Statistik I

Universität Ulm
Abteilung Stochastik

Vorlesungsskript
Prof. Dr. Volker Schmidt
Stand: Sommersemester 2004

Ulm, im Juni 2004

Inhaltsverzeichnis

Literatur

[1] Bickel, P., Doksum, K.
Mathematical Statistics: Basic Ideas and Selected Topics
2nd ed., Vol.1, Prentice Hall, London 2001

[2] Casella, G., Berger, R.L.
Statistical Inference
2nd ed., Duxbury, Pacific Grove (CA) 2002

[3] Georgii, H.-O.
Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik
de Gruyter, Berlin 2002

[4] Karian, Z.A., Tapis, E.
Probability and Statistics Explorations with MAPLE
2nd ed., Prentice Hall, London 1999

[5] Krengel, U.
Einführung in die Wahrscheinlichkeitstheorie und Statistik
6.Aufl., Vieweg-Verlag, Braunschweig 2002

[6] Lehmann, E.L.
Elements of Large-Sample Theory
Springer-Verlag, New York 1999

[7] Lehn, J., Wegmann, H.
Einführung in die Statistik
3.Aufl., Teubner-Verlag, Stuttgart 2000

[8] Overbeck-Larisch, M., Dolejsky, W.
Stochastik mit Mathematica
Vieweg-Verlag, Braunschweig 1998

[9] Pruscha, H.
Vorlesungen über Mathematische Statistik
Teubner-Verlag, Stuttgart 2000

[10] Storm, R.
Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle
Fachbuchverlag Leipzig/Hauser-Verlag München, 2001.

1 Stichproben und Stichprobenfunktionen

Zu den Zielen der mathematischen Statistik gehört die Beschreibung und Untersuchung der Eigenschaften bzw. Gesetzmäßigkeiten von (großen) Datensätzen. Dabei kann es sich einerseits um sogenannte

- *reale Daten*

handeln, die sich z.B. bei der Beobachtung (Messung) von Vorgängen bzw. Strukturen in Natur, Technik oder Wirtschaft ergeben, oder es können

- *synthetische Daten*

sein, die bei der Simulation solcher Vorgänge bzw. Strukturen durch Computeralgorithmen erzeugt werden.

Die grundlegende Idee der Statistik, um diese Zielstellung zu erreichen, ist die *stochastische Modellierung* der vorliegenden Daten.

Dabei nutzt die Statistik Begriffe und Ergebnisse der *Wahrscheinlichkeitsrechnung*, wie zum Beispiel:

- Ereignis und Wahrscheinlichkeit,
- Zufallsvariable und Verteilung,
- Erwartungswert und Varianz,
- stochastische Unabhängigkeit,
- Gesetz der großen Zahlen,
- zentraler Grenzwertsatz,

vgl. das Skript zur Vorlesung „Wahrscheinlichkeitsrechnung" im WS 03/04:

http://www.mathematik.uni-ulm.de/stochastik/lehre/ws03_04/wr/skript/skript.html

Verweise auf dieses Vorlesungsmanuskript werden wir mit dem Zusatz „WR" vor der Nummer der zitierten Abschnitte, Lemmata, Theoreme, Korollare bzw. Formeln kennzeichnen.

1.1 Zufallsstichprobe

Der Vektor der vorliegenden Daten $(x_1, \ldots, x_n)$ kann im allgemeinen eine komplizierte Struktur aufweisen.

- Dabei muss der „Wert" x_i nicht unbedingt eine Zahl sein, sondern x_i kann für jedes $i = 1, \ldots, n$ selbst ein Vektor sein, der beispielsweise die Lage, Größe, Form und Orientierung eines geometrischen Objektes beschreiben kann.
- In dieser einführenden Vorlesung setzen wir jedoch meistens voraus, dass $x_i \in \mathbb{R}$ für jedes $i = 1, \ldots, n$.
- Eine Ausnahme bilden die in der zweiten Hälfte des Semesters diskutierten *Zwei-Stichproben-Probleme*, bei denen der Fall $x_i \in \mathbb{R}^2$ für jedes $i = 1, \ldots, n$ betrachtet wird.

Wir nehmen an, dass die Daten $x_1, \ldots, x_n$ die *Realisierung* eines stochastischen Modells sind.

- Und zwar sei $x_1, \ldots, x_n$ die Realisierung einer Folge von unabhängigen und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n : \Omega \to \mathbb{R}$, die über einem (im allgemeinen nicht näher spezifizierten) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind.
- D.h. insbesondere, dass für ein $\omega \in \Omega$

$$X_i(\omega) = x_i \qquad \forall i \in \{1, \ldots, n\}.$$

- Außerdem setzen wir stets voraus, dass $\mathbb{E}(X_i^2) < \infty$ für $i = 1, \ldots, n$.

Definition

1. Der Vektor $(x_1, \ldots, x_n)$ heißt (konkrete) *Stichprobe*.
2. Die Menge $B \subset \mathbb{R}^n$ aller (potentiell möglichen) Stichproben $(x_1, \ldots, x_n)$ heißt *Stichprobenraum*, wobei wir zur Vereinfachung der Notation annehmen, dass $B = \mathbb{R}^n$.
3. Der Zufallsvektor $(X_1, \ldots, X_n)$ heißt *Zufallsstichprobe*.
4. Für jedes $i = 1, \ldots, n$ heißt x_i *Stichprobenwert* von $(x_1, \ldots, x_n)$. Analog hierzu nennt man X_i *Stichprobenvariable* von $(X_1, \ldots, X_n)$.
5. Die Dimension n von $(x_1, \ldots, x_n)$ bzw. $(X_1, \ldots, X_n)$ heißt *Stichprobenumfang*.

Beachte

- Die allgemeine Zielstellung der statistischen Datenanalyse, die in der Einleitung von Kapitel 1 diskutiert wurde, kann nun wie folgt präzisiert werden: Aus den vorliegenden Daten $x_1, \ldots, x_n$ sollen Schlussfolgerungen über Eigenschaften der (unbekannten) Verteilung der Zufallsstichprobe $(X_1, \ldots, X_n)$ gezogen werden.
- Weil wir voraussetzen, dass die Stichprobenvariablen $X_1, \ldots, X_n$ unabhängig und identisch verteilt sind, wird die Verteilung von $(X_1, \ldots, X_n)$ eindeutig durch die (Rand-) Verteilungsfunktion $F = F_{X_i}$ einer (einzelnen) Stichprobenvariablen X_i bestimmt, vgl. Abschnitt WR–3.3.5 des Skriptes zur Vorlesung „Wahrscheinlichkeitsrechnung" im WS 03/04.
- Aus den Daten $x_1, \ldots, x_n$ sollen deshalb Schlussfolgerungen über Eigenschaften der unbekannten Verteilungsfunktion F gezogen werden.

1.2 Stichprobenfunktionen

Um Eigenschaften der Verteilungsfunktion F zu bestimmen, werden Funktionen $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ betrachtet, die der Stichprobe $(x_1, \ldots, x_n)$ die „Bewertung" $\varphi(x_1, \ldots, x_n) \in \mathbb{R}^m$ zuordnen, vgl. auch Abschnitt WR–3.4. Dies führt zu der folgenden Begriffsbildung.

Definition Sei $m \geq 1$ eine beliebige natürliche Zahl. Eine Borel-messbare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ heißt *Stichprobenfunktion*.

Beachte

- Im allgemeinen ist es üblich, die zusammengesetzte Abbildung $\varphi(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^m$ mit

$$\varphi(X_1, \ldots, X_n)(\omega) = \varphi(X_1(\omega), \ldots, X_n(\omega))$$

Statistik zu nennen, d.h., φ ist dann eine Funktion der Zufallsstichprobe $(X_1, \ldots, X_n)$.
- Manchmal spricht man auch von einer *zufälligen Stichprobenfunktion*.
- Im Zusammenhang mit der Schätzung von Parametern oder anderen Modellcharakteristiken nennt man $\varphi(X_1, \ldots, X_n)$ *Schätzer*. Beim Testen von Hypothesen spricht man dagegen von (Test-) *Statistiken* bzw. von *Testgrößen*, vgl. Abschnitt 4.

1.2.1 Stichprobenmittel

Wir diskutieren zunächst die Frage, wie der Erwartungswert $\mu = \mathbb{E}\, X_i$ der Stichprobenvariablen $X_1, \ldots, X_n$ aus den beobachteten Daten $x_1, \ldots, x_n$ bestimmt werden kann.

- Hierfür betrachten wir die Stichprobenfunktion $\varphi : \mathbb{R}^n \to \mathbb{R}$ mit $\varphi(x_1, \ldots, x_n) = n^{-1} \sum_{i=1}^n x_i$.
- D.h., wir betrachten das arithmetische Mittel

$$\overline{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

  der Stichprobenwerte $x_1, \ldots, x_n$.
- Die Zahl $\overline{x}_n$ wird *Stichprobenmittel* der (konkreten) Stichprobe $(x_1, \ldots, x_n)$ genannt.
- Außerdem betrachten wir das arithmetische Mittel der Stichprobenvariablen $X_1, \ldots, X_n$, dessen Eigenschaften bereits in Abschnitt WR–5.2 im Zusammenhang mit dem Gesetz der großen Zahlen untersucht worden sind.

Definiton Die Zufallsvariable

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \tag{2}$$

heißt *Stichprobenmittel* der Zufallsstichprobe $(X_1, \ldots, X_n)$.

Der Erwartungswert und die Varianz von $\overline{X}_n$ lassen sich wie folgt darstellen.

Theorem 1.1 *Es gilt*

$$\mathbb{E}\, \overline{X}_n = \mu \tag{3}$$

und

$$\operatorname{Var} \overline{X}_n = \frac{\sigma^2}{n}\,, \tag{4}$$

wobei $\mu = \mathbb{E}\, X_i$ und $\sigma^2 = \operatorname{Var} X_i$ den Erwartungswert bzw. die Varianz der Stichprobenvariablen $X_1, \ldots, X_n$ bezeichnen.

Beweis

- Die Formeln (3) und (4) ergeben sich wie folgt aus den Theoremen WR–4.4 bzw. WR–4.10.
- Wegen der Linearität des Erwartungswertes gilt nämlich (vgl. Theorem WR–4.4)

$$\begin{aligned}
\mathbb{E}\, \overline{X}_n &= \mathbb{E}\Bigl(\frac{1}{n} \sum_{i=1}^n X_i\Bigr) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\, X_i \\
&= \frac{1}{n} \sum_{i=1}^n \mu = \mu\,.
\end{aligned}$$

- Aus Theorem WR–4.10 über die Varianz der Summe von unabhängigen Zufallsvariablen ergibt sich, dass

$$\begin{aligned}\operatorname{Var}\overline{X}_n &= \frac{1}{n^2}\operatorname{Var}\Bigl(\sum_{i=1}^n X_i\Bigr)\\ &= \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}X_i\\ &= \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}\,.\end{aligned}$$

□

Beachte

- Weil das Stichprobenmittel $\overline{X}_n$ den Erwartungswert μ hat (vgl. (3)), kann man $\overline{X}_n$ als einen geeigneten „Schätzer” der (im allgemeinen unbekannten) Modellcharakteristik μ ansehen.
- Wegen (3) sagt man, dass bei der Schätzung von μ durch $\overline{X}_n$ kein „systematischer Fehler” begangen wird.
- Der Schätzer $\overline{X}_n$ kann dennoch sehr ungenau sein, wobei man den in (4) gegebenen Wert σ^2/n als Kennzahl für die Schätzgenauigkeit von $\overline{X}_n$ auffassen kann.
- Dabei bedeutet (4), dass die Schätzgenauigkeit mit wachsendem Stichprobenumfang n verbessert wird.
- Es ist jedoch zu beachten, dass σ^2 im allgemeinen ebenfalls unbekannt ist.
- Um eine Vorstellung über die Größenordnung der Schätzgenauigkeit von $\overline{X}_n$ bei vorgegebenem Stichprobenumfang n zu erlangen, muss deshalb auch σ^2 aus den beobachteten Daten $x_1,\ldots,x_n$ geschätzt werden.
- Diese Fragestellung werden wir in Abschnitt 1.2.2 diskutieren.

Neben den Formeln (3) und (4) für Erwartungswert und Varianz des Stichprobenmittels $\overline{X}_n$ sind noch weitere Aussagen über die Verteilung von $\overline{X}_n$ von Interesse bzw. über deren asymptotisches Verhalten für große n.

Theorem 1.2 *Es gilt*

$$P\Bigl(\lim_{n\to\infty}\overline{X}_n=\mu\Bigr)=1 \tag{5}$$

und

$$\lim_{n\to\infty}P\Bigl(\sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma}\le x\Bigr)=\Phi(x) \tag{6}$$

für jedes $x\in\mathbb{R}$, wobei $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung ist.

Der *Beweis* von Theorem 1.2 ergibt sich unmittelbar aus dem starken Gesetz der großen Zahlen bzw. aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen (vgl. die Theoreme WR–5.15 bzw. WR–5.16).

Beachte

- Theorem 1.2 bietet eine Möglichkeit, um die Wahrscheinlichkeit des Ereignisses $\{|\overline{X}_n-\mu|>\varepsilon\}$ zu bestimmen, dass das Stichprobenmittel $\overline{X}_n$ um mehr als einen vorgegebenen Schwellenwert $\varepsilon>0$ von dem zu schätzenden Wert μ abweicht.

- Aus (6) ergibt sich hierfür die folgende Näherungsformel

$$P(|\overline{X}_n - \mu| > \varepsilon) \approx 2\Bigl(1 - \Phi\Bigl(\frac{\varepsilon\sqrt{n}}{\sigma}\Bigr)\Bigr) \tag{7}$$

  für große n, denn es gilt

$$\begin{aligned} P(|\overline{X}_n - \mu| > \varepsilon) &= P\Bigl(\sqrt{n}\,\Bigl|\frac{\overline{X}_n - \mu}{\sigma}\Bigr| > \frac{\varepsilon\sqrt{n}}{\sigma}\Bigr) \\ &= P\Bigl(\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} < -\frac{\varepsilon\sqrt{n}}{\sigma}\Bigr) + P\Bigl(\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} > \frac{\varepsilon\sqrt{n}}{\sigma}\Bigr) \\ &\overset{(6)}{\approx} \Phi\Bigl(-\frac{\varepsilon\sqrt{n}}{\sigma}\Bigr) + 1 - \Phi\Bigl(\frac{\varepsilon\sqrt{n}}{\sigma}\Bigr) \\ &= 2\Bigl(1 - \Phi\Bigl(\frac{\varepsilon\sqrt{n}}{\sigma}\Bigr)\Bigr)\,. \end{aligned}$$

- Falls σ^2 unbekannt ist, dann kann man anstelle der Näherungsformel (7) ein (zufälliges) Intervall angeben, in dem die Abweichung $\overline{X}_n - \mu$ des Stichprobenmittels $\overline{X}_n$ von dem zu schätzenden Wert μ mit einer (näherungsweise) vorgegebenen Wahrscheinlichkeit liegt, vgl. die Anmerkungen am Ende von Abschnitt 1.2.2.

1.2.2 Stichprobenvarianz

Wir untersuchen nun die Frage, wie die Varianz $\sigma^2 = \operatorname{Var} X_i$ der Stichprobenvariablen $X_1, \ldots, X_n$ aus den beobachteten Daten $x_1, \ldots, x_n$ bestimmt werden kann. Dabei gehen wir ähnlich wie in Abschnitt 1.2.1 vor.

- Wir betrachten die Stichprobenfunktion $\varphi : \mathbb{R}^n \to \mathbb{R}$ mit

$$\varphi(x_1, \ldots, x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2\,, \tag{8}$$

  wobei $\overline{x}_n$ in (1) gegeben ist.

- Die in (8) eingeführte Größe wird *Stichprobenvarianz* der (konkreten) Stichprobe $(x_1, \ldots, x_n)$ genannt und mit s_n^2 bezeichnet.

Definition Die Zufallsvariable

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\,, \tag{9}$$

heißt *Stichprobenvarianz* der Zufallsstichprobe $(X_1, \ldots, X_n)$.

Der Erwartungswert und die Varianz von S_n^2 lassen sich wie folgt darstellen.

Theorem 1.3 *Es gilt*

$$\mathbb{E}\,(S_n^2) = \sigma^2\,. \tag{10}$$

Falls $\mathbb{E}\,(X_i^4) < \infty$ für $i = 1, \ldots, n$, dann gilt außerdem

$$\operatorname{Var}(S_n^2) = \frac{1}{n}\Bigl(\mu_4' - \frac{n-3}{n-1}\,\sigma^4\Bigr)\,, \tag{11}$$

wobei $\mu_4' = \mathbb{E}\bigl((X_i - \mu)^4\bigr)$ und $\sigma^4 = (\operatorname{Var} X_i)^2$ das 4-te zentrale Moment bzw. die quadrierte Varianz der Stichprobenvariablen $X_1, \ldots, X_n$ bezeichnen.

Beweis

- Aus der Definitionsgleichung (9) von S_n^2 ergibt sich, dass

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n \Bigl((X_i-\mu) - \frac{1}{n}\sum_{j=1}^n (X_j-\mu)\Bigr)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n (X_i' - \overline{X_n'})^2\,, \end{aligned}$$

d.h.,

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i' - \overline{X_n'})^2\,, \tag{12}$$

wobei $X_i' = X_i - \mu$ mit $\mathbb{E}\, X_i' = 0$.

- Weil die Formel (12) die gleiche Form hat wie die Definitionsgleichung (9) von S_n^2, können (und werden) wir o.B.d.A. voraussetzen, dass

$$\mathbb{E}\, X_i = 0\,. \tag{13}$$

- Hieraus folgt insbesondere, dass

$$\mathbb{E}\, \overline{X}_n = 0\,. \tag{14}$$

- Außerdem ergibt sich aus der Definitionsgleichung (9) von S_n^2, dass

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n (X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) \\ &= \frac{1}{n-1}\Bigl(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\Bigr)\,, \end{aligned}$$

d.h.,

$$S_n^2 = \frac{1}{n-1}\Bigl(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\Bigr)\,. \tag{15}$$

- Wegen der Linearität des Erwartungswertes ergibt sich somit bei Berücksichtigung von (13) und (14), dass

$$\begin{aligned} \mathbb{E}\, S_n^2 &= \frac{1}{n-1}\Bigl(\sum_{i=1}^n \mathbb{E}\,(X_i^2) - n\mathbb{E}\,(\overline{X}_n^2)\Bigr) \\ &= \frac{1}{n-1}\Bigl(\sum_{i=1}^n \mathrm{Var}\, X_i - n\mathrm{Var}\, \overline{X}_n\Bigr) \\ &= \frac{1}{n-1}\Bigl(n\sigma^2 - n\,\frac{\sigma^2}{n}\Bigr) = \sigma^2\,, \end{aligned}$$

wobei sich die vorletzte Gleichheit aus (4) ergibt.

- Damit ist (10) bewiesen.
- Um die Gültigkeit von (11) zu zeigen, berechnen wir zunächst den Erwartungswert $\mathbb{E}\,(S_n^4)$ der quadrierten Stichprobenvarianz $S_n^4 = (S_n^2)^2$.

- Durch Quadrierung beider Seiten von (15) ergibt sich, dass
$$(n-1)^2 S_n^4 = \Bigl(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\Bigr)^2 = \Bigl(\sum_{i=1}^n X_i^2\Bigr)^2 - 2n\overline{X}_n^2 \sum_{i=1}^n X_i^2 + n^2\overline{X}_n^4 .$$
- Für den Erwartungswert des ersten Summanden dieser Summe gilt
$$\begin{aligned}\mathbb{E}\Bigl(\Bigl(\sum_{i=1}^n X_i^2\Bigr)^2\Bigr) &= \mathbb{E}\Bigl(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2\Bigr)\\ &= \mathbb{E}\Bigl(\sum_{i=1}^n X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\Bigr)\\ &= \sum_{i=1}^n \mathbb{E}\,(X_i^4) + \sum_{i\neq j} \mathbb{E}\,(X_i^2 X_j^2)\\ &= \sum_{i=1}^n \mathbb{E}\,(X_i^4) + \sum_{i\neq j} \mathbb{E}\,(X_i^2)\mathbb{E}\,(X_j^2)\\ &= n\mu_4 + n(n-1)\sigma^4 ,\end{aligned}$$
d.h.,
$$\mathbb{E}\Bigl(\Bigl(\sum_{i=1}^n X_i^2\Bigr)^2\Bigr) = n\mu_4 + n(n-1)\sigma^4 . \tag{16}$$
- Für den Erwartungswert des zweiten Summanden gilt
$$\begin{aligned}\mathbb{E}\Bigl(\overline{X}_n^2 \sum_{i=1}^n X_i^2\Bigr) &= \mathbb{E}\Bigl(\Bigl(\frac{1}{n}\sum_{i=1}^n X_i\Bigr)^2 \sum_{j=1}^n X_j^2\Bigr)\\ &= \frac{1}{n^2}\mathbb{E}\Bigl(\Bigl(\sum_{i=1}^n X_i^2 + \sum_{i\neq j} X_i X_j\Bigr)\sum_{j=1}^n X_j^2\Bigr)\\ &= \frac{1}{n^2}\mathbb{E}\Bigl(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2\Bigr) + \frac{1}{n^2}\mathbb{E}\Bigl(\sum_{i\neq j} X_i X_j \sum_{k=1}^n X_k^2\Bigr)\\ &= \frac{1}{n^2}\mathbb{E}\Bigl(\Bigl(\sum_{i=1}^n X_i^2\Bigr)^2\Bigr) + \frac{1}{n^2}\sum_{i\neq j}\sum_{k=1}^n \underbrace{\mathbb{E}\Bigl(X_i X_j X_k^2\Bigr)}_{=0}\\ &\overset{(16)}{=} \frac{1}{n^2}\bigl(n\mu_4 + n(n-1)\sigma^4\bigr)\\ &= \frac{\mu_4 + (n-1)\sigma^4}{n} ,\end{aligned}$$
wobei die vorletzte Gleichheit aus (16), aus der Unabhängigkeit der Stichprobenvariablen $X_1,\ldots,X_n$ und aus der Annahme folgt, dass $\mathbb{E}\,X_i = 0$ für jedes $i = 1,\ldots,n$.
- Schließlich ergibt sich auf ähnliche Weise für den Erwartungswert des dritten Summanden
$$\begin{aligned}\mathbb{E}\Bigl(\overline{X}_n^4\Bigr) &= \mathbb{E}\Bigl(\Bigl(\frac{1}{n}\sum_{i=1}^n X_i\Bigr)^2\Bigl(\frac{1}{n}\sum_{j=1}^n X_j\Bigr)^2\Bigr)\\ &= \frac{1}{n^4}\,\mathbb{E}\Bigl(\Bigl(\sum_{k=1}^n X_k^2 + \sum_{i\neq j} X_i X_j\Bigr)\Bigl(\sum_{r=1}^n X_r^2 + \sum_{s\neq t} X_s X_t\Bigr)\Bigr)\\ &= \frac{1}{n^4}\,\mathbb{E}\Bigl(\sum_{k=1}^n X_k^4 + \sum_{k\neq r} X_k^2 X_r^2 + 2\sum_{i\neq j} X_i^2 X_j^2\Bigr)\\ &= \frac{1}{n^4}\,\bigl(n\mu_4 + 3n(n-1)\sigma^4\bigr) .\end{aligned}$$

- Insgesamt ergibt sich also, dass

$$\begin{aligned}\mathbb{E}\,(S_n^4) &= \frac{n\mu_4}{(n-1)^2}-\frac{2\mu_4}{(n-1)^2}+\frac{\mu_4}{n(n-1)^2}+\sigma^4\Big(\frac{n}{n-1}-\frac{2}{n-1}+\frac{3}{n(n-1)}\Big)\\ &= \frac{(n^2-2n+1)\mu_4}{n(n-1)^2}+\sigma^4\,\frac{n^2-2n+3}{n(n-1)}\\ &= \frac{1}{n}\,\mu_4+\frac{n^2-2n+3}{n(n-1)}\,\sigma^4\,.\end{aligned}$$

- In Theorem WR–4.6 hatten wir gezeigt, dass

$$\mathrm{Var}\,Z=\mathbb{E}\,(Z^2)-(\mathbb{E}\,Z)^2$$

für jede Zufallsvariable Z mit $\mathbb{E}\,(Z^2)<\infty$.

- Hieraus und aus (10) folgt nun, dass

$$\begin{aligned}\mathrm{Var}\,(S_n^2) &= \mathbb{E}\,(S_n^4)-\big(\mathbb{E}\,(S_n^2)\big)^2\\ &= \frac{1}{n}\,\mu_4+\frac{n^2-2n+3}{n(n-1)}\,\sigma^4-\sigma^4\\ &= \frac{1}{n}\Big(\mu_4-\frac{n-3}{n-1}\,\sigma^4\Big)\,.\end{aligned}$$

□

Beachte

- Weil die Stichprobenvarianz S_n^2 den Erwartungswert σ^2 hat (vgl. (10)), kann S_n^2 als ein geeigneter Schätzer der (im allgemeinen unbekannten) Modellcharakteristik σ^2 angesehen werden.
- Wegen (10) sagt man, dass bei der Schätzung von σ^2 durch S_n^2 kein systematischer Fehler begangen wird.
- Darüber hinaus bedeutet (11), dass die Schätzgenauigkeit mit wachsendem Stichprobenumfang n verbessert wird, falls das 4-te zentrale Moment der Stichprobenvariablen endlich ist.

Neben den Formeln (10) und (11) für Erwartungswert und Varianz des Stichprobenvarianz S_n^2 sind erneut weitere Aussagen über die Verteilung von S_n^2 bzw. über deren asymptotisches Verhalten für große n von Interesse.

Theorem 1.4 *Es gilt*

$$P\Big(\lim_{n\to\infty}S_n^2=\sigma^2\Big)=1\,. \tag{17}$$

Falls $\mathbb{E}\,(X_i^4)<\infty$ *für* $i=1,\ldots,n$, *dann gilt außerdem*

$$\lim_{n\to\infty}P\Big(\sqrt{n}\,\frac{S_n^2-\sigma^2}{\sqrt{\mu_4'-\sigma^4}}\le x\Big)=\Phi(x) \tag{18}$$

für jedes $x\in\mathbb{R}$, *wobei* $\mu_4'=\mathbb{E}\,\big((X_i-\mu)^4\big)$ *und* $\sigma^4=(\mathrm{Var}\,X_i)^2$ *das 4-te zentrale Moment bzw. die quadrierte Varianz der Stichprobenvariablen* $X_1,\ldots,X_n$ *bezeichnen und* $\Phi(x)$ *die Verteilungsfunktion der Standardnormalverteilung ist.*

Beweis

- Aus der Voraussetzung, dass die Stichprobenvariablen $X_1, \ldots, X_n$ unabhängig und identisch verteilt sind, ergibt sich, dass auch die Zufallsvariablen $X_1^2, \ldots, X_n^2$ unabhängig und identisch verteilt sind, vgl. Theorem WR–3.18.
- Deshalb ergibt sich aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15), dass mit Wahrscheinlichkeit 1
$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^2 = \mathbb{E}\,(X_1^2)\,.$$
- Außerdem ergibt sich aus (5), dass mit Wahrscheinlichkeit 1
$$\lim_{n\to\infty} \overline{X}_n^2 = \lim_{n\to\infty}\Bigl(\frac{1}{n}\sum_{i=1}^n X_i\Bigr)^2 = \Bigl(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i\Bigr)^2 = (\mathbb{E}\,X_1)^2\,.$$
- Hieraus und aus (15) ergibt sich nun, dass mit Wahrscheinlichkeit 1
$$\begin{aligned}\lim_{n\to\infty} S_n^2 &= \lim_{n\to\infty}\frac{1}{n-1}\Bigl(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\Bigr)\\ &= \lim_{n\to\infty}\frac{n}{n-1}\Bigl(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2\Bigr)\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i^2 - \lim_{n\to\infty}\overline{X}_n^2\\ &= \mathbb{E}\,(X_1^2) - (\mathbb{E}\,X_1)^2 = \sigma^2\,.\end{aligned}$$
- Damit ist (17) bewiesen.
- Um die Gültigkeit von (18) zu zeigen, benutzen wir die Darstellungsformeln (12) und (15) der Stichprobenvarianz S_n^2.
- Dabei ergibt sich, dass
$$S_n^2 = \frac{1}{n-1}\Bigl(\sum_{i=1}^n (X_i')^2 - n(\overline{X'}_n)^2\Bigr) = \frac{n}{n-1}\Bigl(\frac{1}{n}\sum_{i=1}^n (X_i')^2 - (\overline{X'}_n)^2\Bigr)\,,$$
wobei $X_i' = X_i - \mu$.
- Aus Formel (5) in Theorem 1.2 ergibt sich, dass $(\overline{X'}_n)^2 \xrightarrow{\text{f.s.}} 0$.
- Deshalb ergibt sich aus dem Satz von Slutsky für die Addition bzw. Multiplikation (vgl. die Theoreme WR–5.9 und WR–5.11), dass für jedes $x \in \mathbb{R}$
$$\begin{aligned}\lim_{n\to\infty} P\Bigl(\sqrt{n}\,\frac{S_n^2-\sigma^2}{\sqrt{\mu_4'-\sigma^4}} \le x\Bigr) &= \lim_{n\to\infty} P\Bigl(\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^n (X_i')^2 - \sigma^2}{\sqrt{\mu_4'-\sigma^4}} \le x\Bigr)\\ &= \lim_{n\to\infty} P\Bigl(\frac{\sum_{i=1}^n (X_i')^2 - n\sigma^2}{\sqrt{n(\mu_4'-\sigma^4)}} \le x\Bigr)\\ &= \lim_{n\to\infty} P\Bigl(\frac{\sum_{i=1}^n (X_i')^2 - n\mathbb{E}\,\bigl((X_1')^2\bigr)}{\sqrt{n\mathrm{Var}\,\bigl((X_1')^2\bigr)}} \le x\Bigr)\\ &= \Phi(x)\,,\end{aligned}$$

wobei sich die letzte Gleichheit aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen ergibt (vgl. Theorem WR–5.16). □

Korollar 1.1 *Es gilt*

$$\lim_{n\to\infty} P\Big(\sqrt{n}\,\frac{\overline{X}_n-\mu}{S_n}\le x\Big)=\Phi(x) \tag{19}$$

für jedes $x\in\mathbb{R}$, wobei $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung ist.

Beweis

- Aus (17) in Theorem 1.4 ergibt sich, dass $S_n \xrightarrow{\text{f.s.}} \sigma$.
- Aus (6) in Theorem 1.2 ergibt sich nun die Behauptung mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11). □

Beachte

- Korollar 1.1 bietet die Möglichkeit, ein (zufälliges) Intervall anzugeben, in dem die Abweichung $\overline{X}_n-\mu$ des Stichprobenmittels $\overline{X}_n$ von dem zu schätzenden Wert μ mit einer (näherungsweise) vorgegebenen Wahrscheinlichkeit liegt, vgl. Formel (20).
- Sei $\alpha\in(0,1)$, und sei $z_\alpha\in\mathbb{R}$ die (eindeutig bestimmte) Lösung der Gleichung $\Phi(z_\alpha)=\alpha$.
- Dann heißt z_α das *α-Quantil* der Standardnormalverteilung $\mathrm{N}(0,1)$.
- Für $\alpha\ge 0.5$ kann man das α-Quantil der $\mathrm{N}(0,1)$-Verteilung aus Tabelle 1 entnehmen, vgl. Abschnitt 6.
- Aus (19) ergibt sich nun, dass

$$P\Big(\frac{-z_{1-\alpha/2}S_n}{\sqrt{n}}<\overline{X}_n-\mu<\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\Big)\approx 1-\alpha\,, \tag{20}$$

denn

$$\begin{aligned}
P\Big(\frac{-z_{1-\alpha/2}S_n}{\sqrt{n}}<\overline{X}_n-\mu<\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\Big) &= P\Big(-z_{1-\alpha/2}<\sqrt{n}\,\frac{\overline{X}_n-\mu}{S_n}<z_{1-\alpha/2}\Big)\\
&= P\Big(\sqrt{n}\,\frac{\overline{X}_n-\mu}{S_n}<z_{1-\alpha/2}\Big)-P\Big(\sqrt{n}\,\frac{\overline{X}_n-\mu}{S_n}\le -z_{1-\alpha/2}\Big)\\
&\overset{(19)}{\approx} \Phi(z_{1-\alpha/2})-\Phi(-z_{1-\alpha/2})\\
&= \Phi(z_{1-\alpha/2})-\Phi(z_{\alpha/2})=1-\alpha\,,
\end{aligned}$$

wobei in der vorletzten Gleichheit die Symmetrieeigenschaft

$$z_{\alpha/2}=-z_{1-\alpha/2} \tag{21}$$

der Quantile der $\mathrm{N}(0,1)$-Verteilung genutzt wurde.

1.3 Beispiel: Normalverteilte Stichprobenvariablen

- In Abschnitt 1.3.3 werden wir voraussetzen, dass die Stichprobenvariablen $X_1,\ldots,X_n$ normalverteilt sind, d.h., $X_i\sim\mathrm{N}(\mu,\sigma^2)$ für jedes $i\in\{1,\ldots,n\}$, wobei $\mu\in\mathbb{R}$ und $\sigma^2>0$.
- Anstelle der Näherungsformeln (7) und (20) können wir dann „exakte" Formeln für die Wahrscheinlichkeit herleiten, dass die Abweichung $\overline{X}_n-\mu$ des Stichprobenmittels $\overline{X}_n$ von dem zu schätzenden Wert μ in einem bestimmten Intervall liegt.
- Hierfür betrachten wir zunächst eine Klasse von Verteilungen, die mit der Gammafunktion eng zusammenhängt.

1.3.1 Gammaverteilung und χ^2-Verteilung

- Sei $\Gamma : (0,\infty) \to (0,\infty)$ die *Gammafunktion* mit

$$\Gamma(p) = \int\limits_0^\infty e^{-y}\, y^{p-1}\, dy\,, \qquad p > 0\,. \tag{22}$$

- Durch partielle Integration ergibt sich, dass für jedes $p > 0$

$$\Gamma(p+1) = \int\limits_0^\infty e^{-y}\, y^{p}\, dy = \Bigl[-e^{-y}\, y^p\Bigr]_0^\infty + p\int\limits_0^\infty e^{-y}\, y^{p-1}\, dy\,,$$

- Also gilt für jedes $p > 0$

$$\Gamma(p+1) = p\Gamma(p)\,. \tag{23}$$

- Weil

$$\Gamma(1) = \int\limits_0^\infty e^{-y}\, dy = \Bigl[-e^{-y}\Bigr]_0^\infty = 1\,,$$

ergibt sich aus (23), dass $\Gamma(n+1) = n!$ für jedes $n \in \mathbb{N}$.

Beachte

- Mit der Substitution $y' = b^{-1}y$, wobei $b > 0$, geht (22) über in

$$\Gamma(p) = b^p \int\limits_0^\infty e^{-by}\, y^{p-1}\, dy$$

bzw.

$$1 = \int\limits_0^\infty \frac{b^p}{\Gamma(p)}\, e^{-by}\, y^{p-1}\, dy\,, \tag{24}$$

wobei der Integrand als Wahrscheinlichkeitsdichte aufgefasst werden kann.

- Dies führt zu der folgenden Begriffsbildung.

Definition Man sagt, dass die Zufallsvariable $Y : \Omega \to \mathbb{R}$ *gammaverteilt* ist mit den Parametern $b > 0$ und $p > 0$, wenn Y absolutstetig ist und wenn die Dichte $f_Y : \mathbb{R} \to [0,\infty)$ von Y gegeben ist durch

$$f_Y(y) = \begin{cases} \dfrac{b^p}{\Gamma(p)}\, e^{-by}\, y^{p-1}\,, & \text{falls } y > 0,\\[2ex] 0\,, & \text{falls } y \le 0. \end{cases} \tag{25}$$

Schreibweise: $Y \sim \Gamma(b,p)$.

Theorem 1.5 *Die Zufallsvariable $Y : \Omega \to \mathbb{R}$ sei gammaverteilt mit den Parametern $b > 0$ und $p > 0$. Für die momenterzeugende Funktion $\psi_Y : (-\infty, b) \to \mathbb{R}$ mit $\psi_Y(t) = \mathbb{E}\, e^{tY}$ bzw. für die charakteristische Funktion $\varphi_Y : \mathbb{R} \to \mathbb{C}$ mit $\varphi_Y(t) = \mathbb{E}\, e^{\mathrm{i}\, tY}$ gilt dann*

$$\psi_Y(t) = \frac{1}{\Bigl(1-\frac{t}{b}\Bigr)^p}\,, \quad \forall\, t \in (-\infty, b) \qquad \textit{bzw.} \qquad \varphi_Y(t) = \frac{1}{\Bigl(1-\frac{\mathrm{i}\, t}{b}\Bigr)^p}\,, \quad \forall\, t \in \mathbb{R}\,. \tag{26}$$

Außerdem gilt für jedes $k \in \mathbb{N}$

$$\mathbb{E}\,(Y^k) = \frac{p(p+1)\ldots(p+k-1)}{b^k}\,. \tag{27}$$

Beweis

- Aus der Definition der momenterzeugenden Funktion und aus (25) ergibt sich, dass

$$\psi_Y(t) = \mathbb{E}\, e^{tY} = \int_0^\infty e^{ty} f_Y(y)\, dy = \int_0^\infty \frac{b^p}{\Gamma(p)}\, e^{ty} e^{-by}\, y^{p-1}\, dy = \int_0^\infty \frac{b^p}{\Gamma(p)}\, e^{-(b-t)y}\, y^{p-1}\, dy\,.$$

- Ähnlich wie bei der Herleitung von (24) betrachten wir nun die Substitution $y = (b-t)y'$, wobei $b > 0$ und $t < b$, und erhalten die Gleichung

$$1 = \int_0^\infty \frac{(b-t)^p}{\Gamma(p)}\, e^{-(b-t)y}\, y^{p-1}\, dy\,.$$

- Hieraus folgt, dass für jedes $t < b$

$$\psi_Y(t) = \frac{b^p}{(b-t)^p} = \frac{1}{\left(1 - \dfrac{t}{b}\right)^p}\,,$$

  d.h., die in dem Gebiet $\{z = t_1 + \mathrm{i}\, t_2 : t_1 < b\}$ holomorphen Funktionen $\varphi_1(z) = \mathbb{E}\, e^{zY}$ und $\varphi_2(z) = (1 - z/b)^{-p}$ besitzen für jedes $z \in (-\infty, b)$ jeweils die gleichen Funktionswerte.
- Aus dem Identitätssatz für holomorphe Funktionen (vgl. Abschnitt 8.1 in R. Remmert (1992) *Funktionentheorie 1*, Springer, Berlin) ergibt sich nun, dass $\varphi_1(\mathrm{i}\, t) = \varphi_2(\mathrm{i}\, t)$ für jedes $t \in \mathbb{R}$ gilt.
- Damit ist (26) bewiesen.
- Aus (26) ergibt sich nun, dass $\varphi_Y(t)$ unendlich oft differenzierbar ist und dass die k-te Ableitung $\varphi_Y^{(k)}(t)$ von $\varphi_Y(t)$ gegeben ist durch

$$\varphi_Y^{(k)}(t) = \frac{p(p+1)\ldots(p+k-1)}{b^k}\, \mathrm{i}^{\,k} \frac{1}{\left(1 - \dfrac{\mathrm{i}\, t}{b}\right)^{p+k}}\,, \qquad \forall t \in \mathbb{R}.$$

- Weil außerdem $\mathbb{E}\,(|Y|^k) < \infty$ für jedes $k \in \mathbb{N}$, gilt (vgl. Theorem WR–5.21)

$$\mathbb{E}\,(Y^k) = \frac{\varphi_Y^{(k)}(0)}{\mathrm{i}^{\,k}}\,, \qquad \forall\, k \in \mathbb{N}.$$

- Hieraus folgt die Gültigkeit von (27). □

Aus Theorem 1.5 ergibt sich insbesondere, dass die Familie der Gammaverteilungen die folgende Eigenschaft der *Faltungsstabilität* besitzt.

Korollar 1.2 *Seien $Y_1, Y_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen mit $Y_1 \sim \Gamma(b, p_1)$ und $Y_2 \sim \Gamma(b, p_2)$, wobei $b, p_1, p_2 > 0$. Dann gilt $Y_1 + Y_2 \sim \Gamma(b, p_1 + p_2)$.*

Beweis

- Für die charakteristische Funktion $\varphi_{Y_1+Y_2}(t) = \mathbb{E}\, e^{\mathrm{i}\, t(Y_1+Y_2)}$ der Summe $Y_1 + Y_2$ gilt wegen der Unabhängigkeit von Y_1 und Y_2 (vgl. Theorem WR–5.18), dass

$$\varphi_{Y_1+Y_2}(t) = \varphi_{Y_1}(t)\varphi_{Y_2}(t)\,, \qquad \forall t \in \mathbb{R}.$$

- Aus (26) ergibt sich also, dass

$$\begin{aligned}\varphi_{Y_1+Y_2}(t) &= \frac{1}{\left(1-\dfrac{\mathrm{i}\,t}{b}\right)^{p_1}}\;\frac{1}{\left(1-\dfrac{\mathrm{i}\,t}{b}\right)^{p_2}}\\ &= \frac{1}{\left(1-\dfrac{\mathrm{i}\,t}{b}\right)^{p_1+p_2}}\,.\end{aligned}$$

- Die charakteristische Funktion $\varphi_{Y_1+Y_2}(t)$ der Summe Y_1+Y_2 besitzt somit die in (26) hergeleitete Form der charakteristischen Funktion der Gammaverteilung $\Gamma(b, p_1+p_2)$.
- Aus dem Eindeutigkeitssatz für charakteristische Funktionen (vgl. Korollar WR–5.5) folgt nun, dass $Y_1+Y_2 \sim \Gamma(b, p_1+p_2)$. $\square$

Beachte

- Aus der Definitionsgleichung (25) der Dichte der Gammaverteilung ergibt sich, dass die Exponentialverteilung $\mathrm{Exp}(b)$ eine spezielle Gammaverteilung ist, wobei

$$\mathrm{Exp}(b) = \Gamma(b,1)\,.$$

- Wegen Korollar 1.2 ist die Summe $Y_1+\ldots+Y_n$ von n unabhängigen und identisch verteilten Zufallsvariablen $Y_1,\ldots,Y_n$ mit $Y_i \sim \mathrm{Exp}(b)$ für jedes $i \in \{1,\ldots,n\}$ ebenfalls gammaverteilt, und zwar gilt

$$Y_1+\ldots+Y_n \sim \Gamma(b,n)\,.$$

  Ein solche Gammaverteilung heißt *Erlangverteilung* der Ordnung n.
- Außer den Klassen der Exponential- bzw. Erlangverteilungen werden noch weitere Teilklassen von Gammaverteilungen betrachtet.
- Eine solche Teilklasse von Gammaverteilungen spielt bei der Bestimmung der Verteilung der Stichprobenvarianz S_n^2 von normalverteilten Stichprobenvariablen eine wichtige Rolle.
- Es ist dies die Familie der χ^2-Verteilungen, die zu den sogenannten *statistischen Prüfverteilungen* gehören und die wie folgt definiert sind.

Definition

- Sei $r \in \mathbb{N}$ eine beliebige natürliche Zahl, und seien $X_1,\ldots,X_r : \Omega \to \mathbb{R}$ unabhängige und $\mathrm{N}(0,1)$-verteilte Zufallsvariablen.
- Dann sagt man, dass die Zufallsvariable $U_r = \sum_{i=1}^r X_i^2$ eine χ^2-*Verteilung* mit r Freiheitsgraden hat. (Schreibweise: $U_r \sim \chi^2_r$)

Theorem 1.6 *Sei $r \ge 1$ eine beliebige natürliche Zahl, und sei U_r eine χ^2-verteilte Zufallsvariable mit r Freiheitsgraden. Dann ist die Dichte von U_r gegeben durch*

$$f_{U_r}(x) = \begin{cases} \dfrac{x^{(r-2)/2}e^{-x/2}}{2^{r/2}\,\Gamma(r/2)}\,, & \textit{falls } x>0,\\ 0\,, & \textit{sonst,}\end{cases} \tag{28}$$

wobei $\Gamma(1)=1$, $\Gamma(1/2)=\sqrt{\pi}$ und $\Gamma(p+1)=p\,\Gamma(p)$.

Beweis

- Wir betrachten zunächst den Fall $r = 1$, d.h., es gelte $U_1 = X^2$ mit $X \sim \mathrm{N}(0,1)$.
- Für die Dichte $f_{U_1}(x)$ von U_1 ergibt sich dann aus Theorem WR–3.15, dass

$$f_{U_1}(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}\big(f_X(\sqrt{x}) + f_X(-\sqrt{x})\big)\,, & \text{falls } x > 0, \\ 0\,, & \text{falls } x \le 0. \end{cases}$$

- Weil

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)\,,$$

ergibt sich somit, dass

$$f_{U_1}(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi x}} \exp\Big(-\dfrac{x}{2}\Big)\,, & \text{falls } x > 0, \\ 0\,, & \text{falls } x \le 0. \end{cases} \tag{29}$$

- Damit ist (28) für den Fall $r = 1$ bewiesen.
- Ein Vergleich von Formel (29) mit der Definitionsgleichung (25) der Gammaverteilung zeigt außerdem, dass

$$U_1 = X^2 \sim \Gamma(1/2, 1/2)\,. \tag{30}$$

- Seien nun $X_1, \ldots, X_r : \Omega \to \mathbb{R}$ unabhängige und $\mathrm{N}(0,1)$-verteilte Zufallsvariablen.
- Aus (30) und aus Korollar 1.2 folgt dann, dass

$$X_1^2 + \ldots + X_r^2 \sim \Gamma(1/2, r/2)\,.$$

- Hieraus und aus der Definitionsgleichung (25) der Gammaverteilung ergibt sich nun die Gültigkeit von (28) für jedes $r \in \mathbb{N}$. □

Beachte

- Für $U_r \sim \chi^2_r$ und $\alpha \in (0,1)$ sei $\chi^2_{r,\alpha}$ die (eindeutig bestimmte) Lösung der Gleichung $F_{U_r}(\chi^2_{r,\alpha}) = \alpha$.
- Dann heißt $\chi^2_{r,\alpha}$ das *α-Quantil* der χ^2-Verteilung mit r Freiheitsgraden.
- Quantile der χ^2-Verteilung mit r Freiheitsgraden sind in Tabelle 2 gegeben, vgl. Abschnitt 6.

1.3.2 Unabhängigkeit und Transformation von Zufallsvektoren

- In Abschnitt 1.3.3 benötigen wir einige Eigenschaften von Zufallsvektoren, die wir hier lediglich erwähnen (ohne sie im einzelnen zu beweisen).
- In den Abschnitten WR–3 bzw. WR–4 der Vorlesung „Wahrscheinlichkeitsrechnung" im WS 01/02 sind solche Eigenschaften für den Fall reellwertiger Zufallsvariablen hergeleitet worden.

Definition In Verallgemeinerung des Begriffes der Unabhängigkeit von reellwertigen Zufallsvariablen, der in Abschnitt WR–3.3.5 eingeführt wurde, sagen wir, dass die Zufallsvektoren $X_1 : \Omega \to \mathbb{R}^{m_1}, \ldots, X_n : \Omega \to \mathbb{R}^{m_n}$ *unabhängig* sind, falls

$$F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = F_{X_1}(x_1)\ldots F_{X_n}(x_n) \qquad \forall x_1 \in \mathbb{R}^{m_1}, \ldots, x_n \in \mathbb{R}^{m_n}\,. \tag{31}$$

Analog zu Theorem WR–3.11 ergibt sich dann die folgende Charakterisierung der Unabhängigkeit absolutstetiger Zufallsvektoren.

Theorem 1.7 *Seien $(X_1,\ldots,X_n):\Omega\to\mathbb{R}^{m_1+\ldots+m_n}$ und damit auch $X_1:\Omega\to\mathbb{R}^{m_1},\ldots,X_n:\Omega\to\mathbb{R}^{m_n}$ absolutstetige Zufallsvektoren. Die Komponenten $X_1,\ldots,X_n$ des Vektors $(X_1,\ldots,X_n)$ sind genau dann unabhängig, wenn für fast alle $(x_1,\ldots,x_n)\in\mathbb{R}^{m_1+\ldots+m_n}$*

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)=f_{X_1}(x_1)\ldots f_{X_n}(x_n)\,. \tag{32}$$

Analog zu Theorem WR–3.18 ergibt sich der folgende Satz über die Unabhängigkeit zusammengesetzter Abbildungen.

Theorem 1.8 *Die Zufallsvektoren $X_1:\Omega\to\mathbb{R}^{m_1},\ldots,X_n:\Omega\to\mathbb{R}^{m_n}$ seien unabhängig. Für beliebige Borel-messbare Funktionen $\varphi_1:\mathbb{R}^{m_1}\to\mathbb{R}^{m_1'},\ldots,\varphi_n:\mathbb{R}^{m_n}\to\mathbb{R}^{m_n'}$ sind die Zufallsvektoren $\varphi_1(X_1),\ldots,\varphi_n(X_n)$ dann auch unabhängig.*

Schließlich gilt der folgende Transformationssatz für die Dichte von absolutstetigen Zufallsvektoren.

Theorem 1.9

- *Sei $X=(X_1,\ldots,X_n):\Omega\to\mathbb{R}^n$ ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte $f_X:\mathbb{R}^n\to[0,\infty)$, und sei $\varphi=(\varphi_1,\ldots,\varphi_n):\mathbb{R}^n\to\mathbb{R}^n$ eine Borel-messbare Abbildung mit stetigen partiellen Ableitungen $\partial\varphi_i/\partial x_j(x_1,\ldots,x_n)$.*
- *Außerdem gebe es einen n-dimensionalen Quader $B\in\mathcal{B}(\mathbb{R}^n)$ mit*

  $$\{x\in\mathbb{R}^n: f_X(x)\neq 0\}\subset B$$

  und

  $$\det\Bigl(\frac{\partial\varphi_i}{\partial x_j}(x_1,\ldots,x_n)\Bigr)\neq 0\qquad\forall (x_1,\ldots,x_n)\in B\,,$$

  so dass die Einschränkung $\varphi:B\to C$ von φ auf die Menge B eine eineindeutige Abbildung ist, wobei $C=\{\varphi(x):x\in B\}$.
- *Sei $\varphi^{-1}=(\varphi_1^{-1},\ldots,\varphi_n^{-1}):C\to B$ die Umkehrung der Abbildung $\varphi:B\to C$.*
- *Dann ist auch der Zufallsvektor $Y=\varphi(X)$ absolutstetig, und für die Dichte $f_Y(y)$ von Y gilt*

$$f_Y(y)=\begin{cases} f_X(\varphi_1^{-1}(y),\ldots,\varphi_n^{-1}(y))\,\Bigl|\det\Bigl(\dfrac{\partial\varphi_i^{-1}}{\partial y_j}(y_1,\ldots,y_n)\Bigr)\Bigr|, & \text{falls } y=(y_1,\ldots,y_n)\in C,\\[2ex] 0\,, & \text{falls } y\notin C. \end{cases} \tag{33}$$

1.3.3 Verteilung von Stichprobenmittel und Stichprobenvarianz

- Wir bestimmen nun die (gemeinsame) Verteilung des Stichprobenmittels $\overline{X}_n$ und der Stichprobenvarianz S_n^2 bei normalverteilten Stichprobenvariablen $X_1,\ldots,X_n$.
- Zunächst zeigen wir, dass $\overline{X}_n$ und S_n^2 unabhängig sind.

Theorem 1.10 *Sei $(X_1,\ldots,X_n)$ eine normalverteilte Zufallsstichprobe mit $X_i\sim\mathrm{N}(\mu,\sigma^2)$. Dann sind $\overline{X}_n$ und S_n^2 unabhängige Zufallsvariablen.*

Beweis

- Zur Erinnerung: Mit der Schreibweise $X_i' = X_i - \mu$ gilt
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i' + \mu \qquad \text{und} \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i' - \overline{X_n'})^2 \,. \tag{34}$$
- Wegen Theorem 1.8 können (und werden) wir deshalb o.B.d.A. voraussetzen, dass $\mu = 0$ und $\sigma^2 = 1$, d.h. $X_i \sim \mathrm{N}(0,1)$.
- Um die Unabhängigkeit von $\overline{X}_n$ und S_n^2 zu zeigen, nutzen wir die Tatsache, dass sich die Stichprobenvarianz S_n^2 wie folgt darstellen lässt:
$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\Big((X_1 - \overline{X}_n)^2 + \sum_{i=2}^n (X_i - \overline{X}_n)^2\Big) \\ &= \frac{1}{n-1}\Big(\Big(\sum_{i=2}^n (X_i - \overline{X}_n)\Big)^2 + \sum_{i=2}^n (X_i - \overline{X}_n)^2\Big)\,, \end{aligned}$$
wobei sich die letzte Gleichheit aus der Identität $\sum_{i=1}^n (X_i - \overline{X}_n) = 0$ ergibt.
- Die Stichprobenvarianz S_n^2 ist also eine Funktion des Zufallsvektors $\big(X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n\big)$, d.h., es gilt
$$S_n^2 = \widetilde{\varphi}\big(X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n\big)\,, \tag{35}$$
wobei
$$\widetilde{\varphi}(x_2, \ldots, x_n) = \frac{1}{n-1}\Big(\Big(\sum_{i=2}^n x_i\Big)^2 + \sum_{i=2}^n x_i^2\Big)\,. \tag{36}$$
- Wir zeigen nun zunächst, dass der Zufallsvektor $\big(X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n\big)$ unabhängig von $\overline{X}_n$ ist.
- Wegen der Unabhängigkeit und N(0,1)–Verteiltheit der Stichprobenvariablen $X_1, \ldots, X_n$ ist die (gemeinsame) Dichte $f_X(x_1, \ldots, x_n)$ des Zufallsvektors $X = (X_1, \ldots, X_n)$ gegeben durch
$$f_X(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}}\,\exp\Big(-\frac{1}{2}\sum_{i=1}^n x_i^2\Big)\,, \qquad \forall (x_1, \ldots, x_n)\mathbb{R}^n\,.$$
- Wir betrachten die lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ mit
$$\varphi_1(x) = \overline{x}_n\,, \quad \varphi_2(x) = x_2 - \overline{x}_n\,, \quad \ldots, \quad \varphi_n(x) = x_n - \overline{x}_n\,.$$
- Für die Umkehrabbildung $\varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ gilt dann für jedes $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$
$$\varphi_1^{-1}(y) = y_1 - \sum_{i=2}^n y_i\,, \quad \varphi_2^{-1}(y) = y_1 + y_2\,, \quad \ldots, \quad \varphi_n^{-1}(y) = y_1 + y_n\,.$$
- Für die Jacobi-Determinante der Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ bzw. der Umkehrabbildung $\varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ gilt
$$\det\Big(\frac{\partial \varphi_i}{\partial x_j}(x_1, \ldots, x_n)\Big) = \frac{1}{n}$$
für jedes $(x_1, \ldots, x_n) \in \mathbb{R}^n$ bzw.
$$\det\Big(\frac{\partial \varphi_i^{-1}}{\partial y_j}(y_1, \ldots, y_n)\Big) = n$$
für jedes $(y_1, \ldots, y_n) \in \mathbb{R}^n$.

- Aus Theorem 1.9 ergibt sich somit für die Dichte $f_{(\overline{X}_n, X_2-\overline{X}_n,\ldots,X_n-\overline{X}_n)}(y_1,\ldots,y_n)$ des Zufallsvektors
$$\left(\overline{X}_n, X_2-\overline{X}_n,\ldots,X_n-\overline{X}_n\right)=\varphi(X_1,\ldots,X_n)$$
die folgende Darstellungsformel.
- Für jedes $(y_1,\ldots,y_n)\in\mathbb{R}^n$ gilt
$$\begin{aligned}
f_{(\overline{X}_n, X_2-\overline{X}_n,\ldots,X_n-\overline{X}_n)}(y_1,\ldots,y_n) &= \frac{n}{(2\pi)^{n/2}}\ \exp\Bigl(-\frac{1}{2}\bigl(y_1-\sum_{i=2}^n y_i\bigr)^2\Bigr)\ \exp\Bigl(-\frac{1}{2}\sum_{i=2}^n (y_i+y_1)^2\Bigr)\\
&= \Bigl(\Bigl(\frac{n}{2\pi}\Bigr)^{1/2}\exp\Bigl(-\frac{1}{2}\,ny_1^2\Bigr)\Bigr)\ \Bigl(\Bigl(\frac{n}{(2\pi)^{n-1}}\Bigr)^{1/2}\exp\Bigl(-\frac{1}{2}\Bigl(\sum_{i=2}^n y_i^2+\Bigl(\sum_{i=2}^n y_i\Bigr)^2\Bigr)\Bigr)\Bigr).
\end{aligned}$$
- Wegen dieser Produktdarstellung der Dichte $f_{(\overline{X}_n, X_2-\overline{X}_n,\ldots,X_n-\overline{X}_n)}(y_1,\ldots,y_n)$ ergibt sich nun aus Theorem 1.7, dass $\left(X_2-\overline{X}_n,\ldots,X_n-\overline{X}_n\right)$ unabhängig von $\overline{X}_n$ ist.
- Wegen (35) ergibt sich somit aus Theorem 1.8, dass auch die Zufallsvariablen S_n^2 und $\overline{X}_n$ unabhängig sind. □

Theorem 1.11 *Sei $(X_1,\ldots,X_n)$ eine normalverteilte Zufallsstichprobe mit $X_i\sim \mathrm{N}(\mu,\sigma^2)$ für $i\in\{1,\ldots,n\}$. Dann gilt*
$$\overline{X}_n\sim \mathrm{N}(\mu,\sigma^2/n) \tag{37}$$
und
$$\frac{(n-1)S_n^2}{\sigma^2}\sim\chi^2_{n-1}\,. \tag{38}$$

Beweis

- Aus Theorem WR–3.14 (bzw. aus dessen vektorieller Version in Theorem 1.9) ergibt sich, dass
$$\frac{X_i}{n}\sim \mathrm{N}(\mu/n,\sigma^2/n^2)$$
für jedes $i\in\{1,\ldots,n\}$, vgl. auch das Beispiel in Abschnitt WR–3.4.2.
- Weil mit $X_1,\ldots,X_n$ auch die Zufallsvariablen $X_1/n,\ldots,X_n/n$ unabhängig sind, ergibt sich nun aus der Faltungsstabilität der Normalverteilung (vgl. Korollar WR–3.2), dass
$$\overline{X}_n=\frac{X_1}{n}+\ldots+\frac{X_n}{n}\sim \mathrm{N}(\mu,\sigma^2/n).$$
- Damit ist (37) bewiesen.
- Um (38) zu beweisen, betrachten wir die Identität
$$\begin{aligned}
\sum_{i=1}^n (X_i-\mu)^2 &= \sum_{i=1}^n \bigl((X_i-\overline{X}_n)+(\overline{X}_n-\mu)\bigr)^2\\
&= \sum_{i=1}^n (X_i-\overline{X}_n)^2+2(\overline{X}_n-\mu)\cdot\underbrace{\sum_{i=1}^n (X_i-\overline{X}_n)}_{=0}+n(\overline{X}_n-\mu)^2\\
&= \sum_{i=1}^n (X_i-\overline{X}_n)^2+n(\overline{X}_n-\mu)^2\,.
\end{aligned}$$

- Weil die Zufallsvariablen $\sigma^{-1}(X_1 - \mu), \ldots, \sigma^{-1}(X_n - \mu)$ unabhängig und $\mathrm{N}(0,1)$-verteilt sind, ergibt sich somit aus der Definition der χ^2-Verteilung, dass

$$\underbrace{\sum_{i=1}^{n}\Bigl(\frac{X_i - \mu}{\sigma}\Bigr)^2}_{\sim\chi^2_n} = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2} + \underbrace{\Bigl(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}\Bigr)^2}_{\sim\chi^2_1} . \tag{39}$$

- Wegen Theorem 1.10 sind die beiden Summanden RS_1 und RS_2 auf der rechten Seite dieser Gleichung unabhängig.
- Wegen Theorem WR–5.18 gilt deshalb für die charakteristische Funktion $\varphi_{\mathrm{LS}}(t)$ der linken Seite LS von (39), dass

$$\varphi_{\mathrm{LS}}(t) = \varphi_{\mathrm{RS}_1}(t)\varphi_{\mathrm{RS}_2}(t)$$

  für jedes $t \in \mathbb{R}$, wobei $\varphi_{\mathrm{RS}_i}(t)$ die charakteristische Funktion von RS_i bezeichnet.
- Aus Theorem 1.5 folgt somit, dass für jedes $t \in \mathbb{R}$

$$\begin{aligned}\varphi_{\mathrm{RS}_1}(t) &= \frac{\varphi_{\mathrm{LS}}(t)}{\varphi_{\mathrm{RS}_2}(t)}\\ &= \frac{1}{\bigl(1-2\mathrm{i}\,t\bigr)^{(n-1)/2}} .\end{aligned}$$

- Die erneute Anwendung von Theorem 1.5 und des Eindeutigkeitssatzes für charakteristische Funktionen (vgl. Korollar WR–5.5) ergibt nun, dass

$$\frac{(n-1)S_n^2}{\sigma^2} = \mathrm{RS}_1 \sim \chi^2_{n-1} .$$

□

1.3.4 t-Verteilung

Wir führen nun eine weitere Klasse von statistischen Prüfverteilungen ein, die wie folgt definiert sind.

Definition Sei $r \in \mathbb{N}$ eine beliebige natürliche Zahl, und seien X und U_r unabhängige Zufallsvariablen mit $X \sim \mathrm{N}(0,1)$ und $U_r \sim \chi^2_r$. Dann sagt man, dass die Zufallsvariable

$$V_r = X \Big/ \sqrt{\frac{U_r}{r}} \tag{40}$$

t-*verteilt* ist mit r Freiheitsgraden. (Schreibweise: $V_r \sim \mathrm{t}_r$)

Theorem 1.12 *Sei $V_r \sim \mathrm{t}_r$. Für die Dichte $f_{V_r}(v)$ von V_r gilt dann*

$$f_{V_r}(v) = \frac{\Gamma((r+1)/2)}{\Gamma(r/2)}\,\frac{1}{\sqrt{r\pi}\,(1+v^2/r)^{(r+1)/2}} \tag{41}$$

für jedes $v \in \mathbb{R}$.

Beweis

- Seien X und U_r unabhängige Zufallsvariablen mit $X \sim \mathrm{N}(0,1)$ und $U_r \sim \chi^2_r$.

- Für die gemeinsame Dichte $f_{(X,U_r)}(x,u)$ von (X,U_r) gilt dann:

$$f_{(X,U_r)}(x,u) = \frac{1}{\sqrt{2\pi}} \exp\Bigl(-\frac{x^2}{2}\Bigr) \frac{u^{(r-2)/2} e^{-u/2}}{2^{r/2}\,\Gamma(r/2)} \tag{42}$$

für beliebige $x \in \mathbb{R}$ und $u > 0$.
- Wir betrachten die Abbildung $(x,u) \to (v,w)$ mit

$$v = \frac{x}{\sqrt{u/r}}\,, \qquad w = u\,.$$

- Für die Jacobi-Determinante der Umkehrabbildung gilt $(w/r)^{1/2}$.
- Aus Theorem 1.9 ergibt sich somit für die (Rand-) Dichte $f_{V_r}(v)$ von V_r, dass

$$\begin{aligned} f_{V_r}(v) &= \int\limits_0^\infty f_{(X,U_r)}\bigl(v(w/r)^{1/2}, w\bigr)\,(w/r)^{1/2}\,dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2}\,\Gamma(r/2)} \int\limits_0^\infty \exp\Bigl(-\frac{v^2 w}{2r}\Bigr) w^{(r-2)/2} \exp\Bigl(-\frac{w}{2}\Bigr)\Bigl(\frac{w}{r}\Bigr)^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2}\,r^{1/2}\,\Gamma(r/2)} \int\limits_0^\infty \exp\Bigl(-\frac{(1+v^2/r)w}{2}\Bigr) w^{((r+1)/2)-1}\,dw\,. \end{aligned}$$

- Für das Integral ergibt sich aus der Darstellungsformel (24) der Gammafunktion, dass

$$\int\limits_0^\infty \exp\Bigl(-\frac{(1+v^2/r)w}{2}\Bigr) w^{((r+1)/2)-1}\,dw = \frac{\Gamma((r+1)/2)}{\Bigl(\frac{(1+v^2/r)}{2}\Bigr)^{(r+1)/2}}\,.$$

- Hieraus folgt, dass für jedes $v \in \mathbb{R}$

$$f_{V_r}(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2}\,r^{1/2}\,\Gamma(r/2)} \frac{\Gamma((r+1)/2)}{\Bigl(\frac{(1+v^2/r)}{2}\Bigr)^{(r+1)/2}}\,.$$

$\square$

Beachte

- Sei $V_r \sim \mathrm{t}_r$. Für $\alpha \in (0,1)$ wird dann die (eindeutige) Lösung $t_{r,\alpha}$ der Gleichung $F_{V_r}(t_{r,\alpha}) = \alpha$ das α-*Quantil* der t-Verteilung mit r Freiheitsgraden genannt, vgl. Tabelle 3 in Abschnitt 6.
- Analog zu der Symmetrieeigenschaft $z_\alpha = -z_{1-\alpha}$ der Quantile z_α der Standardnormalverteilung gilt auch

$$t_{r,\alpha} = -t_{r,1-\alpha} \tag{43}$$

für beliebige $r \in \mathbb{N}$ und $\alpha \in (0,1)$, weil die in (41) gegebene Dichte der t-Verteilung eine bezüglich des Nullpunktes symmetrische Funktion ist.

Theorem 1.13 *Sei $(X_1,\ldots,X_n)$ eine normalverteilte Zufallsstichprobe mit $X_i \sim \mathrm{N}(\mu,\sigma^2)$ für $i \in \{1,\ldots,n\}$. Dann gilt*

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}\,. \tag{44}$$

Beweis

- In Theorem 1.10 hatten wir gezeigt, dass die Zufallsvariablen $\overline{X}_n$ und S_n^2 unabhängig sind.
- Aus Theorem WR–3.18 (bzw. aus dessen vektorieller Version in Theorem 1.8) ergibt sich somit, dass auch die Zufallsvariablen
$$\sqrt{n}\Big(\frac{\overline{X}_n-\mu}{\sigma}\Big) \qquad \text{und} \qquad \sqrt{\frac{1}{n-1}\frac{(n-1)S_n^2}{\sigma^2}}$$
unabhängig sind.
- Außerdem ergibt sich aus Theorem 1.11, dass
$$\sqrt{n}\,\Big(\frac{\overline{X}_n-\mu}{\sigma}\Big) \sim \mathrm{N}(0,1)$$
und
$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}\,.$$
- Deshalb ergibt sich unmittelbar aus der Definition der t-Verteilung, dass
$$\frac{\sqrt{n}\,(\overline{X}_n-\mu)}{S_n} = \frac{\sqrt{n}\,\dfrac{\overline{X}_n-\mu}{\sigma}}{\sqrt{\dfrac{1}{n-1}\dfrac{(n-1)S_n^2}{\sigma^2}}} \sim \mathrm{t}_{n-1}\,.$$
□

Beachte

- Theorem 1.13 bietet die Möglichkeit, ein (zufälliges) Intervall anzugeben, in dem die Abweichung $\overline{X}_n-\mu$ des Stichprobenmittels $\overline{X}_n$ von dem zu schätzenden Wert μ mit einer vorgegebenen Wahrscheinlichkeit liegt.
- Von besonderer Bedeutung ist dabei die Tatsache, dass die Kenntnis der (im allgemeinen unbekannten) Parameter μ und σ^2 *nicht* zur Konstruktion dieses Intervalls erforderlich ist.
- Aus Theorem 1.13 ergibt sich nämlich auf die gleiche Weise wie bei der Herleitung von (20), dass
$$P\Big(\frac{-t_{n-1,1-\alpha/2}S_n}{\sqrt{n}} < \overline{X}_n-\mu < \frac{t_{n-1,1-\alpha/2}S_n}{\sqrt{n}}\Big) = 1-\alpha \tag{45}$$
für jedes $\alpha \in (0,1)$.

1.4 Ordnungsstatistiken

- Außer dem Stichprobenmittel $\overline{X}_n$ und der Stichprobenvarianz S_n^2, deren Eigenschaften in den Abschnitten 1.2 bzw. 1.3 diskutiert wurden, gibt es noch weitere Stichprobenfunktionen, die bei der statistischen Datenanalyse von Interesse sind.
- Eine solche Klasse von Stichprobenfunktionen sind die sogenannten Ordnungsstatistiken, die mit Hilfe der folgenden Borel-messbaren Abbildung $\varphi:\mathbb{R}^n\to\mathbb{R}^n$ definiert werden:
$$(x_1,\ldots,x_n)\to(x_{(1)},\ldots,x_{(n)})=\varphi(x_1,\ldots,x_n)\,, \tag{46}$$
wobei für jedes $i\in\{1,\ldots,n\}$
$$x_{(i)}=\min\big\{x_j : \#\{k : x_k\le x_j\}\ge i\big\}\,. \tag{47}$$

Beachte Die in (46) und (47) gegebene Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ ist eine Borel-messbare *Permutation* der Komponenten des Vektors $(x_1, \ldots, x_n)$, so dass

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} .$$

Definition

- Sei $(X_1, \ldots, X_n)$ eine beliebige Zufallsstichprobe, und
- für jedes $\omega \in \Omega$ sei
$$(X_{(1)}(\omega), \ldots, X_{(n)}(\omega)) = \varphi\bigl(X_1(\omega), \ldots, X_n(\omega)\bigr)$$
die in (46) und (47) gegebene (messbare) Permutation von $(X_1(\omega), \ldots, X_n(\omega))$, so dass
$$X_{(1)}(\omega) \le \ldots \le X_{(n)}(\omega) . \tag{48}$$
- Die auf diese Weise definierten Zufallsvariablen $X_{(1)}, \ldots, X_{(n)} : \Omega \to \mathbb{R}$ heißen die *Ordnungsstatistiken* von $(X_1, \ldots, X_n)$.

Beachte

- Insbesondere gilt
$$X_{(1)} = \min_{1 \le i \le n} X_i \qquad \text{und} \qquad X_{(n)} = \max_{1 \le i \le n} X_i , \tag{49}$$
d.h., $X_{(1)}$ bzw. $X_{(n)}$ sind das *Minimum* bzw. das *Maximum* der Stichprobenvariablen $X_1, \ldots, X_n$.
- In diesem Zusammenhang wird auch die *Stichprobenspannweite* $R_n = X_{(n)} - X_{(1)}$ betrachtet.
- Anstelle des Stichprobenmittels $\overline{X}_n$ wird manchmal der *Stichprobenmedian* M_n betrachtet, wobei
$$M_n = \begin{cases} X_{((n+1)/2)} , & \text{falls } n \text{ ungerade,} \\ \\ \bigl(X_{(n/2)} + X_{((n/2)+1)}\bigr)/2 , & \text{falls } n \text{ gerade,} \end{cases} \tag{50}$$
- Der Stichprobenmedian ist also ebenfalls ein Mittelwert: Jeweils etwa die Hälfte der Stichprobenvariablen $X_1, \ldots, X_n$ ist kleiner bzw. größer als der Stichprobenmedian.
- Ein Vorteil des Stichprobenmedians M_n besteht darin, dass M_n wesentlich weniger als $\overline{X}_n$ von den extremalen Variablen $X_{(1)}$ und $X_{(n)}$ abhängt.

1.4.1 Diskrete Stichprobenvariablen

Wir untersuchen nun die Verteilung von Ordnungsstatistiken, wobei wir zunächst den Fall betrachten, dass die Verteilung der Stichprobenvariablen $X_1, \ldots, X_n$ diskret ist.

Theorem 1.14

- *Die Stichprobenvariablen $X_1, \ldots, X_n$ seien diskret, d.h., es gebe eine (abzählbare) Menge*
$$C = \{x_{-m}, \ldots, x_{-1}, x_0, x_1, x_2, \ldots, x_{m'}\} \subset \mathbb{R}$$
mit $x_{-m} < \ldots < x_{-1} < x_0 < x_1 < x_2 < \ldots < x_{m'}$ und
$$P(X_i \in C) = 1 \qquad \forall\, i \in \{1, \ldots, n\} ,$$
wobei wir $x_{-m} = -\infty$ bzw. $x_{m'} = \infty$ setzen, falls $m = \infty$ bzw. $m' = \infty$ und $\lim_{n \to -\infty} x_n = -\infty$ bzw. $\lim_{n \to \infty} x_n = \infty$.

- *Dann gilt für beliebige $i \in \{1, \ldots, n\}$ und $j \in \{-m, \ldots, m'\}$*

$$P(X_{(i)} \le x_j) = \sum_{k=i}^{n} \binom{n}{k} P_j^k (1 - P_j)^{n-k} \tag{51}$$

bzw.

$$P(X_{(i)} = x_j) = \sum_{k=i}^{n} \binom{n}{k} \Bigl(P_j^k (1 - P_j)^{n-k} - P_{j-1}^k (1 - P_{j-1})^{n-k} \Bigr), \tag{52}$$

wobei

$$P_j = \sum_{k=-m}^{j} p_k \qquad \textit{und} \qquad p_k = P(X_i = x_k).$$

Beweis

- Für beliebige $i \in \{1, \ldots, n\}$ und $j \in \{-m, \ldots, m'\}$ deuten wir das Ereignis $\{X_i \le x_j\}$ als „Erfolg" und das Ereignis $\{X_i > x_j\}$ als „Misserfolg".
- Weil die Stichprobenvariablen $X_1, \ldots, X_n$ unabhängig und identisch verteilt sind, ergibt sich somit, dass die Zufallsvariable $Y = \#\{i : X_i \le x_j\}$ binomialverteilt ist mit $Y \sim \mathrm{Bin}(n, P_j)$.
- Weil außerdem

$$\{X_{(i)} \le x_j\} = \{Y \ge i\},$$

ergibt sich hieraus, dass

$$\begin{aligned} P(X_{(i)} \le x_j) &= P(Y \ge i) \\ &= \sum_{k=i}^{n} \binom{n}{k} P_j^k (1 - P_j)^{n-k}. \end{aligned}$$

- Damit ist (51) bewiesen.
- Außerdem gilt

$$\begin{aligned} P(X_{(i)} = x_j) &= P(X_{(i)} \le x_j) - P(X_{(i)} \le x_{j-1}) \\ &= \sum_{k=i}^{n} \binom{n}{k} \Bigl(P_j^k (1 - P_j)^{n-k} - P_{j-1}^k (1 - P_{j-1})^{n-k} \Bigr). \end{aligned}$$

□

1.4.2 Absolutstetige Stichprobenvariablen

- Über die Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$ der Stichprobenvariablen $X_1, \ldots, X_n$ setzen wir nun voraus, dass

$$F(x) = \int_{-\infty}^{x} f(y)\, dy,$$

wobei $f(y) \ge 0$ für jedes $y \in \mathbb{R}$ und $\int_{\mathbb{R}} f(y)\, dy = 1$.

- Zusätzlich setzen wir noch voraus, dass die Dichte $f : \mathbb{R} \to [0, \infty)$ stückweise stetig ist, d.h., die Dichte f habe höchstens abzählbar viele Unstetigkeitsstellen, die sich nirgendwo im Endlichen häufen mögen.
- Die Verteilungsfunktion F ist dann stückweise stetig differenzierbar.

Theorem 1.15

- *Die Stichprobenvariablen $X_1,\ldots,X_n$ seien absolutstetig mit der stückweise stetigen Dichte $f:\mathbb{R}\to[0,\infty)$.*
- *Dann ist auch die Ordnungsstatistik $X_{(i)}$ für jedes $i\in\{1,\ldots,n\}$ eine absolutstetige Zufallsvariable, deren Dichte $f_{X_{(i)}}:\mathbb{R}\to[0,\infty)$ gegeben ist durch*

$$f_{X_{(i)}}(x)=\frac{n!}{(i-1)!\,(n-i)!}\,f(x)\bigl(F(x)\bigr)^{i-1}\bigl(1-F(x)\bigr)^{n-i}. \tag{53}$$

Beweis

- Ähnlich wie beim Beweis von (51) in Theorem 1.14 setzen wir $Y=\#\{i:X_i\le x\}$.
- Weil die Stichprobenvariablen $X_1,\ldots,X_n$ unabhängig und identisch verteilt sind, ergibt sich nun, dass Y binomialverteilt ist mit $Y\sim\mathrm{Bin}(n,F(x))$.
- Wegen $\{X_{(i)}\le x\}=\{Y\ge i\}$ erhalten wir also, dass für jedes $x\in\mathbb{R}$

$$P(X_{(i)}\le x)=\sum_{k=i}^{n}\binom{n}{k}\bigl(F(x)\bigr)^{k}\bigl(1-F(x)\bigr)^{n-k}. \tag{54}$$

- Weil F stückweise stetig differenzierbar ist, gilt dies somit auch für die Verteilungsfunktion $F_{X_{(i)}}:\mathbb{R}\to[0,1]$ von $X_{(i)}$ mit $F_{X_{(i)}}(x)=P(X_{(i)}\le x)$ für jedes $x\in\mathbb{R}$.
- Insbesondere ist also $X_{(i)}$ absolutstetig.
- Beiderseitiges Differenzieren von (54) ergibt nun, dass

$$\begin{aligned}
f_{X_{(i)}}(x) &= \frac{d}{dx}\,F_{X_{(i)}}(x)\\
&= \frac{d}{dx}\,P(X_{(i)}\le x)\\
&= \sum_{k=i}^{n}\binom{n}{k}\Bigl(k\bigl(F(x)\bigr)^{k-1}\bigl(1-F(x)\bigr)^{n-k}f(x)-(n-k)\bigl(F(x)\bigr)^{k}\bigl(1-F(x)\bigr)^{n-k-1}f(x)\Bigr).
\end{aligned}$$

- Weil $n-k=0$ für $k=n$, ergibt sich hieraus, dass

$$\begin{aligned}
f_{X_{(i)}}(x) &= \binom{n}{i}if(x)\bigl(F(x)\bigr)^{i-1}\bigl(1-F(x)\bigr)^{n-i}+\sum_{k=i+1}^{n}\binom{n}{k}k\bigl(F(x)\bigr)^{k-1}\bigl(1-F(x)\bigr)^{n-k}f(x)\\
&\quad-\sum_{k=i}^{n-1}\binom{n}{k}(n-k)\bigl(F(x)\bigr)^{k}\bigl(1-F(x)\bigr)^{n-k-1}f(x).
\end{aligned}$$

- Durch die Substitution $k'=k-1$ in der ersten Summe erhalten wir, dass

$$\begin{aligned}
f_{X_{(i)}}(x) &= \frac{n!}{(i-1)!\,(n-i)!}\,f(x)\bigl(F(x)\bigr)^{i-1}\bigl(1-F(x)\bigr)^{n-i}\\
&\quad+\sum_{k=i}^{n-1}\binom{n}{k+1}(k+1)\bigl(F(x)\bigr)^{k}\bigl(1-F(x)\bigr)^{n-k-1}f(x)\\
&\quad-\sum_{k=i}^{n-1}\binom{n}{k}(n-k)\bigl(F(x)\bigr)^{k}\bigl(1-F(x)\bigr)^{n-k-1}f(x)\\
&= \frac{n!}{(i-1)!\,(n-i)!}\,f(x)\bigl(F(x)\bigr)^{i-1}\bigl(1-F(x)\bigr)^{n-i},
\end{aligned}$$

wobei sich die letzte Gleichheit aus der Identität

$$\binom{n}{k+1}(k+1) = \frac{n!}{k!\,(n-k-1)!} = \binom{n}{k}(n-k)$$

ergibt. $\square$

In Verallgemeinerung von Theorem 1.15 kann man die folgende Formel für die gemeinsame Dichte von zwei Ordnungsstatistiken herleiten.

Theorem 1.16

- *Die Stichprobenvariablen $X_1, \ldots, X_n$ seien absolutstetig mit der stückweise stetigen Dichte $f : \mathbb{R} \to [0, \infty)$.*
- *Für $1 \le i < j \le n$ ist die gemeinsame Dichte $f_{X_{(i)},X_{(j)}} : \mathbb{R}^2 \to [0, \infty)$ der Ordnungsstatistiken $X_{(i)}, X_{(j)}$ gegeben durch*

$$f_{X_{(i)},X_{(j)}}(x_i, x_j) = \begin{cases} \dfrac{n!\, f(x_i) f(x_j) \bigl(F(x_i)\bigr)^{i-1} \bigl(F(x_j) - F(x_i)\bigr)^{j-1-i} \bigl(1 - F(x_j)\bigr)^{n-j}}{(i-1)!\,(j-1-i)!\,(n-j)!}\,, & \\ & \text{falls } -\infty < x_i < x_j < \infty, \\ 0\,, & \text{sonst.} \end{cases} \tag{55}$$

Beweis

- Der Beweis von Theorem 1.16 verläuft ähnlich wie der Beweis von Theorem 1.15. Wir geben deshalb hier lediglich die Beweisidee an und lassen die Details weg (vgl. auch Übungsaufgabe 4.1).
- Und zwar sei $Y = \#\{i : 1 \le i \le n,\ X_i \le y\}$ und $Z = \#\{i : 1 \le i \le n,\ y \le X_i \le z\}$ für beliebige $y < z$.
- Dann kann man leicht zeigen, dass für $1 \le k, \ell \le n$

$$P(Y = k,\ Z = \ell) = \frac{n!}{k!\,\ell!\,(n-k-\ell)!}\,\bigl(F(y)\bigr)^k \bigl(F(z) - F(y)\bigr)^\ell \bigl(1 - F(z)\bigr)^{n-k-\ell}.$$

- Außerdem lässt sich die gemeinsame Verteilungsfunktion $F_{X_{(i)},X_{(j)}} : \mathbb{R}^2 \to [0, 1]$ der Ordnungsstatistiken $X_{(i)}, X_{(j)}$ darstellen durch

$$F_{X_{(i)},X_{(j)}}(y, z) = P(Y \ge i,\ Y + Z \ge j) = \sum_{k=i}^{j-1} \sum_{\ell=j-k}^{n-k} P(Y = k,\ Z = \ell) + P(Y \ge j)\,.$$

- Durch Einsetzen der vorhergenden Formel für $P(Y = k,\ Z = \ell)$ bzw. $P(Y \ge j)$ ergibt sich nun die Behauptung (55). $\square$

Beachte

- Es lassen sich auch Formeln für die gemeinsame Dichte von drei und mehr Ordnungsstatistiken herleiten.
- Insbesondere gilt für die gemeinsame Dichte $f_{X_{(1)},\ldots,X_{(n)}} : \mathbb{R}^n \to [0, \infty)$ aller Ordnungsstatistiken $X_{(1)}, \ldots, X_{(n)}$:

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n!\, f(x_1) \ldots f(x_n)\,, & \text{falls } -\infty < x_1 < \ldots < x_n < \infty, \\ 0\,, & \text{sonst,} \end{cases} \tag{56}$$

vgl. Übungsaufgabe 4.2.

1.4.3 Beispiel: gleichverteilte Stichprobenvariablen

- Wir illustrieren nun die in Abschnitt 1.4.2 hergeleiteten Ergebnisse für das folgende Beispiel.
- Die Stichprobenvariablen $X_1, \ldots, X_n$ seien gleichverteilt im Intervall $(0, \theta)$ für ein $\theta > 0$, d.h.

$$f(x) = \begin{cases} \theta^{-1}, & \text{falls } x \in (0, \theta), \\ 0, & \text{falls } x \notin (0, \theta). \end{cases}$$

1. *Dichte, Erwartungswert und Varianz von $X_{(i)}$*
   - Aus Theorem 1.15 ergibt sich durch Einsetzen in (53), dass

   $$f_{X_{(i)}}(x) = \begin{cases} \dfrac{n!}{(i-1)!\,(n-i)!}\,\theta^{-n} x^{i-1} (\theta - x)^{n-i}, & \text{falls } x \in (0, \theta), \\ 0, & \text{falls } x \notin (0, \theta). \end{cases}$$

   - Hieraus folgt insbesondere, dass

   $$\mathbb{E}\, X_{(i)} = \frac{i\theta}{n+1} \qquad \text{und} \qquad \operatorname{Var} X_{(i)} = \frac{i(n-i+1)\theta^2}{(n+1)^2(n+2)}\,.$$

2. *Gemeinsame Dichte von $X_{(1)}$ und $X_{(n)}$*
   - Aus Theorem 1.16 ergibt sich durch Einsetzen in (55), dass für $0 < x_1 < x_n < \theta$

   $$f_{X_{(1)},X_{(n)}}(x_1, x_n) = \frac{n(n-1)}{\theta^2} \left(\frac{x_n}{\theta} - \frac{x_1}{\theta}\right)^{n-2} = \frac{n(n-1)(x_n - x_1)^{n-2}}{\theta^n}\,. \tag{57}$$

3. *Gemeinsame Dichte von $R_n = X_{(n)} - X_{(1)}$ und $Z_n = (X_{(1)} + X_{(n)})/2$*
   - Wir betrachten nun die Stichprobenspannweite $R_n = X_{(n)} - X_{(1)}$ und das arithmetische Mittel $Z_n = (X_{(1)} + X_{(n)})/2$ der extremalen Ordnungsstatistiken $X_{(1)}$ und $X_{(n)}$.
   - Ähnlich wie das Stichprobenmittel $\overline{X}_n$ bzw. der Stichprobenmedian M_n ist auch das Mittel $Z_n = (X_{(1)} + X_{(n)})/2$ eine sogenannte *Lokationskenngröße*.
   - Für die gemeinsame Dichte $f_{R_n,Z_n} : \mathbb{R}^2 \to [0, \infty)$ des Zufallsvektors (R_n, Z_n) gilt

   $$f_{R_n,Z_n}(r.z) = \begin{cases} \dfrac{n(n-1)r^{n-2}}{\theta^n}, & \text{falls } r \in (0, \theta) \text{ und } r/2 < z < \theta - r/2, \\ 0, & \text{sonst.} \end{cases} \tag{58}$$

   - Dabei ergibt sich (58) aus (57) und aus dem in Theorem 1.9 angegebenen Transformationssatz für die Dichte von absolutstetigen Zufallsvektoren.
   - Denn durch die Abbildung

   $$(x_{(1)}, x_{(n)}) \to (r, z) : \mathbb{R}^2 \to \mathbb{R}^2$$

   mit

   $$(r, z) = \left(x_{(n)} - x_{(1)}\,,\ \frac{x_{(1)} + x_{(n)}}{2}\right)$$

   wird die Menge $\{(x_{(1)}, x_{(n)}) : 0 < x_{(1)} < x_{(n)} < \theta\}$ auf die Menge $\{(r, z) : 0 < r < \theta, r/2 < z < \theta - r/2\}$ abgebildet, und

- für die Umkehrabbildung
$$(r,z) \to (x_{(1)}, x_{(n)}) : \mathbb{R}^2 \to \mathbb{R}^2$$
mit
$$(x_{(1)}, x_{(n)}) = \Bigl(z - \frac{r}{2}\,,\, z + \frac{r}{2}\Bigr)$$
ist die Jacobi-Determinante gleich -1.

4. *Dichte von* $R_n = X_{(n)} - X_{(1)}$
   - Für die (Rand-) Dichte $f_{R_n} : \mathbb{R} \to [0,\infty)$ von $R_n = X_{(n)} - X_{(1)}$ gilt
$$f_{R_n}(r) = \begin{cases} \dfrac{n(n-1)r^{n-2}(\theta - r)}{\theta^n}\,, & \text{falls } r \in (0,\theta), \\ 0\,, & \text{sonst,} \end{cases} \tag{59}$$
   - denn aus (58) und aus Theorem WR–3.9 über die Integraldarstellung von Randdichten ergibt sich, dass für jedes $r \in (0,\theta)$
$$\begin{aligned} f_{R_n}(r) &= \int\limits_{r/2}^{\theta - r/2} \frac{n(n-1)r^{n-2}}{\theta^n}\, dz \\ &= \frac{n(n-1)r^{n-2}(\theta - r)}{\theta^n}\,. \end{aligned}$$

5. *Dichte von* $Z_n = (X_{(1)} + X_{(n)})/2$
   - Auf die gleiche Weise wie bei der Herleitung von (59) kann man zeigen, dass
$$f_{Z_n}(z) = \begin{cases} \dfrac{n(2z)^{n-1}}{\theta^n}\,, & \text{falls } z \in (0,\theta/2], \\ \dfrac{n\bigl(2(\theta - z)\bigr)^{n-1}}{\theta^n}\,, & \text{falls } z \in (\theta/2,\theta), \end{cases} \tag{60}$$
   - denn aus (58) ergibt sich, dass für $z \in (0,\theta/2]$
$$f_{Z_n}(z) = \int\limits_0^{2z} \frac{n(n-1)r^{n-2}}{\theta^n}\, dr = \frac{n(2z)^{n-1}}{\theta^n}$$
bzw. für $z \in (\theta/2,\theta)$
$$f_{Z_n}(z) = \int\limits_0^{2(\theta - z)} \frac{n(n-1)r^{n-2}}{\theta^n}\, dr = \frac{n\bigl(2(\theta - z)\bigr)^{n-1}}{\theta^n}\,.$$

1.5 Empirische Verteilungsfunktion

- Wir betrachten nun noch eine andere Klasse von Stichprobenfunktionen, mit deren Hilfe die Verteilungsfunktion F der Stichprobenvariablen $X_1,\ldots,X_n$ aus den vorliegenden Daten $x_1,\ldots,x_n$ bestimmt werden kann.
- Hierfür betrachten wir für jedes $x \in \mathbb{R}$ die Stichprobenfunktion $\varphi_x : \mathbb{R}^n \to \mathbb{R}$ mit
$$\varphi_x(x_1,\ldots,x_n) = \frac{\#\{i : 1 \le i \le n,\, x_i \le x\}}{n}\,, \tag{61}$$
- wobei $\varphi_x(x_1,\ldots,x_n)$ die *relative Häufigkeit* derjenigen Stichprobenwerte ist, die den Schwellenwert x nicht überschreiten.

1.5.1 Definition und elementare Eigenschaften

- Man kann sich leicht überlegen, dass für jeden Vektor $(x_1,\ldots,x_n)\in\mathbb{R}^n$ die in (61) definierte Abbildung
$$x\to\varphi_x(x_1,\ldots,x_n) \tag{62}$$
die Eigenschaften einer Verteilungsfunktion hat.
- Die in (62) gegebene Abbildung wird deshalb *empirische Verteilungsfunktion* der (konkreten) Stichprobe $(x_1,\ldots,x_n)$ genannt.

Dies führt zu der folgenden Begriffsbildung.

Definition Die Abbildung $\widehat{F}_n:\mathbb{R}\times\Omega\to[0,1]$ mit
$$\widehat{F}_n(x,\omega)=\frac{\#\{i:\,1\le i\le n,\,X_i(\omega)\le x\}}{n} \tag{63}$$
heißt *empirische Verteilungsfunktion* der Zufallsstichprobe $(X_1,\ldots,X_n)$.

Beachte

- Die in (63) gegebene Abbildung kann man als eine Familie $\{\widehat{F}_n(x),\,x\in\mathbb{R}\}$ von Zufallvariablen $\widehat{F}_n(x):\Omega\to[0,1]$ auffassen.
- Eine solche Familie von Zufallsvariablen wird *empirischer Prozess* genannt. Empirische Prozesse bilden eine spezielle Klasse *stochastischer Prozesse*, d.h., eine Familie von Zufallsvariablen, die über einunddemselben Wahrscheinlichkeitsraum definiert sind.

