_{n,p} \|T_n(x_0) - f(x_0)\|_p^2] \geq c(\beta, L, p_*)$$

abhängig von der Wahl der Norm $\|\cdot\|_p$, $p \in \{2, \infty\}$ zwei unterschiedliche Konvergenzraten erhalten:

$$\psi_{n,2} = n^{\frac{\beta}{2\beta+1}}, \psi_{n,\infty} = \left(\frac{\log(n)}{n}\right)^{\frac{\beta}{2\beta+1}}$$

Daraus folgt insbesondere, dass mit den entsprechenden Anforderungen an die zu schätzende Dichtefunktion (Hölderstetigkeit), gezeigt wurde, dass die Rate $\psi_n = \frac{1}{2\beta+1}$ unter allen Schätzern bereits optimal ist.

4 Verwendete Sätze und Definitionen

Lemma 6. [Tsy09] (Varshamov-Gilbert Schranke)

Sei $m \geq 8$. Dann existiert eine Teilmenge $\{\omega^{(0)}, \dots, \omega^{(M)}\}$ von Ω , so dass $\omega^{(0)} = (0, \dots, 0)$,

$$\rho(\omega^{(j)}, \omega^{(k)}) \geq \frac{m}{8}, \quad \forall 0 \leq j < k \leq M,$$

und

$$M \geq 2^{m/8}$$

Definition 15. [Tsy09] (Kullback Abstand)

Der Kullback Abstand zwischen zwei Maßen P, Q ist folgendermaßen definiert:

$$K(P, Q) = \begin{cases} \int \log\left(\frac{dP}{dQ}\right) dP, & \text{wenn } P \ll Q. \\ +\infty, & \text{sonst.} \end{cases}$$

Man kann zeigen, dass der Kullback Abstand immer kleiner ist, als Gesamtabweichung $V(P, Q)$ ist.

Satz 2. [Tsy09] (Generalisierte Minkowski Ungleichung)
Für eine beliebige Borel-messbare Funktion auf $\mathbb{R} \times \mathbb{R}$, gilt

$$\int \left(\int g(u, x) du \right)^2 dx \leq \left[\int \left(\int g^2(u, x) dx \right)^{\frac{1}{2}} du \right]^2$$

Definition 16. [For11] (Taylorentwicklung)

Sei $I \subset \mathbb{R}$ ein offenes Intervall, $p : I \rightarrow \mathbb{R}$ eine glatte Funktion und $a \in I$. Dann heißt die unendliche Reihe

$$T_p(x; a) = \sum_{n=0}^{\infty} \frac{p^{(n)}(a)}{n!} (x - a)^n$$

die Taylorreihe von p mit Entwicklungstelle a .

Definition 17. [Tsy09] (Verlustfunktion) Sei

$$\omega : [0, \infty) \rightarrow [0, \infty)$$

monoton steigend und $\omega(0) = 0$, ω nicht konstant null, dann heißt ω Verlustfunktion.

Definition 18. [Pfa14] (Bedingter Erwartungswert)

Zu einem gegebenen Wahrscheinlichkeitsraum $(\mathcal{X}, \mathcal{A}, P)$ und einer Teilmenge der σ -Algebra $\mathcal{B} \in \mathcal{A}$ ist der bedingte Erwartungswert $\mathbb{E}[X|\mathcal{B}] = \mathcal{Z}$ einer Zufallsvariable X wieder eine Zufallsvariable, die folgende Bedingungen erfüllt:

- (i) \mathcal{Z} ist \mathcal{B} -messbar
- (ii) $\forall B \in \mathcal{B}$ ist $\mathbb{E}[\mathbb{1}_B \mathcal{Z}] = \mathbb{E}[\mathbb{1}_B X]$

Zudem ist die bedingte Wahrscheinlichkeit eines Ereignisses $A \in \mathcal{A}$ gegeben \mathcal{B} definiert, als die Zufallsvariable

$$P(A|\mathcal{B}) = \mathbb{E}[\mathbb{1}_A|\mathcal{B}]. \quad (59)$$

Theorem 11. [Tsy09] Nehmen wir an, f gehört zur Hölderklasse $\Sigma(\beta, L)$ auf $[0, 1]$ und es gilt $\beta, L > 0$. Sei \tilde{f}_n der $LP(l)$ Schätzer von Ordnung $l = \lfloor \beta \rfloor$ mit zugehöriger Bandbreite

$$h = \alpha \left(\frac{\log(n)}{n} \right)^{\frac{1}{2\beta+1}}.$$

Außerdem stellen wir vier folgende Bedingungen:

- (i) die Designpunkte X_1, \dots, X_n sind deterministisch.
- (ii) Die LP annahmen aus Tsybakov gelten
- (iii) Die Zufallsvariablen ξ sind i.i.d. Normalverteilt mit $\mu = 0$ und σ_ξ^2 , mit endlicher Varianz.
- (iv) K ist ein Lipschitzkern $K \in \Sigma(1, L_K)$ auf \mathbb{R} . Es gilt außerdem $0 < L_K < \infty$.

Unter genannten Bedingungen existiert eine Konstante $C < \infty$, welche als obere Schranke für den Limes Superior des maximalen Risikos fungiert. Das heißt es gilt:

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, l)} E_f \left[\psi_n^{-2} \|\tilde{f}_n - f\|_\infty^2 \right] \leq C.$$

wobei für die Konvergenzrate gilt:

$$\psi_n = \left(\frac{\log(n)}{n} \right)$$

Satz 3. [Rĭ6](Markov Ungleichung)

Es sei $(\mathcal{X}, \mathcal{A}, P_\theta)$ ein statistisches Modell. $X : \mathcal{X} \rightarrow \mathbb{R}$ eine reellwertige Zufallsvariable und a eine reelle Konstante, desweiteren sei eine monoton wachsende Funktion gegeben $\omega : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, dann gilt:

$$\mathbb{E}_\theta[\omega(X)] \geq \omega(a)P[X \geq a]$$

Beweis.

$$\omega(a)P_\theta[X \geq a] := \omega(a) \int \mathbb{1}_{\{X \geq a\}} dP_\theta \stackrel{\omega \text{ monoton}}{\leq} \int \mathbb{1}_{\{X \geq a\}} \omega(X) dP_\theta = \mathbb{E}_\theta[\omega(X)]$$

□

Definition 19. (Pseudometrik)

Wir nennen eine beliebige Funktion d Pseudometrik auf Θ , falls gilt

$$d : \Theta \times \Theta \rightarrow [0, \infty)$$

- (i) $d(\theta, \theta) = 0$
- (ii) $d(\theta, \theta') = d(\theta', \theta)$
- (iii) $d(\theta, \theta'') \leq d(\theta, \theta') + d(\theta', \theta'')$

der Definitheitsbegriff wird hier ausgelassen, das heißt, es darf Elemente aus Θ geben, die verschieden sind, aber dennoch Abstand null haben.

Satz 4. [Kle13](Zerlegungssatz von Lebesgue)

Gegeben sei ein Meßraum (X, \mathcal{A}) und ein σ -endliches Maß μ und ein σ -endliches signiertes Maß ν auf diesem Messraum. Dann existiert eine eindeutige Zerlegung

$$\nu = \nu_a + \nu_s,$$

in zwei σ -endliche Maße ν_a, ν_s , sodass:

- (i) $\nu_a \ll \mu$
- (ii) $\nu_s \perp \mu$

(ii) bedeutet, dass die Maße zueinander singulär sind. Für Wahrscheinlichkeitsmaße bedeutet das insbesondere, dass ihr Hellinger Abstand gleich eins ist.

5 Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig verfasst habe. Ich versichere, dass ich keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe, und dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.

Literatur

- [ELL98] George Casella E. L. Lehmann. *Theory of Point Estimation*. Springer-Verlag, 1998.
- [For11] Otto Forster. *Analysis 1*. Vieweg+Teubner Verlag, 2011.
- [Kle13] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer Berlin Heidelberg, 2013.
- [LD85] Laszlo Györfi Luc Devroye. *Nonparametric Density Estimation*. John Wiley & Sons, 1985.
- [Pfa14] Peter Pfaffelhuber. *Wahrscheinlichkeitstheorie, Stochastische Prozesse, Stochastische Integration und Finanzmathematik*, 2014.
- [Ri16] Ludger Rüschendorf. *Wahrscheinlichkeitstheorie*. Springer Berlin Heidelberg, 2016.
- [Rei12] Markus Reiß. *Nichtparametrische Statistik*, 2012.
- [Tsy09] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer New York, 2009.