

Angewandte Stochastik

Dr. Matthias Birkner

Vorlesung
Sommersemester 2008

Felix Hoffmann
Universität Bonn

22. Juli 2009

Vorwort

Dieses Werk ist unter einem Creative Commons Namensnennung-NichtKommerziell-Weitergabe unter gleichen Bedingungen-Lizenzvertrag lizenziert. Die vollständige Lizenz ist einzusehen unter:

<http://creativecommons.org/licenses/by-nc-sa/2.0/de/deed.de>



Dieses Skript wurde von FELIX HOFFMANN während der Vorlesung Angewandte Stochastik von MATTHIAS BIRKNER im Sommersemester 2008 an der Rheinischen Friedrich-Wilhelms Universität Bonn mitgeschrieben und anschließend von Matthias Birkner durchgesehen und ergänzt.

Bei Fragen, Wünschen, Korrektur- oder Verbesserungsvorschlägen freuen wir uns über jede E-Mail an

`birkner(AT)wias-berlin.de,`

`felix.hoffmann(AT)iam.uni-bonn.de`

Inhaltsverzeichnis

I. Statistik	1
1. Einführung in die Statistik	3
1.0.1. Beispiel:	3
1.1. Definition und Sprechweise: statistisches Modell	3
1.2. Definition und Sprechweise: parametrisches Modell	4
1.2.1. Bemerkung: Uniforme Verteilung auf einer Cantormenge	4
1.3. Definition: Produktmodell zu $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$:	4
1.4. Definiton: Statistik	4
1.5. Beispiel: „Taxiproblem“, Schätzung des Wertebereichs:	4
1.5.1. Bemerkung:	5
2. Das Maximum-Likelihood-Prinzip:	7
2.0.2. Beispiel: Qualitätskontrolle	7
2.1. Definition: Likelihood-Funktion	7
2.1.1. Beispiele:	7
2.1.2. Beispiel: Uniforme Verteilung auf $[0, \vartheta]$	8
2.1.3. Beispiel: Normales Modell	8
2.2. Sprechweisen: Erwartungstreu, Mittlerer quadratischer Fehler, Bias	9
2.3. Satz: Erwartungstreue Schätzer für Varianz und Mittelwert	9
2.3.1. Definition und Beobachtung: Konsistenz	10
3. Exponentielle Familien, Fisher-Information und beste Schätzer:	11
3.1. Definition: regulär	11
3.1.1. Bemerkung:	11
3.2. Definition: Fisher-Information	11
3.2.1. Bemerkung:	11
3.3. Definition: regulärer Schätzer	12
3.4. Satz: Cramér-Rao-Schranke	12
3.4.1. Beispiele:	13
3.5. Definition: Cramér-Rao-effizient	13
3.6. Definition: exponentielles Modell	13
3.6.1. Bemerkung:	14
3.7. Satz über die „Gutartigkeit“ von exponentiellen Modellen	14
3.8. Korollar:	16
3.8.1. Beispiel:	16
3.9. Konsistenz des ML-Schätzers im Produktmodell:	16
3.9.1. Bemerkung:	17
3.10. Satz und Definition: relative Entropie (Kullback-Leibler-Abstand)	17
3.10.1. Beispiele: Poisson- und Exponentialverteilung	18

3.11. Satz:	18
3.11.1. Bemerkung: Asymptotische Konfidenzintervalle für den ML-Schätzer:	19
3.11.2. Beobachtung:	19
4. Exkurs: Bedingte Erwartung und Anwendungen	21
4.1. Definition: Bedingte Erwartung	21
4.2. Satz: Existenz und Eindeutigkeit der bedingten Erwartung:	21
4.3. Eigenschaften der bedingten Erwartung:	22
4.3.1. Bemerkung: Zur Existenz der regulären bedingten Verteilung	23
4.4. Suffizienz:	23
4.5. Satz von Rao-Blackwell	24
4.6. Definition: Vollständigkeit	24
4.6.1. Beispiel:	24
4.6.2. Bemerkung: Eindeutigkeit der Rao-Blackwellisierung	24
4.7. Satz von Lehmann-Scheffé	24
4.7.1. Bemerkung:	25
5. Bayes-Schätzer	27
5.1. Definition: A-priori-Dichte und -Maß	27
5.2. Satz:	28
6. Konfidenzintervalle:	29
6.1. Definition: Konfidenzbereich zum Irrtumsniveau α	29
6.1.1. Beispiel: Konfidenzintervalle im Binomialmodell	29
6.2. Definition: Quantil, Fraktile und Median	30
6.2.1. Lemma:	30
6.2.2. Satz: exaktes Konfidenzintervall im Binomialmodell	32
6.3. Konfidenzintervalle im Gauß'schen Modell	32
6.3.1. Satz:	33
7. Exkurs: mehrdimensionale Normalverteilung	35
7.1. Dichtentransformationssatz:	35
7.2. Erinnerung: $\Gamma_{a,b}$ -Verteilung	35
7.2.1. Proposition:	35
7.2.2. Korollar:	36
7.2.3. Proposition: Fisher-Verteilung mit m und n Freiheitsgraden	36
7.2.4. Korollar: Student'sches T	37
7.3. Satz:	37
8. Konfidenzintervalle für den Median	39
8.1. Satz über die Konfidenzintervalle für den Median	39
8.2. Korollar: Vorzeichenstest für den Median	40
8.3. Die 2-Stichproben-Situation:	40
8.3.1. Wilcoxon-Rangsummen-Test	40
8.3.2. Proposition:	40
8.4. Satz:	42
8.5. Definition: stochastisch kleiner	43
8.5.1. Satz:	44

8.5.2.	Beispiel: Verkehrstote in Großbritannien 1969–1984	45
9.	Testen von Hypothesen:	47
9.1.	Definition: Test, Randomisierung und Gütefunktion	47
9.2.	Definition: gleichmäßig bester Test	47
9.3.	Alternativtest und das Neyman-Pearson-Lemma	47
9.3.1.	Definition: Neyman-Pearson-Test	48
9.3.2.	Satz: (Neyman-Pearson-Lemma)	48
9.4.	Satz: (Lemma von Stein)	49
9.4.1.	Beispiel:	51
10.	Monotone Likelihood-Quotienten und beste einseitige Tests	53
10.1.	Definition: Likelihood-Quotient $R_{\vartheta',\vartheta}(x)$	53
10.1.1.	Beobachtung:	53
10.2.	Satz:	53
10.2.1.	Beispiel:	54
10.3.	Tests im (2-parametrischen) Gauß'schen Modell	54
10.3.1.	Satz: Linksseitiger χ^2 -Test für die Varianz im Gauß'schen Modell	55
10.3.2.	Bemerkung: (Der rechtsseitige χ^2 -Test für die Varianz)	56
10.3.3.	Satz: Einseitiger Student'scher t -Test	56
11.	χ^2-Anpassungstest	59
11.1.	Test auf eine vorgegebene Häufigkeitsverteilung	59
11.1.1.	Satz:	60
11.1.2.	Korollar:	62
11.1.3.	χ^2 -Anpassungstest	62
11.2.	χ^2 -Test auf Unabhängigkeit:	62
11.2.1.	Satz:	62
11.2.2.	Korollar: χ^2 -Test auf Unabhängigkeit	64
12.	Regression und das lineare Modell	67
12.1.	„Einfache“ Regression:	67
12.1.1.	Satz:	68
12.2.	Das lineare Modell:	69
12.2.1.	Satz von Gauß-Markov	70
12.3.	Das normale lineare Modell:	71
12.3.1.	Verallgemeinerung des Satzes von Student	72
12.3.2.	Korollar: Konfidenzbereiche im normalverteilten linearen Modell	73
12.3.3.	Korollar: Tests im normalverteilten linearen Modell	74
12.3.4.	Beispiel: Polynomielle Regression	75
12.4.	Varianzanalyse:	76
12.4.1.	Satz: Konfidenzbereich für $\gamma = (m_1, \dots, m_s)^t$	77
12.4.2.	Satz: F-Test auf Gleichheit der Gruppenmittelwerte	78
12.4.3.	Korollar: t -Test auf Gleichheit der Mittelwerte für zwei ungepaarte (normalverteilte) Stichproben	78
12.5.	Zweifaktorielle Varianzanalyse	79
12.5.1.	Beispiel:	79

II. Finanzmathematik - Mathematische Modelle und Methoden zur Bewertung von Derivaten	81
13. Einführung in die Finanzmathematik	83
13.1. Europäische Call- und Put-Option	83
13.2. Binomialmodell:	84
13.2.1. Mehrperioden-Binomialmodell (CRR(=Cox-Ross-Rubinstein)-Modell (1979))	85
13.2.2. Beobachtung:	85
13.2.3. Satz:	86
13.2.4. Beispiel: Europäische Call-Option	88
13.2.5. Bemerkung: Call-Put-Parität	88
13.3. Black-Scholes-Formel	88
13.3.1. Satz:	90
13.3.2. Korollar: Black-Scholes-Formel für den Preis eines Calls	90
13.4. Fundamentalsatz der Preistheorie (in endlichen Märkten)	90
13.4.1. Marktmodell	90
13.4.2. Definition: selbstfinanzierend	91
13.4.3. Definition: (duplizierbarer) Claim, Hedge, vollständiges Marktmodell, Arbitrage	92
13.4.4. Korollar:	92
13.4.5. Definition: Äquivalentes Martingalmaß	92
13.4.6. Beobachtung:	92
13.4.7. Fundamentalsatz der Preistheorie:	93
14. Exkurs: Stochastische Integration	95
14.0.8. Definition: Stochastischer Prozess	95
14.0.9. Bemerkung:	95
14.0.10. Lemma:	96
14.1. Definition: Stochastischer Integrator, Itô-Integral	97
14.1.1. Lemma:	97
14.1.2. Definition: Fast sicher endliche Variation	97
14.1.3. Bemerkung:	98
14.1.4. Satz:	99
14.2. Satz:	101
14.2.1. Erinnerung: Satz vom optionalen Stoppen	101
14.2.2. Erinnerung: Doobs \mathcal{L}^2 -Ungleichung:	101
14.3. Itô-Formel	103
14.3.1. Satz:	103
15. Die geometrische Brown'sche Bewegung und die Black-Scholes-Formel	105
III. Anhang	109
A. Verwendeter R-Code	111
A.1. Konfidenzintervalle im Binomialmodell: Für Abbildung 6.3 verwendeter R-Code . . .	111
A.2. Temperaturdiagramm	112

Teil I.
Statistik

1. Einführung in die Statistik

1.0.1. Beispiel:

n u.a. Messungen (z.B. Gewicht eines Goldbarrens), aus denen Rückschlüsse auf den „wahren“ Wert gezogen werden sollen.

Annahme: die Ergebnisse X_1, \dots, X_n seien normalverteilt $\mathcal{N}(m, \sigma^2)$ mit bekannter Varianz σ^2 und unbekanntem Mittelwert m (d.h. X_i haben Dichte $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$).
Wir beobachten x_1, \dots, x_n .

Punktschätzer:

$$M(X) = \frac{1}{n} \sum_{i=1}^n X_i$$

Konfidenzintervall: $\mathbf{P}_{(m, \sigma^2)} [\{[M(X) - \varepsilon(X), M(X) + \varepsilon(X)] \ni m\}] \geq 1 - \alpha \quad \forall m \in \mathbb{R}$

Beobachtung: $M(X) = \frac{1}{n} \sum_{i=1}^n X_i \stackrel{!}{=} \frac{1}{n} \sum_{i=1}^n (\sigma Z_i + m) = m + \frac{\sigma}{n} (Z_1 + \dots + Z_n) \sim \mathcal{N}(m, \frac{\sigma^2}{n})$

$$\forall m : \mathbf{P}_{(m, \sigma^2)} \left[\underbrace{\left\{ \left| \frac{M(X) - m}{\sqrt{\frac{\sigma^2}{n}}} \right| \leq a \right\}}_{=\{M(X) - \frac{\sigma}{n}a \leq m \leq M(X) + \frac{\sigma}{n}a\}} \right] = 2\Phi(a) - 1$$

Testen von Hypothesen: z.B.: Sprechen die Beobachtungen für $m \leq m_0$?

Regel: Ist $M(X) \leq c$ entscheide für Nullhypothese „ $m \leq m_0$ “

Ist jedoch $M(X) > c$ entscheide für Alternative „ $m > m_0$ “

Verlange: $\mathbf{P}_{(m, \sigma^2)} [M(X) > c] \leq \alpha \quad \forall m \leq m_0$ (Wahrscheinlichkeit eines Fehlers 1. Art klein)

und $\mathbf{P}_{(m, \sigma^2)} [M(X) > c] \stackrel{!}{=} \text{möglichst groß} \quad \forall m > m_0$ (Wahrscheinlichkeit eines Fehlers 2. Art klein)

Vorschlag: c so, dass $\mathbf{P}_{(m, \sigma^2)} [M(X) > c] = \mathbf{P}[Z > \frac{c - m_0}{\sigma}] = \alpha$

1.1. Definition und Sprechweise: statistisches Modell

Ein *statistisches Modell* ist ein Tripel $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$, wo $(\mathcal{X}, \mathcal{F})$ ein messbarer Raum (d.h. $\mathcal{F} \subset \mathbf{P}(\mathcal{X})$ ist eine σ -Algebra) und \mathbf{P}_ϑ auf $(\mathcal{X}, \mathcal{F})$ für $\vartheta \in \Theta$ ein Wahrscheinlichkeitsmaß ist, wobei die Indexmenge Θ nicht einpunktig ist.

1.2. Definition und Sprechweise: parametrisches Modell

Das Modell heißt *parametrisch*, wenn $\Theta \subseteq \mathbb{R}^d$ und *einparametrisch*, wenn $\Theta \subseteq \mathbb{R}$, *diskret*, wenn \mathcal{X} abzählbar, *stetig*, wenn $\mathcal{X} \subseteq \mathbb{R}^k$ (messbar) und jedes \mathbf{P}_ϑ eine Dichte f_ϑ bzgl. des Lebesgue-Maß besitzt.

Ein Modell heißt *Standardmodell*, falls es diskret oder stetig ist.

1.2.1. Bemerkung: Uniforme Verteilung auf einer Cantormenge

Wähle $(X_n)_n$ i.i.d. $\sim \text{Ber}(\frac{1}{2})$, $S = \sum_{n=1}^{\infty} 2X_n \cdot 3^{-n}$, die Verteilung von S ist weder diskret noch stetig, sie besitzt keine Dichte bzgl. des Lebesgue-Maßes.

1.3. Definition: Produktmodell zu $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$:

Ein Produktmodell ist ein statistisches Modell der Form $(\mathcal{X}^n, \mathcal{F}^n, \mathbf{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)$.

1.4. Definition: Statistik

- (i) Eine Zufallsgröße von $(\mathcal{X}, \mathcal{F})$ nach (Σ, \mathcal{S}) heißt eine *Statistik*.
- (ii) $\tau : \Theta \rightarrow \Sigma$, $\tau(\vartheta)$ ist eine Kenngröße oder ein „Parametermerkmal“, $T : \mathcal{X} \rightarrow \Sigma$ (messbar) heißt ein Schätzer (für τ).

1.5. Beispiel: „Taxiproblem“, Schätzung des Wertebereichs:

Angenommen alle Taxis einer Stadt sind mit den Nummern $1, \dots, N$ durchnummeriert und man beobachtet die Nummern x_1, \dots, x_n mit $n \ll N$ und den x_i paarweise verschieden. Was wäre dann ein guter Schätzer für N ?

Wir modellieren kontinuierlich: Unter \mathbf{P}_ϑ sind X_1, \dots, X_n unabhängig, uniform verteilt auf $[0, \vartheta]$. Wir betrachten folgende Schätzer für ϑ :

$$\begin{aligned} T_n(x) &:= \frac{2}{n} \sum_{i=1}^n X_i, \\ \widetilde{T}_n(x) &:= \max\{X_1, \dots, X_n\}, \\ T'_n(x) &:= \widetilde{T}_n(x) + \min\{x_1, \dots, x_n\}. \end{aligned}$$

Es gilt

$$\begin{aligned} \mathbf{E}_\vartheta[T_n] &= \frac{2}{n} \sum_{i=1}^n \mathbf{E}_\vartheta[X_i] = \vartheta, \\ \mathbf{E}_\vartheta[\widetilde{T}_n] &= \int_0^\vartheta t \cdot \frac{nt^{n-1}}{\vartheta^n} \mathbf{1}_{[0, \vartheta]}(t) dt \\ &= \frac{n}{n+1} \vartheta, \end{aligned}$$

$$\begin{aligned}
\mathbf{E}_\vartheta[T'_n] &= \mathbf{E}_\vartheta[\widetilde{T}_n] + \underbrace{\mathbf{E}_\vartheta[\min\{X_1, \dots, X_n\}]}_{\substack{= \mathbf{E}_\vartheta[\min\{\vartheta - X_1, \dots, \vartheta - X_n\}] \\ \vartheta + \min\{-X_1, \dots, -X_n\} = \vartheta + \max\{X_1, \dots, X_n\}}} \\
&= \frac{n}{n+1}\vartheta + \vartheta - \frac{n}{n+1}\vartheta = \vartheta.
\end{aligned}$$

Und die Varianz?

$$\begin{aligned}
\text{Var}_\vartheta(T_n) &= \frac{4}{n^2} n \underbrace{\text{Var}_\vartheta(X_i)}_{= \frac{\vartheta^2}{12}} \\
&= \frac{\vartheta^2}{3n},
\end{aligned}$$

$$\begin{aligned}
\text{Var}_\vartheta(\widetilde{T}_n) &= \mathbf{E}[\widetilde{T}_n^2] - \mathbf{E}_\vartheta[\widetilde{T}_n]^2 \\
&= \int_0^\vartheta t^2 \cdot \frac{nt^{n-1}}{\vartheta^n} dt - \left(\int_0^\vartheta t \cdot \frac{nt^{n-1}}{\vartheta^n} dt \right)^2 \\
&= \frac{\vartheta^2 n}{n+2} - \frac{\vartheta^2 n^2}{(n+1)^2} \\
&= \frac{n}{(n+1)^2(n+2)} \vartheta^2.
\end{aligned}$$

1.5.1. Bemerkung:

$T^* := \frac{n+1}{n} \widetilde{T}_n$ hat

$$\mathbf{E}_\vartheta T_n^* = \vartheta$$

mit

$$\text{Var}_\vartheta(T_n^*) = \frac{\vartheta^2}{n(n+2)}$$

Übungsaufgabe: Varianz von T'_n berechnen.

Wir hätten gerne, dass

$$\mathbf{E}_\vartheta[(T - \vartheta)^2] \stackrel{!}{=} \text{klein, glm für alle } \vartheta,$$

d.h. der erwartete quadratische Fehler sollte möglichst (gleichmäßig) klein sein.

¹ $\mathbf{P}_\vartheta[\widetilde{T}_n \leq c] = \left(\frac{c}{\vartheta}\right)^n$, d.h. \widetilde{T}_n hat die Dichte $\frac{nt^{n-1}}{\vartheta^n} \mathbf{1}_{[0, \vartheta]}(t)$

2. Das Maximum-Likelihood-Prinzip:

2.0.2. Beispiel: Qualitätskontrolle

Es sind N Orangen angeliefert worden, darunter ϑ viele faul. Wir nehmen eine Stichprobe von n Orangen, von denen x faul sind.

$$\Theta = \{0, 1, \dots, N\}, \quad \rho_{\vartheta}(x) = \frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{N}{n}},$$

für welche(s) ϑ ist $\rho_{\vartheta}(x)$ maximal?

Betrachte:

$$\frac{\rho_{\vartheta}(x)}{\rho_{\vartheta-1}(x)} = \frac{\frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{N}{n}}}{\frac{\binom{\vartheta-1}{x} \binom{N-\vartheta+1}{n-x}}{\binom{N}{n}}} = \frac{\vartheta}{\vartheta-x} \frac{(N-\vartheta+1-n+x)}{(N-\vartheta+1)} \geq 1 \iff \vartheta n \leq (N+1)x,$$

demnach ist $T(x) = \lfloor \frac{x}{n}(N+1) \rfloor$ (ein) Maximum-Likelihood-Schätzer für ϑ . (Der Maximum-Likelihood-Schätzer ist hier eindeutig, es sei denn $\frac{x}{n}(N+1) \in \mathbb{N}$, dann sind $\frac{x}{n}(N+1)$ und $\frac{x}{n}(N+1) - 1$ beides ML-Schätzer.)

2.1. Definition: Likelihood-Funktion

Sei $\mathcal{M} = (\mathcal{X}, \mathcal{F}, \mathbf{P}_{\vartheta} : \vartheta \in \Theta)$ ein statistisches Standardmodell. Die Funktion

$$\rho : \mathcal{X} \times \Theta \rightarrow \mathbb{R}_+,$$

wobei $x \mapsto \rho(x, \vartheta)$ die Dichte von \mathbf{P}_{ϑ} bezüglich des Referenzmaßes ist, heißt *Likelihood-Funktion*. Wir schreiben auch $\rho(x, \vartheta) = \rho_{\vartheta}(x) = \rho_x(\vartheta)$. Ein Schätzer $T : \mathcal{X} \rightarrow \Theta$ heißt (ein) Maximum-Likelihood-Schätzer (MLE¹), wenn

$$\rho(x, T(x)) = \max_{\vartheta \in \Theta} \rho(x, \vartheta) \quad \forall x \in \mathcal{X}.$$

2.1.1. Beispiele:

Erfolgswahrscheinlichkeit beim Münzwurf (Schätzen des Erfolgsparameters einer Binomialverteilung): Wir werfen eine Münze n mal und zählen die Anzahl Erfolge. Statistisches Modell:

$$(\{0, \dots, n\}, \mathbf{P}(\{0, \dots, n\}), \text{Bin}(n, \vartheta : \vartheta \in [0, 1])).$$

Beobachte x Erfolge:

$$\rho_x(\vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \text{ ist maximal} \iff$$

¹MLE="maximum likelihood estimator", wird häufig mit $\hat{\vartheta}_{ML}$ bezeichnet

$$\frac{d}{d\vartheta} \log \rho_x(\vartheta) \cdot \frac{d}{d\vartheta} (\log \binom{n}{x}) + x \log \vartheta + (n-x) \log(1-\vartheta) = \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} \stackrel{!}{=} 0$$

$\iff \vartheta = \frac{x}{n}$, also ist $T(x) = \frac{x}{n}$ der ML-Schätzer für ϑ

2.1.2. Beispiel: Uniforme Verteilung auf $[0, \vartheta]$

x_1, \dots, X_n unter \mathbf{P}_ϑ i.i.d., $\sim \text{unif}([0, \vartheta])$.

$$\rho(x, \vartheta) = \frac{1}{\vartheta} \mathbf{1}_{[0, \vartheta]}(x), \rho^{\otimes n}((x_1, \dots, x_n), \vartheta) = \begin{cases} \frac{1}{\vartheta^n} & \text{falls } x_1, \dots, x_n \in \vartheta \\ 0 & \text{sonst} \end{cases}$$

Also ist $\max\{X_1, \dots, X_n\}$ der Maximum-Likelihood-Schätzer.

2.1.3. Beispiel: Normales Modell

Sei $\vartheta := (m, v) \in (-\infty, \infty) \times (0, \infty)$. $\mathbf{P}_\vartheta = \mathcal{N}_{(m, v)}$

$$\rho(x, \vartheta) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-m)^2}{2v}},$$

n Beobachtungen: $(\mathbf{Q}_\vartheta = \mathbf{P}_\vartheta^{\otimes n})$,

$$\rho^{\otimes n}((x_1, \dots, x_n), \vartheta) = \frac{1}{\sqrt{2\pi v}^n} \exp\left(-\sum_{i=1}^n \frac{(x_i - m)^2}{2v}\right)$$

Gesucht: ML-Schätzer für $\vartheta = (m, v)$, basierend auf den Beobachtungen X_1, \dots, X_n :

$$\log \rho^{\otimes n}((X_1, \dots, X_n), \vartheta) = -\frac{n}{2} \log 2\pi v - \frac{1}{2v} \sum_{i=1}^n (X_i - m)^2$$

Beobachtung: Steiner'scher Verschiebungssatz (siehe auch Abbildung 2.1)

Sei $M(\mathbf{x}) = M(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i$ der empirische Mittelwert der x_i . Es gilt für jedes $m \in \mathbb{R}$:

$$\frac{1}{n} \sum_{i=1}^n (x_i - m)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - M(\mathbf{x}))^2 + (M(\mathbf{x}) - m)^2,$$

denn die rechte Seite ist gleich

$$\frac{1}{n} \sum_{i=1}^n x_i^2 - 2M(\mathbf{x}) \frac{1}{n} \sum_{j=1}^n x_j + M(\mathbf{x})^2 + (M(\mathbf{x})^2 - 2M(\mathbf{x})m + m^2) = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2M(\mathbf{x})m + m^2.$$

Zur Wahl von v :

$$0 \stackrel{!}{=} \frac{d}{dv} \left(-\frac{n}{2} \log(2\pi v) \right) - \frac{1}{2v} \sum_{i=1}^n (X_i - M(\mathbf{X}))^2 = \frac{n}{2v} - \frac{1}{2v^2} \sum_{i=1}^n (X_i - M(\mathbf{X}))^2 \iff v = \frac{1}{n} \sum_{i=1}^n (X_i - M(\mathbf{X}))^2,$$

d.h. empirischer Mittelwert und empirische Varianz sind die ML-Schätzer im normalen Produktmodell.

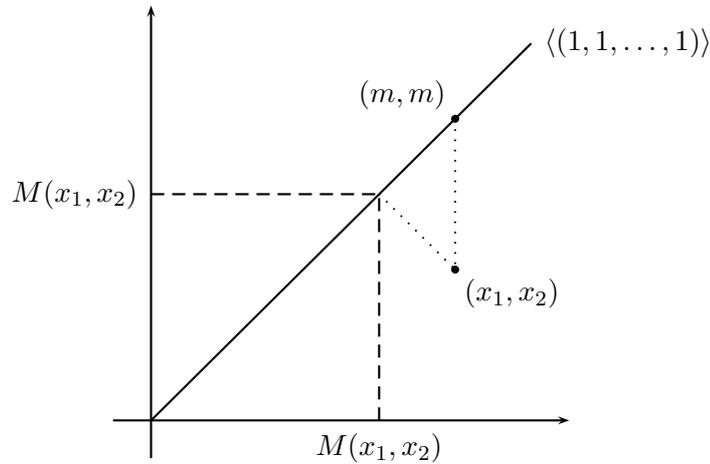


Abbildung 2.1.: Unter allen Punkten auf der Diagonalen $\langle(1, 1, \dots, 1)\rangle$ hat $(M(\mathbf{x}), M(\mathbf{x}), \dots, M(\mathbf{x}))$ den kleinsten euklidischen Abstand von (x_1, \dots, x_n) .

2.2. Sprechweisen: Erwartungstreu, Mittlerer quadratischer Fehler, Bias

Sei T ein Schätzer für ein reelles Parametermerkmal $\tau(\vartheta)$. T heißt *erwartungstreu*, wenn $\mathbf{E}_\vartheta[T] = \tau(\vartheta)$, $B_\vartheta(T) = \mathbf{E}_\vartheta[T] - \tau(\vartheta)$ heißt *Bias* (Verzerrung).

Mittlerer quadratischer Fehler: $F_\vartheta(T) = \mathbf{E}_\vartheta[(T - \tau(\vartheta))^2] = \text{Var}_\vartheta(T) + B_\vartheta(T)^2$.

2.3. Satz: Erwartungstreue Schätzer für Varianz und Mittelwert

Betrachte ein Produktmodell $(\mathbb{R}^n, \mathcal{B}^{\otimes n}(\mathbb{R}), \mathbf{Q}_\vartheta^{\otimes n} : \vartheta \in \Theta)$ (d.h. seien X_1, \dots, X_n i.i.d. gem \mathbf{Q}_ϑ , $\vartheta \in \Theta$), wo \mathbf{Q}_ϑ Mittelwert $m(\vartheta) = \int_{\mathbb{R}} x \mathbf{Q}_\vartheta(dx)$ und Varianz $v(\vartheta) = \int_{\mathbb{R}} (x - m(\vartheta))^2 \mathbf{Q}_\vartheta(dx)$ hat.

$M := \frac{1}{n} \sum_{i=1}^n X_i$ ist ein erwartungstreuer Schätzer für $m(\vartheta)$ und $V^* := \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2$ ist ein erwartungstreuer Schätzer für $v(\vartheta)$.

Beweis:

$\mathbf{E}_\vartheta[M] = m(\vartheta)$ (Linearität für den Erwartungswert),

$$\begin{aligned} V &= \frac{1}{n} \sum_{i=1}^n (X_i - M)^2, \\ \mathbf{E}_\vartheta[V] &= \frac{1}{n} \sum_{i=1}^n \mathbf{E}_\vartheta[(X_i - M)^2] \\ &= \text{Var}_\vartheta(X_1 - M) = \text{Var}_\vartheta\left(\frac{n-1}{n}X_1 - \frac{1}{n} \sum_{i=2}^n X_i\right) \\ &= \left(\frac{n-1}{n}\right)^2 + \frac{n-1}{n^2} \text{Var}_\vartheta(X_1) \\ &= \frac{n-1}{n} v(\vartheta). \end{aligned}$$

□

2.3.1. Definition und Beobachtung: Konsistenz

Man nennt eine Folge von Schätzern T_n für $\tau(\vartheta)$ konsistent, wenn

$$\mathbf{P}_\vartheta[|T_n - \tau(\vartheta)| > \varepsilon] \rightarrow 0 \quad \forall \varepsilon > 0, \forall \vartheta.$$

Sei $\mathbf{P}_\vartheta = \mathbf{Q}_\vartheta^{\otimes n}$, $M_n = \frac{1}{n} \sum_{i=1}^n X_i$, $V_n^* = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2$. Es gilt mit dem Schwachen GGZ:

$$\mathbf{P}_\vartheta[|M_n - m(\vartheta)| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0, \forall \vartheta$$

und

$$\mathbf{P}_\vartheta[|V_n^* - v(\vartheta)| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0, \forall \vartheta$$

(Übung, benutze schwGGZ), d.h. im Produktmodell ist $(M_n)_{n \in \mathbb{N}}$ konsistent für $m(\vartheta)$, $(V_n^*)_{n \in \mathbb{N}}$ (und ebenso $(V_n)_{n \in \mathbb{N}}$) konsistent für $v(\vartheta)$.

3. Exponentielle Familien, Fisher-Information und beste Schätzer:

Ein erwartungstreuer Schätzer T für ein reelles Parametermerkmal $\tau(\vartheta)$ heißt *varianzminimierend* („gleichmäßig bester Schätzer“, UMVU¹-estimator), falls

$$\text{Var}_{\vartheta}(T) \leq \text{Var}_{\vartheta}(S) \quad \forall \vartheta \in \Theta, \forall \text{ anderen erwartungstreuen Schätzer } S.$$

3.1. Definition: regulär

Ein einparametriges Standardmodell heißt *regulär*, falls

- (i) $\Theta \subset \mathbb{R}$ offenes Intervall,
- (ii) $\rho(x, \vartheta)$ ist strikt positiv auf $\mathcal{X} \times \Theta$, für jedes x ist $\vartheta \mapsto \rho(x, \vartheta)$ stetig diffbar,
- (iii) $U_{\vartheta}(x) := \frac{d}{d\vartheta} \log \rho(x, \vartheta) \left(= \frac{\frac{d\rho(x, \vartheta)}{d\vartheta}}{\rho(x, \vartheta)} \right)$ erfüllt $\text{Var}_{\vartheta}(U_{\vartheta}) \in (0, \infty)$ (U_{ϑ} heißt die *Scorefunktion*)

und es gilt $\int_{\mathcal{X}} \frac{d}{d\vartheta} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \int_{\mathcal{X}} \rho(x, \vartheta) dx (= 0)$.

3.1.1. Bemerkung:

$$\mathbf{E}_{\vartheta}[U_{\vartheta}] = \int_{\mathcal{X}} \frac{d}{dx} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \int_{\mathcal{X}} \rho(x, \vartheta) dx = 0$$

3.2. Definition: Fisher-Information

$I(\vartheta) := \text{Var}_{\vartheta}(U_{\vartheta})$ heißt *Fisher-Information*.

3.2.1. Bemerkung:

- (i) Wenn $I(\vartheta) \equiv 0$ auf einem Teilintervall $\Theta_1 \subset \Theta$, so ist $\rho(x, \vartheta) = \rho(x, \vartheta')$ für $\vartheta, \vartheta' \in \Theta_1$.
- (ii) \mathcal{M} ein reguläres Standardmodell mit Fisher-Information $I(\vartheta)$, dann ist das n -fache Produktmodell ebenfalls ein reguläres Standardmodell mit Fisher-Information $I^{\otimes n}(\vartheta) = n \cdot I(\vartheta)$.

(Denn $\rho^{\otimes n}((x_1, \dots, x_n), \vartheta) = \prod_{i=1}^n \rho(x_i, \vartheta)$,

$$U_{\vartheta}^{\otimes n}(x_1, \dots, x_n) = \sum_{i=1}^n \frac{d}{d\vartheta} \log \rho(x_i, \vartheta),$$

¹UMVU = uniformly minimum variance unbiased

$$\text{Var}_{\vartheta}(U_{\vartheta}^{\otimes n}) = \sum_{i=1}^n \text{Var}_{\vartheta}(U_{\vartheta}(X_i)) = n \cdot I(\vartheta).$$

3.3. Definition: regulärer Schätzer

Ein Schätzer T heißt *regulär*, wenn für alle $\vartheta \in \Theta$ gilt: $\frac{d}{d\vartheta} \int T(x) \rho(x, \vartheta) dx = \int T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx$.

$\underbrace{\hspace{10em}}_{= \mathbf{E}_{\vartheta}[T]}$

3.4. Satz: Cramér-Rao-Schranke

Sei \mathcal{M} ein reguläres Standardmodell, $\tau : \Theta \rightarrow \mathbb{R}$ ein stetig differenzierbares Parametermerkmal, T ein regulärer, erwartungstreuer Schätzer für τ (d.h. $\mathbf{E}_{\vartheta}[T] = \tau(\vartheta)$).

Dann gilt (Cramér-Rao-Schranke):

$$\text{Var}_{\vartheta}(T) \geq \frac{(\tau'(\vartheta))^2}{I(\vartheta)} \quad \forall \vartheta \in \Theta. \quad (*)$$

Gleichheit gilt für alle ϑ genau dann, wenn

$$T(x) - \tau(\vartheta) = \frac{\tau'(\vartheta) U_{\vartheta}(x)}{I(\vartheta)},$$

d.h. wenn die Likelihood-Funktion die Form $\rho(x, \vartheta) = h(x) \exp(a(\vartheta)T(x) - b(\vartheta))$ hat. Hierbei ist $a : \Theta \rightarrow \mathbb{R}$ eine Stammfunktion von $\frac{1}{\tau}$, $h : \mathcal{X} \rightarrow (0, \infty)$ meßbar, $b(\vartheta) = \log \int e^{a(\vartheta)T(x)} h(x) dx$

Beweis:

$$\begin{aligned} \text{Cov}_{\vartheta}(T, U_{\vartheta}) &= \mathbf{E}_{\vartheta}[T \cdot U_{\vartheta}] - \mathbf{E}_{\vartheta}[T] \cdot \underbrace{\mathbf{E}_{\vartheta}[U_{\vartheta}]}_{=0} \\ &= \int T(x) \frac{d}{d\vartheta} \log \rho(x, \vartheta) \cdot \rho(x, \vartheta) dx \\ &= \int T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx \\ &= \frac{d}{d\vartheta} \int T(x) \rho(x, \vartheta) dx \\ &= \frac{d}{d\vartheta} \mathbf{E}_{\vartheta}[T] \\ &= \tau'(\vartheta) \end{aligned}$$

Sei $c(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)}$. Es gilt

$$(**) \quad 0 \leq \text{Var}_{\vartheta}(T - c(\vartheta)U_{\vartheta}) = \text{Var}_{\vartheta}(T) + \underbrace{c(\vartheta)^2 \cdot \text{Var}_{\vartheta}(U_{\vartheta})}_{= \frac{\tau'(\vartheta)^2}{I(\vartheta)}} - 2 \underbrace{c(\vartheta) \cdot \text{Cov}_{\vartheta}(T, U_{\vartheta})}_{= \frac{\tau'(\vartheta)^2}{I(\vartheta)}} = \text{Var}_{\vartheta}(T) - \frac{(\tau'(\vartheta))^2}{I(\vartheta)},$$

was die Ungleichung in (*) beweist. Für Gleichheit in (**), und damit in (*), (für ein ϑ) muss $T - c(\vartheta)U_\vartheta$ \mathbf{P}_ϑ -f.s. konstant (und somit gleich seinem Erwartungswert $\tau(\vartheta)$) sein, d.h.

$$\{x : T(x) - c(\vartheta)U_\vartheta(x) \neq \tau(\vartheta)\} \text{ hat Maß } 0 \text{ unter } \mathbf{P}_\vartheta,$$

da $c(\vartheta), U_\vartheta, \tau(\vartheta)$ als Funktion von ϑ stetig sind:

$$\mu(\{x : T(x) - c(\vartheta)U_\vartheta(x) \neq \tau(\vartheta) \text{ für ein } \vartheta \in \Theta\}) = 0$$

Also: Gleichheit in (**) \iff

$$\frac{d}{d\vartheta} \log \rho(x, \vartheta) = U_\vartheta(x) = \frac{T(x) - \tau(\vartheta)}{c(\vartheta)},$$

$$\log \rho(x, \vartheta) = \log \rho(x, \vartheta_0) + T(x) \cdot \underbrace{\int_{\vartheta_0}^{\vartheta} -\frac{I(\sigma)}{\tau'(\sigma)} d\sigma}_{=a(\vartheta)-a(\vartheta_0)} - \int_{\vartheta_0}^{\vartheta} \frac{I(\sigma)}{\tau'(\sigma)} \cdot \tau(\sigma) d\sigma$$

□

3.4.1. Beispiele:

(i) Binomialverteilung: $\mathbf{P}_\vartheta = \text{Bin}_{n,\vartheta}, \vartheta \in (0, 1)$. Dann ist

$$\rho(x, \vartheta) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x} = \binom{n}{x} \exp \left(\underbrace{\frac{x}{n}}_{T(x)} \underbrace{n \log\left(\frac{\vartheta}{1-\vartheta}\right)}_{=a(\vartheta)} + \underbrace{n \log(1-\vartheta)}_{=b(\vartheta)} \right)$$

(ii) Poissonverteilung: $\mathbf{P}_\vartheta = \text{Pois}(\vartheta), \vartheta \in (0, \infty)$:

$$p(x, \vartheta) = e^{-\vartheta} \frac{\vartheta^x}{x!} = \underbrace{\frac{1}{x!}}_{=h(x)} e^{\underbrace{x}_{T(x)} \underbrace{\log \vartheta}_{=a(\vartheta)} - \underbrace{\vartheta}_{b(\vartheta)}}$$

3.5. Definition: Cramér-Rao-effizient

Ein Schätzer, der die Gleichheit in der Cramér-Rao-Ungleichung (Schranke) realisiert, heißt *Cramér-Rao-effizient*.

3.6. Definition: exponentielles Modell

Ein einparametrisches statistisches Standardmodell $\mathcal{M} = (\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$ (mit $\Theta \subseteq \mathbb{R}$ offenes Intervall) heißt *exponentielles Modell* bzgl. der Statistik $T : \mathcal{X} \rightarrow \mathbb{R}$, wenn die Likelihood-Funktion von der Form

$$\rho(x, \vartheta) = h(x) \cdot \exp(a(\vartheta) \cdot T(x) - b(\vartheta))$$

mit einer stetig differenzierbaren Funktion $a : \Theta \rightarrow \mathbb{R}$ mit $a' \neq 0$ und $h : \mathcal{X} \rightarrow \mathbb{R}_+$ messbar.²

$(\mathbf{P}_\vartheta)_{\vartheta \in \Theta}$ heißt „*exponentielle Familie*“.