Theorem 1.17 *Für jedes $x\in\mathbb{R}$ gilt:*

1. *Die Zufallsvariable $n\,\widehat{F}_n(x)$ ist binomialverteilt mit den Parametern n und $p=F(x)$, d.h., es gilt*
$$P(n\,\widehat{F}_n(x)=k)=\binom{n}{k}(F(x))^k(1-F(x))^{n-k}\,,\qquad\forall\,k\in\{0,1,\ldots,n\}\,. \tag{64}$$
2. *Insbesondere gilt*
$$\mathbb{E}\,\widehat{F}_n(x)=F(x)\qquad\textit{und}\qquad\operatorname{Var}\widehat{F}_n(x)=\frac{F(x)(1-F(x))}{n}\,. \tag{65}$$
3. *Mit Wahrscheinlichkeit 1 gilt*
$$\lim_{n\to\infty}\widehat{F}_n(x)=F(x)\,. \tag{66}$$
4. *Falls $0<F(x)<1$, dann gilt außerdem für jedes $y\in\mathbb{R}$*
$$\lim_{n\to\infty}P\Bigl(\sqrt{n}\frac{\widehat{F}_n(x)-F(x)}{\sqrt{F(x)(1-F(x))}}\le y\Bigr)=\Phi(y)\,, \tag{67}$$
wobei $\Phi(y)$ die Verteilungsfunktion der Standardnormalverteilung ist.

Beweis

- So wie in Abschnitt 1.4 deuten wir das Ereignis $\{X_i\le x\}$ als „Erfolg" und das Ereignis $\{X_i>x\}$ als „Misserfolg".
- Aus der Definitionsgleichung (63) der empirischen Verteilungsfunktion $\widehat{F}_n$ ergibt sich somit, dass wir die Zufallsvariable $n\,\widehat{F}_n(x)$ als die Anzahl der Erfolge beim n-maligen Münzwurf mit den identischen Erfolgswahrscheinlichkeiten $a_1=a_2=\ldots=a_n=F(x)$ auffassen können, vgl. Beispiel 2 in Abschnitt WR–3.2.2.

- Hieraus ergibt sich, dass $n\,\widehat{F}_n(x)$ binomialverteilt ist mit $n\,\widehat{F}_n(x) \sim \mathrm{Bin}(n, F(x))$.
- Damit ist (64) bewiesen.
- Aus den Formeln für den Erwartungswert bzw. die Varianz der Binomialverteilung ergibt sich nun die Gültigkeit von (65), vgl. jeweils Beispiel 1 in den Abschnitten WR–4.1.1 bzw. WR–4.2.2.
- Aus der Definitionsgleichung (63) der empirischen Verteilungsfunktion $\widehat{F}_n$ folgt außerdem, dass

$$n\,\widehat{F}_n(x) = \sum_{i=1}^n Y_i\,,$$

wobei

$$Y_i = \begin{cases} 1\,, & \text{falls } X_i \le x,\\ 0\,, & \text{falls } X_i > x. \end{cases}$$

- Wegen dieser Darstellungsmöglichkeit von $n\,\widehat{F}_n(x)$ als Summe von n unabhängigen und identisch (Bernoulli-) verteilten Zufallsvariablen Y_i mit $\mathbb{E}\,Y_i = F(x)$ ergibt sich (66) unmittelbar aus dem starken Gesetz der großen Zahlen; vgl. Theorem WR–5.15.
- Mit der gleichen Begründung ergibt sich (67) unmittelbar aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen; vgl. Theorem WR–5.16. $\square$

Beachte Weil die Zufallsvariable $\widehat{F}_n(x)$ den Erwartungswert $F(x)$ hat (vgl. (65)), kann $\widehat{F}_n(x)$ als ein geeigneter Schätzer von $F(x)$ angesehen werden.

1.5.2 Satz von Gliwenko-Cantelli

- In diesem Abschnitt betrachten wir die Abweichung

$$D_n = \sup_{x\in\mathbb{R}} |\,\widehat{F}_n(x) - F(x)|\,. \tag{68}$$

der empirischen Verteilungsfunktion $\widehat{F}_n$ von der zu schätzenden Verteilungsfunktion F der Stichprobenvariablen $X_1,\ldots,X_n$.

- Die in (68) definierte Zufallsvariable D_n wird *Kolmogorow-Abstand* von $\widehat{F}_n$ und F genannt.
- Weil die empirische Verteilungsfunktion $\widehat{F}_n$ eine Treppenfunktion ist (vgl. (63)) und weil F monoton nichtfallend und rechtsstetig ist, lässt sich (68) auch wie folgt schreiben:

$$D_n = \max_{i\in\{1,\ldots,n\}} \max\Bigl\{\Bigl|\frac{i-1}{n} - F(X_{(i)}-0)\Bigr|\,,\,\Bigl|\frac{i}{n} - F(X_{(i)})\Bigr|\Bigr\}$$

bzw.

$$D_n = \max_{i\in\{1,\ldots,n\}} \max\Bigl\{F(X_{(i)}-0) - \frac{i-1}{n}\,,\,\frac{i}{n} - F(X_{(i)})\Bigr\} \tag{69}$$

wobei $X_{(i)}$ die i-te Ordnungsstatistik von $X_1,\ldots,X_n$ ist, vgl. Übungsaufgabe 5.3.

- Wir können D_n als den *maximalen Schätzfehler* bei der Schätzung der Funktionswerte $F(x)$ von F durch $\widehat{F}_n(x)$ auffassen.

Beachte

- In Theorem 1.17 wurde gezeigt (vgl. (66)), dass die empirische Verteilungsfunktion $\widehat{F}_n$ *punktweise* mit Wahrscheinlichkeit 1 gegen F konvergiert, d.h.,

$$P\Bigl(\lim_{n\to\infty} \widehat{F}_n(x) = F(x)\Bigr) = 1 \qquad \forall\, x\in\mathbb{R}\,.$$

- Darüber hinaus kann man zeigen, dass sich auch die Verteilungsfunktion F insgesamt durch den empirischen Prozess $\{\widehat{F}_n(x),\, x \in \mathbb{R}\}$ im Sinne der fast sicheren Konvergenz approximieren lässt, falls der Stichprobenumfang n unendlich groß wird.
- Damit ist die folgende Eigenschaft des Kolmogorow-Abstandes D_n gemeint, die in der Literatur *Satz von Gliwenko-Cantelli* genannt wird.

Theorem 1.18 *Es gilt*

$$P\Bigl(\lim_{n\to\infty} D_n = 0\Bigr) = 1\,. \tag{70}$$

Beweis

- Wir nehmen zunächst an, dass die Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ stetig ist.
- Für jede natürliche Zahl $m \in \mathbb{N}$ gibt es dann reelle Zahlen $z_1, \ldots, z_{m-1} \in \mathbb{R}$, so dass

  $$z_0 = -\infty < z_1 < \ldots < z_{m-1} < z_m = \infty$$

  und

  $$F(z_0) = 0\,,\; F(z_1) = \frac{1}{m}\;, \ldots,\; F(z_k) = \frac{k}{m}\;, \ldots,\; F(z_{m-1}) = \frac{m-1}{m}\;,\; F(z_m) = 1\,. \tag{71}$$

- Mit der Schreibweise $\varepsilon = 1/m$ ergibt sich dann hieraus, dass für jedes $z \in [z_k, z_{k+1})$

  $$\begin{array}{rcccl} \widehat{F}_n(z) - F(z) & \le & \widehat{F}_n(z_{k+1}) - F(z_k) & = & \widehat{F}_n(z_{k+1}) - F(z_{k+1}) + \varepsilon\,, \\[2ex] \widehat{F}_n(z) - F(z) & \ge & \widehat{F}_n(z_k) - F(z_{k+1}) & = & \widehat{F}_n(z_k) - F(z_k) - \varepsilon\,. \end{array} \tag{72}$$

- Für beliebige $m \in \mathbb{N}$ und $k \in \{0, 1, \ldots, m\}$ sei

  $$A_{m,k} = \Bigl\{\omega \in \Omega : \lim_{n\to\infty} \widehat{F}_n(z_k, \omega) = F(z_k)\Bigr\}\,.$$

- Aus Teilaussage 3 von Theorem 1.17 ergibt sich dann, dass

  $$P(A_{m,k}) = 1 \tag{73}$$

  für beliebige $m \in \mathbb{N}$ und $k \in \{0, 1, \ldots, m\}$.
- Damit gilt auch

  $$P(A_m) = 1\,, \tag{74}$$

  wobei

  $$A_m = \bigcap_{k=0}^{m} A_{m,k}\,,$$

  denn aus (73) ergibt sich, dass

  $$\begin{aligned} P(A_m) &= 1 - P(A_m^c) \\ &= 1 - P\Bigl(\bigcup_{k=0}^{m} A_{m,k}^c\Bigr) \\ &\ge 1 - \sum_{k=0}^{m} P(A_{m,k}^c) = 1\,. \end{aligned}$$

- Für jedes $\omega \in A_m$ gibt es nun eine natürliche Zahl $n(\omega) \in \mathbb{N}$, so dass
$$|\widehat{F}_n(z_k,\omega) - F(z_k)| < \varepsilon$$
für jedes $n \geq n(\omega)$ und für jedes $k \in \{0,1,\ldots,m\}$.
- Hieraus und aus (72) folgt, dass
$$\sup_{z\in\mathbb{R}} |\widehat{F}_n(z,\omega) - F(z)| < 2\varepsilon \tag{75}$$
für jedes $\omega \in A_m$ und für jedes $n \geq n(\omega)$.
- Dies bedeutet, dass es für jedes $\omega \in A = \bigcap_{m=1}^{\infty} A_m$ und für jedes $\varepsilon > 0$ eine natürliche Zahl $n(\omega,\varepsilon) \in \mathbb{N}$ gibt, so dass (75) für jedes $n \geq n(\omega,\varepsilon)$ gilt.
- Dabei ergibt sich genauso wie im Beweis von (74), dass
$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P\Bigl(\bigcup_{m=1}^{\infty} A_m^c\Bigr) \\ &\geq 1 - \sum_{m=1}^{\infty} P(A_m^c) = 1\,, \end{aligned}$$
weil $P(A_m^c) = 0$ für jedes $m \in \mathbb{N}$.
- Weil $\varepsilon > 0$ beliebig klein gewählt werden kann, ist somit die Behauptung (70) für den Fall bewiesen, dass die Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ stetig ist.
- Im Fall einer beliebigen (nichtnotwendig stetigen) Verteilungsfunktion F lässt sich die Gültigkeit von (70) auf ähnliche Weise zeigen.
- Anstelle von (71) nutzen wir nun die Tatsache, dass es für jede natürliche Zahl $m \in \mathbb{N}$ reelle Zahlen $z_1,\ldots,z_{m-1} \in \mathbb{R}$ gibt, so dass
$$z_0 = -\infty < z_1 < \ldots < z_{m-1} < z_m = \infty$$
und für jedes $k \in \{0,1,\ldots,m-1\}$
$$F(z_{k+1} - 0) - F(z_k) \leq \varepsilon\,, \tag{76}$$
wobei $\varepsilon = 1/m$.
- Außerdem ergibt sich genauso wie bei der Herleitung von (72), dass für jedes $z \in [z_k, z_{k+1})$
$$\begin{aligned} \widehat{F}_n(z) - F(z) &\leq \widehat{F}_n(z_{k+1} - 0) - F(z_{k+1} - 0) + \varepsilon\,, \\ \widehat{F}_n(z) - F(z) &\geq \widehat{F}_n(z_k) - F(z_k) - \varepsilon\,. \end{aligned} \tag{77}$$
- Aus Teilaussage 3 von Theorem 1.17 ergibt sich ähnlich wie (73), dass $P(A'_{m,k}) = 1$ für beliebige $m \in \mathbb{N}$ und $k \in \{0,1,\ldots,m\}$, wobei
$$A'_{m,k} = \Bigl\{\omega \in \Omega : \lim_{n\to\infty} \widehat{F}_n(z_k - 0,\omega) = F(z_k - 0)\Bigr\}\,.$$
- Hieraus und aus (77) folgt dann, dass (75) für jedes $\omega \in A'_m$ und für jedes $n \geq n(\omega)$ gilt, wobei
$$A'_m = \bigcap_{k=0}^{m} \bigl(A_{m,k} \cap A'_{m,k}\bigr)\,.$$
- Weil $P(A'_m) = 1$ für jedes $m \in \mathbb{N}$, ergibt sich nun die Behauptung genauso wie im ersten Teil des Beweises. $\square$

1.5.3 Verteilung des maximalen Schätzfehlers

- Um die Verteilung des maximalen Schätzfehlers D_n bei der Schätzung der Funktionswerte $F(x)$ von F durch $\widehat{F}_n(x)$ näher zu untersuchen, benötigen wir den folgenden Begriff.
- Das Intervall $I = (a, b] \subset \mathbb{R}$ mit $a < b$ heißt *Konstanzbereich* der Verteilungsfunktion F von $X_1, \ldots, X_n$, falls $P(X_1 \in I) = 0$ und falls es kein größeres Intervall mit dieser Eigenschaft gibt, das I enthält.
- Für den Fall, dass die Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$ der Stichprobenvariablen $X_1, \ldots, X_n$ stetig ist, zeigen wir nun, dass der Kolmogorow-Abstand D_n eine sogenannte *verteilungsfreie Stichprobenfunktion* ist.
- Damit ist gemeint, dass die Verteilung von D_n nicht von der speziellen Ausprägung der stetigen Verteilungsfunktion F abhängt.

Theorem 1.19 *Für jede stetige Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$ gilt*

$$D_n \stackrel{\mathrm{d}}{=} \sup_{y \in [0,1]} \left| \widehat{G}_n(y) - y \right| , \tag{78}$$

wobei $\widehat{G}_n : \mathbb{R} \times \Omega \to [0, 1]$ die empirische Verteilungsfunktion einer beliebigen Zufallsstichprobe $(Y_1, \ldots, Y_n)$ ist, die aus n unabhängigen und in dem Intervall $[0, 1]$ gleichverteilten Stichprobenvariablen $Y_1, \ldots, Y_n$ besteht.

Beweis

- Sei B die Vereinigung aller Konstanzbereiche von F. Dann gilt mit Wahrscheinlichkeit 1
$$D_n = \sup_{x \in B^c} | \widehat{F}_n(x) - F(x)| . \tag{79}$$
- Außerdem gilt für jedes $x \in B^c$
$$\{X_i \le x\} = \{F(X_i) \le F(x)\} \tag{80}$$
- Wir setzen nun $Y_i = F(X_i)$ für jedes $i \in \{1, \ldots, n\}$.
- Aus Theorem 1.8 (vgl. auch Theorem WR–3.18) ergibt sich dann, dass die Zufallsvariablen $Y_1, \ldots, Y_n$ unabhängig sind.
- Weil F stetig ist, gibt es für jedes $y \in (0, 1)$ ein $x_y \in \mathbb{R}$, so dass
$$x_y = \inf\{x' : F(x') = y\} \in B^c .$$
- Folglich gilt für jedes $y \in (0, 1)$
$$P(Y_i \le y) = P(F(X_i) \le F(x_y)) = P(X_i \le x_y) = F(x_y) = y ,$$
wobei sich die zweite Gleichheit aus (80) ergibt.
- Die Zufallsvariablen $Y_1, \ldots, Y_n$ sind also unabhängig und im Intervall $[0, 1]$ gleichverteilt.
- Wegen (80) gilt somit, dass $\widehat{F}_n(x) = \widehat{G}_n(F(x))$ für jedes $x \in B^c$.
- Hieraus und aus (79) folgt, dass
$$\begin{aligned} D_n &\stackrel{(79)}{=} \sup_{x \in B^c} | \widehat{F}_n(x) - F(x)| \\ &= \sup_{x \in B^c} | \widehat{G}_n(F(x)) - F(x)| \\ &= \sup_{x \in \mathbb{R}} | \widehat{G}_n(F(x)) - F(x)| \\ &= \sup_{y \in [0,1]} | \widehat{G}_n(y) - y| , \end{aligned}$$
wobei in der letzten Gleichheit erneut die Voraussetzung genutzt wurde, dass F stetig ist. $\square$

Korollar 1.3 *Für jede stetige Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ gilt*

$$D_n \stackrel{\mathrm{d}}{=} \max_{i\in\{1,\ldots,n\}} \max\Bigl\{Y_{(i)} - \frac{i-1}{n}\,,\,\frac{i}{n} - Y_{(i)}\Bigr\}\,, \tag{81}$$

wobei $Y_{(i)}$ die i-te Ordnungsstatistik der in $[0,1]$ gleichverteilten Stichprobenvariablen $Y_1,\ldots,Y_n$ ist.

Beweis

- Wegen Theorem 1.19 können wir o.B.d.A. annehmen, dass $F(x) = x$ für jedes $x \in [0,1]$.
- Aus der Darstellungsformel (69) für D_n ergibt sich dann, dass

$$\begin{aligned} D_n &= \max_{i\in\{1,\ldots,n\}} \max\Bigl\{F(X_{(i)}-0) - \frac{i-1}{n}\,,\,\frac{i}{n} - F(X_{(i)})\Bigr\}\\ &= \max_{i\in\{1,\ldots,n\}} \max\Bigl\{X_{(i)} - \frac{i-1}{n}\,,\,\frac{i}{n} - X_{(i)}\Bigr\}\\ &\stackrel{\mathrm{d}}{=} \max_{i\in\{1,\ldots,n\}} \max\Bigl\{Y_{(i)} - \frac{i-1}{n}\,,\,\frac{i}{n} - Y_{(i)}\Bigr\}\,. \end{aligned}$$

□

Beachte

- Für den Fall, dass F stetig ist, ergibt sich aus (81) die folgende Integraldarstellung für die Verteilungsfunktion von D_n:

$$P\Bigl(D_n \le \frac{1}{2n} + x\Bigr) = \begin{cases} 0\,, & \text{falls } x \le 0,\\ \displaystyle\int_{\frac{1}{2n}-x}^{\frac{1}{2n}+x}\int_{\frac{3}{2n}-x}^{\frac{3}{2n}+x}\cdots\int_{\frac{2n-1}{2n}-x}^{\frac{2n-1}{2n}+x} g(y_1,\ldots,y_n)\,dy_n\ldots dy_1\,, & \text{falls } 0 < x < \frac{2n-1}{2n},\\ 1\,, & \text{falls } x \ge \frac{2n-1}{2n}, \end{cases}$$

  wobei

$$g(y_1,\ldots,y_n) = \begin{cases} n!\,, & \text{falls } 0 < y_1 < \ldots < y_n < 1,\\ 0\,, & \text{sonst.} \end{cases}$$

- Diese Integraldarstellung kann mit (nichtelementaren) kombinatorischen Überlegungen hergeleitet werden; vgl. beispielsweise das Buch von J.D. Gibbons und S. Chakrabarti (1992) *Nonparametric Statistical Inference*, 3rd ed., Marcel Dekker, New York.
- Für große n kann man anstelle dieser Integraldarstellung eine (asymptotische) Näherungsformel zur Bestimmung der Verteilungsfunktion von D_n verwenden.
- Und zwar kann man (ähnlich wie beim zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen; vgl. Theorem WR–5.16) zeigen, dass auch D_n bei entsprechend gewählter Normierung gegen einen nichtdeterministischen, d.h. zufälligen Grenzwert (im Sinne der Verteilungskonvergenz) strebt.
- Dies ist die Aussage des folgenden *Satzes von Kolmogorow.*

Theorem 1.20 *Falls die Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ der Stichprobenvariablen $X_1, \ldots, X_n$ stetig ist, dann gilt für jedes $x \in \mathbb{R}$*

$$\lim_{n\to\infty} P\bigl(\sqrt{n}\, D_n \le x\bigr) = K(x)\,, \tag{82}$$

wobei

$$K(x) = \begin{cases} 1 - 2\sum\limits_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2x^2)\,, & \textit{falls } x > 0, \\ 0\,, & \textit{falls } x \le 0. \end{cases} \tag{83}$$

Der *Beweis* von Theorem 1.20 ist tiefliegend und geht über den Rahmen dieser einführenden Vorlesung hinaus; vgl. beispielsweise Kapitel 13 in L. Breiman (1992), *Probability*, 2nd ed., SIAM, Philadelphia.

Beachte

- Wegen (82) ist klar, dass die in (83) gegebene Funktion $K : \mathbb{R} \to [0,1]$ die Eigenschaften einer Verteilungsfunktion hat.
- Für $n > 40$ liefern (82) und (83) eine brauchbare Näherungsformel zur Bestimmung der Verteilungsfunktion von D_n:
$$P(D_n \le x) \approx K(x\sqrt{n}) \qquad \forall\, x > 0\,.$$

2 Punktschätzer

2.1 Parametrisches Modell

- Die Verteilungsfunktion F der Stichprobenvariablen $X_1, \ldots, X_n$ möge zu einer vorgegebenen (d.h. bekannten) *parametrischen Familie* von Verteilungsfunktionen $\{F_\theta,\, \theta \in \Theta\}$ gehören,
- wobei die Menge $\Theta \subset \mathbb{R}^m$ *Parameterraum* genannt wird und $m \in \mathbb{N}$ eine beliebige, jedoch vorgegebene natürliche Zahl ist.
- Mit anderen Worten: Es gelte $F = F_\theta$ für ein $\theta \in \Theta$, wobei jedoch der Parametervektor $\theta = (\theta_1, \ldots, \theta_m)$ (bzw. ein Teil seiner Komponenten) unbekannt sei und aus den beobachteten Daten $x_1, \ldots, x_n$ geschätzt werden soll.
- Dabei werden wir stets voraussetzen, dass
  - der Parameterraum Θ eine Borel-Menge ist, d.h., $\Theta \in \mathcal{B}(\mathbb{R}^m)$.
  - die Parametrisierung $\theta \to F_\theta$ *identifizierbar* ist, d.h., es gelte $F_{\theta_1} \neq F_{\theta_2}$, falls $\theta_1 \neq \theta_2$.

Wir nehmen an, dass der (gemeinsame) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, über dem die Stichprobenvariablen $X_1, X_2, \ldots$ definiert sind, der sogenannte *kanonische Wahrscheinlichkeitsraum* ist, vgl. auch das in Abschnitt WR–4.1.1 diskutierte Beispiel des wiederholten Würfelns. D. h., wir setzen

$$\Omega = \mathbb{R} \times \mathbb{R} \times \ldots = \{\omega : \omega = (\omega_1, \omega_2, \ldots),\ \omega_i \in \mathbb{R}\ \ \forall i = 1, 2, \ldots\}$$

und $\mathcal{F} = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \ldots$, wobei das Wahrscheinlichkeitsmaß P gegeben ist durch

$$P(\{\omega : \omega \in \Omega, \omega_{i_1} \le x_{i_1}, \ldots, \omega_{i_k} \le x_{i_k}\}) = F(x_{i_1}) \cdot \ldots \cdot F(x_{i_k})\,. \tag{1}$$

Die Stichprobenvariablen $X_i : \Omega \to \mathbb{R}$ sind gegeben durch $X(\omega) = \omega_i$, d.h. durch die *Projektion* auf die i-te Komponente ω_i von ω.

Beachte

- Das in (1) definierte Wahrscheinlichkeitsmaß P bezeichnen wir im folgenden mit P_θ, um zu betonen, dass P von dem (unbekannten) Parametervektor $\theta \in \Theta$ abhängt.
- Entsprechend verwenden wir die Bezeichnungen $\mathbb{E}_\theta$ und Var_θ für Erwartungswert bzw. Varianz.
- Falls keine Verwechslung möglich ist, dann bezeichnen wir auch die Verteilung einer (einzelnen) Stichprobenvariablen X_i mit P_θ, d.h. $P_\theta(X_i \le x_i) = P_\theta((-\infty, x_i]) = F_\theta(x_i)$.
- Zur Erinnerung: Für jede Borel-messbare Funktion $\widehat{\theta} : \mathbb{R}^n \to \mathbb{R}^m$ wird der Zufallsvektor $\widehat{\theta}(X_1, \ldots, X_n)$ Statistik bzw. zufällige Stichprobenfunktion genannt.
- Bei der Schätzung von Parametern nennt man $\widehat{\theta}(X_1, \ldots, X_n)$ *Punktschätzer* für den Parameter θ. Dabei wird meistens vorausgesetzt, dass $P_\theta(\widehat{\theta}(X_1, \ldots, X_n) \in \Theta) = 1$. Manchmal wird jedoch zugelassen, dass die Werte des Schätzers $\widehat{\theta}(X_1, \ldots, X_n)$ mit einer (kleinen) positiven Wahrscheinlichkeit außerhalb des Parameterraumes Θ liegen können.

Beispiel Eine der wichtigsten parametrischen Verteilungsfamilien $\{P_\theta,\, \theta \in \Theta\}$ von Stichprobenvariablen, die in dieser Vorlesung betrachtet werden, ist die *Normalverteilungsfamilie* $\{\mathrm{N}(\mu, \sigma^2),\, \mu \in \mathbb{R},\, \sigma^2 > 0\}$. In diesem Fall ist $m = 2$, $\Theta = \mathbb{R} \times (0, \infty)$ und $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$.

Beachte Weitere Beispiele von Familien parametrischer Verteilungen, die bisher in dieser Vorlesung (bzw. in der Vorlesung „Wahrscheinlichkeitsrechnung” des WS 01/02) betrachtet wurden, sind die Familien der

- (diskreten bzw. stetigen) Gleichverteilung,
- Binomialverteilung,
- Poisson-Verteilung,
- Exponentialverteilung,
- χ^2-Verteilung,
- Gammaverteilung,
- t-Verteilung.

2.2 Methoden zur Gewinnung von Punktschätzern

Wir diskutieren nun Methoden, die ein systematisches Vorgehen bei der Wahl einer Borel-messbaren Abbildung $\widehat{\theta} : \mathbb{R}^n \to \Theta$ ermöglichen, um den unbekannten Parametervektor θ auf geeignete Weise zu schätzen.

2.2.1 Momenten-Methode

- Die Momentenmethode ist eines der ältesten Verfahren zur Gewinnung von Schätzern für die unbekannten Komponenten des Parametervektors $\theta = (\theta_1, \ldots, \theta_m)$.
- Sie wurde von Karl Pearson (1857–1936) gegen Ende des 19. Jahrhunderts eingeführt und
- beruht auf dem Vergleich von Momenten der Stichprobenvariablen $X_1, \ldots, X_n$ mit den entsprechenden empirischen Momenten der konkreten Stichprobe $x_1, \ldots, x_n$.

Modellannahmen

- Es gelte $\mathbb{E}_\theta(|X_1|^r) < \infty$ für jedes $\theta \in \Theta$ und für eine natürliche Zahl $r \geq m$.
- Außerdem wird angenommen, dass für jedes $k = 1, \ldots, r$ das k-te Moment
$$\mu_k = \mathbb{E}_\theta(X_1^k) \tag{2}$$
der Stichprobenvariablen $X_1, \ldots, X_n$ eine (bekannte) Funktion des Parametervektors $\theta = (\theta_1, \ldots, \theta_m)$ ist.
- Für jedes $k = 1, \ldots, r$ gebe es also eine Borel-messbare Funktion $g_k : \Theta \to \mathbb{R}$, so dass
$$\mu_k = g_k(\theta), \qquad \forall\, \theta \in \Theta. \tag{3}$$

Lösungsansatz

1. Für jedes $k \in \{1, \ldots, r\}$ bestimmen wir das k-te *empirische Moment* $\widehat{m}_k = \widehat{m}_k(x_1, \ldots, x_n)$ der konkreten Stichprobe $(x_1, \ldots, x_n)$, wobei
$$\widehat{m}_k(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i^k .$$
2. Danach bilden wir das Gleichungssystem
$$\widehat{m}_k(x_1, \ldots, x_n) = g_k(\theta), \qquad k \in \{1, \ldots, r\} \tag{4}$$
mit dem unbekannten Vektor $\theta = (\theta_1, \ldots, \theta_m)$.
3. Es wird vorausgesetzt, dass dieses Gleichungssystem für jedes $(x_1, \ldots, x_n) \in \mathbb{R}^n$ eine eindeutig bestimmte Lösung $\widehat{\theta}(x_1, \ldots, x_n) \in \Theta$ besitzt, die von der konkreten Stichprobe $(x_1, \ldots, x_n)$ abhängt, und dass

4. die Abbildung $\widehat{\theta}: \mathbb{R}^n \to \Theta$ mit
$$(x_1,\ldots,x_n) \to \widehat{\theta}(x_1,\ldots,x_n)\,, \tag{5}$$
die den Stichprobenraum $\mathbb{R}^n$ in den Parameterraum $\Theta \subset \mathbb{R}^m$ abbildet, Borel-messbar ist, d.h., $\widehat{\theta}$ ist eine Stichprobenfunktion.

Definition Der durch (5) gegebene Zufallsvektor $\widehat{\theta}(X_1,\ldots,X_n)$ heißt *M-Schätzer* des Parametervektors θ, wobei in (5) die Zufallsstichprobe $(X_1,\ldots,X_n)$ anstelle der konkreten Stichprobe $(x_1,\ldots,x_n)$ eingesetzt wird.

Beachte

- Mit Hilfe des starken Gesetzes der großen Zahlen kann die folgende Begründung für die Anwendung der Momentenmethode gegeben werden.
- Für das empirische Moment $\widehat{m}_k(X_1,\ldots,X_n)$ der Zufallsstichprobe $(X_1,\ldots,X_n)$ gilt
$$\widehat{m}_k(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow{\text{f.s.}} \mu_k\,, \qquad \forall\, k=1,\ldots,r\,, \tag{6}$$
falls $n\to\infty$; vgl. Theorem WR–5.15.
- Falls die in (3) gegebene Abbildung $g=(g_1,\ldots,g_m):\Theta\to C$ mit $C=g(\Theta)=\{g(\theta):\theta\in\Theta\}\subset\mathbb{R}^m$ eineindeutig ist und falls die Umkehrabbildung $g^{-1}: C\to\Theta$ stetig ist, dann gilt wegen (6) für den M-Schätzer $\widehat{\theta}(X_1,\ldots,X_n)$, dass für jedes $\theta\in\Theta$ mit Wahrscheinlichkeit 1
$$\lim_{n\to\infty}\widehat{\theta}(X_1,\ldots,X_n)=\theta\,. \tag{7}$$
- Diese (asymptotische) Güteeigenschaft des Schätzers $\widehat{\theta}(X_1,\ldots,X_n)$ wird *starke Konsistenz* genannt.
- Weitere Güteeigenschaften von M-Schätzern werden wir in Abschnitten 2.3 und 2.4 diskutieren.

Beispiele

1. *Normalverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\,\theta\in\Theta\}=\{\mathrm{N}(\mu,\sigma^2),\,\mu\in\mathbb{R},\,\sigma^2>0\}$.
   - Dabei nehmen wir an, dass beide Komponenten μ und σ^2 des Vektors (μ,σ^2) unbekannt sind.
   - Dann ist $m=2$ mit $\Theta=\mathbb{R}\times(0,\infty)$ und $\theta=(\theta_1,\theta_2)=(\mu,\sigma^2)$.
   - Außerdem gilt $g_1(\mu,\sigma^2)=\mu$ und $g_2(\mu,\sigma^2)=\sigma^2+\mu^2$.
   - Das Gleichungssystem (4) hat also die Form
   $$\frac{1}{n}\sum_{i=1}^n x_i=\mu\,,\qquad \frac{1}{n}\sum_{i=1}^n x_i^2=\sigma^2+\mu^2\,.$$
   - Hieraus ergibt sich die Lösung $\widehat{\theta}=(\widehat{\mu},\widehat{\sigma}^2)$ mit
   $$\widehat{\mu}=\frac{1}{n}\sum_{i=1}^n x_i$$
   und
   $$\begin{aligned}\widehat{\sigma}^2 &= \Bigl(\frac{1}{n}\sum_{i=1}^n x_i^2\Bigr)-\Bigl(\frac{1}{n}\sum_{i=1}^n x_i\Bigr)^2\\ &= \frac{1}{n}\Bigl(\Bigl(\sum_{i=1}^n x_i^2\Bigr)-n\Bigl(\frac{1}{n}\sum_{i=1}^n x_i\Bigr)^2\Bigr)=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x}_n)^2\,,\end{aligned}$$
   wobei sich die letzte Gleichheit aus der Darstellungsformel (1.15) der Stichprobenvarianz ergibt, die in Abschnitt 1.2.2 hergeleitet wurde.

- Bei normalverteilten Stichprobenvariablen $X_1, \ldots, X_n$ ergeben sich somit die M-Schätzer

$$\widehat{\mu}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{8}$$

und

$$\widehat{\sigma}^2(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \bigl(X_i - \overline{X}_n\bigr)^2 \tag{9}$$

für die unbekannten Modellparameter μ bzw. σ^2.

2. *Binomialverteilte Stichprobenvariablen*
   - Es gelte nun $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Bin}(k,p),\, k \in \mathbb{N},\, p \in [0,1]\}$.
   - Dabei nehmen wir erneut an, dass *beide* Komponenten k und p des Parametervektors (k,p) unbekannt sind.
   - Dann ist $m = 2$ und $\Theta = \mathbb{N} \times [0,1]$ mit $\theta = (\theta_1, \theta_2) = (k,p)$.
   - Außerdem ist

$$g_1(k,p) = kp \qquad \text{und} \qquad g_2(k,p) = kp(1-p) + k^2p^2\,.$$

   - Das Gleichungssystem (4) hat also die Form

$$\frac{1}{n} \sum_{i=1}^{n} x_i = kp\,, \qquad \frac{1}{n} \sum_{i=1}^{n} x_i^2 = kp(1-p) + k^2p^2\,.$$

   - Durch Einsetzen der ersten Gleichung in die zweite Gleichung ergibt sich, dass

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 = \overline{x}_n(1-p) + \overline{x}_n^2\,.$$

   - Falls nicht sämtliche Stichprobenwerte $x_1, \ldots, x_n$ gleich Null sind, dann ergibt sich hieraus die Lösung $\widehat{\theta} = (\,\widehat{k},\, \widehat{p})$ mit

$$\widehat{k} = \frac{\overline{x}_n}{\widehat{p}}$$

und

$$\widehat{p} = \frac{\overline{x}_n - \dfrac{1}{n}\Bigl(\displaystyle\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Bigr)}{\overline{x}_n} = \frac{\overline{x}_n - \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\bigl(x_i - \overline{x}_n\bigr)^2}{\overline{x}_n}\,,$$

wobei sich die letzte Gleichheit erneut aus der Darstellungsformel (1.15) der Stichprobenvarianz ergibt.

   - Bei binomialverteilten Stichprobenvariablen $X_1, \ldots, X_n$ ergeben sich also die M-Schätzer

$$\widehat{k}(X_1, \ldots, X_n) = \frac{\overline{X}_n^2}{\overline{X}_n - \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2} \tag{10}$$

und

$$\widehat{p}(X_1, \ldots, X_n) = \frac{\overline{X}_n}{\widehat{k}(X_1, \ldots, X_n)} \tag{11}$$

für die unbekannten Modellparameter k bzw. p, falls $(X_1, \ldots, X_n) \neq (0, \ldots, 0)$.

3. *Gammaverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\Gamma(b,p),\, b > 0,\, p > 0\}$.

- Dann ist $m = 2$ und $\Theta = (0,\infty)^2$ mit $\theta = (\theta_1,\theta_2) = (b,p)$.
- Aus Theorem 1.5 folgt, dass
$$\mu_1 = \frac{p}{b}\,, \qquad \mu_2 = \frac{p(p+1)}{b^2}\,.$$
- Hieraus ergibt sich das Geichungssystem
$$\widehat{m}_1 = \frac{p}{b}\,, \qquad \widehat{m}_2 = \frac{p(p+1)}{b^2}$$
mit der Lösung
$$\widehat{b} = \frac{\widehat{m}_1}{\widehat{m}_2 - \big(\widehat{m}_1\big)^2}\,, \qquad \widehat{p} = \frac{\big(\widehat{m}_1\big)^2}{\widehat{m}_2 - \big(\widehat{m}_1\big)^2}\,.$$

Beachte

- Es gibt Beispiele parametrischer Verteilungsfamilien, so dass das Gleichungssystem (4) für $r = m$ nicht eindeutig lösbar ist, für $r > m$ jedoch eine eindeutig bestimmte Lösung besitzt.
- D.h., bei der Anwendung der Momentenmethode kann die Anzahl der zu betrachtenden Momente $\mu_1,\ldots,\mu_r$ größer als die Anzahl der (unbekannten) Parameterkomponenten $\theta_1,\ldots,\theta_m$ sein, vgl. die Übungsaufgabe 6.2.b.

2.2.2 Maximum-Likelihood-Schätzer

- Eine andere Methode zur Gewinnung von Schätzern für die unbekannten Komponenten des Parametervektors $\theta = (\theta_1,\ldots,\theta_m)$ ist die Maximum-Likelihood-Methode.
- Genauso wie bei der Momentenmethode wird auch bei der Maximum-Likelihood-Methode das Ziel verfolgt, θ so zu schätzen, dass eine möglichst gute Anpassung der Modellverteilung P_θ bzw. der Verteilungsfunktion F_θ an die beobachteten Daten $x_1,\ldots,x_n$ erreicht wird.

Wir betrachten hier nur die beiden (grundlegenden) Fälle, dass die Stichprobenvariablen $X_1,\ldots,X_n$ entweder diskret oder absolutstetig sind. D.h., für jedes $\theta \in \Theta$ gelte entweder

- $P_\theta(X_1 \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$, wobei wir mit $\{p(x;\theta),\ x \in C\}$ die Wahrscheinlichkeitsfunktion von X_1 bezeichnen, d.h. $p(x;\theta) = P_\theta(X_1 = x)$ für jedes $x \in \mathbb{R}$, oder
- $P_\theta(B) = \int\limits_B f(y;\theta)\,dy$ für jedes $B \in \mathcal{B}(\mathbb{R})$, wobei $f(\cdot\,;\theta)$ die Dichte von P_θ ist.

Dabei wird bei der Maximum-Likelihood-Methode der Parametervektor θ so gewählt, dass

- im diskreten Fall die Wahrscheinlichkeit $P_\theta(X_1 = x_1,\ldots,X_n = x_n) = p(x_1;\theta)\ldots p(x_n;\theta)$ des Ereignisses $\{X_1 = x_1,\ldots,X_n = x_n\}$ bzw.
- im absolutstetigen Fall die „infinitesimale" Wahrscheinlichkeit
$$P_\theta(X_1 \in (x_1,x_1+dx_1),\ldots,X_n \in (x_n,x_n+dx_n)) = f(x_1;\theta)\ldots f(x_n;\theta)\,dx_1\ldots dx_n$$
maximiert wird.

Die Maximum-Likelihood-Methode wurde bereits im Jahre 1821 von Carl Friedrich Gauss (1777–1855) erwähnt. Sir Ronald Aylmer Fisher (1890–1962) hat diese Methode dann im Jahre 1922 wiederentdeckt und mit der Untersuchung ihrer Eigenschaften begonnen.

Definition Die Abbildung $L : \mathbb{R}^n \times \Theta \to [0,\infty)$ sei durch die folgende Vorschrift gegeben.

- Falls P_θ diskret ist, dann sei

$$L(x_1,\ldots,x_n;\theta)=p(x_1;\theta)\ldots p(x_n;\theta)\,,\qquad \forall(x_1,\ldots,x_n)\in\mathbb{R}^n\,. \tag{12}$$

- Falls P_θ absolutstetig ist, dann sei

$$L(x_1,\ldots,x_n;\theta)=f(x_1;\theta)\ldots f(x_n;\theta)\,,\qquad \forall(x_1,\ldots,x_n)\in\mathbb{R}^n\,. \tag{13}$$

Für jeden Vektor $(x_1,\ldots,x_n)\in\mathbb{R}^n$ heißt die Abbildung $\theta\to L(x_1,\ldots,x_n;\theta)$, die den Parameterraum Θ nach $[0,\infty)$ abbildet, die *Likelihood-Funktion* der Stichprobe $(x_1,\ldots,x_n)$.

Die Idee der Maximum-Likelihood-Methode besteht nun darin, für jede (konkrete) Stichprobe $(x_1,\ldots,x_n)$ einen Parametervektor $\theta\in\Theta$ zu bestimmen, so dass der Wert $L(x_1,\ldots,x_n;\theta)$ der Likelihood-Funktion möglichst groß wird. Dies führt zu der folgenden Begriffsbildung.

Definition Sei $\widehat{\theta}:\mathbb{R}^n\to\Theta\subset\mathbb{R}^m$ eine Stichprobenfunktion mit

$$L(x_1,\ldots,x_n;\theta)\le L(x_1,\ldots,x_n;\,\widehat{\theta}(x_1,\ldots,x_n))\,,\qquad \forall(x_1,\ldots,x_n)\in\mathbb{R}^n,\ \theta\in\Theta\,. \tag{14}$$

Der Zufallsvektor $\widehat{\theta}(X_1,\ldots,X_n)$ wird dann *Maximum-Likelihood-Schätzer* für θ (bzw. kurz: ML-Schätzer) genannt.

Beachte

- Manchmal ist der in (14) definierte Maximum-Likelihood-Schätzer $\widehat{\theta}(X_1,\ldots,X_n)$ nicht eindeutig bestimmt bzw. die Likelihood-Funktion $L(x_1,\ldots,x_n;\theta)$ ist zu kompliziert, so dass sich das Optimierungsproblem (14) nicht analytisch lösen lässt.
- Dann muss man auf numerische Algorithmen zurückgreifen.
- Es gibt aber auch eine Reihe von (einfachen) Beispielen parametrischer Verteilungsfamilien, für die sich das Optimierungsproblem (14) analytisch (durch Differenzieren) lösen lässt.
- Anstelle (14) wird dabei oft die (äquivalente) Bedingung

$$\log L(x_1,\ldots,x_n;\theta)\le\log L(x_1,\ldots,x_n;\,\widehat{\theta}(x_1,\ldots,x_n))\,,\qquad \forall(x_1,\ldots,x_n)\in\mathbb{R}^n,\ \theta\in\Theta \tag{15}$$

  betrachtet.
- Für jeden Vektor $(x_1,\ldots,x_n)\in\mathbb{R}^n$ wird die Abbildung $\theta\to\log L(x_1,\ldots,x_n;\theta)$ die *Loglikelihood-Funktion* der Stichprobe $(x_1,\ldots,x_n)$ genannt.
- Sie bietet den Vorteil, dass beim Übergang zur Loglikelihood-Funktion die Produkte in (12) bzw. (13) in (einfacher handhabbare) Summen übergehen.

Beispiele

1. *Wer war vermutlich der Absender?*
   - Eine Warenlieferung eines unbekannten Herstellers bestehe aus 12 Exemplaren eines Artikels.
   - Dabei sei festgestellt worden, dass eines der 12 Exemplare Ausschuss ist.
   - Es sei bekannt, dass nur drei potentiell mögliche Hersteller in Frage kommen und dass deren Lieferungen erfahrungsgemäß jeweils einen Ausschussanteil von $\theta_1=0,05$, $\theta_2=0,10$ bzw. $\theta_3=0,15$ aufweisen.
   - *Frage*: Welcher der drei Hersteller war vermutlich der Absender der Warenlieferung?

- *Modell*: Betrachten die Stichprobenvariablen $X_1, \ldots, X_{12}$ mit

$$X_i(\omega) = \begin{cases} 1\,, & \text{falls das } i\text{-te Exemplar der Lieferung Ausschuss ist,} \\ 0\,, & \text{sonst} \end{cases}$$

  und die Familie der drei Bernoulli-Verteilungen $\{\mathrm{Bin}(1,\theta_1),\mathrm{Bin}(1,\theta_2),\mathrm{Bin}(1,\theta_3)\}$, d.h. $m = 1$ und $\Theta = \{\theta_1, \theta_2, \theta_3\}$.
- *Lösung*: Die Stichprobenfunktion

$$\widehat{\theta} : \{0,1\}^{12} \to \{\theta_1, \theta_2, \theta_3\}$$

  wird so gewählt, dass für jeden Vektor $(x_1, \ldots, x_{12}) \in \{0,1\}^{12}$ mit $\#\{i : x_i = 1\} = 1$ die Wahrscheinlichkeit

$$P_\theta((X_1, \ldots, X_{12}) = (x_1, \ldots, x_{12})) = \theta(1-\theta)^{11}$$

  maximal ist.
- Es gilt

θ	$P_\theta((X_1, \ldots, X_{12}) = (x_1, \ldots, x_{12}))$
$0,05$	$0,028$
$0,10$	$0,031$
$0,15$	$0,025$

- Das Maximum $0,031$ steht in der zweiten Zeile dieser Tabelle.
- Also ist $\widehat{\theta}(x_1, \ldots, x_{12}) = \theta_2$ für jeden Vektor $(x_1, \ldots, x_{12}) \in \{0,1\}^{12}$ mit $\#\{i : x_i = 1\} = 1$, d.h., der Hersteller mit dem Ausschussanteil $\theta_2 = 0.10$ war vermutlich der Absender der Lieferung.

2. *Bernoulli-verteilte Stichprobenvariablen* (Fortsetzung)
   - Betrachten die Familie $\{P_\theta,\ \theta \in \Theta\} = \{\mathrm{Bin}(1,p),\ p \in [0,1]\}$ der Bernoulli-Verteilungen.
   - Dann gilt

$$p(x;p) = \begin{cases} p^x(1-p)^{1-x}\,, & \text{falls } x \in \{0,1\}, \\ 0\,, & \text{sonst} \end{cases}$$

   - Die Likelihood-Funktion L ist also gegeben durch

$$L(x_1, \ldots, x_n; p) = \begin{cases} \prod\limits_{i=1}^n p^{x_i}(1-p)^{1-x_i}\,, & \text{falls } (x_1, \ldots, x_n) \in \{0,1\}^n, \\ 0\,, & \text{sonst} \end{cases}$$

   - Falls $x_1 = \ldots = x_n = 0$ bzw. $x_1 = \ldots = x_n = 1$, dann sieht man leicht, dass die Abbildung $p \to L(x_1, \ldots, x_n; p)$ an der Stelle $p = 0$ bzw. $p = 1$ ein (eindeutig bestimmtes) Maximum hat.
   - Sei nun $(x_1, \ldots, x_n) \in \{0,1\}^n$ mit $0 < \sum\limits_{i=1}^n x_i < n$. Dann ist

$$p \to \log L(x_1, \ldots, x_n; p) = \Bigl(\sum_{i=1}^n x_i\Bigr) \log p + \Bigl(n - \sum_{i=1}^n x_i\Bigr) \log(1-p)$$

     eine stetige Funktion im Intervall $(0,1)$, und es gilt

$$\lim_{p\to 0} \log L(x_1, \ldots, x_n; p) = -\infty \qquad \text{bzw.} \qquad \lim_{p \to 1} \log L(x_1, \ldots, x_n; p) = -\infty\,.$$

- Die Abbildung $p \to \log L(x_1, \ldots, x_n; p)$ hat also ein Maximum im Intervall $(0, 1)$.
- Durch Differenzieren nach p ergibt sich

$$\frac{\partial \log L(x_1, \ldots, x_n; p)}{\partial p} = \Bigl(\sum_{i=1}^n x_i\Bigr)\frac{1}{p} - \Bigl(n - \sum_{i=1}^n x_i\Bigr)\frac{1}{1-p}\,.$$

- Weil die Gleichung

$$\Bigl(\sum_{i=1}^n x_i\Bigr)\frac{1}{p} - \Bigl(n - \sum_{i=1}^n x_i\Bigr)\frac{1}{1-p} = 0$$

die (eindeutig bestimmte) Lösung

$$\widehat{p}(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i \qquad \Bigl(= \overline{x}_n\Bigr)$$

hat, nimmt die Abbildung $p \to \log L(x_1, \ldots, x_n; p)$ an der Stelle $p = \overline{x}_n$ ihr Maximum an.
- Also ist der Maximum-Likelihood-Schätzer für den Parameter p gegeben durch

$$\widehat{p}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i \qquad \Bigl(= \overline{X}_n\Bigr)\,.$$

3. *Binomialverteilte Stichprobenvariablen*
   - Für eine beliebige, jedoch vorgegebene (d.h. bekannte) natürliche Zahl $n_0 \geq 1$ betrachten wir nun die Familie $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Bin}(n_0, p),\, p \in [0, 1]\}$ von Binomialerteilungen.
   - Dann gilt

$$p(x; p) = \begin{cases} \dbinom{n_0}{x} p^x (1-p)^{n_0 - x}\,, & \text{falls } x \in \{0, 1, \ldots, n_0\}, \\ 0\,, & \text{sonst.} \end{cases}$$

   - Genauso wie in Beispiel 2 ergibt sich der Maximum-Likelihood-Schätzer

$$\widehat{p}(X_1, \ldots, X_n) = \frac{\overline{X}_n}{n_0}$$

   für den (unbekannten) Parameter p.

4. *Poisson-verteilte Stichprobenvariablen*
   - Betrachten die Familie $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Poi}(\lambda),\, \lambda \geq 0\}$ der Poisson-Verteilungen.
   - Dann gilt

$$p(x; \lambda) = \begin{cases} \dfrac{\lambda^x}{x\,!}\, e^{-\lambda}\,, & \text{falls } x \in \{0, 1, \ldots\}, \\ 0\,, & \text{sonst.} \end{cases}$$

   - Auf die gleiche Weise wie in den Beispielen 2 und 3 ergibt sich der Maximum-Likelihood-Schätzer

$$\widehat{\lambda}(X_1, \ldots, X_n) = \overline{X}_n$$

   für den Parameter λ.

5. *Normalverteilte Stichprobenvariablen*
   - Betrachten nun die Familie $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{N}(\mu, \sigma^2),\, \mu \in \mathbb{R}, \sigma^2 > 0\}$ der Normalverteilungen.

- Dann gilt
$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\Bigl(-\frac{1}{2}\Bigl(\frac{x-\mu}{\sigma}\Bigr)^2\Bigr).$$
- Die Likelihood-Funktion L ist somit gegeben durch
$$L(x_1,\ldots,x_n;\mu,\sigma^2)=\Bigl(\frac{1}{\sqrt{2\pi}\sigma}\Bigr)^n\exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\Bigr).$$
- Für die Loglikelihood-Funktion gilt
$$\log L(x_1,\ldots,x_n;\mu,\sigma^2)=-n\log(\sqrt{2\pi}\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2.$$
- Durch Differenzieren nach μ ergibt sich
$$\frac{\partial\log L(x_1,\ldots,x_n;\mu,\sigma^2)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu)\qquad\text{bzw.}\qquad\frac{\partial^2\log L(x_1,\ldots,x_n;\mu,\sigma^2)}{\partial^2\mu}=-\frac{n}{\sigma^2}<0.$$
- Für jedes (fest vorgegebene) $\sigma^2>0$ nimmt also die Abbildung
$$\mu\to\log L(x_1,\ldots,x_n;\mu,\sigma^2)$$
ihr Maximum an der Stelle $\mu=\overline{x}_n$ an.
- Es ist nun noch das Maximum der Abbildung
$$\sigma^2\to\log L(x_1,\ldots,x_n;\overline{x}_n,\sigma^2)\tag{16}$$
zu bestimmen.
- Weil $P(X_1=\ldots=X_n)=0$ gilt, können wir annehmen, dass nicht alle Stichprobenwerte $x_1,\ldots,x_n$ gleich sind.
- Beachte: Die Abbildung (16) ist stetig für alle $\sigma^2>0$, und es gilt
$$\lim_{\sigma^2\to0}\log L(x_1,\ldots,x_n;\overline{x}_n,\sigma^2)=-\infty\qquad\text{bzw.}\qquad\lim_{\sigma^2\to\infty}\log L(x_1,\ldots,x_n;\overline{x}_n,\sigma^2)=-\infty.$$
- Die Abbildung (16) hat also ein Maximum im Intervall $(0,\infty)$.
- Durch Differenzieren nach σ^2 ergibt sich
$$\frac{\partial\log L(x_1,\ldots,x_n;\overline{x}_n,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n(x_i-\overline{x}_n)^2.$$
- Weil vorausgesetzt wird, dass nicht alle Stichprobenwerte $x_1,\ldots,x_n$ gleich sind, gilt
$$\sum_{i=1}^n(x_i-\overline{x}_n)^2>0.$$
- Deshalb hat die Gleichung
$$-\frac{n}{2\,\widehat{\sigma}^2}+\frac{1}{2\,\widehat{\sigma}^4}\sum_{i=1}^n(x_i-\overline{x}_n)^2=0$$
die (eindeutig bestimmte) Lösung
$$\widehat{\sigma}^2(x_1,\ldots,x_n)=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x}_n)^2.$$

- Hieraus ergeben sich die Maximum-Likelihood-Schätzer

$$\widehat{\mu}(X_1,\ldots,X_n)=\overline{X}_n\,,\qquad \widehat{\sigma}^2(X_1,\ldots,X_n)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

  für die Parameter μ und σ^2.

6. *Gleichverteilte Stichprobenvariablen*
   - Betrachten die Familie $\{P_\theta,\,\theta\in\Theta\}=\{\mathrm{U}(0,b),\,b>0\}$ von Gleichverteilungen.
   - Dann gilt

$$f(x;b)=\begin{cases}\dfrac{1}{b}\,, & \text{falls } 0\le x\le b,\\ 0\,, & \text{sonst.}\end{cases}$$

   - Die Likelihood-Funktion L ist somit gegeben durch

$$L(x_1,\ldots,x_n;b)=\begin{cases}\dfrac{1}{b^n}\,, & \text{falls } 0\le x_1,\ldots,x_n\le b,\\ 0\,, & \text{sonst.}\end{cases}$$

   - Weil die Abbildung $b\to L(x_1,\ldots,x_n;b)$ monoton fallend ist für $b>\max\{x_1,\ldots,x_n\}\ge 0$, ergibt sich der Maximum-Likelihood-Schätzer

$$\widehat{b}(X_1,\ldots,X_n)=\max\{X_1,\ldots,X_n\}$$

   für den Parameter b.

2.2.3 Bayes-Schätzer

- Zur Erinnerung: Bei der Momenten-Methode in Abschnitt 2.2.1 und auch bei der Maximum-Likelihood-Methode in Abschnitt 2.2.2 betrachteten wir das parametrische Modell, das in Abschnitt 2.1 eingeführt worden ist.
- Dabei setzten wir voraus, dass die Verteilungsfunktion F der Stichprobenvariablen $X_1,\ldots,X_n$ zu einer vorgegebenen (d.h. bekannten) parametrischen Familie von Verteilungsfunktionen $\{F_\theta,\,\theta\in\Theta\}$ gehört.
- Mit anderen Worten: Es wurde angenommen, dass $F=F_\theta$ für *ein* $\theta\in\Theta$, wobei der Parametervektor $\theta=(\theta_1,\ldots,\theta_m)$ (bzw. ein Teil seiner Komponenten) unbekannt sei und aus den beobachteten Daten $x_1,\ldots,x_n$ geschätzt werden soll.
- Wir modifizieren nun dieses parametrische Modell folgendermaßen.
- Anstelle der bisherigen Modellannahme, dass der Parameter θ ein zwar unbekannter, jedoch deterministischer Vektor ist, setzen wir jetzt voraus, dass der Parameter selbst ein Zufallsvektor ist, den wir mit $\widetilde{\theta}:\Omega\to\Theta$ bezeichnen.
- Dabei nehmen wir an, dass wir bereits (vor der Erhebung von Daten) eine gewisse Vorkenntnis über die Verteilung von $\widetilde{\theta}$ besitzen.
- Diese Vorkenntnis modellieren wir durch eine Verteilung $Q:\mathcal{B}(\Theta)\to[0,1]$ über dem Parameterraum $(\Theta,\mathcal{B}(\Theta))$, die wir *a-priori-Verteilung* von $\widetilde{\theta}$ nennen.

Beachte

- Im folgenden werden wir nur den Fall betrachten, dass die a-priori Verteilung Q von $\widetilde{\theta}$ entweder diskret oder absolutstetig ist, wobei $q : \Theta \to [0,\infty)$ entweder die Wahrscheinlichkeitsfunktion von Q ist, d.h.,
$$\sum_{\theta \in C} q(\theta) = 1$$
für eine abzählbare Menge $C \subset \Theta$, oder die Dichte von Q ist, d.h.
$$Q(B) = \int_B q(\theta)\, d\theta\,, \qquad \forall B \in \mathcal{B}(\Theta)\,.$$
- Die Familie $\{P_\theta,\, \theta \in \Theta\}$ der (potentiell möglichen) Verteilungen der Stichprobenvariablen $X_1, \ldots, X_n$ kann ebenfalls (so wie bisher stets angenommen) entweder eine Familie diskreter Verteilungen oder eine Familie absolutstetiger Verteilungen sein.

Definition (*a-posteriori-Verteilung*)

- Die in Abschnitt 2.2.2 eingeführte Likelihood-Funktion $L : \mathbb{R}^n \times \Theta \to [0,\infty)$ mit
$$L(x_1, \ldots, x_n; \theta) = \begin{cases} p(x_1;\theta)\ldots p(x_n;\theta) & \text{im diskreten Fall,} \\ f(x_1;\theta)\ldots f(x_n;\theta) & \text{im absolutstetigen Fall} \end{cases} \tag{17}$$
sei Borel-messbar in sämtlichen $n+m$ Argumenten $x_1, \ldots, x_n$ und $\theta = (\theta_1, \ldots, \theta_m)$, weil jetzt auch θ als Variable aufgefasst wird.
- Für jeden (Daten-) Vektor $(x_1, \ldots, x_n) \in \mathbb{R}^n$, für den $f(x_1, \ldots, x_n) > 0$ gilt, wobei
$$f(x_1, \ldots, x_n) = \begin{cases} \sum\limits_{\theta \in C} L(x_1, \ldots, x_n; \theta)\, q(\theta)\,, & \text{falls } Q \text{ diskret,} \\ \int\limits_{\Theta} L(x_1, \ldots, x_n; \theta)\, q(\theta)\, d\theta\,, & \text{falls } Q \text{ absolutstetig,} \end{cases} \tag{18}$$
heißt dann die Funktion $q_{(x_1,\ldots,x_n)} : \Theta \to [0,\infty)$ mit
$$q_{(x_1,\ldots,x_n)}(\theta) = \frac{L(x_1, \ldots, x_n; \theta)\, q(\theta)}{f(x_1, \ldots, x_n)} \tag{19}$$
a-posteriori-Wahrscheinlichkeitsfunktion bzw. *a-posteriori-Dichte* von $\widetilde{\theta}$; bei Vorliegen der (konkreten) Stichprobe $(x_1, \ldots, x_n)$.

Beachte

Die Funktion $q_{(x_1,\ldots,x_n)} : \Theta \to [0,\infty)$ kann als bedingte Wahrscheinlichkeitsfunktion bzw. bedingte Dichte von $\widetilde{\theta}$ angesehen werden, unter der Bedingung, dass $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$, denn

- die Likelihood-Funktion $L(x_1, \ldots, x_n; \theta)$ können wir als bedingte Wahrscheinlichkeitsfunktion bzw. bedingte Dichte von $(X_1, \ldots, X_n)$ ansehen, unter der Bedingung, dass $\widetilde{\theta} = \theta$,
- das Produkt $L(x_1, \ldots, x_n; \theta)\, q(\theta)$ als die gemeinsame (unbedingte) Wahrscheinlichkeitsfunktion bzw. Dichte des Zufallsvektors $(X_1, \ldots, X_n, \widetilde{\theta})$, und
- die in (18) gegebene Funktion $f(x_1, \ldots, x_n)$ als die (Rand-) Wahrscheinlichkeitsfunktion bzw. Dichte von $(X_1, \ldots, X_n)$.

Definition (*Bayes-Schätzer*)

- Die Funktion $q_{(x_1,\ldots,x_n)} : \Theta \to [0,\infty)$ kann nun zur Konstruktion von Punktschätzern für den unbekannten Parameter genutzt werden.
- Dabei verallgemeinern wir den in Abschnitt 2.1 eingeführten Begriff des Punktschätzers, indem wir den Zufallsvektor $\varphi(X_1,\ldots,X_n)$ für *jede* Borel-messbare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ als Punktschätzer für θ auffassen, wobei nicht notwendigerweise $\varphi(\mathbb{R}^n) \subset \Theta$ gelten muss.
- Falls $m = 1$, dann kann beispielsweise der Erwartungswert

$$\widehat{\theta}(x_1,\ldots,x_n) = \begin{cases} \sum\limits_{\theta\in C} \theta\, q_{(x_1,\ldots,x_n)}(\theta)\,, & \text{falls } Q \text{ diskret,} \\ \int\limits_{\Theta} \theta\, q_{(x_1,\ldots,x_n)}(\theta)\, d\theta\,, & \text{falls } Q \text{ absolutstetig,} \end{cases} \tag{20}$$

  betrachtet und $\widehat{\theta}(X_1,\ldots,X_n)$ als *Bayes-Schätzer* des unbekannten Parameters angesehen werden.

Man kann zeigen (vgl. Übungsaufgabe WR–8.3), dass der in (20) gegebene Schätzwert $\widehat{\theta}(x_1,\ldots,x_n)$ den *mittleren quadratischen Fehler*

$$g_{(x_1,\ldots,x_n)}(a) = \begin{cases} \sum\limits_{\theta\in C} (\theta-a)^2\, q_{(x_1,\ldots,x_n)}(\theta)\,, & \text{falls } Q \text{ diskret,} \\ \int\limits_{\Theta} (\theta-a)^2\, q_{(x_1,\ldots,x_n)}(\theta)\, d\theta\,, & \text{falls } Q \text{ absolutstetig,} \end{cases}$$

minimiert, d.h., es gilt $g_{(x_1,\ldots,x_n)}(a) \ge g_{(x_1,\ldots,x_n)}(\widehat{\theta}(x_1,\ldots,x_n))$ für jedes $a \in \mathbb{R}$, vgl. auch Abschnitt 2.3.1.

Beachte

- Sei $\varphi : \mathbb{R}^n \to \mathbb{R}$ eine beliebige Stichprobenfunktion.
- Anstelle der in (19) gegebenen a-posteriori Wahrscheinlichkeitsfunktion bzw. Dichte $q_{(x_1,\ldots,x_n)}$ von $\widetilde{\theta}$ betrachtet man dann auch die a-posteriori Wahrscheinlichkeitsfunktion bzw. Dichte $q_y : \Theta \to [0,\infty)$ von $\widetilde{\theta}$ unter der Bedingung, dass $\varphi(X_1,\ldots,X_n) = y$ für $y \in \mathbb{R}$.
- Dabei ist

$$q_y(\theta) = \frac{f_{\theta,\varphi(X_1,\ldots,X_n)}(y)\, q(\theta)}{f_{\varphi(X_1,\ldots,X_n)}(y)}\,, \tag{21}$$

  wobei $f_{\theta,\varphi(X_1,\ldots,X_n)}(y)$ die Wahrscheinlichkeitsfunktion bzw. Dichte von $\varphi(X_1,\ldots,X_n)$ für *eine vorgegebene* Verteilung P_θ der Stichprobenvariablen $X_1,\ldots,X_n$ bezeichnet und

$$f_{\varphi(X_1,\ldots,X_n)}(y) = \begin{cases} \sum\limits_{\theta\in C} f_{\theta,\varphi(X_1,\ldots,X_n)}(y)\, q(\theta)\,, & \text{falls } Q \text{ diskret,} \\ \int\limits_{\Theta} f_{\theta,\varphi(X_1,\ldots,X_n)}(y)\, q(\theta)\, d\theta\,, & \text{falls } Q \text{ absolutstetig,} \end{cases}$$

  die „unbedingte" Wahrscheinlichkeitsfunktion bzw. Dichte von $\varphi(X_1,\ldots,X_n)$ ist.

Beispiel (*Bernoulli-verteilte Stichprobenvariablen*)

- Betrachten die Familie $\{P_\theta,\ \theta\in\Theta\} = \{\mathrm{Bin}(1,p),\ p\in[0,1]\}$ der Bernoulli-Verteilungen.
- Die a-priori-Verteilung von $\widetilde{p} : \Omega \to [0,1]$ sei die *Betaverteilung* $\mathrm{Beta}(\alpha,\beta)$, wobei die Parameter $\alpha,\beta > 0$ vorgegeben seien.
- Die a-priori-Dichte $q : [0,1] \to (0,\infty)$ von $\widetilde{p}$ ist also gegeben durch

$$q(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,, \qquad \forall\, p\in[0,1]\,. \tag{22}$$

- Wir betrachten die Stichprobenfunktion $Y = \sum_{i=1}^n X_i$, d.h., Y ist die Anzahl der „Erfolge” bei n Versuchen.
- Dann gilt für jedes $p \in [0,1]$
$$f_{p,Y}(y) = \binom{n}{y} p^y (1-p)^{n-y}\,, \qquad \forall\, y \in \{0,1,\ldots,n\}$$
und somit
$$f_Y(y) = \int_0^1 f_{p,Y}(y) q(p)\, dp = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}\,, \qquad \forall\, y \in \{0,1,\ldots,n\}\,.$$
- Für die a-posteriori Dichte $q_y : [0,1] \to [0,\infty)$ von $\widetilde{p}$ ergibt sich hieraus, dass
$$q_y(p) = \frac{f_{p,Y}(y)q(p)}{f_Y(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\, p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$
für beliebige $y \in \{0,1,\ldots,n\}$ und $p \in [0,1]$.
- Dies ist die Dichte der Betaverteilung $\mathrm{Beta}(y+\alpha, n-y+\beta)$.
- Für den Erwartungswert $\widetilde{p}(y)$ dieser Verteilung gilt
$$\widetilde{p}(y) = \int_0^1 p\, q_y(p)\, dp = \frac{y+\alpha}{\alpha+\beta+n}\,.$$
- Dies ergibt den Bayes-Schätzer $\widetilde{p}(Y)$ mit
$$\widetilde{p}(Y) = \frac{Y+\alpha}{\alpha+\beta+n}$$
bzw.
$$\widetilde{p}(Y) = \frac{n}{\alpha+\beta+n}\,\frac{Y}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\,\frac{\alpha}{\alpha+\beta}\,. \tag{23}$$
- Der Bayes-Schätzer $\widetilde{p}(Y)$ ist also eine Linearkombination
  1. des Erwartungswertes $\alpha/(\alpha+\beta)$ der a-priori-Verteilung von $\widetilde{p}$, der ein natürlicher Schätzer wäre, wenn wir über keinerlei Daten verfügen würden, und
  2. des Stichprobenmittels Y/n, das ein natürlicher Schätzer wäre, wenn keine a-priori-Verteilung von $\widetilde{p}$ zur Verfügung stehen würde.

2.3 Güteeigenschaften von Punktschätzern

- In Abschnitt 2.2 diskutierten wir verschiedene Methoden, die ein systematisches Vorgehen bei der Wahl einer Borel-messbaren Abbildung $\widehat{\theta} : \mathbb{R}^n \to \Theta$ ermöglichen, um den unbekannten Parametervektor θ zu schätzen.
- Es kann also vorkommen, dass für einunddassselbe Schätzproblem verschiedene Punktschätzer zur Verfügung stehen.
- Deshalb ist es nützlich, Güteeigenschaften zur Bewertung von Punktschätzern zu untersuchen.
- Wir betrachten nun eine beliebige (Borel-messbare) Stichprobenfunktion $\widehat{\theta} : \mathbb{R}^n \to \Theta \subset \mathbb{R}^m$ und den zugehörigen Schätzer $\widehat{\theta}(X_1,\ldots,X_n)$ des Parametervektors θ.
- Dabei nehmen wir so wie in den Abschnitten 2.2.1 und 2.2.2 an, dass $F = F_\theta$ für *ein* $\theta \in \Theta$, wobei der Parametervektor $\theta = (\theta_1,\ldots,\theta_m)$ (bzw. ein Teil seiner Komponenten) unbekannt sei.

2.3.1 Erwartungstreue; mittlerer quadratischer Fehler

Definition

- Es gelte
$$\mathbb{E}_\theta |\widehat{\theta}(X_1,\ldots,X_n)| < \infty\,, \qquad \forall\,\theta\in\Theta\,,$$
wobei $|\widehat{\theta}(X_1,\ldots,X_n)|$ die Länge des Zufallsvektors $\widehat{\theta}(X_1,\ldots,X_n)$ bezeichnet.
- Dann heißt der Schätzer $\widehat{\theta}(X_1,\ldots,X_n)$
  - *erwartungstreu*, falls $\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)=\theta$ für jedes $\theta\in\Theta$,
  - *asymptotisch erwartungstreu*, falls $\lim\limits_{n\to\infty}\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)=\theta$ für jedes $\theta\in\Theta$.

Beachte Die folgende Sprechweise ist üblich:

- Die *Verzerrung* bzw. der *Bias* von $\widehat{\theta}(X_1,\ldots,X_n)$ unter P_θ ist die Differenz
$$\mathrm{Bias}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)=\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)-\theta\,.$$
- Falls
$$\mathbb{E}_\theta(|\widehat{\theta}(X_1,\ldots,X_n)-\theta|^2)<\infty\,, \qquad \forall\,\theta\in\Theta\,,$$
dann heißt der Erwartungswert $\mathbb{E}_\theta(|\widehat{\theta}(X_1,\ldots,X_n)-\theta|^2)$ die *mittlere quadratische Abweichung* des Schätzers $\widehat{\theta}(X_1,\ldots,X_n)$ von dem zu schätzenden Parametervektor $\theta\in\Theta\subset\mathbb{R}^m$.
- Anstelle „mittlere quadratische Abweichung" sagen wir auch kurz *MQ-Fehler* von $\widehat{\theta}(X_1,\ldots,X_n)$.

Im Fall eines eindimensionalen Parameters $\theta\in\mathbb{R}$ kann man den MQ-Fehler von $\widehat{\theta}(X_1,\ldots,X_n)$ wie folgt zerlegen.

Theorem 2.1 *Falls $m=1$, d.h. $\Theta\subset\mathbb{R}$, dann gilt*
$$\mathbb{E}_\theta\bigl(\bigl(\widehat{\theta}(X_1,\ldots,X_n)-\theta\bigr)^2\bigr)=\mathrm{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)+\bigl(\mathrm{Bias}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\bigr)^2\,. \tag{24}$$
Falls $\widehat{\theta}(X_1,\ldots,X_n)$ erwartungstreu ist, dann gilt insbesondere
$$\mathbb{E}_\theta\bigl(\bigl(\widehat{\theta}(X_1,\ldots,X_n)-\theta\bigr)^2\bigr)=\mathrm{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\,. \tag{25}$$

Beweis

- Es gilt
$$\begin{aligned}\mathbb{E}_\theta\bigl((\widehat{\theta}(X_1,\ldots,X_n)-\theta)^2\bigr) &= \mathbb{E}_\theta\bigl((\widehat{\theta}(X_1,\ldots,X_n))^2\bigr)-2\theta\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)+\theta^2\\ &= \mathbb{E}_\theta\bigl((\widehat{\theta}(X_1,\ldots,X_n))^2\bigr)-\bigl(\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\bigr)^2\\ &\quad+\bigl(\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\bigr)^2-2\theta\mathbb{E}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)+\theta^2\\ &= \mathrm{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)+\bigl(\mathrm{Bias}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\bigr)^2\,.\end{aligned}$$
- Damit ist (24) bewiesen, und (25) ergibt sich unmittelbar aus der Definition der Erwartungstreue. □

Beachte

- Aus der Zerlegung (24) ergibt sich, dass der MQ-Fehler eines Schätzers $\widehat{\theta}(X_1,\ldots,X_n)$ genau dann klein ist, wenn sowohl die Varianz $\mathrm{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)$ als auch die Verzerrung $\mathrm{Bias}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)$ von $\widehat{\theta}(X_1,\ldots,X_n)$ klein sind.

- Um zu einem Schätzer mit einem kleinen MQ-Fehler zu gelangen, könnte man also unter den verfügbaren erwartungstreuen Schätzern denjenigen Schätzer mit der kleinsten Varianz auswählen.
- Manchmal ist es jedoch sinnvoll, den MQ-Fehler noch weiter zu verringern, indem man bei der Minimierung des MQ-Fehlers nicht nur erwartungstreue Schätzer betrachtet, sondern auch verzerrte Schätzer mit in die Betrachtung einbezieht.

Beispiele

1. *Normalverteilte Stichprobenvariablen*
   - Sei $(X_1,\ldots,X_n)$ eine normalverteilte Zufallsstichprobe mit $X_i \sim \mathrm{N}(\mu,\sigma^2)$.
   - In den Theoremen 1.1 bzw. 1.3 hatten wir gezeigt, dass $\overline{X}_n$ bzw. S_n^2 erwartungstreue Schätzer für μ bzw. σ^2 sind.
   - Außerdem hatten wir in den Theoremen 1.1 bzw. 1.3 gezeigt, dass die Varianz dieser Schätzer gegeben ist durch
   $$\mathrm{Var}_\theta \overline{X}_n = \frac{\sigma^2}{n} \qquad \text{bzw.} \qquad \mathrm{Var}_\theta(S_n^2) = \frac{1}{n}\Bigl(\mu_4' - \frac{n-3}{n-1}\,\sigma^4\Bigr)\,,$$
   wobei $\mu_4' = \mathbb{E}\bigl((X_i-\mu)^4\bigr)$; $\theta = (\mu,\sigma^2)$.
   - Dabei vereinfacht sich die letzte Formel bei normalverteilten Stichprobenvariablen, denn ähnlich wie in Übungsaufgabe 1.2 kann μ_4' durch Differenzieren der charakteristischen Funktion der $\mathrm{N}(0,\sigma^2)$-Verteilung bestimmt werden.
   - Hieraus ergibt sich, dass
   $$\mathrm{Var}_\theta(S_n^2) = \frac{2\sigma^4}{n-1}\,.$$
   - In den Abschnitten 2.2.1 bzw. 2.2.2 wurde der folgende (alternative) Schätzer für σ^2 konstruiert:
   $$\widehat{\sigma}^2(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n \bigl(X_i - \overline{X}_n\bigr)^2$$
   - Dieser Schätzer ist nicht erwartungstreu, sondern nur asymptotisch erwartungstreu.
   - Für die Varianz von $\widehat{\sigma}^2(X_1,\ldots,X_n)$ gilt jedoch:
   $$\mathrm{Var}_\theta\,\widehat{\sigma}^2(X_1,\ldots,X_n) = \Bigl(\frac{n-1}{n}\Bigr)^2 \mathrm{Var}_\theta(S_n^2) = \frac{2(n-1)\sigma^4}{n^2} < \frac{2\sigma^4}{n-1} = \mathrm{Var}_\theta(S_n^2)\,.$$
   - Aus Theorem 2.1 ergibt sich außerdem für den MQ-Fehler von $\widehat{\sigma}^2(X_1,\ldots,X_n)$ bzw. S_n^2, dass
   $$\begin{aligned}\mathbb{E}_\theta\bigl((\widehat{\sigma}^2(X_1,\ldots,X_n)-\sigma^2)^2\bigr) &= \mathrm{Var}_\theta\,\widehat{\sigma}^2(X_1,\ldots,X_n) + \bigl(\mathrm{Bias}_\theta\,\widehat{\sigma}^2(X_1,\ldots,X_n)\bigr)^2\\ &= \frac{2(n-1)\sigma^4}{n^2} + \Bigl(\frac{n-1}{n}\,\sigma^2 - \sigma^2\Bigr)^2\\ &= \frac{(2n-1)\sigma^4}{n^2}\\ &< \frac{2\sigma^4}{n-1}\\ &= \mathbb{E}_\theta\bigl((S_n^2-\sigma^2)^2\bigr)\,.\end{aligned}$$
   - Der Schätzer $\widehat{\sigma}^2(X_1,\ldots,X_n)$ für σ^2 hat also einen kleineren MQ-Fehler als S_n^2.
   - Andererseits besteht ein Nachteil des Schätzers $\widehat{\sigma}^2(X_1,\ldots,X_n)$ darin, dass $\widehat{\sigma}^2(X_1,\ldots,X_n)$ die Modellvarianz σ^2 systematisch unterschätzt.
2. *Bernoulli-verteilte Stichprobenvariablen*
   - Sei nun $(X_1,\ldots,X_n)$ eine Bernoulli-verteilte Zufallsstichprobe mit $X_i \sim \mathrm{Bin}(1,p)$.

- Der Maximum-Likelihood-Schätzer
$$\widehat{p}(X_1,\ldots,X_n)=\overline{X}_n$$
für den Parameter p, den wir in Abschnitt 2.2.2 hergeleitet hatten, ist erwartungstreu.
- Für den MQ-Fehler dieses Schätzers gilt somit, dass
$$\mathbb{E}_p\big((\widehat{p}(X_1,\ldots,X_n)-p)^2\big)=\mathrm{Var}_p\overline{X}_n=\frac{p(1-p)}{n}\,. \tag{26}$$
- Außerdem hatten wir in Abschnitt 2.2.3 den Bayes-Schätzer $\widetilde{p}(Y)$ für p konstruiert mit
$$\widetilde{p}(Y)=\frac{Y+\alpha}{\alpha+\beta+n}\,,\qquad Y=\sum_{i=1}^n X_i\,.$$
- Aus Theorem 2.1 ergibt sich für den MQ-Fehler dieses Bayes-Schätzers, dass
$$\begin{aligned}\mathbb{E}_p\big((\widetilde{p}(Y)-p)^2\big) &= \mathrm{Var}_p\,\widetilde{p}(Y)+\big(\mathrm{Bias}_p\,\widetilde{p}(Y)\big)^2\\ &= \mathrm{Var}_p\Big(\frac{Y+\alpha}{\alpha+\beta+n}\Big)+\Big(\mathbb{E}_p\,\frac{Y+\alpha}{\alpha+\beta+n}-p\Big)^2\\ &= \frac{np(1-p)}{(\alpha+\beta+n)^2}+\Big(\frac{np+\alpha}{\alpha+\beta+n}-p\Big)^2.\end{aligned}$$
- Falls keine spezielle a-priori-Information über den Parameter p vorliegt, dann erscheint es sinnvoll, α und β so wählen, dass der MQ-Fehler des Bayes-Schätzers $\widetilde{p}(Y)$ nicht von p abhängt.
- Für $\alpha=\beta=\sqrt{n/4}$ gilt nämlich (vgl. Übungsaufgabe 6.1), dass
$$\mathbb{E}_p\big((\widetilde{p}(Y)-p)^2\big)=\frac{n}{4(n+\sqrt{n})^2}\,. \tag{27}$$
- Aus (26) und (27) ergibt sich dann, dass der MQ-Fehler des Bayes-Schätzers $\widetilde{p}(Y)$ bei kleinem Stichprobenumfang ($n\approx 4$) deutlich kleiner als der MQ-Fehler von $\overline{X}_n$ ist (es sei denn, dass p nahe bei 0 oder 1 liegt).
- Umgekehrt ist der MQ-Fehler von $\overline{X}_n$ bei großem Stichprobenumfang ($n\approx 400$) deutlich kleiner als der MQ-Fehler von $\widetilde{p}(Y)$ (es sei denn, dass p nahe bei $1/2$ liegt).

2.3.2 Ungleichung von Cramér-Rao

- In diesem Abschnitt setzen wir voraus, dass der Parameter θ eine relle Zahl ist, d.h.,
- es gelte $m=1$ bzw. $\Theta\subset\mathbb{R}$.
- Aus Theorem 2.1 ergibt sich, dass in der Klasse der erwartungstreuen Schätzer die Minimierung des MQ-Fehlers gleichbedeutend mit der Minimierung der Varianz des Schätzers ist.
- Wir werden deshalb die folgende Sprechweise verwenden.

Definition

- Seien $\widehat{\theta}_1,\widehat{\theta}_2:\mathbb{R}^n\to\Theta$ zwei Stichprobenfunktionen, so dass für jedes $\theta\in\Theta$
$$\mathbb{E}_\theta\widehat{\theta}_1^2(X_1,\ldots,X_n)<\infty\qquad \mathbb{E}_\theta\widehat{\theta}_2^2(X_1,\ldots,X_n)<\infty$$
und
$$\mathbb{E}_\theta\widehat{\theta}_1(X_1,\ldots,X_n)=\mathbb{E}_\theta\widehat{\theta}_2(X_1,\ldots,X_n)=\theta\,.$$

- Falls
$$\mathrm{Var}_\theta\widehat{\theta}_1(X_1,\ldots,X_n) \le \mathrm{Var}_\theta\,\widehat{\theta}_2(X_1,\ldots,X_n)\,, \qquad \forall\,\theta\in\Theta\,,$$
dann sagen wir, dass der Schätzer $\widehat{\theta}_1(X_1,\ldots,X_n)$ *besser* als der Schätzer $\widehat{\theta}_2(X_1,\ldots,X_n)$ für θ ist.

Beispiel *Poisson-verteilte Stichprobenvariablen*

- Sei $(X_1,\ldots,X_n)$ eine Poisson-verteilte Zufallsstichprobe mit $X_i \sim \mathrm{Poi}(\lambda)$.
- Weil sowohl der Erwartungswert als auch die Varianz der $\mathrm{Poi}(\lambda)$-Verteilung gleich λ sind, gilt
$$\mathbb{E}_\lambda\overline{X}_n = \mathbb{E}_\lambda S_n^2 = \lambda\,.$$
- D.h., sowohl $\overline{X}_n$ als auch S_n^2 sind erwartungstreue Schätzer für λ.
- Aus den Theoremen 1.1 bzw. 1.3 folgt außerdem, dass die Varianz dieser Schätzer gegeben ist durch
$$\mathrm{Var}_\lambda\overline{X}_n = \frac{\lambda}{n} \qquad \text{bzw.} \qquad \mathrm{Var}_\lambda(S_n^2) = \frac{1}{n}\Bigl(\mu_4' - \frac{n-3}{n-1}\,\lambda^2\Bigr)\,,$$
wobei $\mu_4' = \mathbb{E}\bigl((X_i-\lambda)^4\bigr)$.
- Für das vierte zentrierte Moment μ_4' der $\mathrm{Poi}(\lambda)$-Verteilung gilt
$$\mu_4' = \lambda + 3\lambda^2\,.$$
- Hieraus ergibt sich, dass die Varianz von $\overline{X}_n$ kleiner ist als die Varianz von S_n^2, denn es gilt:
$$\mathrm{Var}_\lambda\overline{X}_n = \frac{\lambda}{n} < \frac{\lambda}{n}\Bigl(1+\Bigl(3-\frac{n-3}{n-1}\Bigr)\lambda\Bigr) = \mathrm{Var}_\lambda(S_n^2)\,.$$
- Wir werden nun die Frage untersuchen, ob es erwartungstreue Schätzer für λ gibt, deren Varianz kleiner als λ/n ist.

Zunächst leiten wir eine allgemeine untere Schranke für die Varianz von Schätzern mit gewissen Regularitätseigenschaften her, die in der Literatur *Ungleichung von Cramér-Rao* genannt wird.

Das parametrische Modell $\{P_\theta, \theta\in\Theta\}$ genüge den folgenden *Regularitätsbedingungen*:

1. Die Familie $\{P_\theta, \theta\in\Theta\}$ bestehe entweder nur aus diskreten Verteilungen oder nur aus absolutstetigen Verteilungen, wobei $\Theta\subset\mathbb{R}$ ein offenes Intervall sei.
2. Die Menge $B = \{x\in\mathbb{R} : L(x;\theta) > 0\}$ hänge nicht von $\theta\in\Theta$ ab, wobei die Likelihood-Funktion $L(x;\theta)$ gegeben ist durch
$$L(x;\theta) = \begin{cases} p(x;\theta) & \text{im diskreten Fall,}\\ f(x;\theta) & \text{im absolutstetigen Fall}\end{cases}$$
und $p(x;\theta)$ bzw. $f(x;\theta)$ die Wahrscheinlichkeitsfunktion bzw. Dichte von P_θ ist.
3. Die Ableitung $\frac{\partial}{\partial\theta}L(x;\theta)$ existiere für beliebige $\theta\in\Theta$ und $x\in B$.
4. Vertauschbarkeit von Ableitung und Summe/Integral: Für jedes $\theta\in\Theta$ gelte
$$\sum_{x\in B}\frac{\partial}{\partial\theta}L(x;\theta) = \frac{d}{d\theta}\sum_{x\in B}L(x;\theta) = 0 \qquad \text{im diskreten Fall} \tag{28}$$
bzw.
$$\int\limits_B \frac{\partial}{\partial\theta}L(x;\theta)\,dx = \frac{d}{d\theta}\int\limits_B L(x;\theta)\,dx = 0 \qquad \text{im absolutstetigen Fall.} \tag{29}$$

Theorem 2.2 *Sei $\{P_\theta,\, \theta \in \Theta\}$ eine Familie von Verteilungen, die den Regularitätsbedingungen 1–4 genügt, und sei $\widehat{\theta} : \mathbb{R}^n \to \Theta$ eine Stichprobenfunktion, so dass für jedes $\theta \in \Theta$*

- $\mathbb{E}_\theta\big((\widehat{\theta}(X_1,\ldots,X_n))^2\big) < \infty$,
- *die Ableitung* $\frac{d}{d\theta}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)$ *existiert und*

$$\frac{d}{d\theta}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n) = \begin{cases} \displaystyle\sum_{x_1,\ldots,x_n\in B} \widehat{\theta}(x_1,\ldots,x_n)\frac{\partial}{\partial\theta}L(x_1,\ldots,x_n;\theta) & \text{im diskreten Fall,} \\ \displaystyle\int_{B^n} \widehat{\theta}(x_1,\ldots,x_n)\frac{\partial}{\partial\theta}L(x_1,\ldots,x_n;\theta)\,dx_1\ldots dx_n & \text{im absolutstetigen Fall,} \end{cases} \tag{30}$$

wobei $L(x_1,\ldots,x_n;\theta)$ die Likelihood-Funktion ist mit

$$L(x_1,\ldots,x_n;\theta) = \begin{cases} p(x_1;\theta)\ldots p(x_n;\theta) & \text{im diskreten Fall,} \\ f(x_1;\theta)\ldots f(x_n;\theta) & \text{im absolutstetigen Fall.} \end{cases}$$

Dann gilt für jedes $\theta \in \Theta$

$$\mathrm{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n) \ge \frac{\Big(\frac{d}{d\theta}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)\Big)^2}{n\,\mathbb{E}_\theta\Big(\Big(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Big)^2\Big)}\,. \tag{31}$$

Beweis

- Wir betrachten zunächst den diskreten Fall.
- Für jedes $\theta \in \Theta$ sei die Abbildung $\varphi_\theta : \mathbb{R}^n \to \mathbb{R}$ gegeben durch

$$\varphi_\theta(x_1,\ldots,x_n) = \Big(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n L(x_i;\theta)\Big)\mathbb{1}_{B\times\ldots\times B}(x_1,\ldots,x_n)\,.$$

- Dann gilt für jedes $\theta \in \Theta$

$$\begin{aligned} \mathrm{Var}_\theta\varphi_\theta(X_1,\ldots,X_n) &= \mathrm{Var}_\theta\Big(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n L(X_i;\theta)\Big) \\ &= \mathrm{Var}_\theta\Big(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log L(X_i;\theta)\Big) \\ &= \sum_{i=1}^n\mathrm{Var}_\theta\Big(\frac{\partial}{\partial\theta}\log L(X_i;\theta)\Big) \\ &= n\,\mathrm{Var}_\theta\Big(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Big)\,, \end{aligned}$$

  wobei sich die letzten beiden Gleichheiten aus der Unabhängigkeit bzw. identischen Verteiltheit der Stichprobenvariablen $X_1,\ldots,X_n$ ergeben.
- Außerdem gilt

$$\begin{aligned} \mathbb{E}_\theta\Big(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Big) &= \sum_{x\in B}\Big(\frac{\partial}{\partial\theta}\log L(x;\theta)\Big)L(x;\theta) \\ &= \sum_{x\in B}\frac{\partial}{\partial\theta}L(x;\theta) \\ &\overset{(28)}{=} \frac{\partial}{\partial\theta}\sum_{x\in B}L(x;\theta) = 0\,. \end{aligned}$$

- Somit gilt
$$\operatorname{Var}_\theta\varphi_\theta(X_1,\ldots,X_n)=n\mathbb{E}_\theta\Bigl(\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)^2\Bigr)\,.$$
- Aus der Ungleichung von Cauchy-Schwarz (vgl. Theorem WR-4.11) ergibt sich nun, dass
$$\begin{aligned}\Bigl(\operatorname{Cov}_\theta(\varphi_\theta(X_1,\ldots,X_n),\widehat{\theta}(X_1,\ldots,X_n))\Bigr)^2 &\le \operatorname{Var}_\theta\varphi_\theta(X_1,\ldots,X_n)\operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n)\\ &= n\mathbb{E}_\theta\Bigl(\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)^2\Bigr)\operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n)\,.\end{aligned}$$
- Wegen $\mathbb{E}_\theta\varphi_\theta(X_1,\ldots,X_n)=0$ ergibt sich andererseits, dass
$$\begin{aligned}&\operatorname{Cov}_\theta(\varphi_\theta(X_1,\ldots,X_n),\widehat{\theta}(X_1,\ldots,X_n))=\mathbb{E}_\theta(\varphi_\theta(X_1,\ldots,X_n)\,\widehat{\theta}(X_1,\ldots,X_n))\\ &=\sum_{x_1,\ldots,x_n\in B}\widehat{\theta}(x_1,\ldots,x_n)\Bigl(\frac{\partial}{\partial\theta}\log L(x_1,\ldots,x_n;\theta)\Bigr)L(x_1,\ldots,x_n;\theta)\\ &=\sum_{x_1,\ldots,x_n\in B}\widehat{\theta}(x_1,\ldots,x_n)\frac{\partial}{\partial\theta}L(x_1,\ldots,x_n;\theta)\\ &\overset{(30)}{=}\frac{d}{d\theta}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)\,.\end{aligned}$$
- Damit ist (31) für den diskreten Fall bewiesen. Im absolutstetigen Fall verläuft der Beweis analog. □

Korollar 2.1 *Sei $\{P_\theta,\,\theta\in\Theta\}$ eine Familie von Verteilungen und $\widehat{\theta}:\mathbb{R}^n\to\Theta$ eine Stichprobenfunktion, die den Bedingungen von Theorem 2.2 genügen. Falls $\widehat{\theta}(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer für θ ist, dann gilt für jedes $\theta\in\Theta$*
$$\operatorname{Var}_\theta\,\widehat{\theta}(X_1,\ldots,X_n)\ge\frac{1}{n\,\mathbb{E}_\theta\Bigl(\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)^2\Bigr)}\,. \tag{32}$$

Beweis

- Weil $\widehat{\theta}(X_1,\ldots,X_n)$ erwartungstreu ist, gilt
$$\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)=\theta$$
bzw.
$$\frac{d}{d\theta}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)=1\,.$$
- Die Behauptung ergibt sich nun unmittelbar aus der Cramér-Rao-Ungleichung (31). □

Wir diskutieren nun zwei Beispiele von parametrischen Verteilungsfamilien, für die die Bedingungen von Theorem 2.2 bzw. Korollar 2.1 erfüllt sind. Außerdem zeigen wir, dass für diese beiden Verteilungsfamilien das Stichprobenmittel $\overline{X}_n$ ein Schätzer ist, der „optimal" im Sinne der Ungleichungen (31) bzw. (32) ist.

Beispiele

1. *Normalverteilte Stichprobenvariablen*
   - Falls $X_i\sim\mathrm{N}(\mu,\sigma^2)$, wobei die Varianz σ^2 bekannt sei, dann gilt $\Theta=B=\mathbb{R}$ und für jedes $x\in\mathbb{R}$
$$L(x;\mu)=\frac{1}{\sqrt{2\pi}\sigma}\,\exp\Bigl(-\frac{(x-\mu)^2}{2\sigma^2}\Bigr) \tag{33}$$
bzw.
$$\frac{d}{d\mu}L(x;\mu)=\frac{1}{\sqrt{2\pi}\sigma}\,\frac{x-\mu}{\sigma^2}\,\exp\Bigl(-\frac{(x-\mu)^2}{2\sigma^2}\Bigr)\,. \tag{34}$$

- Somit gilt

$$\int_{\mathbb{R}} \frac{d}{d\mu} L(x;\mu)\,dx = \frac{d}{d\mu}\int_{\mathbb{R}} L(x;\mu)\,dx = 0\,. \tag{35}$$

- Die Regularitätsbedingungen 1–4, die unmittelbar vor Theorem 2.2 formuliert wurden, sind also erfüllt. Wir zeigen nun, dass der Schätzer $\overline{X}_n$ für μ die Bedingungen von Theorem 2.2 erfüllt.
- Weil $\overline{X}_n$ normalverteilt und ein erwartungstreuer Schätzer für μ ist, gilt offenbar $\mathbb{E}_\mu(\overline{X}_n^2) < \infty$, und die Ableitung $(d/d\mu)\,\mathbb{E}_\mu\overline{X}_n = 1$ ist wohldefiniert.
- Die Gültigkeit von (30) mittels vollständiger Induktion,
  - denn aus (33) und (34) folgt, dass für $n = 1$

$$\int_{\mathbb{R}} x\,\frac{d}{d\mu} L(x;\mu)\,dx = \int_{\mathbb{R}} x\,\frac{x-\mu}{\sigma^2}\,L(x;\mu)\,dx = \frac{1}{\sigma^2}\,\big(\mathbb{E}_\mu(X_i^2) - \mu^2\big) = 1\,. \tag{36}$$

  - Für ein $n \in \mathbb{N}$ gelte nun

$$\int_{\mathbb{R}^n} (x_1 + \ldots + x_n)\,\frac{d}{d\mu} L(x_1,\ldots,x_n;\mu)\,d(x_1,\ldots,x_n) = n\,. \tag{37}$$

  - Weil

$$\begin{aligned}&\frac{d}{d\mu} L(x_1,\ldots,x_{n+1};\mu)\,d(x_1,\ldots,x_{n+1})\\ &= \frac{d}{d\mu} L(x_1,\ldots,x_n;\mu)\,L(x_{n+1};\mu) + L(x_1,\ldots,x_n;\mu)\,\frac{d}{d\mu} L(x_{n+1};\mu)\end{aligned}$$

    und weil

$$\int_{\mathbb{R}} \frac{d}{d\mu} L(x_i;\mu)\,dx_i = 0\,, \qquad \forall\, i = 1,\ldots,n+1\,,$$

    ergibt sich nun aus (36) und (37), dass

$$\int_{\mathbb{R}^{n+1}} (x_1 + \ldots + x_{n+1})\,\frac{d}{d\mu} L(x_1,\ldots,x_{n+1};\mu)\,d(x_1,\ldots,x_{n+1}) = n+1\,.$$

- Außerdem gilt

$$\frac{d}{d\mu}\log L(x,\mu) = \frac{d}{d\mu}\Big(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}\Big) = \frac{x-\mu}{\sigma^2}$$

  und somit

$$\mathrm{Var}_\mu\,\overline{X}_n = \frac{\sigma^2}{n} = \frac{1}{n\,\mathbb{E}_\mu\Big(\Big(\frac{d}{d\mu}\log L(X_1;\mu)\Big)^2\Big)}\,.$$

2. *Poisson-verteilte Stichprobenvariablen*
   - Falls $X_i \sim \mathrm{Poi}(\lambda)$, dann gilt $B = \mathbb{N}$ und $L(x;\lambda) = (\lambda^x/x!)\,e^{-\lambda}$ für jedes $x \in \mathbb{N}$ bzw.

$$\frac{d}{d\lambda} L(x;\lambda) = \begin{cases} -e^{-\lambda}, & \text{falls } x = 0,\\ \Big(1 - \frac{\lambda}{x}\Big)\frac{\lambda^{x-1}}{(x-1)!}\,e^{-\lambda}, & \text{falls } x > 0. \end{cases} \tag{38}$$

   - Somit gilt

$$\sum_{x\in\mathbb{N}} \frac{d}{d\lambda} L(x;\lambda) = -e^{-\lambda} + \sum_{k=1}^{\infty}\Big(1 - \frac{\lambda}{k}\Big)\frac{\lambda^{k-1}}{(k-1)!}\,e^{-\lambda} = -e^{-\lambda} + e^{-\lambda} = 0\,. \tag{39}$$

- Die Regularitätsbedingungen 1–4 von Theorem 2.2 sind also erfüllt.
- Außerdem genügt der Schätzer $\overline{X}_n$ für λ den Bedingungen von Theorem 2.2, denn für jedes $\lambda > 0$ gilt
$$\mathbb{E}_\lambda|\overline{X}_n| = \mathbb{E}_\lambda \overline{X}_n = \lambda\,, \qquad \mathrm{Var}_\lambda \overline{X}_n = \frac{\lambda}{n}$$
und somit $(d/d\lambda)\,\mathbb{E}_\lambda \overline{X}_n = 1$.
- Hieraus und aus (38) ergibt sich die Gültigkeit der Bedingung (30) erneut mittels vollständiger Induktion:
  - Für $n = 1$ gilt
$$\sum_{x\in\mathbb{N}} x\,\frac{d}{d\lambda}\,L(x;\lambda) = \sum_{k=1}^{\infty} k\Bigl(1-\frac{\lambda}{k}\Bigr)\frac{\lambda^{k-1}}{(k-1)!}\,e^{-\lambda} = \lambda + 1 - \lambda = 1\,. \tag{40}$$
  - Wir nehmen nun an, dass
$$\sum_{x_1,\ldots,x_{n-1}\in\mathbb{N}} (x_1+\ldots+x_{n-1})\,\frac{d}{d\lambda}\,L(x_1,\ldots,x_{n-1};\lambda) = n-1\,. \tag{41}$$
  - Dann gilt
$$\begin{aligned}
&\sum_{x_1,\ldots,x_n\in\mathbb{N}} (x_1+\ldots+x_n)\,\frac{d}{d\lambda}\,L(x_1,\ldots,x_n;\lambda)\\
= \quad & \sum_{x_1,\ldots,x_{n-1}\in\mathbb{N}}\ \sum_{x_n\in\mathbb{N}} \bigl((x_1+\ldots+x_{n-1})+x_n\bigr)\Bigl(\Bigl(\frac{d}{d\lambda}\,L(x_1,\ldots,x_{n-1};\lambda)\Bigr)L(x_n;\lambda)\\
& + L(x_1,\ldots,x_{n-1};\lambda)\frac{d}{d\lambda}\,L(x_n;\lambda)\Bigr)\\
\overset{(39),(40)}{=} \quad & \sum_{x_1,\ldots,x_{n-1}\in\mathbb{N}} \Bigl((x_1+\ldots+x_{n-1})\frac{d}{d\lambda}\,L(x_1,\ldots,x_{n-1};\lambda)\\
& + \lambda\frac{d}{d\lambda}\,L(x_1,\ldots,x_{n-1};\lambda) + L(x_1,\ldots,x_{n-1};\lambda)\Bigr)\\
\overset{(41),(39)}{=} \quad & (n-1)+0+1 = n\,.
\end{aligned}$$
- Außerdem gilt für jedes $x \in \mathbb{N}$
$$\frac{d}{d\lambda}\,\log L(x;\lambda) = \frac{x}{\lambda} - 1 = \frac{1}{\lambda}(x-\lambda)\,.$$
- Hieraus folgt, dass
$$\frac{1}{n\,\mathbb{E}_\lambda\Bigl(\Bigl(\frac{d}{d\lambda}\,\log L(X_1;\lambda)\Bigr)^2\Bigr)} = \frac{\lambda^2}{n\,\mathbb{E}_\lambda\bigl((X_1-\lambda)^2\bigr)} = \frac{\lambda}{n}\,.$$
- Weil $\mathrm{Var}_\lambda \overline{X}_n = \lambda/n$ gilt, ist damit gezeigt, dass in der Klasse derjenigen Schätzer, die die Bedingungen von Theorem 2.2 erfüllen, das Stichprobenmittel $\overline{X}_n$ bester erwartungstreuer Schätzer für λ ist.

Beachte Es gibt jedoch Familien $\{P_\theta, \theta \in \Theta\}$ von Verteilungen,

- die den Regularitätsbedingungen 1–4 nicht genügen, und
- für die man Beispiele von erwartungstreuen Schätzern konstruieren kann, deren Varianz kleiner als die untere Schranke in (32) ist.

Beispiel *Gleichverteilte Stichprobenvariablen*

- Sei $(X_1,\ldots,X_n)$ eine gleichverteilte Zufallsstichprobe mit $X_i \sim \mathrm{U}(0,\theta)$.
- Dann ist $B=(0,\theta)$ (also eine Menge, die entgegen der zweiten Regularitätsbedingung von der spezifischen Ausprägung des Parameters θ abhängt), und es gilt

$$L(x;\theta)=\frac{1}{\theta}\,,\qquad \forall\, x\in(0,\theta)\,.$$

- Hieraus folgt, dass auch die Regularitätsbedingung (29) nicht erfüllt ist, denn

$$\begin{aligned}\frac{d}{d\theta}\int\limits_0^\theta L(x;\theta)\,dx &= \frac{d}{d\theta}\int\limits_0^\theta \frac{1}{\theta}\,dx=\frac{1}{\theta}+\int\limits_0^\theta \frac{\partial}{\partial\theta}\Bigl(\frac{1}{\theta}\Bigr)\,dx\\ &\neq \int\limits_0^\theta \frac{\partial}{\partial\theta}\Bigl(\frac{1}{\theta}\Bigr)\,dx=\int\limits_0^\theta \frac{\partial}{\partial\theta}L(x;\theta)\,dx\,.\end{aligned}$$

- Außerdem gilt für jedes $x\in(0,\theta)$

$$\frac{\partial}{\partial\theta}\log L(x;\theta)=-\frac{1}{\theta}\,.$$

- Hieraus folgt, dass

$$\frac{1}{n\,\mathbb{E}_\theta\Bigl(\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)^2\Bigr)}=\frac{\theta^2}{n}\,.$$

- Wir konstruieren nun einen erwartungstreuen Schätzer für θ, dessen Varianz kleiner als θ^2/n ist.
- Sei

$$\widehat{\theta}(X_1,\ldots,X_n)=\frac{n+1}{n}\max\{X_1,\ldots,X_n\}\,.$$

Dann gilt (vgl. Übungsaufgabe 3.3 bzw. 6.4)

$$\mathbb{E}\,\widehat{\theta}(X_1,\ldots,X_n)=\frac{n+1}{n}\,\mathbb{E}\,\max\{X_1,\ldots,X_n\}=\theta$$

und

$$\begin{aligned}\operatorname{Var}\widehat{\theta}(X_1,\ldots,X_n) &= \frac{(n+1)^2}{n^2}\operatorname{Var}\max\{X_1,\ldots,X_n\}\\ &= \frac{(n+1)^2}{n^2}\,\frac{n\theta^2}{(n+1)^2(n+2)}\\ &= \frac{\theta^2}{n(n+2)}<\frac{\theta^2}{n}\,.\end{aligned}$$

2.3.3 Suffizienz

Sei $\{P_\theta,\,\theta\in\Theta\}$ ein beliebiges parametrisches Modell mit $\Theta\subset\mathbb{R}^m$, und sei $(X_1,\ldots,X_n)$ eine Zufallsstichprobe über dem (kanonischen) Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},P_\theta)$.

- Zur Erinnerung: Unter gewissen Regularitätsbedingungen hatten wir in Abschnitt 2.3.2 für eine Klasse von erwartungstreuen Schätzern eine untere Schranke für ihre Varianz, die sogenannte *Ungleichung von Cramér-Rao*, hergeleitet; vgl. Korollar 2.1.
- In Abschnitt 2.3.5 werden wir die Frage untersuchen, unter welchen Bedingungen erwartungstreue Schätzer mit minimaler Varianz existieren und welche Eigenschaften solche Schätzer haben.

- In Zusammenhang hiermit wird die Suffizienz von Schätzern eine wichtige Rolle spielen, wobei diese Eigenschaft intuitiv wie folgt beschrieben werden kann.
- Man sagt, dass der Schätzer $\varphi(X_1,\ldots,X_n)$ für θ suffizient ist, falls (in einem noch zu präzisierenden Sinne) sämtliche Information über θ, die in der Zufallsstichprobe $(X_1,\ldots,X_n)$ enthalten ist, auch in $\varphi(X_1,\ldots,X_n)$ enthalten ist.

Beachte

- Typischerweise erfolgt ein „Informationsverlust" beim Übergang von $(X_1,\ldots,X_n)$ nach $\varphi(X_1,\ldots,X_n)$ insofern, dass
- die Funktionswerte der Abbildung $(X_1,\ldots,X_n):\Omega\to\mathbb{R}^n$ im allgemeinen *nicht* aus den Funktionswerten der Abbildung $\varphi(X_1,\ldots,X_n):\Omega\to\mathbb{R}^m$ rekonstruiert werden können.
- Von Suffizienz spricht man, wenn dieser „Informationsverlust" in einem gewissen (stochastischen) Sinne nicht wesentlich ist.

Wir präzisieren dies zunächst für den *diskreten Fall*: Für jedes $\theta\in\Theta$ gelte

- $P_\theta(X_1\in C)=1$ für eine abzählbare Menge $C\subset\mathbb{R}$, die nicht von θ abhängt, wobei wir mit $\{p(x;\theta),\ x\in C\}$ die Wahrscheinlichkeitsfunktion von X_1 bezeichnen und o.B.d.A. annehmen, dass $p(x;\theta)>0$ für jedes $x\in C$.
- Dann gilt auch $P_\theta(\varphi(X_1,\ldots,X_n)\in C')=1$ für jedes $\theta\in\Theta$, wobei $C'\subset\mathbb{R}^m$ die minimale abzählbare Menge mit dieser Eigenschaft sei, d.h., dass $P_\theta(\varphi(X_1,\ldots,X_n)=t)>0$ für jedes $t\in C'$.

Definition Der Schätzer $\varphi(X_1,\ldots,X_n)$ für θ heißt *suffizient*, falls für beliebige $x_1,\ldots,x_n\in C$ und $t\in C'$ die bedingten Wahrscheinlichkeiten

$$P_\theta(X_1=x_1,\ldots,X_n=x_n\mid\varphi(X_1,\ldots,X_n)=t)=\frac{P_\theta(X_1=x_1,\ldots,X_n=x_n,\varphi(X_1,\ldots,X_n)=t)}{P_\theta(\varphi(X_1,\ldots,X_n)=t)}\tag{42}$$

nicht von θ abhängen.

Beachte

- Falls (42) gilt, dann ist es nicht möglich, der (konkreten) Stichprobe $(x_1,\ldots,x_n)$ mit Hilfe der (bedingten) Likelihood-Funktion

  $$L_t(x_1,\ldots,x_n;\theta)=P_\theta(X_1=x_1,\ldots,X_n=x_n\mid\varphi(X_1,\ldots,X_n)=t)$$

  einen (eindeutig bestimmten) ML-Schätzwert $\widehat{\theta}(x_1,\ldots,x_n)$ für θ zuzuordnen.
- Mit anderen Worten: Falls $\varphi(X_1,\ldots,X_n)$ ein suffizienter Schätzer für θ ist und falls bekannt ist, dass $\varphi(X_1,\ldots,X_n)=t$, dann kann aus der zusätzlichen Kenntnis der einzelnen Stichprobenwerte $x_1,\ldots,x_n$ keine weitere Information über den unbekannten Parametervektor θ gewonnen werden.

Zunächst zeigen wir, dass die Suffizienz bei (bijektiven) zusammengesetzten Stichprobenfunktionen nicht verloren geht.

Lemma 2.1 *Sei $\varphi(X_1,\ldots,X_n)$ ein suffizienter Schätzer für θ, und sei $g:\mathbb{R}^m\to\mathbb{R}^m$ eine bijektive Borel-messbare Abbildung. Dann ist auch $g\circ\varphi(X_1,\ldots,X_n)$ ein suffizienter Schätzer für θ.*

Beweis Für jedes $t \in \mathbb{R}^m$ mit $P_\theta(g \circ \varphi(X_1, \ldots, X_n) = t) > 0$ gilt

$$\begin{aligned}
& P_\theta(X_1 = x_1, \ldots, X_n = x_n \mid g \circ \varphi(X_1, \ldots, X_n) = t) \\
= \; & \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, g \circ \varphi(X_1, \ldots, X_n) = t)}{P_\theta(g \circ \varphi(X_1, \ldots, X_n) = t)} \\
= \; & \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, \varphi(X_1, \ldots, X_n) = g^{-1}(t))}{P_\theta(\varphi(X_1, \ldots, X_n) = g^{-1}(t))} \\
= \; & P_\theta(X_1 = x_1, \ldots, X_n = x_n \mid \varphi(X_1, \ldots, X_n) = g^{-1}(t)) \, ,
\end{aligned}$$

wobei der letzte Ausdruck nicht von θ abhängt, weil $\varphi(X_1, \ldots, X_n)$ suffizient ist. □

Beachte

- Wir leiten nun eine (notwendige und hinreichende) Bedingung für die Suffizienz von Schätzern her, die sich unmittelbar aus der Definition dieser Güteeigenschaft ergibt.
- Dabei verwenden wir die abkürzende Schreibweise

$$p(x_1, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n) \, , \qquad \forall \, x_1, \ldots, x_n \in C \, ,$$

  für die Wahrscheinlichkeitsfunktion der Zufallsstichprobe $(X_1, \ldots, X_n)$, und

$$q(t; \theta) = P_\theta(\varphi(X_1, \ldots, X_n) = t) \, , \qquad \forall \, t \in C' \, ,$$

  für die Wahrscheinlichkeitsfunktion von $\varphi(X_1, \ldots, X_n)$.

Lemma 2.2 *Der Schätzer $\varphi(X_1, \ldots, X_n)$ für θ ist genau dann suffizient, wenn für beliebige $x_1, \ldots, x_n \in C$ der Quotient $p(x_1, \ldots, x_n; \theta)/q(\varphi(x_1, \ldots, x_n); \theta)$ nicht von $\theta \in \Theta$ abhängt.*

Beweis

- Sei $x_1, \ldots, x_n \in C$ und $t \in C'$.
- Weil die bedingten Wahrscheinlichkeiten in (42) gleich 0 sind, falls $t \neq \varphi(x_1, \ldots, x_n)$, genügt es, den Fall $t = \varphi(x_1, \ldots, x_n)$ zu betrachten.
- Offenbar gilt

$$\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\} \subset \{\varphi(X_1, \ldots, X_n) = \varphi(x_1, \ldots, x_n)\} \, .$$

- Hieraus folgt, dass

$$\begin{aligned}
& P_\theta(X_1 = x_1, \ldots, X_n = x_n \mid \varphi(X_1, \ldots, X_n) = \varphi(x_1, \ldots, x_n)) \\
= \; & \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, \varphi(X_1, \ldots, X_n) = \varphi(x_1, \ldots, x_n))}{P_\theta(\varphi(X_1, \ldots, X_n) = \varphi(x_1, \ldots, x_n))} \\
= \; & \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n)}{P_\theta(\varphi(X_1, \ldots, X_n) = \varphi(x_1, \ldots, x_n))} \\
= \; & \frac{p(x_1, \ldots, x_n; \theta)}{q(\varphi(x_1, \ldots, x_n); \theta)} \, .
\end{aligned}$$

□

Beispiel *Bernoulli-verteilte Stichprobenvariablen*

- Es gelte $\{P_\theta, \, \theta \in \Theta\} = \{\mathrm{Bin}(1, \theta), \, \theta \in (0,1)\}$.
- Wir zeigen, dass $\overline{X}_n$ ein suffizienter Schätzer für θ ist.
- Wegen Lemma 2.1 genügt es zu zeigen, dass der Schätzer $Y = X_1 \ldots + X_n$ suffizient ist.

- Mit der abkürzenden Schreibweise $t = x_1 + \ldots + x_n$ gilt für beliebige $\theta \in (0,1)$ und $x_1, \ldots, x_n \in \{0,1\}$

$$
\begin{aligned}
\frac{p(x_1,\ldots,x_n;\theta)}{q(t;\theta)} &= \frac{\prod\limits_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\theta^{\sum_i x_i}(1-\theta)^{\sum_i (1-x_i)}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{1}{\binom{n}{t}} = \frac{1}{\binom{n}{x_1+\ldots+x_n}}\,.
\end{aligned}
$$

- Weil der letzte Quotient nicht von θ abhängt, ergibt sich aus Lemma 2.2, dass der Schätzer $Y = X_1 + \ldots + X_n$ suffizient ist.

Beachte

- Die folgende (verallgemeinerte) Fassung von Lemma 2.2 wird in der Literatur *Faktorisierungssatz von Neyman-Fisher* genannt.
- Dieser Faktorisierungssatz beinhaltet eine andere (notwendige und hinreichende) Bedingung für die Suffizienz, die gegenüber der in Lemma 2.2 angegebenen Bedingung den Vorteil besitzt, dass die Wahrscheinlichkeitsfunktion von $\varphi(X_1, \ldots, X_n)$ nicht explizit berechnet werden muss.

Theorem 2.3 *Der Schätzer $\varphi(X_1, \ldots, X_n)$ für θ ist genau dann suffizient, wenn es Borel-messbare Funktionen $g : \mathbb{R}^m \times \Theta \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}$ gibt, so dass*

$$
p(x_1,\ldots,x_n;\theta) = g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr)\, h(x_1,\ldots,x_n)\,, \qquad \forall\, x_1,\ldots,x_n \in \mathbb{R},\ \theta \in \Theta\,. \tag{43}
$$

Beweis

- Die Notwendigkeit der Bedingung (43) ergibt sich unmittelbar aus Lemma 2.2.
- Wir nehmen nun an, dass sich die Wahrscheinlichkeitsfunktion $p(x_1, \ldots, x_n; \theta)$ auf die in (43) gegebene Weise darstellen lässt.
- So wie bisher sei $q(t; \theta)$ die Wahrscheinlichkeitsfunktion von $\varphi(X_1, \ldots, X_n)$.
- Mit der Schreibweise $B_{\varphi(x_1,\ldots,x_n)} = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : \varphi(y_1, \ldots, y_n) = \varphi(x_1, \ldots, x_n)\}$ ergibt sich dann aus (43), dass

$$
\begin{aligned}
\frac{p(x_1,\ldots,x_n;\theta)}{q(\varphi(x_1,\ldots,x_n);\theta)} &\overset{(43)}{=} \frac{g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr)\, h(x_1,\ldots,x_n)}{q(\varphi(x_1,\ldots,x_n);\theta)} = \frac{g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr)\, h(x_1,\ldots,x_n)}{\sum\limits_{(y_1,\ldots,y_n)\in B_{\varphi(x_1,\ldots,x_n)}} p(y_1,\ldots,y_n;\theta)} \\
&\overset{(43)}{=} \frac{g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr)\, h(x_1,\ldots,x_n)}{\sum\limits_{(y_1,\ldots,y_n)\in B_{\varphi(x_1,\ldots,x_n)}} g\bigl(\varphi(y_1,\ldots,y_n),\theta\bigr)\, h(y_1,\ldots,y_n)} \\
&= \frac{g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr)\, h(x_1,\ldots,x_n)}{g\bigl(\varphi(x_1,\ldots,x_n),\theta\bigr) \sum\limits_{(y_1,\ldots,y_n)\in B_{\varphi(x_1,\ldots,x_n)}} h(y_1,\ldots,y_n)} \\
&= \frac{h(x_1,\ldots,x_n)}{\sum\limits_{(y_1,\ldots,y_n)\in B_{\varphi(x_1,\ldots,x_n)}} h(y_1,\ldots,y_n)}\,.
\end{aligned}
$$

- Weil der letzte Ausdruck nicht von θ abhängt, ergibt sich nun aus Lemma 2.2, dass der Schätzer $\varphi(X_1,\ldots,X_n)$ für θ suffizient ist. $\square$

Beispiel *Poissonverteilte Stichprobenvariablen*

- Es gelte $\{P_\theta,\, \theta\in\Theta\} = \{\mathrm{Poi}(\lambda),\, \lambda>0\}$, d.h., die Stichprobenvariablen $X_1,\ldots,X_n$ sind Poissonverteilt, wobei λ eine (unbekannte) positive Zahl ist.
- Wir zeigen mit Hilfe von Theorem 2.3, dass $\overline{X}_n$ ein suffizienter Schätzer für λ ist.
- Es gilt

$$p(x_1,\ldots,x_n;\lambda) = \begin{cases} \frac{\lambda^{x_1}}{x_1!}\, e^{-\lambda} \ldots \frac{\lambda^{x_n}}{x_n!}\, e^{-\lambda}\,, & \text{falls } x_1,\ldots,x_n\in\mathbb{N},\\[2ex] 0\,, & \text{sonst.}\end{cases}$$

- Hieraus ergibt sich, dass

$$p(x_1,\ldots,x_n;\lambda) = \underbrace{\lambda^{x_1+\ldots+x_n} e^{-n\lambda}}_{=g(x_1+\ldots+x_n,\lambda)}\ \underbrace{\frac{1}{x_1!\ldots x_n!}\, \mathbb{I}\bigl((x_1,\ldots,x_n)\in\mathbb{N}^n\bigr)}_{=h(x_1,\ldots,x_n)}\,.$$

- Die Wahrscheinlichkeitsfunktion $p(x_1,\ldots,x_n;\lambda)$ lässt sich deshalb auf die in (43) gegebene Weise darstellen.
- Wegen Theorem 2.3 ist somit $\overline{X}_n$ ein suffizienter Schätzer für λ.

Wir betrachten nun den *absolutstetigen Fall*, d.h.,

- für jedes $\theta\in\Theta$ gelte

$$P_\theta(B) = \int_B f(y;\theta)\,dy\,,\qquad \forall\, B\in\mathcal{B}(\mathbb{R})\,,$$

  wobei $f(\cdot\,;\theta)$ die Dichte von P_θ ist.

- Die Menge $\{x\in\mathbb{R} : f(x;\theta)>0\}$ sei die Vereinigung von endlich vielen Intervallen, die nicht von $\theta\in\Theta$ abhängen.

- Außerdem sei $\varphi:\mathbb{R}^n\to\mathbb{R}^m$ eine Stichprobenfunktion, so dass auch die Verteilung von $\varphi(X_1,\ldots,X_n)$ absolutstetig ist. Die Dichte $f_{\varphi(X_1,\ldots,X_n)}(t;\theta)$ von $\varphi(X_1,\ldots,X_n)$ sei stetig in t.

- Dann gilt

$$P_\theta(\varphi(X_1,\ldots,X_n)=t) = 0\,,\qquad \forall\, t\in\mathbb{R}^m\,.$$

- Um den Begriff der Suffizienz definieren zu können, wird deshalb ein allgemeineres Konzept für bedingte Wahrscheinlichkeiten benötigt, als es in Abschnitt WR–2.6 eingeführt worden ist.

- Dabei erfolgt die (maßtheoretische) Definition der bedingten Wahrscheinlichkeit $P_\theta\bigl((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\bigr)$ mit Hilfe des *Satzes von Radon-Nikodym*, vgl. Abschnitt 2.3.7.

- Wir wollen hier zunächst einen anderen (teilweise heuristischen, dafür aber intuitiveren) Zugang zu diesem Begriff betrachten und die bedingte Wahrscheinlichkeit $P_\theta\bigl((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\bigr)$ direkt als Grenzwert auffassen (ohne den allgemeinen maßtheoretischen Zugang zu verwenden): Falls

$$f_{\varphi(X_1,\ldots,X_n)}(t;\theta)>0 \qquad\text{und somit}\qquad P_\theta\bigl(\varphi(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\bigr)>0\,,\qquad \forall\,\varepsilon>0\,, \tag{44}$$

dann setzen wir

$$P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\big)=\lim_{\varepsilon\downarrow 0} P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\big)\,, \tag{45}$$

wobei vorausgesetzt wird, dass der Grenzwert existiert. Ansonsten setzen wir

$$P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\big)=0\,.$$

Definition Der Schätzer $\varphi(X_1,\ldots,X_n)$ für θ heißt *suffizient*, falls für beliebige $B\in\mathcal{B}(\mathbb{R}^n)$ und $t\in\mathbb{R}^m$ die bedingte Wahrscheinlichkeit $P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\big)$ *nicht* von θ abhängt.

Beachte

- Ähnlich wie im diskreten Fall (vgl. Lemma 2.2) kann man auch im absolutstetigen Fall zunächst eine (hinreichende) Bedingung für die Suffizienz von Schätzern angeben, die direkt an die Definition dieser Güteeigenschaft anknüpft.
- Um die Gültigkeit dieser Bedingung überprüfen zu können, ist dabei jedoch die Kenntnis der Dichte des Schätzers $\varphi(X_1,\ldots,X_n)$ erforderlich.

Lemma 2.3 *Sei $f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\theta)$ die Dichte der Zufallsstichprobe $(X_1,\ldots,X_n)$, und $f_{\varphi(X_1,\ldots,X_n)}(t;\theta)$ sei die Dichte von $\varphi(X_1,\ldots,X_n)$. Falls für beliebige $x_1,\ldots,x_n\in\mathbb{R}$ mit*

$$f_{\varphi(X_1,\ldots,X_n)}(\varphi(x_1,\ldots,x_n);\theta)>0 \tag{46}$$

der Quotient $f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\theta)/f_{\varphi(X_1,\ldots,X_n)}(\varphi(x_1,\ldots,x_n);\theta)$ nicht von $\theta\in\Theta$ abhängt, dann ist durch $\varphi(X_1,\ldots,X_n)$ ein suffizienter Schätzer für θ gegeben.

Beweis

- Sei $t\in\mathbb{R}^m$ so gewählt, dass (44) gilt.
- Dann ist

$$\begin{aligned}
&P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\big)\\
&=\lim_{\varepsilon\downarrow 0} P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\big)\\
&=\lim_{\varepsilon\downarrow 0}\frac{P_\theta\big((X_1,\ldots,X_n)\in B,\ \varphi(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\big)}{P_\theta\big(\varphi(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\big)}\\
&=\lim_{\varepsilon\downarrow 0}\frac{\int_B \mathbb{1}\big(\varphi(x_1,\ldots,x_n)\in(t-\varepsilon,t+\varepsilon)\big)f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\theta)\,d(x_1,\ldots,x_n)}{\displaystyle\int_{(t-\varepsilon,t+\varepsilon)} f_{\varphi(X_1,\ldots,X_n)}(y;\theta)\,dy}\\
&=\lim_{\varepsilon\downarrow 0}\frac{1}{(2\varepsilon)^m}\int_B \mathbb{1}\big(\varphi(x_1,\ldots,x_n)\in(t-\varepsilon,t+\varepsilon)\big)\frac{f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\theta)}{f_{\varphi(X_1,\ldots,X_n)}(\varphi(x_1,\ldots,x_n);\theta)}\,d(x_1,\ldots,x_n)\,.
\end{aligned}$$

- Hieraus ergibt sich, dass die bedingte Wahrscheinlichkeit $P_\theta\big((X_1,\ldots,X_n)\in B \mid \varphi(X_1,\ldots,X_n)=t\big)$ nicht von $\theta\in\Theta$ abhängt, falls der Quotient $f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\theta)/f_{\varphi(X_1,\ldots,X_n)}(\varphi(x_1,\ldots,x_n);\theta)$ diese Eigenschaft für alle $x_1,\ldots,x_n\in\mathbb{R}$ besitzt, für die (46) gilt. □

Beispiele

1. *Normalverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta\in\Theta\} = \{\mathrm{N}(\mu,\sigma^2),\, \mu\in\mathbb{R}\}$, wobei die Varianz $\sigma^2>0$ bekannt sei.
   - Wir zeigen, dass $\overline{X}_n$ ein suffizienter Schätzer für μ ist.
   - Für die gemeinsame Dichte $f_{(X_1,\ldots,X_n)}$ der Zufallsstichprobe $(X_1,\ldots,X_n)$ gilt

$$\begin{aligned} f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\exp\Bigl(-\frac{(x_i-\mu)^2}{2\sigma^2}\Bigr)\\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Bigl(-\sum_{i=1}^n\frac{(x_i-\mu)^2}{2\sigma^2}\Bigr)\\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Bigl(-\sum_{i=1}^n\frac{(x_i-\overline{x}_n+\overline{x}_n-\mu)^2}{2\sigma^2}\Bigr)\\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Bigl(-\sum_{i=1}^n\frac{(x_i-\overline{x}_n)^2}{2\sigma^2}-\frac{n(\overline{x}_n-\mu)^2}{2\sigma^2}\Bigr), \end{aligned}$$

   wobei in der letzten Gleichheit die Tatsache genutzt wurde, dass

$$\sum_{i=1}^n(x_i-\overline{x}_n)(\overline{x}_n-\mu) = (\overline{x}_n-\mu)\sum_{i=1}^n(x_i-\overline{x}_n) = 0\,.$$

   - Weil $\overline{X}_n\sim\mathrm{N}(\mu,\sigma^2/n)$, gilt somit

$$\begin{aligned} \frac{f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\mu)}{f_{\overline{X}_n}(\overline{x}_n;\mu)} &= \frac{\dfrac{1}{(2\pi\sigma^2)^{n/2}}\exp\Bigl(-\sum_{i=1}^n\dfrac{(x_i-\overline{x}_n)^2}{2\sigma^2}-\dfrac{n(\overline{x}_n-\mu)^2}{2\sigma^2}\Bigr)}{\dfrac{1}{(2\pi\sigma^2/n)^{1/2}}\exp\Bigl(\dfrac{-n(\overline{x}_n-\mu)^2}{2\sigma^2}\Bigr)}\\ &= \frac{1}{\sqrt{n}(2\pi\sigma^2)^{(n-1)/2}}\exp\Bigl(-\sum_{i=1}^n\frac{(x_i-\overline{x}_n)^2}{2\sigma^2}\Bigr), \end{aligned}$$

   wobei der letzte Ausdruck nicht von μ abhängt.
   - Aus Lemma 2.3 ergibt sich nun, dass $\overline{X}_n$ ein suffizienter Schätzer für μ ist.

2. *Exponentialverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta\in\Theta\} = \{\mathrm{Exp}(\lambda),\, \lambda>0\}$.
   - Wir zeigen zunächst, dass $Y=X_1+\ldots+X_n$ ein suffizienter Schätzer für λ ist.
   - Für die gemeinsame Dichte $f_{(X_1,\ldots,X_n)}$ der Zufallsstichprobe $(X_1,\ldots,X_n)$ gilt

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\lambda) = \lambda^n\exp\Bigl(-\lambda\sum_{i=1}^n x_i\Bigr)\,.$$

   - Weil Y Erlang-verteilt ist mit der Dichte

$$f_Y(y;\lambda) = \frac{\lambda^n y^{n-1}e^{-\lambda y}}{(n-1)!}\,,\qquad \forall\, y>0\,,$$

   ergibt sich somit, dass

$$\begin{aligned} \frac{f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n;\lambda)}{f_Y(x_1+\ldots+x_n;\lambda)} &= \frac{\lambda^n e^{-\lambda(x_1+\ldots+x_n)}}{\dfrac{\lambda^n(x_1+\ldots+x_n)^{n-1}e^{-\lambda(x_1+\ldots+x_n)}}{(n-1)!}}\\ &= \frac{(n-1)!}{(x_1+\ldots+x_n)^{n-1}}\,. \end{aligned}$$

- Aus Lemma 2.3 ergibt sich nun, dass Y ein suffizienter Schätzer für λ ist.
- Mit einer ähnlichen Überlegung wie im Beweis von Lemma 2.1 kann man auch im absolutstetigen Fall zeigen, dass $g \circ \varphi(X_1, \ldots, X_n)$ ein suffizienter Schätzer für θ ist, falls $\varphi(X_1, \ldots, X_n)$ diese Eigenschaft besitzt und $g : \mathbb{R}^m \to \mathbb{R}^m$ eine bijektive Borel-messbare Abbildung ist.
- Hieraus ergibt sich, dass auch n/Y ein suffizienter Schätzer für λ ist, für den außerdem
$$n/Y \xrightarrow{\text{f.s.}} \lambda$$
für $n \to \infty$ (wegen des starken Gesetzes der großen Zahlen) gilt; vgl. auch Abschnitt 2.4.1.

Beachte

- Es ist jedoch oft schwierig, die Dichte des Schätzers $\varphi(X_1, \ldots, X_n)$ zu bestimmen und die Gültigkeit der in Lemma 2.3 angegebenen Bedingung nachzuprüfen.
- So wie im diskreten Fall liefert dann die folgende Variante des *Faktorisierungssatzes von Neyman-Fisher* eine alternative (leichter nachprüfbare) Bedingung für die Suffizienz.

Theorem 2.4 *Der Schätzer $\varphi(X_1, \ldots, X_n)$ für θ ist genau dann suffizient, wenn es Borel-messbare Funktionen $g : \mathbb{R}^m \times \Theta \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}$ gibt, so dass*
$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \theta) = g\bigl(\varphi(x_1, \ldots, x_n), \theta\bigr)\, h(x_1, \ldots, x_n)\,, \qquad \forall\, x_1, \ldots, x_n \in \mathbb{R},\ \theta \in \Theta\,.$$

Der *Beweis* von Theorem 2.4 geht über den Rahmen dieser einführenden Vorlesung hinaus; vgl. beispielsweise Abschnitt 3.3 in H. Pruscha (2000) *Vorlesungen über Mathematische Statistik*, Teubner-Verlag, Stuttgart, oder Abschnitt 2.6 in E.L. Lehmann (1997) *Testing Statistical Hypotheses*, 2nd ed., Springer-Verlag, New York.

Beispiel *Normalverteilte Stichprobenvariablen* (Fortsetzung)

- Es gelte $\{P_\theta,\ \theta \in \Theta\} = \{\mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}$, wobei sowohl $\mu \in \mathbb{R}$ als auch $\sigma^2 > 0$ unbekannt sei.
- Wir zeigen, dass
$$\widehat{\theta}(X_1, \ldots, X_n) = \Bigl(\overline{X}_n,\ S_n^2\Bigr)$$
ein suffizienter Schätzer für $\theta = (\mu, \sigma^2)$ ist.
- Zur Erinnerung: Für die gemeinsame Dichte $f_{(X_1,\ldots,X_n)}$ der Zufallsstichprobe $(X_1, \ldots, X_n)$ gilt
$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Bigl(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}\Bigr)\,.$$
- Außerdem gilt
$$\begin{aligned}\sum_{i=1}^n (x_i-\mu)^2 &= \sum_{i=1}^n \bigl(x_i - \overline{x}_n + (\overline{x}_n - \mu)\bigr)^2 \\ &= (n-1)s_n^2 + n(\overline{x}_n - \mu)^2\,,\end{aligned}$$
d.h., $f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \mu, \sigma^2)$ lässt sich als Funktion von $(\overline{x}_n, s_n^2)$ und θ darstellen.
- Wegen Theorem 2.4 ist somit $\widehat{\theta}(X_1, \ldots, X_n) = \bigl(\overline{X}_n, S_n^2\bigr)$ ein suffizienter Schätzer für $\theta = (\mu, \sigma^2)$.

2.3.4 Vollständigkeit

Um in Abschnitt 2.3.5 die Frage untersuchen zu können, unter welchen Bedingungen erwartungstreue Schätzer mit minimaler Varianz existieren und wie man solche Schätzer gewinnen kann, benötigen wir noch eine weitere Eigenschaft von Punktschätzern.

Definition Sei $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ eine beliebige Stichprobenfunktion. Der Schätzer $\varphi(X_1, \ldots, X_n)$ für θ heißt *vollständig*, falls für jede messbare Funktion $g : \mathbb{R}^m \to \mathbb{R}$ aus der Gültigkeit von

$$\mathbb{E}_\theta \big|g\big(\varphi(X_1, \ldots, X_n)\big)\big| < \infty \qquad \text{und} \qquad \mathbb{E}_\theta g\big(\varphi(X_1, \ldots, X_n)\big) = 0\,, \qquad \forall\, \theta \in \Theta$$

folgt, dass $P_\theta\big(g\big(\varphi(X_1, \ldots, X_n)\big) = 0\big) = 1$ für jedes $\theta \in \Theta$.

Lemma 2.4 *Für jede messbare Funktion $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ und für jede messbare Abbildung $g' : \mathbb{R}^m \to \mathbb{R}^m$ ist der Schätzer $g'\big(\varphi(X_1, \ldots, X_n)\big)$ für θ vollständig, falls $\varphi(X_1, \ldots, X_n)$ diese Eigenschaft besitzt.*

Der *Beweis* von Lemma 2.4 ergibt sich unmittelbar aus der Definition der Vollständigkeit.

Beachte Für einige Verteilungsfamilien $\{P_\theta,\, \theta \in \Theta\}$ ergibt sich die Vollständigkeit von Punktschätzern direkt aus der Definition dieses Begriffes, ohne dass weitere analytische Hilfsmittel erforderlich sind.

Beispiele

1. *Bernoulli-verteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Bin}(1, p),\, p \in (0, 1)\}$.
   - Zur Erinnerung: In Abschnitt 2.3.3 hatten wir gezeigt, dass $Y = X_1 \ldots + X_n$ und damit auch $\overline{X}_n$ suffiziente Schätzer für p sind.
   - Wir zeigen nun, dass $Y = X_1 \ldots + X_n$ und damit (wegen Lemma 2.4) auch $\overline{X}_n$ vollständige Schätzer für p sind.
   - Sei $g : \mathbb{R} \to \mathbb{R}$ eine Funktion, so dass $\mathbb{E}_p g(Y) = 0$ für jedes $p \in (0, 1)$.
   - Dann gilt für jedes $p \in (0, 1)$

     $$\begin{aligned} 0 = \mathbb{E}_p g(Y) &= \sum_{i=0}^{n} g(i) \binom{n}{i} p^i (1-p)^{n-i} \\ &= (1-p)^n \sum_{i=0}^{n} g(i) \binom{n}{i} \Big(\frac{p}{1-p}\Big)^i \end{aligned}$$

     und damit für jedes $t \in (0, \infty)$

     $$0 = \sum_{i=0}^{n} g(i) \binom{n}{i} t^i\,.$$

   - Hieraus folgt, dass jeder Koeffizient des Polynoms auf der rechten Seite dieser Gleichung Null sein muss, d.h.

     $$g(i) = 0\,, \qquad \forall\, i \in \{0, 1, \ldots, n\}\,.$$

2. *Poisson-verteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Poi}(\lambda),\, \lambda > 0\}$.
   - Dann ist $Y = X_1 + \ldots + X_n$ für λ vollständig, denn

- es gilt $Y \sim \mathrm{Poi}(n\lambda)$, vgl. Übungsaufgabe WR–5.4, und somit gilt
$$0 = \mathbb{E}_\lambda g(Y) = \sum_{k=0}^{\infty} g(k) \frac{(n\lambda)^k}{k!}\, e^{-n\lambda}\,, \qquad \forall\, \lambda > 0$$
bzw. äquivalent hierzu
$$0 = \sum_{k=0}^{\infty} g(k) \frac{(n\lambda)^k}{k!}\,, \qquad \forall\, \lambda > 0$$
genau dann, wenn $g(k) = 0$ für jedes $k = 0, 1, \ldots$.
- Zur Erinnerung: In Abschnitt 2.3.3 hatten wir außerdem mit Hilfe des Faktorisierungssatzes von Neyman-Fisher (vgl.Theorem 2.3) gezeigt, dass $Y = X_1 + \ldots + X_n$ auch suffizient für λ ist.

3. *Gleichverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{U}(0,\theta),\, \theta > 0\}$, d.h., die Stichprobenvariablen $X_1, \ldots, X_n$ sind gleichverteilt über dem Intervall $(0,\theta)$, wobei $\theta > 0$ eine unbekannte Zahl ist.
   - Wir zeigen nun, dass $\max\{X_1, \ldots, X_n\}$ ein vollständiger Schätzer für θ ist.
   - Für die Dichte $f_{\max\{X_1,\ldots,X_n\}}(t;\theta)$ von $\max\{X_1, \ldots, X_n\}$ gilt (vgl. Abschnitt 1.4.3)
$$f_{\max\{X_1,\ldots,X_n\}}(t;\theta) = \begin{cases} nt^{n-1}\theta^{-n}\,, & \text{falls } t \in (0,\theta), \\ 0 & \text{sonst.} \end{cases}$$
   - Sei $g : \mathbb{R} \to \mathbb{R}$ eine Funktion, so dass $\mathbb{E}_\theta g(\max\{X_1, \ldots, X_n\}) = 0$ für jedes $\theta \in (0,\infty)$.
   - Dann gilt für jedes $\theta \in (0,\infty)$
$$\begin{aligned} 0 = \frac{d}{d\theta}\mathbb{E}_\theta g(\max\{X_1, \ldots, X_n\}) &= \frac{d}{d\theta} \int_0^\theta g(t) nt^{n-1}\theta^{-n}\, dt \\ &\overset{\text{Produktregel}}{=} \theta^{-n} \frac{d}{d\theta} \int_0^\theta g(t) nt^{n-1}\, dt + \Bigl(\frac{d}{d\theta}\theta^{-n}\Bigr) \underbrace{\int_0^\theta g(t) nt^{n-1}\, dt}_{0} \\ &= \theta^{-n} n g(\theta)\theta^{n-1} + 0 \\ &= \theta^{-1} n g(\theta)\,. \end{aligned}$$
   - Hieraus folgt, dass $g(\theta) = 0$ für jedes $\theta \in (0,\infty)$.

Beachte

- Für einige Verteilungsfamilien $\{P_\theta,\, \theta \in \Theta\}$ ergibt sich die Vollständigkeit von Punktschätzern *nicht* direkt aus der Definition dieses Begriffes, sondern es sind zusätzliche analytische Hilfsmittel erforderlich.
- Insbesondere ist der folgende *Eindeutigkeitssatz für die Laplace-Transformation* von σ-endlichen Maßen nützlich.
- Dabei setzen wir voraus, dass $m = 1$ und dass $\Theta \subset \mathbb{R}$ ein (offenes) Intervall ist.

Lemma 2.5 *Seien G_1 und G_2 zwei σ-endliche Maße über $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, die entweder beide diskret oder beide absolutstetig sind. Falls die Integrale*
$$I_j(\theta) = \int_{\mathbb{R}} e^{\theta y} G_j(dy)\,, \qquad \forall\, \theta \in \Theta,\ j = 1,2 \tag{47}$$
existieren (und endlich sind) und falls $I_1(\theta) = I_2(\theta)$ für jedes $\theta \in \Theta$, dann gilt $G_1 = G_2$.

Beweis

- Wir nehmen zunächst an, dass $[-a,a] \subset \Theta$ für ein $a > 0$. Dann gilt insbesondere

$$I_j(0) = \int_{\mathbb{R}} G_j(dy) < \infty\,, \qquad j = 1,2\,,$$

d.h., G_1 und G_2 sind endliche Maße mit $G_1(\mathbb{R}) = G_2(\mathbb{R})$.

  - Wir können deshalb o.B.d.A. annehmen, dass G_1 und G_2 Wahrscheinlichkeitsmaße sind.
  - Außerdem kann man die Funktionen

$$I'_j(z) = \int_{\mathbb{R}} e^{(s+it)y} G_j(dy)$$

der komplexen Variablen $z = s + it$ definieren, die dann in dem Teilgebiet $(-a,a) \times (-\infty,\infty)$ der komplexen Ebene holomorph sind.

  - Weil

$$I'_1(z) = I'_2(z)$$

für jedes $z = s + it$ mit $s \in [-a,a]$ und $t = 0$, gilt diese Gleichung wegen des Identitätssatzes für holomorphe Funktionen (vgl. Abschnitt 8.1 in R. Remmert (1992) *Funktionentheorie 1*, Springer, Berlin) wegen der eindeutigen Fortsetzbarkeit auch für jedes $z = s + it$ aus $(-a,a) \times (-\infty,\infty)$.

  - Insbesondere gilt somit

$$\int_{\mathbb{R}} e^{ity} G_1(dy) = \int_{\mathbb{R}} e^{ity} G_2(dy)\,, \qquad \forall\, t \in \mathbb{R}\,. \tag{48}$$

  - Wegen des Eindeutigkeitssatzes für charakteristische Funktionen (vgl. Korollar WR–5.5) ergibt sich nun aus (48), dass $G_1 = G_2$.

- Falls es kein $a > 0$ gibt, so dass $[-a,a] \subset \Theta$, dann betrachten wir die (endlichen) Maße G_1^*, G_2^* mit $G_j^*(dy) = e^{\theta_0 y} G_j(dy)$ für ein $\theta_0 \in \Theta$.
  - Dann gibt es ein $a > 0$, so dass G_1^*, G_2^* die Bedingungen des Lemmas für alle $\theta \in [-a,a]$ erfüllen.
  - Aus dem ersten Teil des Beweises folgt somit, dass $G_1^* = G_2^*$.
  - Im diskreten Fall ergibt sich hieraus für jedes Atom $x \in \mathbb{R}$ von G_1^* bzw. G_2^*, dass

$$G_1(\{x\}) = e^{-\theta_0 x} G_1^*(\{x\}) = e^{-\theta_0 x} G_2^*(\{x\}) = G_2(\{x\})$$

und somit $G_1 = G_2$.

  - Im absolutstetigen Fall impliziert $G_1^* = G_2^*$ die Gültigkeit von $G_1 = G_2$ wegen der Eindeutigkeitseigenschaft von Radon-Nikodym-Dichten. □

Beispiele

1. *Exponentialverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Exp}(\lambda),\, \lambda > 0\}$. Dann ist $Y = X_1 + \ldots + X_n$ Erlang-verteilt mit der Dichte

$$f_Y(y;\lambda) = \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!}\,, \qquad \forall\, y > 0\,.$$

   - Hieraus folgt, dass Y nicht nur ein suffizienter Schätzer (vgl. Abschnitt 2.3.3), sondern auch ein vollständiger Schätzer für λ ist, denn es gilt

$$0 = \mathbb{E}_\lambda g(Y) = \lambda^n \int_0^\infty \frac{g(y) y^{n-1}\, e^{-\lambda y}}{(n-1)!}\, dy\,, \qquad \forall\, \lambda > 0$$

genau dann, wenn

$$\int_0^\infty e^{-\lambda y} g^+(y) y^{n-1}\, dy = \int_0^\infty e^{-\lambda y} g^-(y) y^{n-1}\, dy\,, \qquad \forall\, \lambda > 0\,, \tag{49}$$

wobei $g^+(y) = \max\{0, g(y)\}$ und $g^-(y) = \max\{0, -g(y)\}$.

- Weil dabei $\mathbb{E}_\lambda |g(Y)| < \infty$ vorausgesetzt wird, sind mit $G_1(B) = \int_B g^+(y) y^{n-1} e^{-y}\, dy$ und $G_2(B) = \int_B g^-(y) y^{n-1} e^{-y}\, dy$ zwei σ-endliche Maße G_1 und G_2 gegeben.
- Wegen Lemma 2.5 gilt somit (49) genau dann, wenn $G_1 = G_2$, d.h., $g^+(y) = g^-(y)$ bzw. $g(y) = 0$ für fast jedes $y \in \mathbb{R}$.

2. *Normalverteilte Stichprobenvariablen*
   - Es gelte $\{P_\theta,\ \theta \in \Theta\} = \{\mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R}\}$, wobei die Varianz $\sigma^2 > 0$ bekannt sei.
   - Wir zeigen, dass $\overline{X}_n$ ein vollständiger Schätzer für μ ist.
   - Weil $\overline{X}_n \sim \mathrm{N}(\mu, \sigma^2/n)$, gilt somit

$$f_{\overline{X}_n}(y;\mu) = \frac{1}{(2\pi\sigma^2/n)^{1/2}} \exp\Bigl(\frac{-n(y-\mu)^2}{2\sigma^2}\Bigr)$$

   - Hieraus folgt, dass

$$0 = \mathbb{E}_\mu g(\overline{X}_n) = \int_{\mathbb{R}} g(y) \frac{1}{(2\pi\sigma^2/n)^{1/2}} \exp\Bigl(\frac{-n(y-\mu)^2}{2\sigma^2}\Bigr)\, dy\,, \qquad \forall\, \mu \in \mathbb{R}$$

   genau dann, wenn

$$0 = \int_{\mathbb{R}} \exp\Bigl(\frac{n\mu}{\sigma^2}\, y\Bigr) g(y) \exp\Bigl(\frac{-ny^2}{2\sigma^2}\Bigr)\, dy\,, \qquad \forall\, \mu \in \mathbb{R}\,.$$

   - Hieraus und aus Lemma 2.5 ergibt sich nun (genauso wie in dem vorhergehenden Beispiel exponentialverteilter Stichprobenvariablen), dass dies genau dann gilt, wenn $g(y) = 0$ für fast jedes $y \in \mathbb{R}$.

2.3.5 Beste erwartungstreue Schätzer

In diesem Abschnitt setzen wir erneut voraus, dass der Parameter θ eine relle Zahl ist, d.h., es gelte $m = 1$ bzw. $\Theta \subset \mathbb{R}$.

Definition

- Sei $\theta^* : \mathbb{R}^n \to \Theta$ eine Stichprobenfunktion, so dass für jedes $\theta \in \Theta$

$$\mathbb{E}_\theta \theta^*(X_1, \ldots, X_n) = \theta \qquad \text{und} \qquad \mathrm{Var}_\theta \theta^*(X_1, \ldots, X_n) < \infty\,.$$

- Falls für jede Stichprobenfunktion $\widehat{\theta} : \mathbb{R}^n \to \Theta$ mit

$$\mathbb{E}_\theta \widehat{\theta}(X_1, \ldots, X_n) = \theta\,, \qquad \forall\, \theta \in \Theta$$

  die Ungleichung

$$\mathrm{Var}_\theta \theta^*(X_1, \ldots, X_n) \le \mathrm{Var}_\theta \widehat{\theta}(X_1, \ldots, X_n)\,, \qquad \forall\, \theta \in \Theta$$

  gilt, dann heißt $\theta^*(X_1, \ldots, X_n)$ *bester erwartungstreuer Schätzer* für θ.

Beachte

- Bevor wir zeigen, welche erwartungstreuen Schätzer beste erwartungstreue Schätzer sind, diskutieren wir zunächst einige grundlegende Eigenschaften solcher Schätzer.
- Als erstes zeigen wir den folgenden *Eindeutigkeitssatz* für beste erwartungstreue Schätzer.

Lemma 2.6 *Es gibt höchstens einen besten erwartungstreuen Schätzer für θ, d.h., falls $\theta^*(X_1,\ldots,X_n)$ ein bester erwartungstreuer Schätzer für θ ist, dann ist $\theta^*(X_1,\ldots,X_n)$ mit Wahrscheinlichkeit 1 eindeutig bestimmt.*

Beweis

- Wir führen einen indirekten Beweis und nehmen an, dass $\theta^*(X_1,\ldots,X_n)$ ein bester erwartungstreuer Schätzer für θ ist und dass es noch einen weiteren besten erwartungstreuen Schätzer $\theta'(X_1,\ldots,X_n)$ für θ gibt.
- Dann ist auch
$$\theta''(X_1,\ldots,X_n)=\frac{1}{2}\bigl(\theta^*(X_1,\ldots,X_n)+\theta'(X_1,\ldots,X_n)\bigr)$$
ein erwartungstreuer Schätzer für θ, und es gilt
$$\begin{aligned}
\operatorname{Var}_\theta\theta''(X_1,\ldots,X_n) &= \operatorname{Var}_\theta\Bigl(\frac{1}{2}\,\theta^*(X_1,\ldots,X_n)+\frac{1}{2}\,\theta'(X_1,\ldots,X_n)\Bigr)\\
&\overset{\text{Theorem WR-4.13}}{=} \frac{1}{4}\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)+\frac{1}{4}\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n)\\
&\qquad+\frac{1}{2}\operatorname{Cov}_\theta\bigl(\theta^*(X_1,\ldots,X_n),\theta'(X_1,\ldots,X_n)\bigr)\\
&\overset{\text{Theorem WR-4.11}}{\le} \frac{1}{4}\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)+\frac{1}{4}\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n)\\
&\qquad+\frac{1}{2}\Bigl(\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n)\Bigr)^{1/2}\\
&= \operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)\,,
\end{aligned}$$
wobei in der letzten Gleichheit die Annahme genutzt wurde, dass
$$\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)=\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n)\,.$$
- Hieraus folgt, dass die Ungleichung in dieser Rechnung eine Gleichheit sein muss, weil ansonsten $\theta^*(X_1,\ldots,X_n)$ nicht bester erwartungstreuer Schätzer für θ wäre.
- Mit anderen Worten: Der Korrelationskoeffizient von $\theta^*(X_1,\ldots,X_n)$ und $\theta'(X_1,\ldots,X_n)$ ist gleich 1.
- Wegen Theorem WR–4.12 gibt es somit Zahlen $a(\theta),b(\theta)\in\mathbb{R}$, so dass mit Wahrscheinlichkeit 1
$$\theta'(X_1,\ldots,X_n)=a(\theta)\theta^*(X_1,\ldots,X_n)+b(\theta)\,.$$
- Aus den Rechenregeln für die Kovarianz (vgl. Theorem WR–4.11) ergibt sich nun, dass
$$\begin{aligned}
\operatorname{Cov}_\theta\bigl(\theta^*(X_1,\ldots,X_n),\theta'(X_1,\ldots,X_n)\bigr) &= \operatorname{Cov}_\theta\bigl(\theta^*(X_1,\ldots,X_n),a(\theta)\theta^*(X_1,\ldots,X_n)+b(\theta)\bigr)\\
&= \operatorname{Cov}_\theta\bigl(\theta^*(X_1,\ldots,X_n),a(\theta)\theta^*(X_1,\ldots,X_n)\bigr)\\
&= a(\theta)\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)\,.
\end{aligned}$$
- Weil wir andererseits gezeigt hatten, dass
$$\begin{aligned}
\operatorname{Cov}_\theta\bigl(\theta^*(X_1,\ldots,X_n),\theta'(X_1,\ldots,X_n)\bigr) &= \Bigl(\operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n)\Bigr)^{1/2}\\
&= \operatorname{Var}_\theta\theta^*(X_1,\ldots,X_n)\,,
\end{aligned}$$
folgt somit, dass $a(\theta)=1$.

- Wegen $\mathbb{E}_\theta\theta'(X_1,\ldots,X_n) = \mathbb{E}_\theta\theta^*(X_1,\ldots,X_n)$ ergibt sich hieraus, dass $b(\theta)=0$ bzw.
$$\theta'(X_1,\ldots,X_n) = \theta^*(X_1,\ldots,X_n)\,.$$
$\square$

Beachte

- Wir leiten nun ein Kriterium dafür her, dass ein erwartungstreuer Schätzer für θ gleichzeitig bester erwartungstreuer Schätzer ist.
- Das ist zunächst ein formales (d.h. nicht direkt nachprüfbares) Kriterium, welches jedoch bei der Herleitung des Hauptergebnisses dieses Abschnittes in Theorem 2.5 bzw. bei dessen Verallgemeinerung in Theorem 2.6 sehr nützlich ist.
- Hierfür benötigen wir noch den folgenden Begriff.

Definition Sei $\varphi:\mathbb{R}^n\to\mathbb{R}$ eine beliebige Stichprobenfunktion mit $\mathbb{E}_\theta\bigl((\varphi(X_1,\ldots,X_n))^2\bigr)<\infty$ für jedes $\theta\in\Theta$. Falls $\mathbb{E}_\theta\varphi(X_1,\ldots,X_n)=0$ für jedes $\theta\in\Theta$, dann sagen wir, dass $\varphi(X_1,\ldots,X_n)$ ein *erwartungstreuer Schätzer für* 0 ist.

Lemma 2.7 *Sei $\widehat\theta(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer für θ mit*
$$\mathbb{E}_\theta\Bigl(\bigl(\widehat\theta(X_1,\ldots,X_n)\bigr)^2\Bigr)<\infty\,,\qquad \forall\,\theta\in\Theta\,.$$
Dann ist $\widehat\theta(X_1,\ldots,X_n)$ genau dann bester erwartungstreuer Schätzer für θ, wenn $\widehat\theta(X_1,\ldots,X_n)$ unkorreliert ist mit jedem erwartungstreuen Schätzer für 0.

Beweis

- Sei $\widehat\theta(X_1,\ldots,X_n)$ bester erwartungstreuer Schätzer für θ, und sei $\varphi(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer für 0.
- Für jedes $a\in\mathbb{R}$ ist dann auch $\theta'(X_1,\ldots,X_n)=\widehat\theta(X_1,\ldots,X_n)+a\varphi(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer für θ, und es gilt
$$\begin{aligned}\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n) &= \operatorname{Var}_\theta\widehat\theta(X_1,\ldots,X_n)+2a\operatorname{Cov}_\theta\bigl(\widehat\theta(X_1,\ldots,X_n),\,\varphi(X_1,\ldots,X_n)\bigr)\\ &\quad +a^2\operatorname{Var}_\theta\varphi(X_1,\ldots,X_n)\,.\end{aligned}$$
- Wir führen einen indirekten Beweis und nehmen an, dass es ein $\theta_0\in\Theta$ gibt, so dass
$$\operatorname{Cov}_{\theta_0}\bigl(\widehat\theta(X_1,\ldots,X_n),\,\varphi(X_1,\ldots,X_n)\bigr)\neq 0\,,$$
- Dann gibt es ein $a\in\mathbb{R}$, so dass
$$2a\operatorname{Cov}_{\theta_0}\bigl(\widehat\theta(X_1,\ldots,X_n),\,\varphi(X_1,\ldots,X_n)\bigr)+a^2\operatorname{Var}_{\theta_0}\varphi(X_1,\ldots,X_n)<0\,.$$
- Hieraus folgt, dass
$$\operatorname{Var}_{\theta_0}\theta'(X_1,\ldots,X_n)<\operatorname{Var}_{\theta_0}\widehat\theta(X_1,\ldots,X_n)\,,$$
was im Widerspruch zu der Annahme steht, dass $\widehat\theta(X_1,\ldots,X_n)$ bester erwartungstreuer Schätzer für θ ist.
- Also ist
$$\operatorname{Cov}_\theta\bigl(\widehat\theta(X_1,\ldots,X_n),\,\varphi(X_1,\ldots,X_n)\bigr)=0\,,\qquad\forall\,\theta\in\Theta\,. \tag{50}$$
- Damit ist die Notwendigkeit der Bedingung (50) bweisen.
- Es gelte nun (50) für jeden erwartungstreuen Schätzer $\varphi(X_1,\ldots,X_n)$ für 0.

- Sei $\theta'(X_1,\ldots,X_n)$ ein anderer erwartungstreuer Schätzer für θ mit
$$\mathbb{E}_\theta\bigl((\theta'(X_1,\ldots,X_n))^2\bigr) < \infty\,, \qquad \forall\theta\in\Theta\,.$$
- Wegen der Identität
$$\theta'(X_1,\ldots,X_n) = \widehat{\theta}(X_1,\ldots,X_n) + \bigl(\theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)\bigr)$$
ergibt sich aus Theorem WR-4.13, dass
$$\begin{aligned}\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n) &= \operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n) + \operatorname{Var}_\theta\bigl(\theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)\bigr)\\ &\quad +2\operatorname{Cov}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n),\, \theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)\bigr)\\ &= \operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n) + \operatorname{Var}_\theta\bigl(\theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)\bigr)\,,\end{aligned}$$
wobei sich die letzte Gleichheit aus (50) ergibt, weil $\theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer $\varphi(X_1,\ldots,X_n)$ für 0 ist.
- Weil
$$\operatorname{Var}_\theta\bigl(\theta'(X_1,\ldots,X_n) - \widehat{\theta}(X_1,\ldots,X_n)\bigr) \ge 0\,,$$
ergibt sich nun, dass
$$\operatorname{Var}_\theta\theta'(X_1,\ldots,X_n) \ge \operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n)\,,$$
d.h., $\widehat{\theta}(X_1,\ldots,X_n)$ ist bester erwartungstreuer Schätzer für θ. □

Mit Hilfe von Lemma 2.7 lässt sich nun das folgende Ergebnis herleiten, das in der Literatur *Satz von Lehmann-Scheffé* genannt wird.

Theorem 2.5

- *Sei $\{P_\theta,\,\theta\in\Theta\}$ ein beliebiges parametrisches Modell mit $\Theta\subset\mathbb{R}$, und sei $\widehat{\theta}:\mathbb{R}^n\to\mathbb{R}$ eine beliebige Stichprobenfunktion, so dass $\widehat{\theta}(X_1,\ldots,X_n)$ ein vollständiger und suffizienter Schätzer für θ ist.*
- *Falls $\widehat{\theta}(X_1,\ldots,X_n)$ erwartungstreu für θ ist und falls*
$$\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)^2\bigr) < \infty\,, \qquad \forall\,\theta\in\Theta\,,$$
dann ist $\widehat{\theta}(X_1,\ldots,X_n)$ bester erwartungstreuer Schätzer für θ.

Beweis

- Sei $\widehat{\theta}(X_1,\ldots,X_n)$ ein vollständiger und suffizienter Schätzer für θ, der erwartungstreu für θ ist und dessen Varianz für jedes $\theta\in\Theta$ endlich ist.
- Wegen Lemma 2.7 genügt es zu zeigen, dass $\widehat{\theta}(X_1,\ldots,X_n)$ unkorreliert ist mit jedem erwartungstreuen Schätzer $\varphi(X_1,\ldots,X_n)$ für 0.
- Hierfür genügt es zu zeigen, dass
$$\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\,\varphi(X_1,\ldots,X_n)\bigr) = 0\,, \qquad \forall\,\theta\in\Theta\,, \tag{51}$$
weil $\mathbb{E}_\theta\varphi(X_1,\ldots,X_n) = 0$ für jedes $\theta\in\Theta$.

- Wir benutzen die folgende Formel der totalen Wahrscheinlichkeit:
$$\begin{aligned}
&P_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\varphi(X_1,\ldots,X_n)\in B\bigr)\\
&= \int\limits_{\mathbb{R}} P_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\varphi(X_1,\ldots,X_n)\in B \mid \widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\,, \qquad \forall\, B\in\mathcal{B}(\mathbb{R})\,,
\end{aligned}$$
die sich unmittelbar aus Theorem WR–2.6 ergibt, falls $\widehat{\theta}(X_1,\ldots,X_n)$ eine diskrete Zufallsvariable ist (ansonsten kann man die Gültigkeit dieser Formel mittels Grenzwertbildung so wie in (45) erklären).
- Hieraus folgt, dass
$$\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\,\varphi(X_1,\ldots,X_n)\bigr)=\int\limits_{\mathbb{R}} t\mathbb{E}_\theta\bigl(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\,,$$
wobei $\mathbb{E}_\theta\bigl(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)$ den Erwartungswert der (bedingten) Verteilung
$$\bigl\{P_\theta\bigl(\varphi(X_1,\ldots,X_n)\in B\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\,,\; B\in\mathcal{B}(\mathbb{R})\bigr\}$$
bezeichnet mit
$$\begin{aligned}
&P_\theta\bigl(\varphi(X_1,\ldots,X_n)\in B\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\\
&= \lim_{\varepsilon\downarrow 0} P_\theta\bigl(\varphi(X_1,\ldots,X_n)\in B\mid\widehat{\theta}(X_1,\ldots,X_n)\in(t-\varepsilon,t+\varepsilon)\bigr)\,.
\end{aligned}$$
- Weil $\widehat{\theta}(X_1,\ldots,X_n)$ suffizient ist, hängt $\mathbb{E}_\theta\bigl(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)$ lediglich von t, jedoch nicht von θ ab, d.h., es gibt eine (messbare) Funktion $g:\mathbb{R}\to\mathbb{R}$, so dass
$$g(t)=\mathbb{E}_\theta\bigl(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\,,\qquad \forall\,\theta\in\Theta\,.$$
- Somit gilt
$$\begin{aligned}
\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\,\varphi(X_1,\ldots,X_n)\bigr) &= \int\limits_{\mathbb{R}} tg(t)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\, g(\widehat{\theta}(X_1,\ldots,X_n))\bigr)\,,
\end{aligned}$$
d.h. $\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\,\varphi(X_1,\ldots,X_n)\bigr)=\mathbb{E}_\theta\bigl(\widehat{\theta}(X_1,\ldots,X_n)\, g(\widehat{\theta}(X_1,\ldots,X_n))\bigr)$.
- Hieraus ergibt sich die Gültigkeit von (51), denn aus der Formel der totalen Wahrscheinlichkeit folgt, dass
$$\begin{aligned}
\mathbb{E}_\theta g(\widehat{\theta}(X_1,\ldots,X_n)) &= \int\limits_{\mathbb{R}} g(t)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \int\limits_{\mathbb{R}} \mathbb{E}_\theta\bigl(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\bigr)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \mathbb{E}_\theta\varphi(X_1,\ldots,X_n)=0
\end{aligned}$$
und somit dass $P_\theta\bigl(g(\widehat{\theta}(X_1,\ldots,X_n))=0\bigr)=1$ für jedes $\theta\in\Theta$, weil $\widehat{\theta}(X_1,\ldots,X_n)$ vollständig ist. $\square$

2.3.6 Bedingte Erwartung; Ungleichung von Rao-Blackwell

- In den Abschnitten 2.3.3 und 2.3.4 hatten wir Beispiele von Schätzern $\widehat{\theta}(X_1,\ldots,X_n)$ für θ betrachtet, die zwar vollständig und suffizient, jedoch nicht erwartungstreu sind.

- Wenn außerdem noch ein weiterer Schätzer $\widetilde{\theta}(X_1,\ldots,X_n)$ für θ gegeben ist, der zwar erwartungstreu ist, jedoch weder vollständig noch suffizient sein muss, dann ergibt sich durch die folgende Verknüpfung von $\widetilde{\theta}(X_1,\ldots,X_n)$ und $\widehat{\theta}(X_1,\ldots,X_n)$ ein bester erwartungstreuer Schätzer für θ.
- Ähnlich wie im Beweis von Theorem 2.5 sei
$$\widetilde{g}(t) = \mathbb{E}_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n) = t\bigr) \tag{52}$$
der Erwartungswert der bedingten Verteilung
$$\bigl\{P_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \in B \mid \widehat{\theta}(X_1,\ldots,X_n) = t\bigr)\,,\ B \in \mathcal{B}(\mathbb{R})\bigr\}\,,$$
wobei
$$\begin{aligned} &P_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \in B \mid \widehat{\theta}(X_1,\ldots,X_n) = t\bigr) \\ &\quad= \lim_{\varepsilon\downarrow 0} P_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \in B \mid \widehat{\theta}(X_1,\ldots,X_n) \in (t-\varepsilon, t+\varepsilon)\bigr)\,. \end{aligned} \tag{53}$$
- Weil $\widehat{\theta}(X_1,\ldots,X_n)$ suffizient ist, hängt $\widetilde{g}(t)$ nicht von θ ab.
- Wir können somit die zusammengesetzte Stichprobenfunktion $\widetilde{g}\circ\widehat{\theta} : \mathbb{R}^n \to \mathbb{R}$ betrachten mit
$$\bigl(\widetilde{g}\circ\widehat{\theta}\bigr)(x_1,\ldots,x_n) = \widetilde{g}\bigl(\widehat{\theta}(x_1,\ldots,x_n)\bigr)\,.$$

Definition Die zufällige Stichprobenfunktion $\bigl(\widetilde{g}\circ\widehat{\theta}\bigr)(X_1,\ldots,X_n)$ wird in der Literatur die *bedingte Erwartung* von $\widetilde{\theta}(X_1,\ldots,X_n)$ bezüglich $\widehat{\theta}(X_1,\ldots,X_n)$ genannt und mit $\mathbb{E}\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n)\bigr)$ bezeichnet.

Auf ähnliche Weise wie Theorem 2.5 lässt sich dann die folgende Aussage aus Lemma 2.7 herleiten.

Theorem 2.6

- *Sei $\{P_\theta,\ \theta\in\Theta\}$ ein beliebiges parametrisches Modell mit $\Theta\subset\mathbb{R}$, und seien $\widetilde{\theta}, \widehat{\theta} : \mathbb{R}^n \to \mathbb{R}$ beliebige Stichprobenfunktionen, so dass $\widetilde{\theta}(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer und $\widehat{\theta}(X_1,\ldots,X_n)$ ein vollständiger und suffizienter Schätzer für θ ist, wobei $\mathbb{E}_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n)^2\bigr) < \infty$ für jedes $\theta\in\Theta$.*
- *Dann ist die bedingte Erwartung $\mathbb{E}\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n)\bigr)$ bester erwartungstreuer Schätzer für θ.*

Beweis

- Wegen Lemma 2.7 genügt es zu zeigen, dass $\mathbb{E}\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n)\bigr)$ quadratisch integierbar, erwartungstreu und unkorreliert mit jedem erwartungstreuen Schätzer $\varphi(X_1,\ldots,X_n)$ für 0 ist.
- Es gilt
$$\begin{aligned} \mathbb{E}_\theta\Bigl(\mathbb{E}\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n)\bigr)\Bigr) &= \mathbb{E}_\theta\Bigl(\bigl(\widetilde{g}\circ\widehat{\theta}\bigr)(X_1,\ldots,X_n)\Bigr) = \int_{\mathbb{R}} \widetilde{g}(t)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t) \\ &= \int_{\mathbb{R}} \mathbb{E}_\theta\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n) = t\bigr)\, dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t) = \mathbb{E}_\theta\widetilde{\theta}(X_1,\ldots,X_n) = \theta\,, \end{aligned}$$
d.h., $\mathbb{E}\bigl(\widetilde{\theta}(X_1,\ldots,X_n) \mid \widehat{\theta}(X_1,\ldots,X_n)\bigr)$ ist erwartungstreu.

- Auf die gleiche Weise ergibt sich für jeden erwartungstreuen Schätzer $\varphi(X_1,\ldots,X_n)$ für 0, dass

$$
\begin{aligned}
\mathbb{E}_\theta\Big(\mathbb{E}\,(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n))\varphi(X_1,\ldots,X_n)\Big) &= \mathbb{E}_\theta\Big((\widetilde{g}\circ\widehat{\theta})(X_1,\ldots,X_n)\varphi(X_1,\ldots,X_n)\Big)\\
&= \int\limits_{\mathbb{R}} \mathbb{E}_\theta\Big((\widetilde{g}\circ\widehat{\theta})(X_1,\ldots,X_n)\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\Big)\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \int\limits_{\mathbb{R}} \mathbb{E}_\theta\Big((\widetilde{g}(t)\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\Big)\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \int\limits_{\mathbb{R}} \widetilde{g}(t)\mathbb{E}_\theta\Big((\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\Big)\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \int\limits_{\mathbb{R}} \widetilde{g}(t)g(t)\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\,,
\end{aligned}
$$

wobei $g(t)=\mathbb{E}_\theta\big(\varphi(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\big)$ für jedes $\theta\in\Theta$.

- Hieraus ergibt sich nun genauso wie im Beweis von Theorem 2.5, dass

$$
\mathbb{E}_\theta\Big(\mathbb{E}\,(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n))\varphi(X_1,\ldots,X_n)\Big)=0\,.
$$

- Außerdem gilt, dass

$$
\begin{aligned}
\mathbb{E}_\theta\Big(\mathbb{E}\,(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n))^2\Big) &= \mathbb{E}_\theta\Big((\widetilde{g}\circ\widehat{\theta})^2(X_1,\ldots,X_n)\Big)\\
&= \int\limits_{\mathbb{R}} (\widetilde{g}^2(t))\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \int\limits_{\mathbb{R}} \Big(\mathbb{E}_\theta\big(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\big)\Big)^2\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&\le \int\limits_{\mathbb{R}} \mathbb{E}_\theta\big((\widetilde{\theta}(X_1,\ldots,X_n))^2\mid\widehat{\theta}(X_1,\ldots,X_n)=t\big)\,dF_{\widehat{\theta}(X_1,\ldots,X_n)}(t)\\
&= \mathbb{E}_\theta\widetilde{\theta}^2(X_1,\ldots,X_n)<\infty\,,
\end{aligned}
$$

wobei in der vorletzten Ungleichung genutzt wurde, dass $(\mathbb{E}\,X)^2\le\mathbb{E}\,(X^2)$ für jede Zufallsvariable X mit $\mathbb{E}\,(X^2)<\infty$. □

Korollar 2.2

- *Sei $\{P_\theta,\,\theta\in\Theta\}$ ein beliebiges parametrisches Modell mit $\Theta\subset\mathbb{R}$, und sei $\widehat{\theta}(X_1,\ldots,X_n)$ ein vollständiger und suffizienter Schätzer für θ.*
- *Falls $g:\mathbb{R}\to\mathbb{R}$ eine messbare Funktion ist, so dass*

$$
\mathbb{E}_\theta\big(g\big(\widehat{\theta}(X_1,\ldots,X_n)\big)^2\big)<\infty\,,\qquad\forall\,\theta\in\Theta
$$

und dass $g\big(\widehat{\theta}(X_1,\ldots,X_n)\big)$ ein erwartungstreuer Schätzer für θ ist, dann ist $g\big(\widehat{\theta}(X_1,\ldots,X_n)\big)$ bester erwartungstreuer Schätzer für θ.

Beweis Die Behauptung ergibt sich unmittelbar aus Theorem 2.6, denn kann man zeigen, dass

$$
\mathbb{E}\,\big(g\big(\widehat{\theta}(X_1,\ldots,X_n)\big)\mid\widehat{\theta}(X_1,\ldots,X_n)\big)=g\big(\widehat{\theta}(X_1,\ldots,X_n)\big)\,,
$$

vgl. Teilaussage 4 in Theorem 2.8. □

Wir betrachten nun noch eine abgeschwächte Version von Theorem 2.6, die besagt, dass durch den Übergang zur bedingten Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ auch dann eine Verbesserung der Schätzgenauigkeit erzielt werden kann, wenn *nicht* vorausgesetzt wird, dass $\widehat{\theta}(X_1,\ldots,X_n)$ vollständig ist.

Theorem 2.7

- *Seien $\widetilde{\theta},\widehat{\theta}:\mathbb{R}^n\to\mathbb{R}$ beliebige Stichprobenfunktionen, so dass $\widetilde{\theta}(X_1,\ldots,X_n)$ ein erwartungstreuer Schätzer und $\widehat{\theta}(X_1,\ldots,X_n)$ ein suffizienter Schätzer für θ ist, wobei $\mathbb{E}_\theta\left(\widetilde{\theta}(X_1,\ldots,X_n)^2\right)<\infty$ für jedes $\theta\in\Theta$.*
- *Dann ist die bedingte Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ ein erwartungstreuer Schätzer für θ mit*

$$\operatorname{Var}_\theta\Bigl(\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)\Bigr)\le\operatorname{Var}_\theta\widetilde{\theta}(X_1,\ldots,X_n)\,,\tag{54}$$

Beweis

- Im Beweis von Theorem 2.6 hatten wir gezeigt, dass

$$\mathbb{E}_\theta\Bigl(\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)\Bigr)=\mathbb{E}_\theta\widetilde{\theta}(X_1,\ldots,X_n)=\theta$$

und

$$\mathbb{E}_\theta\Bigl(\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)^2\Bigr)\le\mathbb{E}_\theta\widetilde{\theta}^2(X_1,\ldots,X_n)\,.$$

- Hieraus ergibt sich die Behauptung, weil $\operatorname{Var}X=\mathbb{E}(X^2)-(\mathbb{E}X)^2$ für jede Zufallsvariable X mit $\mathbb{E}(X^2)<\infty$. □

Beachte

- Die Ungleichung (54) wird in der Literatur *Ungleichung von Rao-Blackwell* genannt.
- Weil in Theorem 2.7 nicht vorausgesetzt wird, dass $\widehat{\theta}(X_1,\ldots,X_n)$ vollständig ist, muss die bedingte Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ nicht notwendig bester erwartungstreuer Schätzer für θ sein.

2.3.7 Maßtheoretische Definition der bedingten Erwartung

- In Abschnitt 2.3.6 wurde der Begriff der bedingten Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ eingeführt, wobei zunächst die Grenzwerte $P_\theta\left(\widetilde{\theta}(X_1,\ldots,X_n)\in B\mid\widehat{\theta}(X_1,\ldots,X_n)=t\right)$ in (53) betrachtet wurden.
- Diese Vorgehensweise besitzt den Vorteil, dass intuitiv klar wird, wieso $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ als Funktion von $\widehat{\theta}(X_1,\ldots,X_n)$ aufgefasst werden kann.
- Allerdings wurde dabei nicht näher untersucht,
  - ob bzw. unter welchen Bedingungen die Grenzwerte $P_\theta\left(\widetilde{\theta}(X_1,\ldots,X_n)\in B\mid\widehat{\theta}(X_1,\ldots,X_n)=t\right)$ in (53) existieren,
  - ob sie (bei gegebenem t) eine σ-additive Mengenfunktion bezüglich B, d.h. ein Wahrscheinlichkeitsmaß über $\mathcal{B}(\mathbb{R})$ bilden und
  - ob die Abbildung $t\to\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)=t\right)$ Borel-messbar ist.

Der folgende (maßtheoretische) Zugang zur bedingten Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1,\ldots,X_n)\mid\widehat{\theta}(X_1,\ldots,X_n)\right)$ ist zwar weniger intuitiv, besitzt jedoch den Vorteil, dass er mit den allgemeinen Rechenregeln der Maß- und Integrationstheorie begründet werden kann.

- Sei $X : \Omega \to \mathbb{R}$ eine beliebige Zufallsvariable über dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\mathbb{E}\,|X| < \infty$,
- und sei $\mathcal{G} \subset \mathcal{F}$ eine beliebige Teil-σ-Algebra von $\mathcal{F}$.
- Dann können wir das (endliche) signierte Maß $Q : \mathcal{G} \to \mathbb{R}$ betrachten, wobei

$$Q(A) = \int_A X(\omega) P(d\omega)\,, \qquad \forall\, A \in \mathcal{G}\,. \tag{55}$$

- Für jedes $A \in \mathcal{G}$ gilt somit die Implikation: $Q(A) = 0$, falls $P(A) = 0$.
- Mit anderen Worten: Das in (55) definierte signierte Maß Q ist *absolutstetig* bezüglich der Einschränkung von P auf die Teil-σ-Algebra $\mathcal{G}$.
- Aus dem *Satz von Radon-Nikodym* folgt nun, dass es eine $(\mathcal{G}, \mathcal{B}(\mathbb{R}))$-messbare Abbildung $Z : \Omega \to \mathbb{R}$ gibt, so dass

$$\int_A Z(\omega) P(d\omega) = \int_A X(\omega) P(d\omega)\,, \qquad \forall\, A \in \mathcal{G}\,, \tag{56}$$

  wobei die Funktionswerte der Abbildung Z mit Wahrscheinlichkeit 1 eindeutig bestimmt sind.
- Jede $(\mathcal{G}, \mathcal{B}[\mathbb{R}])$-messbare Abbildung $Z : \Omega \to \mathbb{R}$, für die (56) gilt, heißt eine Version der *bedingten Erwartung* von X bezüglich $\mathcal{G}$ und wird mit $\mathbb{E}\,(X \mid \mathcal{G})$ bezeichnet.

Aus der Definitionsgleichung (56) und aus den allgemeinen Rechenregeln für das Lebesgue-Integral ergeben sich die folgenden Eigenschaften der bedingten Erwartung, die wir hier lediglich (ohne Beweis) erwähnen.

Theorem 2.8 *Seien $X, Y : \Omega \to \mathbb{R}$ beliebige Zufallsvariablen über $(\Omega, \mathcal{F}, P)$ mit*

$$\mathbb{E}\,|X| < \infty\,, \qquad \mathbb{E}\,|Y| < \infty\,, \qquad \mathbb{E}\,|XY| < \infty\,,$$

und sei $\mathcal{G} \subset \mathcal{F}$ eine beliebige Teil-σ-Algebra von $\mathcal{F}$. Dann gilt

1. $\mathbb{E}\,(X \mid \{\emptyset, \Omega\}) = \mathbb{E}\,X$, $\mathbb{E}\,(X \mid \mathcal{F}) = X$,
2. $\mathbb{E}\,(aX + bY \mid \mathcal{G}) = a\,\mathbb{E}\,(X \mid \mathcal{G}) + b\,\mathbb{E}\,(Y \mid \mathcal{G})$ *für beliebige* $a, b \in \mathbb{R}$,
3. $\mathbb{E}\,(X \mid \mathcal{G}) \le \mathbb{E}\,(Y \mid \mathcal{G})$, *falls* $X \le Y$,
4. $\mathbb{E}\,(XY \mid \mathcal{G}) = Y\mathbb{E}\,(X \mid \mathcal{G})$, *falls Y eine $(\mathcal{G}, \mathcal{B}(\mathbb{R}))$-messbare Zufallsvariable ist,*
5. $\mathbb{E}\,\bigl(\mathbb{E}\,(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\bigr) = \mathbb{E}\,(X \mid \mathcal{G}_1)$, *falls $\mathcal{G}_1$ und $\mathcal{G}_2$ Teil-σ-Algebren von $\mathcal{F}$ sind mit $\mathcal{G}_1 \subset \mathcal{G}_2$,*
6. $\mathbb{E}\,(X \mid \mathcal{G}) = \mathbb{E}\,X$, *falls die σ-Algebren $\mathcal{G}$ und $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$ unabhängig sind, d.h., falls $P(A \cap A') = P(A)P(A')$ für beliebige $A \in \mathcal{G}$ und $A' \in \sigma(X)$.*

Beachte

- Aus den Teilaussagen 1 und 5 ergibt sich, dass $\mathbb{E}\,\bigl(\mathbb{E}\,(X \mid \mathcal{G})\bigr) = \mathbb{E}\,X$, wenn $\mathcal{G}_1 = \{\emptyset, \Omega\}$ und $\mathcal{G}_2 = \mathcal{G}$ gesetzt wird.
- Aus Teilaussage 2 ergibt sich, dass $\mathbb{E}\,(a \mid \mathcal{G}) = a$ für beliebige $a \in \mathbb{R}$ und $\mathcal{G} \subset \mathcal{F}$.
- Hieraus und aus Teilaussage 4 folgt, dass $\mathbb{E}\,(Y \mid \mathcal{G}) = Y$ für jede $(\mathcal{G}, \mathcal{B}(\mathbb{R}))$-messbare Zufallsvariable Y (wenn $X = 1$ gesetzt wird).

Beispiel

- Seien $X, Y : \Omega \to \mathbb{R}$ beliebige Zufallsvariablen über $(\Omega, \mathcal{F}, P)$ mit $\mathbb{E}\,|X| < \infty$, und sei $\mathcal{G} = Y^{-1}(\mathcal{B}(\mathbb{R}))$ die Teil-σ-Algebra von $\mathcal{F}$, die durch Y erzeugt wird.
- Dann heißt $\mathbb{E}\,(X \mid \mathcal{G})$ die bedingte Erwartung von X bezüglich Y, wobei auch die Schreibweise $\mathbb{E}\,(X \mid Y)$ benutzt wird.
- Insbesondere stimmt die bedingte Erwartung $\mathbb{E}\left(\widetilde{\theta}(X_1, \ldots, X_n) \mid \widehat{\theta}(X_1, \ldots, X_n)\right)$ von $\widetilde{\theta}(X_1, \ldots, X_n)$ bezüglich $\widehat{\theta}(X_1, \ldots, X_n)$ mit dem in Abschnitt 2.3.6 eingeführten Begriff überein, falls die Grenzwerte in (53) existieren und Wahrscheinlichkeitsmaße bilden, deren Erwartungswerte eine in t Borel-messbare Funktion ergeben.

2.4 Asymptotische Eigenschaften von Punktschätzern

- In den Abschnitten 2.3.1–2.3.5 untersuchten wir Güteeigenschaften von Punktschätzern, wobei in den meisten Fällen vorausgesetzt wurde, dass der Stichprobenumfang n beliebig, jedoch fest vorgegeben ist.
- Eine Ausnahme bildete lediglich die asymptotische Erwartungstreue, die eine *asymptotische* Güteeigenschaft ist.
- Wir diskutieren nun noch einige weitere asymptotische Eigenschaften von Punktschätzern für $n \to \infty$.

2.4.1 Konsistenz

Sei $\{P_\theta,\, \theta \in \Theta\}$ eine beliebige Verteilungsfamilie mit $\Theta \subset \mathbb{R}^m$.

Definition Für jedes $n \geq 1$ sei $\widehat{\theta} : \mathbb{R}^n \to \mathbb{R}^m$ eine beliebige Stichprobenfunktion. Der Schätzer $\widehat{\theta}(X_1, \ldots, X_n)$ für θ heißt

- *schwach konsistent*, falls $\widehat{\theta}(X_1, \ldots, X_n) \xrightarrow{\text{P}} \theta$ für jedes $\theta \in \Theta$ und für $n \to \infty$, d.h.,
$$\lim_{n\to\infty} P_\theta(|\,\widehat{\theta}(X_1, \ldots, X_n) - \theta| > \varepsilon) = 0\,, \qquad \forall\, \varepsilon > 0,\, \theta \in \Theta\,.$$
- *stark konsistent*, falls $\widehat{\theta}(X_1, \ldots, X_n) \xrightarrow{\text{f.s.}} \theta$ für jedes $\theta \in \Theta$ und für $n \to \infty$, d.h.,
$$P_\theta\Bigl(\lim_{n\to\infty} |\widehat{\theta}(X_1, \ldots, X_n) - \theta| = 0\Bigr) = 1\,, \qquad \forall\, \theta \in \Theta\,.$$

Beachte Wegen Korollar WR–5.1 ist jeder stark konsistente Schätzer gleichzeitig auch schwach konsistent.

Beispiel *Normalverteilte Stichprobenvariablen*

- Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}$, wobei sowohl μ als auch σ^2 unbekannt sei.
- Dann ergibt sich aus den Theoremen 1.2 bzw. 1.4, dass für $n \to \infty$
$$\overline{X}_n \xrightarrow{\text{f.s.}} \mu \qquad \text{und} \qquad S_n^2 \xrightarrow{\text{f.s.}} \sigma^2\,, \qquad \forall\, \mu \in \mathbb{R},\ \sigma^2 > 0\,. \tag{57}$$
- Wegen der Erhaltung der (fast sicheren) Konvergenz bei Addition bzw. Multiplikation (vgl. jeweils die Teilaussage 1 der Theoreme WR–5.8 bzw. WR–5.10) ergibt sich aus (57), dass für $n \to \infty$
$$\bigl|\bigl(\overline{X}_n, S_n^2\bigr) - (\mu, \sigma^2)\bigr| \xrightarrow{\text{f.s.}} 0\,, \tag{58}$$
d.h., der (erwartungstreue) Schätzer $\widehat{\theta}(X_1, \ldots, X_n)) = \bigl(\overline{X}_n, S_n^2\bigr)$ für $\theta = (\mu, \sigma^2)$ ist stark (und damit auch schwach) konsistent.

- Hieraus ergibt sich außerdem, dass auch der in Beispiel 5 des Abschnittes 2.2.2 hergeleitete (asymptotisch erwartungstreue) Maximum-Likelihood-Schätzer $\big(\widehat{\mu}(X_1,\ldots,X_n),\widehat{\sigma}^2(X_1,\ldots,X_n)\big)$ für $\theta=(\mu,\sigma^2)$ stark konsistent ist.

Beachte

- Weil die Theoreme 1.2 und 1.4 nicht nur für normalverteilte Stichprobenvariablen gelten, ist der Schätzer $\widehat{\theta}(X_1,\ldots,X_n))=\big(\overline{X}_n,S_n^2\big)$ für *jede* parametrische Verteilungsfamile konsistent, für die der Erwartungswert μ bzw. die Varianz σ^2 der Stichprobenvariablen zu den Komponenten des Parametervektors θ gehören.
- Falls $m=1$, dann kann man die folgende allgemeine Bedingung für die schwache Konsistenz von (asymptotisch erwartungstreuen) Schätzern formulieren.

Theorem 2.9 *Sei $\Theta\subset\mathbb{R}$, und für jedes $n\ge 1$ sei $\widehat{\theta}:\mathbb{R}^n\to\mathbb{R}$ eine beliebige Stichprobenfunktion, so dass*

$$\mathbb{E}_\theta\widehat{\theta}^2(X_1,\ldots,X_n)<\infty\,,\qquad\forall\,\theta\in\Theta\,.$$

Falls $\widehat{\theta}(X_1,\ldots,X_n)$ asymptotisch erwartungstreu ist, d.h.,

$$\lim_{n\to\infty}\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)=\theta\,,\qquad\forall\,\theta\in\Theta\,,\tag{59}$$

und falls die Schätzvarianz mit wachsendem Stichprobenumfang gegen 0 konvergiert, d.h.,

$$\lim_{n\to\infty}\operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n)=0\,,\qquad\forall\,\theta\in\Theta\,,\tag{60}$$

dann ist der Schätzer $\widehat{\theta}(X_1,\ldots,X_n)$ schwach konsistent.

Beweis Für beliebige $\theta\in\Theta$, $\varepsilon>0$ und $n\ge 1$ ergibt sich aus der Markov-Ungleichung (vgl. Formel WR–(4.73)), aus den Monotonie- und Linearitätseigenschaften des Erwartungswertes (vgl. Theorem WR–4.4) bzw. aus der Ungleichung von Cauchy-Schwarz (vgl. Theorem WR–4.11), dass

$$\begin{aligned}
P_\theta(|\,\widehat{\theta}(X_1,\ldots,X_n)-\theta|>\varepsilon) &\overset{\text{Formel WR}-(4.73)}{\le} \frac{\mathbb{E}_\theta\big(|\widehat{\theta}(X_1,\ldots,X_n)-\theta|\big)}{\varepsilon}\\
&\overset{\text{Theorem WR}-4.4}{\le} \frac{\mathbb{E}_\theta\big(|\widehat{\theta}(X_1,\ldots,X_n)-\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)|\big)}{\varepsilon}\\
&\qquad+\frac{|\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)-\theta|}{\varepsilon}\\
&\overset{\text{Theorem WR}-4.11}{\le} \frac{\big(\operatorname{Var}_\theta\widehat{\theta}(X_1,\ldots,X_n)\big)^{1/2}}{\varepsilon}+\frac{|\mathbb{E}_\theta\widehat{\theta}(X_1,\ldots,X_n)-\theta|}{\varepsilon}\,,
\end{aligned}$$

wobei der letzte Ausdruck für $n\to\infty$ gegen 0 konvergiert, falls (59) und (60) erfüllt sind. $\square$

Beachte

- Wir hatten bereits in Abschnitt 2.2.1 erwähnt, dass mit der Momenten-Methode konstruierte Schätzer stark konsistent sind, falls sie eindeutig bestimmt sind, d.h., falls das Gleichungssystem (3) eine eindeutig bestimmte Lösung in Θ besitzt, und falls diese Lösung stetig von den empirischen Momenten $\widehat{m}_k(x_1,\ldots,x_n)$ abhängt.
- Wir zeigen nun, dass Maximum-Likelihood-Schätzer unter gewissen Regularitätsbedingungen schwach konsistent sind. Hierfür führen wir zunächst den folgenden Begriff ein.

Definition

- Ähnlich wie in Abschnitt 2.3.2 setzen wir voraus, dass die Familie $\{P_\theta, \theta \in \Theta\}$ entweder nur aus diskreten Verteilungen oder nur aus absolutstetigen Verteilungen besteht, wobei $\Theta \subset \mathbb{R}$ ein offenes Intervall ist.
- Die Likelihood-Funktion $L(x;\theta)$ ist somit gegeben durch

$$L(x;\theta) = \begin{cases} p(x;\theta) & \text{im diskreten Fall,} \\ f(x;\theta) & \text{im absolutstetigen Fall,} \end{cases}$$

  wobei $p(x;\theta)$ bzw. $f(x;\theta)$ die Wahrscheinlichkeitsfunktion bzw. Dichte von P_θ ist.
- So wie bisher in Abschnitt 2 stets angenommen, gelte

$$P_\theta \neq P_{\theta'} \qquad \text{genau dann, wenn} \qquad \theta \neq \theta' . \tag{61}$$

- Für beliebige $\theta, \theta' \in \Theta$ heißt dann die Größe

$$H(P_\theta; P_{\theta'}) = \begin{cases} \displaystyle\int_{\mathbb{R}} L(x;\theta) \log \frac{L(x;\theta)}{L(x;\theta')}\, dx\,, & \text{falls } P_\theta(x : L(x;\theta') = 0) = 0, \\ \infty & \text{sonst,} \end{cases} \tag{62}$$

  die *relative Entropie* bzw. *Kullback-Leibler-Information* von P_θ bezüglich $P_{\theta'}$.

Beachte

- Im diskreten Fall ist das Integral in (62) durch eine Summe zu ersetzen.
- In diesem Abschnitt wird zugelassen, dass die Menge $B = \{x \in \mathbb{R} : L(x;\theta) > 0\}$ von θ abhängt.

Lemma 2.8 *Die in* (62) *betrachtete Größe* $H(P_\theta; P_{\theta'})$ *ist wohldefiniert, und es gilt stets* $H(P_\theta; P_{\theta'}) \geq 0$ *sowie*

$$H(P_\theta; P_{\theta'}) = 0 \qquad \textit{genau dann, wenn} \qquad P_\theta = P_{\theta'} . \tag{63}$$

Beweis

- Der Kürze wegen diskutieren wir hier nur den absolutstetigen Fall. Im diskreten Fall kann man völlig analog vorgehen, wobei lediglich Summen anstelle von Integralen betrachtet werden müssen.
- Zum Nachweis der Wohldefiniertheit von $H(P_\theta; P_{\theta'})$ muss lediglich gezeigt werden, dass das Integral in (62) wohldefiniert ist, falls $P_\theta(x : L(x;\theta') = 0) = 0$.
- Wir betrachten nun diesen Fall und setzen für jedes $x \in \mathbb{R}$

$$f(x) = \begin{cases} \dfrac{L(x;\theta)}{L(x;\theta')}\,, & \text{falls } L(x;\theta') > 0, \\ 1 & \text{sonst.} \end{cases}$$

- Dann können wir auch $L(x;\theta')f(x)$ als Likelihood-Funktion von P_θ auffassen und o.B.d.A. annehmen, dass

$$L(x;\theta) = L(x;\theta')f(x) .$$

- Beachte: Die Funktion $g : (0,\infty) \to [0,\infty)$ mit $g(x) = 1 - x + x \log x$ ist nichtnegativ, strikt konvex und nimmt ihr Minimum $g(1) = 0$ an der Stelle 1 an.

- Weil g nichtnegativ ist, ist der Erwartungswert
$$\mathbb{E}_{\theta'}(g \circ f) = \int_{\mathbb{R}} g(f(x))L(x;\theta')\,dx$$
wohldefiniert und nichtnegativ, wobei der Fall $\mathbb{E}_{\theta'}(g \circ f) = \infty$ nicht ausgeschlossen ist.
- Außerdem gilt $\mathbb{E}_{\theta'}(1-f) = 1 - \mathbb{E}_{\theta'}f = 1 - \mathbb{E}_{\theta}1 = 0$.
- Hieraus folgt durch Differenzbildung, dass auch der Erwartungswert
$$\mathbb{E}_{\theta'}(f \log f) = \int_{\mathbb{R}} L(x;\theta')f(x)\log f(x)\,dx$$
wohldefiniert ist, wobei $0 \le \mathbb{E}_{\theta'}(g \circ f) \le \infty$.
- Wegen $L(x;\theta) = L(x;\theta')f(x)$ ist damit auch gezeigt, dass das Integral $H(P_\theta;\,P_{\theta'})$ wohldefiniert ist und dass $H(P_\theta;\,P_{\theta'}) \ge 0$.
- Es ist klar, dass $H(P_\theta;\,P_{\theta'}) = 0$, falls $P_\theta = P_{\theta'}$.
- Falls nun umgekehrt $H(P_\theta;\,P_{\theta'}) = 0$, dann ergibt sich aus den Überlegungen im ersten Teil des Beweises, dass auch $\mathbb{E}_{\theta'}(g \circ f) = 0$.
- Weil g nichtnegativ ist, ergibt sich hieraus, dass $P_{\theta'}\big((g \circ f) = 0\big) = 1$.
- Weil das Minimum $g(1) = 0$ nur an der Stelle 1 angenommen wird, folgt hieraus, dass $P_{\theta'}(f = 1) = 1$, d.h., dass $P_\theta = P_{\theta'}$. $\square$

Die in (62) definierte relative Entropie $H(P_\theta;\,P_{\theta'})$ ist ein wichtiges Hilfsmittel, um zu zeigen, dass Maximum-Likelihood-Schätzer schwach konsistent sind, falls die Likelihood-Funktion unimodal in θ ist.

Theorem 2.10

- *Für beliebige $n \ge 1$ und $x_1,\ldots,x_n \in \mathbb{R}$ besitze die Maximum-Likelihood-Ungleichung* (14) *mindestens eine Lösung $\widehat{\theta}(x_1,\ldots,x_n) \in \Theta$.*
- *Die Likelihood-Funktion $L(x_1,\ldots,x_n;\theta)$ sei unimodal in θ, d.h., für jede Lösung $\widehat{\theta}(x_1,\ldots,x_n)$ von* (14) *sei die Funktion $\theta \to L(x_1,\ldots,x_n;\theta)$ wachsend für alle $\theta < \widehat{\theta}(x_1,\ldots,x_n)$ und fallend für alle $\theta > \widehat{\theta}(x_1,\ldots,x_n)$.*
- *Dann ist die Folge $\{\widehat{\theta}(X_1,\ldots,X_n),\ n \ge 1\}$ von Maximum-Likelihood-Schätzern für θ schwach konsistent.*

Beweis

- Sei $\theta \in \Theta$ beliebig, jedoch fest vorgegeben, und sei $\varepsilon > 0$ hinreichend klein, so dass $\theta \pm \varepsilon \in \Theta$.
- Wegen (61) und (63) gibt es dann ein $\delta > 0$, so dass $\delta < H(P_\theta;\,P_{\theta\pm\varepsilon})$.
- Außerdem gilt
$$\begin{aligned}\bigcap_{c=\pm1}\Big\{\frac{1}{n}\log\frac{L(x_1,\ldots,x_n;\theta)}{L(x_1,\ldots,x_n;\theta+c\varepsilon)} > \delta\Big\} &= \bigcap_{c=\pm1}\Big\{\frac{L(x_1,\ldots,x_n;\theta)}{L(x_1,\ldots,x_n;\theta+c\varepsilon)} > e^{n\delta}\Big\}\\ &\subset \big\{L(x_1,\ldots,x_n;\theta-\varepsilon) < L(x_1,\ldots,x_n;\theta) > L(x_1,\ldots,x_n;\theta+\varepsilon)\big\}\\ &\subset \{\theta-\varepsilon < \widehat{\theta}(x_1,\ldots,x_n) < \theta+\varepsilon\}\,.\end{aligned}$$
- Es genügt somit zu zeigen, dass
$$\lim_{n\to\infty} P_\theta\Big(\frac{1}{n}\log\frac{L(X_1,\ldots,X_n;\theta)}{L(X_1,\ldots,X_n;\theta+\varepsilon)} > \delta\Big) = 1 \tag{64}$$
und
$$\lim_{n\to\infty} P_\theta\Big(\frac{1}{n}\log\frac{L(X_1,\ldots,X_n;\theta)}{L(X_1,\ldots,X_n;\theta-\varepsilon)} > \delta\Big) = 1\,. \tag{65}$$

- Wir zeigen hier nur die Gültigkeit von (64), denn der Beweis von (65) verläuft analog.
- Wir betrachten zunächst den Fall, dass $H(P_\theta;\, P_{\theta+\varepsilon}) < \infty$.
- Für die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit
$$f(x) = \begin{cases} \dfrac{L(x;\theta)}{L(x;\theta+\varepsilon)}\,, & \text{falls } L(x;\theta+\varepsilon) > 0, \\ 1 & \text{sonst} \end{cases} \tag{66}$$
gilt dann $\log f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\theta)$.
- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15) ergibt sich somit, dass
$$\frac{1}{n}\log\frac{L(X_1,\ldots,X_n;\theta)}{L(X_1,\ldots,X_n;\theta+\varepsilon)} = \frac{1}{n}\sum_{i=1}^n \log f(X_i) \stackrel{P_\theta\text{-f.s.}}{\longrightarrow} \mathbb{E}_\theta(\log f) = H(P_\theta;\, P_{\theta+\varepsilon})\,.$$
- Damit ist (64) in diesem Fall bewiesen.
- Es gelte nun $H(P_\theta;\, P_{\theta+\varepsilon}) = \infty$ und $P_\theta(x:\, L(x;\theta+\varepsilon)=0) = 0$.
- Dann ist
  - die in (66) eingeführte Funktion für fast jedes $x \in \mathbb{R}$ gegeben durch $f(x) = L(x;\theta)/L(x;\theta+\varepsilon)$,
  - es gilt $\log\min\{f,c\} \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\theta)$ für jedes $c > 1$,
  - und aus dem Satz über die monotone Konvergenz ergibt sich, dass für $c \to \infty$
$$\mathbb{E}_\theta\bigl(\log\min\{f,c\}\bigr) \to H(P_\theta;\, P_{\theta+\varepsilon}) = \infty\,.$$
- Es gibt somit ein $c > 1$ mit $\mathbb{E}_\theta\bigl(\log\min\{f,c\}\bigr) > \delta$, und so wie im ersten Fall ergibt sich aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15), dass für $n \to \infty$
$$P_\theta\Bigl(\frac{1}{n}\log\frac{L(X_1,\ldots,X_n;\theta)}{L(X_1,\ldots,X_n;\theta+\varepsilon)} > \delta\Bigr) \ge P_\theta\Bigl(\frac{1}{n}\sum_{i=1}^n \log\min\{f(X_i),c\} > \delta\Bigr) \to 1\,.$$
- Damit ist (64) auch in diesem Fall bewiesen.
- Falls schließlich $P_\theta(x:\, L(x;\theta+\varepsilon)=0) = a > 0$, dann gilt wegen $P_\theta(x:\, L(x;\theta) > 0) = 1$, dass für $n \to \infty$
$$P_\theta\Bigl(\frac{1}{n}\log\frac{L(X_1,\ldots,X_n;\theta)}{L(X_1,\ldots,X_n;\theta+\varepsilon)} = \infty\Bigr) = 1 - (1-a)^n \to 1\,.$$
- Hieraus ergibt sich erneut die Gültigkeit von (64). □

Beachte

- Die in Theorem 2.10 vorausgesetzte Unimodalität der Abbildung $\theta \to L(x_1,\ldots,x_n;\theta)$ ist in zahlreichen Fällen erfüllt.
- Hinreichend hierfür ist, dass die Abbildung $\theta \to \log L(x;\theta)$ für jedes $x \in \mathbb{R}$ konkav ist.

Beispiele

1. *Poisson-verteilte Stichprobenvariablen*
   - Sei $\{P_\theta,\, \theta \in \Theta\} = \{\,\mathrm{Poi}(\lambda),\, \lambda > 0\}$.
   - Dann gilt für jedes $x \in \mathbb{R}$
$$L(x;\lambda) = \begin{cases} \frac{\lambda^x}{x!}\, e^{-\lambda}\,, & \text{falls } x \in \mathbb{N}, \\ 0 & \text{sonst} \end{cases}$$
   und somit
$$\log L(x;\lambda) = \begin{cases} x\log\lambda - \lambda - \log(x!)\,, & \text{falls } x \in \mathbb{N}, \\ -\infty & \text{sonst,} \end{cases}$$
   d.h., $\log L(x;\lambda)$ ist konkav in λ.

- Aus Theorem 2.10 ergibt sich nun, dass die Folge $\widehat{\lambda} = \overline{X}_n$ von Maximum-Likelihood-Schätzern für λ schwach konsistent ist.
- Beachte: In Theorem 1.2 hatten wir allerdings bereits mit Hilfe des starken Gesetzes der großen Zahlen gezeigt, dass $\widehat{\lambda} = \overline{X}_n$ nicht nur schwach, sondern sogar stark konsistent ist.

2. *Exponentialverteilte Stichprobenvariablen*
   - Sei $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Exp}(\lambda),\, \lambda > 0\}$.
   - Dann gilt für jedes $x \in \mathbb{R}$
   $$L(x;\lambda) = \begin{cases} \lambda\, e^{-\lambda x}\,, & \text{falls } x > 0, \\ 0\,, & \text{sonst} \end{cases}$$
   und somit
   $$\log L(x;\lambda) = \begin{cases} -x\lambda + \log\lambda\,, & \text{falls } x > 0, \\ -\infty\,, & \text{sonst,} \end{cases}$$
   d.h., $\log L(x;\lambda)$ ist konkav in λ.
   - Aus Theorem 2.10 ergibt sich nun, dass die Folge $\widehat{\lambda} = 1/\overline{X}_n$ von Maximum-Likelihood-Schätzern für λ schwach konsistent ist.
   - Beachte: Aus dem starken Gesetz der großen Zahlen (vgl. Theorem 1.2) ergibt sich darüber hinaus, dass $\widehat{\lambda} = 1/\overline{X}_n$ nicht nur schwach, sondern sogar stark konsistent ist.
3. *Gleichverteilte Stichprobenvariablen*
   - Sei $\{P_\theta,\, \theta \in \Theta\} = \{\,\mathrm{U}(\theta),\, \theta > 0\}$.
   - Dann gilt für jedes $x \in \mathbb{R}$
   $$L(x;\theta) = \begin{cases} \theta^{-1}\,, & \text{falls } 0 < x \le \theta, \\ 0\,, & \text{sonst} \end{cases}$$
   und somit ist die Likelihood-Funktion
   $$L(x_1,\ldots,x_n;\theta) = \begin{cases} \theta^{-n}\,, & \text{falls } 0 < x_1,\ldots,x_n \le \theta, \\ 0\,, & \text{sonst} \end{cases}$$
   unimodal in θ.
   - Wegen Theorem 2.10 ist somit durch $\widehat{\theta}(X_1,\ldots,X_n) = \max\{X_1,\ldots,X_n\}$ eine (schwach) konsistente Folge von Maximum-Likelihood-Schätzern gegeben.

2.4.2 Asymptotische Normalverteiltheit

- In diesem Abschnitt diskutieren wir die asymptotische Normalverteiltheit von Parameterschätzern für den Fall, dass der Stichprobenumfang n unendlich groß wird.
- Wir beginnen mit dem folgenden Beispiel, das Aussagen enthält, die wir bereits in den Abschnitten 1.2 bzw. 1.3 (teilweise unter allgemeineren Modellannahmen) hergeleitet hatten.

Beispiel (*Normalverteilte Stichprobenvariablen*)

- Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{N}(\mu,\sigma^2),\, \mu \in \mathbb{R},\, \sigma^2 > 0\}$, wobei sowohl μ als auch σ^2 unbekannt sei.
- Weil für das 4-te zentrale Moment μ_4' normalverteilter Stichprobenvariablen $X_1,\ldots,X_n$
$$\mu_4' - \sigma^4 = 2\sigma^4$$

gilt, ergibt sich aus den Theoremen 1.4 bzw. 1.11, dass für beliebige $\mu \in \mathbb{R}$, $\sigma^2 > 0$

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \stackrel{\mathrm{d}}{=} Y \qquad \text{und} \qquad \sqrt{n}\,\frac{S_n^2 - \sigma^2}{\sqrt{2\sigma^2}} \stackrel{\mathrm{d}}{\longrightarrow} Y \quad \text{für } n \to \infty, \tag{67}$$

wobei $Y \sim \mathrm{N}(0,1)$.

- Weil die Zufallsvariablen $\overline{X}_n$ und S_n^2 unabhängig sind (vgl. Theorem 1.10), ergibt sich hieraus, dass für beliebige $y_1, y_2 \in \mathbb{R}$

$$\lim_{n\to\infty} P_{(\mu,\sigma^2)}\Bigl(\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \le y_1,\ \sqrt{n}\,\frac{S_n^2 - \sigma^2}{\sqrt{2\sigma^2}} \le y_2\Bigr) = \Phi(y_1)\,\Phi(y_2)\,, \tag{68}$$

  wobei $\Phi : \mathbb{R} \to [0,1]$ die Verteilungsfunktion der $\mathrm{N}(0,1)$-Verteilung ist.

- Hieraus folgt außerdem, dass auch

$$\lim_{n\to\infty} P_{(\mu,\sigma^2)}\Bigl(\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\sigma} \le y_1,\ \sqrt{n}\,\frac{\widehat{\sigma}_n^2 - \sigma^2}{\sqrt{2\sigma^2}} \le y_2\Bigr) = \Phi(y_1)\,\Phi(y_2)\,, \tag{69}$$

  wobei $(\widehat{\mu}_n, \widehat{\sigma}_n^2)$ der Maximum-Likelihood-Schätzer für (μ, σ^2) ist mit

$$\widehat{\mu}_n = \overline{X}_n\,, \qquad \widehat{\sigma}_n^2(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\,.$$

Wir beweisen nun einen *zentralen Grenzwertsatz* für Maximum-Likelihood-Schätzer, wobei wir voraussetzen, dass die folgenden Bedingungen erfüllt sein mögen.

- Die Familie $\{P_\theta, \theta \in \Theta\}$ bestehe entweder nur aus diskreten Verteilungen oder nur aus absolutstetigen Verteilungen, wobei $\Theta \subset \mathbb{R}$ ein offenes Intervall sei.

- Die Menge $B = \{x \in \mathbb{R} : L(x;\theta) > 0\}$ hänge nicht von $\theta \in \Theta$ ab, wobei die Likelihood-Funktion $L(x;\theta)$ gegeben ist durch

$$L(x;\theta) = \begin{cases} p(x;\theta) & \text{im diskreten Fall,} \\ f(x;\theta) & \text{im absolutstetigen Fall} \end{cases}$$

  und $p(x;\theta)$ bzw. $f(x;\theta)$ die Wahrscheinlichkeitsfunktion bzw. Dichte von P_θ ist.

- Es gelte

$$P_\theta \ne P_{\theta'} \qquad \text{genau dann, wenn} \qquad \theta \ne \theta'\,.$$

- Außerdem sei die Abbildung $\theta \to L(x;\theta)$ für jedes $x \in B$ dreimal stetig differenzierbar, und es gelte

$$\frac{d^k}{d\theta^k}\int\limits_B L(x;\theta)\,dx = \int\limits_B \frac{\partial^k}{\partial\theta^k}\,L(x;\theta)\,dx\,, \qquad \forall\, k = 1,2 \text{ und } \theta \in \Theta, \tag{70}$$

  wobei die Integrale im diskreten Fall durch die entsprechenden Summen zu ersetzen sind.

- Für jedes $\theta_0 \in \Theta$ gebe es eine Konstante $c_{\theta_0} > 0$ und eine messbare Funktion $g_{\theta_0} : B \to [0,\infty)$, so dass

$$\Bigl|\frac{\partial^3}{\partial\theta^3}\,\log L(x;\theta)\Bigr| \le g_{\theta_0}(x)\,, \qquad \forall\, x \in B \text{ und } |\theta - \theta_0| < c_{\theta_0} \tag{71}$$

  und

$$\mathbb{E}_{\theta_0} g_{\theta_0}(X_1) < \infty\,. \tag{72}$$

Beachte

- Das Integral
$$I(\theta) = \mathbb{E}_\theta\Bigl(\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)^2\Bigr)\,, \qquad \forall\,\theta\in\Theta\,, \tag{73}$$
das wir bereits in Abschnitt 2.3.2 betrachtet hatten (vgl. (31) bzw. (32)), wird in der Literatur *Fisher-Information* genannt.
- Bei der Herleitung der Ungleichung von Cramér-Rao (vgl. den Beweis von Theorem 2.2) hatten wir gezeigt, dass
$$\mathbb{E}_\theta\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr) = 0 \tag{74}$$
und somit
$$I(\theta) = \operatorname{Var}_\theta\Bigl(\frac{\partial}{\partial\theta}\log L(X_1;\theta)\Bigr)\,. \tag{75}$$

Theorem 2.11 *Falls*
$$0 < I(\theta) < \infty\,, \qquad \forall\,\theta\in\Theta\,,$$
dann gilt für jede schwach konsistente Folge $\{\widehat\theta(X_1,\ldots,X_n),\ n\ge 1\}$ *von Maximum-Likelihood-Schätzern für* θ, *dass*
$$\bigl(n\,I(\theta)\bigr)^{1/2}\,\bigl(\widehat\theta(X_1,\ldots,X_n)-\theta\bigr) \xrightarrow{\mathrm{d}} Y \sim \mathrm{N}(0,1), \qquad \forall\,\theta\in\Theta\,. \tag{76}$$

Beweis

- Wir benutzen die abkürzende Schreibweise $l_n(\theta) = \log L(X_1,\ldots,X_n;\theta)$ und $l_n^{(1)}(\theta)$, $l_n^{(2)}(\theta)$, $l_n^{(3)}(\theta)$ für die entsprechenden Ableitungen bezüglich θ.
- Sei $\widehat\theta = \widehat\theta(X_1,\ldots,X_n)$, $n\ge 1$, eine schwach konsistente Folge von Maximum-Likelihood-Schätzern für θ.
- Für jedes $\omega\in\Omega$ entwickeln wir nun $l_n^{(1)}(\widehat\theta)$ in eine Taylor-Reihe an der Stelle $\theta\in\Theta$ und erhalten, dass
$$l_n^{(1)}(\widehat\theta) = l_n^{(1)}(\theta) + (\widehat\theta-\theta)l_n^{(2)}(\theta) + \frac{1}{2}(\widehat\theta-\theta)^2 l_n^{(3)}(\widehat\theta^*)\,,$$
wobei $\widehat\theta^*$ zwischen θ und $\widehat\theta$ liegt.
- Weil $l_n^{(1)}(\widehat\theta) = 0$ gilt, ergibt sich hieraus, dass
$$\sqrt{n}\,(\widehat\theta-\theta) = \frac{l_n^{(1)}(\theta)/\sqrt{n}}{-\dfrac{1}{n}\,l_n^{(2)}(\theta) - \dfrac{1}{2n}\,(\widehat\theta-\theta)l_n^{(3)}(\widehat\theta^*)}\,. \tag{77}$$
- Es genügt zu zeigen, dass
$$\frac{1}{\sqrt{n}}\,l_n^{(1)}(\theta) \xrightarrow{\mathrm{d}} Y \sim\ \mathrm{N}(0, I(\theta)), \tag{78}$$
dass
$$-\frac{1}{n}\,l_n^{(2)}(\theta) \xrightarrow{\mathrm{P}} I(\theta) \tag{79}$$
und dass es eine Konstante $c<\infty$ gibt, so dass
$$\lim_{n\to\infty} P_\theta\Bigl(\Bigl|\frac{1}{n}\,l_n^{(3)}(\widehat\theta^*)\Bigr| < c\Bigr) = 1\,. \tag{80}$$
- Denn wegen $\widehat\theta \xrightarrow{\mathrm{P}} \theta$ ergibt sich aus (79) und (80), dass der Nenner der rechten Seite von (77) in Wahrscheinlichkeit gegen $I(\theta)$ konvergiert.

- Hieraus und aus (78) ergibt sich dann die Gültigkeit von (76) mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11).
- Um (78) zu zeigen, genügt es zu beachten, dass
$$l_n^{(1)}(\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta}\,\log L(X_i;\theta)\,.$$
- Wegen (74) und (75) ergibt sich (78) dann aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen; vgl. Theorem WR–5.16.
- Um (79) zu zeigen, genügt es zu beachten, dass
$$\begin{aligned}-\frac{1}{n}\,l_n^{(2)}(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\bigl(L^{(1)}(X_i;\theta)\bigr)^2 - L(X_i;\theta)L^{(2)}(X_i;\theta)}{\bigl(L(X_i;\theta)\bigr)^2}\\ &= \frac{1}{n}\sum_{i=1}^n \Bigl(\frac{L^{(1)}(X_i;\theta)}{L(X_i;\theta)}\Bigr)^2 - \frac{1}{n}\sum_{i=1}^n \frac{L^{(2)}(X_i;\theta)}{L(X_i;\theta)}\,,\end{aligned}$$
wobei
$$L^{(k)}(X_i;\theta) = \frac{\partial^k}{\partial\theta^k}\,L(X_i;\theta)\,.$$
- Aus dem starken Gesetz der großen Zahlen für Summen von unabhängigen und identisch verteilten Zufallsvariablen (vgl. Theorem WR–5.15) ergibt sich nun, dass
$$\begin{aligned}-\frac{1}{n}\,l_n^{(2)}(\theta) &\stackrel{\text{f.s.}}{\longrightarrow} \mathbb{E}_\theta\Bigl(\Bigl(\frac{L^{(1)}(X_1;\theta)}{L(X_1;\theta)}\Bigr)^2\Bigr) - \mathbb{E}_\theta\Bigl(\frac{L^{(2)}(X_1;\theta)}{L(X_1;\theta)}\Bigr)\\ &= I(\theta) - \mathbb{E}_\theta\Bigl(\frac{L^{(2)}(X_1;\theta)}{L(X_1;\theta)}\Bigr)\\ &= I(\theta)\,,\end{aligned}$$
wobei sich die vorletzte Gleichheit aus der Definitionsgleichung (73) der Fisher-Information $I(\theta)$ ergibt und
$$\begin{aligned}\mathbb{E}_\theta\Bigl(\frac{L^{(2)}(X_1;\theta)}{L(X_1;\theta)}\Bigr) &= \int_B \frac{\partial^2}{\partial\theta^2}\,L(x;\theta)\,dx\\ &\stackrel{(70)}{=} \frac{d^2}{d\theta^2}\int_B L(x;\theta)\,dx = \frac{d^2}{d\theta^2}\,1 = 0\,.\end{aligned}$$
- Schließlich ergibt sich aus (71), dass
$$\Bigl|\frac{1}{n}\,l_n^{(3)}(\widehat{\theta}^*)\Bigr| \le \frac{1}{n}\sum_{i=1}^n g_\theta(X_i)\,, \tag{81}$$
falls $|\widehat{\theta}-\theta| < c_\theta$.
- Weil $|\widehat{\theta}-\theta| \stackrel{\text{P}}{\longrightarrow} 0$ gilt, konvergiert die Wahrscheinlichkeit des in (81) betrachteten Ereignisses gegen 1 für $n\to\infty$.
- Weil andererseits wegen (72)
$$\frac{1}{n}\sum_{i=1}^n g_\theta(X_i) \stackrel{\text{f.s.}}{\longrightarrow} \mathbb{E}_\theta g_\theta(X_1) < \infty\,,$$
ergibt sich hieraus die Gültigkeit von (80). □

Beispiel (*Poisson-verteilte Stichprobenvariablen*)

- Es gelte $\{P_\theta,\, \theta \in \Theta\} = \{\mathrm{Poi}(\lambda),\, \lambda > 0\}$.
- Dann ist die Likelihood-Funktion $L(x;\lambda) = \bigl(\lambda^x/x!\bigr)\, e^{-\lambda}$ für jedes $x \in \mathbb{N}$ beliebig oft differenzierbar in λ, und die Vertauschbarkeitsbedingung (70) ist erfüllt, denn es gilt für jedes $x \in \mathbb{N}$

$$\frac{d}{d\lambda}\Bigl(\sum_{x=0}^{\infty} L(x;\lambda)\Bigr) = \frac{d}{d\lambda}\Bigl(\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}\, e^{-\lambda}\Bigr) = \frac{d}{d\lambda}\, 1 = 0$$

und

$$\begin{aligned}\sum_{x=0}^{\infty} \frac{\partial}{\partial\lambda}\, L(x;\lambda) &= \sum_{x=0}^{\infty} \frac{\partial}{\partial\lambda}\, \Bigl(\frac{\lambda^x}{x!}\, e^{-\lambda}\Bigr)\\ &= \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}\, e^{-\lambda} - \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}\, e^{-\lambda} = 0\,.\end{aligned}$$

- Außerdem gilt für beliebige $x \in \mathbb{N}$ und $\lambda > 0$

$$\frac{\partial^3}{\partial\lambda^3}\, \log L(x;\lambda) = \frac{\partial^3}{\partial\lambda^3}\, (x\log\lambda - \lambda) = \frac{2x}{\lambda^3}\,.$$

- Hieraus folgt, dass für jedes $\lambda_0 > 0$

$$\Bigl|\frac{\partial^3}{\partial\lambda^3}\, \log L(x;\lambda)\Bigr| \le g_{\lambda_0}(x)\,, \qquad \forall\, x \in \mathbb{N} \text{ und } |\lambda - \lambda_0| < c_{\lambda_0}\,,$$

wobei $c_{\lambda_0} = \lambda_0/2$ und

$$g_{\lambda_0}(x) = \frac{2x}{(\lambda_0/2)^3} \qquad \text{mit} \qquad \mathbb{E}_{\lambda_0} g_{\lambda_0}(X_1) = \frac{16}{\lambda_0^2} < \infty\,.$$

- Die Bedingungen der Theoreme 2.10 und 2.11 sind somit erfüllt.
- Hieraus folgt, dass der Maximum-Likelihood-Schätzer $\widehat{\lambda}(X_1,\ldots,X_n) = \overline{X}_n$ für λ asymptotisch normalverteilt ist.
- Dies ergibt sich jedoch in diesem Fall auch direkt aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen; vgl. Theorem 1.2.

3 Konfidenzintervalle

3.1 Modellbeschreibung

In diesem Abschnitt setzen wir erneut voraus, dass

- die Verteilung der Stichprobenvariablen $X_1, \ldots, X_n$ zu einer vorgegebenen *parametrischen Familie* von Verteilungen $\{P_\theta,\, \theta \in \Theta\}$ gehört; $\Theta \subset \mathbb{R}^m$.

Dabei nehmen wir (zur Vereinfachung der Darlegungen) an, dass

- jeweils nur eine einzelne Komponente θ_j des Parametervektors $\theta = (\theta_1, \ldots, \theta_m)$ aus den beobachteten Daten $x_1, \ldots, x_n$ geschätzt werden soll; $j \in \{1, \ldots, m\}$.

Genauso wie in den Abschnitten 1 und 2 nehmen wir an, dass

- die Stichprobenvariablen $X_1, X_2, \ldots$ unabhängig und identisch verteilt sind und dass
- der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, über dem die Stichprobenvariablen $X_1, X_2, \ldots$ definiert sind, der kanonische Wahrscheinlichkeitsraum dieser Zufallsvariablen ist.

3.1.1 Konfidenzintervall und Konfidenzniveau

Anstelle jeweils eine (einzelne) Stichprobenfunktion zu betrachten, wie wir es in den Abschnitten 1 und 2 getan haben, betrachten wir nun

- *zwei* Stichprobenfunktionen $\underline{\theta} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ und $\overline{\theta} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, so dass
$$\underline{\theta}(x_1, \ldots, x_n) \le \overline{\theta}(x_1, \ldots, x_n) \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n\,. \tag{1}$$

Definition

- Sei $\gamma \in (0,1)$ eine beliebige, jedoch fest vorgegebene Zahl.
- Dann heißt das zufällige Intervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ *Konfidenzintervall* für θ_j zum *Niveau* γ, falls
$$P_\theta(\underline{\theta}(X_1, \ldots, X_n) \le \theta_j \le \overline{\theta}(X_1, \ldots, X_n)) \ge \gamma \tag{2}$$
für jedes $\theta \in \Theta$.

Beachte

- Die Stichprobenfunktionen $\underline{\theta}$ und $\overline{\theta}$ sind nicht eindeutig bestimmt.
- Sie sollten so gewählt werden, dass das Konfidenzintervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ bei gegebenem Niveau γ möglichst kurz ist.
- Weil das Konfidenzintervall möglichst kurz sein soll, wird anstelle von (2) manchmal auch die folgende Bedingung betrachtet:
$$\inf_{\theta \in \Theta} P_\theta(\underline{\theta}(X_1, \ldots, X_n) \le \theta_j \le \overline{\theta}(X_1, \ldots, X_n)) = \gamma\,. \tag{3}$$
- Falls (3) gilt, dann nennt man $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ *minimales Konfidenzintervall* für θ_j zum Niveau γ).

- Wenn (bei vorgegebenem θ) die Verteilung der Zufallsvariablen $\underline{\theta}(X_1,\ldots,X_n)$ und $\overline{\theta}(X_1,\ldots,X_n)$ schwierig handhabbar bzw. unbekannt ist, d.h., wenn es nicht möglich ist, die Gültigkeit von (2) bzw. (3) nachzuweisen, dann kann beispielsweise die folgende Bedingung betrachtet werden:

$$\lim_{n\to\infty} P_\theta\bigl(\underline{\theta}(X_1,\ldots,X_n) \le \theta_j \le \overline{\theta}(X_1,\ldots,X_n)\bigr) \ge \gamma \tag{4}$$

  für jedes $\theta \in \Theta$.
- Falls (4) gilt, dann nennt man $\bigl(\underline{\theta}(X_1,\ldots,X_n), \overline{\theta}(X_1,\ldots,X_n)\bigr)$ Konfidenzintervall für θ_j zum asymptotischen Niveau γ (bzw. *asymptotisches Konfidenzintervall* zum Niveau γ).

Die *praktische Berechnung* eines (konkreten) Konfidenzintervalls $\bigl(\underline{\theta}(x_1,\ldots,x_n), \overline{\theta}(x_1,\ldots,x_n)\bigr)$ für θ_j auf der Basis einer (konkreten) Stichprobe $(x_1,\ldots,x_n)$ besteht aus den folgenden Schritten:

1. Bestimme zwei Stichprobenfunktionen $\underline{\theta} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ und $\overline{\theta} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, so dass
   - (1) gilt und
   - eine der Bedingungen (2)–(4) erfüllt ist.
2. Berechne die Funktionswerte $\underline{\theta}(x_1,\ldots,x_n)$ und $\overline{\theta}(x_1,\ldots,x_n)$.

3.1.2 Quantilfunktion

Bei der Konstruktion von Konfidenzintervallen benötigen wir den Begriff der Quantilfunktion, den wir bereits in Abschnitt WR–4.1.4 betrachtet hatten.

Definition

- Sei $F : \mathbb{R} \to [0,1]$ eine beliebige Verteilungsfunktion.
- Die Funktion $F^{-1} : [0,1] \to \mathbb{R}$ mit

$$F^{-1}(y) = \inf\{x : F(x) \ge y\} \tag{5}$$

  heißt *Quantilfunktion* der Verteilungsfunktion F (bzw. der zugehörigen Verteilung).
- Sei $\alpha \in (0,1)$. Die Zahl $F^{-1}(\alpha)$ wird dann α-*Quantil* von F genannt.

Beachte

- Wenn $F : \mathbb{R} \to \mathbb{R}$ eine beliebige monoton wachsende rechtsstetige Funktion ist (die nicht unbedingt eine Verteilungsfunktion sein muss), dann heißt die in (5) definierte Funktion F^{-1} *verallgemeinerte inverse Funktion* von F.
- Quantilfunktionen sind also spezielle verallgemeinerte inverse Funktionen.

Beispiele

Um Konfidenzintervalle für die Parameter von normalverteilten Stichprobenvariablen konstruieren zu können, werden insbesondere die Quantilfunktionen der Standardnormalverteilung, der χ^2-Verteilung und der t-Verteilung benötigt, die bereits in den Abschnitten 1.2.2, 1.3.1 bzw 1.3.4 eingeführt worden sind und an die wir hier zunächst erinnern wollen.

1. *Quantile der* $\mathrm{N}(0,1)$-*Verteilung*
   - Das α-Quantil der Verteilungsfunktion Φ (d.h. der Standardnormalverteilung $\mathrm{N}(0,1)$) wird mit z_α bezeichnet.
   - Mit anderen Worten: Für jedes $\alpha \in (0,1)$ ist z_α die Lösung der *Quantilgleichung* $\Phi(z_\alpha) = \alpha$.

- Für $\alpha \geq 0,50$ kann man das α-Quantil z_α aus Tabelle 1 entnehmen; vgl. Abschnitt 6.
- Aus der Symmetrieeigenschaft $\Phi(-x) = 1 - \Phi(x)$ für jedes $x \in \mathbb{R}$ ergibt sich außerdem, dass für jedes $\alpha \in (0,1)$
$$z_\alpha = -z_{1-\alpha}\,. \tag{6}$$

2. *Quantile der χ^2-Verteilung mit r Freiheitsgraden*
   - Die Zufallsvariable U_r sei χ^2-verteilt mit r-Freiheitsgraden, d.h., $U_r \sim \chi^2_r$.
   - Für $\alpha \in (0,1)$ sei $\chi^2_{r,\alpha}$ die (eindeutig bestimmte) Lösung der Gleichung $F_{U_r}(\chi^2_{r,\alpha}) = \alpha$, wobei die Dichte f_{U_r} von U_r in Theorem 1.6 gegeben ist.
   - Dann heißt $\chi^2_{r,\alpha}$ das α-*Quantil* der χ^2-Verteilung mit r Freiheitsgraden.
   - Quantile der χ^2-Verteilung mit r Freiheitsgraden sind in Tabelle 2 gegeben, vgl. Abschnitt 6.
3. *Quantile der* t-*Verteilung mit r Freiheitsgraden*
   - Die Zufallsvariable V_r sei t-verteilt mit r-Freiheitsgraden, d.h., $V_r \sim \mathrm{t}_r$, wobei die Dichte f_{V_r} von V_r in Theorem 1.12 gegeben ist.
   - Für $\alpha \in (0,1)$ wird dann die Lösung $t_{r,\alpha}$ der Gleichung $F_{V_r}(t_{r,\alpha}) = \alpha$ das α-*Quantil* der t-Verteilung mit r Freiheitsgraden genannt, vgl. Tabelle 3 in Abschnitt 6.
   - Analog zu der Symmetrieeigenschaft $z_\alpha = -z_{1-\alpha}$ der Quantile z_α der $\mathrm{N}(0,1)$-Verteilung gilt auch
   $$t_{r,\alpha} = -t_{r,1-\alpha} \tag{7}$$
   für beliebige $r \in \mathbb{N}$ und $\alpha \in (0,1)$.

3.1.3 F-Verteilung

Wir führen nun noch eine weitere Klasse von statistischen Prüfverteilungen ein: die Klasse der F-*Verteilungen*, die zum Beispiel bei der Konstruktion von Konfidenzintervallen für Zwei-Stichproben-Probleme benötigt werden; vgl. Abschnitt 3.3.

Definition

- Seien $r, s \geq 1$ beliebige natürliche Zahlen, und seien $U_r, U'_s : \Omega \to (0,\infty)$ unabgängige χ^2-verteilte Zufallsvariablen mit $U_r \sim \chi^2_r$ und $U'_s \sim \chi^2_s$.
- Man sagt dann, dass die Zufallsvariable
$$W_{r,s} = \frac{U_r/r}{U'_s/s}$$
F-*verteilt* ist mit (r,s) Freiheitsgraden. (Schreibweise: $W_{r,s} \sim \mathrm{F}_{r,s}$)

Theorem 3.1 *Seien $r, s \geq 1$ beliebige natürliche Zahlen, und sei $W_{r,s}$ eine* F-*verteilte Zufallsvariable mit (r,s) Freiheitsgraden. Dann ist die Dichte von $W_{r,s}$ gegeben durch*

$$f_{W_{r,s}}(x) = \begin{cases} \dfrac{\Gamma\Big(\dfrac{r+s}{2}\Big)}{\Gamma\Big(\dfrac{r}{2}\Big)\Gamma\Big(\dfrac{s}{2}\Big)} \Big(\dfrac{r}{s}\Big)^{r/2} \dfrac{x^{(r/2)-1}}{\Big(1+\dfrac{r}{s}\,x\Big)^{(r+s)/2}}\,, & \textit{falls } x > 0, \\ 0\,, & \textit{sonst,} \end{cases} \tag{8}$$

wobei $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$ *und* $\Gamma(p+1) = p\,\Gamma(p)$.

Beweis

- Aus Theorem 1.6 ergibt sich für die Dichten der χ^2-verteilten Zufallsvariablen mit U_r und U'_s, dass
$$f_{U_r}(x) = \begin{cases} \dfrac{x^{(r-2)/2}e^{-x/2}}{2^{r/2}\,\Gamma(r/2)}\,, & \text{falls } x > 0,\\ 0\,, & \text{sonst}\end{cases}$$
und
$$f_{U'_s}(x) = \begin{cases} \dfrac{x^{(s-2)/2}e^{-x/2}}{2^{s/2}\,\Gamma(s/2)}\,, & \text{falls } x > 0,\\ 0\,, & \text{sonst.}\end{cases}$$
- Hieraus folgt, dass
$$f_{U_r/r}(x) = \begin{cases} \dfrac{r(rx)^{(r-2)/2}e^{-rx/2}}{2^{r/2}\,\Gamma(r/2)}\,, & \text{falls } x > 0,\\ 0\,, & \text{sonst}\end{cases} \tag{9}$$
und
$$f_{U'_s/s}(x) = \begin{cases} \dfrac{s(sx)^{(s-2)/2}e^{-sx/2}}{2^{s/2}\,\Gamma(s/2)}\,, & \text{falls } x > 0,\\ 0\,, & \text{sonst.}\end{cases} \tag{10}$$
- Wegen Theorem WR–3.17 gilt außerdem für die Dichte $f_{W_{r,s}}$ des Quoienten $W_{r,s}$ der unabhängigen Zufallsvariablen U_r/r und U'_s/s, dass
$$f_{W_{r,s}}(z) = \int_{-\infty}^{\infty} |t| f_{U_r/r}(zt) f_{U'_s/s}(t)\, dt\,, \qquad \forall\, z \in \mathbb{R}\,.$$
- Durch Einsetzen von (9) und (10) ergibt sich somit, dass für $z > 0$
$$\begin{aligned} f_{W_{r,s}}(z) &= \int_0^\infty t\, \frac{r(zrt)^{(r-2)/2}e^{-zrt/2}}{2^{r/2}\,\Gamma(r/2)}\, \frac{s(st)^{(s-2)/2}e^{-st/2}}{2^{s/2}\,\Gamma(s/2)}\, dt\\ &= \frac{r^{r/2}s^{s/2}z^{r/2-1}}{2^{(r+s)/2}\Gamma(r/2)\Gamma(s/2)} \int_0^\infty t^{(r+s)/2-1}\, e^{-(rz+s)t/2}\, dt\\ &= \frac{r^{r/2}s^{s/2}z^{r/2-1}}{2^{(r+s)/2}\Gamma(r/2)\Gamma(s/2)}\, \frac{\Gamma((r+s)/2)}{\big((rz/2)+(s/2)\big)^{(r+s)/2}}\\ &= \frac{\Gamma\Big(\dfrac{r+s}{2}\Big)}{\Gamma\Big(\dfrac{r}{2}\Big)\Gamma\Big(\dfrac{s}{2}\Big)} \Big(\frac{r}{s}\Big)^{r/2} \frac{z^{(r/2)-1}}{\Big(1+\dfrac{r}{s}\,z\Big)^{(r+s)/2}}\,, \end{aligned}$$
wobei sich die vorletzte Gleichheit aus der in Abschnitt 1.3.1 hergeleiteten Darstellungsformel (1.24) für die Gammafunktion ergibt, gemäß der für beliebige $b, p > 0$
$$\Gamma(p) = b^p \int_0^\infty e^{-by}\, y^{p-1}\, dy\,.$$

□

Beachte

- Seien $r, s \geq 1$ beliebige natürliche Zahlen, und sei $\alpha \in (0,1)$.
- Falls $W_{r,s} \sim \mathrm{F}_{r,s}$, dann wird die Lösung $F_{r,s,\alpha}$ der Gleichung $F_{W_{r,s}}(F_{r,s,\alpha}) = \alpha$ das α-*Quantil* der F-Verteilung mit (r,s) Freiheitsgraden genannt, vgl. die Tabellen 4a-4f in Abschnitt 6.

3.2 Konfidenzintervalle bei Normalverteilung

In diesem Abschnitt nehmen wir an, dass die Stichprobenvariablen $X_1, \ldots, X_n$ normalverteilt sind, d.h., es gelte $\{P_\theta,\ \theta \in \Theta\} = \{\mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}$ mit $\theta = (\mu, \sigma^2)$.

3.2.1 Konfidenzintervalle für den Erwartungswert

Um Konfidenzintervalle für den Erwartungswert der normalverteilten Stichprobenvariablen $X_1, \ldots, X_n$ herzuleiten, nehmen wir zunächst an, dass die Varianz bekannt ist, d.h., es gelte $X_i \sim \mathrm{N}(\mu, \sigma^2)$ für ein (unbekanntes) $\mu \in \mathbb{R}$ und ein (bekanntes) $\sigma^2 > 0$.

- In Theorem 1.11 hatten wir gezeigt, dass $\sqrt{n}(\overline{X}_n - \mu)/\sigma \sim \mathrm{N}(0,1)$.
- Hieraus ergibt sich, dass für $\alpha \in (0,1)$
$$P_\theta\Big(z_{\alpha/2} \leq \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq z_{1-\alpha/2}\Big) = 1 - \alpha\,. \tag{11}$$
- Unter Berücksichtigung von (6) ergibt sich somit, dass für jedes $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$
$$P_\theta\Big(-z_{1-\alpha/2} \leq \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq z_{1-\alpha/2}\Big) = 1 - \alpha$$
bzw.
$$P_\theta\Big(\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha\,.$$
- Also ist mit
$$\underline{\theta}(X_1, \ldots, X_n) = \overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1, \ldots, X_n) = \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{12}$$
ein Konfidenzintervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ für μ zum Niveau $\gamma = 1 - \alpha$ gegeben.
- Für die Länge $\ell(X_1, \ldots, X_n)$ dieses Konfidenzintervalls gilt
$$\begin{aligned} \ell(X_1, \ldots, X_n) &= \overline{\theta}(X_1, \ldots, X_n) - \underline{\theta}(X_1, \ldots, X_n) \\ &= z_{1-\alpha/2}\frac{2\sigma}{\sqrt{n}}\,. \end{aligned}$$
- Hieraus ergibt sich insbesondere, dass $\ell(X_1, \ldots, X_n)$ nicht vom Zufall abhängt.
- Außerdem erkennen wir, dass die Länge $\ell(X_1, \ldots, X_n)$ des in (12) gegebenen Konfidenzintervalls klein ist, falls das Niveau $\gamma = 1 - \alpha$ klein, die Varianz σ^2 klein bzw. der Stichprobenumfang n groß ist.
- Man kann sich leicht überlegen, dass $\ell(X_1, \ldots, X_n) \leq \varepsilon$ gilt, falls
$$n \geq \Big(\frac{2\sigma z_{1-\alpha/2}}{\varepsilon}\Big)^2, \tag{13}$$
wobei $\varepsilon > 0$ ein vorgegebener Schwellenwert ist.

Beispiel

- Für eine (konkrete) Stichprobe $(x_1,\ldots,x_n)$, die wir als Realisierung der Zufallsstichprobe $(X_1,\ldots,X_n)$ auffassen, ergibt sich nun das „konkrete" Konfidenzintervall $\bigl(\underline{\theta}(x_1,\ldots,x_n),\overline{\theta}(x_1,\ldots,x_n)\bigr)$ durch Einsetzen in (12).
- Falls beispielsweise für $\gamma = 0.95$ und $\sigma = 0,10$ ein solches Konfidenzintervall für die folgenden Daten 41.60, 41.48, 42.34, 41.95, 41.86, 42.18, 41.72, 42.26, 41.81, 42.04 ermittelt werden soll,
- dann entnehmen wir zunächst das Quantil $z_{0,975} = 1,96$ aus Tabelle 1
- und erhalten somit für $n = 10$

$$\begin{aligned}\underline{\theta}(x_1,\ldots,x_n) &= \overline{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = \quad 41.924 - 1.96\,\frac{0.1}{\sqrt{10}} &&= 41.862\\ \overline{\theta}(x_1,\ldots,x_n) &= \overline{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = \qquad\qquad \ldots &&= 41.986\end{aligned}$$

- Somit ergibt sich das (konkrete) Konfidenzintervall $(41.862, 41.986)$ für den Erwartungswert μ.

Beachte

- Das in (12) betrachtete (symmetrische) Konfidenzintervall kann dahingehend verallgemeinert werden, dass anstelle von (11) die folgende Gleichung für beliebige $\alpha_1,\alpha_2 \in [0,1/2)$ mit $\alpha_1+\alpha_2 = \alpha \in (0,1)$ betrachtet wird:

$$P_\theta\Bigl(z_{\alpha_2} \le \sqrt{n}\frac{\overline{X}_n-\mu}{\sigma} \le z_{1-\alpha_1}\Bigr) = 1-\alpha\,. \tag{14}$$

- Hieraus ergibt sich dann das (asymmetrische) Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für μ zum Niveau $\gamma = 1-\alpha$ mit

$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - z_{1-\alpha_1}\frac{\sigma}{\sqrt{n}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + z_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}\,. \tag{15}$$

- Für $\alpha_1 = 0$ ergibt sich insbesondere das sogenannte *einseitige Konfidenzintervall* $\bigl(-\infty,\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für μ zum Niveau $\gamma = 1-\alpha$ mit

$$\overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\,.$$

- Analog ergibt sich für $\alpha_2 = 0$ das einseitige Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\infty\bigr)$ für μ zum Niveau $\gamma = 1-\alpha$ mit

$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\,.$$

Im Unterschied zu dem oben diskutierten Fall nehmen wir nun an, dass $X_i \sim \mathrm{N}(\mu,\sigma^2)$, wobei sowohl $\mu\in\mathbb{R}$ als auch $\sigma^2 > 0$ unbekannt seien.

- Dabei lassen sich auf ähnliche Weise (wie bei bekanntem σ^2) symmetrische, allgemeine bzw. einseitige Konfidenzintervalle für μ gewinnen.
- Die bisher konstruierten Konfidenzintervalle für μ sind jetzt jedoch nicht mehr geeignet, weil sie die (unbekannte) Größe σ enthalten.
- Wir nutzen vielmehr das Ergebnis von Theorem 1.13 und erkennen (vgl. auch die Formel (1.45)), dass für beliebige $\alpha\in(0,1)$ und $\theta = (\mu,\sigma^2)\in\mathbb{R}\times(0,\infty)$

$$P_\theta\Bigl(t_{n-1,\alpha/2} \le \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n} \le t_{n-1,1-\alpha/2}\Bigr) = 1-\alpha \tag{16}$$

bzw. wegen (7)

$$P_\theta\Bigl(-t_{n-1,1-\alpha/2} \le \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n} \le t_{n-1,1-\alpha/2}\Bigr) = 1-\alpha$$

- D.h.
$$P_\theta\Big(\overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \le \mu \le \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\Big) = 1-\alpha\,.$$
- Also ist mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \tag{17}$$
ein (symmetrisches) Konfidenzintervall für μ zum Niveau $\gamma = 1-\alpha$ gegeben.

Beachte

- Das in (17) betrachtete (symmetrische) Konfidenzintervall kann dahingehend verallgemeinert werden, dass anstelle von (16) die folgende Gleichung für beliebige $\alpha_1,\alpha_2 \in [0,1/2)$ mit $\alpha_1+\alpha_2 = \alpha \in (0,1)$ betrachtet wird:
$$P_\theta\Big(t_{n-1,\alpha_2} \le \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n} \le t_{n-1,1-\alpha_1}\Big) = 1-\alpha\,. \tag{18}$$
- Hieraus ergibt sich dann das (asymmetrische) Konfidenzintervall $\big(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\big)$ für μ zum Niveau $\gamma = 1-\alpha$ mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - t_{n-1,1-\alpha_1}\frac{S_n}{\sqrt{n}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + t_{n-1,1-\alpha_2}\frac{S_n}{\sqrt{n}}\,. \tag{19}$$
- Für $\alpha_1 = 0$ ergibt sich insbesondere das einseitige Konfidenzintervall $\big(-\infty,\overline{\theta}(X_1,\ldots,X_n)\big)$ für μ zum Niveau $\gamma = 1-\alpha$ mit
$$\overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + t_{n-1,1-\alpha}\frac{S_n}{\sqrt{n}}\,.$$
- Analog ergibt sich für $\alpha_2 = 0$ das einseitige Konfidenzintervall $\big(\underline{\theta}(X_1,\ldots,X_n),\infty\big)$ für μ zum Niveau $\gamma = 1-\alpha$ mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - t_{n-1,1-\alpha}\frac{S_n}{\sqrt{n}}\,.$$

3.2.2 Konfidenzintervalle für die Varianz

In diesem Abschnitt diskutieren wir Konfidenzintervalle für die Varianz σ^2, wobei wir zunächst annehmen, dass der Erwartungswert μ unbekannt ist.

- Aus Theorem 1.11 ergibt sich dann, dass für beliebige $\alpha_1,\alpha_2 \in [0,1/2)$ mit $\alpha_1+\alpha_2 = \alpha \in (0,1)$
$$P_\theta\Big(\chi^2_{n-1,\alpha_2} \le \frac{(n-1)S_n^2}{\sigma^2} \le \chi^2_{n-1,1-\alpha_1}\Big) = 1-\alpha\,. \tag{20}$$
- Dies ergibt das Konfidenzintervall $\big(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\big)$ für σ^2 zum Niveau $\gamma = 1-\alpha$ mit
$$\underline{\theta}(X_1,\ldots,X_n) = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha_1}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1,\ldots,X_n) = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha_2}}\,. \tag{21}$$
- Für die Länge $\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n)$ dieses Konfidenzintervalls gilt
$$\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n) = (n-1)S_n^2\Big(\frac{1}{\chi^2_{n-1,\alpha_2}} - \frac{1}{\chi^2_{n-1,1-\alpha_1}}\Big) \tag{22}$$
mit
$$\mathbb{E}\left(\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n)\right) = (n-1)\sigma^2\Big(\frac{1}{\chi^2_{n-1,\alpha_2}} - \frac{1}{\chi^2_{n-1,1-\alpha_1}}\Big)\,.$$

Falls der Erwartungswert μ bekannt ist, dann lässt sich mit Hilfe der modifizierten Stichprobenvarianz $\widetilde{S}_n^2$ ein weiteres Konfidenzintervall für die Varianz σ^2 konstruieren, wobei

$$\widetilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \,.$$

- Weil $(X_i - \mu)/\sigma \sim \mathrm{N}(0,1)$, ergibt sich unmittelbar aus der Definition der χ^2-Verteilung, dass $n\widetilde{S}^2/\sigma^2$ eine χ^2-verteilte Zufallsvariable mit n Freiheitsgraden ist.
- Für beliebige $\alpha_1, \alpha_2 \in [0, 1/2)$ mit $\alpha_1 + \alpha_2 = \alpha \in (0,1)$ gilt also

$$P_\theta\Bigl(\chi^2_{n,\alpha_2} \le \frac{n\widetilde{S}_n^2}{\sigma^2} \le \chi^2_{n,1-\alpha_1}\Bigr) = 1 - \alpha \,. \tag{23}$$

- Dies ergibt das folgende (modifizierte) Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n), \overline{\theta}(X_1,\ldots,X_n)\bigr)$ für σ^2 zum Niveau $\gamma = 1-\alpha$ mit

$$\underline{\theta}(X_1,\ldots,X_n) = \frac{n\widetilde{S}_n^2}{\chi^2_{n,1-\alpha_1}} \qquad \text{bzw.} \qquad \overline{\theta}(X_1,\ldots,X_n) = \frac{n\widetilde{S}_n^2}{\chi^2_{n,\alpha_2}} \,, \tag{24}$$

wobei

$$\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n) = n\,\widetilde{S}_n^2\Bigl(\frac{1}{\chi^2_{n,\alpha_2}} - \frac{1}{\chi^2_{n,1-\alpha_1}}\Bigr) \tag{25}$$

mit

$$\mathbb{E}\,\bigl(\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n)\bigr) = n\,\sigma^2\Bigl(\frac{1}{\chi^2_{n,\alpha_2}} - \frac{1}{\chi^2_{n,1-\alpha_1}}\Bigr)\,.$$

Beachte

- Die in (25) gegebene Länge $\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n)$ des modifizierten Konfidenzintervalls kann *größer* als die Länge (22) des Intervalls sein, das in (21) für den Fall betrachtet wurde, dass der Erwartungswert μ unbekannt ist.
- Der Grund hierfür ist, dass „untypische" Stichprobenwerte, die große Abweichungen vom Erwartungswert μ aufweisen, besser durch das Stichprobenmittel $\overline{X}_n$ als durch μ kompensiert werden.

3.3 Zwei-Stichproben-Probleme

3.3.1 Modellbeschreibung

In Verallgemeinerung der Situation, die bisher in dieser Vorlesung betrachtetet wurde, sollen nun gleichzeitig zwei Datensätze $(x_{11},\ldots,x_{1n_1})$ und $(x_{21},\ldots,x_{2n_2})$ stochastisch modelliert werden.

- Dabei können die Stichprobenumfänge $n_1, n_2 \in \mathbb{N}$ beliebige natürliche Zahlen sein, die nicht notwendig gleich sein müssen.
- Die (konkreten) Stichproben $(x_{11},\ldots,x_{1n_1})$ und $(x_{21},\ldots,x_{2n_2})$ fassen wir als Realisierungen von zwei Zufallsstichproben $(X_{11},\ldots,X_{1n_1})$ bzw. $(X_{21},\ldots,X_{2n_2})$ auf.

Hierfür betrachten wir zwei (unendliche) Folgen $X_{11}, X_{12},\ldots$ und $X_{21}, X_{22},\ldots$ von Zufallsvariablen und nehmen an, dass

- die (zweidimensionalen) Zufallsvektoren $X_1, X_2, \ldots$ mit $X_i = (X_{1i}, X_{2i})$ unabhängig und identisch verteilt sind,

- die (gemeinsame) Verteilung von X_i zu einer parametrischen Familie von Verteilungen $\{P_\theta,\, \theta \in \Theta\}$ gehört, $\Theta \subset \mathbb{R}^m$,
- der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, über dem die Zufallsvektoren $X_1, X_2, \ldots$ definiert sind, der kanonische Wahrscheinlichkeitsraum dieser Zufallsvektoren ist.

Beachte Die Komponenten X_{1i} und X_{2i} von X_i müssen jedoch im allgemeinen *weder* unabhängig *noch* identisch verteilt sein.

Im Rahmen dieser einführenden Vorlesung betrachten wir lediglich den Fall, dass

- ein (nichtvektorieller) Funktionswert $g(\theta) \in \mathbb{R}$ des Parametervektors $\theta = (\theta_1, \ldots, \theta_m)$ aus den beobachteten Daten $(x_{11}, \ldots, x_{1n_1})$ bzw. $(x_{21}, \ldots, x_{2n_2})$ geschätzt werden soll, wobei $g : \Theta \to \mathbb{R}$ eine vorgegebene Borel-messbare Funktion sei.
- Dabei diskutieren wir insbesondere die Frage, wie das in Abschnitt 3.1 eingeführte Modell des Konfidenzintervalls verallgemeinert werden kann, um zu Konfidenzintervallen ausgehend von (vektoriellen) Stichprobenvariablen $X_i = (X_{1i}, X_{2i})$ zu gelangen.

Zur Vereinfachung der Schreibweise setzen wir $n = \max\{n_1, n_2\}$ und betrachten

- *zwei* Stichprobenfunktionen $\underline{\theta} : \mathbb{R}^{2n} \to \mathbb{R}$ und $\overline{\theta} : \mathbb{R}^{2n} \to \mathbb{R}$, so dass
$$\underline{\theta}(x_1, \ldots, x_n) \le \overline{\theta}(x_1, \ldots, x_n) \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^{2n}, \tag{26}$$
- wobei die Funktionswerte $\underline{\theta}(x_1, \ldots, x_n)$ und $\overline{\theta}(x_1, \ldots, x_n)$ jedoch nur von den (beobachteten) Komponenten $(x_{11}, \ldots, x_{1n_1})$ und $(x_{21}, \ldots, x_{2n_2})$ des Vektors $(x_1, \ldots, x_n)$ abhängen mögen.

Definition Sei $\gamma \in (0, 1)$ eine beliebige, jedoch fest vorgegebene Zahl. Dann heißt das zufällige Intervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ *Konfidenzintervall* für $g(\theta)$ zum *Niveau* γ, falls
$$P_\theta(\underline{\theta}(X_1, \ldots, X_n) \le g(\theta) \le \overline{\theta}(X_1, \ldots, X_n)) \ge \gamma, \qquad \forall\, \theta \in \Theta. \tag{27}$$

Beachte

- Bisher hatten wir in Abschnitt 3 stets den Fall betrachtet, dass $g(\theta) = \theta_j$ für ein $j \in \{1, \ldots, m\}$.
- In den folgenden Beispielen hat $g(\theta)$ die Form $g(\theta) = \theta_j - \theta_k$ bzw. $g(\theta) = \theta_j / \theta_k$ für ein vorgegebenes Paar $j, k \in \{1, \ldots, m\}$ von Indizes.

3.3.2 Konfidenzintervalle für Differenzen bzw. Quotienten von Parameterkomponenten

Die Stichprobenvariablen $X_i = (X_{1i}, X_{2i})$ seien nun normalverteilte Zufallsvektoren, wobei wir in diesem Abschnitt außerdem voraussetzen, dass die Komponenten X_{1i} und X_{2i} von X_i unabhängige (jedoch im allgemeinen nicht identisch verteilte) Zufallsvariablen sind.

Wir betrachten die folgenden zwei Beispiele.

1. *Konfidenzintervall für die Differenz zweier Erwartungswerte (bei bekannten Varianzen)*
   - Für zwei beliebige, jedoch vorgegebene Zahlen $n_1, n_2 \in \mathbb{N}$ betrachten wir zwei unabhängige Zufallsstichproben $(X_{11}, \ldots, X_{1n_1})$ und $(X_{21}, \ldots, X_{2n_2})$.

- Dabei nehmen wir an, dass $X_{1i} \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_{2i} \sim \mathrm{N}(\mu_2, \sigma_2^2)$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$ und (bekannte) $\sigma_1^2, \sigma_2^2 > 0$, d.h., $\theta = (\mu_1, \mu_2)$.
- Dann sind die Stichprobenmittel $\overline{X}_{1n_1}$ und $\overline{X}_{2n_2}$ unabhängige Zufallsvariablen mit

$$\overline{X}_{1n_1} \sim \mathrm{N}(\mu_1, \sigma_1^2/n_1)\,, \qquad \overline{X}_{2n_2} \sim \mathrm{N}(\mu_2, \sigma_2^2/n_2)\,,$$

  vgl. Theorem 1.11.
- Hieraus folgt, dass

$$\overline{X}_{1n_1} - \overline{X}_{2n_2} \sim \mathrm{N}\Bigl(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Bigr)$$

  bzw.

$$\frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)\,.$$

- Also gilt für jedes $\alpha \in (0,1)$

$$P_\theta\Biggl(z_{\frac{\alpha}{2}} \le \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \le z_{1-\frac{\alpha}{2}}\Biggr) = 1 - \alpha$$

  bzw.

$$P_\theta\Biggl(\underbrace{\overline{X}_{1n_1} - \overline{X}_{2n_2} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{=\underline{\theta}} \le \mu_1 - \mu_2 \le \underbrace{\overline{X}_{1n_1} - \overline{X}_{2n_2} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{=\overline{\theta}}\Biggr) = 1 - \alpha\,.$$

- Die auf diese Weise bestimmten Stichprobenfunktionen $\underline{\theta}(X_1, \ldots, X_n)$ und $\overline{\theta}(X_1, \ldots, X_n)$ ergeben ein Konfidenzinterval für $\mu_1 - \mu_2$ zum Niveau $\gamma = 1 - \alpha$.

2. *Konfidenzintervall für den Quotienten zweier Varianzen (bei unbekannten Erwartungswerten)*

   - Ähnlich wie in dem vohergehenden Beispiel betrachten wir für zwei beliebige, jedoch vorgegebene Zahlen $n_1, n_2 \in \mathbb{N}$ zwei unabhängige Zufallsstichproben $(X_{11}, \ldots, X_{1n_1})$ und $(X_{21}, \ldots, X_{2n_2})$.
   - Dabei nehmen wir an, dass $X_{1i} \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_{2i} \sim \mathrm{N}(\mu_2, \sigma_2^2)$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$ und (unbekannte) $\sigma_1^2, \sigma_2^2 > 0$, d.h., $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.
   - Dann sind die Stichprobenvarianzen $S_{1n_1}^2$ und $S_{2n_2}^2$ unabhängige Zufallsvariablen und wegen Theorem 1.11 gilt

$$\frac{(n_1 - 1)S_{1n_1}^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1} \qquad \text{bzw.} \qquad \frac{(n_2 - 1)S_{2n_2}^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1}\,.$$

   - Hieraus und aus der Definition der F-Verteilung, die in Abschnitt 3.1.3 eingeführt wurde, ergibt sich

$$\frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2} \sim \mathrm{F}_{n_2-1, n_1-1}\,.$$

   - Also gilt für jedes $\alpha \in (0,1)$

$$P_\theta\Bigl(F_{n_2-1, n_1-1, \frac{\alpha}{2}} \le \frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2} \le F_{n_2-1, n_1-1, 1-\frac{\alpha}{2}}\Bigr) = 1 - \alpha$$

bzw.

$$P_\theta\Big(\underbrace{\frac{S^2_{1n_1}}{S^2_{2n_2}}\,F_{n_2-1,n_1-1,\frac{\alpha}{2}}}_{=\underline{\theta}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \underbrace{\frac{S^2_{1n_1}}{S^2_{2n_2}}\,F_{n_2-1,n_1-1,1-\frac{\alpha}{2}}}_{=\overline{\theta}}\Big) = 1-\alpha\,,$$

wobei $F_{n_2-1,n_1-1,\gamma}$ das γ-Quantil der F-Verteilung mit (n_2-1, n_1-1) Freiheitsgraden bezeichnet, vgl. Tabellen 4a–4f.

- Die auf diese Weise bestimmten Stichprobenfunktionen $\underline{\theta}(X_1,\ldots,X_n)$ und $\overline{\theta}(X_1,\ldots,X_n)$ ergeben ein Konfidenzinterval für σ_1^2/σ_2^2 zum Niveau $\gamma = 1-\alpha$.

3.3.3 Verbundene Stichproben

- Die in Abschnitt 3.3.2 gemachte Modellannahme, dass $(X_{11},\ldots,X_{1n_1})$ und $(X_{21},\ldots,X_{2n_2})$ unabhängige Zufallsstichproben sind, ist manchmal unrealistisch.
- Betrachten beispielsweise die folgende Tabelle der Körpergrößen $x_{11},\ldots,x_{1\,12}$ einer (konkreten) Stichprobe von 12 Müttern und der Körpergrößen $x_{21},\ldots,x_{2\,12}$ ihrer ältesten Töchter.

x_{1i}	1.65	1.63	1.67	1.64	1.68	1.62	1.70	1.66	1.68	1.67	1.69	1.71
x_{2i}	1.68	1.66	1.68	1.65	1.69	1.66	1.68	1.65	1.71	1.67	1.68	1.70

- Aus dieser Tabelle ist ersichtlich (und dies stimmt mit unserer Intuition überein), dass vermutlich ein Zusammenhang zwischen den Körpergrößen $x_{11},\ldots,x_{1\,12}$ der Mütter und den Körpergrößen $x_{21},\ldots,x_{2\,12}$ ihrer ältesten Töchter besteht.
- In diesem Fall ist es nicht sinnvoll anzunehmen, dass $(x_{11},\ldots,x_{1\,12})$ und $(x_{21},\ldots,x_{2\,12})$ die Realisierungen von zwei unabhängigen Zufallsstichproben $(X_{11},\ldots,X_{1\,12})$ bzw. $(X_{21},\ldots,X_{2\,12})$ sind.

Dies führt zum Begriff verbundener Stichproben, der in dem folgenden Beispiel erläutert wird, in dem wir ein Konfidenzintervall für die Differenz der Erwartungswerte von verbundenen Stichproben herleiten.

- Betrachten die Zufallsvektoren $X_1,\ldots,X_n$ mit $X_i = (X_{1i}, X_{2i})$, wobei $X_1,\ldots,X_n$ (so wie bisher stets angenommen) unabhängig und identisch verteilt seien.
- Die (einzelnen) Zufallstichproben $(X_{11},\ldots,X_{1n})$ und $(X_{21},\ldots,X_{2n})$ müssen jedoch jetzt *nicht* unabhängig sein, weshalb man von *verbundenen Stichproben* spricht.
- Wir nehmen an, dass $\mathbb{E}\,X_{1i} = \mu_1$ und $\mathbb{E}\,X_{2i} = \mu_2$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$.
- Außerdem seien die Differenzen $Y_1,\ldots,Y_n$ mit $Y_i = X_{1i} - X_{2i}$ unabhängige und identisch (normal-) verteilte Zufallsvariablen mit $Y_i \sim \mathrm{N}(\mu_1-\mu_2, \sigma^2)$ für ein (unbekanntes) $\sigma^2 > 0$, d.h., $\theta = (\mu_1-\mu_2, \sigma^2)$.
- Die gleichen Überlegungen wie in Abschnitt 3.2.1 ergeben nun, dass für jedes $\alpha \in (0,1)$

$$P_\theta\Big(-t_{n-1,1-\alpha/2} \le \frac{\sqrt{n}(\overline{Y}_n - (\mu_1-\mu_2))}{S_{Y,n}} \le t_{n-1,1-\alpha/2}\Big) = 1-\alpha \tag{28}$$

bzw.

$$P_\theta\Big(\overline{Y}_n - t_{n-1,1-\alpha/2}\frac{S_{Y,n}}{\sqrt{n}} \le \mu_1-\mu_2 \le \overline{Y}_n + t_{n-1,1-\alpha/2}\frac{S_{Y,n}}{\sqrt{n}}\Big) = 1-\alpha\,,$$

wobei $\overline{Y}_n = \frac{1}{n}\sum\limits_{i=1}^n Y_i$ und $S^2_{Y,n} = \frac{1}{n-1}\sum\limits_{i=1}^n (Y_i - \overline{Y}_n)^2$.

- Also ist mit

$$\underline{\theta}(X_1,\ldots,X_n)=\overline{Y}_n-t_{n-1,1-\alpha/2}\frac{S_{Y,n}}{\sqrt{n}}\qquad\text{bzw.}\qquad\overline{\theta}(X_1,\ldots,X_n)=\overline{Y}_n+t_{n-1,1-\alpha/2}\frac{S_{Y,n}}{\sqrt{n}}\tag{29}$$

ein (symmetrisches) Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für $\mu_1-\mu_2$ zum Niveau $\gamma=1-\alpha$ gegeben.

Beachte

- Für die oben betrachtete (konkrete) Stichprobe $(x_1,\ldots,x_{12})$ der Körpergrößen von 12 Müttern und ihrer ältesten Töchter gilt

$$\overline{y}_{12}=-0.0092\,,\qquad s^2_{y,12}=0.00039\quad\text{bzw.}\quad s_{y,12}=0.0197\,.$$

- Für $\gamma=0.95$ entnehmen wir das Quantil $t_{11,0.975}=2.201$ aus Tabelle 3
- und erhalten

$$\begin{aligned}\underline{\theta}(x_1,\ldots,x_{12})=\overline{y}_{12}-t_{11,0.975}\frac{s_{y,12}}{\sqrt{12}}&=\;-0.0092-2.201\,\frac{0.0197}{\sqrt{12}}\;=-0.022\\ \overline{\theta}(x_1,\ldots,x_{12})=\overline{y}_{12}+t_{11,0.975}\frac{s_{y,12}}{\sqrt{12}}&=\;-0.0092+2.201\,\frac{0.0197}{\sqrt{12}}\;=0.0033\end{aligned}$$

- Somit ergibt sich das (konkrete) Konfidenzintervall $(-0.022,\,0.0033)$ für $\mu_1-\mu_2$.

3.4 Asymptotische Konfidenzintervalle

Wir zeigen nun, wie man mit Hilfe des zentralen Grenzwertsatzes (vgl. Theorem WR–5.16) und des starken Gesetzes der großen Zahlen (vgl. Theorem WR–5.15) asymptotische Konfidenzintervalle auf einfache Weise herleiten kann.

3.4.1 Ein-Stichproben-Probleme

Wir kehren zunächst zur Betrachtung von Ein-Stichproben-Problemen zurück. D.h., wir nehmen an, dass *ein* Datensatz $(x_1,\ldots,x_n)$ beobachtet wird, den wir als Realisierung *einer* Zufallsstichprobe $(X_1,\ldots,X_n)$ auffassen.

Theorem 3.2 *Der Erwartungswert μ der Stichprobenvariablen $X_1,\ldots,X_n$ sei eine Komponente des Parameters $\theta\in\Theta$, und $\alpha\in(0,1)$ sei eine beliebige, jedoch vorgegebene Zahl. Dann ist durch $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ mit*

$$\underline{\theta}(X_1,\ldots,X_n)=\overline{X}_n-\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\,,\qquad\overline{\theta}(X_1,\ldots,X_n)=\overline{X}_n+\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}$$

ein asymptotisches Konfidenzintervall für μ zum Niveau $\gamma=1-\alpha$ gegeben, wobei $\overline{X}_n$ bzw. S_n das Stichprobenmittel bzw. die Wurzel der Stichprobenvarianz sind.

Beweis Die Behauptung ergibt sich unmittelbar aus Formel (1.20), denn

$$\begin{aligned}\lim_{n\to\infty}P_\theta\Bigl(\overline{X}_n-\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}<\mu<\overline{X}_n+\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\Bigr)&=\lim_{n\to\infty}P_\theta\Bigl(\frac{-z_{1-\alpha/2}S_n}{\sqrt{n}}<\overline{X}_n-\mu<\frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\Bigr)\\&=1-\alpha\,.\end{aligned}$$

□

Für spezielle Beispiele von Familien parametrischer Verteilungen kann man asymptotische Konfidenzintervalle für den Erwartungswert μ der Stichprobenvariablen $(X_1,\ldots,X_n)$ konstruieren, ohne die Stichprobenstreuung S_n^2 betrachten zu müssen.

1. *Konfidenzintervall für den Erwartungswert λ bei Poisson-Verteilung*

   - Wir nehmen an, dass $X_i \sim \mathrm{Poi}(\lambda)$ für ein (unbekanntes) $\lambda > 0$.
   - Weil dann $\mathbb{E}\, X_i = \lambda$ und $\operatorname{Var} X_i = \lambda$ gilt, ergibt sich aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen (vgl. Theorem WR–5.16), dass für jedes $\alpha \in (0,1)$
$$\lim_{n\to\infty} P_\lambda\Bigl(-z_{1-\alpha/2} \le \sqrt{n}\frac{\overline{X}_n-\lambda}{\sqrt{\lambda}} \le z_{1-\alpha/2}\Bigr) = 1-\alpha\,. \tag{30}$$
   - Durch Quadrierung der Ungleichungen in (30) und anschließende Auflösung nach λ ergibt sich, dass (30) äquivalent ist mit
$$\lim_{n\to\infty} P_\lambda\left(\Bigl|\lambda - \Bigl(\overline{X}_n + \frac{z^2_{1-\alpha/2}}{2n}\Bigr)\Bigr| \le \sqrt{\frac{\overline{X}_n z^2_{1-\alpha/2}}{n} + \frac{z^4_{1-\alpha/2}}{4n^2}}\right) = 1-\alpha\,.$$
   - Also ist mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + \frac{z^2_{1-\alpha/2}}{2n} - \sqrt{\frac{\overline{X}_n z^2_{1-\alpha/2}}{n} + \frac{z^4_{1-\alpha/2}}{4n^2}}\,,$$
$$\overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + \frac{z^2_{1-\alpha/2}}{2n} + \sqrt{\frac{\overline{X}_n z^2_{1-\alpha/2}}{n} + \frac{z^4_{1-\alpha/2}}{4n^2}}$$
     ein asymptotisches Konfidenzintervall $(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n))$ für λ zum Niveau $\gamma = 1-\alpha$ gegeben.
   - Für die Länge dieses Konfidenzintervalls gilt
$$\overline{\theta}(X_1,\ldots,X_n) - \underline{\theta}(X_1,\ldots,X_n) = 2\sqrt{\frac{\overline{X}_n z^2_{1-\alpha/2}}{n} + \frac{z^4_{1-\alpha/2}}{4n^2}}\,. \tag{31}$$

   Beachte

   - Ein weiteres asymptotisches Konfidenzintervall für λ ergibt sich, wenn die Größe $\sqrt{\lambda}$ im Nenner von (30) ersetzt wird durch $\sqrt{\overline{X}_n}$.
   - Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15) ergibt sich nämlich, dass
$$\overline{X}_n \xrightarrow{\text{f.s.}} \lambda\,.$$
   - Mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11) ergibt sich hieraus und aus (30) die Gültigkeit von
$$\lim_{n\to\infty} P_\lambda\Bigl(-z_{1-\alpha/2} \le \sqrt{n}\,\frac{\overline{X}_n-\lambda}{\sqrt{\overline{X}_n}} \le z_{1-\alpha/2}\Bigr) = 1-\alpha \tag{32}$$
     für jedes $\alpha \in (0,1)$.
   - Also ist mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n}\,, \qquad \overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n} \tag{33}$$
     ein weiteres asymptotisches Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für λ zum Niveau $\gamma = 1-\alpha$ gegeben.

- Dabei ist allerdings zu beachten, dass die Zufallsvariable $\underline{\theta}(X_1,\ldots,X_n)$ in (33) auch negative Werte annehmen kann, obwohl bei der Poisson-Verteilung vorausgesetzt wird, dass $\lambda > 0$.
- Deshalb kann man anstelle von (33) die Zufallsvariablen

$$\underline{\theta}(X_1,\ldots,X_n) = \max\Bigl\{0, \overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n}\Bigr\}\,, \qquad \overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n}$$

  als Endpunkte des Konfidenzintervalls $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für λ betrachten.
- Die Länge dieses Konfidenzintervalls ist stets *kleiner* als die in (31) gegebene Länge des Intervalls, das aus dem Ansatz (30) resultiert.
- Der Grund hierfür ist, dass „untypische" Stichprobenwerte, die große Abweichungen vom Erwartungswert λ aufweisen, besser durch das Stichprobenmittel $\overline{X}_n$ im Nenner von (32) als durch den Erwartungswert λ im Nenner von (30) kompensiert werden.

2. *Konfidenzintervall für die „Erfolgswahrscheinlichkeit" p bei Bernoulli-Verteilung*

- Wir nehmen an, dass $X_i \sim \mathrm{Bin}(1,p)$ für ein (unbekanntes) $p \in (0,1)$.
- Weil dann $\mathbb{E}\,X_i = p$ und $\mathrm{Var}\,X_i = p(1-p)$ gilt, ergibt sich aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen (vgl. Theorem WR–5.16), dass für jedes $\alpha \in (0,1)$

$$\lim_{n\to\infty} P_p\Bigl(-z_{1-\alpha/2} \le \sqrt{n}\,\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \le z_{1-\alpha/2}\Bigr) = 1-\alpha\,. \tag{34}$$

- Durch Quadrierung der Ungleichungen in (34) und anschließende Auflösung nach p ergibt sich, dass (34) äquivalent ist mit

$$\lim_{n\to\infty} P_p\bigl(\underline{\theta}(X_1,\ldots,X_n) \le p \le \overline{\theta}(X_1,\ldots,X_n)\bigr) = 1-\alpha\,,$$

  wobei

$$\underline{\theta}(X_1,\ldots,X_n) = \frac{1}{n+z^2_{1-\alpha/2}}\left(n\overline{X}_n + \frac{z^2_{1-\alpha/2}}{2} - z_{1-\alpha/2}\sqrt{n\overline{X}_n(1-\overline{X}_n) + \frac{z^2_{1-\alpha/2}}{4}}\right)$$

  bzw.

$$\overline{\theta}(X_1,\ldots,X_n) = \frac{1}{n+z^2_{1-\alpha/2}}\left(n\overline{X}_n + \frac{z^2_{1-\alpha/2}}{2} + z_{1-\alpha/2}\sqrt{n\overline{X}_n(1-\overline{X}_n) + \frac{z^2_{1-\alpha/2}}{4}}\right).$$

- Auf diese Weise ergibt sich ein asymptotisches Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für p zum Niveau $\gamma = 1-\alpha$.

Beachte

- Ähnlich wie in dem vorhergehenden Beispiel erhält man ein einfacheres asymptotisches Konfidenzintervall für p, wenn man berücksichtigt, dass wegen des starken Gesetzes der großen Zahlen (vgl. Theorem WR–5.15)

$$\overline{X}_n \xrightarrow{\text{f.s.}} p\,.$$

- Hieraus und aus (34) ergibt sich nun mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11), dass

$$\lim_{n\to\infty} P_p\Bigl(-z_{1-\alpha/2} \le \sqrt{n}\,\frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \le z_{1-\alpha/2}\Bigr) = 1-\alpha\,. \tag{35}$$

- Also ist mit
$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n(1-\overline{X}_n)}$$
bzw.
$$\overline{\theta}(X_1,\ldots,X_n) = \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n(1-\overline{X}_n)}$$
ein weiteres asymptotisches Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n)\bigr)$ für p zum Niveau $\gamma = 1-\alpha$ gegeben.
- Dabei ist jedoch zu beachten, dass die Zufallsvariable $\underline{\theta}(X_1,\ldots,X_n)$ negative Werte und $\overline{\theta}(X_1,\ldots,X_n)$ Werte größer als 1 annehmen kann, obwohl bei der Bernoulli-Verteilung vorausgesetzt wird, dass $0 < p < 1$.
- Deshalb betrachtet man die modifizierten Schätzer
$$\underline{\theta}(X_1,\ldots,X_n) = \max\Bigl\{0, \overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n(1-\overline{X}_n)}\Bigr\}$$
bzw.
$$\overline{\theta}(X_1,\ldots,X_n) = \min\Bigl\{1, \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n(1-\overline{X}_n)}\Bigr\}$$
als Endpunkte des Konfidenzintervalls $(\underline{\theta}(X_1,\ldots,X_n),\overline{\theta}(X_1,\ldots,X_n))$ für p.

3.4.2 Zwei-Stichproben-Probleme

Wir betrachten nun noch zwei Beispiele für asymptotische Konfidenzintervalle bei *Zwei-Stichproben-Problemen.*

- Seien $n_1, n_2 \in \mathbb{N}$ zwei beliebige, jedoch vorgegebene natürliche Zahlen.
- So wie in Abschnitt 3.3.2 fassen wir die (konkreten) Stichproben $(x_{11},\ldots,x_{1n_1})$ und $(x_{21},\ldots,x_{2n_2})$ als Realisierungen von zwei unabhängigen (d.h. nicht verbundenen) Zufallsstichproben $(X_{11},\ldots,X_{1n_1})$ bzw. $(X_{21},\ldots,X_{2n_2})$ auf.

1. *Konfidenzintervall für die Differenz zweier Erwartungswerte (bei Poisson-Verteilung)*
   - Wir nehmen an, dass $X_{1i} \sim \mathrm{Poi}(\lambda_1)$ und $X_{2i} \sim \mathrm{Poi}(\lambda_2)$ für (unbekannte) $\lambda_1, \lambda_2 > 0$.
   - Um ein asymptotisches Konfidenzintervall für die Differenz $\lambda_1 - \lambda_2$ herzuleiten, wenden wir den *zentralen Grenzwertsatz von Ljapunow* für Summen von unabhängigen (jedoch nichtnotwendig identisch verteilten) Zufallsvariablen an; vgl. Theorem WR–5.23.
   - *Zur Erinnerung*: Für jedes $n \in \mathbb{N}$ sei $Y_{n1},\ldots,Y_{nn}$ eine Folge von unabhängigen Zufallsvariablen, so dass für jedes $k \in \{1,\ldots,n\}$
   $$\mathbb{E}\, Y_{nk} = 0\,, \qquad 0 < \sigma^2_{nk} = \mathrm{Var}\, Y_{nk} < \infty\,, \qquad \sum_{k=1}^n \sigma^2_{nk} = 1 \tag{36}$$
   und für ein $\delta > 0$
   $$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left(|Y_{nk}|^{2+\delta}\right) = 0\,. \tag{37}$$
   - In Theorem WR–5.23 hatten wir gezeigt, dass dann für jedes $x \in \mathbb{R}$
   $$\lim_{n\to\infty} P_{(\lambda_1,\lambda_2)}(Y_{n1} + \ldots + Y_{nn} \le x) = \Phi(x)\,, \tag{38}$$
   wobei $\Phi : \mathbb{R} \to [0,1]$ die Verteilungsfunktion der $\mathrm{N}(0,1)$-Verteilung ist.

- Hieraus ergibt sich nun der folgende Hilfssatz.

Lemma 3.1 *Falls $n_1, n_2 \to \infty$, dann gilt*

$$\frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - (\lambda_1 - \lambda_2)}{\sqrt{\dfrac{\lambda_1}{n_1} + \dfrac{\lambda_2}{n_2}}} \xrightarrow{d} Y \sim \mathrm{N}(0,1)\,, \tag{39}$$

wobei

$$\overline{X}_{in_i} = \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i}\,, \qquad i = 1, 2\,.$$

Beweis

- Mit der Schreibweise $n = n_1 + n_2$ und

$$Y_{nk} = \frac{X_{1k} - \lambda_1}{n_1\sqrt{\dfrac{\lambda_1}{n_1} + \dfrac{\lambda_2}{n_2}}} \qquad \forall\, k = 1, \ldots, n_1$$

  bzw.

$$Y_{nk} = -\frac{X_{2,k-n_1} - \lambda_2}{n_2\sqrt{\dfrac{\lambda_1}{n_1} + \dfrac{\lambda_2}{n_2}}} \qquad \forall\, k = n_1 + 1, \ldots, n_1 + n_2$$

  gilt offenbar für jedes $k \in \{1, \ldots, n\}$

$$\mathbb{E}\, Y_{nk} = 0\,, \qquad 0 < \sigma_{nk}^2 = \mathrm{Var}\, Y_{nk} < \infty\,, \qquad \sum_{k=1}^{n} \sigma_{nk}^2 = 1\,,$$

  d.h., die Bedingung (36) ist erfüllt.

- Außerdem gilt für jedes $\delta > 0$ und für $n_1, n_2 \to \infty$

$$\begin{aligned}
\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left(|Y_{nk}|^{2+\delta}\right) &= \lim_{n_1,n_2\to\infty} \left(\frac{\mathbb{E}\left(|X_{11} - \lambda_1|^{2+\delta}\right)}{n_1^{1+\delta}\left(\dfrac{\lambda_1}{n_1} + \dfrac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} + \frac{\mathbb{E}\left(|X_{21} - \lambda_2|^{2+\delta}\right)}{n_2^{1+\delta}\left(\dfrac{\lambda_1}{n_1} + \dfrac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} \right) \\
&= \lim_{n_1,n_2\to\infty} \left(\frac{\mathbb{E}\left(|X_{11} - \lambda_1|^{2+\delta}\right)}{n_1^{\delta/2}\left(\lambda_1 + \dfrac{n_1\lambda_2}{n_2}\right)^{1+\delta/2}} + \frac{\mathbb{E}\left(|X_{21} - \lambda_2|^{2+\delta}\right)}{n_2^{\delta/2}\left(\dfrac{n_2\lambda_1}{n_1} + \lambda_2\right)^{1+\delta/2}} \right) \\
&\le \lim_{n_1,n_2\to\infty} \left(\frac{\mathbb{E}\left(|X_{11} - \lambda_1|^{2+\delta}\right)}{n_1^{\delta/2}\lambda_1^{1+\delta/2}} + \frac{\mathbb{E}\left(|X_{21} - \lambda_2|^{2+\delta}\right)}{n_2^{\delta/2}\lambda_2^{1+\delta/2}} \right) \\
&= 0\,.
\end{aligned}$$

- Damit ist auch die Gültigkeit der Bedingung (37) gezeigt, und die Behauptung (39) ergibt sich somit aus (38). □

Beachte

- Nun können wir genauso wie bei den beiden Beispielen vorgehen, die in Abschnitt 3.4.1 betrachtet worden sind.
- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15) ergibt sich nämlich, dass für $n_1, n_2 \to \infty$

$$\overline{X}_{in_i} \xrightarrow{\text{f.s.}} \lambda_i\,, \qquad i = 1, 2\,.$$

- Mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11) ergibt sich hieraus und aus Lemma 3.1, dass für jedes $\alpha \in (0,1)$

$$\lim_{n_1,n_2\to\infty} P_{(\lambda_1,\lambda_2)}\Bigl(-z_{1-\alpha/2} \le \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - (\lambda_1-\lambda_2)}{\sqrt{\dfrac{\overline{X}_{1n_1}}{n_1} + \dfrac{\overline{X}_{2n_2}}{n_2}}} \le z_{1-\alpha/2}\Bigr) = 1-\alpha \tag{40}$$

- