Skript zur Vorlesung
Maß- und Wahrscheinlichkeitstheorie

Teil II: Wahrscheinlichkeitstheorie

Christina Schneider

Wintersemester 2009/2010

Institut für Statistik
LMU München

Inhaltsverzeichnis

11. Gesetze der Großen Zahlen, Zentraler Grenzwertsatz

Aus dem Bachelorstudium sind verschiedene Gesetze der Großen Zahlen und der Zentrale Grenzwertsatz bereits bekannt. Sie seien nachstehend zur Wiederholung aufgeführt. Wesentlich für Grenzwertsätze und die Gesetze der Großen Zahlen sind die verschiedenen, dort verwendeten Konvergenzbegriffe.

11.1. Konvergenzbegriffe

Im Folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ (kurz: (X_n)) eine Folge reeller Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ sowie X eine weitere Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Mit $P_{X_n} = X_n(P)$ bzw. $P_X = X(P)$ werden die Verteilungen der Zufallsvariablen bezeichnet.

11.1.1. Fast sichere Konvergenz oder starke Konvergenz

Die Folge (X_n) *konvergiert fast sicher* (oder *konvergiert stark*) gegen $X :\Longleftrightarrow$

$$\forall \varepsilon > 0: \quad \lim_{n\to\infty} P\left(\sup_{m\geq n} |X_m - X| > \varepsilon\right) = 0$$

Das bedeutet, die Menge $\{\omega \in \Omega : \sup_{m\geq n} |X_m(\omega) - X(\omega)| > \varepsilon\}$ konvergiert gegen eine Nullmenge.

Eine äquivalente Bedingung ist

$$\forall \varepsilon > 0: \quad P\left(\limsup_{n\to\infty} |X_n - X| > \varepsilon\right) = 0\,,$$

das heißt, $P\,(\omega \in \Omega : \text{größter Häufungspunkt von } |X_n - X| > \varepsilon) = 0$.

Eine andere Schreibweise für fast sichere Konvergenz ist

$$P(\{\omega \in \Omega : X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\}) = 1$$

oder kurz

$$X_n \xrightarrow{n\to\infty} X \quad P - f.s.$$

Fast sichere Konvergenz entspricht der punktweisen Konvergenz von deterministischen Funktionen und kann nur auf Nullmengen schiefgehen.

11.1.2. $\mathcal{L}^p$-Konvergenz oder Konvergenz im p-ten Mittel

Die Folge (X_n) ist $\mathcal{L}^p$-konvergiert gegen X oder *konvergiert im p-ten Mittel* gegen X $(1 \leq p < \infty)$ $:\Longleftrightarrow$

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = \lim_{n\to\infty} \int_\Omega |X_n(\omega) - X(\omega)|^p \, \mu(\mathrm{d}\omega) = 0$$

Für $p = 1$ spricht man von “Konvergenz im Mittel„.

11.1.3. Stochastische Konvergenz oder Konvergenz nach Wahrscheinlichkeit

Die Folge (X_n) *konvergiert stochastisch* oder *konvergiert nach Wahrscheinlichkeit* gegen X $:\Longleftrightarrow$

$$\forall \varepsilon > 0 : \quad \lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

Eine äquivalente Bedingung ist

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \,.$$

Eine Kurzschreibweise für stochastische Konvergenz ist

$$p \lim_{n\to\infty} X_n = X$$

Stochastische Konvergenz gegen die Null entspricht der stochastischen Konvergenz gegen die P-fast sicher konstante Zufallsvariable mit dem Wert 0, d.h.

$P(|X_n| \geq \varepsilon) \xrightarrow{n\to\infty} 0$, $\forall \varepsilon > 0$, oder kurz $p\lim_{n\to\infty} X_n = 0$.

11.1.4. Schwache Konvergenz oder Konvergenz in Verteilung

Die Folge (P_{X_n}) der Verteilungen der Folge (X_n) *konvergiert schwach* gegen die Verteilung P_X von X $:\Longleftrightarrow$ Für alle stetigen und beschränkten Funktionen f gilt

$$\lim_{n\to\infty} \mathbb{E}(f \circ X_n) = \mathbb{E}(f \circ X)$$

Wegen

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}(f \circ X_n) = \mathbb{E}(f \circ X) \quad &\Longleftrightarrow \quad \int f \circ X_n \,\mathrm{d}P \xrightarrow{n\to\infty} \int f \circ X \,\mathrm{d}P \\ &\Longleftrightarrow \quad \int f \,\mathrm{d}X_n(P) \xrightarrow{n\to\infty} \int f \,\mathrm{d}X(P)\end{aligned}$$

kann durch Verallgemeinerung der letzte Darstellung auch (ohne abbildende Zufallsvariaben) die stochastische (oder schwache) Konvergenz von Wahrscheinlichkeitsmaßen μ_n auf $(\Omega, \mathcal{A})$ gegen ein Wahrscheinlichkeitsmaß μ auf $(\Omega, \mathcal{A})$ definiert werden.

11.1.5. Satz (Konvergenz-Implikationen)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reeller Zufallsvariablen und X eine reelle Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Konvergiert (X_n) stochastisch gegen X, dann konvergiert die Folge der Verteilungen (P_{X_n}) schwach gegen P_X.

Ist X P-fast sicher konstant, so gilt hiervon auch die Umkehrung.

Beweis: **(author?)** (Bauer 2001, S. 36) □

Insgesamt gelten folgende Implikationen:

$$\boxed{\begin{array}{l}\mathcal{L}^p\text{-Konvergenz} \Rightarrow \mathcal{L}^q\text{-Konvergenz } (q \leq p) \\ \Rightarrow \text{stochastische Konvergenz} \Rightarrow \text{schwache Konvergenz}\end{array}}$$

sowie

$$\boxed{\text{fast sichere Konvergenz} \Rightarrow \text{stochastische Kovergenz}}$$

Beweis: Teilweise in der Übung. □

11.2. Gesetze der Großen Zahlen

$(X_n)_{n\in\mathbb{N}}$ bezeichne wieder eine Folge reeller Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Ferner sei die Summe der ersten n Zufallsvariablen definiert durch

$$S_n := \sum_{i=1}^{n} X_i$$

Somit lässt sich die Folge $(S_n)_{n\in\mathbb{N}}$ der Summen der jeweils ersten n Zufallsvariablen betrachten.

11.2.1. Definition (Schwaches Gesetz der Großen Zahlen)

Eine Folge (X_n) genügt dem *Schwachen Gesetz der Großen Zahlen*, falls die Folge

$$\frac{1}{n} S_n$$

stochastisch gegen 0 konvergiert, in Zeichen:

$$p \lim_{n\to\infty} \frac{1}{n} S_n = 0$$

11.2.2. Definition (Starkes Gesetz der Großen Zahlen)

Eine Folge (X_n) genügt dem *Starken Gesetz der Großen Zahlen*, falls die Folge

$$\frac{1}{n} S_n$$

fast sicher gegen 0 konvergiert, in Zeichen:

$$\lim_{n\to\infty} \frac{1}{n} S_n = 0 \quad P - f.s.$$

Hinsichtlich der schwachen Konvergenz von Folgen gilt der folgende Satz:

11.2.3. Satz

Sei (X_n) eine Folge paarweise unkorrelierter Zufallsvariablen mit $\mathbb{E}(X_n) = 0$ und $\mathrm{Var}(X_n) < \infty$ für alle $n \in \mathbb{N}$. Ist zudem (a_n) eine Folge positiver reeller

Zahlen, für welche gilt:

$$\frac{1}{{a_n}^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = 0$$

so folgt

$$p\lim_{n\to\infty}\frac{1}{a_n}\sum_{i=1}^{n} X_i = 0\,,$$

das heißt, die Folge (X_n) konvergiert stochastisch gegen 0.

Beweis: $\rightsquigarrow$ Übung. $\square$

Hinweis:

Die Bedingung ist mit $a_n = n$ sicher erfüllt, wenn die Zufallsvariablen identisch verteilt sind.

Beweis: $\rightsquigarrow$ Übung. $\square$

Hinweis:

Die Bedingung ist hinreichend, aber nicht notwendig. Beispiel: (X_n) sei eine Folge von unabhängigen und identisch Cauchy-verteilten Zufallsvariablen mit Skalenparameter 1. Dann konvergiert die Folge $S_n = \frac{1}{a_n}\sum_{i=1}^{n} X_i$ stochastisch gegen 0, falls $\lim_{n\to\infty}\frac{a_n}{n} = 0$ gilt.

Die Bedingung kann aber gar nicht erfüllt sein, da für Cauchy-Verteilungen keine Erwartungswerte existieren.

Beweis: $\rightsquigarrow$ Übung. $\square$

11.3. Die Kolmogoroffschen Sätze

Diese Sätze geben Antwort auf die Frage unter welchen Bedingungen eine Folge von Zufallsvariablen dem Starken Gesetz der Großen Zahlen genügen.

11.3.1. Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und integrierbarer reeller Zufallsvariablen mit

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Var}(X_n) < \infty \,,$$

so genügt die Folge dem Starken Gesetz der Großen Zahlen.

11.3.2. Satz

Jede unabhängige Folge $(X_n)_{n\in\mathbb{N}}$ quadratintegrierbarer reeller und identisch verteilter Zufallsvariablen genügt dem Starken Gesetz der Großen Zahlen.

11.4. Der Zentrale Grenzwertsatz

Sei $(X_n)_{n\in\mathbb{N}}$ eine unabhängige Folge reeller quadratintegrierbarer Zufallsvariablen mit $\operatorname{Var}(X_n) > 0$. Die Zufallsvariable

$$S_n := \frac{\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))}{\sqrt{\operatorname{Var}(X_1 + \cdots + X_n)}}$$

heißt *standardisierte Summe*.

11.4.1. Definition (Zentraler Grenzwertsatz)

Für die Folge $(X_n)_{n\in\mathbb{N}}$ gilt der *Zentrale Grenzwertsatz*, falls die Folge der Verteilungen $P_{S_n} = S_n(P)$ der standardisierten Summen schwach gegen die $N(0,1)$-Verteilung konvergiert (Konvergenz in Verteilung).

Die Gültigkeit des Zentralen Grenzwertsatzes ist an Bedingungen geknüpft. Wichtige Bedingungen sind die Lindeberg-Bedingung und die Fellersche Bedingung.

11.4.2. Definition (Lindeberg Bedingung, Fellersche Bedingung)

Sei abkürzend

$$
\begin{aligned}
\sigma_n &:= \sigma(X_n) = (\mathrm{Var}(X_n))^{\frac{1}{2}} \\
s_n &:= \sigma(X_1 + \cdots + X_n) = (\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n))^{\frac{1}{2}} \\
\eta_n &:= \mathbb{E}(X_n)
\end{aligned}
$$

Die Folge (X_n) genügt der *Lindeberg-Bedingung* $:\Longleftrightarrow$

$$\forall \varepsilon > 0: \quad L_n(\varepsilon) := \frac{1}{{s_n}^2} \sum_{i=1}^{n} \int_{|x-\eta_i| \geq \varepsilon s_n} (x-\eta)^2 \, [X_i(P)](\mathrm{d}x) \quad \xrightarrow{n\to\infty} \quad 0$$

Die Folge (X_n) genügt der *Fellerschen Bedingung* $:\Longleftrightarrow$

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{\sigma_i}{s_n} = 0$$

11.4.3. Definition (Asymptotische Vernachlässigbarkeit)

Eine Familie $(X_{ni})_{n=1,2,\ldots,\, i=1,\ldots,k_n}$ heißt *asymptotisch vernachlässigbar* $:\Longleftrightarrow$

$$\forall \varepsilon > 0: \quad \lim_{n\to\infty} \max_{1\leq i\leq k_n} P(|X_{ni}| \geq 0) = 0$$

11.4.4. Satz von Lindeberg-Feller

Für jede Folge unabhängiger quadratisch integrierbarer Zufallsvariablen (X_n) sind die folgenden Aussagen gleichwertig:

1. Die Folge genügt der Lindeberg-Bedingung.
2. Es gilt der Zentrale Grenzwertsatz und die Folge genügt der Fellerschen Bedingung.
3. Es gilt der Zentrale Grenzwertsatz und die Folge (X_{ni}) ist asymptotisch vernachlässigbar.

Hinweis:

Der Satz von de Moivre-Laplace ist ein Spezialfall hiervon.

12. Martingale

- Anwendung und Erweiterung des Arbeitens mit bedingten Erwartungen
- Möglichkeit, Grenzwertsätze einheitlicher und allgemeiner zu beweisen und zu formulieren (insbesondere können Unabhängigkeitsbedingungen wegfallen)

Hauptanwendungsgebiet ist die Theorie stochastischer Prozesse. Woher die Bezeichnung "Martingal„kommt, ist nicht ganz klar: Ist es Zaumzeug, eine Spielstrategie aus Martigue (Südfrankreich) oder ein Teil der Tackelage bei Segelschiffen? Das folgende einführende Beispiel veranschaulicht die passendste Herkunft.

Einführendes Beispiel: Spielstrategie

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Münzwürfen mit

$$X_n = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p, \\ -1 & \text{mit Wahrscheinlichkeit } 1-p. \end{cases}$$

Im n-ten Durchgang ist der Einsatz $e_n > 0$.

> Eine Spielstrategie ist eine Vorschrift zur Festlegung des Einsatzes e_n im n-ten Spiel.

Die Festlegung erfolgt in Abhängigkeit der vorherigen Ergebnisse und wird durch eine Abbildung

$$E_{n-1} : \{-1; 1\}^{n-1} \to \mathbb{R}_+,$$

definiert. Die *Zufallsvariable* $E_{n-1} \circ (X_1, \ldots, X_{n-1})$ spiegelt den Einsatz im n-ten Spiel wider. Als Realisierung erhält man $e_n = E_{n-1} \circ (X_1, \ldots, X_{n-1})(\omega)$.

Sei nun $a > 0$ das Startkapital, e_1 der Einsatz beim 1. Spiel und S_n das Guthaben nach dem n-ten Wurf. Damit gilt:

$$\begin{aligned} S_1 &= a + X_1 \cdot e_1 \\ &\vdots \\ S_n &= S_{n-1} + X_n \cdot E_{n-1} \circ (X_1, \ldots, X_{n-1}) \end{aligned} \tag{12.1}$$

Mit der σ-Algebra $\mathcal{A}_n := \sigma(X_1, \ldots, X_n)$ ist S_n $\mathcal{A}_n$-messbar und $\mathbb{E}(S_{n+1} \,|\, \mathcal{A}_n)$ wohldefiniert:

$$\begin{aligned} \mathbb{E}(S_{n+1} \,|\, \mathcal{A}_n) &\overset{(12.1)}{=} \mathbb{E}(S_n + X_{n+1} \cdot E_n \circ (X_1, \ldots, X_n) \,|\, \mathcal{A}_n) \\ &= \underbrace{\mathbb{E}(S_n \,|\, \mathcal{A}_n)}_{\substack{\text{keine Einschränkung} \\ \text{der Information}}} + \mathbb{E}(X_{n+1} \cdot \underbrace{E_n \circ (X_1, \ldots, X_n)}_{\mathcal{A}_n\text{-messbare ZV}} \,|\, \mathcal{A}_n) \\ &= S_n + \mathbb{E}(X_{n+1} \,|\, \mathcal{A}_n) \cdot E_n \circ (X_1, \ldots, X_n) \\ &\underset{\text{unabh.}}{\overset{X_i}{=}} S_n + \underbrace{\mathbb{E}(X_{n+1})}_{=2p-1} \cdot E_n \circ (X_1, \ldots, X_n) \end{aligned}$$

Damit gilt:

$$\mathbb{E}(S_{n+1} \,|\, \mathcal{A}_n) \begin{cases} \leq S_n & \text{falls } p < \frac{1}{2} \quad \rightsquigarrow \quad \text{Super-Martingal} \\ = S_n & \text{falls } p = \frac{1}{2} \quad \rightsquigarrow \quad \text{Martingal} \\ \geq S_n & \text{falls } p > \frac{1}{2} \quad \rightsquigarrow \quad \text{Sub-Martingal} \end{cases}$$

Hier waren die Zufallsvariablen X_n zwar unabhängig. Im Allgemeinen definiert der Martingalbegriff Abhängigkeiten zwischen Zufallsvariablen über die sukzessiven bedingten Erwartungen.

12.0. Vorbereitende Begriffe

geordnete Menge Sei T eine Menge auf welcher eine zweistellige Relation "$\leq$„definiert ist. T heißt *geordnet*, wenn für $s, t, u \in T$ gilt:

$$t \leq t \qquad \text{(Reflexivität)}$$

$$s \leq t \text{ und } t \leq s \;\Rightarrow\; t = s \qquad \text{(Antisymmetrie)}$$

$$s \leq t \text{ und } t \leq u \;\Rightarrow\; s \leq u \qquad \text{(Transitivität)}$$

total geordnete Menge Die Menge T (in der Anwendung: Zeit) heißt *total geordnet*, wenn für jedes Paar $(s,t) \in T \times T$ $s \leq t$ oder $t \leq s$ gilt. Dann ist auch eine zeitliche Interpretation möglich.

isotone/wachsende Familie von Sub-σ-Algebren Sei T eine geordnete Menge, $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(\mathcal{A}_t)_{t\in T}$ eine Familie von Sub-σ-Algebren. Wenn gilt

$$s \leq t \;\Rightarrow\; \mathcal{A}_s \subset \mathcal{A}_t \quad (s,t \in T)$$

dann heißt die Familie $(\mathcal{A}_t)_{t\in T}$ *isoton* (oder *wachsend*).

adaptiert Sei ferner $(X_t)_{t\in T}$ mit $X_t : (\Omega, \mathcal{A}_t) \to (\Omega', \mathcal{A}')$ eine Familie $\mathcal{A}_t/\mathcal{A}'$-messbarer Zufallsvariablen. Dann heißt die Familie von Zufallsvariablen der Familie $(\mathcal{A}_t)_{t\in T}$ *adaptiert*.

natürliche Filtration $(X_t)_{t\in T}$ ist immer der Familie $\sigma(X_s : s \leq t)$ adaptiert. $\sigma(X_s : s \leq t)$ ist die von allen Zufallsvariablen X_s, $s \leq t$, erzeugte Sub-σ-Algebra von $\mathcal{A}$, d.h. – mit der Interpretation von T als Zeit – die von der Vergangenheit erzeugte σ-Algebra (sie enthält alle Urbilder der $\mathcal{A}'$-messbaren Mengen bzgl. aller X_s, $s \leq t$). Häufig wird sie deshalb auch intrinsische, kanonische oder natürliche Filtration genannt.

gleichgradig integrierbar Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\mathcal{F}$ eine Familie $\mathcal{A}$-messbarer reeller Funktionen auf Ω. $\mathcal{F}$ heißt *p-fach* $(1 \leq p < \infty)$ *gleichgradig integrierbar* (bzgl. μ), wenn gilt:

$$\forall \varepsilon > 0 \quad \exists\, g \in \mathcal{L}_p(\Omega, \mathcal{A}, \mu), g \geq 0 \quad \forall f \in \mathcal{F} : \int\limits_{\{\omega : |f(\omega)| \geq g(\omega)\}} |f|^p \,\mathrm{d}\mu \leq \varepsilon$$

Für $p = 1$ spricht man von *gleichgradig integrierbar.*

Abbildung 12.1 illustriert die "Einschränkung„der Funktionenfamilie $\mathcal{F}$ durch die Funktion g.

Zur Erinnerung: In Definition 6.3 wurde die Integrierbarkeit von Funktionen definiert. Mit der Definition

$$\mathcal{L}_p(\Omega, \mathcal{A}, \mu) := \left\{ f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathfrak{B}) \;\middle|\; \int |f|^p \,\mathrm{d}\mu < \infty \right\}$$

bezeichnet $\mathcal{L}(\Omega, \mathcal{A}, \mu) := \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$ die Menge der bzgl. μ integrierbaren

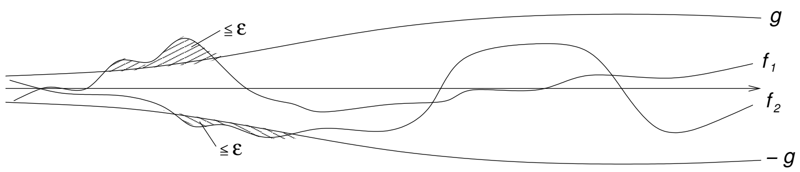

Abbildung 12.1.: Die Funktionen der Familie $\mathcal{F}$ sind durch g “gedeckelt„

Funktionen und $\mathcal{L}_2(\Omega, \mathcal{A}, \mu)$ die Menge der bzgl. μ quadratintegrierbaren Funktionen.

Wiederholung: Von Zufallsvariablen erzeugte (Sub-)σ-Algebren Ist $(X_t)_{t \in T}$ mit $X_t : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathfrak{B})$ ein stochastischer Prozess, so ist $X_t(\omega)$

- für festes ω ein *Pfad* (als Funktion von t)
- für festes t eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$

Abbildung 12.2 zeigt drei Pfade (Realisationen) einer einfachen Irrfahrt.

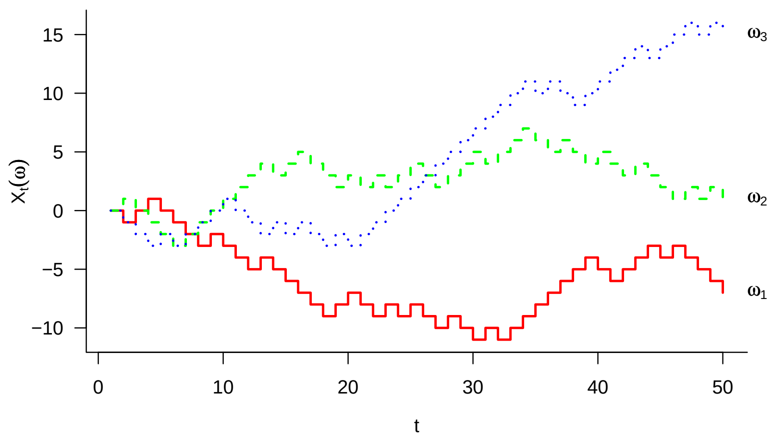

Abbildung 12.2.: Pfade eines stochastischen Prozesses (hier: einfache Irrfahrt)

Die Menge $X_t^{-1}(\mathfrak{B})$ ist die gröbste σ-Algebra für welche die Zufallsvariable X_t messbar ist.

Beispiel:

Sei $\Omega = [0; 1]$ sowie

$$X_1(\omega) = \begin{cases} 1 & 0 \le \omega < \frac{1}{2} \\ 2 & \frac{1}{2} \le \omega \le 1 \end{cases} \quad \text{und} \quad X_2(\omega) = \begin{cases} 0 & 0 \le \omega < \frac{1}{3} \\ 4 & \frac{1}{3} \le \omega \le 1 \end{cases}.$$

Dann ist für $B \in \mathfrak{B}$

$$\begin{aligned} X_1^{-1}(B) &= \begin{cases} \emptyset & 1 \notin B \text{ und } 2 \notin B \\ [0; \frac{1}{2}) & 1 \in B \text{ und } 2 \notin B \\ [\frac{1}{2}; 1] & 1 \notin B \text{ und } 2 \in B \\ [0; 1] & 1 \in B \text{ und } 2 \in B \end{cases} \\ X_2^{-1}(B) &= \begin{cases} \emptyset & 0 \notin B \text{ und } 4 \notin B \\ [0; \frac{1}{3}) & 0 \in B \text{ und } 4 \notin B \\ [\frac{1}{3}; 1] & 0 \notin B \text{ und } 4 \in B \\ [0; 1] & 0 \in B \text{ und } 4 \in B \end{cases} \end{aligned}$$

$X_1^{-1}(\mathfrak{B}) = \{\emptyset, [0; \frac{1}{2}), [\frac{1}{2}; 1], [0; 1]\}$ bildet die σ-Algebra $\sigma(X_1)$.
$X_2^{-1}(\mathfrak{B}) = \{\emptyset, [0; \frac{1}{3}), [\frac{1}{3}; 1], [0; 1]\}$ bildet die σ-Algebra $\sigma(X_2)$.
Die von den Zufallsvariablen X_1, X_2 erzeugte Sub-σ-Algebra von $\mathcal{A}$ ist $\sigma(X_1, X_2) = \{\emptyset, [0; 1], [0; \frac{1}{2}), [\frac{1}{2}; 1], [0; \frac{1}{3}), [\frac{1}{3}; 1], [0; \frac{1}{3}) \cup [\frac{1}{2}; 1], [0; \frac{1}{2}) \cap [\frac{1}{3}; 1]\}$.

12.1. Definition von Martingalen

12.1.1. Definition

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(\mathcal{A}_t)_{t \in T}$ eine isotone Familie von Sub-σ-Algebren, T geordnet und $(X_t)_{t \in T}$ eine Familie von $\mathcal{A}_t/\mathfrak{B}$-messbaren Zufallsvariablen mit $X_t \in \mathcal{L}(\Omega, \mathcal{A}, P)$, $t \in T$. Dann heißt $(X_t)_{t \in T}$

$$\begin{array}{c} Sub-Martingal \\ Martingal \\ Super-Martingal \end{array} \quad \text{bzgl. } (\mathcal{A}_t)_{t \in T} \;:\Longleftrightarrow\; \forall s \le t:\; X_s \begin{array}{c} \le \\ = \\ \ge \end{array} \mathbb{E}(X_t \,|\, \mathcal{A}_s), \;\; P-f.s.$$

12.1.2. Bemerkungen

- Jedes (Sub-, Super-) Martingal ist auch ein (Sub-, Super-) Martingal bzgl. der Familie $(\sigma(X_s : s \le t) \subset \mathcal{A}_t)_{t \in T}$, d.h. bzgl. der natürlichen Filtration

(σ-Algebra der Vergangenheit bis inklusive t). Man sagt dann (Sub-, Super-) Martingal (schlechthin).

- $(X_t)_{t\in T}$ ist genau dann ein Sub-Martingal, wenn $(-X_t)_{t\in T}$ ein Super-Martingal ist.

  Beweis: $X_s \leq \mathbb{E}(X_t \,|\, \mathcal{A}_s) \Longleftrightarrow -X_s \geq -\mathbb{E}(X_t \,|\, \mathcal{A}_s) = \mathbb{E}(-X_t \,|\, \mathcal{A}_s)$ □

- Die (Sub-, Super-) Martingaleigenschaft ist genau dann erfüllt, wenn

$$\int_{F_s} X_s \,\mathrm{d}P \;\begin{array}{c}\leq\\=\\\geq\end{array}\; \int_{F_s} X_t \,\mathrm{d}P \qquad \text{für alle } s \leq t,\ F_s \in \mathcal{A}_s$$

  Mit $F_s = \Omega$ folgt, dass sich die jeweiligen Erwartungswerte im relevanten Sinn monoton verhalten.

12.1.3. Beispiele

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

(a) Eine monoton wachsende Familie $(X_t)_{t\in T}$ von P-integrierbaren Zufallsvariablen (d.h. $X_s(\omega) \leq X_t(\omega)$, $\forall s \leq t$, $P - f.s.$) ist ein Sub-Martingal schlechthin (bzgl. der natürlichen Filtration).

Beweis: Für alle $s \leq t$ gilt

$$X_s = \mathbb{E}(X_s \,|\, \mathcal{A}_s) \leq \mathbb{E}(X_t \,|\, \mathcal{A}_s) \qquad P - f.s.$$

Die Gleichung gilt wegen der $\mathcal{A}_s$-Messbarkeit von X_s und mit Satz 10.7[3]. Die Ungleichung gilt mit Satz 10.7[7]. □

(b) Sei $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$ und $(\mathcal{A}_t)_{t\in T}$ eine monoton wachsende Familie von Sub-σ-Algebren von $\mathcal{A}$. Dann wird durch $X_t := \mathbb{E}(X \,|\, \mathcal{A}_t)$, $t \in T$ ein Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$ definiert.

Beweis: X_t ist nach Definition $\mathcal{A}_t$-messbar und für alle $s \leq t$ gilt

$$\mathbb{E}(X_t \,|\, \mathcal{A}_s) \stackrel{\text{Def}}{=} \mathbb{E}(\mathbb{E}(X \,|\, \mathcal{A}_t) \,|\, \mathcal{A}_s) \stackrel{\text{Satz 10.11}}{=} \mathbb{E}(X \,|\, \mathcal{A}_s) \stackrel{\text{Def}}{=} X_s \qquad P - f.s.,$$

da $\mathcal{A}_s \subset \mathcal{A}_t$ („$\mathcal{A}_s$ ist gröber als $\mathcal{A}_t$"). □

(c) Sei $(X_n)_{n\in\mathbb{N}}$ eine Familie von unabhängigen und identisch verteilten Zufallsvariablen mit $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathfrak{B})$ sowie $\mathbb{E}(X_n) =: \mu < \infty$ und $\operatorname{Var}(X_n) < \infty$. Ist $(X_n)_{n\in\mathbb{N}}$ ein (Sub-, Super-) Martingal (bzgl. der kanonischen Filtration)?

$$X_n \overset{\overset{?}{\leq}}{\underset{\underset{?}{\geq}}{=}} \mathbb{E}(X_{n+k} \,|\, X_1, \ldots, X_n) = \mathbb{E}(X_{n+k}) = \mu \qquad P-f.s.,\ k \geq 1$$

Also ist $(X_n)_{n\in\mathbb{N}}$ nur dann ein Martingal, wenn die X_n f.s. konstant gleich μ sind (degenierte Zufallsvariablen). Eine iid-Folge von (echten) Zufallsvariablen besitzt also nicht die Martingal-Eigenschaft!

Was passiert unter der Einschränkung $a_n \leq X_n \leq b_n$, d.h. mit beschränktem Träger? Definiere

$$\varphi_n(X_n) := \frac{X_n - a_n}{b_n - a_n} \in [0; 1] \quad \forall n \in \mathbb{N}$$

Diese transformierten Zufallsvariablen $\varphi_n(X_n)$ sind weiterhin unabhängig und es gilt auch

$$\mathbb{E}(\varphi_n(X_n)) \in [0; 1] \quad \forall n \in \mathbb{N}$$

$$\Longrightarrow \begin{cases} \varphi_{n-1}(X_{n-1}) + (n-1) < \mathbb{E}\left[\varphi_n(X_n) + n\right] & \text{(Sub-Martingal)} \\ \varphi_{n-1}(X_{n-1}) - (n-1) > \mathbb{E}\left[\varphi_n(X_n) - n\right] & \text{(Super-Martingal)} \end{cases}$$

Fazit: Bei beschränkten Trägern kann man beispielsweise durch geeignete Transformationen (hier: Addition von n bzw. $-n$ zu $\varphi_n(X_n)$) die (Sub-, Super-) Martingaleigenschaft erzwingen.

12.1.4. Satz

Sei $(X_t)_{t\in T}$ ein Sub-Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$ (also $X_s \leq \mathbb{E}(X_t \,|\, \mathcal{A}_s)$ für $s \leq t$), $I \subset \mathbb{R}$ ein Intervall und $X_t(\Omega) \subset I$ für alle $t \in T$. Ferner sei $\varphi : I \to \mathbb{R}$ eine konvexe Funktion[1] mit $\varphi \circ X_t \in \mathcal{L}(\Omega, \mathcal{A}, P)$, für alle $t \in T$.

Dann gilt:

(i) Ist φ monoton wachsend, so ist auch $(\varphi \circ X_t)_{t\in T}$ ein Sub-Martingal bzgl.

[1] Eine Funktion φ heißt *konvex* $:\Leftrightarrow\ \varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y),\ \lambda \in [0; 1]$

$(\mathcal{A}_t)_{t\in T}$.

(ii) Ist $(X_t)_{t\in T}$ sogar ein Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$, so kann in (i) auf die Monotonie von φ verzichtet werden. Dann ist $(\varphi\circ X_t)_{t\in T}$ trotzdem wieder ein Sub-Martingal.

12.1.5. Korollar

(i) $(X_t)_{t\in T}$ sei ein Martingal bzgl. $(A_t)_{t\in T}$ mit $X_t \in \mathcal{L}_p(\Omega,\mathcal{A},P)$ für alle $t\in T$ $(1\le p<\infty)$. Dann ist $(|X_t|^p)_{t\in T}$ ein Sub-Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$.

(ii) Für jedes $c\in\mathbb{R}$ und Sub-Martingal $(X_t)_{t\in T}$ ist auch $(\max\{c,X_t\})_{t\in T}$ ein Sub-Martingal bzgl. $(A_t)_{t\in T}$. Insbesondere ist mit $c=0$ dann auch $(X_t^+)_{t\in T}$ ein Sub-Martingal.

(iii) Ist $(X_t)_{t\in T}$ ein Super-Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$, so ist $(X_t^-)_{t\in T}$ ein Sub-Martingal bzgl. $(\mathcal{A}_t)_{t\in T}$. Zur Erinnerung: $X_t^- := -\min\{0,X_t\}$.

Beweis: $(X_t)_{t\in T}$ Super-Martingal $\Rightarrow$ $(-X_t)_{t\in T}$ Sub-Martingal. Weiter mit (ii). $\square$

12.1.6. Lemma

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von integrierbaren Zufallsvariablen über $(\Omega,\mathcal{A},P)$. Sei $(\mathcal{A}_n)_{n\in\mathbb{N}}$ eine monoton wachsende Folge von Sub-σ-Algebren von $\mathcal{A}$ und X_n sei $\mathcal{A}_n/\mathfrak{B}$-messbar für alle $n\in\mathbb{N}$.

Dann gilt:

$$(X_n)_{n\in\mathbb{N}} \text{ ist ein (Sub-) Martingal bzgl. } (\mathcal{A}_n)_{n\in\mathbb{N}}$$
$$\Longleftrightarrow$$
$$X_n \overset{(\leq)}{=} \mathbb{E}(X_{n+1}\,|\,\mathcal{A}_n) \qquad P-f.s.,\ \forall n\in\mathbb{N}$$

12.2. Stoppzeit – Optional Stopping Theorem

12.2.1. Definition (Stoppzeit)

Sei $(\mathcal{A}_t)_{t\in T}$ eine isotone Familie von Sub-σ-Algebren bzgl. $(\Omega,\mathcal{A})$ und T eine (total) geordnete Menge aus $\mathbb{R}$ (beispielsweise $T=\{0;1;2;\ldots;\infty\}$).

Eine Zufallsvariable $\tau : \Omega \to T$ heißt *Stoppzeit*
$:\iff \quad \forall t \in T: \ \tau^{-1}((0;t]) = \{\omega : \tau(\omega) \le t\} \in \mathcal{A}_t$, d.h. τ ist $\mathcal{A}_t/\mathfrak{B}$-messbar
$\Big[\iff \quad \forall t \in T: \ \{\omega : \tau(\omega) \ge t\} \in \mathcal{A}_t \Big]$

12.2.2. Definition (Spielsystem)

Ein *Spielsystem* ist eine Folge von Zufallsvariablen $(X_t)_{t\in\mathbb{N}}$ auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(\mathbb{R}, \mathfrak{B})$ und

$$X_1 = W_1\Delta_1 \qquad X_{t+1} = X_t + W_{t+1}\Delta_{t+1} \quad (t \ge 1)\,,$$

wobei

Δ_t: unabhängige Zufallsvariable mit $\mathbb{E}(\Delta_t) = 0$. Diese Zufallsvariablen repräsentieren eine unabhängige Folge von fairen Spielen, deren Ausgang vom Spieler nicht beeinflusst werden kann, z.B. $\Delta_t \overset{i.i.d.}{\sim} U(\{-1;1\})$.

X_t: *kumulierter Spielgewinn* nach dem t-ten Spiel.

W_t: *Einsatz*, den der Spieler für das t-te Spiel leistet.

Die Spieleinsätze W_t können jeweils in Abhängigkeit des Spielverlaufs bis zum t-ten Spiel gewählt werden: $W_t = g_t(X_{t-1}, \ldots, X_1)$ ist $\mathcal{A}_{t-1}^X := \sigma(X_{t-1}, \ldots, X_1)$-messbar, wobei g_t eine "deterministische„Funktion ist. Man sagt dann auch, $(W_t, \mathcal{A}_t)_{t\in\mathbb{N}}$ ist eine *vorhersagbare* Folge.

12.2.3. Satz

Unter obigen Voraussetzungen bildet die Folge $(X_t)_{t\in\mathbb{N}}$ der kumulierten Spielgewinne ein Martingal.

Beweis:

$$\begin{aligned}
\mathbb{E}(X_{t+1} \,|\, \mathcal{A}_t^X) &= \mathbb{E}(X_t + W_{t+1}\Delta_{t+1} \,|\, \mathcal{A}_t^X) \\
&= \mathbb{E}(X_t \,|\, \mathcal{A}_t^X) + \mathbb{E}(W_{t+1}\Delta_{t+1} \,|\, \mathcal{A}_t^X) \\
&= X_t + W_{t+1}\underbrace{\mathbb{E}(\Delta_{t+1} \,|\, \mathcal{A}_t^X)}_{=\mathbb{E}(\Delta_{t+1})=0} \\
&= X_t
\end{aligned}$$

Die dritte Zeile folgt, da X_t und W_{t+1} (nach Voraussetzung) beide $\mathcal{A}_t^X$-messbar sind. Anschließend wird noch benutzt, dass die Zufallsvariable Δ_{t+1} unabhängig von der "Vergangenheit„$\mathcal{A}_t^X$ der Spielgewinne ist. □

12.2.4. Ein Spielsystem beim Roulette (Herkunft des Martingalbegriffs)

Spielsystem:

1. Setze auf Rot und beginne mit dem Einsatz 1; verdopple den Einsatz nach jedem Spiel.
2. Verdopple solange, bis zum ersten Mal Rot erscheint.

Das heißt

$$\Delta_t = \begin{cases} +1 & \text{falls Rot erscheint (Gewinn)} \\ -1 & \text{falls Schwarz erscheint (Verlust)} \end{cases}$$

In der ersten Spielphase ist für $t = 1, 2, \dots$

$$\begin{aligned} W_t &= 2^{t-1} && \text{(Einsatz im } t\text{-ten Spiel)} \\ X_t &= \sum_{i=1}^{t} 2^{i-1}\Delta_i && \text{(kumulierter Spielgewinn nach dem } t\text{-ten Spiel)} \end{aligned}$$

Nach 12.2.3 bildet die Folge der kumulierten Spielgewinne ein Martingal wegen $\mathbb{E}(X_t \,|\, \sigma(X_s : s \le t-1)) = X_{t-1}$. Hier mit $\mathbb{E}(X_t) = \sum_{i=1}^{t} 2^{i-1}\mathbb{E}(\Delta_i) = 0$.

Die zweite Spielphase bedeutet die Einführung einer Stoppzeit τ mit

$$\tau(\omega) = \min\{t : \Delta_t(\omega) = 1\}$$

Diese ist Geom$(\frac{1}{2})$ verteilt, d.h. $P(\{\omega : \tau(\omega) = t\}) = \frac{1}{2^t}$, $t = 1, 2, \dots$.

Der Gewinn in Abhängigkeit von der Stoppzeit ist dann

$$X_\tau = \begin{cases} \sum_{i=1}^{\tau} 2^{i-1}\Delta_i & \tau < \infty \\ \text{undefiniert} & \tau = \infty \quad (P(\tau = \infty) = 0) \end{cases}$$

und es gilt für $\omega \in \{\omega : \tau(\omega) < \infty\}$ wegen $\Delta_1 = \cdots = \Delta_{\tau-1} = -1$ und $\Delta_\tau = 1$

$$X_\tau(\omega) = -\sum_{i=1}^{\tau(\omega)-1} 2^{i-1} + 2^{\tau(\omega)-1} \underset{\text{Summe}}{\overset{\text{geom.}}{=}} -\frac{2^0 - 2^{\tau(\omega)-1}}{1-2} + 2^{\tau(\omega)-1} = 1\,,$$

also $P(X_\tau = 1) = 1$.

Man kann also durch diese Stoppzeit das Spiel so einrichten, dass man mit Wahrscheinlichkeit 1 den Betrag 1 gewinnt. Dann würde das Casino aber bankrott gehen. Deshalb begrenzen Casinos die Anzahl der Verdoppelungen.

12.2.5. Optional Stopping Theorem

Für ein Martingal gilt $\mathbb{E}(X_1) = \mathbb{E}(X_t)$ für jedes *feste* t. Kann man diese Gleichung durch Einführung einer Stoppzeit überlisten? Beim Spielsystem 12.2.4 ging das: $\mathbb{E}(X_1) = \mathbb{E}(X_t) = 0$, aber $\mathbb{E}(X_\tau) = 1$. Der folgende Satz sagt aus, in welchen Situationen das Martingal auch für eine (*zufällige*) Stoppzeit nicht überlistet werden kann.

Satz (Optional Stopping Theorem)

Sei $(X_t)_{t\in\mathbb{N}}$ ein Martingal und τ eine Stoppzeit. Es gelte *eine* der folgenden Bedingungen:

(i) τ ist beschränkt, d.h. $\exists\, k < \infty \;\; \forall\, \omega \in \Omega : |\tau(\omega)| \leq k$

(ii) $(X_t)_{t\in\mathbb{N}}$ ist beschränkt, d.h. $\exists\, k < \infty \;\; \forall\, \omega \in \Omega : |X_t(\omega)| \leq k$

(iii) $\mathbb{E}(\tau) < \infty$ und $(X_t - X_{t-1})_{t\in\mathbb{N}}$ ist beschränkt

Dann gilt

$$\mathbb{E}(X_\tau) = \mathbb{E}(X_1)\,.$$

Bemerkung

Beim Spielsystem 12.2.4 sind alle drei Bedingungen verletzt: τ, $(X_t)_{t\in\mathbb{N}}$ und $(X_t - X_{t-1})_{t\in\mathbb{N}} \equiv (2^{t-1})_{t\in\mathbb{N}}$ sind nicht beschränkt.

12.3. Martingaldifferenzenfolgen

12.3.1. Beispiel

Sei $\eta_n \in \mathcal{L}(\Omega, \mathcal{A}, P)$, $n \in \mathbb{N}$, mit $\mathcal{A}_n := \sigma(\eta_1, \ldots, \eta_n)$ und $a \in \mathbb{R}$ beliebig. Definiere

$$X_1 := \eta_1 - a \quad \text{und} \quad X_{n+1} := X_n + \eta_{n+1} - \mathbb{E}(\eta_{n+1} \,|\, \mathcal{A}_n) \quad (n \geq 1)\,.$$

Dann gilt $P - f.s.$

$$\begin{aligned} \mathbb{E}(X_{n+1} \,|\, \mathcal{A}_n) &= \mathbb{E}(X_n \,|\, \mathcal{A}_n) + \mathbb{E}(\eta_{n+1} \,|\, \mathcal{A}_n) - \mathbb{E}(\mathbb{E}(\eta_{n+1} \,|\, \mathcal{A}_n) \,|\, \mathcal{A}_n) \\ &= X_n + \mathbb{E}(\eta_{n+1} \,|\, \mathcal{A}_n) - \mathbb{E}(\eta_{n+1} \,|\, \mathcal{A}_n) \\ &= X_n \end{aligned}$$

Das heißt, die Folge $(X_n)_{n\in\mathbb{N}}$ bildet ein Martingal (bzgl. $\mathcal{A}_n$).

Ist umgekehrt $(X_n)_{n\in\mathbb{N}}$ als Martingal (bzgl. $\sigma(X_1, \ldots, X_n)$) vorausgesetzt und definiert man

$$\eta_1 := X_1 \qquad \eta_n := X_n - X_{n-1} \quad (n \geq 2)$$

dann gilt

$$\begin{aligned} \mathbb{E}(\eta_{n+1} \,|\, \eta_1, \ldots, \eta_n) &= \mathbb{E}(X_{n+1} - X_n \,|\, \eta_1, \ldots, \eta_n) \\ &= \mathbb{E}(X_{n+1} - X_n \,|\, X_1, \ldots, X_n) \\ &= \underbrace{\mathbb{E}(X_{n+1} \,|\, X_1, \ldots, X_n)}_{=X_n,\ \text{da } (X_n) \text{ Martingal}} - X_n \\ &= 0\,. \end{aligned}$$

Das heißt, die Folge der Martingaldifferenzen ist um null zentriert. Deswegen ist folgende Definition möglich:

12.3.2. Definition (Martingaldifferenzenfolge)

Eine Folge reeller integrierbarer Zufallsvariablen $(\eta_n)_{n\in\mathbb{N}}$ heißt *Martingaldifferenzenfolge* (kurz: MDF), falls

$$\mathbb{E}(\eta_{n+1} \,|\, \eta_1, \ldots, \eta_n) = 0 \qquad P - f.s., \text{ für alle } n \in \mathbb{N}$$

Ergänzend wird gefordert, dass $\mathbb{E}(\eta_1 \,|\, \mathcal{A}_0) = 0$, mit $\mathcal{A}_0 := \{\emptyset, \Omega\}$.

12.3.3. Satz

Es gelten die folgenden Zusammenhänge zwischen MDF und Unabhängigkeit bzw. Unkorreliertheit:

(i) Wenn $(\eta_n)_{n\in\mathbb{N}}$ eine MDF ist mit $\eta_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$, $n \in \mathbb{N}$, dann sind die η_n paarweise unkorreliert.

(ii) Wenn $(\eta_n)_{n\in\mathbb{N}}$ eine MDF ist und für alle $I \in e(\mathbb{N})$ (= Menge aller endlichen Teilmengen von $\mathbb{N}$) $(\eta_i)_{i\in I}$ (multivariat) normalverteilt ist, dann sind die η_n paarweise unabhängig.

(iii) Wenn $(\eta_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger, integrierbarer und zentrierter Zufallsvariablen ist, dann ist die Folge eine MDF.

Bemerkung

Für einelementiges I ist die Implikation (ii) mit Hilfe von Aussage (i) klar, denn: η_n eindimensional normalverteilt $\Rightarrow$ Varianz existiert $\Leftrightarrow$ $\eta_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P) \overset{(i)}{\Rightarrow}$ $(\eta_n)_{n\in\mathbb{N}}$ paarweise unkorreliert $\overset{\eta_n \sim \mathrm{N}(\cdot,\cdot)}{\Longleftrightarrow}$ $(\eta_n)_{n\in\mathbb{N}}$ paarweise unabhängig.

Teil (iii) des Satzes ist trivial, da hier $\mathbb{E}(\eta_{n+1} \,|\, \eta_1, \ldots, \eta_n) = \mathbb{E}(\eta_{n+1}) = 0$.

12.3.4. Korollar

Sei $(\eta_n)_{n\in\mathbb{N}}$ eine MDF mit $\eta_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$, $n \in \mathbb{N}$. Dann gilt für das zugehörige Martingal der Partialsummen $X_n = \sum_{i=1}^n \eta_i$ (vgl. 12.3.1):

(i) $\left(X_n^2\right)_{n\in\mathbb{N}}$ ist ein Sub-Martingal

(ii) $\mathbb{E}\left(X_n^2\right) = \sum\limits_{i=1}^{n} \mathbb{E}\left(\eta_i^2\right)$

Beweis: (i) folgt mit 12.1.5 (i).

(ii):

$$\begin{aligned}
\mathbb{E}\left(X_n^2\right) &= \mathbb{E}\left(\left(\sum_{i=1}^n \eta_i\right)^2\right) = \mathbb{E}\left(\sum_{i=1}^n\sum_{j=1}^n \eta_i\eta_j\right) \\
&= \sum_{i=1}^n\sum_{j=1}^n \mathbb{E}(\eta_i\eta_j) = \sum_{i=1}^n \mathbb{E}(\eta_i\eta_i) + \sum_{i=1}^n\sum_{j\neq i} \mathbb{E}(\eta_i\eta_j) \\
&= \sum_{i=1}^n \mathbb{E}\left(\eta_i^2\right) ,
\end{aligned}$$

wobei verwendet wurde, dass $\mathbb{E}(\eta_i\eta_j) = 0$ $(i \neq j)$, da die η_i nach Satz 12.3.3 (i) unkorreliert sind und (wegen MDF) den Erwartungswert 0 besitzen. □

12.4. Grenzwertsätze für Martingale bzw. Martingaldifferenzenfolgen

12.4.1. Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine quadratintegrierbare MDF[2] auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $(a_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen mit $0 < a_n \nearrow \infty$, sodass

$$\sum_{n=1}^{\infty} a_n^{-2}\,\mathbb{E}(X_n^2) < \infty\,.$$

Dann gilt

$$\lim_{n\to\infty}\left(\frac{1}{a_n}\sum_{i=1}^n X_i\right) = 0 \qquad P - f.s.$$

Beispiel:

Mit $a_n := n$ gilt $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} 0$. Allgemeiner gilt folgender Grenzwertsatz.

[2] Dies impliziert die Existenz der Varianzen (und Erwartungswerte) der X_n, sowie $\mathbb{E}(X_n) = 0$.

12.4.2. Satz von Chow

Sei $(S_n)_{n\in\mathbb{N}}$ ein nicht-negatives Sub-Martingal mit $\mathbb{E}(S_n^\alpha) < \infty$, für alle $n \in \mathbb{N}$ und ein $\alpha \geq 1$. Sei ferner $0 < a_n \nearrow \infty$ eine Folge reeller Zahlen mit

$$\sum_{n=2}^{\infty} a_n^{-\alpha}\, \mathbb{E}(S_n^\alpha - S_{n-1}^\alpha) < \infty \,.$$

Dann gilt

$$\lim_{n\to\infty} \left(\frac{1}{a_n} S_n\right) = 0 \qquad P - f.s.$$

Hinweis

Sei nun $(X_n)_{n\in\mathbb{N}}$ eine beliebige quadratintegrierbare Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Dann ist mit

$$\tilde{X}_1 := X_1 - \mathbb{E}(X_1) \quad \text{und} \quad \tilde{X}_{i+1} := X_{i+1} - \mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1)$$

eine MDF gegeben und $\tilde{X}_i$ gehört zu $\mathcal{L}_2(\Omega, \mathcal{A}, P)$. Außerdem gilt:

$$\begin{aligned}
\mathbb{E}(\tilde{X}_{i+1}^2) &= \mathbb{E}\left(X_{i+1}^2\right) - 2\,\mathbb{E}(X_{i+1} \cdot \mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1)) \\
&\quad + \mathbb{E}\left((\mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1))^2\right) \\
&= \mathbb{E}\left(X_{i+1}^2\right) - \mathbb{E}\left((\mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1))^2\right) \\
&\leq \mathbb{E}\left(X_{i+1}^2\right)
\end{aligned}$$

Die zweite Gleichheit folgt mit Satz 10.10 (Glättungssatz) oder 10.11 (iteriertes Bedingen), denn:

$$\begin{aligned}
\mathbb{E}(X_{i+1} \cdot \mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1)) &= \mathbb{E}\left(\mathbb{E}(X_{i+1} \cdot \mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1)) \,|\, X_i, \ldots, X_1\right) \\
&= \mathbb{E}\left(\mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1) \cdot \mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1)\right) \\
&= \mathbb{E}\left((\mathbb{E}(X_{i+1} \,|\, X_i, \ldots, X_1))^2\right)
\end{aligned}$$

Damit folgt als Erweiterung von 12.4.1 auf nicht-zentrierte Zufallsvariablen sofort der nächste Grenzwertsatz.

12.4.3. Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen aus $\mathcal{L}_2(\Omega, \mathcal{A}, P)$ und $0 < a_n \nearrow \infty$ eine Folge reeller Zahlen mit $\sum_{n=1}^{\infty} a_n^{-2}\, \mathbb{E}(X_n^2) < \infty$. Dann folgt:

$$\lim_{n\to\infty} \left(\frac{1}{a_n} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i \,|\, X_{i-1}, \dots, X_1)) \right) = 0 \qquad P - f.s.$$

Beweis: $\rightsquigarrow$ Übung. □

12.5. Gesetz vom iterierten Logarithmus

Dieses Gesetz stellt für i.i.d. Folgen von Zufallsvariablen einen Bezug zur Streuung her. Es gilt zwar das starke Gesetz der großen Zahlen und der klassische Zentrale Grenzwertsatz, aber beide sagen nichts über die Fluktuation von $S_n = X_1 + \cdots + X_n$ aus. Das Gesetz vom iterierten Logarithmus verschärft die Aussage

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d} \mathrm{N}(0; \sigma^2) ,$$

wobei $\sigma^2 := \mathrm{Var}(X_i)$.

12.5.1. Satz vom iterierten Logarithmus

Für jede i.i.d. Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ auf $(\Omega, \mathcal{A}, P)$, die quadratintegrierbar und zentriert ist (d.h. die Varianzen existieren und $\mathbb{E}(X_n) = 0$), gelten für die Folge der Partialsummen $(S_n)_{n\in\mathbb{N}}$, $S_n := \sum_{i=1}^{n} X_i$ folgende Beschränkungseigenschaften $P - f.s.$:

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\, \log\log n}} = +\sqrt{\mathrm{Var}(X_i)} = +\sigma$$
$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\, \log\log n}} = -\sqrt{\mathrm{Var}(X_i)} = -\sigma$$

Interpretation:

Für fast alle ω konvergiert die Folge $(S_n)_{n\in\mathbb{N}}$ *nicht*, sie besitzt jedoch $\pm\sigma$ als extremste Häufungspunkte.

12.6. Zentraler Grenzwertsatz für Martingaldifferenzenfolgen

Ziel: Zentraler Grenzwertsatz für *abhängige* Variablen, hier speziell für MDF-Dreiecksschemata.

12.6.1. Definition (Dreiecksschema)

Ein *Dreiecksschema* ist eine doppelt indizierte Familie von Zufallsvariablen $(X_{nj} : 1 \leq j \leq n,\, n \in \mathbb{N})$ auf $(\Omega, \mathcal{A}, P)$ mit der kanonischen Filtration

$$\begin{aligned} \mathcal{A}_{n0} &:= \{\emptyset, \Omega\} \\ \mathcal{A}_{nj} &:= \sigma(X_{n1}, \ldots, X_{nj}) \qquad 1 \leq j \leq n,\, n \in \mathbb{N}\,. \end{aligned}$$

und $X_{n0} = 0$ (falls nötig). Außerdem seien

$$\begin{array}{l} \mu_{nj} := \mathbb{E}(X_{nj} \,|\, \mathcal{A}_{n,j-1}) \\ \sigma_{nj}^2 := \mathbb{E}(X_{nj}^2 \,|\, \mathcal{A}_{n,j-1}) \end{array} \qquad \Rightarrow \qquad \operatorname{Var}(X_{nj} \,|\, \mathcal{A}_{n,j-1}) = \sigma_{nj}^2 - \mu_{nj}^2$$

sowie

$$\begin{aligned} V_{nj}^2 &:= \sum_{i=1}^{j} \sigma_{ni}^2 = \sum_{i=1}^{j} \mathbb{E}(X_{ni}^2 \,|\, \mathcal{A}_{n,i-1}) \\ V_n^2 &:= V_{nn}^2 \end{aligned}$$

Skizze

$$\begin{array}{ccccccc} & & X_{11} & & & n=1 & 1 \leq j \leq 1 \\ & X_{21} & & X_{22} & & n=2 & 1 \leq j \leq 2 \\ X_{31} & & X_{32} & & X_{33} & n=3 & 1 \leq j \leq 3 \\ \vdots & & \vdots & & \vdots & \vdots & \end{array}$$

In der Anwendung sind häufig die Zufallsvariablen $X_{n1}, \ldots, X_{nn}$ einer Zeile unabhängig und die Abhängigkeit besteht entlang der Zeilen, d.h. in der Skizze in vertikaler Richtung. In diesem Fall ist $\mu_{nj} = \mathbb{E}(X_{nj})$ und $V_{nj}^2 = \sum_{i=1}^{j} \mathbb{E}(X_{ni}^2)$. Später werden dann die Partialsummen S_n als Zeilensummen $\sum_{j=1}^{n} X_{nj}$ definiert und deren Konvergenzverhalten für $n \to \infty$ betrachtet.

12.6.2. Definition (Konditionierte Lindeberg-Bedingung (KL))

Ein Dreiecksschema $(X_{nj}, \mathcal{A}_{nj})_{1\leq j\leq n,\, n\in\mathbb{N}}$ erfüllt die *konditionierte Lindeberg-Bedingung* (KL) $:\Longleftrightarrow$

$$\forall \varepsilon > 0: \quad KL_n(\varepsilon) := \sum_{j=1}^{n} \mathbb{E}\left(X_{nj}^2 \cdot I_{\{|X_{nj}|>\varepsilon\}} \,\middle|\, \mathcal{A}_{n,j-1}\right) \quad \xrightarrow[n\to\infty]{p} \quad 0\,,$$

das heißt, die Folge $(KL_n(\varepsilon))_{n\in\mathbb{N}}$ konvergiert für alle ε stochastisch gegen Null.

Hinweis:

Die zugehörige "klassische„Lindeberg-Bedingung (hier für standardisierte, zentrierte, unabhängige Zufallsvariablen $X_{n1}, \ldots, X_{nn}$, $n \in \mathbb{N}$) lautet

$$\forall \varepsilon > 0: \quad L_n(\varepsilon) := \sum_{j=1}^{n} \int\limits_{\{|X_{nj}|>\varepsilon\}} X_{nj}^2 \mathrm{d}P = \sum_{j=1}^{n} \mathbb{E}\left(X_{nj}^2 \cdot I_{\{|X_{nj}|>\varepsilon\}}\right) \quad \xrightarrow{n\to\infty} \quad 0$$

(vgl. 11.4.2).

12.6.3. Satz (Zentraler Grenzwertsatz für Martingaldifferenzenfolgen)

Sei $(X_{nj}, \mathcal{A}_{nj})_{1\leq j\leq n,\, n\in\mathbb{N}}$ ein *MDF*-Dreiecksschema, d.h.

$$\mu_{nj} = \mathbb{E}(X_{nj} \,|\, \mathcal{A}_{n,j-1}) = 0 \qquad \text{für alle } 1 \leq j \leq n,\ n \in \mathbb{N}\,.$$

Dies impliziert für die Zeilensummen $S_n := \sum_{j=1}^{n} X_{nj}$, dass $\mathbb{E}(S_n \,|\, \mathcal{A}_{n,n-1}) = 0$. (Eine äquivalente Forderung wäre: Die Partialsummen S_n bilden ein Martingal.)

Außerdem erfülle das Dreiecksschema die konditionierte Lindeberg-Bedingung und es gelte

$$V_n^2 = \sum_{j=1}^{n} \mathbb{E}(X_{nj}^2 \,|\, \mathcal{A}_{n,j-1}) \xrightarrow{p} 1\,.$$

Dann folgt

$$S_n \xrightarrow{d} \mathrm{N}(0;1)\,.$$

13. Verteilungsfamilien

13.0. Einführung

Ganz allgemein versteht man unter einer Verteilungsannahme eine Klasse (oder Menge) $\mathcal{P}$ von Verteilungen über einem Stichprobenraum $(\mathfrak{X}, \mathfrak{B})$ (Messraum), in Zeichen: $(\mathfrak{X}, \mathfrak{B}, \mathcal{P})$. $\mathcal{P}$ ist eine Teilmenge der Klasse aller Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathfrak{B})$.

Das Ziel statistischer Inferenz ist es dann, eine bestimmte Teilmenge von $\mathcal{P}$ aufgrund einer Beobachtung aus $\mathfrak{X}$ als besonders plausibel zu erklären.

Oft wird so eine Klasse von Verteilungen $\mathcal{P}$ mit einem *Parameter* $\theta \in \Theta$ indiziert. Die Gesamtheit Θ solcher Parameterwerte heißt auch *Parameterraum*. Unter einer *Parametrisierung* versteht man ganz allgemein eine bijektive Abbildung von $\mathcal{P}$ in Θ. Solch eine Abbildung existiert immer.

Interessant sind "vernünftige„Parametrisierungen, z.B.

- $\{\mathrm{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$, mit $\Theta = \mathbb{R} \times \mathbb{R}_+ \subset \mathbb{R}^2$ (zweiparametrisch)
- $\{\mathrm{Po}(\lambda) : \lambda > 0\}$ mit $\Theta = \mathbb{R}_+ \subset \mathbb{R}$ (einparametrisch)
- $\{\mathrm{B}(n, p) : p \in [0; 1]\}$ mit n fest und $\Theta = [0; 1] \subset \mathbb{R}$ (einparametrisch)

also Parameterräume, die eine Teilmenge von $\mathbb{R}^k$ $(k \geq 1)$ bilden. $\Theta \subset \mathbb{R}^k$ kann als "gutartiger„Parameterraum angesehen werden.

Aber auch die Forderung einer bijektiven Abbildung genügt noch nicht ganz. Beispielsweise soll aus $P_n \to P_0$ (in $\mathcal{P}$, konvergent) auch für die entsprechenden Parameter $\theta_n \to \theta_0$ (in Θ, konvergent) folgen, und umgekehrt. Deshalb möchte man sogenannte *stetige Parametrisierungen*, d.h. $\theta_n \to \theta_0 \Rightarrow P_{\theta_n} \to P_{\theta_0}$. Hierzu braucht man aber Konvergenzbegriffe auf $\mathcal{P}$. Wir induzieren diese durch Metriken und betrachten nur sogenannte dominierte Verteilungsfamilien, d.h. Verteilungsfamilien die nach dem Satz von Radon-Nikodym (Satz 7.4) auch Dichten bzgl. eines σ-finiten Maßes besitzen.

Einschub: Totalvariationsabstand, Hellinger-Abstand

Seien P, Q Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathfrak{B})$, welche bzgl. eines σ-endlichen Maßes μ auf $(\mathfrak{X}, \mathfrak{B})$ die Dichten p bzw. q haben. Dann ist mit

$$\|P - Q\| := \frac{1}{2} d_1(P, Q) = \frac{1}{2} \int |p - q| \ \mathrm{d}\mu = \sup_{B \in \mathfrak{B}} |P(B) - Q(B)|$$

der sogenannte *Totalvariationsabstand* definiert.

Mit

$$H(P, Q) := \left(\frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2 \ \mathrm{d}\mu \right)^{\frac{1}{2}} = \left(1 - \int \sqrt{pq} \, \mathrm{d}\mu \right)^{\frac{1}{2}}$$

ist der sogenannte *Hellinger-Abstand* definiert.

Beim Totalvariationsabstand integriert man also über den Abstand der beiden Dichten und erhält damit einen Ausdruck der an die Kolmogorow-Smirnow-Teststatistik erinnert. Manchmal wird auch behauptet, der Totalvariationsabstand sei besser in der Testtheorie und der Hellinger-Abstand besser für die Schätztheorie. Jedoch sind die Abstände zumindest topologisch äquivalent. Es gilt:

$$\begin{gathered} 0 \le \|P - Q\| \le 1 \quad \text{und} \quad 0 \le H(P, Q) \le 1 \\ \|P - Q\| = 0 \Longleftrightarrow P = Q \Longleftrightarrow H(P, Q) = 0 \\ \|P - Q\| = 1 \Longleftrightarrow P \perp Q \Longleftrightarrow H(P, Q) = 1 \end{gathered}$$

(Wobei $P \perp Q$ bedeutet, dass Nullmengen von P keine Nullmengen von Q sind und umgekehrt.)

Deshalb befassen wir uns im folgenden nur mit dem Totalvariationsabstand.

Eine Verteilungsannahme heißt nun *stetig parametrisiert*, wenn es einen metrischen Raum (Θ, d) und eine bijektive und stetige Abbildung

$$\tau : (\Theta, d) \to (\mathcal{P}, \|\cdot\|)$$

gibt (und damit auch nach $(\mathcal{P}, H(\cdot,\cdot))$). Das heißt insbesondere

$$d(\theta_n, \theta_0) \to 0 \Longrightarrow \|P_{\theta_n} - P_{\theta_0}\| \to 0$$

Die Klasse der Verteilungen wird dann mit $\mathcal{P} = (P_\theta : \theta \in \Theta)$ notiert. "Gutartig„ist $\Theta \subset \mathbb{R}^k$ mit euklidischem Abstand d auf Θ. Um dies zu präzisieren benötigt man den Begriff der "Separabilität„.

Vorbereitende Begriffe

dicht Sei $(V, \mathcal{T})$ ein topologischer Raum. Eine Teilmenge $W \subset V$ heißt *dicht* $\Longleftrightarrow \forall v \in V$ $\forall$ Umgebungen $U(v) : U(v) \cap W \neq \emptyset$. Klassisches Beispiel einer dichten Teilmenge ist die Menge der rationalen Zahlen $\mathbb{Q} \subset \mathbb{R}$. Man sagt, $\mathbb{Q}$ liegt dicht in $\mathbb{R}$, weil in jeder ε-Umgebung jeder reellen Zahl eine rationale Zahl liegt.

separabel Ein topologischer Raum V heißt *separabel*, wenn es eine dichte und abzählbare Teilmenge $W \subset V$ gibt. Klassisches Beispiel eines separablen topologischen Raums ist $\mathbb{R}$ mit $W = \mathbb{Q}$, denn $\mathbb{Q}$ ist abzählbar!

polnisch Ein *polnischer Raum* ist ein vollständiger, separabler Raum. Klassisches Beispiel ist wieder $\mathbb{R}$.

lineare Unabhängigkeit von Funktionen Sei V ein Vektorraum über dem Körper K und für $i \in \{1, \ldots, n\}$ seien $l_i : Y \to V$ Funktionen mit Werten in diesem Vektorraum, sowie $\alpha_i \in K$. Die Funktionen $l_1, \ldots, l_n$ heißen *linear unabhängig* $:\Longleftrightarrow$

$$\sum_{i=1}^{n} \alpha_i \, l_i(y) \equiv \mathbf{0} \quad (\forall y \in Y) \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_n = 0$$

In diesem Fall lässt sich keine der Funktionen durch Linearkombination der anderen darstellen.

Borel-σ-Algebra über Θ Sei Θ mit einer Topologie $\mathcal{T}$ versehen, wobei die Elemente der Topologie "offenen Mengen„(bzgl. $\mathcal{T}$) heißen. Eine Borel-σ-Algebra über Θ ist nun die kleinste σ-Algebra über Θ, welche alle offenen Mengen enthält.

abzählbar erzeugt Eine σ-Algebra $\mathcal{A}$ heißt *abzählbar erzeugt*, wenn $\mathcal{A}$ einen abzählbaren Erzeuger enthält. Klassisches Beispiel ist die von den Intervallen mit rationalen Eckpunkten erzeugte Borelsche σ-Algebra. (Der Erzeuger ist abzählbar, da er wegen der Abzählbarkeit von $\mathbb{Q}$ nur abzählbar viele Intervalle enthält; die Intervalle selbst sind überabzählbar.)

Vervollständigung einer σ-Algebra $\mathcal{A}$ bzgl. eines Maßes ν Dies ist die kleinste σ-Algebra $\mathcal{D} \supset \mathcal{A}$, welche alle Teilmengen von ν-Nullmengen enthält.

Gruppe Eine Paar $(\mathcal{G}, \circ)$ mit einer Menge $\mathcal{G}$ und einer inneren zweistelligen Verknüpfung $\circ$ – d.h. $g_1, g_2 \in \mathcal{G} \Rightarrow g_1 \circ g_2 \in \mathcal{G}$ – heißt *Gruppe* $:\Longleftrightarrow$

1. $\forall\, g_1, g_2, g_3 \in \mathcal{G}: \quad (g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ (Assoziativität)
2. $\exists\, id \in \mathcal{G}\ \forall\, g \in \mathcal{G}: \quad g \circ id = id \circ g = g$ (Neutrales Element)
3. $\forall\, g \in \mathcal{G}\ \exists\, g^{-1} \in \mathcal{G}: \quad g \circ g^{-1} = g^{-1} \circ g = id$ (Inverses Element)

Beispiel: Die Menge $\{T_a : a \in \mathbb{R}\}$ der Verschiebungsabbildungen mit $T_a(x) = x + a$ und der Verkettung $(g_1 \circ g_2)(x) = g_1(g_2(x))$ ist eine Gruppe mit dem neutralen Element T_0 und dem zu T_a inversen Element T_{-a}.

Symmetrische Gruppe Die Gruppe $\mathcal{S}_n$, die aus allen Permutationen (Vertauschungen) einer n-elementigen Menge besteht, heißt *symmetrische Gruppe* (oder Permutationsgruppe) (auf n). Das heißt formal:

$$\mathcal{S}_n := \{\pi : \{1, \ldots, n\} \to \{1, \ldots, n\} \,|\, \pi \text{ bijektiv}\}$$

13.1. Parametrische Verteilungsfamilien mit Parameterraum $\Theta \subset \mathbb{R}^k$

13.1.1. Beispiele

(i) Eindimensionale Normalverteilungen $\{\mathrm{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ mit Parameterraum $\Theta = \mathbb{R} \times \mathbb{R}_+ \subset \mathbb{R}^2$ (zweiparametrisch)

(ii) Binomialverteilungen $\{\mathrm{B}(n, p) : p \in [0; 1]\}$ mit festem n und Parameterraum $\Theta = [0; 1]$ oder $(0; 1) \subset \mathbb{R}$ (einparametrisch)

13.1.2. Parametrische Verteilungsfamilien, Parametrisierung

Eine Familie von Verteilungen $\mathcal{P}$ auf einem Messraum $(\mathfrak{X}, \mathfrak{B})$ heißt (formal) *parametrisch*, wenn es eine bijektive Abbildung von einem $\Theta \subset \mathbb{R}^k$ in $\mathcal{P}$ gibt. Gemäß der Dimension k spricht man auch von *k-parametrisch* (k-parametrig).

Das Bild von $\theta \in \Theta$ wird mit P_θ bezeichnet und $\mathcal{P}$ wird als $\{P_\theta : \theta \in \Theta\}$ notiert. Die Abbildung heißt auch *Parametrisierung*.

Da die Abbildung (Parametrisierung) stetig sein soll, braucht man Topologien. Die Stetigkeitsfrage ist einfach zu handhaben, wenn die Verteilungsfamilien dominiert sind (und damit entsprechende Dichten besitzen):

13.1.3. Dominierte Verteilungsfamilien

Eine Familie $\mathcal{P}$ von Verteilungen heißt *dominiert*, wenn es ein σ-finites Maß ν auf $(\mathfrak{X}, \mathfrak{B})$ gibt, sodass jedes $P \in \mathcal{P}$ von ν dominiert wird. D.h.

$$\forall P \in \mathcal{P}: \quad \nu(B) = 0 \implies P(B) = 0 \qquad \text{(vgl. Definition 7.3).}$$

Nach dem Satz von Radon-Nikodym (Satz 7.4) besitzt dann jedes P einer von ν dominierten Verteilungsfamilie eine ν-Dichte f. Ist $\mathcal{P}$ außerdem parametrisierbar, dann lässt sich $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ auch durch eine Familie entsprechender Dichten darstellen: $\{f(x;\theta) : \theta \in \Theta\}$. Im Folgenden werden wir nur solche Verteilungen betrachten.

Für die Statistik wichtig ist nämlich der "Standardfall„$\mathfrak{X} = \mathbb{R}^n$, $\mathfrak{B} = \mathfrak{B}^{\otimes n}$ mit den zwei folgenden Typen von Maßen:

n-dimensionales Lebesgue-Maß ($\nu = \lambda^n$): Hierbei ist $\{f(x;\theta) : \theta \in \Theta\}$ eine Familie von Lebesgue-Dichten und $\mathcal{P}$ ist eine Familie von Lebesgue-stetigen (kurz: stetigen) Verteilungen.
Beispiel: eindimensionale Normalverteilungen $\mathrm{N}(\mu, \sigma^2)$ mit $\theta = (\mu, \sigma^2)$ und $f(x;\theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2})$

endliches bzw. abzählbares Zählmaß ($\nu = \#$): In diesem Fall besteht $\mathcal{P}$ aus diskreten Verteilungen mit der Eigenschaft, dass $\{x \,|\, \exists\, \theta \in \Theta : P_\theta(x) > 0\}$ eine abzählbare Menge ist.

Beispiel: Binomialverteilungen $B(n,p)$ mit $\theta = p \in \Theta = (0;1)$ (bzw. $[0;1]$). x beschreibt die Anzahl der Erfolge bei n Versuchen. $\mathfrak{X} = \{0, \dots, n\}$.

In diesen beiden Fällen für ν ist die "Stetigkeitsfrage„besonders einfach, denn:

13.1.4. Stetige Parametrisierung

Eine dominierte Verteilungsfamilie $\mathcal{P}$ heißt *stetig parametrisiert (im starken Sinn)*, wenn die Dichten $f(x;\theta)$ stetig in θ sind für $P - f.a.\ x$.

Ist $\mathcal{P}$ "nur„darstellbar durch die Verteilungsfunktionen $\{F(x;\theta) : \theta \in \Theta\}$, dann heißt die Verteilungsfamilie $\mathcal{P}$ *stetig parametrisiert im schwachen Sinn*, wenn $F(x;\theta)$ stetig in θ ist für $P - f.a.\ x$.

In den obigen Beispielen sieht man, dass die λ-Dichte der eindimensionalen Normalverteilung stetig in (μ, σ^2) ist und dass die #-Dichte der Binomialverteilung $\binom{n}{x} p^x (1-p)^{n-x}$ stetig in p ist (Polynom). Deshalb heißen diese beiden Familien stetig parametrisiert.

Hinweise:

- "Für $P - f.a.\ x$„heißt, "für alle x bis auf eine Menge $N \subset \mathfrak{X}$, für welche $P(N) = 0$ für alle $P \in \mathcal{P}$ gilt.„
- Für $\mathfrak{X} \subset \mathbb{R}^n$ gilt: Nicht-parametrische Verteilungsfamilien sind solche, die weder im starken noch im schwachen Sinn stetig parametrisierbar sind.
- Für entsprechende Verteilungsfamilien fallen schwache und starke Stetigkeit sowie Stetigkeit bzgl. dem Totalvariationsabstand und dem Hellinger-Abstand zusammen.

Parametrisierung durch Teilmengen des $\mathbb{R}^k$ haben andere schöne Eigenschaften für dominierte Familien von Verteilungen. Dies hängt wesentlich an der Separabilität von $\mathbb{R}^k$.

13.2. Exponentialfamilien

13.2.1. Definition (Exponentialfamilie)

Eine dominierte Verteilungsfamilie $\mathcal{P}$ auf $(\mathfrak{X}, \mathfrak{B})$ heißt *Exponentialfamilie*, wenn $\mathcal{P}$ parametrisch ist, d.h. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, und wenn sich die jeweiligen Dichten bzgl. eines geeigneten $\mathcal{P}$ dominierenden Maßes ν in der Gestalt

$$f(x;\theta) = b(x)\, c(\theta)\, \exp\left(\sum_{i=1}^{k} \gamma_i(\theta)\, T_i(x)\right) \tag{13.1}$$

darstellen lassen. Hierbei gilt:

- b, $T_1, \ldots, T_k$ sind reellwertige, messbare Funktionen auf $\mathfrak{X}$
- c, $\gamma_1, \ldots, \gamma_k$ sind reellwertige Funktionen auf Θ
- $b(x) \geq 0$ und $c(\theta) > 0$

Mit den vektorwertigen Funktionen

$$\begin{aligned} \gamma : \Theta &\to \mathbb{R}^k & \qquad T : \mathfrak{X} &\to \mathbb{R}^k \\ \theta &\mapsto (\gamma_1(\theta), \ldots, \gamma_k(\theta)) & x &\mapsto (T_1(x), \ldots, T_k(x)) \end{aligned}$$

spricht man auch von einer Exponentialfamilie in T und γ.

Der transformierte Parameter $\gamma = \gamma(\theta)$ heißt auch *natürlicher Parameter*.

Hinweis:

Nach Definition ist $b(x)$ der einzige Faktor, für welchen die Dichten 0 werden können und dieser hängt nicht von θ ab. Somit haben alle Verteilungen einer Exponentialfamilie den gleichen Träger, oder anders ausgedrückt: Für je zwei Verteilungen P_θ und $P_{\theta'}$ aus einer Exponentialfamilie gilt für jedes messbare $B \in \mathfrak{B}$

$$P_\theta(B) = 0 \Longleftrightarrow P_{\theta'}(B) = 0\,.$$

Im Umkehrschluss gilt: Verteilungen, deren Träger von einem Verteilungsparameter abhängen, können keiner Exponentialfamilie angehören.

13.2.2. Bemerkung

Sei $\mathcal{P}$ eine Exponentialfamilie mit Dichten wie in (13.1) bzgl. eines dominierenden Maßes ν. Dann ist durch

$$\tilde{\nu}(B) := \int_B b(x)\,\nu(\mathrm{d}x) \qquad (B \in \mathfrak{B})$$

ein weiteres $\mathcal{P}$ dominierendes Maß $\tilde{\nu}$ definiert (σ-finit!) und die Dichten bzgl. $\tilde{\nu}$ haben die Gestalt

$$\tilde{f}(x;\theta) = c(\theta)\,\exp\left(\sum_{i=1}^{k} \gamma_i(\theta)\,T_i(x)\right). \tag{13.2}$$

Mit $\gamma_0(\theta) := \log(c(\theta))$ lassen sich die Dichten schließlich wie folgt schreiben:

$$\tilde{f}(x;\theta) = \exp\left(\gamma_0(\theta)\cdot 1 + \sum_{i=1}^{k} \gamma_i(\theta)\,T_i(x)\right).$$

Diese Transformationen haben weniger praktischen Nutzen, sie sollen vielmehr verdeutlichen, dass alle wesentlichen Eigenschaften von der Gestalt des *Exponenten* abhängen.

Hinweis:

Für jedes $\theta \in \Theta$ und $B \in \mathfrak{B}$ gilt

$$P_\theta(B) = 0 \Longleftrightarrow \tilde{\nu}(B) = 0\,,$$

d.h. $\tilde{\nu}$ ist äquivalent mit jedem P_θ.

13.2.3. Defintion

Eine Exponentialfamilie heißt *(strikt) k-parametrisch*, wenn ihre Dichten in die Gestalt (13.1) bzw. (13.2) gebracht werden können und dabei für alle $\theta \in \Theta$ gilt:

- $1, \gamma_1(\theta), \ldots, \gamma_k(\theta)$ sind linear unabhängig
- $1, T_1(x), \ldots, T_k(x)$ sind auf dem Komplement einer jeden $\tilde{\nu}$-Nullmenge

linear unabhängig

Inhaltlich bedeutet dies, dass die Anzahl k der Parameter der Exponentialfamilie nicht reduzierbar ist. Jeder (natürliche) Parameter $\gamma_1, \ldots, \gamma_k$ leistet seinen eigenen Beitrag.

13.2.4. Satz

Ist $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ eine k-parametrische Exponentialfamilie in $T = T(x)$ und $\gamma = \gamma(\theta)$, so bildet $\mathcal{P}^T = \{P_\theta^T : \theta \in \Theta\}$ eine k-parametrische Exponentialfamilie in der Identität $V(t) = t$ und γ.

Bei Parametrisierung durch γ können die entsprechenden Dichten von $\mathcal{P}^T$ auf die Gestalt $\tilde{c}(\gamma)\, \exp\left(\sum_{i=1}^k \gamma_i\, t_i\right)$ (bzgl. $\tilde{\nu}^T$) gebracht werden.

13.3. Invariante Verteilungsfamilien

Motivation: Beispielsweise bei Längen- oder Temperaturmessung soll die Verteilungsklasse unabhängig von der verwendeten Skala sein.

13.3.1. Definition (invariante Verteilungsfamilie)

Sei $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ eine parametrische Verteilungsfamilie auf $(\mathfrak{X}, \mathfrak{B})$ und $\mathcal{G}$ eine Gruppe von zulässigen Transformationen $g : \mathfrak{X} \to \mathfrak{X}$, d.h. g ist bijektiv und g und g^{-1} sind messbar.

Dann heißt $\mathcal{P}$ *invariant* bzgl. $\mathcal{G}$, wenn $\mathcal{P}^g = \{P_\theta^g : \theta \in \Theta\} = \mathcal{P}$, für alle $g \in \mathcal{G}$, wobei $P_\theta^g := g(P_\theta)$.

13.3.2. Satz

Zum Nachweis der Invarianz einer Verteilungsfamilie $\mathcal{P}$ bzgl. einer Gruppe $\mathcal{G}$ genügt es $\mathcal{P}^g \subset \mathcal{P}$ für jedes $g \in \mathcal{G}$ zu zeigen.

Beweis: $\rightsquigarrow$ Übung. $\square$

13.3.3. Beispiel

Die Familie der Normalverteilungen $\mathcal{P} = \left\{ \mathrm{N}(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma \in \mathbb{R}_+ \right\}$ ist invariant gegenüber der Gruppe der affin-linearen Transformationen $T(x) = b\,x + a$, mit $a \in \mathbb{R}$, $b \in \mathbb{R}_+$.

Beweis:

a) Diese Transformationen bilden eine Gruppe.
   - $T_1 \circ T_2$ ist wieder eine affin-lineare Abbildung, denn $(T_1 \circ T_2)(x) = T_1(T_2(x)) = T_1(b_2\,x + a_2) = b_1\,(b_2\,x + a_2) + a_1 = b_1\,b_2\,x + (b_1\,a_2 + a_1)$.
   - Das neutrale Element ist die Identitätsabbildung mit $a = 0$ und $b = 1$.
   - Das inverse Element ist die durch $T^{-1}(y) = \frac{y-a}{b} = \frac{1}{b}\,y - \frac{a}{b}$ definierte affin-lineare Abbildung, denn $T(T^{-1}(y)) = b\,(\frac{1}{b}\,y - \frac{a}{b}) + a = y$.

b) Diese Transformationen sind zulässig, denn sie sind bijektiv und messbar.

c) Die Verteilungsfamilie ist invariant, da

$$\mathcal{P}^T = \{\mathrm{N}(b\,\mu + a, b^2\,\sigma^2) : \mu, a \in \mathbb{R},\ \sigma, b \in \mathbb{R}_+\} \subset \mathcal{P}\,,$$

denn mit $y = T(x)$ ist

$$\begin{aligned} f_Y(y) &= \left|\frac{\mathrm{d}}{\mathrm{d}y} T^{-1}(y)\right| \cdot f_X\left(T^{-1}(y); \mu, \sigma^2\right) \\ &= \frac{1}{b} \cdot \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sigma}\,e^{-\frac{1}{2}\frac{1}{\sigma^2}(T^{-1}(y)-\mu)^2} \\ &= \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sigma\,b}\,e^{-\frac{1}{2}\frac{1}{\sigma^2}(\frac{1}{b}\,y-\frac{a}{b}-\mu)^2} \\ &= \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sigma\,b}\,e^{-\frac{1}{2}\frac{1}{b^2\,\sigma^2}(y-(a+b\,\mu))^2}\,, \end{aligned}$$

d.h. die Transformation der Stichprobe mit $T(x) = b\,x + a$ überträgt sich auf die Parameter und die transformierte Dichte ist wieder in $\mathcal{P}$.

□

Dieser Zusammenhang ist natürlich schon aus dem Bachelorstudium bekannt:

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right) \quad \Longrightarrow \quad b\,X + a \sim \mathrm{N}\left(b\,\mu + a, b^2\,\sigma^2\right)$$

Die wichtigen Arten der Invarianz gegenüber Verschiebungs- und/oder Skalierungs-Transformationen besitzen eigene Bezeichnungen:

13.3.4. Defintion (Lokationsfamilie)

Eine Verteilungsfamilie $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ mit den Eigenschaften

1. Invarianz gegenüber Transformationen der Art $T_a(x) = x + a$, mit $a \in \mathbb{R}$
2. Existenz einer "festen„Verteilung $P_0 \in \mathcal{P}$, sodass $P_\theta(B) = P_0(T_\theta^{-1}(B))$

heißt *Lokationsfamilie.*

Beispiel

$\left\{\mathrm{N}(\mu, \sigma_0^2) : \mu \in \mathbb{R}\right\}$: Familie der Normalverteilungen mit bekannter Varianz σ_0^2. Die Standard-Verteilung ist $P_0 \equiv \mathrm{N}(0, \sigma_0^2)$.

13.3.5. Definition (Skalenfamilie)

Eine Verteilungsfamilie $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}_+\}$ mit den Eigenschaften

1. Invarianz gegenüber Transformationen der Art $T_b(x) = b\,x$, mit $b \in \mathbb{R}_+$
2. Existenz einer "festen„Verteilung $P_1 \in \mathcal{P}$, sodass $P_\theta(B) = P_1(T_\theta^{-1}(B))$

heißt *Skalenfamilie.*

Beispiel

$\left\{\mathrm{N}(\mu_0, \sigma^2) : \sigma \in \mathbb{R}_+\right\}$: Familie der Normalverteilungen mit bekanntem Erwartungswert μ_0. Die Standard-Verteilung ist $P_1 \equiv \mathrm{N}(\mu_0, 1)$.

13.3.6. Satz

Ist $\mathcal{P}$ eine Lokations- *bzw.* Skalenfamilie und ist $X \sim P_\theta \in \mathcal{P}$, dann gilt

$$T(X) = X + a \sim P_{\theta+a} \qquad \textit{bzw.} \qquad T(X) = b\,X \sim P_{b\,\theta}\,.$$

Beweis: $\rightsquigarrow$ Übung. $\square$

13.3.7. Definition (Lokations- und Skalenfamilie)

Eine Verteilungsfamilie $\mathcal{P} = \{P_{\eta,\tau} : (\eta, \tau) = \theta \in \Theta \subset \mathbb{R} \times \mathbb{R}_+\}$ mit den Eigenschaften

1. Invarianz gegenüber Transformationen der Art $T_{\eta,\tau}(x) = \tau\, x + \eta$, mit $\eta \in \mathbb{R}$, $\tau \in \mathbb{R}_+$
2. Existenz einer festen Verteilung $P_{0,1} \in \mathcal{P}$, sodass $P_{\eta,\tau}(B) = P_{0,1}(T_{\eta,\tau}^{-1}(B))$

heißt *Lokations- und Skalenfamilie.*

Beispiel

$\{\mathrm{N}(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma \in \mathbb{R}_+\}$: Familie der Normalverteilungen (vgl. 13.3.3) mit Lokationsparameter μ und Skalenparameter $\sigma > 0$. Die Standard-Verteilung ist $P_{0,1} \equiv \mathrm{N}(0, 1)$.

13.3.8. Satz

Ist $\mathcal{P}$ eine Lokations- und Skalenfamilie mit Lokationsparameter η und Skalenparameter τ und ist $X \sim P_{\eta,\tau} \in \mathcal{P}$, dann gilt

$$T(X) = b\, X + a \sim P_{b\eta+a,b\tau} \, .$$

Beweis: $\rightsquigarrow$ Übung. $\square$

13.3.9. Definition (permutationsinvariante Verteilungsfamilie, Vertauschbarkeit)

Sei $(X_1, \ldots, X_n)$ ein Zufallsvektor mit gemeinsamer Verteilung P. Gilt dann für jede Permutation π von $\{1, \ldots, n\}$ (d.h. für alle Elemente der symmetrischen Gruppe $\mathcal{S}_n$)

$$P^\pi := \pi(P) = P \, ,$$

dann heißt P *permutationsinvariant* und $(X_1, \ldots, X_n)$ *vertauschbar* (“exchangeable„).

Bemerkungen

- Die Bedingung $P^\pi = P$ ist äquivalent mit

$$\forall B : P((X_{\pi(1)}, \dots, X_{\pi(n)}) \in B) = P((X_1, \dots, X_n) \in B)$$

- Die Eigenschaft der Permutationsinvarianz ist schwächer als die der Unabhängigkeit. Beispielsweise sind Ziehungen aus der hypergeometrischen Verteilung zwar permutationsinvariant, da es – wie auch bei der Binomialverteilung – nicht auf die Reihenfolge der Ziehungen ankommt. Sie sind jedoch nicht unabhängig, da die Ziehungen ohne Zurücklegen vorgenommen werden.

- Unabhängige Produktverteilungen sind automatisch permutationsinvariant.

- Orderstatistiken sind quasi das Gegenteil von permutationsinvarianten Verteilungen.

13.3.10. Korollar

Sei $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ eine parametrische Verteilungsfamilie mit permutationsinvarianten P_θ. Dann ist $\mathcal{P}$ invariant bzgl. der symmetrischen Gruppe.

13.4. Motivation zur Bayes-Inferenz

Gegeben ist eine parametrisierte, (vom σ-finiten Maß ν) dominierte Verteilungsfamilie, die durch die entsprechenden ν-Dichten $\{p(x;\theta) : \theta \in \Theta\}$ dargestellt wird. x_0 sei eine Realisation. Außerdem ist eine Priori-Verteilung für θ gegeben, auch dargestellt durch eine ν-Dichte $\pi(\theta)$.

Dann ist die Posteriori-Verteilung von θ gegeben durch die Dichte

$$P(\theta; x_0) = \frac{p(x_0;\theta) \cdot \pi(\theta)}{\int_\Theta p(x_0;\theta)\pi(\theta)\, \nu(\mathrm{d}\theta)} ,$$

wobei im Zähler die gemeinsame Dichte von x_0 und θ steht, und im Nenner die Randdichte $f_R(x_0)$ der Beobachtung x_0.

Die Posteriori-Dichte wird als eine Dichte der bedingten Verteilung von θ unter x_0 aufgefasst. Somit ist vorausgesetzt, dass (x, θ) eine gemeinsame Verteilung besitzen. Gewöhnlich wird $p(x; \theta)$ wie eine "bedingte Dichte„behandelt. Dies ist möglich, wenn

$$K(B, \theta) := \int_B p(x; \theta) \nu(\mathrm{d}x)$$

ein Markov-Kern vom Parameterraum $(\Theta, \mathcal{A}_\Theta)$ in den Stichprobenraum $(\mathfrak{X}, \mathfrak{B})$ ist. Hierbei ist $\mathcal{A}_\Theta$ eine geeignete σ-Algebra.

Dann ist – vermöge dem Verallgemeinerten Satz von Fubini (Satz 10.24) – durch K und π eine gemeinsame Verteilung auf $\mathfrak{X} \times \Theta$ definiert, deren bedingte Verteilung unter θ fast sicher $K(\cdot; \theta)$ ist.

"Umgekehrt„kann dann aus der gemeinsamen Verteilung die bedingte Verteilung (bzw. eine Version davon) unter x_0 angegeben werden – die Posteriori. Dass dies in allen wesentlichen Situationen gut geht, ist Aussage des folgenden Satzes.[1]

13.4.1. Satz

Es seien $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ eine dominierte Verteilungsfamilie mit μ-Dichten $\{p(x; \theta) : \theta \in \Theta\}$ und $p(x; \cdot)$ für jedes x eine stetige Funktion in θ.[2] $\mathcal{A}_\Theta$ bezeichne eine σ-Algebra über Θ.

Dann gilt:

(a) Ist Θ ein separabler, metrischer Raum[3] und $\mathcal{A}_\Theta$ die Borel-σ-Algebra über Θ, dann ist die Abbildung

$$(x, \theta) \longmapsto p(x; \theta)$$

$\mathfrak{B} \otimes \mathcal{A}_\Theta$-messbar.

Die Abbildung

$$(B, \theta) \longmapsto P_\theta(B)$$

ist ein Markov-Kern von $(\Theta, \mathcal{A}_\Theta)$ nach $(\mathfrak{X}, \mathfrak{B})$.

[1] In dem Satz tauchen viele Begriffe auf, die im Abschnitt "Vorbereitende Begriffe„auf Seite 32 erklärt wurden. Bei Bedarf kann man dort also nochmal nachlesen.

[2] Dies ist noch eine etwas stärkere Forderung als "sei $\mathcal{P}$ stetig parametrisiert„(im Sinne von Abschnitt 13.1.4).

[3] d.h. topologisch so gut wie $\mathbb{R}^k$

(b) Sei ν ein σ-finites Maß auf $(\Theta, \mathcal{A}_\Theta)$ und $\mathfrak{B}$ abzählbar erzeugt sowie die Abbildung $(B, \theta) \longmapsto P_\theta(B)$ ein Markov-Kern von $(\Theta, \mathcal{A}_\Theta)$ nach $(\mathfrak{X}, \mathfrak{B})$. Sei $\mathcal{D}$ die Vervollständigung von $\mathfrak{B} \times \mathcal{A}_\Theta$ bzgl. $\mu \otimes \nu$, dann existiert eine $\mathcal{D}$-messbare Funktion q auf $\mathfrak{X} \times \Theta$, sodass $q(\cdot, \theta)$ für jedes $\theta \in \Theta$ eine Version der μ-Dichte von P_θ ist.

13.4.2. Satz (Bayes-Formel für Dichten)

Seien $X : \Omega \to \mathbb{R}^l$ und $Y : \Omega \to \mathbb{R}^k$ Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$.[4]
ν sei ein σ-finites Maß auf $\mathbb{R}^l$ und λ sei ein σ-finites Maß auf $\mathbb{R}^k$.
$\nu \otimes \lambda$ sei das assoziierte (ebenfalls σ-finite) Produktmaß.

(a) Wird die gemeinsame Verteilung von X und Y durch $\nu \otimes \lambda$ dominiert, so werden die bedingten Verteilungen von X unter $Y = y$ durch ν und die bedingten Verteilungen von Y unter $X = x$ von λ dominiert. Es gilt für die entsprechenden Dichten die Bayes-Formel

$$f_{Y\,|\,X}(y\,|\,x) = \frac{1}{f_X(x)} \cdot f_{X\,|\,Y}(x\,|\,y)\; f_Y(y)$$

mit

$$f_X(x) = \int f_{X\,|\,Y}(x\,|\,y)\, f_Y(y)\; \lambda(\mathrm{d}y)\,.$$

(b) Werden die bedingten Verteilungen von X unter $Y = y$ durch ν und die bedingten Verteilungen von Y unter $X = x$ durch λ dominiert, so wird die gemeinsame Verteilung von X und Y durch $\nu \otimes \lambda$ dominiert. Für die entsprechende Dichte gilt

$$f_{X,Y}(x, y) = f_{X\,|\,Y}(x\,|\,y) \cdot f_Y(y)$$

Beispiel:

Sei $X\,|\,Y = \theta$ (“X bedingt durch $Y = \theta$„) verteilt nach $B(n, \theta)$ mit Parameter $Y = \theta$ und Realisierung x, sodass $p(x; \theta) = \binom{n}{x}\, \theta^x\, (1 - \theta)^{n-x}$.
Das diskrete $X\,|\,Y = \theta$ wird durch das Zählmaß $\nu = \#$ dominiert und das stetige $Y\,|\,X = x$ wird durch das eindimensionale Lebesgue-Maß λ dominiert.

[4] Die Bezeichnung Y statt θ ist beabsichtigt, da in der Bayesianischen Ideologie Stichprobe und Parameter strukturell gleich sind.

Dann wird nach Teil (b) die gemeinsame Verteilung von X und Y von $\# \otimes \lambda$ dominiert; es existiert also eine gemeinsame Dichte bzgl. $\# \otimes \lambda$.

Nimmt man dies wiederum als Voraussetzung, so weiß man umgekehrt mit Teil (a), dass $X \,|\, Y = \theta$ durch das Zählmaß $\nu = \#$ dominiert wird und dass $Y \,|\, X = x$ durch das Lebesgue-Maß λ dominiert wird, womit die Posteriori-Verteilung eine entsprechende λ-Dichte $P(\theta; x)$ besitzt.

13.5. Konjugierte Verteilungsfamilien, konjugierte Paare

13.5.1. Einführendes Beispiel

Sei $\mathcal{P}$ die Familie der $\mathrm{B}(n,p)$-Verteilungen, dargestellt über die zugehörigen $\#$-Dichten

$$\left\{ \tbinom{n}{x} p^x (1-p)^{n-x} : p \in [0;1] \right\} .$$

Im Rahmen der Bayes-Inferenz lässt sich die jeweilige Wahrscheinlichkeitsfunktion zu einem Parameter p als bedingte Verteilung von x unter p auffassen, wobei die zugehörige gemeinsame Verteilung von x und p mittels einer Priori-Verteilung über p erzeugt wird: $\pi(p)$ (λ-Dichte auf $[0;1]$). Dann ist

$$\tbinom{n}{x} p^x (1-p)^{n-x} \cdot \pi(p)$$

die gemeinsame Dichte bzgl. $\# \otimes \lambda$. Die Posteriori-Dichte ist eine Dichte bzgl. λ und lautet

$$P(p \,|\, x) = \tbinom{n}{x} p^x (1-p)^{n-x} \cdot \pi(p) \cdot \frac{1}{\underbrace{\displaystyle\int_{[0;1]} \tbinom{n}{x} p^x (1-p)^{n-x} \pi(p) \, \lambda(\mathrm{d}p)}_{=f(x)}} .$$

Dabei bezeichnet $f(x)$ die Randdichte von x bei herausintegriertem Parameter θ. Wie diese zu interpretieren ist, kann diskutiert werden.

Sei nun $\pi(p)$ die Dichte einer Beta-Verteilung mit den *bekannten* Parametern a

und b $(a, b \in \mathbb{R}_+)$, d.h.

$$\begin{array}{rrcl} & \pi(p) & = & \dfrac{1}{B(a;b)}\, p^{a-1}(1-p)^{b-1}\, I_{[0;1]}(p) \\ \text{wobei} & B(a;b) & = & \displaystyle\int_0^1 p^{a-1}(1-p)^{b-1}\,\lambda(\mathrm{d}p) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ & \Gamma(x) & = & \displaystyle\int_0^\infty t^{x-1}\, e^{-t}\,\mathrm{d}t = (x-1)\,\Gamma(x-1) \qquad (\Gamma(1)=1) \\ \text{und} & \mathbb{E}(p) & = & \dfrac{a}{a+b} \\ & \mathrm{Var}(p) & = & \dfrac{ab}{(a+b)^2(a+b+1)} \end{array}$$

Abbildung 13.1 zeigt die Dichten einiger Beta-Verteilungen. Für $a = 1 = b$ erhält man die stetige Gleichverteilung auf dem Intervall $[0; 1]$, ansonsten ist die Verteilung unimodal. Die Beta-Verteilung mit den Parametern $a = b = 0.5$ ist Jeffreys' nicht-informative Priori-Verteilung für die $\mathrm{B}(n, p)$-Verteilung. Diese besitzt die Eigenschaft invariant bzgl. der Gruppe der zulässigen Transformationen zu sein (vgl. 13.3.1). Eine "naive„nicht-informative Priori-Verteilung ist die Gleichverteilung, die diese Eigenschaft nicht besitzt und daher bevorzugt als flache Priori bezeichnet wird.

Für die Posteriori ergibt sich nun

$$\begin{array}{rrcl} & P(p\,|\,x) & = & \binom{n}{x} p^x(1-p)^{n-x} \cdot \dfrac{1}{B(a;b)}\, p^{a-1}(1-p)^{b-1} \cdot \dfrac{1}{f(x)} \\ \text{mit} & f(x) & = & \displaystyle\int \frac{\binom{n}{x}}{B(a;b)}\, p^{a+x-1}(1-p)^{b+n-x-1}\,\lambda(\mathrm{d}p) \\ & & = & \dfrac{\binom{n}{x}}{B(a;b)}\, B(a+x, b+n-x) \end{array}$$

also

$$P(p\,|\,x) = \frac{1}{B(a+x, b+n-x)}\, p^{a+x-1}(1-p)^{b+n-x-1}$$

und mithin eine Beta-Verteilung mit den "aufdatierten„Parametern $a + x$ und $b + n - x$. In der Statistik bezeichnet man dieses Szenario als *Beta-Binomial-Modell.*

Das heißt im Bayesianischen Sinn: Auch durch "Zulernen„verbleibt man in der gleichen Verteilungsfamilie wie die Priori. Dieses Verhältnis führt – etwas allgemeiner – zu folgender Definition.

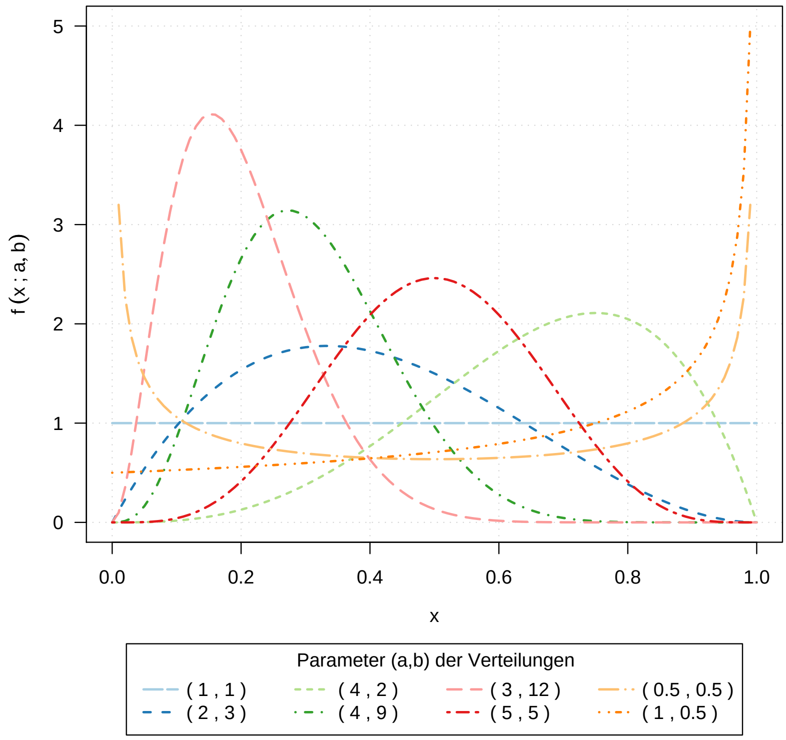

Abbildung 13.1.: Dichtefunktionen einiger Beta-Verteilungen

13.5.2. Definition

Eine Familie $\mathcal{V}$ von Priori-Verteilungen heißt zu einer Familie von Verteilungsannahmen $\mathcal{P}$ (Beobachtungsmodell) *konjugiert*, wenn für alle $\pi \in \mathcal{V}$ und alle $P \in \mathcal{P}$ die Posteriori-Verteilung wieder zu $\mathcal{V}$ gehört.

Ist T eine Statistik und $\mathcal{P}^T$ die Familie der Bildverteilungen, dann wird die Konjugiertheit von $\mathcal{P}^T$ zu $\mathcal{V}$ in analoger Weise erklärt.

Hinweis:

Ist T eine suffiziente Statistik, dann ist $\mathcal{V}$ zu $\mathcal{P}$ genau dann konjugiert, wenn $\mathcal{V}$ zu $\mathcal{P}^T$ konjugiert ist.

Beispiel: $\mathcal{P}$ Bernoulli-Verteilungen, $\mathcal{V}$ Beta-Verteilungen, $X_i \overset{i.i.d.}{\sim} \mathrm{B}(1, p)$ und $T = \sum_{i=1}^n X_i \sim \mathrm{B}(n, p)$ suffizient.

13.5.3. Weitere Beispiele

(a) **Modell:** Familie $\mathcal{P} = \{f(x;\mu,1) : \mu \in \mathbb{R}\}$ der Normalverteilungen mit bekannter Varianz 1, dargestellt durch die entsprechenden Lebesgue-Dichten f

Priori: $\mu \sim \mathrm{N}(\nu, \tau^2)$, d.h.

$$\pi(\mu) \propto e^{-\frac{1}{2\tau^2}(\mu-\nu)^2}$$

Posteriori:

$$\begin{aligned} P(\mu \,|\, x) &\propto e^{-\frac{1}{2}(x-\mu)^2} \cdot e^{-\frac{1}{2\tau^2}(\mu-\nu)^2} \\ &\propto e^{-\frac{1}{2}\frac{\tau^2+1}{\tau^2}(\mu-\frac{x\tau^2+\nu}{\tau^2+1})^2}, \end{aligned}$$

d.h. die Posteriori-Verteilung gehört wieder (wie die Priori-Verteilung) zur Familie der Normalverteilungen.

Beweis: $\rightsquigarrow$ Übung. □

Somit ist die Normalverteilungsfamilie mit Parametern ν und τ^2 (als Priori) zur Normalverteilungsfamilie mit fester Varianz 1 (als Modell) konjugiert.

Da bei $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathrm{N}(\mu, \sigma_0^2)$ mit bekannter Varianz σ_0^2 die Statistik $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ suffizient ist, gilt dieser Zusammenhang auch für $\bar{X}$.

(b) **Modell:** Gleichverteilungen auf dem Intervall $[0;\theta]$ mit Parameter $\theta \in \mathbb{R}_+$

Priori: Pareto-Verteilung $\theta \sim \mathrm{Par}(a,b)$ mit $a, b \in \mathbb{R}_+$ bekannt und Dichte

$$\pi(\theta) = \begin{cases} \frac{a\,b^a}{\theta^{a+1}} & \text{falls } \theta \geq b \\ 0 & \text{sonst} \end{cases}$$

(Der Parameter b begrenzt den Träger von θ also nach unten, vgl. Abbildung 13.2.)

Posteriori:

$$P(\theta \,|\, x) = \frac{1}{\theta} I_{[0;\theta]}(x) \cdot \frac{a\,b^a}{\theta^{a+1}} I_{[b;\infty)}(\theta) \cdot \frac{1}{f(x)}$$

Der Träger der Posteriori-Verteilung ist also $\{\theta \in \mathbb{R}_+ : \theta \geq \max\{x, b\}\}$ und $f(x) = \int_{\max\{x,b\}}^{\infty} \frac{a\,b^a}{\theta^{a+2}}\,\lambda(\mathrm{d}\theta)$.

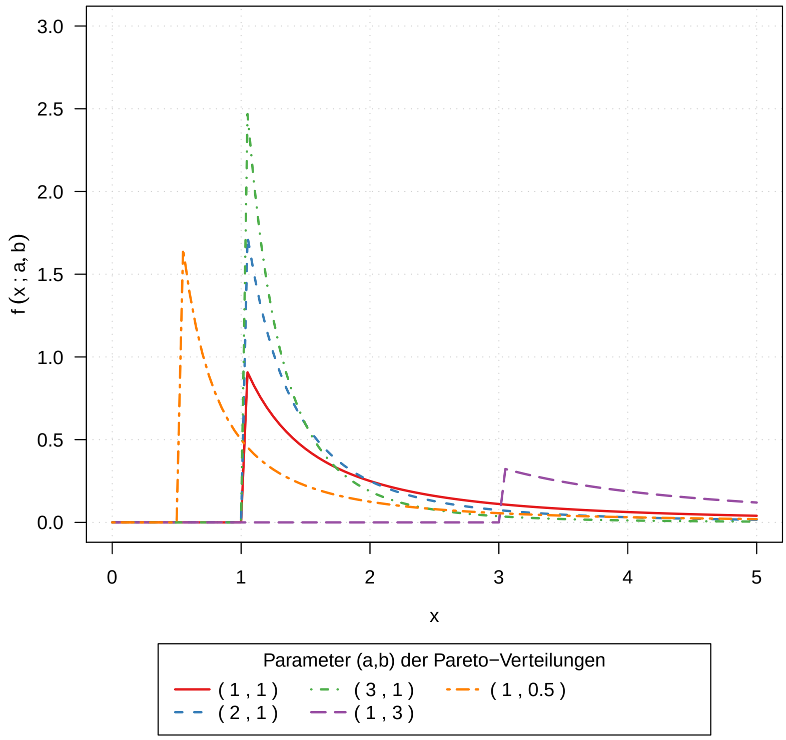

Abbildung 13.2.: Dichtefunktionen einiger Pareto-Verteilungen

Es folgt

$$P(\theta \,|\, x) = \begin{cases} \frac{(a+1)\, m^{a+1}}{\theta^{a+2}} & \text{falls } \theta \geq \max\{x, b\} =: m \\ 0 & \text{sonst} \end{cases}$$

und somit gehört die Posteriori wieder zur Familie der Pareto-Verteilungen. Die Familie der Pareto-Verteilungen ist also konjugiert zur Familie der stetigen Gleichverteilungen auf $[0; \theta]$.

Beweis: $\rightsquigarrow$ Übung. $\square$

13.6. Mischverteilungen

Anwendung finden Mischverteilungen z.B. bei der statistischen Modellierung heterogener Populationen sowie in der Bayes-Inferenz, insbesondere in hierarchischen Modellen zur Formulierung multimodaler Prioris. Man kann dabei finite (diskrete) und stetige Mischverteilungen unterscheiden. Dieser Abschnitt gibt eine kurze Einführung mit Beispielen in diese zwei Varianten von Mischverteilungen.

13.6.1. Finite Mischverteilungen

Modell für heterogene Populationen

Eine Population bestehe aus K Gruppen, die zufällig im Verhältnis der relativen Gruppenanteile $\eta_1, \ldots, \eta_K$ vermischt sind ($\sum_{k=1}^{K} \eta_k = 1$). Dazu sei eine Indikatorvariable $S \in \{1, \ldots, K\}$ definiert, sodass $P(S = k) = \eta_k$ gleich der Wahrscheinlichkeit für eine Beobachtung ("selection„) aus der Gruppe k ist.

Sei Y eine interessierende (stetige) Zufallsvariable, dann ist die gemeinsame Dichte von Y und S bzgl. $\lambda \otimes \#$ gleich

$$f_{Y,S}(y, k) = f_{Y \mid S}(y|k) \cdot \eta_k \,.$$

Falls die bedingten Verteilungen $Y \mid \{S = k\}$ für alle $k \in \{1, \ldots, K\}$ der gleichen parametrischen Verteilungsfamilie (mit Parameterraum Θ) angehören, so lässt sich dies auch als

$$f_{Y,S}(y, k) = f(y; \theta_k) \cdot \eta_k$$

schreiben, wobei $\theta_k \in \Theta$ der zur k-ten Gruppe gehörige Parameter(vektor) ist.

Eine finite Mischverteilung für Y resultiert, falls der Gruppenindikator S nicht beobachtbar (latent, "hidden„) ist. Dann besitzt Y die marginale Dichte

$$f(y) = \sum_{k=1}^{K} f(y; \theta_k)\, \eta_k \,.$$

Beispiele (vgl. Abbildung 13.3):

- Fischerei-Daten: Die empirische Verteilung deutet auf eine trimodale Verteilung hin. Mögliche Mischverteilung ist eine *Normalverteilungsmischung*

  $$Y \sim \eta_1\, \mathrm{N}(\mu_1, \sigma_1^2) + \eta_2\, \mathrm{N}(\mu_2, \sigma_2^2) + \eta_3\, \mathrm{N}(\mu_3, \sigma_3^2)\,,$$

  evtl. mit der Vereinfachung $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$
- Wartezeiten beim Old Faithful Geysir im Yellow Stone Nationalpark: bimodale Normalverteilungsmischung

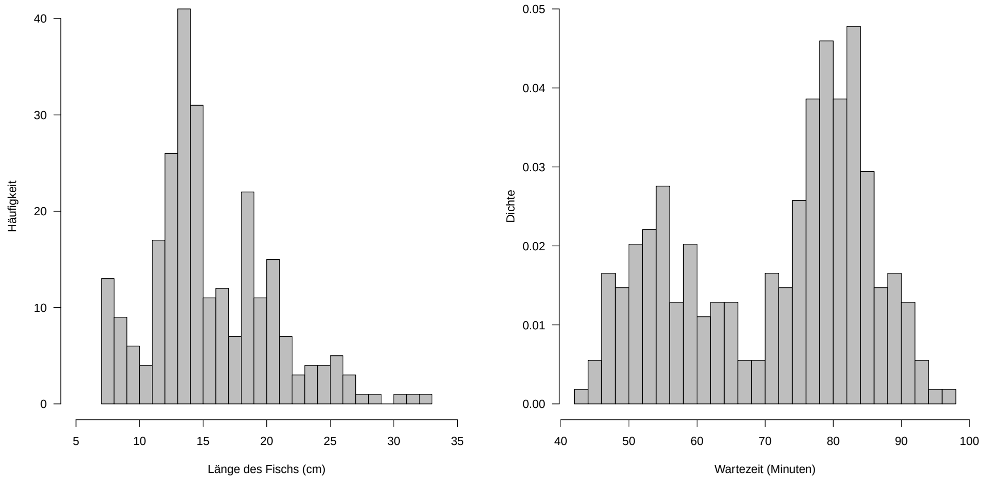

Abbildung 13.3.: Histogramme: Fischerei-Daten (links), Geysir (rechts)

Im Folgenden wird die bayesianische Schreibweise verwendet, d.h. die Parameter θ_k werden als zufällig angesehen und wir schreiben $f(y \,|\, \theta_k)$ statt $f(y; \theta_k)$.

Definition (Finite Mischverteilung)

Die Zufallsvariable $Y \sim f(y \,|\, \psi)$ besitzt eine *finite Mischverteilung*, falls

$$f(y \,|\, \psi) = \eta_1 \, f_1(y \,|\, \theta_1) + \cdots + \eta_K \, f_K(y \,|\, \theta_K) \, .$$

Dabei ist $\psi := (\eta_1, \ldots, \eta_K, \theta_1, \ldots, \theta_K)$, mit $\sum_{k=1}^{K} \eta_k = 1$ und die Dichten $f_k(y \,|\, \theta_k)$ sind diskret oder stetig. Die Parameter η_k werden als *Mischungsanteile* und die Dichten $f_k(y \,|\, \theta_k)$ als *Mischungskomponenten* bezeichnet.

Typische Verteilungsfamilien für die Mischungskomponenten sind Normal-, Exponential- oder Poissonverteilungen oder allgemein Exponentialfamilien.

Meistens gehören die Mischungskomponenten der gleichen Verteilungsfamilie an und ψ ist ein unbekannter Parametervektor, den man schätzen möchte. Bei einer i.i.d.-Stichprobe $Y_1, \ldots, Y_n$ geht das konzeptionell mit Likelihood-Inferenz (und dem EM-Algorithmus) oder mit Bayes-Inferenz. Auch die Anzahl K der Mischungskomponenten kann unbekannt sein oder man ist an Schätzwerten für die latenten Gruppenindikatoren S_i, $i = 1, \ldots, n$ interessiert (“Clusteranalyse„). Details findet man in **(author?)** (Frühwirth-Schnatter 2006).

Momente finiter Mischverteilungen

Sei $h(Y)$ eine Transformation von Y, z.B. $h(Y) = Y$ oder $h(Y) = (Y - \mathbb{E}(Y))^2$. Dann gilt wegen $f(y \,|\, \psi) = \sum_{k=1}^{K} \eta_k \, f_k(y \,|\, \theta_k)$

$$\mathbb{E}(h(Y) \,|\, \psi) = \sum_{k=1}^{K} \eta_k \, \mathbb{E}(h(Y) \,|\, \theta_k) \, ,$$

wobei $\mathbb{E}(h(Y) \,|\, \theta_k)$ den Erwartungswert von $h(Y)$ bzgl. der Verteilung von $Y \,|\, \theta_k$ bezeichnet, d.h. der Erwartungswert von $h(Y)$ in der k-ten Gruppe.

Für $h(Y) = Y$ bzw. $h(Y) = (Y - \mathbb{E}(Y))^2$ erhält man mit $\mu_k := \mathbb{E}(Y \,|\, \theta_k)$ und $\sigma_k^2 := \mathrm{Var}(Y \,|\, \theta_k)$

$$\mu := \mathbb{E}(Y \,|\, \psi) = \sum_{k=1}^{K} \eta_k \, \mu_k \tag{13.3}$$

bzw.

$$\sigma^2 := \mathrm{Var}(Y \,|\, \psi) = \sum_{k=1}^{K} \eta_k \, \underbrace{(\mu_k^2 + \sigma_k^2)}_{=\mathbb{E}(Y^2 \,|\, \theta_k)} - \mu^2 \, . \tag{13.4}$$

Finite Mischung von Poisson-Verteilungen

$$Y \sim \eta_1 \; \mathrm{Po}(\lambda_1) + \cdots + \eta_K \; \mathrm{Po}(\lambda_K)$$

$$\psi = (\eta_1, \ldots, \eta_K, \lambda_1, \ldots, \lambda_K)$$

Diese Mischverteilung kann benutzt werden um Überdispersion in Poisson-Modellen zu berücksichtigen. Bei $Y_0 \sim \mathrm{Po}(\lambda)$ gilt nämlich $\mathbb{E}(Y_0) = \lambda = \mathrm{Var}(Y_0)$, jedoch zeigen empirische Zähldaten oft eine Überdispersion, d.h.

$$S^2 = \widehat{\mathrm{Var}}(Y_0) > \widehat{\mathbb{E}}(Y_0) = \bar{Y}$$

Für die Poisson-Mischung gilt mit (13.3) und (13.4) ($\mu_k = \lambda_k$ und $\sigma_k^2 = \lambda_k$)

$$\begin{aligned} \mathbb{E}(Y \,|\, \psi) &= \sum_{k=1}^{K} \eta_k \lambda_k \\ \mathrm{Var}(Y \,|\, \psi) &= \sum_{k=1}^{K} \eta_k \lambda_k (1 + \lambda_k) - [\mathbb{E}(Y \,|\, \psi)]^2 \, . \end{aligned}$$

Die Varianz lässt sich auch Umformen zu

$$\operatorname{Var}(Y \mid \psi) = \mathbb{E}(Y \mid \psi) + B(\psi) ,$$

wobei

$$B(\psi) = \operatorname{Var}(Y \mid \psi) - \mathbb{E}(Y \mid \psi) = \sum_{k=1}^{K} \eta_k \left(\lambda_k - \mathbb{E}(Y \mid \psi)\right)^2$$

ein Maß für die Heterogenität zwischen den Gruppen ist.

Als "Spezialfall„einer Poisson-Mischung ($\lambda_1 = 0$) erhält man das Modell für "Exzess-Nullen„

$$Y \sim \eta\, \delta_0 + (1-\eta)\ \operatorname{Po}(\lambda_2)$$

mit der Einpunktverteilung δ_0 (Dirac-Maß in 0, vgl. Abschnitt 4.1), sodass

$$P(Y = y) = \begin{cases} \eta + (1-\eta)\, e^{-\lambda_2} & \text{für } y = 0 , \\ (1-\eta)\, \frac{\lambda_2{}^y}{y!}\, e^{-\lambda_2} & \text{sonst } (y \in \mathbb{N}) . \end{cases}$$

Mischverteilungen als Priori-Verteilungen in hierarchischen Bayes-Modellen

(a) Mischung von Beta-Verteilungen als Priori-Verteilung für π im Bernoulli-Experiment (Bernardo und Smith 1994):

$$X_1 \mid \pi, \ldots, X_n \mid \pi \overset{i.i.d}{\sim} \operatorname{Bin}(1, \pi)$$

Die konjugierte Priori-Verteilung ist die Beta-Verteilung $\operatorname{Be}(\alpha, \beta)$ (vgl. Abschnitt 13.5.1) und mit $Z := \sum_{i=1}^{n} X_i$ ergibt sich die Posteriori-Verteilung

$$\pi \mid X_1 = x_1, \ldots, X_n = x_n, \alpha, \beta \sim \operatorname{Be}(\alpha + z,\, \beta + n - z) .$$

In folgendem Beispiel wird aber eine multimodale Priori benötigt:

> Eine Münze wird nicht geworfen, sondern wie ein Kreisel gedreht, und fällt dann auf eine Seite. Die Erfahrung (Vorwissen) zeigt, dass die Chance für Kopf für manche Münzen nicht 1:1, sondern 2:1 oder 1:2 zu sein scheint.

Ein mögliche Priori für die Kopf-Wahrscheinlichkeit ist eine Mischung von Beta-Verteilungen:

$$\pi \,|\, (\eta_k, \alpha_k, \beta_k)_{k=1,\ldots,K} \sim \sum_{k=1}^{K} \eta_k \; \mathrm{Be}(\alpha_k, \beta_k)$$

Jede Mischungskomponente repräsentiert einen der Münztypen, z.B. könnte mit der Priori-Mischverteilung

$$0.5 \; \mathrm{Be}(10, 20) + 0.2 \; \mathrm{Be}(15, 15) + 0.3 \; \mathrm{Be}(20, 10)$$

das Vorwissen ausgedrückt werden, dass etwa 20% der Münzen symmetrisch sind, etwa 50% eine 1:2 Chance und etwa 30% eine 2:1 Chance für Kopf besitzen. Für die Posteriori-Verteilung lässt sich zeigen:

$$\pi \,|\, x_1, \ldots, x_n, (\eta_k^*, \alpha_k^*, \beta_k^*)_{k=1,\ldots,K} \sim \sum_{k=1}^{K} \eta_k^* \; \mathrm{Be}(\alpha_k^*, \beta_k^*) \,,$$

mit $\alpha_k^* = \alpha_k + z$, $\beta_k^* = \beta_k + n - z$ und

$$\eta_k^* \propto \eta_k \, \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\,\Gamma(\beta_k)} \, \frac{\Gamma(\alpha_k^*)\,\Gamma(\beta_k^*)}{\Gamma(\alpha_k^* + \beta_k^*)} \qquad (\eta_1^* + \cdots + \eta_K^* = 1) \,.$$

(b) Bayesianische Variablenselektion im Linearen Modell:

$$y_i = \beta_0 + \beta_1 \, x_{i1} + \cdots + \beta_p \, x_{ip} + \varepsilon_i \qquad \varepsilon_i \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma^2)$$

Der Einfachheit halber sei σ^2 bekannt. Eine übliche Priori-Annahme ist

$$\beta_j \overset{ind}{\sim} \mathrm{N}(0; k_j \, \sigma^2) \qquad j = 0, \ldots, p \,.$$

Eine mögliche Priori-Verteilung zur Variablenselektion ist

$$\beta_j \overset{ind}{\sim} \eta_1 \, \delta_0 + \eta_2 \; \mathrm{N}(0; k_j \, \sigma^2) \qquad j = 0, \ldots, p \,,$$

also eine diskret-stetige Mischverteilung (vgl. 7.5: "Gemischte Verteilungen„), wobei $\eta_1 = P(\beta_j = 0)$ und $\eta_2 = 1 - \eta_1 = P(\beta_j \neq 0)$, d.h. es gibt eine Punktmasse auf der Null. Dies kann auch mit Hilfe von latenten Indikatorvariablen $\gamma_j \overset{i.i.d}{\sim} \mathrm{Bin}(1, \eta_1)$ formuliert werden.

Ein anderes Beispiel für eine Bayesianische Schrumpfungs-Priori bzw. Variablenselektion ist das “Spike and Slab„-Modell (Ishwaran und Rao 2005).

13.6.2. Stetige Mischungen

Sei $f(x \mid \theta)$ die Dichte einer (stetigen oder diskreten) Zufallsvariable X gegeben der Parameter θ und evtl. weitere Parameter. Sei θ eine stetige Zufallsvariable mit Dichte $w(\theta)$ (“Priori„), die evtl. von weiteren (Hyper-)Parametern abhängt. Dann ist die marginale Dichte

$$f(x) = \int_\Theta f(x \mid \theta)\, w(\theta)\, \mathrm{d}\theta$$

eine *stetige Mischung*.

Es folgen einige Beispiele für stetige Mischungen.

Binomial-Beta-Verteilung

Die Binomial-Beta-Verteilung $\mathrm{Bb}(a, b, n)$ wird generiert durch die stetige Mischung der Binomial-Verteilung $\mathrm{Bin}(n, \pi)$ mit dem $\mathrm{Be}(a, b)$-verteilten Parameter π. Ihre Dichte ist also (in symbolischer Schreibweise)

$$\mathrm{Bb}(x \mid a, b, n) = \int_0^1 \mathrm{Bin}(x \mid n, \pi)\ \mathrm{Be}(\pi \mid a, b)\, \mathrm{d}\pi\,.$$

Als Ergebnis erhält man

$$\mathrm{Bb}(x \mid a, b, n) = c \binom{n}{x} \Gamma(a + x)\, \Gamma(b + n - x)\, I_{\{0;1;\ldots;n\}}(x)\,,$$

mit

$$c = \frac{\Gamma(a + b)}{\Gamma(a)\, \Gamma(b)\, \Gamma(a + b + n)}\,.$$

Erwartungswert und Varianz der Binomial-Beta-Verteilung ergeben sich zu

$$\begin{aligned} \mathbb{E}(X) &= n\, \frac{a}{a + b} \\ \mathrm{Var}(X) &= \frac{nab(a + b + n)}{(a + b)^2\, (a + b + 1)} \end{aligned}$$

Für $a = 1 = b$ erhält man die diskrete Gleichverteilung auf $\{0; 1; \ldots; n\}$.

Poisson-Gamma-Verteilung

Die Poisson-Gamma-Verteilung $\mathrm{Pg}(a, b, \nu)$ entsteht durch die Mischung der $\mathrm{Po}(\nu\lambda)$-Verteilung über $\lambda \sim \mathrm{Ga}(a, b)$:

$$\begin{aligned} \mathrm{Pg}(x \mid a, b, \nu) &= \int_0^\infty \mathrm{Po}(x \mid \nu\lambda)\ \mathrm{Ga}(\lambda \mid a, b)\, \mathrm{d}\lambda \\ &= c\,\frac{\Gamma(a+x)}{x!}\,\frac{\nu^x}{(b+\nu)^{a+x}}\,, \quad \text{für } x \in \mathbb{N}_0\,, \end{aligned}$$

mit $c = \frac{b^a}{\Gamma(a)}$. Dabei ist ν ein sogenannter "offset„-Parameter und λ wird herausintegriert. Als Spezialfall für $a \in \mathbb{N}$ erhält man eine Negativ-Binomial-Verteilung $\mathrm{Nb}(x \mid a, \frac{b}{b+\nu})$.

Erwartungswert und Varianz der Poisson-Gamma-Verteilung sind

$$\begin{aligned} \mathbb{E}(X) &= \nu\,\frac{a}{b} \\ \mathrm{Var}(X) &= \frac{\nu a(b+\nu)}{b^2} \end{aligned}$$

d.h. Überdispersion ist hier möglich (im Gegensatz zur Poisson-Verteilung).

Skalenmischung von Normalverteilungen

Man erhält eine Student-(t-)Verteilung $\mathrm{St}(\mu, \lambda, \alpha)$, wenn man die Normalverteilung $\mathrm{N}(\mu, \lambda\,\sigma^2)$ durch $\sigma^2 \sim \mathrm{Ga}(\frac{\alpha}{2}, \frac{\alpha}{2})$ mischt. ($\rightsquigarrow$ Übung)

Negativ-Binomial-Beta-Verteilung

Die Negativ-Binomial-Beta-Verteilung $\mathrm{Nbb}(a, b, r)$ wird generiert durch die stetige Mischung der Negativ-Binomial-Verteilung $\mathrm{Nb}(\pi, r)$ durch den Parameter $\pi \sim \mathrm{Be}(a, b)$:

$$\begin{aligned} \mathrm{Nbb}(x \mid a, b, r) &= \int_0^1 \mathrm{Nb}(x \mid \pi, r)\ \mathrm{Be}(\pi \mid a, b)\, \mathrm{d}\pi \\ &= c\binom{r+x-1}{r-1}\frac{\Gamma(b+x)}{\Gamma(a+b+x+r)}\,, \quad \text{für } x \in \mathbb{N}_0\,, \end{aligned}$$

wobei

$$c = \frac{\Gamma(a+b)\,\Gamma(a+r)}{\Gamma(a)\,\Gamma(b)}\,.$$

14. Fourier-Transformationen, Charakteristische Funktionen

14.1. Fourier-Transformation eines Wahrscheinlichkeitsmaßes

14.1.1. Vorbereitung

Es wird auf folgende Begriffe und Zusammenhänge hingewiesen:

a) Sind $x = (x_1, \ldots, x_p)' \in \mathbb{R}^p$ und $y = (y_1, \ldots, y_p)' \in \mathbb{R}^p$, so bezeichnet $\langle x, y \rangle = \sum_{i=1}^{p} x_i y_i$ das *Skalarprodukt* von x und y.

b) Sei $f : \Omega \to \mathbb{C}$ eine komplexwertige Funktion der Form

$$f(\omega) = u(\omega) + i\, v(\omega)$$

mit rellen Funktionen u und v, sowie $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt

$$\int f \,\mathrm{d}\mu := \int u \,\mathrm{d}\mu + i \int v \,\mathrm{d}\mu$$

wenn die Integrale auf der rechten Seite existieren (vgl. Satz 6.4 [1] und [2]).

c) Es gilt in Verallgemeinerung von Satz 6.4 [6]: Eine komplexwertige, messbare Funktion f auf Ω ist genau dann μ-integrierbar, wenn $|f|$ μ-integrierbar ist.

d) Für $f(\omega) = u(\omega) + i\, v(\omega)$ wird $\bar{f}(\omega) := u(\omega) - i\, v(\omega)$ als *konjugiert komplexe Funktion* bezeichnet.

14.1.2. Definition (Fourier-Transformierte)

a) Sei P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$. Dann heißt die komplexwertige Funktion

$$\hat{P}(x) := \int e^{i\,\langle x,y\rangle}\, P(\mathrm{d}y) \qquad (x \in \mathbb{R}^p)$$

Fourier-Transformierte von P.

b) Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, also $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^p, \mathfrak{B}^{\otimes p})$, und P_X deren Verteilung. Dann heißt die Fourier-Transformierte

$$\hat{P}_X(x) := \int e^{i\,\langle x,y\rangle} P_X(\mathrm{d}y) = \mathbb{E}(e^{i\,\langle x,X\rangle})$$

auch *charakteristische Funktion* von X und wird auch mit $\varphi_X(x)$ bezeichnet.

Hinweis:

Wegen $\left|e^{i\,t}\right| = 1$ für alle $t \in \mathbb{R}$ sind die Integrabilitätsbedingungen erfüllt.

14.1.3. Beispiele

1. Die charakteristische Funktion der Einpunktverteilung in $a \in \mathbb{R}^p$

$$\varepsilon_a(x) = \begin{cases} 1 & x = a \\ 0 & \text{sonst} \end{cases}$$

ist

$$\hat{\varepsilon}_a(x) = e^{i\,\langle x,a\rangle}\ .$$

2. Die charakteristische Funktion einer diskreten Verteilung

$$\mu = \sum_{n=1}^{\infty} \alpha_n\, \varepsilon_{a_n}$$

ist

$$\hat{\mu}(x) = \sum_{n=1}^{\infty} \alpha_n e^{i\,\langle x,a_n\rangle}$$

Beweis: $\rightsquigarrow$ Übung. $\square$

14.1.4. Rechenregeln

Seien μ und ν Wahrscheinlichkeitsmaße auf $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$. Dann gilt:

a) Ist $\mu * \nu$ das Faltungsprodukt, dann gilt

$$\widehat{\mu * \nu} = \hat{\mu} \cdot \hat{\nu} \qquad \text{das heißt} \qquad \widehat{\mu * \nu}\,(x) = \hat{\mu}(x) \cdot \hat{\nu}(x)$$

b) Ist T eine lineare Abbildung auf $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$ und T^t deren Transponierte, dann gilt

$$\widehat{T(\mu)} = \hat{\mu} \circ T^t$$

c) Ist S eine Spiegelung am Nullpunkt, d.h. $S(x) = -x$, dann gilt

$$\widehat{S(\mu)} = \hat{\mu} \circ S = \bar{\hat{\mu}}$$

d) Für jede Translation $T_a(x) = a + x$ $(a, x \in \mathbb{R}^p)$ gilt

$$\widehat{T_a(\mu)} = \hat{\varepsilon}_a \cdot \hat{\mu}$$

14.1.5. Satz

Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$ und ν ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^q, \mathfrak{B}^{\otimes q})$, dann gilt für das Produktmaß $\mu \otimes \nu$ auf $(\mathbb{R}^{p+q}, \mathfrak{B}^{\otimes p} \otimes \mathfrak{B}^{\otimes q})$

$$\widehat{\mu \otimes \nu}\,(x, y) = \hat{\mu}(x) \cdot \hat{\nu}(y)\,.$$

14.1.6. Folgerungen

a) Insbesondere gilt für die charakteristische Funktion der Summe von n i.i.d. Zufallsvariablen $X_1, \ldots, X_n$

$$\varphi_{X_1 + \cdots + X_n} = \varphi_{X_1} \cdots \varphi_{X_n}$$

b) Für jede $\mathbb{R}^p$-wertige Zufallsvariable X und jede lineare Abbildung T von $\mathbb{R}^p$ in sich mit Transformierter T^t sowie für jedes $a \in \mathbb{R}^p$ gilt

$$\varphi_{a + T \circ X}(x) = e^{i\,\langle x, a \rangle} \cdot \varphi_X(T^t(x))$$

c) Aus b) folgt

$$\varphi_{-X}(x) = \varphi_X(-x) = \overline{\varphi_X}(x)$$

14.2. Fourier-Transformierte für Lebesgue-integrierbare Funktionen

Oft sind Wahrscheinlichkeitsmaße darstellbar durch eine Lebesgue-Dichte f: $\nu(A) = \int_A f(\omega)\, \lambda^p(\mathrm{d}\omega)$, $A \in \mathcal{A}$. Für solche Maße lässt sich eine Fourier-Transformierte direkt für die Dichte f angeben.

14.2.1. Definition

Sei $f : \mathbb{R}^p \to \mathbb{C}$ eine λ^p-integrierbare Funktion. Dann heißt die durch

$$\hat{f}(x) := \int e^{i\,\langle x,y\rangle}\, f(y)\, \lambda^p(\mathrm{d}y)$$

definierte Funktion $\hat{f} : \mathbb{R}^p \to \mathbb{C}$ die *Fourier-Transfomierte* von f.

Hinweis:

Ist $f = u + i\,v$, wobei $u = u^+ - u^-$ und $v = v^+ - v^-$, dann gilt

$$\hat{f} = \widehat{u^+} - \widehat{u^-} + i\,(\widehat{v^+} - \widehat{v^-})$$

14.2.2. Satz

Für je zwei Lebesgue-integrierbare komplexwertige Funktionen f und g gilt:

a) $\hat{f}$ ist gleichmäßig stetig.

b) Die Abbildung $f \mapsto \hat{f}$, die jeder Dichte ihre Fourier-Transformierte zuordnet, ist eine lineare Abbildung über $\mathbb{C}$.

c) $\widehat{f * g} = \hat{f} \cdot \hat{g}$.

d) Mit $f^*(x) = \bar{f}(-x)$ ist $\bar{\hat{f}} = \widehat{f^*}$.

e) Die Funktion $h(x,y) = f(x) \cdot g(y)$ ist λ^{p+q}-integrierbar, falls f λ^p-integrierbar und g λ^q-integrierbar ist. Dann gilt: $\hat{h}(x,y) = \hat{f}(x) \cdot \hat{g}(y)$.

14.2.3. Beispiele

1. Für die Dichte f einer $\mathrm{N}(\mu, \sigma^2)$-Verteilung lautet die Fourier-Transformierte
$$\hat{f}(x) = e^{i\mu x - \sigma^2 \frac{x^2}{2}}$$

2. Für die Dichte
$$f(y) = \frac{1}{\pi} \frac{1}{1+y^2}$$
der Standard-Cauchy-Verteilung $\mathrm{C}(0,1)$ lautet die Fourier-Transfomierte
$$\hat{f}(x) = e^{-|x|}$$

Frage: Wie lautet die Fourier-Transformierte der Dichte f_α $(\alpha > 0)$, einer $C(0, \alpha)$-Verteilung?

14.2.4. Satz (Eindeutigkeitssatz für Fourier-Transformierte)

Sind $\hat{\mu}$ und $\hat{\nu}$ Fourier-Transfomierte von zwei Wahrscheinlichkeitsmaßen μ und ν, dann gilt:
$$\hat{\mu} = \hat{\nu} \quad \Longrightarrow \quad \mu = \nu$$

Hieraus folgt der Eindeutigkeitssatz für Fourier-Transformierte $\hat{f}$ und $\hat{g}$ von Dichten f und g bzgl. λ^p:
$$\hat{f} = \hat{g} \quad \Longrightarrow \quad f = g \quad (\lambda^p - \text{fast überall})$$

14.3. Fourier-Transformierte und Momente

14.3.1. Lemma

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, U eine offene Teilmenge von $\mathbb{R}$ bzw. $\mathbb{C}$ sowie $f : U \times \Omega \to \mathbb{C}$ eine Funktion mit folgenden Eigenschaften:

(i) $\omega \mapsto f(t, \omega)$ ist μ-integrierbar für alle $t \in U$

(ii) $t \mapsto f(t,\omega)$ ist differenzierbar in t_0 für alle $\omega \in \Omega$

(iii) Es gibt eine μ-integrierbare Funktion $h \geq 0$ auf Ω mit

$$\left|\frac{f(t,\omega) - f(t_0,\omega)}{t - t_0}\right| \leq h(\omega)$$

für alle $t \in U$

Dann ist die Funktion

$$\begin{aligned} \varphi: \quad U &\longrightarrow \mathbb{C} \\ t &\longmapsto \int f(t,\omega)\, \mu(\mathrm{d}\omega) \end{aligned}$$

differenzierbar in t_0 und die durch $\omega \mapsto f'(t_0,\omega)$ definierte Funktion ist μ-integrierbar, sodass $\varphi'(t_0) = \int f'(t_0,\omega)\, \mu(\mathrm{d}\omega)$.

14.3.2. Definition

Sei μ ein (nicht notwendigerweise endliches) Borel-Maß[1] auf $\mathbb{R}^p$ und seien $k_1, \ldots, k_p \in \mathbb{N}_0$. Ist dann

$$\begin{aligned} \mathbb{R}^p &\longrightarrow \mathbb{R} \\ x = (x_1, \ldots, x_p) &\longmapsto x_1^{k_1} \cdots x_p^{k_p} \end{aligned}$$

μ-integrierbar, so heißt

$$M_{k_1,\ldots,k_p} := \int x_1^{k_1} \cdots x_p^{k_p}\, \mu(\mathrm{d}x)$$

das $(k_1, \ldots, k_p)$-*te Moment* von μ und $\sum_{i=1}^p k_i$ dessen *Ordnung*. Man sagt auch, “das $(k_1, \ldots, k_p)$-te Moment existiert„.

Hinweis

Sind $X_1, \ldots, X_p$ reelle Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, mit der gemeinsamen Verteilung $\mu := P_{X_1} \otimes \cdots \otimes P_{X_p}$, dann existiert das $(k_1, \ldots, k_p)$-te Moment $M_{k_1,\ldots,k_p}$ von μ genau dann, wenn die

[1] d.h. ein Maß auf dem mit der Borel-σ-Algebra versehenen Messraum $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$

Zufallsvariable $X_1^{k_1} \cdots X_p^{k_p}$ P-integrierbar ist. Dann ist

$$M_{k_1,\ldots,k_p} = \mathbb{E}(X_1^{k_1} \cdots X_p^{k_p}) \,.$$

14.3.3. Satz

Sei μ ein W'maß auf $(\mathbb{R}^p, \mathfrak{B}^{\otimes p})$, für welches neben dem $(k_1, \ldots, k_p)$-ten Moment alle Momente $M_{l_1,\ldots,l_p}$ für $0 \leq l_j \leq k_j$, $j = 1, \ldots, p$ existieren. Für alle derartigen $l_1, \ldots, l_p$ existiert dann die partielle Ableitung

$$D_{l_1,\ldots,l_p}\,\hat{\mu} := \frac{\partial^{l_1+\cdots+l_p}}{\partial^{l_1} x_1 \cdots \partial^{l_p} x_p} \hat{\mu}$$

der Fourier-Transformierten auf ganz $\mathbb{R}^p$. Es gilt

$$D_{l_1,\ldots,l_p}\,\hat{\mu}(x) = i^{l_1+\cdots+l_p} \int e^{i\,\langle x,y \rangle}\, y_1^{l_1} \cdots y_p^{l_p}\; \mu(\mathrm{d}y)$$

und insbesondere ist

$$D_{l_1,\ldots,l_p}\,\hat{\mu}(0) = i^{l_1+\cdots+l_p} M_{l_1,\ldots,l_p} \,.$$

Jede dieser Abbildungen ist auf $\mathbb{R}^p$ gleichmäßig stetig und beschränkt.

14.3.4. Korollar

Existiert für ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$ das Moment M_k k-ter Ordnung, so ist $\hat{\mu}$ k-mal differenzierbar. Die k-te Ableitung ist gleichmäßig stetig und beschränkt und es gilt

$$\hat{\mu}^{(k)}(0) = i^k M_k \,.$$

Hinweis

Besitzt die Fourier-Transformierte $\hat{\mu}$ eines Wahrscheinlichkeitsmaßes auf $(\mathbb{R}, \mathfrak{B})$ ein k-tes Moment – und damit alle Momente M_l, $0 < l \leq k$ – dann besitzt $\hat{\mu}$

die Taylor-Darstellung

$$\hat{\mu}(x) = \sum_{l=0}^{k-1} \frac{(ix)^l}{l!} \, M_l + \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} \, \hat{\mu}^{(k)}(t) \, \mathrm{d}t + R_k(x)$$

mit der Restgliedabschätzung

$$|R_k(x)| \leq \sup_{0 \leq \theta \leq 1} \left| \hat{\mu}^{(k)}(\theta x) - \hat{\mu}^{(k)}(0) \right| \cdot \frac{|x|^k}{k!} \, .$$

A. Wiederholung: Komplexe Zahlen

Komplexe Zahlen sind Gegenstand der Analysis-Vorlesungen. Hier werden nur einige Zusammenhänge und Rechenregeln wiederholt, die z.B. für das Arbeiten mit Fourier-Transformationen von Bedeutung sind. Die Darstellung orientiert sich an **(author?)** (Walter 2004, Kapitel 8).

A.1. Definitionen

Auf $\mathbb{R}^2$ wird für $\alpha = (a, b)$ und $\beta = (c, d)$ eine Summe und ein Produkt wie folgt definiert:

$$\alpha + \beta = (a, b) + (c, d) \quad := \quad (a + c, b + d) \tag{A.1}$$

$$\alpha \cdot \beta = (a, b) \cdot (c, d) \quad := \quad (ac - bd, ad + bc) \tag{A.2}$$

Mit diesen Definitionen ist $\mathbb{R}^2$ ein Körper und wird mit $\mathbb{C}$ bezeichnet. $\mathbb{C}$ wird auch *Körper der komplexen Zahlen* genannt.

Mit der Abbildung

$$\begin{aligned} \mathbb{R} &\to \mathbb{C} \\ a &\mapsto (a, 0) \end{aligned}$$

wird $\mathbb{R}$ in $\mathbb{C}$ eingebettet und ist somit als Teilkörper von $\mathbb{C}$ zu betrachten.

Das *Nullelement* (neutrales Element bzgl “+„) von $\mathbb{C}$ ist $(0, 0) =: 0$, das *Einselement* (neutrales Element bzgl “$\cdot$„) von $\mathbb{C}$ ist $(1, 0)$.

Das Element $(0, 1)$ wird mit i bezeichnet und heißt auch *imaginäre Einheit*. Es gilt:

$$i^2 = i \cdot i = (0, 1) \cdot (0, 1) \overset{(A.2)}{=} (-1, 0) = -1 \in \mathbb{R}\,.$$

Wegen

$$\alpha = (a, b) = (a, 0) + (0, 1) \cdot (b, 0)$$

kann man dies auch als

$$\alpha = a + i\,b \qquad a, b \in \mathbb{R}$$

schreiben. Diese Darstellung von komplexen Zahlen vereinfacht das Rechnen in $\mathbb{C}$. Hierbei stellt man sich i als Einheitsvektor in Richtung der y-Achse vor.

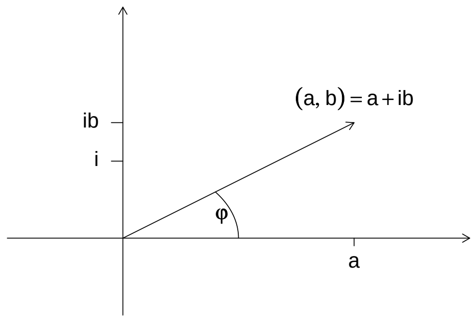

Abbildung A.1.: Darstellung von komplexen Zahlen

Man nennt für $\alpha = a + i\,b$

$$\begin{array}{rl} a & \textit{Realteil} \text{ von } \alpha,\ \mathfrak{Re}\,\alpha \\ b & \textit{Imaginärteil} \text{ von } \alpha,\ \mathfrak{Im}\,\alpha \\ \bar{\alpha} := a - i\,b & \text{zu } \alpha \textit{ konjugierte komplexe Zahl} \\ |\alpha| := \sqrt{a^2 + b^2} & \text{(absoluten) } \textit{Betrag} \text{ von } \alpha \end{array}$$

Außerdem wird das Inverse einer komplexen Zahl $\alpha \neq 0$ definiert durch

$$\alpha^{-1} := \frac{1}{\alpha} := \frac{a - i\,b}{a^2 + b^2} = \frac{\bar{\alpha}}{|\alpha|^2} \tag{A.3}$$

A.2. Rechenregeln

Es seien $\alpha, \beta \in \mathbb{C}$.

$$\mathfrak{Re}\,\alpha = \frac{1}{2}\,(\alpha + \bar{\alpha}) \qquad \mathfrak{Im}\,\alpha = \frac{1}{2\,i}\,(\alpha - \bar{\alpha}) \qquad \mathfrak{Re}(i\,\alpha) = -\mathfrak{Im}\,\alpha \tag{A.4}$$

$$|\alpha| = |\bar{\alpha}| = \sqrt{\alpha\bar{\alpha}} \qquad \bar{\bar{\alpha}} = \alpha \tag{A.5}$$

$$\overline{\alpha \pm \beta} = \bar{\alpha} \pm \bar{\beta} \qquad \overline{\alpha \cdot \beta} = \bar{\alpha} \cdot \bar{\beta} \qquad \overline{\left(\frac{\alpha}{\beta}\right)} = \frac{\bar{\alpha}}{\bar{\beta}} \quad (\beta \neq 0) \tag{A.6}$$

$$|\alpha \cdot \beta| = |\alpha| \cdot |\beta| \qquad \left|\frac{\alpha}{\beta}\right| = \frac{|\alpha|}{|\beta|} \quad (\beta \neq 0) \tag{A.7}$$

$$|\alpha + \beta| \leq |\alpha| + |\beta| \quad \text{(Dreiecksungleichung)} \tag{A.8}$$

A.3. Polarkoordinaten

Es ist bekannt, dass sich Elemente $(a, b) \in \mathbb{R}^2$ auch durch ihre Polarkoordinaten darstellen lassen. Liegt (a, b) auf dem Einheitskreis, dann gibt es genau ein $\varphi \in [0; 2\pi)$ mit $a = \cos\varphi$ und $b = \sin\varphi$. Hierbei ist φ der im Bogenmaß gemessene Winkel zwischen dem Strahl der vom Nullpunkt aus durch (a, b) geht und der positiven reellen Achse (vgl. Abbildung A.1).

Für eine komplexe Zahl $\alpha = a + i\,b \neq 0$ liegt $\frac{\alpha}{|\alpha|}$ auf dem Einheitskreis, womit sie eine eindeutige Darstellung als

$$\alpha = r\,(\cos\varphi + i\,\sin\varphi) \qquad r := |\alpha| > 0\,,\; 0 \leq \varphi < 2\pi$$

besitzt. φ wird auch das *Argument* von α genannt, in Zeichen $\varphi = \arg\alpha$. Üblicherweise setzt man $\arg 0 := 0$. (Auch die Normierung $-\pi < \arg\alpha \leq \pi$ ist üblich.)

Aufgrund der Additionstheoreme für trigonometrische Funktionen gestaltet sich die Multiplikation komplexer Zahlen unter Verwendung der Polarkoordinaten-Darstellung (Polarform) einfach:

$$\left.\begin{array}{l} \alpha = r\,(\cos\varphi + i\,\sin\varphi) \\ \beta = s\,(\cos\psi + i\,\sin\psi) \end{array}\right\} \quad \alpha \cdot \beta = rs\,(\cos(\varphi + \psi) + i\,\sin(\varphi + \psi)) \tag{A.9}$$

Daraus folgt für die Division ($\beta \neq 0$ bzw. $s > 0$) wegen $\arg \bar{\beta} = -\arg \beta$:

$$\begin{aligned}
\frac{\alpha}{\beta} &= \alpha \cdot \beta^{-1} = \alpha \cdot \frac{\bar{\beta}}{|\beta|^2} = \frac{1}{|\beta|^2}\, rs\,(\cos(\varphi - \psi) + i\, \sin(\varphi - \psi)) = \\
&= \frac{r}{s}\,(\cos(\varphi - \psi) + i\, \sin(\varphi - \psi))
\end{aligned} \tag{A.10}$$

A.4. Eulersche Identität

Die Eulersche Identität stellt eine Verknüpfung zwischen trigonometrischen Funktionen und komplexen Zahlen her. Für eine komplexe Zahl $z \in \mathbb{C}$ gilt

$$e^{iz} = \cos z + i\, \sin z \tag{A.11}$$

und damit lassen sich die trigonometrischen Funktionen auch als Linearkombinationen von imaginären Exponentialfunktionen darstellen:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2\, i} \tag{A.12}$$

Die Formeln gelten natürlich ebenso für reelles z.

Wegen der Eulerschen Identität lässt sich $\alpha \in \mathbb{C}$ sowohl in Polarform als auch in Exponentialform darstellen:

$$\alpha = r\,(\cos \varphi + i\, \sin \varphi) = r\, e^{i\,\varphi} \tag{A.13}$$

Mit (A.9) und (A.10) stellen sich Multiplikation und Division komplexer Zahlen in Exponentialform wie folgt dar:

$$\left.\begin{aligned} \alpha &= r\, e^{i\varphi} \\ \beta &= s\, e^{i\psi} \end{aligned}\right\} \quad \alpha \cdot \beta = rs\, e^{i(\varphi+\psi)} \quad \text{und} \quad \frac{\alpha}{\beta} = \frac{r}{s}\, e^{i(\varphi-\psi)} \quad (\beta \neq 0) \tag{A.14}$$

Literaturhinweise

[Bauer 2001] Bauer, Heinz: *Wahrscheinlichkeitstheorie*. 5., durchges. und verb. Aufl. Berlin : de Gruyter, 2001 (de Gruyter Lehrbuch). – ISBN 3–11–017236–4

[Bernardo und Smith 1994] Bernardo, José M. ; Smith, Adrian F.: *Bayesian Theory*. 1. Auflage. Chichester : John Wiley & Sons, März 1994 (Wiley Series in Probability and Statistics). – ISBN 0–471–92416–4

[Frühwirth-Schnatter 2006] Frühwirth-Schnatter, Sylvia: *Finite Mixture and Markov Switching Models*. 1. Auflage. New York : Springer, August 2006 (Springer Series in Statistics). – ISBN 0–387–32909–9

[Ishwaran und Rao 2005] Ishwaran, Hemant ; Rao, J. S.: Spike and Slab Gene Selection for Multigroup Microarray Data. In: *Journal of the American Statistical Association* 100 (2005), September, Nr. 471, S. 764–780. – ISSN 0162–1459

[Walter 2004] Walter, Wolfgang: *Analysis 1*. 7. Auflage. Berlin, Heidelberg : Springer-Verlag, 2004 (Springer-Lehrbuch). – ISBN 3–540–20388–5