(Wir nehmen an, dass T nicht fast überall konstant ist.)

²Wobei $b(\vartheta) = \log \int h(x) e^{a(\vartheta)T(x)} dx$ die Normierungsfunktion ist.

3.6.1. Bemerkung:

Das Produktmodell $\mathcal{M}^{\otimes n}$ eines exponentiellen Modells \mathcal{M} bzgl. der Statistik T ist ebenfalls ein exponentielles Modell, und die zugrundeliegende Statistik ist

$$T_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n T(x_i),$$

denn

$$\rho^{\otimes n}((x_1, \dots, x_n), \vartheta) = \prod_{i=1}^n \rho(x_i, \vartheta) = \prod_{i=1}^n h(x_i) \exp \left(na(\vartheta) \cdot \frac{1}{n} \cdot \sum_{i=1}^n (T(x_i)) - nb(\vartheta) \right).$$

3.7. Satz über die „Gutartigkeit“ von exponentiellen Modellen

Sei \mathcal{M} ein exponentielles Modell, dann gilt:

- (i) b ist stetig differenzierbar auf ϑ mit $b'(\vartheta) = a'(\vartheta) \mathbf{E}_{\vartheta}[T]$.
- (ii) Jede Statistik $S : \mathcal{X} \rightarrow \mathbb{R}$, deren Erwartungswert für jedes $\vartheta \in \Theta$ existiert, ist regulär. Insbesondere sind \mathcal{M} und T regulär und $\tau(\vartheta) := \mathbf{E}_{\vartheta}[T]$ ist stetig differenzierbar mit $\tau'(\vartheta) = a'(\vartheta) \cdot \text{Var}_{\vartheta}(T) > 0$.
- (iii) $I(\vartheta) = a'(\vartheta) \tau'(\vartheta)$

Beweis:

Wir nehmen o.E. an, dass $a(\vartheta) = \vartheta$, ansonsten parametrisieren wir um und verwenden die Kettenregel.

- (i) Sei S eine Statistik, deren Erwartungswert $\mathbf{E}_{\vartheta}[S] = \int_{\mathcal{X}} S(x) h(x) e^{a(\vartheta)T(x) - b(\vartheta)} dx$ für jedes ϑ existiert (und endlich ist).
Es sei weiter $u_S(\vartheta) := e^{b(\vartheta)} \mathbf{E}_{\vartheta}[S] = \int_{\mathcal{X}} S(x) h(x) e^{\vartheta T(x)} dx$.
Wenn t so klein, dass $\vartheta \pm t \in \Theta$, dann:

$$\begin{aligned} & \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \int h(x) |S(x)| \cdot |T(x)|^k e^{\vartheta T(x)} dx \\ & \stackrel{3}{=} \int_{\mathcal{X}} \sum_{k=0}^{\infty} \frac{|t|^k}{k!} h(x) |S(x)| \cdot |T(x)|^k e^{\vartheta T(x)} dx \\ & = \int_{\mathcal{X}} |S(x)| h(x) e^{\vartheta T(x) + |t| T(x)} dx \\ & \leq \underbrace{\int_{\mathcal{X}} |S(x)| h(x) e^{(\vartheta+t)T(x)} dx}_{< \infty} + \underbrace{\int_{\mathcal{X}} |S(x)| h(x) e^{(\vartheta-t)T(x)} dx}_{< \infty} \\ & < \infty \end{aligned}$$

Insbesondere: $ST^k \in \mathcal{L}^1(\mathbf{P}_\vartheta) \forall \vartheta$ und

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \int S(x)h(x)T(x)^k e^{\vartheta T(x)} dx \text{ konvergiert absolut zu } \int S(x)h(x)e^{(\vartheta+t)T(x)} dx = u_S(\vartheta + t)$$

Also: $\vartheta \mapsto u_S(\vartheta)$ ist analytisch, darf unter dem Integral abgeleitet werden.

$$\begin{aligned} u'_S(\vartheta) &= \int S(x)T(x)h(x)e^{\vartheta T(x)} dx \\ &= e^{b(\vartheta)} \mathbf{E}_\vartheta[ST] \end{aligned}$$

und mit der speziellen Wahl $S \equiv 1$

$$\begin{aligned} u'_1(\vartheta) &= e^{b(\vartheta)} \mathbf{E}_\vartheta[T], \quad u''_1(\vartheta) = e^{b(\vartheta)} \mathbf{E}_\vartheta[T^2] \text{ mit} \\ b'(\vartheta) &= \frac{d}{d\vartheta} \log u_1(\vartheta) = \frac{u'_1(\vartheta)}{u_1(\vartheta)} = \mathbf{E}_\vartheta[T] =: \tau(\vartheta), \text{ d.h. (i) gilt und} \\ \tau'(\vartheta) &= b'(\vartheta) = \frac{u''_1(\vartheta)}{u_1(\vartheta)} - \left(\frac{u'_1(\vartheta)}{u_1(\vartheta)} \right)^2 = \mathbf{E}_\vartheta[T^2] - (\mathbf{E}_\vartheta[T])^2 = \text{Var}_\vartheta(T). \end{aligned}$$

(ii) Betrachte nun für eine allgemeine Statistik S :

$$\begin{aligned} \frac{d}{d\vartheta} \mathbf{E}_\vartheta[S] &= \frac{d}{d\vartheta} (e^{-b(\vartheta)} u_S(\vartheta)) \\ &= (u'_S(\vartheta) - u_S(\vartheta)b'(\vartheta))e^{-b(\vartheta)} \\ &= \mathbf{E}_\vartheta[ST] - \mathbf{E}_\vartheta[S]\mathbf{E}_\vartheta[T] = \text{Cov}_\vartheta(S, T) \\ &\stackrel{4}{=} \text{Cov}_\vartheta(S, U_\vartheta) \end{aligned}$$

Wähle nun $S = T$:

$$\begin{aligned} \tau'(\vartheta) &= \frac{d}{d\vartheta} \int T(x)\rho(x, \vartheta) dx = \mathbf{E}_\vartheta[TU_\vartheta] \\ &= \int T(x) \frac{\frac{d}{d\vartheta} \rho(x, \vartheta)}{\rho(x, \vartheta)} \cdot \rho(x, \vartheta) dx \\ &= \int T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx \end{aligned}$$

Und weiter:

$$\frac{d}{d\vartheta} \mathbf{E}_\vartheta[S] = \frac{d}{d\vartheta} \int S(x)\rho(x, \vartheta) dx = \int S(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx,$$

d.h. (ii) gilt.

(iii) $I(\vartheta) = \text{Var}_\vartheta(U_\vartheta) = \text{Var}_\vartheta(T) = \tau'(\vartheta)$

□

³Satz von der monotonen Konvergenz

⁴Funktioniert, da die Kovarianz translationsinvariant ist und $U_\vartheta(x) = \frac{d}{d\vartheta} \log \rho(x, \vartheta) = T(x) - b'(\vartheta)$

3.8. Korollar:

Wie wir bereits anhand der Cramér-Rao-Schranke gesehen haben, gilt: Für ein (reguläres) exponentielles Modell ist T ein bester Schätzer für $\tau(\vartheta) := \mathbf{E}_\vartheta[T] = \frac{b'(\vartheta)}{a'(\vartheta)}$ mit

$$\text{Var}_\vartheta(T) = \frac{\tau'(\vartheta)}{a'(\vartheta)} = \frac{\tau'(\vartheta)^2}{I(\vartheta)}.$$

3.8.1. Beispiel:

Schätzung des Erwartungswertes im normalen Modell bei bekannter Varianz $v > 0$: $\mathbf{P}_\vartheta = \mathcal{N}_{\vartheta, v}$

$$\begin{aligned} \rho(x, \vartheta) &= \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(x - \vartheta)^2\right) \\ &= \frac{1}{\underbrace{\sqrt{2\pi v}}_{=h(x)}} \cdot \exp\left(\underbrace{x}_{=T(x)} \underbrace{\frac{\vartheta}{v}}_{a(\vartheta)} - \underbrace{\frac{\vartheta^2}{2v}}_{=b(\vartheta)}\right) \end{aligned}$$

Probe:

$$\tau(\vartheta) = \mathbf{E}_\vartheta[T] = \vartheta = \frac{\frac{\vartheta}{v}}{\frac{1}{v}} = \frac{b'(\vartheta)}{a'(\vartheta)}, \text{Var}_\vartheta(T) = \frac{\tau'(\vartheta)}{a'(\vartheta)} = \frac{1}{\frac{1}{v}} = v \quad \checkmark$$

$$\Rightarrow I(\vartheta) = \frac{1}{v}$$

(Übung: Schätzung der Varianz im normalen Modell bei bekanntem Mittelwert.)

3.9. Konsistenz des ML-Schätzers im Produktmodell:

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{Q}_\vartheta : \vartheta \in \Theta)$ ein einparametrisches Standardmodell mit Likelihoodfunktion $\rho(x, \vartheta)$, $\mathbf{P}_\vartheta := \mathbf{Q}_\vartheta^{\otimes \mathbb{N}}$. Es gelte weiter:

(i) $\Theta \in \mathbb{R}$ ein offenes Intervall und $\mathbf{Q}_\vartheta \neq \mathbf{Q}_{\vartheta'}$ für $\vartheta \neq \vartheta'$.

(ii) Für jedes $(x_1, \dots, x_n) \in \mathcal{X}^n, n \in \mathbb{N}$ ist

$$\begin{aligned} \vartheta \mapsto \underbrace{\rho^{\otimes n}((x_1, \dots, x_n), \vartheta)}_{= \prod_{i=1}^n \rho(x_i, \vartheta)} \end{aligned}$$

unimodal, d.h. $\exists T_n = T_n(x_1, \dots, x_n)$, sodass $\rho^{\otimes n}((x_1, \dots, x_n), \vartheta)$ wachsend für $\vartheta < T_n$ und fallend für $\vartheta > T_n$.

Dann ist die (eindeutig bestimmte) Folge von (ML-)Schätzern $(T_n)_n$ konsistent für ϑ , d.h.

$$\forall \varepsilon > 0 : \mathbf{P}_\vartheta[|T_n - \vartheta| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0.$$

Beweis:

Sei $\vartheta \in \Theta$, ε so klein, dass $\vartheta \pm \varepsilon \in \Theta$, $0 < \delta < H(\mathbf{Q}_\vartheta | \mathbf{Q}_{\vartheta \pm \varepsilon})$,

$$\left\{ \underbrace{\frac{1}{n} \log \frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta \pm \varepsilon}^{\otimes n}}}_{= \frac{1}{n} \sum_{i=1}^n \log \frac{\rho_\vartheta(x_i)}{\rho_{\vartheta \pm \varepsilon}(x_i)}} > \delta \right\} \subseteq \{ \rho_\vartheta^{\otimes n} > \rho_{\vartheta \pm \varepsilon}^{\otimes n} \} \stackrel{\text{Unimodalitat}}{\subseteq} \{ \vartheta - \varepsilon < T_n < \vartheta + \varepsilon \}$$

Wir erwarten:

$$\frac{1}{n} \sum_{i=1}^n \log \frac{\rho_\vartheta(x_i)}{\rho_{\vartheta \pm \varepsilon}(x_i)} \xrightarrow{\text{GGZ}} \mathbf{E}_{\mathbf{Q}_\vartheta} \left[\log \left(\frac{\rho_\vartheta(x_1)}{\rho_{\vartheta \pm \varepsilon}(x_1)} \right) \right],$$

$$\mathbf{E}_\vartheta \left[\log \left(\frac{\rho_\vartheta(x_1)}{\rho_{\vartheta \pm \varepsilon}(x_1)} \right) \right] = H(\mathbf{Q}_\vartheta | \mathbf{Q}_{\vartheta \pm \varepsilon}) > 0.$$

Wir mussen noch zeigen (o.B.d.A. nur fur $+\varepsilon$):

$$\mathbf{P}_\vartheta \left[\left\{ \frac{1}{n} \log \left(\frac{\rho_\vartheta^{\otimes n}(X_1, \dots, X_n)}{\rho_{\vartheta + \varepsilon}^{\otimes n}(X_1, \dots, X_n)} \right) > \delta \right\} \right] \xrightarrow{n \rightarrow \infty} 1 \quad (*)$$

Falls $H(\mathbf{Q}_\vartheta | \mathbf{Q}_{\vartheta + \varepsilon}) < \infty$, so folgt (*).

Falls $H(\mathbf{Q}_\vartheta | \mathbf{Q}_{\vartheta + \varepsilon}) = \infty$, aber $\mathbf{Q}_\vartheta(\{\rho_{\vartheta + \varepsilon}(X) = 0\}) = 0$.

Betrachte: $h_c(X) := \log \left(\min \left\{ \frac{\rho_\vartheta}{\rho_{\vartheta + \varepsilon}}, c \right\} \right)$ erfullt $\mathbf{E}_\vartheta[h_c(X)] < \infty$, aber gema dem Satz von der monotonen Konvergenz gilt $\lim_{c \rightarrow \infty} \mathbf{E}_\vartheta[h_c(X)] = \infty$.

$$\mathbf{P}_\vartheta \left[\frac{1}{n} \log \frac{\rho_\vartheta^{\otimes n}(X_1, \dots, X_n)}{\rho_{\vartheta + \varepsilon}^{\otimes n}(X_1, \dots, X_n)} > \delta \right] \geq \mathbf{P}_\vartheta \left[\frac{1}{n} \sum_{i=1}^n h_c(X_i) > \delta \right]$$

$$\rightarrow 1 \text{ fur } n \rightarrow \infty$$

Falls $\mathbf{Q}_\vartheta(\rho_{\vartheta + \varepsilon}(X) = 0) =: a > 0$. Dann ist

$$\mathbf{P}_\vartheta \left[\log \frac{\rho_\vartheta^{\otimes n}(X_1, \dots, X_n)}{\rho_{\vartheta + \varepsilon}^{\otimes n}(X_1, \dots, X_n)} = \infty \right] \geq 1 - (1 - a)^n$$

$$\rightarrow 1 \text{ fur } n \rightarrow \infty.$$

3.9.1. Bemerkung:

(ii) ist sicherlich erfullt, wenn $\vartheta \mapsto \log(\rho(x, \vartheta))$ fur jedes x konkav ist.

3.10. Satz und Definition: relative Entropie (Kullback-Leibler-Abstand)

Fur Mae \mathbf{P}, \mathbf{Q} mit Dichten $\rho(x)$ bzw. $\sigma(x)$ bzgl. des Lebesgue-Maes (oder bzgl. des Zahlmaes, wenn \mathcal{X} abzahlbar ist) heit

$$H(\mathbf{P} | \mathbf{Q}) := \int_{\mathcal{X}} \rho(x) \log \frac{\rho(x)}{\sigma(x)} dx$$

die *relative Entropie von \mathbf{P} bzgl. \mathbf{Q}* , es gilt: $H(\mathbf{P} | \mathbf{Q}) \in [0, \infty]$ und $H(\mathbf{P} | \mathbf{Q}) = 0 \iff \mathbf{P} = \mathbf{Q}$.

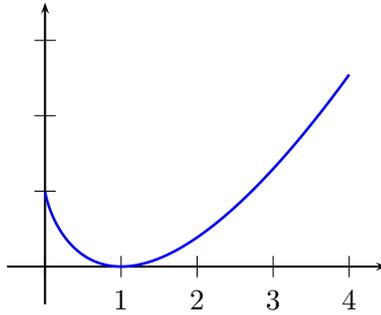


Abbildung 3.1.: Darstellung von $\psi(s) = 1 - s + s \log s$

Die Funktion $\psi(s) := 1 - s + s \log s$ ist konvex, setze $f(x) := \begin{cases} 1 & \text{wenn } \sigma(x) = 0, \\ \frac{\rho(x)}{\sigma(x)} & \text{sonst.} \end{cases}$

Es gilt

$$0 \leq \int \psi(f(x))\sigma(x) dx, \int (1 - f(x))\sigma(x)dx = \int \sigma(x) - \rho(x) dx = 0,$$

folglich

$$\int \psi(f(x))\sigma(x) dx = \int (1 - f(x))\sigma(x)dx + \int \frac{\rho(x)}{\sigma(x)} \log \frac{\rho(x)}{\sigma(x)} \sigma(x) dx,$$

d.h. $H(\mathbf{P}|\mathbf{Q}) \geq 0$ und $H(\mathbf{Q}|\mathbf{Q}) = 0$.

Umgekehrt: Sei $H(\mathbf{P}|\mathbf{Q}) = 0 = \int \psi(f(x))\sigma(x)dx \Rightarrow \mathbf{Q}(\underbrace{\{x : \psi(f(x)) = 0\}}_{=\{x: f(x)=1=\frac{\rho(x)}{\sigma(x)}\}}) = 1$.

3.10.1. Beispiele: Poisson- und Exponentialverteilung

(i) Poissonverteilung: Sei

$\mathbf{E} = \mathbb{Z}_+$, $\mathcal{F} = \mathcal{P}(\mathbb{Z}_+)$, $\vartheta \in (0, \infty)$, $\rho(x, \vartheta) = e^{-\vartheta} \frac{\vartheta^x}{x!}$, $\log \rho(x, \vartheta) = x \log \vartheta - \vartheta - \log(x!)$ ist (strikt) konkav.

Im Produktmodell ist $T_n(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n X_i$ der Maximum-Likelihood-Schätzer für ϑ .

(ii) Exponentialverteilung: Sei $\rho(x, \vartheta) = \vartheta e^{-\vartheta x}$, der ML-Schätzer im Produktmodell ist

$$T_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}.$$

3.11. Satz:

Es gelten die Voraussetzungen des Satzes über die Konsistenz des ML-Schätzers im Produktmodell und

(i) $\vartheta \mapsto \log \rho(x, \vartheta)$ ist dreimal differenzierbar $\forall x \in \mathcal{X}, \vartheta \in \Theta$,

(ii) $\int \frac{d}{d\vartheta} \rho(x, \vartheta) dx = 0 = \int \frac{d^2}{d\vartheta^2} \rho(x, \vartheta) dx \quad \forall \vartheta \in \Theta$,

(iii) $\left| \frac{d^3}{d\vartheta^3} \log \rho(x, \vartheta) \right| \leq M(x)$ für eine Funktion $M : \mathcal{X} \rightarrow \mathbb{R}_+$ mit $\sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta[M(X)] < \infty$.

Dann gilt für den ML-Schätzer T_n im n -fachen Produktmodell:

$$\sqrt{n}(T_n - \vartheta) \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(0, \frac{1}{I(\vartheta)}\right) \text{ in Verteilung unter } \mathbf{P}_\vartheta.$$

3.11.1. Bemerkung: Asymptotische Konfidenzintervalle für den ML-Schätzer:

Für $\alpha > 0$ wähle a so, dass

$$\int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{\alpha}{2}$$

und sei T_n der ML-Schätzer in obigem Szenario.

Dann ist

$$\mathcal{J} := \left[T_n - \frac{a}{\sqrt{nI(T_n)}}, T_n + \frac{a}{\sqrt{nI(T_n)}} \right]$$

ein asymptotisches Konfidenzintervall für ϑ zum Irrtumsniveau α , d.h. für jedes $\vartheta \in \Theta$ gilt

$$\mathbf{P}_\vartheta[\vartheta \in \mathcal{J}] \xrightarrow{n \rightarrow \infty} 1 - \alpha.$$

3.11.2. Beobachtung:

$$\begin{aligned} \mathbf{E}_\vartheta \left[\frac{d^2}{d\vartheta^2} \log \rho(X, \vartheta) \right] &= \int_{\mathcal{X}} \underbrace{\frac{d^2}{d\vartheta^2} (\log \rho(x, \vartheta))}_{\frac{d}{d\vartheta} \left(\frac{\frac{d}{d\vartheta} \rho(x, \vartheta)}{\rho(x, \vartheta)} \right) = \frac{\frac{d^2}{d\vartheta^2} \rho(x, \vartheta)}{\rho(x, \vartheta)} - \left(\frac{\frac{d}{d\vartheta} \rho(x, \vartheta)}{\rho(x, \vartheta)} \right)^2} \rho(x, \vartheta) dx \\ &= \underbrace{\int \frac{d^2}{d\vartheta^2} \rho(x, \vartheta) dx}_{=0} - \underbrace{\int \left(\frac{\frac{d}{d\vartheta} \rho(x, \vartheta)}{\rho(x, \vartheta)} \right)^2 \rho(x, \vartheta) dx}_{=U_\vartheta(x)} \\ &= -\text{Var}_\vartheta(U_\vartheta) = -I(\vartheta) \end{aligned}$$

Beweis des Satzes:

Wir zeigen für $\vartheta \in \Theta$:

$$(*) \quad \sqrt{n}(T_n - \vartheta) \cdot I(\vartheta) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{\frac{d}{d\vartheta} \log \rho(X_i, \vartheta)}_{U_\vartheta(X_i)} \xrightarrow{n \rightarrow \infty} 0 \text{ in Verteilung unter } \mathbf{P}_\vartheta.$$

Aus (*) folgt die Behauptung mit dem ZGWS, denn $\frac{1}{n} \sum U_\vartheta(X_i) \xrightarrow{d} \mathcal{N}(0, I(\vartheta))$ unter \mathbf{P}_ϑ .

Sei $U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) := \frac{d}{d\vartheta} \log \rho_{\vartheta}^{\otimes n}(X_1, \dots, X_n) = \sum_{i=1}^n U_{\vartheta}(X_i)$.

$$\begin{aligned} 0 &= U_{T_n}^{\otimes n}(X_1, \dots, X_n) \\ &= U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) + (T_n - \vartheta) \frac{d}{d\vartheta} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) + \frac{1}{2} (T_n - \vartheta)^2 \frac{d^2}{d\vartheta^2} U_{\tilde{T}_n}^{\otimes n}(X_1, \dots, X_n) \\ &\quad \text{mit einem } \tilde{T}_n \in [\min\{T_n, \vartheta\}, \max\{T_n, \vartheta\}] \text{ (Taylorentwicklung)}. \end{aligned}$$

Es ist

$$\frac{1}{n} \frac{d}{d\vartheta} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\vartheta^2} \log \rho(X_i, \vartheta) \xrightarrow{d} \mathbf{E}_{\vartheta} \left[\frac{d^2}{d\vartheta^2} \log \rho(X_1, \vartheta) \right] = -I(\vartheta)$$

und

$$\left| \frac{1}{n} \frac{d^2}{d\vartheta^2} U_{T_n}^{\otimes n}(X_1, \dots, X_n) \right| = \left| \frac{1}{n} \sum_{i=1}^n \frac{d^3}{d\vartheta^3} \log \rho(X_i, \tilde{T}_n) \right| \leq \frac{1}{n} \sum_{i=1}^n M(X_i),$$

also $\frac{1}{n} \frac{1}{2} (T_n - \vartheta) \frac{d^2}{d\vartheta^2} U_{T_n}^{\otimes n}(X_1, \dots, X_n) \xrightarrow{n \rightarrow \infty} 0$ in Verteilung unter \mathbf{P}_{ϑ} , den (T_n) ist konsistent für ϑ . Weiter gilt

$$\sqrt{n}(T_n - \vartheta) = \frac{-\frac{1}{\sqrt{n}} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n)}{\frac{1}{n} \left(\frac{d}{d\vartheta} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) + \frac{1}{2} (T_n - \vartheta) \frac{d^2}{d\vartheta^2} U_{T_n}^{\otimes n}(X_1, \dots, X_n) \right)},$$

somit

$$\begin{aligned} &\sqrt{n}(T_n - \vartheta) I(\vartheta) - \frac{1}{\sqrt{n}} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n) \\ &= \underbrace{-\frac{1}{\sqrt{n}} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n)}_{\rightarrow \mathcal{N}(0, I(\vartheta))} \dots \\ &\quad \dots \left(1 - \frac{I(\vartheta)}{\underbrace{\frac{1}{n} \frac{d}{d\vartheta} U_{\vartheta}^{\otimes n}(X_1, \dots, X_n)}_{\rightarrow -I(\vartheta)} + \underbrace{\frac{1}{2n} (T_n - \vartheta) \frac{d^2}{d\vartheta^2} U_{T_n}^{\otimes n}(X_1, \dots, X_n)}_{\rightarrow 0}} \right) \\ &\rightarrow 0 \text{ in Verteilung.} \end{aligned}$$

Bemerkung:

Seien Y_n, Z_n ZVN, $Z_n \xrightarrow{d} Z, Y_n \xrightarrow{d} 0$, dann gilt $Y_n \cdot Z_n \xrightarrow{d} 0$.

4. Exkurs: Bedingte Erwartung und Anwendungen

- (o) Diskreter Fall: X, Y reellwertig, mit abzählbarem Wertebereich, gemeinsame Gewichte $\rho_{X,Y}(x, y)$, Marginalgewichte $\rho_X(x) = \sum_y \rho_{X,Y}(x, y)$, $\rho_Y(y) = \sum_x \rho_{X,Y}(x, y)$. Weiter gelte:

$\mathbf{E}[|X|] < \infty$. Sei

$$\varphi(y) := \sum_x \frac{x \rho_{X,Y}(x, y)}{\rho_Y(y)},$$

$\varphi(Y)$ ist in diesem Szenario die bedingte Erwartung von X gegeben Y .

- (i) Seien X, Y reellwertig mit gemeinsamer Dichte $f_{X,Y}$ bzgl. Lebesgue-Maß (d.h. $\mathbf{P}[(X, Y) \in A] = \int_A f_{X,Y}(x, y) d(x, y)$), Marginaldichten f_X , bzw. f_Y , es gelte $\mathbf{E}[|X|] < \infty$.

$$\varphi(y) = \int_{\mathbb{T}} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx,$$

$\varphi(Y)$ würden wir in diesem Szenario plausiblerweise die bedingte Erwartung von X gegeben Y nennen.

- (ii) Allgemeiner Fall: Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, X eine reellwertige ZV auf $(\Omega, \mathcal{F}, \mathbf{P})$ mit endlichem Erwartungswert ($X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$), $\mathcal{G} \subseteq \mathcal{F}$ eine Teil- σ -Algebra.

4.1. Definition: Bedingte Erwartung

Eine reellwertige ZV Z heißt bedingte Erwartung von X gegeben \mathcal{G} (schreibe $\mathbf{E}[X|\mathcal{G}]$), wenn gilt

- (1) Z ist \mathcal{G} -messbar, d.h. $\{Z \in B\} \in \mathcal{G}$ für jede messbare Teilmenge $B \subset \mathbb{R}$,
- (2) $\mathbf{E}[H \cdot Z] = \mathbf{E}[H \cdot X]$ für alle beschränkten \mathcal{G} -messbaren ZVn H .

4.2. Satz: Existenz und Eindeutigkeit der bedingten Erwartung:

Für Y mit $\mathbf{E}[|X|] < \infty$ existiert $\mathbf{E}[X|\mathcal{G}]$ und ist eindeutig (bis auf f.s. Gleichheit)¹.

Beobachtung:

Im Szenario von (o): Betrachte $\mathcal{G} = \sigma(Y) = \{Y^{-1}(B) : B \in \mathbf{P}(E_Y)\}$, mit $E_Y =$ Wertebereich von Y . $Z := \varphi(Y)$ ist $\sigma(Y)$ -messbar. Sei $H = h(Y)$ (Übung: Jede $\sigma(Y)$ -messbare ZV ist hier von dieser Form). Dann gilt

¹Wir fassen X und X' als (fast sicher) gleich auf, wenn $\mathbf{P}[X \neq X'] = 0$.

$$\begin{aligned} \mathbf{E}[H \cdot Z] &= \sum_y h(y) \underbrace{\varphi(y)}_{\rho_Y(y)} \rho_Y(y) = \sum_{x,y} h(y) x \rho_{X,Y}(x,y) = \mathbf{E}[H \cdot X]. \\ &= \frac{1}{\rho_Y(y)} \sum_x x \rho_{X,Y}(x,y) \end{aligned}$$

Beweis:

Eindeutigkeit: Seien Z, Z' bedingte Erwartungen.

$$\begin{aligned} \mathbf{E}[Z \cdot \mathbf{1}_{\{Z > Z'\}}] &= \mathbf{E}[X \cdot \mathbf{1}_{\{Z > Z'\}}] = \mathbf{E}[Z' \mathbf{1}_{\{Z > Z'\}}] \\ &\Rightarrow \mathbf{E}[(Z - Z') \cdot \mathbf{1}_{\{Z > Z'\}}] = 0 \quad \Rightarrow \mathbf{P}[Z > Z'] = 0, \\ \text{analog } \mathbf{P}[Z < Z'] &= 0 \Rightarrow \mathbf{P}[Z = Z'] = 1 \Rightarrow Z = Z' \text{ fast sicher.} \end{aligned}$$

Existenz: Betrachte zunächst X mit $\mathbf{E}[X^2] < \infty$.

Bemerkung: $\mathcal{H} = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ ist ein Hilbertraum mit Skalarprodukt $\langle X, Y \rangle = \mathbf{E}[X \cdot Y]$.

Betrachte nun $\mathcal{H}' := \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P}) \subseteq \mathcal{H}$ ist ein Unterraum, $Z := \text{pr}_{\mathcal{H}'}(X)$, d.h. Z ist dasjenige Element von \mathcal{H}' mit $\mathbf{E}[(X - Z)^2] \leq \mathbf{E}[(X - Z')^2] \forall Z' \in \mathcal{H}'$.

Sei $H \in \mathcal{H}'$ (beschränkt), dann ist $\langle Z - X, H \rangle = 0 = \mathbf{E}[(Z - X)H] \Rightarrow$ (ii) \checkmark .

Rücktransfer nach \mathcal{L}^1 : Für $X \in \mathcal{L}^1$ zerlege $X = X_+ - X_-$, $\mathbf{E}[(\min\{X_+, n\})|\mathcal{G}] =: Z_+$ existiert (Monotonie). Dann folgt

$$\mathbf{E}[X|\mathcal{G}] = Z_+ - Z_-.$$

4.3. Eigenschaften der bedingten Erwartung:

- (i) $\mathbf{E}[aX + bY|\mathcal{G}] = a\mathbf{E}[X|\mathcal{G}] + b\mathbf{E}[Y|\mathcal{G}]$ für $a, b \in \mathbb{R}, X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$
- (ii) $X \geq Y$ f.s. $\Rightarrow \mathbf{E}[X|\mathcal{G}] \geq \mathbf{E}[Y|\mathcal{G}]$
- (iii) $\mathbf{E}[|XY|] < \infty, Y$ \mathcal{G} -messbar $\Rightarrow \mathbf{E}[XY|\mathcal{G}] = Y \cdot \mathbf{E}[X|\mathcal{G}]$
- (iv) Sind X und \mathcal{G} unabhängig ($\mathbf{P}[\{X \in B\} \cap A] = \mathbf{P}[\{X \in B\}]\mathbf{P}[A] \forall A \in \mathcal{G}, B \in \mathcal{B}(\mathbb{R})$), so ist $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.
- (v) $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]] = \mathbf{E}[X]$

Beweis:

$$(ii) A = \{\mathbf{E}[X|\mathcal{G}] < \mathbf{E}[Y|\mathcal{G}]\} \in \mathcal{G},$$

$$0 \leq \mathbf{E}[(X - Y) \cdot \mathbf{1}_A] = \mathbf{E}[(\mathbf{E}[X|\mathcal{G}] - \mathbf{E}[Y|\mathcal{G}])\mathbf{1}_A]$$

$$\Rightarrow \mathbf{P}[A] = 0$$

(iii) Betrachte zunächst Y beschränkt, dann gilt

$$\mathbf{E}[\overbrace{H \cdot Y}^{=H'} \mathbf{E}[X|\mathcal{G}]] = \mathbf{E}[XH'] = \mathbf{E}[(XY)H]$$

für jedes beschränkte \mathcal{G} -messbare H .

Allgemeiner Fall via Approximation und monotone Konvergenz.

4.3.1. Bemerkung: Zur Existenz der regulären bedingten Verteilung

Für $A \in \mathcal{F}$ ist $\mathbf{P}[A|\mathcal{G}] := \mathbf{E}[\mathbb{1}_A|\mathcal{G}]$ die bedingte Wahrscheinlichkeit von A gegeben \mathcal{G} . Im Allgemeinen ist $\mathbf{P}[A|\mathcal{G}]$ nur bis auf \mathbf{P} -Nullmenge definiert, und es ist zunächst unklar, ob die (u.U. überabzählbar vielen) Ausnahme-Nullmengen, die sich ergeben, wenn man A über ganz \mathcal{F} variiert, kompatibel gewählt werden können. Insbesondere wünschen wir uns

$$\mathbf{P} \left[\bigcup_n A_n \middle| \mathcal{G} \right] = \sum_n \mathbf{P}[A_n|\mathcal{G}]$$

für jede abzählbare Familie $(A_n) \subset \mathcal{F}$ von (paarw.) disjunkten A_n .

Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ messbare Räume, $k : \Omega_1 \times \mathcal{F}_2 \rightarrow [0, 1]$ heißt *stochastischer Kern*, wenn

- (i) $\omega_1 \mapsto k(\omega_1, B)$ messbar ist $\forall B \in \mathcal{F}_2$ und
- (ii) $\forall \omega_1 \in \Omega_1 : k(\omega_1, \cdot)$ ein Wahrscheinlichkeitsmaß auf Ω_2 ist.

Sei $(\Omega, \mathcal{A}, \mathbf{P})$, Y reellwertige ZV auf $(\Omega, \mathcal{A}, \mathbf{P})$, $\mathcal{F} \subseteq \mathcal{A}$ Teil- σ -Algebra, dann gibt es einen stochastischen Kern $k_{Y|\mathcal{F}}$ von (Ω, \mathcal{F}) nach $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

$$k_{Y|\mathcal{F}}(\omega, B) = \mathbf{E}[\mathbb{1}_B(Y)|\mathcal{F}], \quad B \in \mathcal{B}(\mathbb{R})$$

für \mathbf{P} -f.a. ω .

Wir konstruieren dazu eine zufällige Verteilungsfunktion. Betrachte $B = (-\infty, r], r \in \mathbb{Q}$, setze

$$F_r := \mathbf{E}[\mathbb{1}_{(-\infty, r]}(Y)|\mathcal{F}].$$

Dann gilt wie gewünscht \mathbf{P} -f.s.: $F_r \leq F_{r'}$, für $r < r', (r, r' \in \mathbb{Q}), \lim_n F_{r+\frac{1}{n}} = F_r$ für $r \in \mathbb{Q}$,

$$\lim_{n \rightarrow \infty} F_n = 1, \quad \lim_{n \rightarrow -\infty} F_n = 0.$$

Wegen der Abzählbarkeit von \mathbb{Q} gibt es $N \in \mathcal{A}$ mit $\mathbf{P}(N) = 0$, so dass obiges für $\omega \in \Omega \setminus N$ und alle $r, r' \in \mathbb{Q}$ gilt. Dann definiert

$$\tilde{F}_s := \begin{cases} \inf\{F_r : r \geq s, r \in \mathbb{Q}\}, & \omega \in \Omega \setminus N, \\ \bar{F}_s, & \omega \in N, \end{cases}$$

wobei \bar{F} irgendeine Verteilungsfunktion ist, die (zufällige) Verteilungsfunktion von $k_{Y|\mathcal{F}}$. Details finden sich beispielsweise in A. Klenke, *Wahrscheinlichkeitstheorie*, Abschn. 8.28, oder in Übungsaufgabe 3 auf Blatt 3, in der wir den allgemeinen Fall polnischer Wertebereiche behandeln haben.

4.4. Suffizienz:

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta, \vartheta \in \Theta)$ ein statistisches (Standard-)Modell. Eine Statistik $T : \mathcal{X} \rightarrow \Sigma$ heißt *suffizient*, wenn $\mathbf{P}_\vartheta[\cdot | T = t] = \mathbf{Q}_t(\cdot)$ für alle $\vartheta \in \Theta$ und t aus dem Wertebereich von T (d.h. wenn die Verteilung von X gegeben T nicht von ϑ abhängt).

Bemerkung:

Für ein Standardmodell mit Likelihood funktion $\rho(x, \vartheta)$ ist T suffizient \iff
 $\rho(x, \vartheta) = h(x) \cdot \Psi_{\vartheta}(T(x))$, für ein $h : \mathcal{X} \rightarrow \mathbb{R}_+$, $\Psi(\cdot) : \Theta \times \Sigma \rightarrow \mathbb{R}_+$. In einer exponentiellen Familie
mit Likelihood-Funktion $\rho(x, \vartheta) = h(x) \cdot \exp(a(\vartheta)T(x) - b(\vartheta))$ ist T eine suffiziente Statistik.

4.5. Satz von Rao-Blackwell

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_{\vartheta} : \vartheta \in \Theta)$ ein statistisches Modell, T eine suffiziente Statistik, S ein erwartungstreuer
Schätzer für $\tau(= \tau(\vartheta))$, $\mathbf{E}_{\vartheta}[S^2] < \infty$, $g_S(t) := \mathbf{E}_{\vartheta}[S|T = t] = \mathbf{E}_{\mathbf{Q}_t}[S]$ und .

Dann ist $\tilde{S} = g_S(T)$ ein erwartungstreuer Schätzer für $\tau(\vartheta)$ mit $\text{Var}_{\vartheta}(\tilde{S}) \leq \text{Var}_{\vartheta}(S) \forall \vartheta \in \Theta$
(Ungleichung ist strikt, es sei denn S ist eine Funktion von T).

Beweis:

$$\mathbf{E}_{\vartheta}[\tilde{S}] = \mathbf{E}_{\vartheta}[\mathbf{E}_{\vartheta}[S|T]] = \mathbf{E}_{\vartheta}[S] = \tau(\vartheta)$$

Es gilt allgemein:

$$\begin{aligned} \text{Var}(X) &= \underbrace{\text{Var}(\mathbf{E}[X|\mathcal{F}])}_{= \mathbf{E}[(\mathbf{E}[X|\mathcal{F}])^2] - \mathbf{E}[X]^2} + \mathbf{E}[\underbrace{V(X|\mathcal{F})}_{:= \mathbf{E}[X^2|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]^2}] \\ &= \mathbf{E}[\mathbf{E}[X^2|\mathcal{F}]] - \mathbf{E}[X]^2 \\ &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \end{aligned}$$

Demnach ist

$$\text{Var}_{\vartheta}(S) = \text{Var}_{\vartheta}(\overbrace{\mathbf{E}_{\vartheta}[S|T]}^{\tilde{S}}) + \underbrace{\mathbf{E}[\text{Var}_{\vartheta}(S|T)]}_{\geq 0} \Rightarrow \text{Var}_{\vartheta}(S) \geq \text{Var}_{\vartheta}(\tilde{S})$$

4.6. Definition: Vollständigkeit

Eine Statistik T heißt vollständig, wenn für alle messbaren $g : \Sigma \rightarrow \mathbb{R}$ mit $\mathbf{E}_{\vartheta}[|g(T)|] < \infty$ aus
 $\mathbf{E}_{\vartheta}[g(T)] = 0 \forall \vartheta \in \Theta$ folgt $g \equiv 0$ \mathbf{P}_{ϑ} -f.s. $\forall \vartheta$.

4.6.1. Beispiel:

Seien X_1, \dots, X_n i.i.d. $Ber(\vartheta), \vartheta \in [0, 1]$. Dann ist $T_n = \sum_{i=1}^n \frac{X_i}{n}$ ist vollständig.

4.6.2. Bemerkung: Eindeutigkeit der Rao-Blackwellisierung

Für ein vollständiges T ist die „Rao-Blackwellisierung“ erwartungstreuer Schätzer für $\tau(\vartheta)$
eindeutig.

4.7. Satz von Lehmann-Scheffé

Sei T suffizient und vollständig, S, S' erwartungstreue Schätzer für $\tau(\vartheta)$, dann gilt
 $\tilde{S} = \mathbf{E}_{\vartheta}[S|T] = \mathbf{E}_{\vartheta}[S'|T]$ und \tilde{S} ist der beste Schätzer für $\tau(\vartheta)$.

Beweis:

$g(t) = \mathbf{E}_{\mathbf{Q}_t}[S] - \mathbf{E}_{\mathbf{Q}_t}[S'], \mathbf{E}_{\vartheta}[g(T)] = \tau(\vartheta) - \tau(\vartheta) = 0$, d.h. $g \equiv 0$. Nach dem Satz von Rao-Blackwell gilt $\text{Var}_{\vartheta}(\tilde{S}) \leq \text{Var}_{\vartheta}(S) \forall \vartheta \forall$ erwartungstreue Schätzer S .

4.7.1. Bemerkung:

In einer (einparametrischen) exponentiellen Familie ($\rho(x, \vartheta) = h(x) \cdot \exp(a(\vartheta)T(x) - b(\vartheta))$) ist T eine vollständige und suffiziente Statistik.

Zur Vollständigkeit: (o.B.d.A. $a(\vartheta) = \vartheta$)

$$0 = \int_{\mathcal{X}} g(T(x))h(x)e^{\vartheta T(x) - b(\vartheta)} dx = \int_{\Sigma} g(t)e^{\vartheta t} \nu(dt) = \int_{\Sigma} g_+(t)e^{\vartheta t} \nu(dt) - \int_{\Sigma} g_-(t)e^{\vartheta t} \nu(dt),$$

wobei ν das Bildmaß von $h(x)dx$ ist unter der Abbildung $x \mapsto T(x)$ ist, und $g = g_+ - g_-$ mit $g_+, g_- \geq 0$. Die Eindeutigkeit der Laplace-Transformierten erzwingt dann $g_+ = g_-$, also $g = 0$, für Details siehe z.B. Erich Lehmann, *Testing statistical hypotheses*, S. 132.

5. Bayes-Schätzer

Erinnerung (Satz von Bayes):

Für Ereignisse A, B , mit $\mathbf{P}[A], \mathbf{P}[B] > 0$ gilt

$$\mathbf{P}[A|B] = \frac{\mathbf{P}[B|A] \cdot \mathbf{P}[A]}{\mathbf{P}[B]}.$$

5.1. Definition: A-priori-Dichte und -Maß

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta \subseteq \mathbb{R}^d)$ ein parametrisches Standardmodell mit $\mathcal{F} \otimes \mathcal{B}(\Theta)$ -messbarer Likelihood-Funktion $\rho(x, \vartheta)$. Wir nennen eine Wahrscheinlichkeitsdichte $\alpha : \Theta \rightarrow \mathbb{R}_+$ eine *A-priori-Dichte*. (Das Maß mit Dichte α heißt *A-priori-Maß* oder *Vorbewertung*)

$$\pi_x(\vartheta) := \frac{\alpha(\vartheta)\rho(x, \vartheta)}{\int_{\Theta} \alpha(t)\rho(x, t) dt}$$

heißt dann die *A-posteriori-Dichte*.

Sei $\tau : \Theta \rightarrow \mathbb{R}$ (mit $\int (\tau(\vartheta))^2 \alpha(\vartheta) d\vartheta < \infty$). Dann heißt

$$T(x) := \mathbf{E}_{\pi_x}[\tau] = \int \tau(\vartheta)\pi_x(\vartheta) d\vartheta$$

Bayes-Schätzer für τ (zur Vorbewertung $\alpha(\vartheta)d\vartheta$).

$$F_\alpha(S) := \int_{\Theta} \int_{\mathcal{X}} (S(x) - \tau(\vartheta))^2 \rho(x, \vartheta) dx \alpha(\vartheta) d\vartheta$$

heißt der *erwartete quadratische Fehler* (zur Vorbewertung $\alpha(\vartheta)d\vartheta$).

Beispiel:

Eine Situation, in der diese Philosophie sinnvoll ist, könnte folgende sein: Nehmen wir an, ein Versicherungsnehmer hat jedes Jahr mit einer gewissen (zu ihm „gehörigen“) Wahrscheinlichkeit ϑ einen Schadensfall (unabhängig über die Jahre), und $\alpha(\vartheta)d\vartheta$ beschreibt die Verteilung der Schadenswahrscheinlichkeiten aller Kunden dieser Versicherung (diese Verteilung sei der Versicherung aus Erfahrungswerten bekannt).

Mit $\rho(x, \vartheta) = \text{Bin}_{n, \vartheta}(x)$ ist dann die Wahrscheinlichkeit, dass ein „typischer Kunde“ in n Jahren k Schadensfälle verursacht $\int_0^1 \alpha(\vartheta)\rho(k, \vartheta) d\vartheta$, und $\pi_k(\vartheta)$ ist die Verteilung der Schadenswahrscheinlichkeit pro Jahr eines Kunden, bedingt darauf, dass er in den letzten n Jahren k Schäden hatte. Diese Information kann die Versicherung beispielsweise für Vertragsverlängerung, Tarifanpassung, etc. benutzen.

5.2. Satz:

Der Bayes-Schätzer T für $\tau(\vartheta)$ minimiert den erwarteten quadratischen Fehler (jeweils zur Vorbewertung $\alpha(\vartheta)d\vartheta$).

Beweis:

$$\begin{aligned}
 F_\alpha(S) - F_\alpha(T) &= \int_{\Theta} \int_{\mathcal{X}} \left[\underbrace{(S(x) - \tau(\vartheta))^2 - (T(x) - \tau(\vartheta))^2}_{=S(x)^2 - 2S(x)\tau(\vartheta) - T(x)^2 + 2T(x)\tau(\vartheta)} \right] \underbrace{\int_{\Theta} \alpha(t) \rho(x, t) dt}_{=: \rho_\alpha(x)} \rho(x, \vartheta) \alpha(\vartheta) \, dx d\vartheta \\
 &\stackrel{\text{Fubini}}{=} \int_{\mathcal{X}} \int_{\Theta} (S(x)^2 - 2S(x)\tau(\vartheta) - T(x)^2 + 2T(x)\tau(\vartheta)) \pi_x(\vartheta) d\vartheta \rho_\alpha(x) dx \\
 &= \int_{\mathcal{X}} \underbrace{(S(x)^2 - 2S(x)T(x) - T(x)^2)}_{=(S(x)-T(x))^2} \rho_\alpha(x) dx \\
 &\geq 0.
 \end{aligned}$$

6. Konfidenzintervalle:

6.1. Definition: Konfidenzbereich zum Irrtumsniveau α

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$ ein statistisches Modell, $\tau : \Theta \rightarrow \Sigma$ ein Parametermerkmal, $\alpha \in (0, 1)$. Eine Abbildung $C : \mathcal{X} \rightarrow \mathcal{P}(\Sigma)$ heißt ein *Konfidenzbereich* zum *Irrtumsniveau* α (oder zum *Sicherheitsniveau* $1 - \alpha$), wenn

$$\forall \vartheta \in \Theta : \mathbf{P}_\vartheta[\{x : C(x) \ni \tau(\vartheta)\}] \geq 1 - \alpha.$$

Bemerkung:

Stets ist $C(x) = \Sigma$ ein (trivialer) Konfidenzbereich.

6.1.1. Beispiel: Konfidenzintervalle im Binomialmodell

Wir betrachten das statistische Modell $(\mathcal{X} = \{0, 1, \dots, n\}, \mathcal{F} = \mathcal{P}(\mathcal{X}), \mathbf{P}_\vartheta = \text{Bin}_{n, \vartheta})$.

Erinnerung:

$T = \frac{X}{n}$ ist der beste Schätzer für ϑ , wenn X die Anzahl der Einsen zählt.

(i) Anwendung der Čebyšev-Ungleichung, Ansatz:

$$C(x) = \left(\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon \right).$$

$$\mathbf{P}_\vartheta[\{C(X) \ni \vartheta\}] = \mathbf{P}_\vartheta \left[\left| \frac{X}{n} - \vartheta \right| > \varepsilon \right],$$

$$1 - \mathbf{P}_\vartheta[\{C(X) \ni \vartheta\}] = \mathbf{P}_\vartheta \left[\left| \frac{X}{n} - \vartheta \right| > \varepsilon \right] \stackrel{\text{Chebyshev}}{\leq} \frac{1}{\varepsilon^2} \text{Var}_\vartheta \left(\frac{X}{n} \right) = \frac{\vartheta(1-\vartheta)}{\varepsilon^2 n} \stackrel{\sup_{\vartheta \in [0,1]} \vartheta(1-\vartheta) = \frac{1}{4}}{\leq} \frac{1}{4\varepsilon^2 n} \stackrel{!}{=} \alpha,$$

d.h. wähle $\varepsilon = \frac{1}{\sqrt{4n\alpha}}$.

(ii) Normalapproximation: $(\frac{X}{n} \approx \mathcal{N}(\vartheta, \frac{\vartheta(1-\vartheta)}{n}))$.

$$\begin{aligned} \mathbf{P}_\vartheta \left[\left| \frac{X}{n} - \vartheta \right| < \varepsilon \right] &\approx \mathbf{P}_\vartheta \left[\left| \sqrt{\frac{\vartheta(1-\vartheta)}{n}} Z \right| < \varepsilon \right] \quad \text{mit } Z \sim \mathcal{N}(0, 1) \\ &= \Phi \left(\varepsilon \cdot \sqrt{\frac{\vartheta(1-\vartheta)}{n}} \right) - \Phi \left(-\varepsilon \cdot \sqrt{\frac{\vartheta(1-\vartheta)}{n}} \right) \\ &= 2\Phi \left(\varepsilon \sqrt{\frac{\vartheta(1-\vartheta)}{n}} \right) - 1, \end{aligned}$$

wähle $\varepsilon \geq \sqrt{\frac{\vartheta(1-\vartheta)}{n}} \Phi^{-1}(1 - \frac{\alpha}{2})$ (praktisch: $\varepsilon := \frac{1}{2\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2})$).

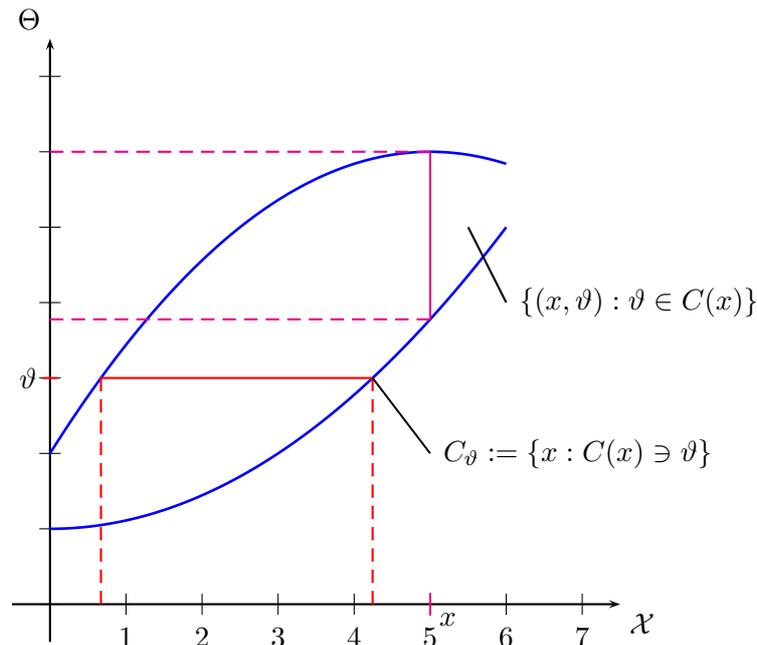


Abbildung 6.1.: Darstellung von Konfidenzbereichen in der x - ϑ -Ebene: Es soll $\mathbf{P}_\vartheta[C_\vartheta] \geq 1 - \alpha$ gelten.

Für ein Standardmodell (mit Likelihood-Funktion $\rho(x, \vartheta)$) setze $C_\vartheta := \{x : \rho(x, \vartheta) \geq c_\vartheta\}$ für c_ϑ so groß wie möglich, sodass immer noch $\mathbf{P}_\vartheta[C_\vartheta] \geq 1 - \alpha$ erfüllt ist.

6.2. Definition: Quantil, Fraktile und Median

Sei μ ein Wahrscheinlichkeitsmaß auf \mathbb{R} , $\alpha \in (0, 1)$. $q \in \mathbb{R}$ heißt α -Quantil von μ , wenn $\mu((-\infty, q)) \geq \alpha$ und $\mu([q, \infty)) \geq 1 - \alpha$ (eindeutig, wenn die Verteilungsfunktion F_μ strikt monoton wächst). Sprechweise: Ein $(1 - \alpha)$ -Quantil heißt auch ein α -Fraktile, ein $\frac{1}{2}$ -Quantil heißt auch *Median*.

Fortsetzung des Beispiels zum Binomialmodell:

(iii) Exakte Konfidenzintervalle im Binomialmodell:

6.2.1. Lemma:

- (1) Für $\vartheta \in (0, 1)$ ist $\{0, \dots, n\} \ni x \mapsto \text{Bin}_{n, \vartheta}(\{x\})$ strikt wachsend auf $\{0, 1, \dots, \lceil (n+1)\vartheta - 1 \rceil\}$, strikt fallend auf $\{\lfloor (n+1)\vartheta \rfloor, \dots, n\}$, also maximal auf $x = \lfloor (n+1)\vartheta \rfloor$ (und auf $(n+1)\vartheta - 1$, wenn $(n+1)\vartheta \in \mathbb{Z}$).
- (2) Für $x \in \{1, \dots, n\}$ ist $[0, 1] \ni \vartheta \mapsto \text{Bin}_{n, \vartheta}(\{x, x+1, \dots, n\})$ stetig, strikt monoton wachsend mit

$$\text{Bin}_{n, \vartheta}(\{x, x+1, \dots, n\}) = \mathcal{B}_{x, n-x+1}([0, \vartheta]),$$

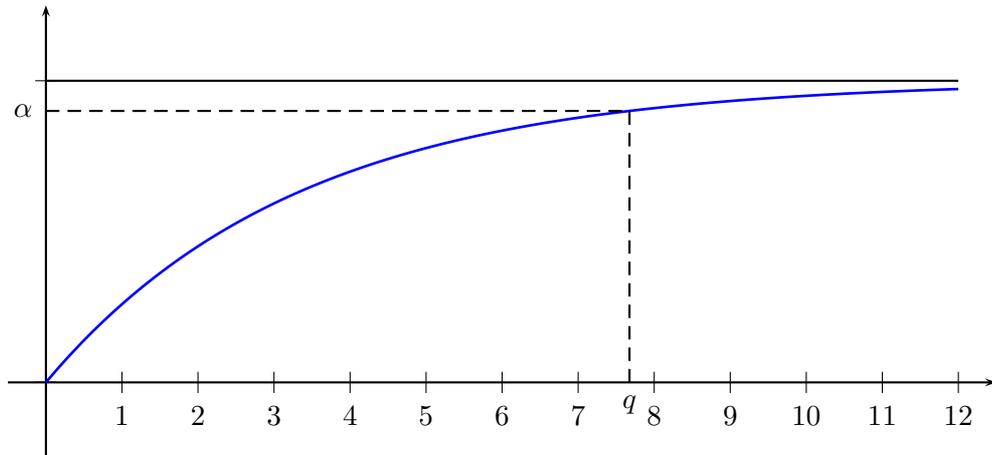


Abbildung 6.2.: Darstellung des α -Quantils q einer Verteilung mit stetiger, strikt monoton wachsender Verteilungsfunktion (die blaue Kurve)

wo $\mathcal{B}_{a,b}$ (die Beta-Verteilung) die Dichte

$$f_{\mathcal{B}_{a,b}}(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}$$

auf $(0,1)$ hat.

Beweis:

(1)

$$\begin{aligned} \frac{Bin_{n\vartheta}(\{x\})}{Bin_{n\vartheta}(\{x-1\})} &= \frac{\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}}{\binom{n}{x-1}\vartheta^{x-1}(1-\vartheta)^{n-x+1}} \\ &= \frac{(n-x+1)\vartheta}{x(1-\vartheta)} \\ &> 1 \\ &\iff x < (n+1)\vartheta \end{aligned}$$

(2) U_1, \dots, U_n unabhängig und uniform auf $[0,1]$,

$$S_\vartheta := \sum_{i=1}^n \mathbf{1}_{[0,\vartheta]}(U_i)$$

ist $Bin_{n,\vartheta}$ -verteilt.

Sei $U_{(1)} < U_{(2)} < \dots < U_{(n)}$ die „Ordnungsstatistik“.

$$\begin{aligned}
& \text{Bin}_{n,\vartheta}(\{x, \dots, n\}) \\
&= \mathbf{P}[S_\vartheta \geq x] \\
&= \mathbf{P}[U_{(x)} \leq \vartheta] \\
&= \sum_{k=1}^n \left(\sum_{\substack{B \subseteq \{1, \dots, n\} \setminus \{k\} \\ |B|=x-1}} \underbrace{\mathbf{P}[U_k \leq \vartheta, U_m \leq U_k \text{ für } m \in B, U_l > U_k \text{ für } l \in \{1, \dots, n\} \setminus (\{k\} \cup B)]}_{= \int_0^\vartheta u^{|B|} (1-u)^{n-|B|-1} du = \int_0^\vartheta u^{x-1} (1-u)^{n-x} du} \right) \\
&= \frac{n!}{(x-1)!(n-x)!} \frac{\Gamma(n+1)}{\Gamma(x)\Gamma(n-x+1)} \int_0^\vartheta u^{x-1} (1-u)^{n-x} du
\end{aligned}$$

Zurück zum Binomialmodell:

Wähle $C_\vartheta := \{x_-(\vartheta), x_-(\vartheta) + 1, \dots, x_+(\vartheta)\}$ mit
 $x_-(\vartheta) = \max\{x : \text{Bin}_{n,\vartheta}(\{0, \dots, x-1\}) \leq \frac{\alpha}{2}\}$ und
 $x_+(\vartheta) = \min\{x : \text{Bin}_{n,\vartheta}(\{x+1, \dots, n\}) \leq \frac{\alpha}{2}\}$.

Es gilt:

- $x \leq x_+(\vartheta) \iff \text{Bin}_{n,\vartheta}(\{x, \dots, n\}) = \mathcal{B}_{x,n-x+1}([0, \vartheta]) > \frac{\alpha}{2}$
 $\iff \vartheta > p_-(x) := \frac{\alpha}{2}\text{-Quantil von } \mathcal{B}_{x,n-x+1}$.
- $x \geq x_+(\vartheta) \iff \text{Bin}_{n,\vartheta}(\{0, \dots, x\}) = 1 - \text{Bin}(\{x+1, \dots, n\}) = \mathcal{B}_{x+1,n-x}([\vartheta, 1]) \geq \frac{\alpha}{2}$
 $\iff \vartheta < p_+(x) := \frac{\alpha}{2}\text{-Fraktile von } \mathcal{B}_{x+1,n-x}$.

6.2.2. Satz: exaktes Konfidenzintervall im Binomialmodell

Die Abbildung $x \mapsto (p_-(x), p_+(x))$ definiert im Binomialmodell ein Konfidenzintervall für ϑ zum Irrtumsniveau α .

Siehe auch Abbildung 6.3 für Illustration und Vergleich der in diesem Abschnitt betrachteten Konfidenzintervalle.

6.3. Konfidenzintervalle im Gauß'schen Modell

Sei X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(m, v)$, $M := \frac{1}{n} \sum_{i=1}^n X_i$, $V^* := \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2$ sind erwartungstreue

Schätzer für m bzw. v .

Ansatz für ein Konfidenzintervall für m : $(M \pm a\sqrt{V^*})$.

Behauptung:

$M - m$ und V^* sind unabhängig, und für jedes (m, v) ist $\frac{\sqrt{n}(M - m)}{\sqrt{V^*}}$ Student-verteilt mit $n - 1$ Freiheitsgraden („ t_{n-1} -verteilt“).

6.3.1. Satz:

Sei $\alpha \in (0, 1)$, $t := \frac{\alpha}{2}$ -Fraktil der t_{n-1} -Verteilung. Dann ist

$$\left(M - t\sqrt{\frac{V^*}{n}}, M + t\sqrt{\frac{V^*}{n}}\right)$$

ein Konfidenzintervall für m zum Sicherheitsniveau $1 - \alpha$.

Beweis:

Klar, denn

$$\mathbf{P}_{m,v} \left[\left| \frac{\sqrt{n}(M - m)}{\sqrt{V^*}} \right| \geq t \right] = \alpha.$$

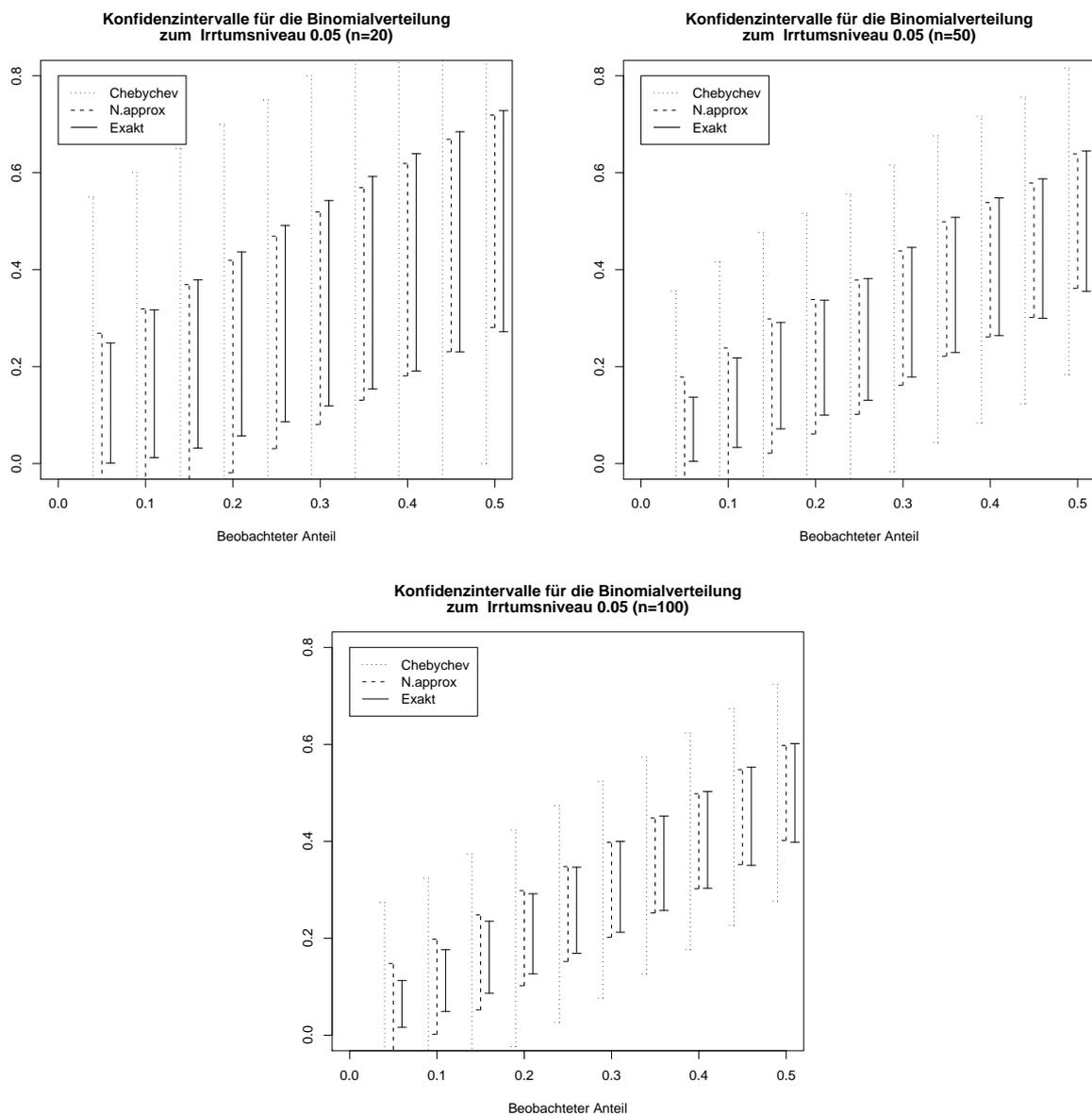


Abbildung 6.3.: Konfidenzintervalle zum Niveau 0.05 für den Erfolgsparameter im Binomialmodell, basierend auf Chebyshev-Ungleichung, Normalapproximation und auf Satz 6.2.2, als Funktion des beobachteten Anteils der Erfolge, für $n \in \{20, 50, 100\}$. Siehe Abschnitt A.1 für den R-Code zur Erzeugung dieser Bilder.

7. Exkurs: mehrdimensionale Normalverteilung

7.1. Dichtentransformationssatz:

Sei $\mathcal{X} \subseteq \mathbb{R}^n$ offen, P ein Wahrscheinlichkeitsmaß auf \mathcal{X} mit Dichte ρ , $\mathcal{Y} \subseteq \mathbb{R}^n$ offen und $T : \mathcal{X} \rightarrow \mathcal{Y}$ Diffeomorphismus¹. Sei

$$DT = \left(\frac{\partial T_j}{\partial x_i} \right) \text{ die Jacobi-Matrix, } \det DT \neq 0 \text{ auf } \mathcal{X}.$$

Dann hat $\mathbf{Q} := P \circ T^{-1}$ die Dichte $|\det DT^{-1}(y)|\rho(T^{-1}(y))$

7.2. Erinnerung: $\Gamma_{a,b}$ -Verteilung

Sei $\Gamma_{a,b}$ die Gamma-Verteilung mit Skalenparameter a und Formparameter b . Sie hat die Dichte

$$\frac{a^b}{\Gamma(b)} x^{b-1} e^{-ax} \text{ auf } (0, \infty)$$

Bemerkung: $X \sim \Gamma_{a,b} \Rightarrow cX \sim \Gamma_{\frac{a}{c}, b}$

Beobachtung:

- $X \sim \mathcal{N}(0, 1)$, so ist $X^2 \sim \Gamma_{\frac{1}{2}, \frac{1}{2}}$ (diese Verteilung heißt auch χ_1^2 -Verteilung).
- $|X|$ hat Dichte

$$\frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ auf } (0, \infty).$$

Betrachte $T : (0, \infty) \rightarrow (0, \infty), x \mapsto x^2$. Dann ist

$$T^{-1}(y) = \sqrt{y}, \frac{d}{dy} T^{-1}(y) = \frac{1}{2\sqrt{y}},$$

also hat X^2 die Dichte

$$\frac{1}{2\sqrt{y}} \frac{2}{\sqrt{2\pi}} e^{-\frac{y}{2}} = \left(\frac{1}{2}\right)^{\frac{1}{2}} \frac{1}{\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}.$$

7.2.1. Proposition:

Sei $a, r, s > 0$. $X \sim \Gamma_{a,r}, Y \sim \Gamma_{a,s}$ unabhängig. Dann sind $X + Y$ und $\frac{X}{X + Y}$ unabhängig und

$$X + Y \sim \Gamma_{a,r+s}, \frac{X}{X + Y} \sim \mathcal{B}_{r,s}^2$$

¹d.h. bijektiv, T und T^{-1} stetig differenzierbar.

²Die Beta-Verteilung $\mathcal{B}_{r,s}$ hat die Dichte

$$\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} u^{r-1} (1-u)^{s-1}.$$

Beweis:

(X, Y) hat Dichte

$$\frac{a^{r+s}}{\Gamma(r)\Gamma(s)} x^{r-1} y^{s-1} e^{-a(x+y)} \text{ auf } (0, \infty)^2.$$

Sei

$$T(x, y) = \begin{pmatrix} x+y \\ \frac{x}{x+y} \end{pmatrix}, T^{-1}(z, u) = \begin{pmatrix} zu \\ z(1-u) \end{pmatrix}, DT^{-1}(z, u) = \begin{pmatrix} u & (1-u) \\ z & -z \end{pmatrix}, |\det DT^{-1}(z, u)| = z$$

Dann ist mit Satz 7.1 und $Z := X + Y, U := \frac{X}{X+Y}$ die Dichte von (Z, U) :

$$\begin{aligned} f_{(Z,U)} &= z \cdot \frac{a^{r+s}}{\Gamma(r)\Gamma(s)} (zu)^{r-1} (z(1-u))^{s-1} e^{-az} \\ &= \underbrace{\frac{a^{r+s}}{\Gamma(r+s)} z^{r+s-1} e^{-az}}_{\text{Dichte der } \Gamma\text{-Verteilung}} \underbrace{\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} u^{r-1} (1-u)^{s-1}}_{\text{Dichte der } \mathcal{B}\text{-Verteilung}} \end{aligned}$$

7.2.2. Korollar:

- (i) $X \sim \Gamma_{a,r}, Y \sim \Gamma_{a,s}$ unabhängig, so ist $X + Y \sim \Gamma_{a,r+s}$.
- (ii) X_1, \dots, X_n i.i.d., $\sim \mathcal{N}(0, 1)$, so ist $\sum_{i=1}^n X_i^2 \sim \Gamma_{\frac{1}{2}, \frac{n}{2}}$ (χ^2 -Verteilung mit n Freiheitsgraden, χ_n^2 -Verteilung).

7.2.3. Proposition: Fisher-Verteilung mit m und n Freiheitsgraden

Seien $X_1, \dots, X_m, Y_1, \dots, Y_n$ unabhängig Standard-normalverteilt. Dann hat

$$F_{m,n} := \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2}$$

die Dichte

$$f_{m,n}(x) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{(m+n)}{2}}}.$$

Beweis:

Sei $X := \sum_{i=1}^m X_i^2 \sim \Gamma_{\frac{1}{2}, \frac{m}{2}}, Y := \sum_{j=1}^n Y_j^2 \sim \Gamma_{\frac{1}{2}, \frac{n}{2}}$ unabhängig, also ist $U := \frac{X}{X+Y} \sim \mathcal{B}_{\frac{m}{2}, \frac{n}{2}}$. Dann ist

$$F_{m,n} = \frac{nX}{mY} = \frac{n}{m} \frac{X}{X+Y} = \frac{n}{m} \frac{U}{1-U},$$

mit

$$T: (0, 1) \rightarrow (0, \infty), u \mapsto \frac{n}{m} \frac{u}{1-u}, T^{-1}(z) = \frac{mz}{n+ mz}, \frac{d}{dz} T^{-1}(z) = \frac{mnz}{(n+ mz)^2}$$

ist $F_{m,n} = T(U)$, hat also die Dichte

$$f_{m,n}(z) = \frac{mnz}{(n+ mz)^2} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{mz}{n+ mz} \right)^{\frac{m}{2}-1} \left(\frac{n}{n+ mz} \right)^{\frac{n}{2}-1}.$$

7.2.4. Korollar: Student'sches T

X, Y_1, \dots, Y_n i.i.d. $\sim \mathcal{N}(0, 1)$, dann hat

$$T := \frac{X}{\sqrt{\frac{1}{n} \sum_{i=1}^n Y_i^2}}$$

die Dichte

$$\tau_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} \frac{1}{\sqrt{n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Diese Verteilung heißt *Student'sche t-Verteilung mit n Freiheitsgraden, t_n-Verteilung*

Beweis:

T^2 hat Dichte $f_{1,n}$,

$$\varphi(z) = \sqrt{z}, \varphi^{-1}(y) = y^2, \frac{d}{dy} \varphi^{-1}(y) = 2y.$$

$|T|$ hat Dichte $2tf_{1,n}(t^2)$, da T symmetrisch um 0 verteilt ist, hat T die Dichte $|t|f_{1,n}(t^2)$.

7.3. Satz:

Seien X_1, \dots, X_n u.i.v. $\sim \mathcal{N}(m, v)$, $M := \frac{1}{n} \sum_{i=1}^n X_i$, $V^* = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2$. Dann gilt:

- (i) M und V^* sind unabhängig,
- (ii) $M \sim \mathcal{N}(m, \frac{v}{n})$, $\frac{n-1}{v} V^* \sim \chi_{n-1}^2$,
- (iii) $T := \frac{\sqrt{n}(M - m)}{\sqrt{V^*}}$ ist Student- $(n-1)$ -verteilt ($\sim t_n$).

Sei

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

mit Z_1, \dots, Z_n unabhängig und $\mathcal{N}(0, 1)$. Die Verteilung heißt die n -dimensionale Standardnormalverteilung mit Dichte

$$(2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n Z_i^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \|z\|^2}$$

auf \mathbb{R}^n . Sei weiter O eine orthogonale $(n \times n)$ -Matrix (d.h. $O^T O = Id_n$). Dann ist $Y := OZ$ wieder n -dimensional Standard-normalverteilt (wie man beispielsweise mit dem Dichtentransformationssatz sieht). Insbesondere sind die Koordinaten Y_1, \dots, Y_n unabhängig.

Beweis:

O.B.d.A. $m = 0, v = 1$, sonst gehe über zu $X'_i = \frac{X_i - m}{\sqrt{v}}$.

Sei

$$z_1 = \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix},$$

ergänze zu Orthonormalbasis z_1, \dots, z_n von \mathbb{R}^n . Sei weiter

$$O = \begin{pmatrix} z_1^t \\ z_2^t \\ \vdots \\ z_n^t \end{pmatrix} \text{ ist orthogonal, } Y = OX \quad \left(, X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \right),$$

Y ist ebenfalls n -dimensional Standard-normalverteilt.

$$Y_1 = (OX)_1 = \sum_{i=1}^n \frac{1}{\sqrt{n}} X_i = \sqrt{n}M \text{ (insbesondere ist } M \sim \mathcal{N}(0, \frac{1}{n})\text{),}$$

$$(n-1)V^* = \sum_{i=1}^n (X_i - M)^2 = \underbrace{\sum_{i=1}^n X_i^2}_{=||X||^2 = ||Y||^2} - \underbrace{nM^2}_{=Y_1^2} = \sum_{i=2}^n Y_i^2 \sim \chi_{n-1}^2,$$

also ist $(n-1)V^* \sim \chi^2$ -verteilt mit $n-1$ Freiheitsgraden und unabhängig von Y_1 .

8. Konfidenzintervalle für den Median

Szenario: Seien X_1, \dots, X_n u.i.v. reellwertig, mit (unbekannter) Verteilung \mathbf{P} , \mathbf{P} habe stetige Verteilungsfunktion. $m(\mathbf{P})$ sei ein Median von \mathbf{P} (d.h. $\mathbf{P}[(-\infty, m(\mathbf{P}))] = \frac{1}{2} = \mathbf{P}[[m(\mathbf{P}), \infty))$). Die zugehörige Ordnungsstatistik ist

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}.$$

Bemerkung:

Da \mathbf{P} stetige Verteilungsfunktion hat, gilt

$$\mathbf{P}^{\otimes n}[X_i = X_j] = 0 \text{ für } i \neq j.$$

Beobachtung:

$\mathbf{P}^{\otimes n}[X_{(k)} > m(\mathbf{P})] = \mathbf{P}^{\otimes n}[\#\{1 \leq i \leq n : X_i \leq m(\mathbf{P})\} \leq k-1] = \text{Bin}_{n, \frac{1}{2}}(\{0, \dots, k-1\})$, analog:
 $\mathbf{P}^{\otimes n}[X_{(n-k-1)} < m(\mathbf{P})] = \text{Bin}_{n, \frac{1}{2}}(\{n-k+1, \dots, n\})$

8.1. Satz über die Konfidenzintervalle für den Median

Zu $\alpha \in (0, 1)$ wähle k maximal, so dass $\text{Bin}_{n, \frac{1}{2}}(\{0, \dots, k-1\}) \leq \frac{\alpha}{2}$, dann ist

$$C(X) := [X_{(k)}, X_{(n-k+1)}]$$

ein Konfidenzintervall für den Median $m(\mathbf{P})$ zum Irrtumsniveau α .

Beweis:

$$\begin{aligned} \mathbf{P}^{\otimes n}[[X_{(k)}, X_{(n-k+1)}] \ni m(\mathbf{P})] &= \mathbf{P}^{\otimes n}[\{X_{(k)} > m(\mathbf{P})\} \cup \{X_{(n-k+1)} < m(\mathbf{P})\}] \\ &\leq \mathbf{P}^{\otimes n}[\{X_{(k)} > m(\mathbf{P})\}] + \mathbf{P}^{\otimes n}[\{X_{(n-k+1)} < m(\mathbf{P})\}] \\ &= 2 \cdot \frac{\alpha}{2} = \alpha \end{aligned}$$

Beobachtung: Konfidenzintervalle und Tests

Sei $C(X)$ ein Konfidenzbereich für $\tau(\mathbf{P})$ zum Sicherheitsniveau $1 - \alpha$, d.h.

$$\mathbf{P}^{\otimes n}[\tau(\mathbf{P}) \in C(X)] \geq 1 - \alpha \quad \forall \mathbf{P} \text{ (aus einer gewissen Klasse).}$$

Betrachte die *Hypothese* $H_0 : \tau(\mathbf{P}) = \tau_0$ und die *Alternative* $H_1 : \tau(\mathbf{P}) \neq \tau_0$ und folgenden Test: Lehne H_0 ab, wenn $C(X) \ni \tau_0$, sonst nehme H_0 an.

Die „Wahrscheinlichkeit, die Nullhypothese fälschlicherweise anzulehnen“, ist

$$\sup_{\mathbf{P}:\tau(\mathbf{P})=\tau_0} \mathbf{P}[„lehne H_0 ab“] = \sup_{\mathbf{P}:\tau(\mathbf{P})=\tau_0} \mathbf{P}[\tau(\mathbf{P}) \notin C(X)] = \sup_{\mathbf{P}:\tau(\mathbf{P})=\tau_0} \mathbf{P}[\tau_0 \notin C(X)] \leq \alpha.$$

Diese Größe nennt man (die Wahrscheinlichkeit eines) *Fehler(s) 1. Art*.

8.2. Korollar: Vorzeichentest für den Median

Seien X_1, \dots, X_n u.i.v. reellwertig, gemäß \mathbf{P} (\mathbf{P} habe stetige Verteilungsfunktion), $\alpha \in (0, 1)$, wähle k maximal mit $\text{Bin}_{n, \frac{1}{2}}(\{0, \dots, k-1\}) \leq \frac{\alpha}{2}$, teste $H_0 : m(\mathbf{P}) = m_0$ gegen $H_1 : m(\mathbf{P}) \neq m_0$ folgendermaßen:

Wenn $[X_{(k)}, X_{(n-k+1)}] \ni m_0$, so nehme H_0 an, sonst lehne H_0 ab und nehme H_1 an. Der Fehler 1. Art dieses Tests ist $\leq \alpha$.

Analog für $H_0 : m(\mathbf{P}) \leq m_0$ gegen $H_1 : m(\mathbf{P}) > m_0$:

Wenn $X_{(k')} > m_0$, so lehne H_0 ab, wo k' so gewählt ist, dass $\text{Bin}_{n, \frac{1}{2}}(\{0, \dots, k'-1\}) \leq \alpha$ (und k' maximal mit dieser Eigenschaft).

8.3. Die 2-Stichproben-Situation:

8.3.1. Wilcoxon-Rangsummen-Test

Sei $n = k + l$, X_1, \dots, X_k u.i.v. gemäß \mathbf{P} und X_{k+1}, \dots, X_{k+l} u.i.v. gemäß \mathbf{Q} , mit \mathbf{P}, \mathbf{Q} Wahrscheinlichkeitsmaße auf \mathbb{R} mit stetiger Verteilungsfunktion.

Wie kann man prüfen, ob unter \mathbf{P} „typischerweise kleinere Werte angenommen werden?“

$$R_i = |\{1 \leq j \leq n : X_j \leq X_i\}|,$$

Dann ist

$$W = \sum_{i=1}^k R_i \text{ die Wilcoxon-Rangsummen-Statistik.}$$

Beobachtung:

- $\sum_{i=1}^n R_i = \frac{n(n+1)}{2}$
- $W = U + \frac{k(k+1)}{2}$, wo $U = U_{k,l} = \sum_{i=1}^k \sum_{j=k+1}^{k+l} \mathbb{1}_{(X_i > X_j)}$

O.B.d.A. nehme an, dass $X_1 < \dots < X_k$ (denn U, W invariant unter Permutation von X_1, \dots, X_k), dann ist

$$R_i = i + \sum_{j=k+1}^{k+l} \mathbb{1}_{(X_i > X_j)}.$$

8.3.2. Proposition:

Sei \mathbf{P} stetige Verteilung auf \mathbb{R} , $n = k + l$, X_1, \dots, X_n i.i.d gemäß \mathbf{P} , U wie oben. Dann gilt für $m \in \{0, 1, \dots, k \cdot l\}$:

$$\mathbf{P}^{\otimes n}[U = m] = \frac{1}{\binom{n}{k}} \cdot N(m; k, l),$$

wobei

$$N(m; k, l) = \left| \left\{ (m_1, \dots, m_k) \in \{0, 1, \dots, l\}^k : m_1 \leq \dots \leq m_k, \sum_{i=1}^k m_i = m \right\} \right|.$$

Ohne Einschränkung seien X_1, \dots, X_n unabhängig $\sim \text{unif}([0, 1])$ ($n = k + l$),

$$U_{k,l} := \sum_{i=1}^k \sum_{j=k+1}^{k+l} \mathbb{1}_{X_i > X_j}.$$

Es gilt $\mathbf{E}[U_{k,l}] = \frac{1}{2}kl$,

$$\begin{aligned} \text{Var}(U_{k,l}) &= \text{Cov}(U_{k,l}, U_{k,l}) \\ &= \sum_{i=1}^k \sum_{j=k+1}^{k+l} \sum_{i'=1}^k \sum_{j'=k+1}^{k+l} \text{Cov}(\mathbb{1}_{(X_i > X_j)}, \mathbb{1}_{(X_{i'} < X_{j'})}) \\ &= \frac{1}{4} \cdot kl + \frac{1}{12}kl(l-1) + \frac{1}{12}lk(k-1) \\ &= \frac{kl(k+l+1)}{12} =: V_{k,l}. \end{aligned}$$

8.4. Satz:

$$U_{k,l}^* := \frac{U_{k,l} - \frac{1}{2}kl}{\sqrt{V_{k,l}}} \xrightarrow[k,l \rightarrow \infty]{\text{in Vert.}} \mathcal{N}(0, 1)$$

Beweis:

Idee: Approximiere $\mathbb{1}_{X_i > X_j} - \frac{1}{2}$ durch $X_i - X_j$. Definiere weiter

$$Z_{k,l} := \sum_{i=1}^k \sum_{j=k+1}^{k+l} (X_i - X_j) = l \sum_{i=1}^k X_i - k \sum_{j=k+1}^{k+l} X_j.$$

Dann ist

$$\mathbf{E}[Z_{k,l}] = 0 \text{ und } \text{Var}(Z_{k,l}) = l^2k \frac{1}{12} + k^2l \frac{1}{12} = \frac{kl(k+l)}{12}.$$

1

$$\text{Cov}(\mathbb{1}_{(X_i > X_j)}, \mathbb{1}_{(X_{i'} < X_{j'})}) = \begin{cases} 0 & \text{wenn } i \neq i' \text{ und } j \neq j' \\ \mathbf{P}[X_1 > X_2, X_1 > X_3] - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} & \text{wenn } i = i' \text{ und } j \neq j' \text{ oder } i \neq i' \text{ und } j = j' \\ \frac{1}{2} - \frac{1}{4} = \frac{1}{4} & \text{wenn } i = i' \text{ und } j = j' \end{cases}$$

Sei nun $Z_{k,l}^* := \frac{Z_{k,l}}{\sqrt{V_{k,l}}}$. Wir wollen dann noch zeigen, dass $\text{Var}(U_{k,l}^* - Z_{k,l}^*) \xrightarrow{k,l \rightarrow \infty} 0$:

$$\begin{aligned} \text{Cov}(U_{k,l}, Z_{k,l}) &= l \sum_{i=1}^k \text{Cov}(U_{k,l}, X_i) - k \sum_{j=k+1}^{k+l} \text{Cov}(U_{k,l}, X_j), \\ \text{Cov}(U_{k,l}, X_i) &= \sum_{i'=1}^k \sum_{j'=k+1}^{k+l} \underbrace{\text{Cov}(\mathbb{1}_{X_{i'} > X_{j'}}, X_i)}_{\substack{\text{für } i \neq i' \\ \text{falls } i = i' \\ = \int_0^1 \int_0^1 \mathbb{1}_{x_1 > x_2} x_1 dx_2 dx_1 - \frac{1}{4} = \frac{1}{12}}} \\ &= l \frac{1}{12}, \\ \text{Cov}(U_{k,l}, X_i) &= -\frac{k}{12} \text{ analog.} \end{aligned}$$

Also $\text{Cov}(U_{k,l}, Z_{k,l}) = \frac{kl(k+l)}{12}$, somit

$$\begin{aligned} \text{Var}(U_{k,l} - Z_{k,l}) &= \text{Var}(U_{k,l}) - 2\text{Cov}(U_{k,l}, Z_{k,l}) + \text{Var}(Z_{k,l}) = \frac{kl}{12} \\ \text{und } \text{Var}(U_{k,l}^* - Z_{k,l}^*) &= \frac{1}{k+l+1} \xrightarrow{k,l \rightarrow \infty} 0. \end{aligned}$$

$$Z_{k,l}^* = \sqrt{\frac{l}{k+l+1}} \cdot S_k^* - \sqrt{\frac{k}{k+l+1}} T_l^*$$

$\underbrace{\hspace{1.5cm}}_{=: a_{k,l}} \quad \underbrace{\hspace{1.5cm}}_{=: b_{k,l}}$

$$\text{mit } S_k^* = \frac{1}{\sqrt{\frac{k}{12}}} \sum_{i=1}^k (X_i - \frac{1}{2}), T_l^* = \frac{1}{\sqrt{\frac{l}{12}}} \sum_{j=k+1}^{k+l} (X_j - \frac{1}{2}).$$

Nach ZGWS:

$$(S_k^*, T_l^*) \xrightarrow[k,l \rightarrow \infty]{\text{in Vert.}} (S, T), \text{ mit } S, T \text{ unabhängig und } \mathcal{N}(0, 1) \text{ verteilt.}$$

Sei $(k_i, l_i)_{i \in \mathbb{N}}$ mit $k_i \rightarrow \infty, l_i \rightarrow \infty$, dann gibt es eine Teilfolge (k_{i_j}, l_{i_j}) , sodass $a_{k_{i_j}, l_{i_j}} \rightarrow a$ und $b_{k_{i_j}, l_{i_j}} \rightarrow b$ und $a + b = 1$.

$$\text{Demnach } Z_{k_{i_j}, l_{i_j}}^* \xrightarrow[j \rightarrow \infty]{\text{in Vert.}} \underbrace{\sqrt{a}S}_{\sim \mathcal{N}(0, a)} + \underbrace{\sqrt{b}T}_{\sim \mathcal{N}(0, b)} \sim \mathcal{N}(0, \underbrace{a+b}_{=1}).$$

8.5. Definition: stochastisch kleiner

Seien \mathbf{P}, \mathbf{Q} zwei Wahrscheinlichkeitsmaße auf \mathbb{R} , so heißt \mathbf{P} *stochastisch kleiner* als \mathbf{Q} , $\mathbf{P} \preceq \mathbf{Q}$, wenn

$$\mathbf{P}[[c, \infty)] \leq \mathbf{Q}[[c, \infty)] \quad \forall c \in \mathbb{R},$$

d.h. für die zugehörigen Verteilungsfunktionen gilt $F_{\mathbf{P}} \geq F_{\mathbf{Q}}$.

Bemerkung:

$\mathbf{P} \preceq \mathbf{Q} \iff \exists$ Zufallsvariablen X, Y (auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$) mit $X \leq Y$ \mathbf{P} -f.s., $\mathbf{P} \circ X^{-1} = \mathbf{P}$ und $\mathbf{P} \circ Y^{-1} = \mathbf{Q}$.

Beweis der Bemerkung:

„ \Leftarrow “: $\mathbf{P}[[c, \infty)] = \mathbf{P}[X \geq c] \leq \mathbf{P}[Y \geq c] = \mathbf{Q}[[c, \infty)]$
„ \Rightarrow “: Sei $F_{\mathbf{P}}^{-1}(u) := \inf\{x : F_{\mathbf{P}}(x) \geq u\}$ und analog $F_{\mathbf{Q}}^{-1}$. Sei weiter U uniform auf $[0, 1]$,
 $X := F_{\mathbf{P}}^{-1}(U) \leq Y := F_{\mathbf{Q}}^{-1}(U)$, denn $F_{\mathbf{P}}^{-1}(u) \leq F_{\mathbf{Q}}^{-1}(u)$.

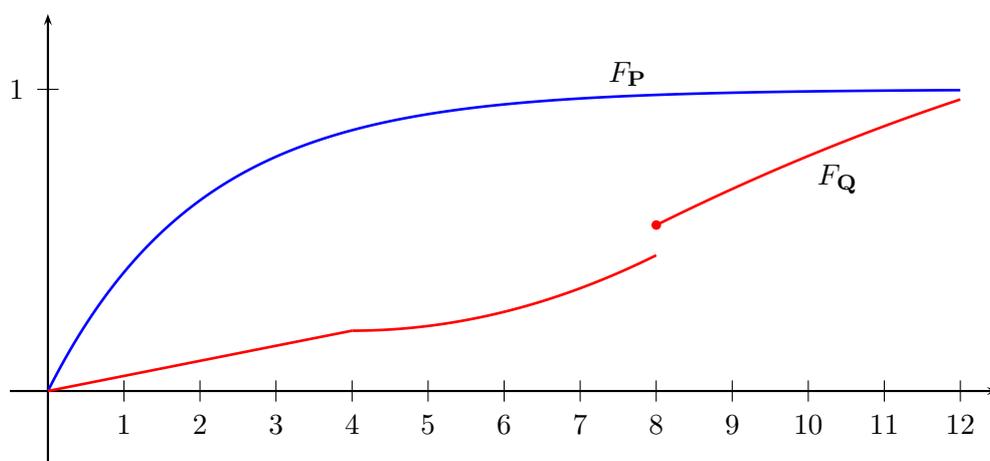


Abbildung 8.2.: Darstellung der Verteilungsfunktionen von \mathbf{P} und \mathbf{Q} , wenn $\mathbf{P} \preceq \mathbf{Q}$

Bemerkung:

$X \sim \mathbf{P}, c > 0, X + c \sim \mathbf{Q}$, dann ist $\mathbf{P} \preceq \mathbf{Q}$, insbesondere ist $\mathcal{N}(m_1, v) \preceq \mathcal{N}(m_2, v)$, wenn $m_1 \leq m_2$.

8.5.1. Satz:

Sei $n = k + l$, X_1, \dots, X_k i.i.d. gemäß \mathbf{Q}_1 und X_{k+1}, \dots, X_{k+l} i.i.d. gemäß \mathbf{Q}_2 , $\mathbf{Q}_1, \mathbf{Q}_2$ stetige Verteilungen auf \mathbb{R} , $\alpha \in (0, 1)$. Bestimme c so, dass

$$\mathbf{P}[U_{k,l} \leq c] \leq \alpha.$$

Test: Berechne $U_{k,l}$ anhand der Daten und lehne $H_0: \mathbf{Q}_1 = \mathbf{Q}_2$ ab, wenn $U_{k,l} \leq c$ zugunsten der Alternative $H_1: \mathbf{Q}_1 \preceq \mathbf{Q}_2$, sonst nehme H_0 an.

Dann ist die Wahrscheinlichkeit für einen Fehler 1. Art (d.h. H_0 fälschlicherweise anzulehnen) $\leq \alpha$. (Dies ist ein Test von H_0 gegen H_1 zum Niveau α .)

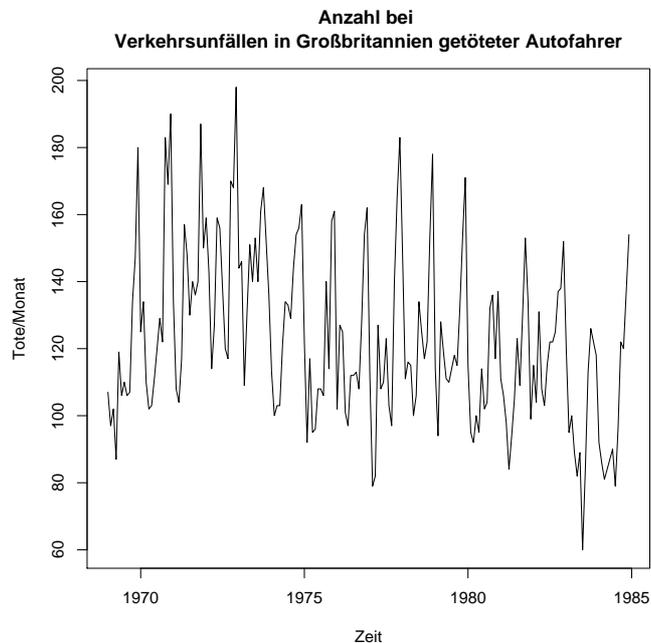
8.5.2. Beispiel: Verkehrstote in Großbritannien 1969–1984

Der mit R mitgelieferte Datensatz `Seatbelts`² enthält (u.A.) für jeden Monat im Zeitraum Januar 1969 bis Dezember 1984 die Anzahl bei Verkehrsunfällen in Großbritannien getöteten Autofahrer.

```
R version 2.4.0 Patched (2006-11-25 r39997)
Copyright (C) 2006 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
```

...

```
> data(Seatbelts)
> plot(Seatbelts[,1], xlab="Zeit", ylab="Tote/Monat",main="Anzahl bei
Verkehrsunfällen in Großbritannien getöteter Autofahrer")
```



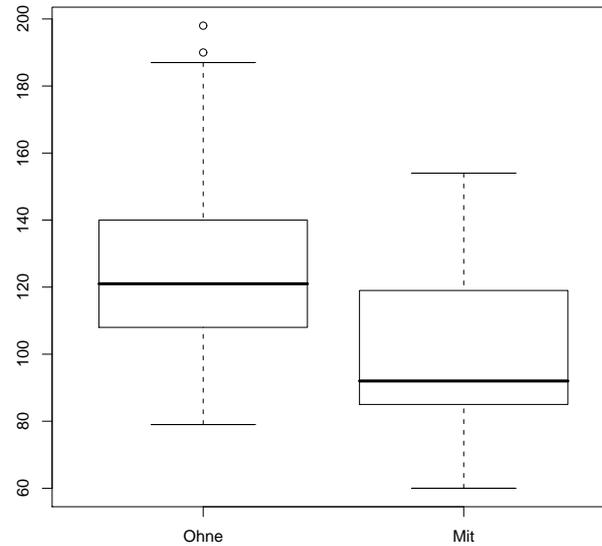
Am 31. Januar 1983 wurde in Großbritannien die Gurtpflicht eingeführt, wir spalten die Daten entsprechend:

```
> A<-Seatbelts[Seatbelts[,8]==0,1]
> B<-Seatbelts[Seatbelts[,8]==1,1]
(Seatbelts[,8] ist 0, wenn im betreffenden Monat keine Gurtpflicht galt, A enthält nun die
Monate ohne Gurtpflicht, B die mit.)
```

Ein Boxplot zeigt, dass die Werte in A tendenziell größer als die in B sind:

```
> boxplot(list("Ohne"=A, "Mit"=B))
```

²Basierend auf A. C. Harvey and J. Durbin, The effects of seat belt legislation on British road casualties: A case study in structural time series modelling, *J. Roy. Stat. Soc. B* **149**, 187–227 (1986).



Ein Rangsummen-Test zeigt, dass die Werte in A tatsächlich signifikant kleiner sind als die in B:

```
> wilcox.test(B,A,alt="less")
```

Wilcoxon rank sum test with continuity correction

data: B and A

W = 829, p-value = 4.171e-06

alternative hypothesis: true location shift is less than 0

Bemerkung: „Machtlosigkeit“ des Rangsummentests gegen unterschiedliche Varianzen

Zur Illustration simulieren wir 100 $\mathcal{N}(0,1)$ - und 100 $\mathcal{N}(0,2)$ -verteilte Werte. Obwohl diese Verteilungen deutlich verschieden sind, schlägt der Rangsummentest hier typischerweise nicht an:

```
> C<-rnorm(100,mean=0,sd=1)
```

```
> D<-rnorm(100,mean=0,sd=sqrt(2))
```

```
> wilcox.test(C,D)
```

Wilcoxon rank sum test with continuity correction

data: C and D

W = 5066, p-value = 0.8728

alternative hypothesis: true location shift is not equal to 0

9. Testen von Hypothesen:

9.1. Definition: Test, Randomisierung und Gütefunktion

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_\vartheta : \vartheta \in \Theta)$ ein statistisches Modell, $\Theta = \Theta_0 \dot{\cup} \Theta_1$ (Zerlegung in „Nullhypothese“ und „Alternative“).

Eine Statistik $\varphi : \mathcal{X} \rightarrow [0, 1]$ (interpretiert als Entscheidungsregel: Bei gegebener Beobachtung x entscheide mit Wahrscheinlichkeit $\varphi(x)$ für die Alternative) heißt ein *Test* von Θ_0 gegen Θ_1 .

Der Test heißt *randomisiert*, wenn $\varphi(\mathcal{X}) \not\subseteq \{0, 1\}$, sonst *nicht-randomisiert*.

Für nicht randomisierte Tests heißt

$$\{x : \varphi(x) = 1\} \text{ der Ablehnungs- oder Verwerfungsbereich.}$$

Weiter heißt

$$G_\varphi(\vartheta) = \mathbf{E}_\vartheta[\varphi] \text{ die Gütefunktion von } \varphi,$$

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta)$$

heißt *Umfang* oder *effektives Niveau* von φ . φ heißt ein Test zum Irrtumsniveau $\alpha \in (0, 1)$, wenn $\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \leq \alpha$. Für $\vartheta \in \Theta_1$ heißt $G_\varphi(\vartheta)$ die *Macht* oder *Schärfe* von φ (engl. „Power“) bei ϑ .

9.2. Definition: gleichmäßig bester Test

Ein Test von Θ_0 gegen Θ_1 heißt ein *gleichmäßig bester Test* zum Niveau $\alpha \in (0, 1)$, wenn er das Niveau α einhält, d.h.

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \leq \alpha,$$

und für jeden Test ψ , der das Niveau α einhält,

$$G_\varphi(\vartheta) \geq G_\psi(\vartheta) \quad \forall \vartheta \in \Theta_1$$

gilt.

9.3. Alternativtest und das Neyman-Pearson-Lemma

Betrachte ein Standardmodell, $\Theta = \{0, 1\}$, d.h.

$$\mathbf{P}_0[dx] = \rho(x, 0)dx, \quad \mathbf{P}_1[dx] = \rho(x, 1) dx$$

(bzw. analog für \mathcal{X} diskret). Setze

$$R(x) := \begin{cases} \frac{\rho(x, 1)}{\rho(x, 0)} & \text{wenn } \rho(x, 0) > 0, \\ \infty & \text{sonst.} \end{cases}$$

R heißt *Likelihood-Quotient*.

9.3.1. Definition: Neyman-Pearson-Test

Ein Test φ von \mathbf{P}_0 gegen \mathbf{P}_1 der Form

$$\varphi(x) = \begin{cases} 1 & \text{für } R(x) > c, \\ 0 & \text{für } R(x) < c, \end{cases}$$

heißt ein *Neyman-Pearson-Test* zum Schwellenwert c .

9.3.2. Satz: (Neyman-Pearson-Lemma)

Betrachte Standardmodell mit einpunktiger Nullhypothese und einpunktiger Alternative, $\alpha \in (0, 1)$.

- (i) Es gibt einen Neyman-Pearson-Test φ mit $\mathbf{E}_0[\varphi] = \alpha$.
- (ii) Jeder Neyman-Pearson-Test φ mit $\mathbf{E}_0[\varphi] = \alpha$ ist ein gleichmäßig bester Test zum Niveau α , und jeder gleichmäßig beste Test zum Niveau α ist ununterscheidbar von einem (gewissen) Neyman-Pearson-Test.

Beweis:

- (i) Wähle c mit $\mathbf{P}_0[R \geq c] \geq \alpha$ und $\mathbf{P}_0[R \leq c] \geq 1 - \alpha$

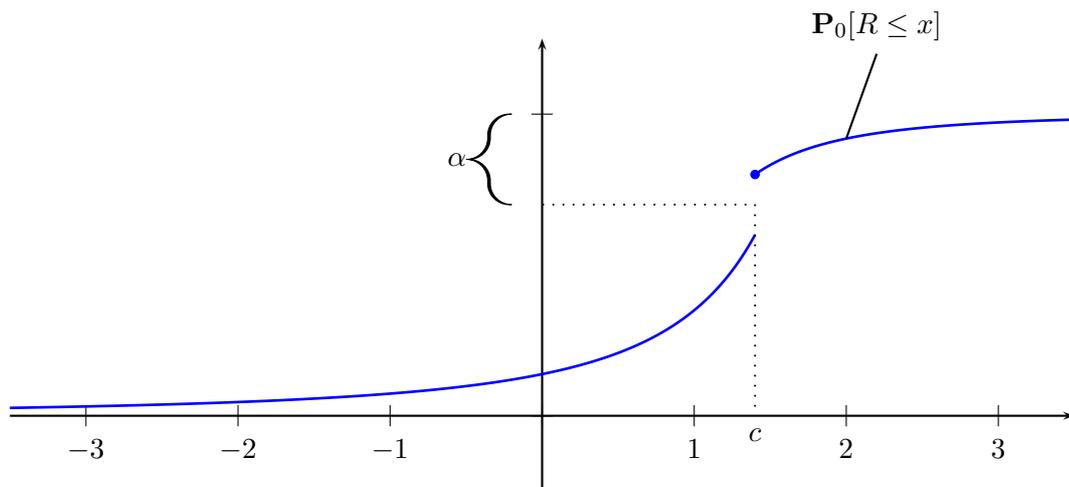


Abbildung 9.1.: Abbildung zur Wahl von c , falls die Verteilungsfunktion gerade an dieser Stelle einen Sprung hat. Deswegen kann es in manchen Fällen auch sinnvoll sein, den Test zu randomisieren.

Falls $\mathbf{P}_0[R = c] = 0$, so ist $\varphi = \mathbf{1}_{R(x) > c}$ ein Neyman-Pearson-Test mit $\mathbf{E}_0[\varphi] = \alpha$ wie gewünscht.

Falls $\mathbf{P}_0[R = c] > 0$, setze $\gamma := \frac{\alpha - \mathbf{P}_0[R > c]}{\mathbf{P}_0[R = c]} \in [0, 1]$ und

$$\varphi := \begin{cases} 1 & \text{wenn } R(x) > c, \\ \gamma & \text{wenn } R(x) = c, \\ 0 & \text{wenn } R(x) < c, \end{cases}$$

denn $\mathbf{P}_0[R > c] = 1 - \mathbf{P}_0[R \leq c] \leq \alpha$ und
 $\mathbf{P}_0[R = c] = \mathbf{P}_0[R \geq c] - \mathbf{P}_0[R > c] \geq \alpha - \mathbf{P}_0[R > c]$.

Dies ist ein Neyman-Pearson-Test mit

$$\mathbf{E}_0[\varphi] = 1 \cdot \mathbf{P}_0[R > c] + \gamma \mathbf{P}_0[R = c] + 0 \cdot \mathbf{P}_0[R < c] = \alpha.$$

- (ii) Sei φ ein Neyman-Pearson-Test (mit Schwellenwert c) mit $\mathbf{E}_0[\varphi] = \alpha$, ψ irgendein Test mit $\mathbf{E}_0[\psi] = \alpha$. Es gilt

$$\forall x : (\varphi(x) - \psi(x))(\rho(x, 1) - c\rho(x, 0)) \geq 0,$$

denn die beiden Faktoren haben dasselbe Vorzeichen (sofern der zweite $\neq 0$):

$\varphi(x) \geq \mathbf{1}(\rho(x, 1) > c\rho(x, 0))$, also

$$(**) \quad f_1(x) := (\varphi(x) - \psi(x))\rho(x, 1) \geq c(\varphi(x) - \psi(x))\rho(x, 0) =: cf_0(x),$$

und demnach

$$(*) \quad \mathbf{E}_1[\varphi] - \mathbf{E}_1[\psi] = \int_{\mathcal{X}} f_1(x) dx \geq c \int_{\mathcal{X}} f_0(x) dx = c(\alpha - \mathbf{E}_0[\psi]) \geq 0$$

Also ist die Macht von φ größergleich der von ψ .

Sei ψ ebenfalls ein bester Test mit $\mathbf{E}_0[\psi] = \alpha$, dann gilt Gleichheit in (*), folglich gilt Gleichheit in (**) (unter Umständen bis auf x aus einer Lebesgue-Nullmenge N), d.h.

$$(\varphi(x) - \psi(x)) \cdot (\rho(x, 1) - c\rho(x, 0)) = 0,$$

d.h. $\varphi = \psi$ für $x \notin N$ mit $R(x) \neq c$. In diesem Sinne sind die beiden Tests ununterscheidbar.

9.4. Satz: (Lemma von Stein)

Im unendlichen Produktmodell $(\mathcal{X}^{\mathbb{N}}, \mathcal{F}^{\otimes \mathbb{N}}, \mathbf{P}_{\vartheta}^{\otimes \mathbb{N}} : \vartheta \in \{0, 1\})$, sei φ_n ein Neyman-Pearson-Test mit $\mathbf{E}_0[\varphi_n] = \alpha$, der nur von den Beobachtungen X_1, \dots, X_n abhängt. Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(1 - \mathbf{E}_1[\varphi_n]) = -H(\mathbf{P}_0 | \mathbf{P}_1),$$

d.h. $\mathbf{E}[\varphi_n] \approx 1 - e^{-nH(\mathbf{P}_0 | \mathbf{P}_1)}$, die Macht von φ konvergiert exponentiell schnell gegen 1.

Beweis:

Sei

$$R_n := \frac{\rho_1^{\otimes n}(X_1, \dots, X_n)}{\rho_0^{\otimes n}(X_1, \dots, X_n)} = \prod_{i=1}^n \frac{\rho_1(X_i)}{\rho_0(X_i)},$$

$$h_n := -\frac{1}{n} \log R_n = \frac{1}{n} \sum_{i=1}^n h(X_i),$$

¹Im Diskreten statt Integral natürlich eine Summe.

wo $h(x) = \log \left(\frac{\rho_0(x)}{\rho_1(x)} \right)$. Dabei bemerken wir, dass

$$\mathbf{E}_0[h(X_1)] = H(\mathbf{P}_0|\mathbf{P}_1).$$

Dann ist

$$\varphi_n = \begin{cases} 1 & \text{wenn } h_n < a_n \\ 0 & \text{wenn } h_n > a_n \end{cases}$$

für ein a_n . Sei $a < \mathbf{E}_0[h]$,

$$\mathbf{P}_0^{\otimes n}[h_n \leq a] \xrightarrow{n \rightarrow \infty} 0 \text{ mit dem Gesetz großer Zahlen,}$$

also $a_n > a$ für n genügend groß,

$$\{1 - \varphi_n > 0\} \subseteq \{h_n \geq a_n\} = \{\rho_0^{\otimes n} \geq e^{na_n} \rho_1^{\otimes n}\}$$

Somit

$$1 \geq \mathbf{E}_0[1 - \varphi_n] \geq e^{na_n} \mathbf{E}_1[1 - \varphi_n],$$

d.h.

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(1 - \mathbf{E}_1[\varphi_n]) \leq -a,$$

also auch

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(1 - \mathbf{E}_1[\varphi_n]) \leq -\mathbf{E}_0[h].$$

Sei $a > \mathbf{E}_0[h]$ (und ohne Einschränkung $\mathbf{E}_0[h] < \infty$), dann ist

$$\mathbf{P}_0^{\otimes n}[\rho_1^{\otimes n} \geq e^{-na} \rho_0^{\otimes n}] = \mathbf{P}_0^{\otimes n}[h_n \leq a] \geq \frac{1 + \alpha}{2}$$

für n genügend groß.

$$\begin{aligned} \mathbf{E}_1[1 - \varphi_n] &= \mathbf{E}_0[(1 - \varphi_n) \frac{\rho_1^{\otimes n}}{\rho_0^{\otimes n}}] \\ &\geq e^{-na} \mathbf{E}_0[(1 - \varphi_n) \mathbf{1}_{h_n \leq a}] \\ &\geq e^{-na} (\mathbf{P}_0^{\otimes n}[h_n \leq a] - \underbrace{\mathbf{E}_0[\varphi_n]}_{=\alpha}) \\ &\geq e^{-na} \frac{1 - \alpha}{2} \end{aligned}$$

Somit

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}_1[1 - \varphi_n] \geq -a,$$

mit $a \searrow \mathbf{E}_0[h]$ folgt die Behauptung.

□

9.4.1. Beispiel:

Gegeben: X_1, \dots, X_n , $\mathbf{P}_0 = \mathcal{N}^{\otimes n}(m_0, v)$, $\mathbf{P}_1 = \mathcal{N}^{\otimes n}(m_1, v)$, $v > 0$ fest, o.B.d.A. $m_0 < m_1$.
Teste $H_0 : m = m_0$ gegen $H_1 : m = m_1$.

Sei

$$R_n = \exp \left(-\frac{1}{2v} \sum_{i=1}^n (X_i - m_1)^2 + \frac{1}{2v} \sum_{i=1}^n (X_i - m_0)^2 \right) = \exp \left(-\frac{n}{2v} (2(m_1 - m_0) \underbrace{M_n}_{=:\frac{1}{n} \sum_{i=1}^n X_i} + m_1^2 - m_0^2) \right)$$

Betrachte nun

$$h_n = -\frac{1}{n} \log R_n = \frac{m_0 - m_1}{v} M_n - \frac{m_1^2 - m_0^2}{2v}.$$

Für $\alpha \in (0, 1)$ wähle a_n so, dass $\mathbf{P}_0[h_n < a_n] = \alpha$, also

$$a_n = \frac{m_0 - m_1}{v} \left(\sqrt{\frac{n}{v}} \Phi^{-1}(1 - \alpha) + \frac{m_1^2 - m_0^2}{2(m_0 - m_1)} + m_0 \right).$$

10. Monotone Likelihood-Quotienten und beste einseitige Tests

10.1. Definition: Likelihood-Quotient $R_{\vartheta', \vartheta}(x)$

Ein Standardmodell $(\mathcal{X}, \mathcal{F}, \mathbf{P}_{\vartheta} : \vartheta \in \Theta)$ mit $\Theta \subseteq \mathbb{R}$ hat wachsende Likelihood-Quotienten (bzgl. einer reellwertigen Statistik T), wenn für alle $\vartheta < \vartheta'$ der Likelihood-Quotient

$$R_{\vartheta', \vartheta}(x) := \frac{\rho_{\vartheta'}(x)}{\rho_{\vartheta}(x)}$$

eine wachsende Funktion von T ist, d.h. $R_{\vartheta', \vartheta}(x) = f_{\vartheta', \vartheta}(T(x)) = (f_{\vartheta', \vartheta} \circ T)(x)$ für eine wachsende Funktion $f_{\vartheta', \vartheta}$.

10.1.1. Beobachtung:

Jedes einparametrische exponentielle Modell hat wachsende Likelihood-Quotienten:

$$\rho(x, \vartheta) = h(x)e^{a(\vartheta)T(x) - b(\vartheta)}$$

mit a strikt wachsend oder strikt fallend, demnach

$$R_{\vartheta', \vartheta}(x) = \frac{\rho_{\vartheta'}(x)}{\rho_{\vartheta}(x)} = \exp((a(\vartheta') - a(\vartheta)T(x) - b(\vartheta') + b(\vartheta))).$$

Klar, falls a streng monoton wachsend, sonst gehe zu $-T$ über.

10.2. Satz:

Sei $(\mathcal{X}, \mathcal{F}, \mathbf{P}_{\vartheta} : \vartheta \in \Theta)$, $\Theta \subseteq \mathbb{R}$ ein Standardmodell mit wachsenden Likelihoodquotienten bzgl. T , $\vartheta_0 \in \Theta$, $\alpha \in (0, 1)$. Ein gleichmäßig bester Test von $H_0 : \vartheta \leq \vartheta_0$ gegen $H_1 : \vartheta > \vartheta_0$ zum Niveau α hat die Form

$$\varphi(x) = \begin{cases} 1 & \iff T(x) > c \\ \gamma & \iff T(x) = c \\ 0 & \iff T(x) < c \end{cases}$$

für gewisse $c \in \mathbb{R}$, $\gamma \in [0, 1]$, die durch die Bedingung $G_{\varphi}(\vartheta_0) = \alpha$ festgelegt sind. Die Gütefunktion ist monoton wachsend in ϑ .

Bemerkung:

Um $H_0 : \vartheta \geq \vartheta_0$ gegen $H_1 : \vartheta < \vartheta_0$ zu testen, vertausche „<“ und „>“ der Definition von φ (formal: multipliziere ϑ und T jeweils mit -1).

Beweis:

Setze $\Theta_0 := \Theta \cap (-\infty, \vartheta_0]$, $\Theta_1 := \Theta \cap (\vartheta_0, \infty)$. Sei ψ irgendein Test von H_0 gegen H_1 der das Niveau α einhält, d.h. $\sup_{\vartheta \in \Theta_0} G_\psi(\vartheta) \leq \alpha$, $\vartheta' \in \Theta_1$. Fasse φ und ψ auf als Tests von $\{\vartheta_0\}$ gegen $\{\vartheta'\}$; in diesem Szenario ist dann φ ein Neyman-Pearson-Test mit $\mathbf{E}_{\vartheta_0}[\varphi] = \alpha$.

Demnach

$$G_\varphi(\vartheta') \geq G_\psi(\vartheta') \quad \forall \vartheta' \in \Theta_1.$$

Zur Monotonie: Sei $\vartheta < \vartheta'$, fasse φ auf als einen (Neyman-Pearson-)Test von $\{\vartheta\}$ gegen $\{\vartheta'\}$ mit Niveau $\beta := G_\varphi(\vartheta)$.

Vergleiche mit dem (trivialen) Test $\psi_\beta(x) \equiv \beta$, der ebenfalls Niveau β hat.

Gemäß dem Neyman-Pearson-Lemma ist $G_\varphi(\vartheta') \geq G_{\psi_\beta}(\vartheta') = \beta = G_\varphi(\vartheta)$, insbesondere ist $\vartheta \mapsto G_\varphi(\vartheta)$ monoton wachsend, also

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \stackrel{(\leq)}{=} \alpha.$$

□

10.2.1. Beispiel:

- (i) Sei X_1, \dots, X_n u.i.v gemäß $\mathcal{N}(\vartheta, v)$, $v > 0$ fest, $\vartheta \in \mathbb{R}$. Teste $\{\vartheta \leq m_0\}$ gegen $\{\vartheta > m_0\}$ folgendermaßen:

Dies ist ein exponentielles Modell bzgl. $M_n = \frac{1}{n} \sum_{i=1}^n X_i$. Der beste Test zum Niveau α hat Ablehnungsbereich

$$\left\{ M_n > m_0 + \sqrt{\frac{v}{n}} \Phi^{-1}(1 - \alpha) \right\}.$$

- (ii) Sei X_1, \dots, X_n u.i.v gemäß $\mathcal{N}(m, \vartheta)$, m fest, $\vartheta > 0$. Teste $\{\vartheta \geq \vartheta_0\}$ gegen $\{\vartheta < \vartheta_0\}$: Dies ist eine exponentielle Familie bzgl. $T := \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$, der beste Test zum Niveau α hat dann den Ablehnungsbereich

$$\left\{ T < \frac{1}{n} \vartheta_0 \cdot \chi_{n,\alpha}^2 \right\},$$

wobei $\chi_{n,\alpha}^2$ das α -Quantil von χ_n^2 .

10.3. Tests im (2-parametrischen) Gauß'schen Modell

Szenario: Sei $(m, v) \in \mathbb{R} \times (0, \infty)$, $(X_1, \dots, X_n) \sim \underbrace{\mathcal{N}(m, v)^{\otimes n}}_{=\mathbf{P}(m,v)}$, $M_n := \frac{1}{n} \sum_{i=1}^n X_i$.

Beobachtung

$$\frac{1}{v} \sum_{i=1}^n (X_i - M_n)^2 \text{ ist unter } \mathbf{P}_{(m,v)} \chi_{n-1}^2 \text{-verteilt.}$$

10.3.1. Satz: Linksseitiger χ^2 -Test für die Varianz im Gauß'schen Modell

Der Test φ mit Ablehnungsbereich¹

$$\left\{ \sum_{i=1}^n (X_i - M_n)^2 > v_0 \chi_{n-1, 1-\alpha}^2 \right\}$$

ist ein gleichmäßig bester Test von $H_0 : v \leq v_0$ gegen $H_1 : v > v_0$ zum Niveau α , wobei $\chi_{n-1, 1-\alpha}^2 = (1 - \alpha)$ -Quantil von χ_{n-1}^2 .

D.h.

$$(*) \quad \mathbf{E}_{(m,v)}[\varphi] \leq \alpha \quad \forall (m,v) \in \Theta_0$$

und für jeden Test ψ , der auch Niveau α einhält, gilt

$$\mathbf{E}_{(m,v)}[\varphi] \geq \mathbf{E}_{(m,v)}[\psi] \quad \forall (m,v) \in \Theta_1.$$

Beweis:

Zu (*):

$$\mathbf{P}_{(m,v)}[\varphi = 1] = \mathbf{P}_{(m,v)} \left[\frac{1}{v} \sum_{i=1}^n (X_i - M_n)^2 > \frac{v_0}{v} \chi_{n-1, 1-\alpha}^2 \right] \leq \alpha \quad \forall (m,v) \in \Theta_0.$$

Fixiere $(m_1, v_1) \in \Theta_1$, für $v \in (0, v_1]$ sei

$$\bar{\mathbf{P}}_v := \int \mathbf{P}_{(m,v)} w_v(\mathrm{d}m)$$

mit $w_v = \mathcal{N}(m_1, \frac{v_1-v}{n})^2$, $\bar{\mathbf{P}}_{v_1} = \mathbf{P}_{(m_1, v_1)}$.

Die Likelihood-Funktion ist

$$\begin{aligned} \bar{p}_v(X_1, \dots, X_n) &= \int_{-\infty}^{\infty} (2\pi v)^{-\frac{n}{2}} e^{-\frac{1}{2v} \sum_{i=1}^n (X_i - m)^2} \cdot \left(2\pi \frac{v_1 - v}{n} \right)^{-\frac{1}{2}} e^{-\frac{(m - m_1)^2}{2(v_1 - v)/n}} \mathrm{d}m \\ &= c_1(v) \int_{-\infty}^{\infty} \exp \left(-\frac{(m - m_1)^2}{2(v_1 - v)/n} - \frac{1}{2v} \sum_{i=1}^n (X_i - m)^2 \right) \mathrm{d}m \\ &\stackrel{3}{=} c_1(v) e^{-\frac{n-1}{2v} \cdot V^*} \underbrace{\int_{-\infty}^{\infty} \exp \left(-\frac{(m - m_1)^2}{2(v_1 - v)/n} - \frac{(m - M_n)^2}{2v/n} \right) \mathrm{d}m}_{=\text{const.} e^{-\frac{(m_1 - M_n)^2}{2v_1/n}}} \\ &= c_2(v) \exp \left(-\frac{n-1}{2v} \cdot V^* - \frac{(m_1 - M_n)^2}{2v_1/n} \right), \end{aligned}$$

¹D.h. $\varphi := \mathbb{1} \left\{ \sum_{i=1}^n (X_i - M_n)^2 > v_0 \chi_{n-1, 1-\alpha}^2 \right\}$

²Bemerkung: M_n unter $\bar{\mathbf{P}}_v$ hat Verteilung $\mathcal{N}(m_1, \frac{v_1}{n})$, denn wenn $Z \sim \mathcal{N}(0, 1)$ betrachte:

$$\sqrt{\frac{v_1}{n}} Z + M$$

d.h. $\bar{\mathbf{P}}_v, v \leq v_1$ sind eine exponentielle Familie (bzgl. V^*).

Demnach: $\tilde{\varphi} = \mathbf{1}_{\{V^* > c\}}$ (wo c so gewählt, dass $\bar{\mathbf{P}}_{v_0}[V^* > c] = \alpha$, mit $\bar{\mathbf{P}}_{v_0}[V^* > c] = \int \mathbf{P}_{(m,v)}[V^* > c] w_v dm = \mathbf{P}_{(0,v_0)}[V^* > c]$) ist ein gleichmäßig bester Test von $\{\bar{\mathbf{P}}_v : v \leq v_0\}$ gegen $\{\bar{\mathbf{P}}_{v_1}\}$ zum Niveau α (und eigentlich $\varphi = \tilde{\varphi}$). $c = \frac{v_0}{n-1} \chi_{n-1; 1-\alpha}^2$

Sei ψ ein Test von Θ_0 gegen Θ_1 zum Niveau α ,

$$\bar{G}_\psi(v) = \int \underbrace{G_\psi((m,v))}_{\leq \alpha} w_v dm \leq \alpha \int w_v dm = \alpha \cdot 1 = \alpha,$$

also liefert die Optimalität von $\tilde{\varphi}$: $G_{\tilde{\varphi}}((m_1, v_1)) = \bar{G}_{\tilde{\varphi}}(v_1) \leq \bar{G}_\psi(v_1) = G_\psi((m_1, v_1))$.

□

10.3.2. Bemerkung: (Der rechtsseitige χ^2 -Test für die Varianz)

Der Test $\varphi = \mathbf{1}_{\{\sum_{i=1}^n (X_i - M_n) < v_0 \chi_{n-1, \alpha}^2\}}$ von $H_0 : v \geq v_0$ gegen $H_1 : v < v_0$ hat das Niveau α und ist unverfälscht, d.h.

$$\sup_{(m,v) \in \Theta_0} G_\varphi((m,v)) \leq \alpha \leq \inf_{(m,v) \in \Theta_1} G_\varphi((m,v)).$$

(Er entscheidet mit größerer Wahrscheinlichkeit für die Alternative, wenn sie zutrifft, als wenn sie nicht zutrifft.) Er hat unter allen unverfälschten Tests von H_0 gegen H_1 zum Niveau α die größte Macht (d.h. die Macht ist punktweise nicht kleiner), aber es gibt zu $\alpha \in (0, 1)$ (viele) Tests, die Niveau α einhalten und auf Teilen der Alternative größere Macht haben. Für $m \in \mathbb{R}$ sei

$$\varphi_m = \mathbf{1}_{\left\{ \sum_{i=1}^n (X_i - m)^2 < v_0 \chi_{n, \alpha}^2 \right\}}.$$

φ_m hat Niveau α : $(m', v) \in \Theta_0$.

$$\mathbf{P}_{(m',v)}[\varphi_m = 1] \leq \mathbf{P}_{(m,v)}[\varphi_m = 1] \leq \mathbf{P}_{(m,v_0)}[\varphi_m = 1] = \alpha$$

(Übung: $\mathcal{N}(0, 1)^{\otimes n}(B_r(\underline{x})) \leq \mathcal{N}(0, 1)^{\otimes n}(B_r(0))$).

Für $\{m\} \times [v_0, \infty)$ gegen $\{m\} \times (0, v_0)$ ist φ_m ein bester Test, insbesondere (Übung: echt) besser als φ .

Andererseits:

$$G_{\varphi_m}((m', v)) \xrightarrow{m' \rightarrow \infty} 0 \quad \text{für jedes } v, \text{ insbesondere für } v < v_0.$$

10.3.3. Satz: Einseitiger Student'scher t -Test

Im Gauß'schen Produktmodell ist $\varphi = \mathbf{1}_{\left\{ \frac{\sqrt{n}(M_n - m_0)}{\sqrt{V^*}} > t_{n-1, 1-\alpha} \right\}}$ ($t_{n-1, 1-\alpha} = (1 - \alpha)$ -Quantil der Student'schen-Verteilung mit $n - 1$ Freiheitsgraden) ein bester unverfälschter Test von $H_0 : m \leq m_0$ gegen $H_1 : m > m_0$ zum Niveau α .

³mit

$$V^* = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2$$

Beweis:

O.B.d.A. $m_0 = 0$. Re-parametrisiere $\mu = \frac{m\sqrt{n}}{v}$, $\eta := \frac{1}{2v}$.

$$\begin{aligned}\rho_{(\mu,\eta)}(X_1, \dots, X_n) &= (2\pi v)^{-\frac{n}{2}} e^{-\frac{1}{2v} \sum_{i=1}^n (X_i - m)^2} \\ &= (2\pi v)^{-\frac{n}{2}} \exp\left(-\frac{1}{2v} \sum_{i=1}^n X_i + \frac{m}{v} n \cdot M_n - \frac{nm^2}{2v}\right) \\ &= c(\mu, \eta) \exp\left(-\mu \widetilde{M}_n - \eta S\right)\end{aligned}$$

mit $\widetilde{M}_n = \sqrt{n}M_n$, $S = \sum_{i=1}^n X_i^2$.

$$\begin{aligned}T_0 &= \frac{\sqrt{n}M_n}{V^*} \\ &= \frac{\widetilde{M}_n \sqrt{n-1}}{\sqrt{S - (\widetilde{M}_n)^2}},\end{aligned}$$

denn $(n-1)V^* = S - nM_n^2$.

$$\{T_0 > t_{n-1,1-\alpha}\} = \left\{ \frac{\widetilde{M}_n}{\sqrt{S - (\widetilde{M}_n)^2}} > \frac{t_{n-1,1-\alpha}}{\sqrt{n-1}} \right\} = \{\widetilde{M}_n > f(s)\}$$

mit $f(S) = r\sqrt{\frac{S}{1+r^2}}$, $r = \frac{t_{n-1,1-\alpha}}{\sqrt{n-1}}$.

Sei ψ ein unverfälschter Test von Θ_0 gegen Θ_1 zum Niveau α . Dann gilt

$$\mathbf{E}_{(0,\eta)}[\psi] = \alpha \quad \forall \eta > 0$$

und

$$\mathbf{E}_{(0,\eta)}[\varphi] = \alpha \quad \forall \eta > 0.$$

Behauptung: (*) $\mathbf{E}_{(0,\eta)}[h(S)(\varphi - \psi)] = 0 \quad \forall h : [0, \infty) \rightarrow \mathbb{R}$, die subexponentiell wachsen, d.h. $h(x)e^{-\delta x} \xrightarrow{x \rightarrow \infty} 0 \quad \forall \delta > 0$.

Zu (*): Sei $\gamma \in \mathbb{R}$, $k \in \mathbb{N}$ $\mathbf{E}_{(0,\gamma+k)}[\varphi - \psi] = \frac{c(0, \gamma+k)}{c(0, \gamma)} \mathbf{E}_{(0,\gamma)}[e^{-kS}(\varphi - \psi)] = 0 \quad \forall k$, also

$$\mathbf{E}_{(0,\eta)}[g(e^{-S})(\varphi - \psi)] = 0$$

für jedes Polynom g , gemäß dem Weierstraß'schen Approximationssatz also für jedes stetige $g : [0, 1] \rightarrow \mathbb{R}$ und jedes $\eta > 0$. Für gegebenes h setze

$$g_\delta(x) := \begin{cases} h(\log(1/x))x^\delta, & 0 < x \leq 1, \\ 0, & x = 0, \end{cases}$$

(g_δ ist stetig, und $g_\delta(e^{-s}) = h(s)e^{-\delta s}$), demnach

$$\mathbf{E}_{(0,\eta)} \left[h(S)e^{-\delta S}(\varphi - \psi) \right] = \mathbf{E}_{(0,\eta)} [g_\delta(e^{-S})(\varphi - \psi)] = 0$$

für jedes $\delta > 0$, was (*) impliziert.

Betrachte ein $(\mu, \eta) \in \Theta_1$, der Likelihood-Quotient hat die Gestalt

$$R_{(\mu,\eta),(0,\eta)} = c(\mu, \eta)e^{\mu\tilde{M}_n}.$$

Demnach hat der Verwerfungsbereich von φ die Form

$$\{R_{(\mu,\eta),(0,\eta)} > h(S)\} \text{ mit } h(S) = c_{(\mu,\eta)}e^{\mu f(S)},$$

also

$$\begin{aligned} \mathbf{E}_{(\mu,\eta)}[\varphi - \psi] &= \mathbf{E}_{(0,\eta)}[R_{(\mu,\eta),(0,\eta)}(\varphi - \psi)] - \underbrace{\mathbf{E}_{(0,\eta)}[h(S)(\varphi - \psi)]}_{=0 \text{ gemäß } (*)} \\ &= \mathbf{E}_{(0,\eta)} \left[\underbrace{(R_{(\mu,\eta),(0,\eta)} - h(S))(\varphi - \psi)}_{\geq 0} \right] \geq 0, \end{aligned}$$

d.h. die Macht von φ ist nicht kleiner als die von ψ .

□

11. χ^2 -Anpassungstest

11.1. Test auf eine vorgegebene Häufigkeitsverteilung

Gregor Mendels Erbsenexperimente

Form: rund oder kantig, Farbe: gelb oder grün. Nennen wir A rund, a kantig und B gelb, b grün. Dabei sind in diesem diploiden Genmodell A und B dominant.

Betrachte die Nachkommen eines $AaBb$: Es gibt nun 4 Möglichkeiten für den Phänotyp, mit (theoretische) Häufigkeiten gemäß den Mendel'schen Vererberungsregeln:

	gelb	grün
rund	$\frac{9}{16}$	$\frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$
kantig	$\frac{3}{16}$	$\frac{1}{16}$

Mendel beobachtete in $n = 556$ Versuchen:

	gelb	grün
rund	315	108
kantig	101	32

oder in relativen Häufigkeiten:

	gelb	grün
rund	0.567	0.194
kantig	0.181	0.058

Experiment mit s möglichen Ausgängen, Ausgang i mit Wahrscheinlichkeit $\rho(i)$, n unabhängige Wiederholungen.

X_m Ausgang des m -ten Experiments

$$h_n(i) := |\{1 \leq m \leq n : X_m = i\}|,$$

$$\mathbf{P}[h_n(1) = a_1, \dots, h_n(s) = a_s] = \begin{cases} \binom{n}{a_1, a_2, \dots, a_s} \rho(1)^{a_1} \cdot \dots \cdot \rho(s)^{a_s} & \text{falls } a_1 + a_2 + \dots + a_s = n \\ 0 & \text{sonst} \end{cases}$$

Diese Verteilung heißt *Multinomialverteilung*.

$$\sum_{i=1}^s \frac{(h_n(i) - n\rho(i))^2}{n\rho(i)} =: D_{n,\rho}(X_1, \dots, X_m)$$

Ziel: Die Verteilung von $D_{n,\rho}(X)$ unter der Multinomial($n; \rho(1), \rho(2), \dots, \rho(s)$)-Verteilung studieren. Sei $h_n^*(i) = \frac{h_n(i) - n\rho(i)}{\sqrt{n\rho(i)}}$. Dann ist

$$\sum_{i=1}^s \sqrt{\rho(i)} \cdot h_n^*(i) = 0$$

und

$$H_\rho := \{x \in \mathbb{R}^s : x^t \cdot u_\rho = 0\}, \quad u_\rho = (\sqrt{\rho(1)}, \dots, \sqrt{\rho(s)})^t,$$

$$\Pi_\rho : \mathbb{R}^2 \rightarrow H_\rho \text{ die Projektion auf } H_\rho.$$

Sei O eine orthogonale $s \times s$ -Matrix, deren letzte Spalte u_ρ ist. Dann wähle

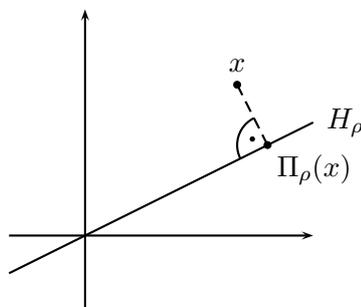


Abbildung 11.1.: Darstellung der orthogonalen Projektion im Zweidimensionalen auf einen eindimensionalen Unterraum

$$\Pi_\rho = O \cdot \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 & \\ & & & 0 \end{pmatrix} O^t$$

Multivariate Standardnormalverteilung auf H_ρ :

$$\mathcal{N}(0, 1)^{\otimes s} \cdot \Pi_\rho^{-1} = \mathcal{N}_\rho$$

Bemerkung:

$$(\mathcal{N}(0, 1)^{\otimes(s-1)} \otimes \delta_0) \circ O^t = \mathcal{N}_\rho$$

11.1.1. Satz:

$$(h_n^*(1), \dots, h_n^*(s))^t \xrightarrow[n \rightarrow \infty]{v} \mathcal{N}_\rho$$

Erinnerung:

$Y_n \xrightarrow[n \rightarrow \infty]{v} \mathbf{Q} : \iff \mathbf{P}[Y_n \in A] \xrightarrow[n \rightarrow \infty]{} \mathbf{Q}(A) \quad \forall A \text{ mit } \mathbf{Q}[\partial A] = 0.$ Für die Konvergenz genügt es, $\mathbf{P}[Y_n \in A] \rightarrow \mathbf{Q}(A)$ für A der Form $(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_s]$ mit $a_i \in \mathbb{R}$ zu zeigen.

Beobachtung:

$Z_k(i), k = 1, 2, \dots, 1 \leq i \leq s$ unabhängig und $Z_k(i) \sim \text{Pois}(\rho(i)).$

Betrachte $S_n(i) := \sum_{k=1}^n Z_k(i), N_n := S_n(1) + S_n(2) + \dots + S_n(s).$ Dann ist $S_n(i) \sim \text{Pois}(n\rho(i)),$ $S_n(i)$ sind unabhängig, $N_n \sim \text{Pois}(n)$ und gegeben $N_n = m$ ist $(S_n(1), \dots, S_n(s))$

Multinomial($n; \rho(1), \dots, \rho(s)$)-verteilt, denn

$$\mathbf{P}[S_n(1) = a_1, \dots, S_n(s) = a_s | N_n = m] = \frac{1}{e^{-n} \frac{n^m}{m!}} \prod_{i=1}^s e^{-\rho(i)n} \frac{(n\rho(i))^{a_i}}{a_i!} = \binom{m}{a_1, \dots, a_s} \rho(1)^{a_1} \dots \rho(s)^{a_s}.$$

Setze

$$S_n^*(i) := \frac{S_n(i) - n\rho(i)}{\sqrt{n\rho(i)}}, \text{ so gilt (mit dem ZGWS) } S_n^* := \begin{pmatrix} S_n^*(1) \\ \vdots \\ S_n^*(s) \end{pmatrix} \xrightarrow{v} \mathcal{N}(0, 1)^{\otimes s}.$$

Sei weiter

$$N_n^* := \sum_{i=1}^s \sqrt{\rho(i)} S_n^*(i) = \frac{N_n - n}{\sqrt{n}}$$

und

$$Y_n^* := O^t S_n^*, \quad Y_n^*(s) = N_n^*.$$

Es gilt: $Y_n^* \xrightarrow[n \rightarrow \infty]{v} \mathcal{N}(0, 1)^{\otimes s}$,

$$\mathbf{P}[h_n^* \in A] = \mathbf{P}[S_n^* \in A | N_n^* = n] = \mathbf{P}[Y_n^* \in O^t A | Y_n^*(s) = 0].$$

Zeige: $\forall \varepsilon > 0$: $\mathbf{P}[h_n^* \in A] \leq \mathbf{P}[S_n^* \in A | N_n^* \in [-\varepsilon, 0]]$ und $\mathbf{P}[h_n^* \in A] \geq \mathbf{P}[S_n^* \in A | N_n^* \in [0, \varepsilon]]$ für A der Form $(-\infty, a_1] \times \dots \times (-\infty, a_s]$.

Sei $h_{m,n}^*(i) := \frac{h_m(i) - n\rho(i)}{\sqrt{n\rho(i)}}$, $q_{m,n} := \mathbf{P}[h_{m,n}^* \in A]$.

$\{h_{m+1,n}^* \in A\} \subset \{h_{m,n}^* \in A\}$, d.h. $q_{m+1,n} \leq q_{m,n}$, also

$$q_{n,n} \geq \sum_{m=n}^{\lfloor n+\varepsilon\sqrt{n} \rfloor} q_{m,n} \mathbf{P}[N_n = m | N_n^* \in [0, \varepsilon]] = \mathbf{P}[S_n^* \in A | N_n^* \in [0, \varepsilon]],$$

analog die andere Ungleichung $q_{n,n} \leq \mathbf{P}[S_n^* \in A | N_n^* \in [-\varepsilon, 0]]$.

$$\mathbf{P}[S_n^* \in A] \rightarrow \prod_{i=1}^s \int_{-\infty}^{a_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Sei weiter $U_\varepsilon := \{x \in \mathbb{R}^s : x_s \in [0, \varepsilon]\}$. Dann gilt:

$$\mathbf{P}[S_n^* \in A | N_n^* \in [0, \varepsilon]] = \frac{\mathbf{P}[\{Y_n^* \in O^t A\} \cap U_\varepsilon]}{\mathbf{P}[Y_n^*(s) \in [0, \varepsilon]]} \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1)^{\otimes s}(O^t A | U_\varepsilon) =: q_\varepsilon$$

Dann ist

$$q_\varepsilon = \frac{1}{\mathcal{N}(0, 1)([0, \varepsilon])} \int_0^\varepsilon \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} \underbrace{\mathcal{N}(0, 1)^{\otimes (s-1)}(\{x \in \mathbb{R}^{s-1} : (x, r) \in O^t A\})}_{\xrightarrow{r \rightarrow 0} \mathcal{N}_\rho(A)} dr \xrightarrow{\varepsilon \rightarrow 0} \mathcal{N}_\rho(A).$$

Demnach $\liminf q_{n,m} \geq \mathcal{N}_\rho(A)$, analog für \limsup .

11.1.2. Korollar:

Unter \mathbf{P}_ρ gilt:

$$\sum_{i=1}^s (h_n^*(i))^2 = D_{n,\rho} \xrightarrow[n \rightarrow \infty]{v} \chi_{s-1}^2,$$

denn

$$\begin{aligned} \mathbf{P}_\rho \left[\sum_{i=1}^s (h_n^*(i))^2 \leq c \right] &= P_\rho[h_n^* \in \{x \in \mathbb{R}^s : \|x\|^2 \leq c\}] \\ &\xrightarrow[n \rightarrow \infty]{} \mathcal{N}_\rho(\{x : \|x\|^2 \leq c\}) = \left(\mathcal{N}(0,1)^{\otimes(s-1)} \otimes \delta_0 \right) (\{x \in \mathbb{R}^s : \|x\|^2 \leq c\}) = \chi_{s-1}^2([0, c]). \end{aligned}$$

11.1.3. χ^2 -Anpassungstest

Zu $\alpha \in (0, 1)$ sei $c = (1 - \alpha)$ -Quantil von χ_{s-1}^2 . Der Test von $H_0 : \vartheta = \rho$ gegen $H_1 : \vartheta \neq \rho$ mit Ablehungsbereich $\{D_n > c\}$ hat (asymptotisches) Niveau α .

Zurück zu Mendel:

$D_{556} = 0.47$, $\chi_3^2([0, 0.47]) = 0.075$, der χ^2 -Test lehnt H_0 nicht ab.

11.2. χ^2 -Test auf Unabhängigkeit:

Seien X_1, X_2, \dots unabhängige Werte in $E = \{1, \dots, a\} \times \{1, \dots, b\} =: A \times B$, $a, b > 1$
 $\mathbf{P}_\vartheta[X_1 = (i, j)] = \vartheta(i, j)$,

$$\Theta = \{\vartheta = (\vartheta(i, j))_{i \in A, j \in B} : \vartheta(i, j) > 0, \sum_{i,j} \vartheta(i, j) = 1\}$$

Marginalverteilungen:

$$\vartheta(i, \cdot) := \sum_{j \in B} \vartheta(i, j), \vartheta(\cdot, j) := \sum_{i \in A} \vartheta(i, j)$$

$$\begin{aligned} \Theta_0 &:= \{\vartheta \in \Theta : \vartheta(i, j) = \vartheta(i, \cdot) \vartheta(\cdot, j) \forall i, j\} \\ &= \{\vartheta \in \Theta : \vartheta(i, j) = \alpha(i) \cdot \beta(j) \text{ für W'keitsvektoren } \alpha \text{ auf } A \text{ und } \beta \text{ auf } B\} \end{aligned}$$

Betrachte nun

$$h_n(i, j) = |\{1 \leq k \leq n : X_k = (i, j)\}|$$

und berechne

$$\tilde{D}_n := \sum_{i,j} \frac{h_n(i, j) - n \cdot \frac{1}{n} h_n(i, \cdot) \cdot \frac{1}{n} h_n(\cdot, j)}{\frac{1}{n} h_n(i, \cdot) \cdot h_n(\cdot, j)} = \sum_{i,j} \frac{h_n(i, j) - \frac{1}{n} h_n(i, \cdot) \cdot h_n(\cdot, j)}{\frac{1}{n} h_n(i, \cdot) \cdot h_n(\cdot, j)}.$$

11.2.1. Satz:

Für $\vartheta = \alpha \otimes \beta \in \Theta_0, c > 0$ gilt

$$\mathbf{P}_\vartheta[\tilde{D}_n \leq c] \xrightarrow[n \rightarrow \infty]{} \chi_{(a-1)(b-1)}^2([0, c]).$$

Beweis:

Sei $s = a \cdot b, r = (a - 1) \cdot (b - 1), \vartheta = \alpha \otimes \beta$

$$\tilde{h}_n(i, j) := \frac{h_n(i, j) - \frac{1}{n}h_n(i, \cdot)h_n(\cdot, j)}{\sqrt{\frac{1}{n}h_n(i, \cdot)h_n(\cdot, j)}}$$

somit $\tilde{D}_n = \|h_n\|^2$.

Beweisidee:

$$h_{n, \alpha\beta}^* := \left(\frac{h_n(i, j) - n\alpha(i)\beta(j)}{\sqrt{n\alpha(i)\beta(j)}} \right)_{ij} \approx \mathcal{N}_{\alpha\otimes\beta} \text{ unter } \mathbf{P}_{\alpha\otimes\beta}$$

(gemäß Satz 11.1.1)

Definiere:

$$h_n^\circ(i, j) = \frac{h_n(i, j) - \frac{1}{n}h_n(i, \cdot)h_n(\cdot, j)}{\sqrt{n\alpha(i)\beta(j)}}$$

sei

$$a_l = (a_l(i, j)) = (\sqrt{\alpha(i)}\delta_{jl}) = \begin{pmatrix} \sqrt{\alpha(1)} \\ \vdots \\ \sqrt{\alpha(a)} \end{pmatrix},$$

wobei $\begin{pmatrix} \sqrt{\alpha(1)} \\ \vdots \\ \sqrt{\alpha(a)} \end{pmatrix}$ der l -te Spaltenvektor von b -vielen Spalten ist.

und

$$b_k = (b_k(i, j)) = (\sqrt{\beta(j)}\delta_{ki}) = \begin{pmatrix} 0 & & \\ \sqrt{\beta(1)} & \dots & \sqrt{\beta(b)} \\ & & 0 \end{pmatrix}$$

Es gilt $h_n^\circ \cdot a_l = 0, h_n^\circ \cdot b_k = 0, M = \text{span}(a_l, b_k, l = 1, \dots, b, k = 1, \dots, a)$

Behauptung: $\dim M = a + b - 1$:

$$\sum_1^b \sqrt{\beta(l)}a_l = \sum_1^a \sqrt{\alpha(k)}b_k = (\sqrt{\alpha(i)} \cdot \sqrt{\beta(j)})_{i,j} =: u_s,$$

also $\dim M \leq a + b - 1$, zeige $a_1, \dots, a_b, b_1, \dots, b_{a-1}$ sind linear unabhängig.(Übung)

$L :=$ orthogonales Komplement von $M, h_n^\circ \in L$.

Ergänze u_s zu Orthonormalbasis $u_{r+1}, u_{r+2}, \dots, u_s$ von M , diese zu Orthonormalbasis u_1, \dots, u_s von $\mathbb{R}^s = L \oplus M$. Setze:

$$\mathcal{O}_{\alpha\beta} := \begin{pmatrix} u_1 & \dots & u_s \\ | & & | \\ | & & | \end{pmatrix} \text{ ist orthogonal, } \Pi_{\alpha\beta} = \mathcal{O}_{\alpha\beta} \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} \mathcal{O}_{\alpha\beta}^t$$

wobei in der Diagonalen nur r -viele 1-er stehen, ist die Projektion auf L .

Zeige: $\left\| h_n^\circ - \Pi_{\alpha\beta} h_{n,\alpha\beta}^* \right\|_{n \rightarrow \infty} \rightarrow 0$ in Verteilung unter $\mathbf{P}_{\alpha\otimes\beta}$.

Es gilt:

$$\Pi_{\alpha\beta} h_{n,\alpha\beta}^*(i, j) = \frac{h_n(i, j) + n\alpha(i)\beta(j) - \alpha(i)h_n(\cdot, j) - h_n(i, \cdot)\beta(j)}{\sqrt{n\alpha(i)\beta(j)}},$$

denn rechte Seite $\perp a_l, b_k$, also rechte Seite $\in L$ und h_n^* - Rechte Seite $\in \text{span}(a_l, b_k) = L^\perp$.

$$(h_n^\circ - \Pi_{\alpha\beta} h_{n,\alpha\beta}^*)(i, j) = \eta_n^A(i) \cdot \eta_n^B(j),$$

wo $\eta_n^A(i) := \frac{n^{\frac{1}{4}}}{\sqrt{\alpha(i)}} \left(\frac{1}{n} h_n(i, \cdot) - \alpha(i) \right)$, $\eta_n^B(j) := \frac{n^{\frac{1}{4}}}{\sqrt{\beta(j)}} \left(\frac{1}{n} h_n(\cdot, j) - \beta(j) \right)$ und es gilt

$$\eta_n^A(i), \eta_n^B(j) \xrightarrow[n \rightarrow \infty]{V} 0 \text{ unter } \mathbf{P}_{\alpha\otimes\beta}$$

(mit ZGWS oder Čebyšev-Ungleichung).

Wir wissen:

$$h_{n,\alpha}^* \xrightarrow[n \rightarrow \infty]{V} \mathcal{N}_{\alpha\otimes\beta} = (\mathcal{N}(0, 1)^{\otimes(s-1)} \otimes \delta_0) \circ \mathcal{O}_{\alpha\beta}^{-1}$$

□

11.2.2. Korollar: χ^2 -Test auf Unabhängigkeit

Sei $\alpha \in (0, 1)$, $c = (1 - \alpha)$ -Quantil von $\chi_{(a-1)(b-1)}^2$. Der Test von $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta \setminus \Theta_0$ mit Ablehnungsbereich $\{\tilde{D}_n > c\}$ hat asymptotisches Niveau α .

Bemerkung:

$\dim \Theta = a \cdot b - 1$ und $\dim(\Theta_0) = a - 1 + b - 1$. Es ist:

$$(a - 1)(b - 1) = ab - 1 - ((a - 1) + (b - 1)).$$

Das Schätzen der unbekanntten Marginalverteilungen „verbraucht“ $a + b - 2$ Freiheitsgrade.

Beispiel: Zulassungsstatistik der Berkeley-Kandidaten, Simpson-Paradoxon

Im Herbst 1973 haben sich an der Universität Berkeley 12763 Kandidaten für ein Studium beworben, davon 8442 Männer und 4321 Frauen. Es kam zu folgenden Zulassungszahlen:

	Aufgenommen	Abgelehnt
Männer	3738	4704
Frauen	1494	2827

Demnach betrug die Zulassungsquote bei den Männern ca. 44%, bei den Frauen nur ca. 35%. Ein χ^2 -Test auf Unabhängigkeit (mit \mathbf{R}) zeigt, dass eine solche Unverhältnismäßigkeit nur mit verschwindend kleiner Wahrscheinlichkeit durch „reinen Zufall“ entsteht:

```

> berkeley <- matrix(c(3738,4704,1494,2827),nrow=2,byrow=TRUE)
> berkeley
      [,1] [,2]
[1,] 3738 4704
[2,] 1494 2827
> chisq.test(berkeley)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data: berkeley
X-squared = 110.8489, df = 1, p-value < 2.2e-16

```

Dieser Fall hat einiges Aufsehen erregt, er wurde u.A. diskutiert in P.J. Bickel, E.A. Hammel, J.W. O'Connell, Sex Bias in Graduate Admissions: Data from Berkeley, *Science*, **187**, no. 4175, 398–404 (1975). Bemerkenswerterweise verschwindet das Ungleichgewicht, wenn man die Zulassungszahlen nach Departments aufspaltet: Es stellt sich heraus, dass innerhalb der Departments die Aufnahmewahrscheinlichkeiten nicht signifikant vom Geschlecht abhängen, aber sich Frauen häufiger bei Departments mit (absolut) niedriger Aufnahmequote beworben haben als Männer – dies ist ein Beispiel für das sogenannte *Simpson-Paradox*. Die genauen nach Departments aufgeschlüsselten Bewerber- und Zulassungszahlen sind leider nicht öffentlich zugänglich (siehe aber Abb. 1 in Bickel et. al, loc. cit, für eine grafische Aufbereitung der Daten, die den Simpson-Effekt zeigt). Bickel et. al demonstrieren das Phänomen mittels eines hypothetischen Beispiels:

	Aufgenommen	Abgelehnt
	<i>Department of machismathics</i>	
Männer	200	200
Frauen	100	100
	<i>Department of social warfare</i>	
Männer	50	100
Frauen	150	300
	<i>Gesamt</i>	
Männer	250	300
Frauen	250	400

12. Regression und das lineare Modell

12.1. „Einfache“ Regression:

Modell:

Seien $X_k = \gamma_0 + \gamma_1 \cdot t_k + \sqrt{v}\xi_k, k = 1, \dots, n$ unsere Beobachtungen (Zielvariablen), wo γ_1, γ_2 die unbekanntenen Koeffizienten, t_1, \dots, t_n bekannte Werte (die Ausgangs- oder Regressorvariable) und ξ_1, \dots, ξ_n Störgrößen (mit $\mathbf{E}[\xi_i] = 0, \text{Var}(\xi_i) = 1$), sowie $v > 0$ der Varianzparameter. (Wir nehmen dabei an, dass nicht alle t_i identisch sind.)

Problem:

Anhand der X_i und t_i auf γ_0 und γ_1 schließen.

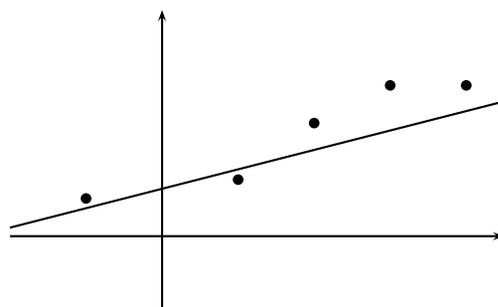


Abbildung 12.1.: Ausgleichsgerade zu gegebenen Punkten

Vorschlag: Regressionsgerade via kleinste Quadrate

Für $\gamma = \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}$ und Realisierungen x_1, \dots, x_n von X_1, \dots, X_n sei

$$\begin{aligned} F_\gamma &:= \frac{1}{n} \sum_{i=1}^n (x_i - (\gamma_0 + \gamma_1 \cdot t_i))^2 \\ &= \frac{1}{n} |X - (\gamma_0 \mathbf{1} + \gamma_1 \vec{t})|^2 \end{aligned}$$

wähle $\hat{\gamma}$ so, dass $F_{\hat{\gamma}} = \min_{\gamma} F_\gamma$.

Bemerkung:

Die Abbildung $\gamma \mapsto F_\gamma$ ist glatt und $\lim_{|\gamma| \rightarrow \infty} F_\gamma = \infty$.

$$\begin{aligned} 0 &= \frac{\partial}{\partial \gamma_0} F_\gamma \\ &= -\frac{2}{n} \sum_{i=1}^n (X_i - (\gamma_0 + \gamma_1 t)) \\ &\quad \text{und} \\ 0 &= \frac{\partial}{\partial \gamma_1} F_\gamma \\ &= -\frac{2}{n} \sum_{i=1}^n t_i (X_i - \gamma_0 - \gamma_1 t) \end{aligned}$$

$$\iff \gamma_0 + \gamma_1 M(t) = M(X) \quad (1),$$

wo $M(u) = \frac{1}{n} \sum_{i=1}^n u_i$ und

$$\gamma_0 M(t) + \gamma_1 \frac{1}{n} \sum_{i=1}^n t_i^2 = \frac{1}{n} \sum_{i=1}^n t_i X_i \quad (2)$$

Weiter gilt:

$$(2) \iff \gamma_0 M(t) + \gamma_1 (V(t) + M(t)^2) = C(X, t) + M(X)M(t) \quad (3),$$

mit $V(t) = \frac{1}{n} \left(\sum_{i=1}^n t_i^2 \right) - M(t)^2$ und $C(X, t) = \frac{1}{n} \left(\sum_{i=1}^n t_i X_i \right) - M(X)M(t)$

Setze (1) in (3) ein:

$$\gamma_1 \cdot V(t) = C(X, t)$$

12.1.1. Satz:

$$\hat{\gamma}_1 = \frac{C(X, t)}{V(t)}, \hat{\gamma}_0 = M(X) - \frac{M(t)}{V(t)} C(X, t)$$

sind die eindeutig bestimmten kleinste-Quadrate-Schätzer für γ . Sie sind erwartungstreue Schätzer für γ .

Beweis:

Eindeutigkeit ist klar.

Zur Erwartungstreue: Sei $\vartheta := (\gamma, v)$. Dann

$$\begin{aligned} \mathbf{E}_\vartheta[\hat{\gamma}_1] &= \frac{1}{V(t)} \mathbf{E} \left[\frac{1}{n} \sum_{i=1}^n t_i X_i - M(t) \frac{1}{n} \sum_{i=1}^n X_i \right] \\ &= \frac{1}{V(t)} \mathbf{E} \left[\frac{1}{n} \sum_{i=1}^n t_i (\gamma_0 + \gamma_1 t_i + \sqrt{v} \xi_i) - M(t) \frac{1}{n} \sum_{i=1}^n (\gamma_0 + \gamma_1 t_i + \sqrt{v} \xi_i) \right] \\ &= \frac{\gamma_1 V(t)}{V(t)} \end{aligned}$$

und ebenso:

$$\mathbf{E}_\vartheta[\hat{\gamma}_0] = \underbrace{\mathbf{E}_\vartheta[M(X)]}_{\gamma_0 + \gamma_1 M(t)} - \underbrace{\frac{M(t)}{V(t)} \mathbf{E}_\vartheta \left[\frac{1}{n} \sum_{i=1}^n t_i (\gamma_0 + \gamma_1 t_i + \sqrt{v} \xi_i) - M(t) \frac{1}{n} \sum_{i=1}^n (\gamma_0 + \gamma_1 t_i + \sqrt{v} \xi_i) \right]}_{= \frac{M(t)}{V(t)} \cdot \gamma_1 V(t)}$$

Bemerkung:

- (i) Die Regressionsgerade verläuft durch den Schwerpunkt $(M(t), M(X))^t$ der „Punktwolke.“
- (ii) Der „Regressionsvektor“

$$\hat{\gamma}_0 \vec{1} + \hat{\gamma}_1 \vec{t} = \hat{\gamma}_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \hat{\gamma}_1 \cdot \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}$$

ist die (orthogonale) Projektion auf $L = \text{span}(\vec{1}, \vec{t})$.

12.2. Das lineare Modell:

Szenario:

Sei $n, s \in \mathbb{N}, s < n$, Parameter $\gamma \in \mathbb{R}^s$, Varianzparameter $v > 0$, „Designmatrix“ $(n \times s)$ -Matrix A mit Rang s , ξ_1, \dots, ξ_n Störgrößen, $\mathbf{E}[\xi_i] = 0, \text{Var}(\xi_i) = 1$, unkorreliert. Beobachtungen $X = (X_1, \dots, X_n)^t$ entsteht als

$$X = A \cdot \gamma + \sqrt{v} \xi.$$

Bemerkung:

Sei $L = \{A\eta : \eta \in \mathbb{R}^s\}$ der Spaltenraum von A , $\Pi_L : \mathbb{R}^n \rightarrow L$ die orthogonale Projektion auf L . Dabei gilt für alle $x \in \mathbb{R}^n$:

(i) $\Pi_L x \in L$ und $|x - \Pi_L x|^2 \leq \min_{u \in L} |x - u|^2$

(ii) $\Pi_L x \in L$ und $x - \Pi_L x \perp L$

(und $\Pi_L x$ ist durch (i) und durch (ii) charakterisiert)

(ii) \Leftrightarrow (ii)' $\Pi_L x \in L$ und $u^t(x - \Pi_L x) = 0$ für u aus einer Basis von $L \iff A^t(x - \Pi_L x) = 0$

Bemerkung:

$A^t A$ ist invertierbar, und $\Pi_L = A(A^t A)^{-1} A^t$.

Beweis der Bemerkung:

Angenommen $A^t A c = 0$ für ein $c \in \mathbb{R}^s \setminus \{0\}$. Dann ist

$$|Ac|^2 = c^t A^t A c = 0 \text{ im } \not\Leftarrow \text{ zu } \text{Rang}(A) = s.$$

Weiterhin ist

$$A(A^t A)^{-1} A^t x \in L, \quad A^t(x - A(A^t A)^{-1} A^t x) = A^t x - A^t x = 0 \text{ für jedes } x \in \mathbb{R}^n.$$

12.2.1. Satz von Gauß-Markov

Die Störgrößen ξ_i erfüllen $E[\xi_i] = 0$, $\text{Var}(\xi_i) = 1$ und seien unkorreliert. Dann gilt:

- (i) Der kleinste Quadrate Schätzer $\hat{\gamma} = (A^t A)^{-1} A^t X$ ist erwartungstreu für γ .
- (ii) Ist $\tau : \mathbb{R}^s \rightarrow \mathbb{R}$ eine lineare Kenngröße von γ (d.h. $\tau(\gamma) = c^t \gamma$ für ein $c \in \mathbb{R}^s$), dann ist $T := c^t \hat{\gamma}$ erwartungstreuer Schätzer für $\tau(\gamma)$ und hat unter allen erwartungstreuen linearen Schätzern für $\tau(\gamma)$ die kleinste Varianz.
- (iii) $V^* := \frac{|X - \Pi_L X|^2}{n - s} = \frac{|X|^2 - |\Pi_L X|^2}{n - s} = \frac{|X - A\hat{\gamma}|^2}{n - s}$ ist ein erwartungstreuer Schätzer für die Varianz v .

Beweis:

Sei $\vartheta = (\gamma, v)$.

(i)

$$\begin{aligned}
 \begin{pmatrix} \mathbf{E}_\vartheta(\hat{\gamma}_1) \\ \vdots \\ \mathbf{E}_\vartheta(\hat{\gamma}_s) \end{pmatrix} &= \mathbf{E}_\vartheta[\hat{\gamma}] \\
 &= \mathbf{E}_\vartheta[(A^t A)^{-1} A^t X] \\
 &= (A^t A)^{-1} A^t \underbrace{\mathbf{E}_\vartheta[X]}_{=A\gamma} \\
 &= \gamma
 \end{aligned}$$

(Beobachtung: $\Pi_L X = \underbrace{A(A^t A)^{-1} A^t X}_{=\hat{\gamma}}$, also ist es der kleinste-Quadrate-Schätzer.)

(ii) $\mathbf{E}_\vartheta[T] = c^t \mathbf{E}_\vartheta[\hat{\gamma}] = c^t \gamma = \tau(\gamma)$, also erwartungstreu für $\tau(\gamma)$.

Sei $a := A(A^t A)^{-1} c (\in L)$. Zeige: $T = a^t X$.

Klar, denn

$$\Pi_L a = a \iff a^t \Pi_L = a^t, A^t a = c \implies a^t A = c^t,$$

somit $\tau(\gamma) = c^t \gamma = a^t A \gamma$, $\tau(\hat{\gamma}) = a^t A (A^t A)^{-1} A X = a^t X$.

Sei $S := b^t X$ irgendein linearer Schätzer mit $\mathbf{E}_\vartheta[S] = \tau(\gamma)$ ($\forall \vartheta = (\gamma, v)$), also

$$b^t A \gamma = \mathbf{E}_\vartheta[b^t X] = \tau(\gamma) = \mathbf{E}_\vartheta[a^t X] = a^t A \gamma \text{ für jedes } \gamma \in \mathbb{R}^s, \text{ d.h. } (b - a)^t u = 0 \quad \forall u \in L,$$

($\Leftrightarrow b - a \perp L$), folglich ist $a = \Pi_L b$, insbesondere $|a|^2 \leq |b|^2$.

$$\begin{aligned} \text{Var}_\vartheta(S) - \text{Var}_\vartheta(T) &= \mathbf{E}_\vartheta[(\underbrace{b^t(X - A\gamma)}_{\sqrt{v\xi}})^2 - (a^t \underbrace{(X - A\gamma)}_{\sqrt{v\xi}})^2] \\ &= v\mathbf{E}_\vartheta[b^t \xi \xi^t b - a^t \xi \xi^t a] \\ &= v(b^t \underbrace{\mathbf{E}_\vartheta[\xi \xi^t]}_{=I_n} b - a^t \underbrace{\mathbf{E}_\vartheta[\xi \xi^t]}_{=I_n} a) \\ &= v(|b|^2 - |a|^2) \geq 0 \end{aligned}$$

(iii) Sei u_1, \dots, u_s eine Orthonormalbasis von L , ergänze zu ONB u_1, \dots, u_n von \mathbb{R}^n . Betrachte

die Matrix $O := \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ | & | & \dots & | \end{pmatrix}$, $\Pi_L = O \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ 0 & & & \ddots \\ & & & & 0 \end{pmatrix} O^t$ mit s Einsen in

der Diagonalen ist die Projektion auf L , $\eta = O^t \xi$.

Es ist

$$\begin{aligned} (n-s)V^* &= |A\gamma + \sqrt{v}\xi - \Pi_L(A\gamma + \sqrt{v}\xi)|^2 \\ &= v|\xi - \Pi_L \xi|^2 \\ &= v \left| O\eta - O \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ 0 & & & \ddots \\ & & & & 0 \end{pmatrix} \underbrace{O^t O \eta}_{=I_n} \right| \\ &= v \left| \eta - \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ 0 & & & \ddots \\ & & & & 0 \end{pmatrix} \eta \right| \\ &= v \sum_{k=s+1}^n \eta_k^2, \end{aligned}$$

und

$$\mathbf{E}_\vartheta[\eta_k^2] = \mathbf{E}_\vartheta \left[\sum_{i,j} O_{ik} O_{jk} \xi_i \xi_j \right] = \sum_i O_{ik} O_{ik} = \sum_i O_{ik} O_{ki}^t = (OO^t)_{ii} = 1.$$

12.3. Das normale lineare Modell:

Wir nehmen nun zusätzlich an, dass die ξ_i u.a., $\mathcal{N}(0, 1)$ -verteilt sind.

12.3.1. Verallgemeinerung des Satzes von Student

Im normalverteilten linearen Modell gilt unter $\mathbf{P}_{(\gamma,v)}$:

- (i) $\hat{\gamma} \sim \mathcal{N}(\gamma, v(A^t A)^{-1})$
- (ii) $\frac{n-s}{v} V^*$ ist χ_{n-s}^2 -verteilt und unabhängig von $\hat{\gamma}$.
- (iii) $\frac{1}{v} |A(\hat{\gamma} - \gamma)|^2 = \frac{1}{v} |\Pi_L X - \mathbf{E}_{(\gamma,v)}[X]|^2$ ist χ_s^2 -verteilt und ist unabhängig von V^* , insbesondere:

$$\frac{|A(\hat{\gamma} - \gamma)|^2}{s \cdot V^*} \text{ ist Fisher}_{s, n-s}\text{-verteilt}$$

- (iv) $H \subset L$ linearer Teilraum von L , $\dim H = r < s$ und $A\gamma \in H$, dann ist $\frac{1}{v} |\Pi_L X - \Pi_H X|^2$ χ_{s-r}^2 -verteilt und es ist unabhängig von V^* , insbesondere ist

$$F_{H,L} := \frac{n-s}{s-r} \frac{|\Pi_L X - \Pi_H X|^2}{|X - \Pi_L X|^2} = \frac{|A\hat{\gamma} - \Pi_H X|^2}{(s-r)V^*} \text{ Fisher}_{s-r, n-s}\text{-verteilt.}$$

Bemerkung:

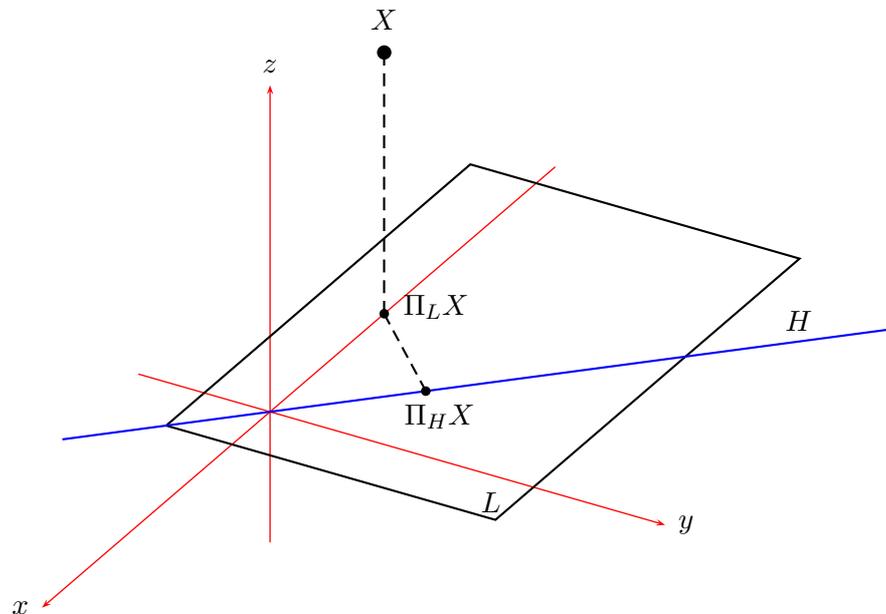


Abbildung 12.2.: Darstellung der Projektionen eines Punktes X im \mathbb{R}^3 auf $H \subset L \subset \mathbb{R}^3$, L ist dabei die Ebene, die durch das Rechteck L dargestellt werden soll und H der Unterraum, der von der blauen Geraden aufgespannt wird.

Es gilt

$$|\Pi_L X - \Pi_H X|^2 = |\Pi_L X|^2 - |\Pi_H X|^2 = |X - \Pi_H X|^2 - |X - \Pi_L X|^2 \quad (\text{Pythagoras})$$

und

$$|X - \Pi_L X|^2 = |X|^2 - |\Pi_L X|^2.$$

Beweis des Satzes:

- (i) $\hat{\gamma}$ ist s -dimensional normalverteilt; es gilt:

$$\mathbf{E}_{(\gamma,v)}[\hat{\gamma}] = (A^t A)^{-1} A \cdot \mathbf{E}_{(\gamma,v)}[X] = \gamma$$

und

$$\text{Cov}(\hat{\gamma}) = (A^t A)^{-1} A \cdot v I_n \cdot A^t ((A^t A)^{-1})^t = v (A^t A)^{-1}.$$

- (ii) Sei o.B.d.A. u_1, \dots, u_n eine ONB von \mathbb{R}^n mit $H = \text{span}(u_1, \dots, u_r)$, $L = \text{span}(u_1, \dots, u_s)$, $O = (u_1, \dots, u_n)$, $\eta := O^t \xi$, η ist $\mathcal{N}(0, 1)^{\otimes n}$ -verteilt.

$$\begin{aligned} \frac{n-s}{v} V^* &= \frac{1}{v} |A\gamma + \sqrt{v}\xi - \Pi_L(A\gamma + \sqrt{v}\xi)|^2 \\ &= |\xi - \Pi_L \xi|^2 \\ &= \sum_{i=s+1}^n \eta_i^2 \sim \chi_{n-s}^2 \end{aligned}$$

- (iii)

$$\frac{1}{v} |\Pi_L X - A\gamma|^2 = |\Pi_L \xi|^2 = \sum_{i=1}^s \eta_i^2 \sim \chi_s^2$$

und ist unabhängig von V^* .

- (iv) $\frac{1}{v} |\Pi_L X - \Pi_H X|^2 = |\Pi_L \xi - \Pi_H \xi|^2 = \sum_{i=r+1}^s \eta_i^2 \sim \chi_{s-r}^2$ und unabhängig von V^* .

12.3.2. Korollar: Konfidenzbereiche im normalverteilten linearen Modell

Sei $\alpha \in (0, 1)$.

- (i) $C_{\hat{\gamma}} := \{\tilde{\gamma} \in \mathbb{R}^s : |A(\tilde{\gamma} - \hat{\gamma})|^2 < s \cdot V^* \cdot f_{s, n-s; 1-\alpha}\}$ ist ein Konfidenzbereich für γ zum Sicherheitsniveau $1 - \alpha$, wobei $f_{s, n-s; 1-\alpha}$ das $(1 - \alpha)$ -Quantil der Fisherverteilung mit s und $n - s$ Freiheitsgraden ist.
- (ii) Sei $\tau(\gamma) = c^t \gamma$ ($c \in \mathbb{R}^s$) ein lineares Parametermerkmal.

$$C_{\tau} = (c^t \hat{\gamma} - \delta \sqrt{V^*}, c^t \hat{\gamma} + \delta \sqrt{V^*}),$$

wo $\delta = \sqrt{c^t (A^t A)^{-1} c} \cdot t_{n-s, 1-\frac{\alpha}{2}}$ mit $t_{n-s, 1-\frac{\alpha}{2}}$ dem $(1 - \frac{\alpha}{2})$ -Quantil der Student'schen Verteilung mit $n - s$ -Freiheitsgraden, ist ein Konfidenzintervall für $\tau(\gamma)$ zum Sicherheitsniveau α .

- (iii) $C_v := (\frac{n-s}{\chi_{n-s, 1-\frac{\alpha}{2}}^2} V^*, \frac{n-s}{\chi_{n-s, \frac{\alpha}{2}}^2} V^*)$ ist ein Konfidenzintervall für v zum Sicherheitsniveau $1 - \alpha$.

Beweis:

- (i) Mit $\frac{|A(\tilde{\gamma} - \hat{\gamma})|^2}{s V^*} \sim f_{s, n-s}$ -verteilt \checkmark
- (ii) $Z := c^t \hat{\gamma} \sim \mathcal{N}(c^t \gamma, v c^t (A^t A)^{-1} c)$, also ist $Z^* = \frac{Z - c^t \gamma}{\sqrt{v c^t (A^t A)^{-1} c}} \sim \mathcal{N}(0, 1)$ und somit $T := \frac{Z^*}{\sqrt{\frac{V^*}{v}}} \sim t_{n-s}$. Dann ist $\mathbf{P}_{(\gamma,v)}[|T| \leq t_{n-s, 1-\frac{\alpha}{2}}] = \alpha$.

(iii)

$$\mathbf{P}_{(\gamma,v)} \left[v \in \left(\frac{n-s}{\chi_{n-s,1-\frac{\alpha}{2}}^2}, \frac{n-s}{\chi_{n-s,\frac{\alpha}{2}}^2} \right) \right] = \mathbf{P}_{(\gamma,v)} \left[\chi_{n-s,\frac{\alpha}{2}}^2 < \frac{n-s}{v} V^* < \chi_{n-s,1-\frac{\alpha}{2}}^2 \right] = 1 - \alpha$$

12.3.3. Korollar: Tests im normalverteilten linearen Modell

Sei $\alpha \in (0, 1)$.

(i) (t -Test für $c^t \gamma = m_0$, bzw. \leq oder \geq)

Der Ablehnungsbereich

$$\left\{ |c^t \hat{\gamma} - m_0| > t_{n-s,1-\frac{\alpha}{2}} \cdot \sqrt{c^t (A^t A)^{-1} c V^*} \right\}$$

definiert einen Test von $H_0 : c^t \gamma = m_0$ gegen $H_1 : c^t \gamma \neq m_0$ zum Irrtumsniveau α .

Analog definiert

$$\left\{ c^t \hat{\gamma} - m_0 \gtrless t_{n-s,1-\alpha} \cdot \sqrt{c^t (A^t A)^{-1} c V^*} \right\}$$

einen (einseitigen) Test von $H_0 : c^t \gamma \gtrless m_0$ gegen $H_1 : c^t \gamma \lessgtr m_0$.

(ii) Sei $H \subseteq L$ linearer Unterraum, $\dim H = r < s$,

$$F_{H,L} := \frac{n-s}{s-r} \frac{|\Pi_L X - \Pi_H X|^2}{|X - \Pi_L|^2}.$$

Der Ablehnungsbereich

$$\{F_{H,L} > f_{s-r,n-s;1-\alpha}\}$$

mit $f_{s-r,n-s;1-\alpha} = (1-\alpha)$ -Quantil der Fisher $_{s-r,n-s}$ -Verteilung definiert einen Test von $H_0 : A\gamma \in H$ gegen $H_1 : A\gamma \notin H$ zum Niveau α .

(iii) Sei $v_0 > 0$. Der Ablehnungsbereich

$$\{(n-s)V^* \gtrless v_0 \chi_{n-s,1-\alpha}^2\}$$

definiert einen Test von $H_0 : v \gtrless v_0$ gegen $H_1 : v \lessgtr v_0$ zum Niveau α .

Beweis:

Wir betrachten exemplarisch (ii): Für jedes $\gamma \in H_0$, $v > 0$ ist $F_{H,L}$ unter $\mathbf{P}_{(\gamma,v)}$ Fisher $_{s-r,n-s}$ -verteilt, demnach gilt

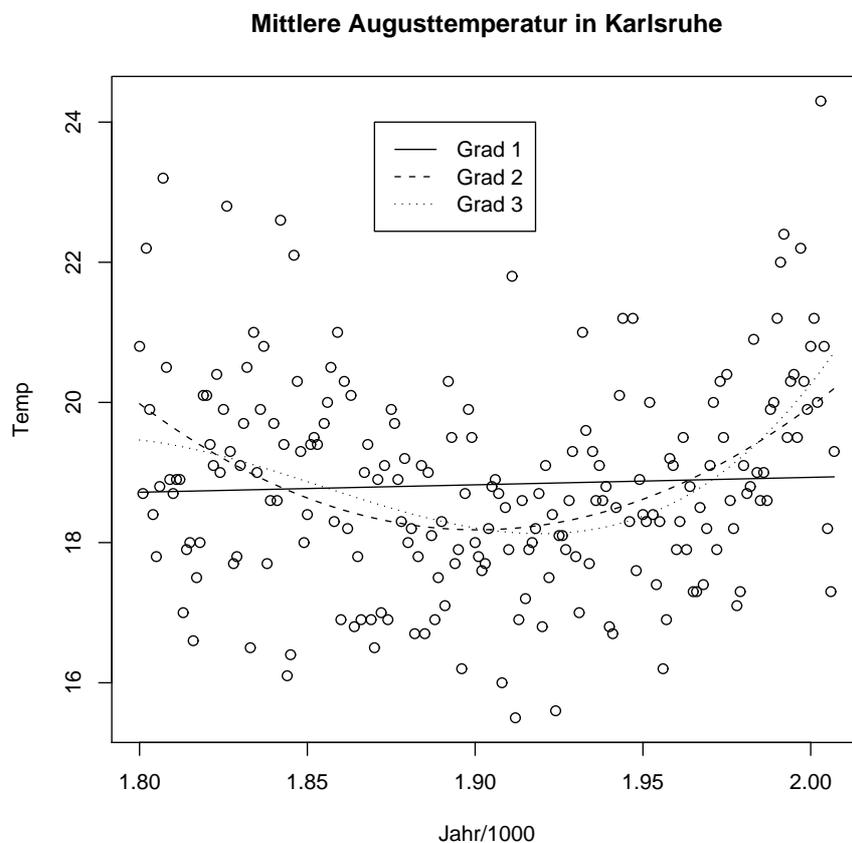
$$\mathbf{P}_{(\gamma,v)}[F_{H,L} > f_{s-r,n-s;1-\alpha}] = \alpha.$$

12.3.4. Beispiel: Polynomielle Regression

Seien t_1, \dots, t_n die (Werte der) Regressor-Variable, $X_i = \gamma_0 + \sum_{j=1}^d \gamma_j t_i^j + \sqrt{v}\xi_i$, $s = d + 1$,

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^d \end{pmatrix}$$

Auf der Website <http://www.klimadiagramme.de> finden sich u.A. die Durchschnittstemperaturen in Karlsruhe im August in den Jahren 1800 bis 2007 ($n = 206$ Beobachtungen, denn die Jahre 1854 und 1945 fehlen):



August-Durchschnittstemperaturen in Karlsruhe und Ausgleichspolynome

$$t_1(x) = 0.064x + 16.804, \quad t_2(x) = 177.9x^2 - 676.3x + 660.8,$$

$$t_3(x) = 1190x^3 - 6618x^2 + 12252x - 7532$$

(t_i ist jeweils das Polynom vom Grad $\leq i$, das die quadratische Abweichung minimiert).

Um beispielsweise zu testen, ob (wenn man normalverteilte Fehlerterme unterstellt) der Koeffizient γ_1 in $t_1(x) = \gamma_1 x + \gamma_0$ signifikant von 0 verschieden ist, können wir Korollar 12.3.3 verwenden: Es ist

$$A^t A = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} = \begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}, \quad (A^t A)^{-1} = \frac{1}{\text{Var}(t_i)} \begin{pmatrix} \sum t_i^2 & \sum t_i \\ \sum t_i & \frac{1}{n} \end{pmatrix}$$

Dann ist $c = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $c^t(A^t A)^{-1}c = \frac{1}{n \cdot \text{Var}(t_i)}$ und

$$\frac{0.064 - 0}{\sqrt{\frac{V^*}{(n \cdot \text{Var}(t_i))}}} \approx 0.614 < t_{206-2, 1-0.01/2} = 2.60,$$

d.h. die Nullhypothese $\gamma_1 = 0$ kann nicht zum Niveau $\alpha = 0.01$ verworfen werden.

Siehe Abschnitt A.2 für den zur Berechnung verwendeten R-Code.

12.4. Varianzanalyse:

Bemerkung: In der englischen Literatur ist das Akronym ANOVA=ANalysis Of VAriance üblich.

Seien X_1, \dots, X_n n Beobachtungen in s Gruppen (gem. s möglichen Werten eines „Faktors“). $n_1 + \dots + n_s = n$.

Modell: Sei $X_{i,k}$ die Beobachtung k in Gruppe i . Dann ist

$$X_{i,k} = m_i + \sqrt{v}\xi_{i,k} \text{ mit } \xi_{i,k} \text{ u.i.v. } \sim \mathcal{N}(0, 1)$$

(„Wörtlicher“ Bezug zum linearen Modell:

$$X = (X_{1,1}, X_{1,2}, \dots, X_{1,n_1}, X_{2,1}, \dots, X_{2,n_2}, \dots, X_{s,1}, \dots, X_{s,n_s}), \gamma = (m_1, \dots, m_s)^t \in \mathbb{R}^s,$$

Designmatrix: $A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & \vdots \\ 0 & 1 & \dots & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & 1 & & \vdots \\ \vdots & 0 & \dots & \vdots \\ \vdots & \vdots & & 0 \\ \vdots & \vdots & & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{pmatrix}$, wobei in der ersten Zeile n_1 viele Einsen stehen und in Zeile

2 n_2 viele Einsen stehen usw; L der Spaltenraum von A . Somit ist $X = A\gamma + \sqrt{v}\xi$ mit $\xi \sim \mathcal{N}(0, 1)^{\otimes n}$

Es gilt:

$$(A^t A) = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & n_s \end{pmatrix}$$

und

$$A^t X = (n_1 M_1, \dots, n_s M_s)^t$$

mit $M_i = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{i,k}$, $\hat{\gamma} = (A^t A)^{-1} A^t X = (M_1, \dots, M_s)^t$,

$$\begin{aligned} V^* &= \frac{1}{n-s} |X - A\hat{\gamma}|^2 \\ &= \frac{1}{n-s} |X - \underbrace{(M_1, \dots, M_1)}_{n_1} \underbrace{(M_2, \dots, M_2)}_{n_2} \dots \underbrace{(M_s, \dots, M_s)}_{n_s}|^2 \\ &= \frac{1}{n-s} \sum_{i=1}^s \sum_{k=1}^{n_i} (X_{i,k} - M_i)^2 \\ &= \sum_{i=1}^s \frac{n_i - 1}{n-s} V_i^* \end{aligned}$$

mit $V_i^* = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (X_{i,k} - M_i)^2$.

Schreibweise: $V_{iG}^* := V_i^*$, „Varianz innerhalb der Gruppen“

Bemerkung: „Streuungszerlegung“

Mit $M := \frac{1}{n} \sum_{i=1}^s \sum_{k=1}^{n_i} X_{i,k}$ ist

$$\begin{aligned} V_{\text{tot}}^* &:= \frac{1}{n-1} \sum_{i=1}^s \sum_{k=1}^{n_i} (X_{i,k} - M)^2 \\ &= \frac{1}{n-1} |X - M \cdot \mathbf{1}|^2 \\ &= \frac{1}{n-1} (|X - \Pi_L X|^2 + |\Pi_L X - M \cdot \mathbf{1}|^2), \end{aligned}$$

also

$$(n-1)V_{\text{tot}}^* = (n-s)V_{iG}^* + (s-1)V_{zG}^*$$

mit $V_{zG}^* := \frac{1}{s-1} \sum_{i=1}^s n_i (M_i - M)^2 = \frac{1}{s-1} |\Pi_L X - M \cdot \mathbf{1}|^2$ („Varianz zwischen den Gruppen“).

12.4.1. Satz: Konfidenzbereich für $\gamma = (m_1, \dots, m_s)^t$

Für $\alpha \in (0, 1)$ ist

$$C := \left\{ \tilde{m} = (\tilde{m}_1, \dots, \tilde{m}_s) : \frac{1}{s} \sum_{i=1}^s n_i (\tilde{m}_i - M_i)^2 < V_{iG}^* \cdot f_{s, n-s; 1-\alpha} \right\}$$

(wo $f_{s, n-s; 1-\alpha} = (1-\alpha)$ -Quantil der Fisher $_{s, n-s}$ -Verteilung) ein Konfidenzbereich (oder Konfidenzellipsoid) zum Irrtumsniveau α .

Beweis:

$$\begin{aligned}
 \mathbf{P}_{(\gamma,v)}[\gamma \in C] &= \mathbf{P}_{(\gamma,v)} \left[\frac{\frac{1}{s} \sum_{i=1}^s n_i (\gamma_i - M_i)^2}{V_{iG}^*} < f_{s,n-s;1-\alpha} \right] \\
 &= \mathbf{P}_{(0,v)} \left[\frac{\frac{1}{v} \frac{1}{s} \sum_{i=1}^s n_i (M_i)^2}{\frac{1}{v} V_{iG}^*} < f_{s,n-s;1-\alpha} \right] \\
 &= 1 - \alpha
 \end{aligned}$$

□

12.4.2. Satz: F-Test auf Gleichheit der Gruppenmittelwerte

$H = \{m \cdot \mathbf{1} : m \in \mathbb{R}\} \subset L(\mathbb{C} \mathbb{R}^n)$, (d.h. $A\gamma \in H \iff m_1 = m_2 = \dots = m_s$).

Der Ablehnungsbereich

$$\{V_{zG}^* > f_{s-1,n-s;1-\alpha} V_{iG}^*\}$$

definiert einen Test von $H_0 : m_1 = \dots = m_s$ gegen H_1 : “nicht alle m_i sind gleich“ zum Niveau α .

Beweis:

Sei $v > 0$, $\gamma \in \mathbb{R}^s$ mit $A\gamma \in H$. Betrachte

$$\frac{V_{zG}^*}{V_{iG}^*} = \frac{\frac{1}{s-1} |\Pi_L X - \Pi_H X|^2}{\frac{1}{n-s} |X - \Pi_L X|^2}, \text{ dies ist unter } \mathbf{P}_{\gamma,v} \text{ Fisher}_{s-1,n-s}\text{-verteilt.}$$

12.4.3. Korollar: t -Test auf Gleichheit der Mittelwerte für zwei ungepaarte (normalverteilte) Stichproben

Seien $X_{1,1}, \dots, X_{1,n_1}$ u.i.v. $\sim \mathcal{N}(m_1, v)$, $X_{2,1}, \dots, X_{2,n_2}$ u.i.v. $\sim \mathcal{N}(m_2, v)$ und

$$M_1 = \frac{1}{n_1} \sum_{k=1}^{n_1} X_{1,k}, M_2 = \frac{1}{n_2} \sum_{k=1}^{n_2} X_{2,k},$$

$$V_{iG}^* = \frac{1}{n-2} \left(\sum_{k=1}^{n_1} (X_{1,k} - M_1)^2 + \sum_{k=1}^{n_2} (X_{2,k} - M_2)^2 \right).$$

Der Ablehnungsbereich

$$\left\{ \frac{|M_1 - M_2|}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2}) V_{iG}^*}} > t_{n-2,1-\frac{\alpha}{2}} \right\}$$

(wo $t_{n-2,1-\frac{\alpha}{2}} = (1 - \frac{\alpha}{2})$ -Quantil der Student'schen t -Verteilung mit $n - 2$ Freiheitsgraden) definiert einen Test von $H_0 : m_1 = m_2$ gegen $H_1 : m_1 \neq m_2$ zum Niveau α .

Beweis:

Es ist $M = \frac{1}{n_1+n_2} \left(\sum_{i=1}^{n_1} X_{1,i} + \sum_{i=1}^{n_2} X_{2,i} \right) = \frac{n_1}{n_1+n_2} M_1 + \frac{n_2}{n_1+n_2} M_2$, also
 $V_{zG}^* = n_1(M_1 - M)^2 + n_2(M_2 - M)^2 = \frac{(M_1 - M_2)^2}{\frac{1}{n_1} + \frac{1}{n_2}}$. Nach obigem ist

$$\frac{V_{zG}^*}{V_{iG}^*} = \frac{(M_1 - M_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) V_{iG}^*}$$

Fisher $_{1,n-2}$ -verteilt, und die „symmetrisch signierte“ Wurzel einer Fisher $_{1,n-2}$ -verteilten Zufallsgröße ist t_{n-2} -verteilt.

12.5. Zweifaktorielle Varianzanalyse

Modell: Es gibt 2 Faktoren mit s_1 bzw. s_2 „Stufen“. $X_{i,j,k}$ ist die k -te Beobachtung zur Stufe i des ersten Faktors und Stufe j des 2. Faktors.

$$X_{i,j,k} = m_{i,j} + \sqrt{v} \xi_{i,j,k}, i \in \{1, \dots, s_1\}, j \in \{1, \dots, s_2\}, 1 \leq k \leq n_{i,j}$$

(insgesamt: $n = \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} n_{i,j}, s = s_1 \cdot s_2$).

$$m_{\cdot,\cdot} = \frac{1}{s} \sum_{i,j} m_{i,j}, m_{i,\cdot} = \frac{1}{s_2} \sum_{j=1}^{s_2} m_{i,j}, m_{\cdot,j} = \frac{1}{s_1} \sum_{i=1}^{s_1} m_{i,j}$$

Schreibe $m_{i,j} = m_{\cdot,\cdot} + \alpha_i + \beta_j + \gamma_{i,j}$ mit $\alpha_i = m_{i,\cdot} - m_{\cdot,\cdot}$ (i -ter Zeileneffekt) und $\beta_j = m_{\cdot,j} - m_{\cdot,\cdot}$ (j -ter Spalteneffekt), sowie $\gamma_{i,j} = m_{i,j} - m_{i,\cdot} - m_{\cdot,j} + m_{\cdot,\cdot}$ („Wechselwirkungseffekt“ bei (i, j)) mit Designmatrix A und ihrem Spaltenraum L .

12.5.1. Beispiel:

Sei $\tilde{H} = \{Am' : m' \in \mathbb{R}^s : m'_{i,j} = m'_{i,\cdot} + m'_{\cdot,j}\}$, und wir nehmen an, dass $n_{ij} = \ell$ für alle i, j .

(Übung: $(\Pi_{\tilde{H}} X)_{i,j,k} = (M_{i,j} - M_{i,\cdot} - M_{\cdot,j} + M_{\cdot,\cdot}), i \in \{1, \dots, s_1\}, j \in \{1, \dots, s_2\}, 1 \leq k \leq n_{i,j}$)

Betrachte nun

$$|\Pi_L X - \Pi_{\tilde{H}} X|^2 = \sum_{i,j} n_{i,j} (M_{\cdot,\cdot} - M_{i,\cdot} - M_{\cdot,j})^2 =: (s_1 \cdot s_2 - (s_1 + s_2 - 1)) V_{zG^{1-2}}^* = (s_1 - 1) \cdot (s_2 - 1) V_{zG^{1-2}}^*,$$

der Ablehnungsbereich

$$\left\{ \frac{V_{zG^{1-2}}^*}{V_{iG}^*} > f_{(s_1-1)(s_2-1), n-s_1 \cdot s_2; 1-\alpha} \right\}$$

definiert einen Test für $H_0 : \gamma_{i,j} \equiv 0$ (keine Wechselwirkungseffekte) gegen $H_1 : \gamma_{i,j} \neq 0$.

Teil II.

**Finanzmathematik - Mathematische
Modelle und Methoden zur Bewertung
von Derivaten**

13. Einführung in die Finanzmathematik

Beispiele:

Eine Aktie habe heute den Wert $S_0 = 400$ Euro, zu einem zukünftigen Zeitpunkt T gelte

$$S_T = \begin{cases} 500 \text{ Euro} & \text{mit Wahrscheinlichkeit } p = \frac{1}{2}, \\ 350 \text{ Euro} & \text{mit Wahrscheinlichkeit } 1 - p = \frac{1}{2}. \end{cases}$$

13.1. Europäische Call- und Put-Option

Eine europäische Call-Option auf die Aktie (mit Fälligkeit („maturity“) T) und Ausübungspreis („strike“) K : Das Recht (aber nicht die Pflicht), zum Zeitpunkt T eine Aktie zum Preis K zu kaufen.

Wert des Calls zum Zeitpunkt T :

$$C_T = (S_T - K)_+$$

Analog: Europäische Put-Option, Wert zur Zeit T :

$$P_T = (K - S_T)_+$$

Was ist der „faire“ Preis der Call-Option?

Beispiel: $K = 440$ Euro.

„Naiver Ansatz“: $\mathbf{E}[C_T] = \frac{1}{2}60\text{Euro} + \frac{1}{2}0\text{Euro} = 30\text{Euro}$

Beobachtung: („Hedging strategy“)

Mit einem Startkapital von 20 Euro kann man das Auszahlungsprofil der Call-Option replizieren.

Zeit 0: Leihe 140 Euro, kaufe 0.4 Aktien (zahle dafür $0.4 \cdot 400\text{Euro} = 20\text{Euro} + 140\text{Euro}$)

Zeit T : Wenn $S_T = 500\text{Euro}$: Wert des Portfolios: $0.4 \cdot 500\text{Euro} - 140\text{Euro} = 60\text{Euro}$

Wenn $S_T = 350\text{Euro}$: Wert des Portfolios: $0.4 \cdot 350\text{Euro} - 140\text{Euro} = 0$

Demnach ist der faire Preis $\pi(C) = 20\text{Euro}$, sonst gäbe es eine Arbitrage-Möglichkeit in diesem Markt.

Beobachtung:

$\pi(C) = \mathbf{E}_{p^*}[C_T]$ für $p^* = \frac{1}{3}$.

Betrachte einen allgemeinen „contingent claim“ H mit Fälligkeit T ,

$$H_T = \begin{cases} H^b & \text{wenn } S_T = 500\text{Euro} \\ H^a & \text{wenn } S_T = 350\text{Euro} \end{cases}$$

Beobachtung:

$\pi(H) = \mathbf{E}_{p^*}[H_T] = p^*H^b + (1 - p^*)H^a$, denn betrachte folgende Handelsstrategie:

Zur Zeit 0 kaufe

$$\Theta^1 := \frac{H^b - H^a}{150\text{Euro}} \text{ Aktien}$$

und verleihe

$$\Theta^0 = H^b - \frac{H^b - H^a}{150\text{Euro}} \cdot 500\text{Euro}$$

Wert des Portfolios zur Zeit T :

$$\Theta^1 S_T + \Theta^0 = \begin{cases} \frac{H^b - H^a}{150} 500 + H^b - \frac{H^b - H^a}{150} 500 = H^b & , \text{ wenn } S_T = 500\text{Euro} \\ \frac{H^b - H^a}{150} 350 + H^b - \frac{H^b - H^a}{150} 500 = H^a & , \text{ wenn } S_T = 350\text{Euro} \end{cases}$$

Wert des Portfolios zur Zeit 0:

$$\Theta^1 \cdot 400\text{Euro} + \Theta^0 = \frac{H^b - H^a}{150} \cdot 400 + H^b - \frac{H^b - H^a}{150} 500 = \frac{1}{3}H^b + \frac{2}{3}H^a = \mathbf{E}_{p^*}[H_T] = \pi(H)$$

Beobachtung:

$$\mathbf{E}_{p^*}\left[\frac{S_T}{1}\right] = \frac{1}{3}500\text{Euro} + \frac{2}{3}350\text{Euro} = 400\text{Euro} = \frac{S_0}{1}$$

Das Wahrscheinlichkeitsmaß P^* heißt auch „risikoneutrales Maß“ oder „Äquivalentes Martingalmaß“.

13.2. Binomialmodell:

(i) Eine Periode:

Handelszeitpunkte $\mathbb{T} = \{0, 1\}$.

Wertpapiere:

- Bankkonto mit Zinssatz $r \geq 0$:

$$S_0^0 = 1, S_1^0 = (1 + r)$$

- risikobehaftetes Wertpapier:

$$S_0^1 \text{ fest, } S_1^1 \text{ ZV mit } \mathbf{P}[S_1^1 = \overbrace{S_0^1(1+b)}^{S^b}] = p = 1 - \mathbf{P}[S_1^1 = \overbrace{S_0^1(1+a)}^{S^a}]$$

Wir nehmen an, dass $a < r < b$, denn wenn $r < \min\{a, b\}$ oder $r > \max\{a, b\}$, so gibt es Arbitrage im Modell.

Portfolio:

$$\Theta_t^0 := \text{Betrag auf Bankkonto im Intervall } (t-1, t]$$

$$\Theta_t^1 := \text{Einheiten der risikobehafteten Anlage im Intervall } (t-1, t]$$

Wertprozess: $V_0 = \Theta_1^0 \cdot S_0^0 + \Theta_1^1 S_0^1$, $V_t = \Theta_t^0 S_t^0 + \Theta_t^1 S_t^1$, ($t \geq 1$)

Betrachte einen contingent claim H mit $H_1 = \begin{cases} H^a & \text{wenn } S_1^1 = S_0^1(1+a), \\ H^b & \text{wenn } S_1^1 = S_0^1(1+b) \end{cases}$.

Gesucht: Hedging-Strategie und fairer Preis $\pi(H) := V_0$ (Wert zur Zeit $t = 0$ des replizierenden Portfolios).

$$H^b = (1+r)V_0 + \Theta_1^1((1+b)S_0^1 - (1+r)S_0^1)$$

$$H^a = (1+r)V_0 + \Theta_1^1((1+a)S_0^1 - (1+r)S_0^1)$$

Dann

$$\Theta_1^1 = \frac{H^b - H^a}{(b-a)S_0^1} = \frac{H^b - H^a}{(b-a)S_0^1} \text{ „Delta“}$$

$$H^b = (1+r)\Theta_1^0 + \frac{H^b - H^a}{(b-a)S_0^1}(1+b)S_0^1 \Rightarrow \Theta_1^0 = \frac{1}{1+r} \frac{(1+b)H^a - (1+a)H^b}{b-a}$$

$$V_0 = \frac{1}{1+r} \left(H^a \underbrace{\frac{b-r}{b-a}}_{1-p^*} + H^b \cdot \underbrace{\frac{r-a}{b-a}}_{=:p^*} \right) = \frac{1}{1+r} \mathbf{E}_{p^*}[H_1] = \mathbf{E}_{p^*}\left[\frac{H_1}{S_0^1}\right]$$

13.2.1. Mehrperioden-Binomialmodell (CRR(=Cox-Ross-Rubinstein)-Modell (1979))

Sei $\mathbb{T} = \{0, 1, \dots, T\}$ und es gebe zwei Wertpapiere:

(i) $S_t^0 = (1+r)^t$

(ii) $S_t^1 = S_0^1 \cdot X_1 \cdots X_t$, wo X_1, \dots, X_T unabhängig mit $\mathbf{P}[X_i = 1+b] = p = 1 - \mathbf{P}[X_i = 1+a]$ mit $-1 < a < r < b$ und $0 < p < 1$

Sei $\mathcal{F}_t = \sigma(S_u^i, 1 \leq u \leq t, i \in \{0, 1\}) (= \sigma(X_1, \dots, X_t))$. Es gilt $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, d.h. die Familie $(\mathcal{F}_t)_{t=0, \dots, T}$ ist eine *Filtration*.

Gegeben: Eine europäische Option H mit Fälligkeit T und Basiswertpapier S^1 , d.h. $H_T = f(S_T^1)$.

Gesucht: Fairer Preis, zugehörige selbstfinanzierende Replikationsstrategie.

Unter dem Maß \mathbf{Q} sei X_1, \dots, X_T unabhängig verteilt mit

$$\mathbf{Q}[X_i = 1+b] = p^* = 1 - \mathbf{Q}[X_i = 1+a], \quad p^* := \frac{r-a}{b-a}.$$

\mathbf{Q} heißt „äquivalentes Martingalmaß“ oder „risikoneutrales Maß“.

13.2.2. Beobachtung:

$$\mathbf{E}_{\mathbf{Q}} \left[\frac{S_{t+1}^1}{S_{t+1}^0} \middle| \mathcal{F}_t \right] = \mathbf{E}_{\mathbf{Q}} \left[\frac{S_0^1 X_1 \cdots X_{t+1}}{(1+r)S_t^0} \middle| \mathcal{F}_t \right] = \frac{S_0^1 X_1 \cdots X_t}{S_t^0} \overbrace{\mathbf{E}_{\mathbf{Q}} \left[\frac{X_{t+1}}{(1+r)} \middle| \mathcal{F}_t \right]}^{=1} = \frac{S_t^1}{S_t^0},$$

$$\frac{1}{1+r} \cdot \left(\frac{r-a}{b-a}(1+b) + \frac{b-r}{b-a}(1+a) \right) = \frac{1+r-b-a-ab+ab+ab-f-ra}{(1+r)(b-a)} = 1$$

d.h. $\left(\frac{S_t^1}{S_t^0}\right)_{t=0,1,\dots,T}$ ist $(\mathbf{Q}, (\mathcal{F}_t))$ -Martingal.

Sei H_t der Wert der Option zur Zeit $t \leq T$ (Schreibe $H_t^{x_1, \dots, x_t}$ für den Wert auf $\{X_1 = x_1, \dots, X_T = x_t\}$).

Notiere Portfolio folgendermaßen:

$\Theta_t^1 \dots$ Anzahl S^1 , die im Intervall $(t-1, t]$ gehalten werden,

$\Theta_t^0 \dots$ Anzahl S^0 , die im Intervall $(t-1, t]$ gehalten werden.

Dann ist

$$V_t = \Theta_t^1 S_t^1 + \Theta_t^0 S_t^0 \quad (\text{notiere } V_0 = \Theta_1^1 S_0^1 + \Theta_1^0 S_0^0)$$

der Wert des Portfolios zur Zeit t .

Es soll die sogenannte „Selbstfinanzierungsbedingung“ gelten:

$$\Theta_{t+1}^1 S_t^1 + \Theta_{t+1}^0 S_t^0 = V_t = \Theta_t^1 S_t^1 + \Theta_t^0 S_t^0, \quad t = 1, \dots, T-1,$$

d.h. während der Laufzeit der Option wird dem Portfolio weder Kapital entnommen noch zugeführt.

Die Strategie soll *prävisibel* sein, d.h. Θ_t^0, Θ_t^1 sind \mathcal{F}_{t-1} -messbar — die Investitionsentscheidungen über das Zeitintervall $(t-1, t]$ müssen zum Zeitpunkt $t-1$ getroffen werden, ohne „Blick in die Zukunft“.

13.2.3. Satz:

Im CRR-Modell gilt

$$\frac{H_t}{S_t^0} = \mathbf{E}_{\mathbf{Q}} \left[\frac{H_T}{S_T^0} \middle| \mathcal{F}_t \right], \quad (t = 0, \dots, T),$$

die zugehörige (selbstfinanzierende) duplizierende Handelsstrategie hat folgende Form (auf dem Ereignis $\{X_1 = x_1, \dots, X_{t-1} = x_{t-1}\}$):

$$\Theta_t^1 = \frac{H_t^{x_1, \dots, x_{t-1}, 1+b} - H_t^{x_1, \dots, x_{t-1}, 1+a}}{(b-a)S_{t-1}^1}, \quad \Theta_t^0 = \frac{1}{1+r} \frac{(1+b)H_t^{x_1, \dots, x_{t-1}, 1+b} - (1+a)H_t^{x_1, \dots, x_{t-1}, 1+a}}{(b-a)S_{t-1}^0}.$$

Beobachtung:

$$\mathbf{E}_{\mathbf{Q}} \left[\frac{H_{t+1}}{S_{t+1}^0} \middle| \mathcal{F}_t \right] = \mathbf{E}_{\mathbf{Q}} \left[\mathbf{E}_{\mathbf{Q}} \left[\frac{H_T}{S_T^0} \middle| \mathcal{F}_t \right] \middle| \mathcal{F}_t \right] = \mathbf{E}_{\mathbf{Q}} \left[\frac{H_T}{S_T^0} \middle| \mathcal{F}_t \right] = \frac{H_t}{S_t^0},$$

d.h. $(H_t/S_t^0)_{t=0,1,\dots,T}$ ist ein \mathbf{Q} -Martingal.

Beweis des Satzes

Beweis durch Rückwärtsinduktion:

Zum Zeitpunkt $T-1$, auf dem Ereignis $\{X_1 = x_1, \dots, X_{T-1} = x_{T-1}\}$ handelt es sich um ein „1-Perioden-(Binomial-)Modell“ (schlage Formeln vom letzten Modell nach).

Angenommen die Behauptung gilt für $t, t+1, \dots, T-1$: Fasse $\tilde{H}_t = H_t$ als Claim in einem 1-Perioden-Modell (von $t-1$ nach t) auf.

Auf $\{X_1 = x_1, \dots, X_{t-1} = x_{t-1}\}$: Kaufe

$$\Theta_t^1 := \frac{H_t^{X_1, \dots, X_{t-1}, 1+b} - H_t^{X_1, \dots, X_{t-1}, 1+a}}{(b-a)S_{t-1}^1}$$

viele Einheiten S^1 und

$$\Theta_t^0 := \frac{1}{r+1} \frac{(1+b)H_t^{X_1, \dots, X_{t-1}, 1+a} - (1+a)H_t^{X_1, \dots, X_{t-1}, 1+b}}{(b-a)S_{t-1}^0}$$

viele Einheiten von S^0 .

Wert dieses Portfolios zur Zeit t :

- Falls $S_t^1 = S_{t-1}^1 \cdot (1+b)$:

$$\begin{aligned} & V_t \\ &= S_{t-1}^1(1+b) \frac{H_t^{X_1, \dots, X_{t-1}, 1+b} - H_t^{X_1, \dots, X_{t-1}, 1+a}}{(b-a)S_{t-1}^1} \\ & \quad + S_{t-1}^0(1+r) \frac{(1+b)H_t^{X_1, \dots, X_{t-1}, 1+a} - (1+a)H_t^{X_1, \dots, X_{t-1}, 1+b}}{(b-a)S_{t-1}^0} \\ &= \frac{(1+b)H_t^{X_1, \dots, X_{t-1}, 1+b} - (1+b)H_t^{X_1, \dots, X_{t-1}, 1+a} + (1+b)H_t^{X_1, \dots, X_{t-1}, 1+a} - (1+a)H_t^{X_1, \dots, X_{t-1}, 1+b}}{b-a} \\ &= H_t^{X_1, \dots, X_{t-1}, 1+b}. \end{aligned}$$

- Ebenso falls $S_t^1 = S_{t-1}^1 \cdot (1+a)$: $V_t = H_t^{X_1, \dots, X_{t-1}, 1+a}$.

Aus den Ergebnissen des Ein-Perioden-Modells folgt darüberhinaus, dass

$$H_{t-1}^{X_1, \dots, X_{t-1}} = \frac{1}{1+r} \mathbf{E}_{\mathbf{Q}}[H_t | \mathcal{F}_{t-1}].$$

Zur Selbstfinanzierungs-Bedingung: Wir müssen zeigen, dass

$$\Theta_{t+1}^1 S_t^1 + \Theta_{t+1}^0 S_t^0 \stackrel{!}{=} V_t = \Theta_t^1 S_t^1 + \Theta_t^0 S_t^0.$$

Rechte Seite $= H_t$ (denn das Portfolio repliziert die Option H).

Die linke Seite ist

$$\begin{aligned} & \frac{1+r}{1+r} \frac{H_t^{X_1, \dots, X_{t-1}, X_t, 1+b} - H_t^{X_1, \dots, X_{t-1}, X_t, 1+a}}{(b-a)S_t^1} S_t^1 \\ & + \frac{1}{1+r} \frac{(1+b)H_t^{X_1, \dots, X_{t-1}, X_t, 1+a} - (1+a)H_t^{X_1, \dots, X_{t-1}, X_t, 1+b}}{(b-a)S_t^0} S_t^0 \\ &= \frac{1}{1+r} \frac{(r-a)H_t^{X_1, \dots, X_{t-1}, X_t, 1+b} + (b-r)H_t^{X_1, \dots, X_{t-1}, X_t, 1+a}}{b-a} \\ &= \frac{1}{1+r} \mathbf{E}_{\mathbf{Q}}[H_{t+1} | \mathcal{F}_t], \end{aligned}$$

was nach Induktionsvoraussetzung $= H_t$ ist.

□

13.2.4. Beispiel: Europäische Call-Option

$C_T = (S_T^1 - K)_+$, $S_T^1 = S_0^1(1+b)^N(1+a)^{T-N}$, wo $N = |\{i \leq T : X_i = 1+b\}|$, unter \mathbf{Q} ist N $Bin(T, p^*)$ -verteilt.

Sei $A = \min\{n \in \mathbb{Z} : S_0^1(1+b)^n(1+a)^{T-n} > K\}$.

$$\begin{aligned} & \frac{1}{(1+r)^T} \mathbf{E}_{\mathbf{Q}} [(S_0^1(1+b)^N(1+a)^{T-N} - K)_+] \\ &= \frac{1}{(1+r)^T} \sum_{n=A}^T \binom{T}{n} (p^*)^n (1-p^*)^{T-n} (S_0^1(1+b)^n(1+a)^{T-n} - K) \\ &= S_0^1 Bin(T, p')(\{A, A+1, \dots, T\}) - \frac{1}{(1+r)^T} K \cdot Bin(T, p^*)(\{A, \dots, T\}) \end{aligned}$$

mit $p' = \frac{p^*(1+b)}{1+r}$

13.2.5. Bemerkung: Call-Put-Parität

$C_T = (S_T^1 - K)_+$, $P_T = (K - S_T^1)_+$, also $C_T - P_T = S_T^1 - K$, d.h. fairer Preis wäre:

$$\pi(C_T - P_T) = \pi(C_T) - \pi(P_T) = \pi(S_T^1 - K) = S_0^1 - (1+r)^{-T} K.$$

(Der Vertrag mit Auszahlungsprofil $S_T^1 - K$ heißt „Forward“.) Demnach bestimmt der Preis des europäischen Calls den des Puts mit demselben Ausübungspreis und umgekehrt.

13.3. Black-Scholes-Formel

(Fisher Black, Merton Scholes 1973, Robert Merton 1973; Merton und Scholes haben für diese Arbeiten 1997 den Nobelpreis für Wirtschaftswissenschaften erhalten.)

Sei T der Zeithorizont, $N \cdot T$ die Anzahl der Handelsperioden (i -te Periode entspricht „Realzeitintervall“ $(\frac{i-1}{N}, \frac{i}{N})$), r die instantane Zinsrate ≥ 0

$r_N := \frac{r}{N}$, $a_N := -\frac{\sigma}{\sqrt{N}}$, $b_N := \frac{\sigma}{\sqrt{N}}$, $p_N := \frac{1}{2} + \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}}$, wo $\sigma > 0$ die „Volatilität“, $\mu \in \mathbb{R}$ der „Renditeparameter“, S_0^1 sei fest gegebener Wert.

Beobachtung:

$$S_{NT, N}^0 = (1 + r_N)^{NT} = \left(1 + \frac{r}{N}\right)^{NT} \xrightarrow{N \rightarrow \infty} e^{rT}$$

Betrachte

$$\begin{aligned}
 S_{tN,N}^1 &= S_0^1 \cdot \prod_{i=1}^{tN} X_i^{(N)} \quad \text{für } tN \in \mathbb{N}_0 \\
 &= S_0^1 \cdot \prod_{i=1}^{tN} (1 + R_i^{(N)}) \quad \text{mit } R_i^{(N)} = X_i^{(N)} - 1 = \frac{S_{i,N}^1 - S_{i-1,N}^1}{S_{i-1,N}^1} \\
 &\quad \text{„Rendite der Aktie über das } i\text{-te Handelsintervall } \left(\frac{(i-1)}{N}, \frac{i}{N}\right)\text{“} \\
 &= S_0^1 \exp\left(\sum_{i=1}^{tN} \log(1 + R_i^{(N)})\right)
 \end{aligned}$$

Beobachtung:

$$R_i^{(N)} = \begin{cases} \frac{\sigma}{\sqrt{N}} & \text{mit Wahrscheinlichkeit } \frac{1}{2} + \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}} \\ -\frac{\sigma}{\sqrt{N}} & \text{mit Wahrscheinlichkeit } \frac{1}{2} - \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}} \end{cases}$$

und $\log(1+r) = r - \frac{1}{2}r^2 + O(r^3)$ für r aus einer Umgebung der 0, also

$$\log\left(\frac{S_{tN,N}^1}{S_0^1}\right) = \sum_{i=1}^{tN} \left(R_i^{(N)} - \frac{1}{2}(R_i^{(N)})^2\right) + \text{Rest}_{N,t}$$

mit

$$|\text{Rest}_{N,t}| \leq \frac{C}{\sqrt{N}} \quad \text{für eine Konstante } C.$$

Es gilt

$$\mathbf{E}[R_i^{(N)}] = \frac{1}{2} \left(1 + \frac{\mu}{\sigma\sqrt{N}}\right) \frac{\sigma}{\sqrt{N}} + \frac{1}{2} \left(1 - \frac{\mu}{\sigma\sqrt{N}}\right) \frac{-\sigma}{\sqrt{N}} = \frac{\mu}{N}$$

und

$$\mathbf{E}[(R_i^{(N)})^2] = \frac{\sigma^2}{N}, \quad \text{Var}(R_i^{(N)}) = \frac{\sigma^2}{N} - \frac{\mu^2}{N^2},$$

also

$$\begin{aligned}
 \mathbf{E}\left[R_i^{(N)} - \frac{1}{2}(R_i^{(N)})^2\right] &= \frac{\mu - \frac{1}{2}\sigma^2}{N}, \\
 \text{Var}\left(R_i^{(N)} - \frac{1}{2}(R_i^{(N)})^2\right) &= \text{Var}(R_i^{(N)}) = \frac{\sigma^2}{N} - \frac{\mu^2}{N^2}.
 \end{aligned}$$

Demnach konvergiert $S_{[Nt],N}^1$ in Verteilung gegen

$$S_0^1 \exp\left(\sigma B_t + t\left(\mu - \frac{1}{2}\sigma^2\right)\right),$$

wo $B_t \sim \mathcal{N}(0, t)$

Es gilt

$$p_N^* = \frac{r_N - a_N}{b_N - a_N} = \frac{\frac{r}{N} + \frac{\sigma}{\sqrt{N}}}{\frac{2\sigma}{\sqrt{N}}} = \frac{1}{2} + \frac{1}{2} \frac{r}{\sigma\sqrt{N}}.$$

Demnach konvergiert unter \mathbf{P}_N^* mit $N \rightarrow \infty$

$$S_{NT,N}^1 \xrightarrow{d} S_0^1 \exp \left(\sigma B_T + \left(r - \frac{\sigma^2}{2} \right) T \right),$$

wobei $B_T \sim \mathcal{N}(0, T)$

13.3.1. Satz:

$f: \mathbb{R} \rightarrow \mathbb{R}$ stetiges, beschränktes Auszahlungsprofil einer Option $H_{T,N} = f(S_{NT,N}^1)$.
Die Folge $\pi_N(H_{T,N})$ der fairen Preise konvergiert gegen

$$e^{-rT} \cdot \int_{-\infty}^{\infty} f \left(S_0^1 e^{\sigma\sqrt{T}y + rT - \frac{1}{2}\sigma^2 T} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

13.3.2. Korollar: Black-Scholes-Formel für den Preis eines Calls

$$f(s) = (s - K)_+,$$

$$\begin{aligned} v(x, T) &= e^{-rT} \int_{-\infty}^{\infty} (x e^{\sigma\sqrt{T}y + rT - \frac{1}{2}\sigma^2 T} - K)_+ \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= x\Phi(d_+(x, T)) - e^{-rT} K\Phi(d_-(x, T)) \end{aligned}$$

mit $d_-(x, T) = \frac{\log \frac{x}{K} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$, $d_+(x, T) = d_-(x, T) + \sigma\sqrt{T} = \frac{\log \frac{x}{K} + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$ und Φ die Verteilungsfunktion von $\mathcal{N}(0, 1)$.

Bemerkung: Das Auszahlungsprofil des Calls ist nicht beschränkt. Um Satz 13.3.1 wörtlich anzuwenden, betrachte zunächst den entsprechenden Put, dessen Auszahlungsprofil beschränkt ist, und benutze dann die Call-Put-Parität.

13.4. Fundamentalsatz der Preistheorie (in endlichen Märkten)

13.4.1. Marktmodell

Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum mit $|\Omega| < \infty$, $\mathbb{T} = \{0, 1, \dots, T\}$ eine endliche Menge von Handelszeiten und $(\mathcal{F}_t)_{t \in \mathbb{T}}$ die zugehörige Filtration, d.h. $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ und $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_T = \mathcal{F}$. Weiter seien S^0, S^1, \dots, S^d die $d+1$ Wertpapiere mit $S_t^i =$ Wert des i -ten Wertpapiers zur Zeit t , welches positive \mathcal{F}_t -messbare Zufallsvariablen sind, d.h. $(S_t^i)_{t \in \mathbb{T}}$ ist $(\mathcal{F}_t)_{t \in \mathbb{T}}$ -adaptiert.

S_t^0 sei > 0 und deterministisch (eine risikolose Anlage).

Betrachte

$$\bar{S}_t^i := \frac{S_t^i}{S_t^0} \quad (\text{„diskontierte“ Preise}).$$

Wir benutzen S^0 als „Numéraire.“

Handelsstrategien: Ein stochastischer Prozess

$$(\Theta_t)_{t=1,\dots,T} = ((\Theta_t^i)_{t=1,\dots,T}, i = 0, \dots, d)$$

ist eine Handelsstrategie, wenn Θ_t^i \mathcal{F}_{t-1} -messbar ist (für $i = 0, \dots, d; t = 1, \dots, T$). Man nennt $((\Theta_t^i)_{t=1,\dots,T}, i = 0, \dots, d)$ dann *prävisibel*.

Interpretation: Halte Θ_t^i Einheiten von Wertpapier i über das Intervall $(t-1, t]$.

Wertprozess: $V_t(\Theta) = \langle \Theta_t, S_t \rangle = \sum_{i=0}^d \Theta_t^i S_t^i, t = 1, \dots, T$ mit $V_0(\Theta) = \Theta_1 \cdot S_0$

Zugewinnprozess: $G_0(\Theta) = 0$ und $G_t(\Theta) = \sum_{n=1}^t \langle \Theta_n, \Delta S_n \rangle$ mit $\Delta S_n = S_n - S_{n-1}$

13.4.2. Definition: selbstfinanzierend

Eine Handelstrategie heißt *selbstfinanzierend*, wenn

$$V_t(\Theta) = V_0(\Theta) + G_t(\Theta) \quad \forall t \in \mathbb{T}$$

gilt. Sei W die Menge der selbstfinanzierenden Handelsstrategien.

Bemerkung:

(i) W ist ein Vektorraum.

(ii) $\Theta \in W \iff \langle \Delta \Theta_t, S_{t-1} \rangle = 0 \quad \forall t \in \{1, \dots, T\}$
 $\iff \langle \Delta \Theta_t, \bar{S}_{t-1} \rangle = 0 \quad \forall t \in \{1, \dots, T\}$

(iii)

$$\bar{V}_t(\Theta) = \frac{V_t(\Theta)}{S_t^0} = \Theta_t^0 + \sum_{i=1}^d \Theta_t^i \bar{S}_t^i = \bar{V}_0(\Theta) + \sum_{s=1}^t \sum_{i=1}^d \Theta_s^i \Delta \bar{S}_s^i,$$

also ist

$$\Theta_t^0 = \bar{V}_0(\Theta) + \sum_{s=1}^t \sum_{i=1}^d \Theta_s^i \Delta \bar{S}_s^i - \sum_{i=1}^d \Theta_t^i \bar{S}_t^i,$$

d.h. für eine selbstfinanzierende Handelsstrategie genügt es, $V_0(\Theta)$ und $\Theta^1, \dots, \Theta^d$ festzulegen.

Beweis der Bemerkung:

Zu (ii):

$$\begin{aligned} \langle \Delta \Theta_t, S_{t-1} \rangle &= \sum_{i=0}^d (\Theta_t^i - \Theta_{t-1}^i) S_{t-1}^i \\ &= \sum_{i=0}^d \Theta_t^i (S_t^i - S_{t-1}^i) - \sum_{i=0}^d \Theta_{t-1}^i S_{t-1}^i \\ &= V_t(\Theta) - \Delta G_t(\Theta) - V_{t-1}(\Theta) \end{aligned}$$

13.4.3. Definition: (duplizierbarer) Claim, Hedge, vollständiges Marktmodell, Arbitrage

Eine Zufallsvariable H auf $(\Omega, \mathcal{F}, \mathbf{P})$, aufgefasst als Auszahlung zum Zeitpunkt T heißt ein *Claim*. H heißt *duplizierbar* (replizierbar, “attainable“), wenn es $\Theta \in W$ gibt mit $V_T(\Theta) = H$.

Ein solches Θ heißt ein *Hedge* für H .

Ein Marktmodell heißt *vollständig*, wenn jedes H duplizierbar ist.

Eine *Arbitragemöglichkeit* ist eine selbstfinanzierende Handelsstrategie mit $V_0(\Theta) = 0, V_T(\Theta) \geq 0$ und $\mathbf{P}[V_T(\Theta) > 0] > 0$. Ein Markt(-modell) heißt *arbitragefrei*, wenn es keine Arbitragemöglichkeiten gibt.

Ist H ein Claim, $(H_t)_{t \in \mathbb{T}}$ ein adaptierter stochastischer Prozess mit $H_T = H$, so dass der um $S_t^{d+1} := (H_t)$ erweiterte Markt arbitragefrei ist, so heißt (H_t) (ein) *Arbitragepreisprozess* für H , $\pi((H)_t) := H_0$ ein *Arbitragepreis*.

Beobachtung:

Sei der Markt arbitragefrei und $\Theta \in W$ mit $V_T(\Theta) = 0$. Dann gilt $V_t(\Theta) = 0 \forall t \in \mathbb{T}$.

Beweis der Beobachtung:

Angenommen es gäbe ein $t \in \mathbb{T}$ und ein $c < 0$, so dass $\mathbf{P}[V_t(\Theta) = c] > 0$, dann beobachte den Markt bis t und steige in den Markt ein, sofern $V_t(\Theta) = c$, verfolge dann Θ bis T . Dies ergibt dann den sicheren Gewinn $(-c) \cdot \frac{S_T^0}{S_t^0}$, im Widerspruch zur angenommenen Arbitragefreiheit.

Analog, wenn $c > 0$: Verfolge $\Theta' = -\Theta$.

13.4.4. Korollar:

In einem arbitragefreien Markt ist der Arbitrageprozess eines replizierbaren Claims H eindeutig festgelegt (als Wertprozess des Hedge).

13.4.5. Definition: Äquivalentes Martingalmaß

Ein Wahrscheinlichkeitsmaß \mathbf{Q} auf (Ω, \mathcal{F}) heißt ein *äquivalentes Martingalmaß*, wenn $\mathbf{P} \sim \mathbf{Q}$, d.h. $\mathbf{P}[\{\omega\}] > 0 \iff \mathbf{Q}[\{\omega\}] > 0 \forall \omega \in \Omega$ in diesem Kontext, und jedes $(\bar{S}_t^i)_{t \in \mathbb{T}}$ ein \mathbf{Q} -Martingal ist, $i = 0, 1, \dots, d$.

$\mathcal{P}^* := \{\mathbf{Q} : \mathbf{Q} \text{ ist äquivalentes Martingalmaß}\}$

13.4.6. Beobachtung:

Sei $\mathbf{Q} \in \mathcal{P}^*, \Theta \in W$, dann ist $(\bar{V}_t(\Theta))_{t \in \mathbb{T}}$ ein \mathbf{Q} -Martingal, denn

$$\mathbf{E}_{\mathbf{Q}}[\bar{V}_t(\Theta) - \bar{V}_{t-1}(\Theta) | \mathcal{F}_{t-1}] = \mathbf{E}_{\mathbf{Q}} \left[\sum_{i=0}^d \Theta_t^i (\bar{S}_t^i - \bar{S}_{t-1}^i) \middle| \mathcal{F}_{t-1} \right] = \sum_{i=0}^d \Theta_t^i \underbrace{\mathbf{E}_{\mathbf{Q}} [\bar{S}_t^i - \bar{S}_{t-1}^i | \mathcal{F}_{t-1}]}_{=0} = 0$$

13.4.7. Fundamentalsatz der Preistheorie:

Das Marktmodell ist arbitragefrei $\iff |\mathcal{P}^*| \geq 1$.

Das Marktmodell ist arbitragefrei und vollständig $\Rightarrow |\mathcal{P}^*| = 1$.

Beweis:

Sei $\mathbf{Q} \in \mathcal{P}^*$, Θ selbstfinanzierende Handelsstrategie mit $V_T(\Theta) \geq 0$ und $\mathbf{P}[V_T(\Theta) > 0] > 0$.

Dann gilt $\mathbf{Q}[V_T(\Theta) > 0] > 0$ und $\frac{V_0(\Theta)}{S_0^0} = \mathbf{E}_{\mathbf{Q}}\left[\frac{V_T(\Theta)}{S_T^0}\right] > 0$, d.h. es gibt keine Arbitragemöglichkeit.

Sei der Markt arbitragefrei:

Betrachte $U := \{\bar{V}_T(\Theta) : \Theta \in W, V_0(\Theta) = 0\} \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) (\cong \mathbb{R}^{|\Omega|})$,

$C := \{X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) : X \geq 0, \mathbf{E}[X] = 1\}$.

Beobachtung: U ist ein Untervektorraum, C ist konvex und kompakt als Teilmenge von $\mathbb{R}^{|\Omega|}$. Aus Arbitragefreiheit folgt: $U \cap C = \emptyset$.

Nach dem Trennungssatz² gibt es $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ mit $\langle X, G \rangle = 0 \ \forall G \in U$ und $\langle X, Y \rangle > 0$ für $Y \in C$.

Sei $A \in \mathcal{F}$ (mit $\mathbf{P}[A] > 0$), $\frac{\mathbf{1}_A}{\mathbf{P}[A]} \in C$, also $\langle X, \frac{\mathbf{1}_A}{\mathbf{P}[A]} \rangle = \frac{1}{\mathbf{P}[A]} \mathbf{E}[X \cdot \mathbf{1}_A] > 0$, d.h. $X > 0$.

Setze $\mathbf{Q}[A] := \frac{\mathbf{1}_A}{\mathbf{E}[X]} \mathbf{E}[X \cdot \mathbf{1}_A]$, es ist $\mathbf{Q} \sim \mathbf{P}$.

Sei $A \in \mathcal{F}_t, i \in \{1, \dots, d\}$, betrachte selbstfinanzierende Handelsstrategie Θ mit $\Theta_s^i = \mathbf{1}_A \mathbf{1}_{\{s > t\}}, \Theta_s^j \equiv 0, j \in \{1, \dots, d\} \setminus \{i\}$.

$$\bar{V}_T(\Theta) - \bar{V}_t(\Theta) = \mathbf{1}_A(\bar{S}_T^i - \bar{S}_t^i) \in U,$$

demnach

$$\begin{aligned} 0 &= \langle X, \bar{V}_T(\Theta) - \bar{V}_t(\Theta) \rangle \\ &= \mathbf{E}[X \cdot (\bar{V}_T(\Theta) - \bar{V}_t(\Theta))] \\ &= \mathbf{E}[X \cdot \mathbf{1}_A(\bar{S}_T^i - \bar{S}_t^i)] \\ &= \mathbf{E}[X] \cdot \mathbf{E}_{\mathbf{Q}}[\mathbf{1}_A(\bar{S}_T^i - \bar{S}_t^i)], \end{aligned}$$

d.h. unter \mathbf{Q} sind die \bar{S}^i Martingale.

Sei der Markt nun arbitragefrei und vollständig.

Sei $A \in \mathcal{F}, \mathbf{1}_A$ ist ein duplizierbarer Claim; also sei Θ eine selbstfinanzierende Handelsstrategie mit $V_T(\Theta) = \mathbf{1}_A$.

Seien $\mathbf{Q}, \mathbf{Q}' \in \mathcal{P}^*$.

$$\frac{V_0(\Theta)}{S_0^0} = \mathbf{E}_{\mathbf{Q}}\left[\frac{V_T(\Theta)}{S_T^0}\right] = \frac{1}{S_T^0} \mathbf{Q}[A]$$

²In diesem endlich-dimensionalen Kontext kann man leicht folgendermaßen argumentieren: Sei $\pi_U : \mathbb{R}^{|\Omega|} \rightarrow U$ die orthogonale Projektion, dann ist $\pi_U(C)$ kompakt, und es gibt $c_0 \in C$, so dass mit $u_0 := \pi_U(c_0)$ gilt

$$\inf\{|u - c| : c \in C, u \in U\} = |u_0 - c_0| > 0.$$

Dann gilt für $X := c_0 - u_0 (\neq 0) : \langle X, u \rangle = \langle c_0, u \rangle - \langle \pi_U(c_0), u \rangle = 0$ für alle $u \in U$; für $c \in C$ und $\lambda \in [0, 1]$ ist $c_0 + \lambda(c - c_0) \in C$ (Konvexität), somit $|u_0 - (c_0 + \lambda(c - c_0))| \geq |u_0 - c_0| > 0$. Also

$$0 \leq \frac{d}{d\lambda} |_{\lambda=0} |u_0 - (c_0 + \lambda(c - c_0))|^2 = -2\langle u_0 - c_0, c - c_0 \rangle = 2\langle X, c - c_0 \rangle,$$

folglich $\langle X, c \rangle \geq \langle X, c_0 \rangle = \langle X, c_0 - u_0 \rangle = \langle X, X \rangle = |X|^2 > 0$ für jedes $c \in C$.

Aber auch

$$\frac{V_0(\Theta)}{S_0^0} = \mathbf{E}_{\mathbf{Q}'} \left[\frac{V_T(\Theta)}{S_T^0} \right] = \frac{1}{S_T^0} \mathbf{Q}'[A],$$

also $\mathbf{Q} = \mathbf{Q}'$.

Bemerkung und Beispiel:

Sei $\Omega = \{1, 2, 3\}$, $S_0^0 = S_1^0 = 1$, $S_0^1 = 1$, $S_1^1 = s_i$ auf Atom i , $\min\{s_1, s_2, s_3\} < 1 < \max\{s_1, s_2, s_3\}$.
Dieses Modell ist arbitragefrei, aber nicht vollständig (und es gibt viele äquivalente Martingalmaße).

14. Exkurs: Stochastische Integration

14.0.8. Definition: Stochastischer Prozess

Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, $(\mathcal{F}_t)_{t \geq 0}$ eine Filtration.

Eine Familie $(X_t)_{t \geq 0}$ von (reellwertigen) Zufallsvariablen heißt ein *stochastischer Prozess*.

Wir wollen annehmen, dass die betreffenden Prozesse (\mathbf{P} -fast-sicher) càdlàg-Pfade (continue à droite, limite à gauche) haben.

$$\forall t \geq 0 : \lim_{h \searrow 0} X_{t+h} = X_t, \quad \forall t > 0 : \lim_{h \searrow 0} X_{t-h} =: X_{t-} \text{ existieren.}$$

Solche Prozesse nennen wir *regulär*. Sei $X = (X_t)_{t \geq 0}$ heißt $((\mathcal{F}_t)$ -) *adaptiert*, wenn X_t \mathcal{F}_t -messbar $\forall t \geq 0$. Eine Zufallsvariable τ mit Werten in $[0, \infty]$ heißt $((\mathcal{F}_t)$ -) *Stoppzeit*, wenn $\{\tau < t\} \in \mathcal{F}_t$ $\forall t \geq 0$.

Für eine Stoppzeit τ heißt

$$\mathcal{F}[\tau] = \{A \in \mathcal{F} : A \cap \{\tau < t\} \in \mathcal{F}_t \quad \forall t \geq 0\}$$

die τ -Vergangenheit.

Bemerkung:

Ist X adaptiert und τ eine Stoppzeit, so ist auch $(X_{\min\{t, \tau\}})_{t \geq 0}$ adaptiert.

Seien τ_1, \dots, τ_k Stoppzeiten und L_1, \dots, L_k die Sprunglevel; L_i sei $\mathcal{F}[\tau_i]$ -messbar.

Für einen regulären Prozess $Y = (Y_t)_{t \geq 0}$ der Gestalt $Y_t = \sum_{i=1}^k L_i \mathbb{1}_{\{\tau_i \leq t\}}$ und einen adaptierten regulären Prozess X definieren wir einen stochastischen Prozess via

$$\int_0^t Y_- dX := \sum_{i=1}^k L_i (X_t - X_{\min\{t, \tau_i\}}) = \sum_{i=1}^k L_i \mathbb{1}_{\{\tau_i < t\}} (X_t - X_{\tau_i}), \quad t \geq 0.$$

Wir nennen dies ein *elementares stochastisches Integral*.

14.0.9. Bemerkung:

(i) $\left(\int_0^t Y_- dX \right)_{t \geq 0}$ ist ein adaptierter, regulärer stochastischer Prozess.

(ii) Für (etwaige) Sprünge gilt:

$$\Delta \left(\int_0^{\cdot} Y_- dX \right)_t = \int_0^t Y_- dX - \int_0^{t-} Y_- dX = Y_{t-} (X_t - X_{t-}) = Y_{t-} \Delta X_t$$

(iii) Wir schreiben $H_t = Y_{t-}$ für die linksstetige Version von Y .

(iv) „kanonische“ Darstellung eines adaptierten linksstetigen H mit höchstens k Sprüngen:

$$\tau_0 := 0, \tau_i := \inf\{t > \tau_{i-1} : H_t \neq H_{\tau_{i-1}}\}, i = 1, 2, \dots, k \quad (\tau_{k+1} = \infty),$$

$H_t = \sum_{i=0}^k H_{\tau_i+} \mathbb{1}_{\{\tau_i < t \leq \tau_{i+1}\}}$. Das stochastische Integral $\int H dX$ hat dann die Gestalt

$$\int_0^t H dX := \sum_i H_{\tau_i+} (X_{\min\{\tau_{i+1}, t\}} - X_{\min\{\tau_i, t\}}).$$

Definition:

Für $f, g : [0, \infty) \rightarrow \mathbb{R}$ sei $d_{lg}(f, g) = \inf\{\varepsilon > 0 : |f(t) - g(t)| \leq \varepsilon \quad \forall t \leq \frac{1}{\varepsilon}\}$, dies ist metrisiert die lokal gleichmäßige Konvergenz.

Wir schreiben $X^n \xrightarrow{\text{stochastisch}} X$, wenn $\mathbf{P}[d_{lg}(X^n, X) > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$

14.0.10. Lemma:

Das elementare stochastische Integral ist wohldefiniert und linear (in Y).

Beweis:

Linearität ist klar, für die Wohldefiniertheit genügt es zu zeigen:

Sei $Y_{t-} = H_t = \sum_{i=1}^k L_i \mathbb{1}_{\{\tau_i < t\}} = 0 \quad \forall t \geq 0$, dann gilt auch $\int_0^t H dX = 0$.

Betrachte

$$\int_0^t H dX = \underbrace{Y_t}_{=0} X_t - \sum_{i=1}^k L_i X_{\tau_i} \mathbb{1}_{\{\tau_i \leq t\}}$$

ist konstant bis auf höchstens k Sprünge, die aufgrund von Bemerkung (ii) die Höhe 0 haben müssen, wegen $\int_0^0 H dX = 0$, also $\int_0^t H dX = 0 \quad \forall t \geq 0$.

Beobachtung:

Jeder adaptierte linksstetiger Prozess H kann durch eine Folge von elementaren Integranden approximiert werden: Sei $\varepsilon_n > 0$ eine Nullfolge,

$$\tau_0^{(n)} := 0, \inf\{t > \tau_i^{(n)} : |H_{t+} - H_{\tau_i^{(n)}+}| > \varepsilon_n\} =: \tau_{i+1}^{(n)}, k_n \in \mathbb{N} \text{ so groß, dass } \mathbf{P}[\tau_{k_n}^{(n)} \geq \frac{1}{\varepsilon_n}] \leq \varepsilon_n.$$

Die Folge der approximierenden Prozesse:

$$H_t^n := \sum_{i=0}^{k_n-1} H_{\tau_i^{(n)}+} \mathbb{1}_{\{\tau_i^{(n)} < t \leq \tau_{i+1}^{(n)}\}}$$

erfüllt

$$\sup_{t < \tau_{k_n}^{(n)}} |H_t^n - H_t| \leq \varepsilon_n,$$

demnach

$$H^n \xrightarrow{\text{stoch.}} H.$$

Bemerkung / Erinnerung:

Sind X_1, X_2, \dots, X reelle Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbf{P})$, dann gilt $X_n \rightarrow X$ stochastisch \iff jede Teilfolge der (X_n) enthält eine Teilteilstolge, die \mathbf{P} -f.s. gegen X konvergiert.

In diesem Sinne hängt die Theorie der stochastischen Integration nur von den Nullmengen von \mathbf{P} ab.

14.1. Definition: Stochastischer Integrator, Itô-Integral

Ein adaptierter regulärer Prozess X heißt ein *stochastischer Integrator*, falls für jeden Integranden H ein adaptierter Prozess I , geschrieben $I_t = \int_0^t H dX$ ($= \int_0^t H_s dX_s$), existiert mit folgender Eigenschaft:

H^n elementare Integranden mit $H^n \xrightarrow{\text{stoch.}} H$, so gilt $\int H^n dX \xrightarrow{\text{stoch.}} I$.

$\int H dX$ heißt *Itô-Integral* von H bezüglich X

14.1.1. Lemma:

X ist genau dann ein Integrator, falls für jede Folge H^n von elementaren Integranden mit $H^n \xrightarrow{\text{stoch.}} 0$ gilt $\int H^n dX \xrightarrow{\text{stoch.}} 0$

Beweis:

Notwendige Bedingung: \checkmark

Hinreichende Bedingung: Angenommen es gibt Folge H^n von elementaren Integranden mit

$H^n \xrightarrow{\text{stoch.}} H$, aber $\int H^n dX \not\xrightarrow{\text{stoch.}} \int H dX$, d.h. $\exists (r_n), (s_n), \varepsilon > 0$ mit $d_{lq}(\int H^{s_n} dX, \int H^{r_n} dX) \geq \varepsilon > 0$.

Dann gilt für $\tilde{H}^n := H^{r_n} - H^{s_n} \xrightarrow{\text{stoch.}} 0$, aber $\int \tilde{H}^n dX$ konvergiert nicht.

14.1.2. Definition: Fast sicher endliche Variation

X hat *f.s. endliche Variation*, wenn es für jedes $t > 0$ eine reelle Zufallsvariable V_t gibt mit

$$\sum_{i=0}^{k-1} |X_{t_{i+1}} - X_{t_i}| \leq V_t \text{ f.s.}$$

für alle Zerlegungen $0 = t_0 < t_1 \leq \dots \leq t_k = t, k \in \mathbb{N}$.

Beobachtung:

Ein Prozess X mit endlicher Variation ist ein (stochastischer) Integrator:

Für elementaren Integranden $H_t = \sum_{i=0}^{k-1} H_{\tau_i+} \cdot \mathbb{1}_{\{\tau_i < t \leq \tau_{i+1}\}}$ ist

$$\begin{aligned} \left| \int_0^t H dX \right| &= \left| \sum_{i=0}^{k-1} H_{\tau_i+} \cdot (X_{\min\{\tau_i, t\}} - X_{\min\{\tau_i, t\}}) \right| \\ &\leq \sup_{s \leq t} |H_s| \cdot \underbrace{\sum_{i=0}^{k-1} |X_{\min\{\tau_i, t\}} - X_{\min\{\tau_i, t\}}|}_{\leq V_t} \end{aligned}$$

demnach gilt $\int H^n dX \xrightarrow{\text{stoch}} 0$, sofern $H^n \xrightarrow{\text{stoch}} 0$.

(Dieses Integral ist das (Lebesgue-)Stieltjes-Integral.)

14.1.3. Bemerkung:

- (i) Die Prozesse mit endlicher Variation sind ein Vektorraum.
- (ii) Wenn X (f.s.) wachsende Pfade hat, so ist X von endlicher Variation.
- (iii) X ist (fast sicher) von endlicher Variation $\iff X = X' - X''$ (f.s.) für X', X'' Prozesse mit wachsenden Pfaden

Betrachte zufällige Partition $\mathbb{T} = (\tau_0, \dots, \tau_k)$ mit $0 =: \tau_0 \leq \tau_1 \leq \dots \leq \tau_k \leq \tau_{k+1} = \infty$ Stoppzeiten.

Definiton

Eine Folge von Partitionen $\mathbb{T}_n := (\tau_0^{(n)}, \dots, \tau_{k_n}^{(n)})$ heißt *asymptotisch fein* auf $[0, t]$, wenn $\forall \varepsilon > 0$

$$\mathbf{P} \left[\max_{i: \tau_i^{(n)} \leq t} |\tau_{i+1}^{(n)} - \tau_i^{(n)}| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0$$

und *asymptotisch fein*, wenn dies für jedes $t > 0$ gilt.

Zu einem Integranden H und Partitionsfolge (\mathbb{T}_n) bilde $H_t^{\mathbb{T}_n} = \sum_{i=0}^{k_n} H_{\tau_i^{(n)}+} \cdot \mathbb{1}_{\{\tau_i^{(n)} < t \leq \tau_{i+1}^{(n)}\}}$, für asymptotisch feine (\mathbb{T}_n) gilt $H^{\mathbb{T}_n} \xrightarrow{\text{stoch.}} H$

Beobachtung:

Es gilt $\int H^{\mathbb{T}_n} dX \xrightarrow{\text{stoch}} \int H dX$.

Beobachtung und Definition:

Sei X ein Integrator und $0 =: \tau_0 \leq \tau_1 \leq \dots \leq \tau_k = t$ Stoppzeiten, so gilt:

$$\begin{aligned} \sum_{i=0}^{k-1} (X_{\tau_{i+1}} - X_{\tau_i})^2 &= \sum_{i=0}^{k-1} (X_{\tau_{i+1}}^2 - X_{\tau_i}^2) - 2 \sum_{i=0}^{k-1} X_{\tau_i} (X_{\tau_{i+1}} - X_{\tau_i}) \\ &= X_t^2 - X_0^2 - 2 \sum_{i=0}^{k-1} X_{\tau_i} (X_{\tau_{i+1}} - X_{\tau_i}) \\ &= X_t^2 - X_0^2 - 2 \int_0^t X_-^{\mathbb{T}} dX \end{aligned}$$

Für eine asymptotisch feine Folge (\mathbb{T}_n) gilt:

$$\sum_{i=0}^{k-1} (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{\tau_i, t\}})^2 \xrightarrow{\text{stoch}} X_t^2 - X_0^2 - 2 \int_0^t X_- dX =: [X]_t$$

Der Prozess $([X]_t)_{t \in \mathbb{R}_+}$ heißt *quadratische Variation* von X .

Für Integratoren X, Y heißt der Prozess $([X, Y]_t)_{t \in \mathbb{R}_+}$ mit

$$[X, Y]_t = X_t Y_t - X_0 Y_0 - \int_0^t Y_- dX - \int_0^t X_- dY$$

quadratische Kovariation von X und Y .

Bemerkung:

$[X]_t = 0$, wenn X endliche Variation hat.

14.1.4. Satz:

Seien X, Y, Z Integratoren und H, K Integranden.

(i) Es gilt (f.s.)

$$[X, Y] = [Y, X], \quad [X, X] = [X], \quad [\lambda X + \mu Y, Z] = \lambda [X, Z] + \mu [Y, Z] \quad \forall \lambda, \mu \in \mathbb{R}, \quad [X, Y] \leq [X] \cdot [Y].$$

(ii) $[X]$ und $[X, Y]$ haben Pfade von endlicher Variation.

(iii) $I := \int H dX$ ist ebenfalls ein Integrator und $\int K dI = \int K H dX$.

(iv) Ist $0 =: \tau_0^{(n)} \leq \tau_1^{(n)} \leq \dots \leq \tau_{k_n}^{(n)}$ asymptotisch fein auf $[0, t]$, so gilt

$$\sum_{i=0}^{k_n-1} H_{\tau_i^{(n)+}} (X_{\tau_{i+1}^{(n)}} - X_{\tau_i^{(n)}}) (Y_{\tau_{i+1}^{(n)}} - Y_{\tau_i^{(n)}}) \xrightarrow{\text{stoch}} \int_0^t H d[X, Y],$$

weiter ist

$$\left[\int H dX \right] = \int H^2 d[X] \quad \text{und} \quad \left[\int H dX, \int K dY \right] = \int H K d[X, Y].$$

Beweis:

Beobachtung: Sei (\mathbb{T}_n) asymptotisch feine Folge. Dann ist

$$\sum_i (X_{\min\{t, \tau_{i+1}^{(n)}\}} - X_{\min\{t, \tau_i^{(n)}\}})(Y_{\min\{t, \tau_{i+1}^{(n)}\}} - Y_{\min\{t, \tau_i^{(n)}\}}) \xrightarrow{\text{stoch}} [X, Y]$$

- (i) Nach Definition und Beobachtung klar. ✓ (verwende Cauchy-Schwarz für die letzte Ungleichung)
- (ii) $[X]$ hat wachsende Pfade, ist folglich von endlicher Variation, $[X, Y] = \frac{1}{4}[X + Y] - \frac{1}{4}[X - Y]$ ist somit ebenfalls von endlicher Variation
- (iii) (Beweisskizze:)
Seien $H_t := L\mathbf{1}_{\{\tau < t\}}$, $K_t := L'\mathbf{1}_{\{\tau' < t\}}$.
Dann ist $Y_t := L(X_t - X_{\min\{\tau, t\}})$,

$$\begin{aligned} \int_0^t K dY &= L'(Y_t - Y_{\min\{t, \tau'\}}) \\ &= L' \cdot L(X_t - X_{\min\{t, \tau\}} - X_{\min\{t, \tau'\}} + X_{\min\{t, \tau, \tau'\}}) \\ &= L' \cdot L(X_t - X_{\min\{t, \max\{\tau, \tau'\}\}}) \\ &= \int_0^t KH dX. \end{aligned}$$

Der allgemeine Fall ergibt sich durch Approximation mit elementaren Integranden.

- (iv) Sei $Z := XY - X_0Y_0$, dies ist ein Integrator¹

$$\begin{aligned} &\sum H_{\tau_i+}(X_{\tau_{i+1}} - X_{\tau_i})(Y_{\tau_{i+1}} - Y_{\tau_i}) \\ &= \sum H_{\tau_i+}(Z_{\tau_{i+1}} - Z_{\tau_i}) - \sum H_{\tau_i+}X_{\tau_i}(Y_{\tau_{i+1}} - Y_{\tau_i}) - \sum H_{\tau_i+}Y_{\tau_i}(X_{\tau_{i+1}} - X_{\tau_i}) \\ &\xrightarrow{\text{stoch}} \int HdZ - \int HX_-dY - \int HY_-dX \end{aligned}$$

für eine asymptotisch feine Folge von Partitionen. (ersetze „ $dZ = d[X, Y] + X_-dY + Y_-dX$ “)

Wir zeigen schließlich $[\int HdX, \int KdY] = \int HKd[X, Y]$:

Sei H elementar, $K \equiv 1$, $I := \int HdX$. Sei weiter \mathbb{T}^n asymptotisch feine Folge von

¹Denn $Z = [X, Y] + \int Y_-dX + \int X_-dY$, $[X, Y]$ hat endliche Variation, $\int Y_-dX$ und $\int X_-dY$ sind nach (iii) ebenfalls Integratoren.

Partitionen (die jeweils die Sprünge von H enthalten). Es gilt

$$\begin{aligned}
I_{\tau_{i+1}^{(n)}} - I_{\tau_i^{(n)}} &= H_{\tau_i^{(n)+}} (X_{\tau_{i+1}^{(n)}} - X_{\tau_i^{(n)}}), \\
&= \sum_{i=0}^{k_n} \underbrace{(I_{\tau_{i+1}^{(n)}} - I_{\tau_i^{(n)}})}_{=H_{\tau_i^{(n)+}}(X_{\tau_{i+1}^{(n)}} - X_{\tau_i^{(n)}})} \cdot (Y_{\tau_{i+1}^{(n)}} - Y_{\tau_i^{(n)}}) \xrightarrow{\text{stoch}} [I, Y] \\
&= \sum_{i=0}^{k_n} H_{\tau_i^{(n)+}} (X_{\tau_{i+1}^{(n)}} - X_{\tau_i^{(n)}}) \cdot (Y_{\tau_{i+1}^{(n)}} - Y_{\tau_i^{(n)}}) \\
&\xrightarrow{\text{stoch}} \int Hd[X, Y].
\end{aligned}$$

Für allgemeines H : Betrachte H^n elementar mit $H^n \xrightarrow{\text{stoch}} H$, also
 $I^n := \int H^n dX \xrightarrow{\text{stoch}} \int HdX$,

$$[I^n, Y]_t = I_t^n Y_t - I_0^n Y_0 - \int_0^t I_-^n dY - \int_0^t Y_- dI^n,$$

$$\left[\int HdX, \int KdY \right] = \left[I, \int KdY \right] = \int Kd[I, Y] = \int KHd[X, Y].$$

14.2. Satz:

Ein Martingal (X_t) mit $\mathbf{E}[X_t^2] < \infty \forall t$ ist ein Integrator und für jeden Integrand H mit

$$\mathbf{E}[\sup_{s \leq t} |H_s|^2] < \infty \forall t \quad (\star)$$

ist dann $\int HdX$ ein Martingal.

14.2.1. Erinnerung: Satz vom optionalen Stoppen

Sei τ f.s. beschränkte Stoppzeit, M ein Martingal, so gilt $\mathbf{E}[M_t | \mathcal{F}[\tau]] = M_{\min\{t, \tau\}}$.

14.2.2. Erinnerung: Doobs \mathcal{L}^2 -Ungleichung:

Sei (M_t) ein \mathcal{L}^2 -Martingal, dann gilt

$$\mathbf{P}[\sup_{s \leq t} |M_s| \geq \varepsilon] \leq \varepsilon^{-2} \cdot \mathbf{E}[M_t^2].$$

Beweis des Satzes (14.2):

- (i) Sei H f.s. beschränkter elementarer Integrand, zeige $\int HdX$ ist Martingal, o.E.
 $H_t = L\mathbb{1}_{\{\tau < t\}}$. Sei $s < t$:

$$\begin{aligned}
 \mathbf{E} \left[\int_0^t HdX \middle| \mathcal{F}[\max\{s, \tau\}] \right] &= \mathbf{E}[L\mathbb{1}_{\{\tau < t\}}(X_t - X_\tau) | \mathcal{F}[\max\{s, \tau\}]] \\
 &= L \cdot \mathbf{E}[\mathbb{1}_{\{\tau < t\}}(X_t - X_\tau) | \mathcal{F}[\max\{s, \tau\}]] \\
 &= L \cdot (X_{\min\{t, \max\{s, \tau\}\}} - X_{\min\{t, \tau, \max\{s, \tau\}\}}) \\
 &= L\mathbb{1}_{\{\tau \leq s\}}(X_s - X_{\min\{s, \tau\}}) + L\mathbb{1}_{\{\tau > s\}} \underbrace{(X_{\min\{t, \tau\}} - X_{\min\{t, \tau\}})}_{=0} \\
 &= L\mathbb{1}_{\{\tau \leq s\}}(X_s - X_{\min\{s, \tau\}}) \text{ ist } \mathcal{F}_s\text{-messbar,}
 \end{aligned}$$

also

$$\begin{aligned}
 \mathbf{E} \left[\int_0^t HdX | \mathcal{F}_s \right] &= \mathbf{E}[\mathbf{E}[\dots | \mathcal{F}[\max\{s, \tau\}]] | \mathcal{F}_s] \\
 &= L\mathbb{1}_{\{\tau \leq s\}}(X_s - X_{\min\{s, \tau\}}) \\
 &= \int_0^s HdX.
 \end{aligned}$$

- (ii) Sei $H_t = \sum_{i=0}^k L_i \mathbb{1}_{\{\tau_i < t \leq \tau_{i+1}\}}$ mit $|L_i| \leq c$ f.s. Für $i < j$ ist

$$\begin{aligned}
 &\mathbf{E}[L_i \cdot (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{t, \tau_i\}}) \cdot L_j \cdot (X_{\min\{t, \tau_{j+1}\}} - X_{\min\{t, \tau_j\}}) | \mathcal{F}[\tau_j]] \\
 &= L_i (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{t, \tau_i\}}) \cdot L_j \cdot \underbrace{\mathbf{E}[X_{\min\{t, \tau_{j+1}\}} - X_{\min\{t, \tau_j\}} | \mathcal{F}[\tau_j]]}_{=0} \\
 &= 0 \text{ f.s.,}
 \end{aligned}$$

also

$$\begin{aligned}
 \mathbf{E} \left[\left(\int_0^t HdX \right)^2 \right] &= \mathbf{E} \left[\sum_{i,j=0}^k L_i (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{t, \tau_i\}}) \cdot L_j (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{t, \tau_j\}}) \right] \\
 &= \sum_{i=0}^k \mathbf{E}[L_i^2 (X_{\min\{t, \tau_{i+1}\}} - X_{\min\{t, \tau_i\}})^2] \\
 &\leq c^2 \mathbf{E}[(X_t - X_0)^2].
 \end{aligned}$$

- (iii) Sei H^n eine Folge von elementaren Integranden mit $H^n \xrightarrow{\text{stoch}} 0$,
 $K^n := \max\{\min\{H^n, c\}, (-c)\}$,

$$\begin{aligned}
 \mathbf{P} \left[\sup_{s \leq t} \left| \int_0^s H^n dX \right| > \varepsilon \right] &\leq \mathbf{P} \left[\sup_{s \leq t} |H_s| > c \right] + \underbrace{\mathbf{P} \left[\sup_{s \leq t} \left| \int_0^s K^n dX \right| > \varepsilon \right]}_{\leq \varepsilon^{-2} \mathbf{E} \left[\left(\int_0^t K^n dX \right)^2 \right] \leq \varepsilon^{-2} c^2 \mathbf{E}[(X_t - X_0)^2]} \\
 &\leq \varepsilon^{-2} \mathbf{E} \left[\left(\int_0^t K^n dX \right)^2 \right] \leq \varepsilon^{-2} c^2 \mathbf{E}[(X_t - X_0)^2]
 \end{aligned}$$

(nach Doobs Ungleichung), folglich

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left[\sup_{s \leq t} \left| \int_0^s H^n dX \right| > \varepsilon \right] = 0 \quad \forall \varepsilon > 0.$$

- (iv) Nach obigem ist $\int H dX$ ein (\mathcal{L}^2) -Martingal für beschränktes H , für allgemeines H , das (\star) erfüllt, ergibt sich dies mittels Approximation. Die entscheidende Beobachtung ist, dass die für die „kanonischen“ Approximanden $H^n \xrightarrow{\text{stoch}} H$ gilt $\sup_{n \in \mathbb{N}} \mathbf{E} \left[\left(\int_0^t H^n dX \right)^2 \right] < \infty$.

Bericht:

Die allgemeinsten stochastischen Integratoren sind sog. Semimartingale, d.h. stochastische Prozesse, die sich als Summe aus einem (lokalen) Martingal und einem Prozess von f.s. endlicher Variation darstellen lassen (siehe z.B. Philip Protter, *Stochastic integration and differential equations. A new approach*. Springer, 1990).

Bemerkung/Bericht:

Die Brown'sche Bewegung ist ein (\mathcal{L}^2) -Martingal.

14.3. Itô-Formel

14.3.1. Satz:

Seien X^1, \dots, X^d Integratoren mit fast sicher stetigen Pfaden, $f \in C^2(\mathbb{R}^d)$. Dann gilt fast sicher:

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t f_{x_i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t f_{x_i x_j}(X_s) d[X_s^i, X_s^j]$$

(mit $f_{x_i} = \frac{\partial}{\partial X_i} f$, $f_{x_i x_j} = \frac{\partial^2}{\partial X_i \partial X_j} f$).

Beweis:

Sei $0 \leq u \leq v \leq t$, $|X_u - X_v| \leq \eta$, $\sup_{s \leq t} |X_s| \leq K$. Gemäß Taylor-Entwicklung gilt

$$f(X_v) = f(X_u) + \sum_{i=1}^d f_{x_i}(X_u)(X_v^i - X_u^i) + \frac{1}{2} \sum_{i,j=1}^d f_{x_i x_j}(X_u)(X_v^i - X_u^i)(X_v^j - X_u^j) + R(u, v),$$

wobei

$$\begin{aligned} |R(u, v)| &\leq \sup_{x \in S(u, v)} \left| \sum_{i,j=1}^d (f_{x_i x_j}(x) - f_{x_i x_j}(X_u))(X_v^i - X_u^i)(X_v^j - X_u^j) \right| \\ &\leq C |X_v - X_u|^2 \cdot \underbrace{\max_{i,j=1, \dots, d} \max_{y, z \in B_K(0), |y-z| \leq \eta} |f_{x_i x_j}(y) - f_{x_i x_j}(z)|}_{\varphi(K, \eta)} \end{aligned}$$

wobei $S(u, v) =$ Strecke zwischen X_u und X_v .

(Beobachtung: Für jedes $K \geq 0$ gilt $\varphi(K, \eta) \xrightarrow{\eta \rightarrow 0} 0$, da die zweiten Ableitungen auf jeder kompakten Menge gleichmäßig stetig sind.)

Sei $\tau_0 \leq \tau_1 \leq \dots \leq \tau_k, \tau_k = t$ eine Partition mit $|X_{\tau_{l+1}} - X_{\tau_l}| \leq \eta$:

$$\begin{aligned} \mathbb{1}_{\sup_{s \leq t} |X_s| \leq K} & \left| f(X_t) - f(X_0) - \sum_{l=0}^{k-1} \sum_{i=1}^d f_{x_i}(X_{\tau_l})(X_{\tau_{l+1}}^i - X_{\tau_l}^i) \right. \\ & \left. - \sum_{l=0}^{k-1} \sum_{i,j=1}^d f_{x_i x_j}(X_{\tau_l})(X_{\tau_{l+1}}^i - X_{\tau_l}^i)(X_{\tau_{l+1}}^j - X_{\tau_l}^j) \right| \\ & \leq C \cdot \varphi(K, \eta) \sum_{l=0}^{k-1} \sum_{i=1}^d (X_{\tau_{l+1}}^i - X_{\tau_l}^i)^2 \end{aligned}$$

Längs einer asymptotisch feinen Folge von Partitionen konvergiert die linke Seite gegen

$$\mathbb{1}_{\sup_{s \leq t} |X_s| \leq K} \left| f(X_t) - f(X_0) - \sum_{i=1}^d \int_0^t f_{x_i}(X_s) dX_s^i - \frac{1}{2} \sum_{i,j=1}^d \int_0^t f_{x_i x_j}(X_s) d[X_s^i, X_s^j] \right|,$$

die rechte Seite gegen

$$C \varphi(K, \eta) \sum_{i=1}^d [X^i]_t.$$

Die Behauptung folgt mit $\eta \rightarrow 0$, dann $K \rightarrow \infty$.

15. Die geometrische Brown'sche Bewegung und die Black-Scholes-Formel

In Abschnitt 13.3 hatten wir die Black-Scholes-Formel als Limes des fairen Preises einer (europäischen) Option in einer Folge von Cox-Ross-Rubinstein-Modellen hergeleitet, in der die Anzahl Handelsperioden in geeigneter Weise unendlich groß wird. Dieser Zugang gestattete, die BS-Formel mit „elementaren“ Methoden (im Wesentlichen dem Zentralen Grenzwertsatz) herzuleiten, andererseits verliert man dabei die Limesdynamik des Preisprozesses und der Hedgingstrategien „aus dem Blick.“ Wir tragen hier nach, wie die Theorie der stochastischen Integration aus Abschnitt 14 diese Lücke füllen kann.

Sei $(B_t)_{t \geq 0}$ eine Brownsche Bewegung, d.h. ein stochastischer Prozess mit unabhängigen Zuwächsen und stetigen Pfaden, bei dem $B_{t+h} - B_t \sim \mathcal{N}(0, h)$ für alle $t, h \geq 0$, und $B_0 = 0$. (Existenz hatten wir in Übungsaufgabe 3 auf Blatt 11 gesehen.) Offenbar ist $(B_t)_{t \geq 0}$ ein Martingal (mit $\mathbf{E} B_t^2 = t < \infty$ für jedes $t \geq 0$), und somit nach Satz 14.2 ein stochastischer Integrator.

Beobachtung (quadratische Variation der Brown'schen Bewegung)

Es gilt $[B]_t = t$, was man gelegentlich auch suggestiv schreibt als „ $(dB_t)^2 = dt$ “.

Beweis:

Wegen

$$\mathbf{E}[B_{t+h}^2 | \mathcal{F}_t] = \mathbf{E}[B_t^2 + 2B_t(B_{t+h} - B_t) + (B_{t+h} - B_t)^2 | \mathcal{F}_t] = B_t^2 + \mathbf{E}[(B_{t+h} - B_t)^2 | \mathcal{F}_t] = B_t^2 + h$$

ist $M_t := B_t^2 - t$ ein stetiges Martingal (mit $\mathbf{E} M_t^2 < \infty$ für jedes $t \geq 0$), andererseits ist nach Satz 14.2 auch

$$M'_t := B_t^2 - [B]_t = 2 \int_0^t B_s dB_s$$

ein Martingal, somit auch $M''_t := M_t - M'_t = [B]_t - t$. Darüberhinaus ist $[M'']_t \equiv 0$, denn M'' hat offensichtlich Pfade von beschränkter Variation, somit ist auch

$$(M''_t)^2 = 2 \int_0^t M''_s dM''_s$$

ein Martingal. Folglich gilt $\mathbf{E} (M''_t)^2 = \mathbf{E} (M''_0)^2 = 0$, d.h. $M'' \equiv 0$ f.s., also $[B]_t = t$.

(Hinter diesem Beweis steckt die allgemeine Beobachtung, dass die quadratische Variation eines Martingals X charakterisiert ist als derjenige Prozess V , für den $X^2 - V$ wiederum ein Martingal ist. Alternativ kann man im Fall der Brown'schen Bewegung auch explizit quadrierte Inkremente längs einer geeigneten Partitionsfolge aufsummieren.)

Geometrische Brown'sche Bewegung

Sei S_0 \mathcal{F}_0 -adaptiert, $\mu \in \mathbb{R}$, $\sigma > 0$. Die Lösung $(S_t)_{t \geq 0}$ der (stochastischen Integral-)Gleichung

$$(*) \quad S_t = S_0 + \int_0^t \mu S_u du + \int_0^t \sigma S_u dB_u, \quad t \geq 0$$

ist gegeben durch

$$(**) \quad S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right), \quad t \geq 0.$$

S heißt *geometrische Brown'sche Bewegung*. Man schreibt (*) auch häufig in der suggestiven Form einer *stochastischen Differentialgleichung*

$$dS_t = \mu S_t dt + \sigma S_t dB_t.$$

Beweis: Anwendung der Itô-Formel auf $f(B_t, t)$ mit $f(b, t) := \exp(\sigma b + (\mu - \sigma^2/2)t)$ ($f_b(b, t) = \sigma f(b, t)$, $f_{bb}(b, t) = \sigma^2 f(b, t)$, $f_t(b, t) = (\mu - \sigma^2/2)f(b, t)$) liefert

$$\begin{aligned} f(B_t, t) &= \underbrace{f(B_0, 0)}_{=1} + \int_0^t \sigma f(B_s, s) dB_s + \int_0^t (\mu - \sigma^2/2) f(B_s, s) ds + \frac{1}{2} \int_0^t \sigma^2 f(B_s, s) d[B]_s \\ &= 1 + \int_0^t \mu f(B_s, s) ds + \int_0^t \sigma f(B_s, s) dB_s, \end{aligned}$$

denn $[B]_s = s$ und mit $I_t = t$ ist $[I] = [I, B] \equiv 0$. Durch Multiplikation mit S_0 folgt (*).

Black-Scholes-Gleichung

Seien eine *Volatilität* $\sigma > 0$, *risikoloser Zinssatz* $r \geq 0$ und ein (beschränktes) stetiges Auszahlungsprofil $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ gegeben. Wie wir in Abschnitt 13.3 gesehen haben, ist der Black-Scholes-Preis der europäischen Option mit Auszahlungsprofil f und Laufzeit $t \geq 0$ bei aktuellem Kurs x des Basiswertpapiers gegeben durch

$$\begin{aligned} v(x, t) &= e^{-rt} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-b^2/(2t)} f(xe^{\sigma b + (r - \sigma^2/2)t}) db \\ &= e^{-rt} \mathbf{E}\left[f(xe^{\sigma B_t + (r - \sigma^2/2)t})\right]. \end{aligned}$$

Die Wertfunktion v löst folgendes Cauchy-Problem (die *Black-Scholes-Gleichung*)

$$\begin{cases} \frac{\partial}{\partial t} v(x, t) = rx \frac{\partial}{\partial x} v(x, t) + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2} v(x, t) - rv(x, t), & x, t > 0, \\ v(x, 0) = f(x), & x > 0. \end{cases}$$

Black-Scholes-Modell, Δ -Hedge

Das Black-Scholes-Modell ist ein zeitkontinuierliches Marktmodell, in dem es zwei Wertpapiere gibt: Eine festverzinsliche Anleihe $S_t^0 = e^{rt}$ mit risikolosem Zinssatz $r \geq 0$, und ein risikobehaftetes Wertpapier S^1 , dessen Dynamik durch eine geometrische Brown'sche Bewegung gegeben ist,

$$dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 dB_t$$

(dieses ist die Limesdynamik der in Abschnitt 13.3 betrachteten Folge von Modellen).

Wir betrachten eine (europäische) Option mit Fälligkeit $T > 0$ und Auszahlungsprofil $f(S_T^1)$. Nach obigem ist der Wert zur Zeit t , $t \in [0, T]$, gegeben durch $V_t := v(S_t^1, T - t)$. Die Itô-Formel liefert

$$\begin{aligned} dV_t &= v_x(S_t^1, T - t)dS_t^1 + \frac{1}{2}v_{xx}(S_t^1, T - t)\underbrace{d[S^1]_t}_{=\sigma^2(S_t^1)^2dt} - v_t(S_t^1, T - t)dt \\ &= v_x(S_t^1, T - t)dS_t^1 + r(v(S_t^1, T - t) - S_t^1v_x(S_t^1, T - t))dt \\ &= \underbrace{v_x(S_t^1, T - t)}_{=:\Theta_t^1}dS_t^1 + \underbrace{\frac{V_t - S_t^1v_x(S_t^1, T - t)}{S_t^0}}_{=:\Theta_t^0}dS_t^0 \end{aligned}$$

(wobei wir benutzen, dass $\frac{1}{2}v_{xx}x - v_t = r(v - xv_x)$ gemäß der Black-Scholes-Gleichung, und $dS_t^0 = rS_t^0dt$ nach Annahme).

Demnach ist das Portfolio, das zur Zeit t aus Θ_t^0 Einheiten S^0 und Θ_t^1 Einheiten S^1 besteht (also den Wert $V_t = \Theta_t^0S_t^0 + \Theta_t^1S_t^1$ hat), *selbstfinanzierend*, d.h. es genügt der Bedingung

$$V_t = V_0 + \int_0^t \Theta_u^0 dS_u^0 + \int_0^t \Theta_u^1 dS_u^1,$$

und es dupliziert das gewünschte Auszahlungsprofil, denn $V_T = v(S_T, T - T) = f(S_T)$.

Diese Replikationsstrategie heißt Δ -Hedge: Die Anzahl gehaltener Aktien ist gerade die Ableitung des aktuellen Werts der Option nach dem Basiskurs, das sogenannte Δ der Option. Eine umfassendere Diskussion findet sich beispielsweise in Abschnitt 5.6 des Buches von Hans Föllmer und Alexander Schied, *Stochastic finance. An introduction in discrete time, 2nd Ed.*, de Gruyter, 2004.

Teil III.
Anhang

A. Verwendeter R-Code

A.1. Konfidenzintervalle im Binomialmodell: Für Abbildung 6.3 verwendeter R-Code

```
# Konfidenzintervalle fuer den Erfolgsparameter bei der
# Binomialverteilung

# Exaktes Konfidenzintervall basierend auf Beta-Quantilen
bin.konf.ex <- function(n, x, a) {
  c(qbeta(a/2,x,n-x+1), qbeta(a/2,x+1,n-x,lower.tail=FALSE))
}

# Auf Normalapproximation basiertes Konfidenzintervall
bin.konf.nappr <- function(n, x, a) {
  c(x/n-qnorm(1-a/2)/(2*sqrt(n)), x/n+qnorm(1-a/2)/(2*sqrt(n)))
}

# Auf Chebychev-Ungleichung basiertes Konfidenzintervall
bin.konf.cheby <- function(n, x, a) {
  c(x/n-1/sqrt(4*n*a), x/n+1/sqrt(4*n*a))
}

anteile <- c(0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5)

erzeuge.bild <- function(n, alpha) {
  linientyp <- c(3,2,1)
  ve <- 0.01
  br <- 0.005

  plot(c(-1,-1), xlim=c(0,0.5), ylim=c(0,0.8),
       xlab="Beobachteter Anteil", ylab="",
       main=c("Konfidenzintervalle für die Binomialverteilung",
             paste("zum Irrtumsniveau ",alpha," (n=",n,")",sep="")))
  abline(1,0)

  for (a in anteile) {
    x <- round(n*a)

    i <- bin.konf.cheby(n,x,alpha)

    lines(c(a-ve,a+ve), i, lty=linientyp[1])
  }
}
```

```

lines(c(a-ve-br,a-ve+br), c(i[1],i[1]), lty=linientyp[1])
lines(c(a-ve-br,a-ve+br), c(i[2],i[2]), lty=linientyp[1])

i <- bin.konf.nappr(n,x,alpha)
lines(c(a,a), i, lty=linientyp[2])
lines(c(a-br,a+br), c(i[1],i[1]), lty=linientyp[2])
lines(c(a-br,a+br), c(i[2],i[2]), lty=linientyp[2])

i <- bin.konf.ex(n,x,alpha)
lines(c(a+ve,a+ve), i, lty=linientyp[3])
lines(c(a+ve-br,a+ve+br), c(i[1],i[1]), lty=linientyp[3])
lines(c(a+ve-br,a+ve+br), c(i[2],i[2]), lty=linientyp[3])
}

legend(0.0, 0.8, c("Chebychev","N.approx","Exakt"), lty=linientyp)
}

erzeuge.bild(20,0.05)
erzeuge.bild(50,0.05)
erzeuge.bild(100,0.05)

```

A.2. Temperaturdiagramm

In Abschnitt 12.3.4 verwendeter R-Code.

```

> a<-read.table("karlsruhe_august.dat", header=TRUE)
> attach(a)
> Zeit<-Jahr/1000
> b<-data.frame(Temp=Temp, Zeit=Zeit, Zeit2=Zeit^2, Zeit3=Zeit^3)
> m1<-lm(Temp~Zeit, data=b)
> m1

```

Call:

```
lm(formula = Temp ~ Zeit, data = b)
```

Coefficients:

(Intercept)	Zeit
16.804	1.064

```

> m2<-lm(Temp~Zeit+Zeit2, data=b)
> m2

```

Call:

```
lm(formula = Temp ~ Zeit + Zeit2, data = b)
```

Coefficients:

(Intercept)	Zeit	Zeit2
660.8	-676.3	177.9

```

> m3<-lm(Temp~Zeit+Zeit2+Zeit3, data=b)
> m3

Call:
lm(formula = Temp ~ Zeit + Zeit2 + Zeit3, data = b)

Coefficients:
(Intercept)      Zeit      Zeit2      Zeit3
      -7532      12252     -6618       1190

> plot(Zeit, Temp, main="Mittlere Augusttemperatur in Karlsruhe",
> xlab="Jahr/1000",ylab="Temp")
> points(Zeit,fitted.values(m1),type='l',lty=1)
> points(Zeit,fitted.values(m2),type='l',lty=2)
> points(Zeit,fitted.values(m3),type='l',lty=3)
> legend(1.87,24,legend=c("Grad 1","Grad 2","Grad 3"),lty=c(1,2,3))
>
> dev.copy2eps(file="karlsruhe.eps")
X11
  2
> Var.ti <- sum(Zeit^2)/length(Zeit)-mean(Zeit)^2
> V.stern <- sum(residuals(m1)^2)/(206-2)
> 1.064/sqrt(V.stern/(206*Var.ti))
[1] 0.6136871
> qt(0.995,df=204)
[1] 2.600144

```


Abbildungsverzeichnis

2.1. Unter allen Punkten auf der Diagonalen $\langle(1, 1, \dots, 1)\rangle$ hat $(M(\mathbf{x}), M(\mathbf{x}), \dots, M(\mathbf{x}))$ den kleinsten euklidischen Abstand von (x_1, \dots, x_n)	9
3.1. Darstellung von $\psi(s) = 1 - s + s \log s$	18
6.1. Darstellung von Konfidenzbereichen in der x - ϑ -Ebene	30
6.2. Darstellung des α -Quantils q	31
6.3. Konfidenzintervalle zum Niveau 0.05 für den Erfolgsparameter im Binomialmodell, basierend auf Chebyshev-Ungleichung, Normalapproximation und auf Satz 6.2.2, als Funktion des beobachteten Anteils der Erfolge, für $n \in \{20, 50, 100\}$. Siehe Abschnitt A.1 für den R-Code zur Erzeugung dieser Bilder.	34
8.1. Darstellung der kombinatorischen Größe $N(m; k, l)$, als Anzahl der Wege durch dieses $k \times l$ -Gitter, so dass unter dem Weg nur $m = \sum m_i$ viele Punkte liegen	41
8.2. Darstellung der Verteilungsfunktionen von \mathbf{P} und \mathbf{Q} , wenn $\mathbf{P} \preceq \mathbf{Q}$	44
9.1. Abbildung zur Wahl von c , falls die Verteilungsfunktion gerade an dieser Stelle einen Sprung hat. Deswegen kann es in manchen Fällen auch sinnvoll sein, den Test zu randomisieren.	48
11.1. Darstellung der orthogonalen Projektion im Zweidimensionalen auf einen eindimensionalen Unterraum	60
12.1. Ausgleichsgerade zu gegebenen Punkten	67
12.2. Darstellung der Projektionen eines Punktes X im \mathbb{R}^3 auf $H \subset L \subset \mathbb{R}^3$, L ist dabei die Ebene, die durch das Rechteck L dargestellt werden soll und H der Unterraum, der von der blauen Geraden aufgespannt wird.	72

Index

- χ^2 -Verteilung, 36
- A-posteriori-Dichte, 27
- A-priori-Dichte, 27
- A-priori-Maß, 27
- adaptiert, 95
- Arbitrage, 83
 - frei, 92
 - möglichkeit, 92
 - preis, 92
 - preisprozess, 92
- Bayes-Schätzer, 27
- Bias, 9
- Call-Put-Parität, 88
- Claim, 92
 - duplizierbarer -, 92
- Contingent claim, 83
- Cramér-Rao-effizient, 13
- Dichte
 - A-posteriori-, 27
 - A-priori-, 27
- Doob
 - $s\mathcal{L}^2$ -Ungleichung, 101
- Entropie
 - relative, 17
- Erwartungstreue, 9
- Erwartungswert
 - bedingter, 21
- Exponentialverteilung, 18
- Fehler
 - 1.Art, 39
- Filtration, 85
- Fisher-Information, 11
- Fisher-Verteilung, 36
- Formel
 - Black-Scholes-, 88
 - Itô-, 103
- Forward, 88
- Fraktil, 30
- Handelsstrategie, 91
- Hedge, 92
- Hypothese
 - Alternativ-, 39
 - Null-, 39
 - Testen von -n, 3
- Integral
 - Itô-, 97
- Integrator
 - stochastischer -, 97
- Irrtumsniveau, 29
- Itô
 - Formel, 103
- Konfidenzbereich, 29, 32
 - e im normalverteilten linearen Modell, 73
- Konfidenzintervall, 3
- Konsistenz, 10, 16
- Kullback-Leibler-Abstand, 17
- Lemma
 - von Stein, 49
 - Neyman-Pearson-, 47
- Likelihood
 - Funktion, 7
 - Quotient, 53
- Maß
 - äquivalentes Martingal-, 92
 - A-priori-, 27
- Marktmodell
 - vollständiges -, 92
- Maximum-Likelihood
 - Prinzip, 7

- Schätzer, 7, 16
- Median, 30
 - Konfidenzintervalle für -, 39
- Mittlerer quadratischer Fehler, 9
- Modell
 - Gauß'sches-, 54
 - CRR-, 85
 - exponentielles, 13
 - lineares -, 69
 - Normales, 8
 - parametrisches, 4
 - Standard-, 4
 - statistisches, 3
- Option
 - Call-, 83
 - Put-, 83
- Parametermerkmal, 4
- Poisson
 - Verteilung, 18
- prävisibel, 86
- Prozess
 - Arbitragepreis-, 92
 - stochastischer -, 95
 - Wert-, 91
 - Zugewinn-, 91
- Punkt-
 - Schätzer, 3
- Quantile, 30
- Rao-Blackwellisierung, 24
- Regression, 67
- Regularität, 11
- Satz
 - vom optionalen Stoppen, 101
 - von Gauß-Markov, 70
 - von Lehmann-Scheffé, 24
 - Dichtentransformations-, 35
 - Fundamentals. der Preistheorie, 90
 - Steiner'scher Verschiebungs-, 8
 - Verallgemeinerter S. von Student, 72
 - von Cramér-Rao, 12
 - von Rao-Blackwell, 24
- Schätzer
 - Bayes-, 27
 - kleinste-Quadrate-, 68
 - Maximum-Likelihood-, 7
 - regulärer, 12
- Scorefunktion, 11
- selbstfinanzierend, 91
- Simpson-Paradoxon, 64
- Statistik, 4
 - Ordnungs-, 31
- stochastisch kleiner, 43
- Stoppzeit, 95
- Suffizienz, 23
- Taxiproblem, 4
- Test, 47
 - t -T. auf Gleichheit der Mittelwerte, 78
 - Alternativ-, 47
 - χ^2 -, 55
 - χ^2 -Anpassungs-, 59
 - χ^2 -T. auf Unabhängigkeit, 62
 - χ^2 -Anpassungs-, 62
 - effektives Niveau eines -s, 47
 - F-T. auf Gleichheit der
 - Gruppenmittelwerte, 78
 - Gütefunktion eines -s, 47
 - gleichmäßig bester -, 47
 - Macht des -s, 47
 - Neyman-Pearson-, 48
 - randomisierter -, 47
 - Student'scher t -, 56
 - Umfang eines -s, 47
- unverfälscht, 56
- Varianz
 - analyse, 76
- Variation
 - f.s. endliche -, 97
- Verteilung
 - χ_n^2 , 36
 - Beta-, 31
 - Fisher, 36
 - Multinomial-, 59
 - Multivariate Standardnormal-, 60
 - Student'sche-, 37
- Vollständigkeit, 24
- Wilcoxon
 - Rangsummen-Statistik, 40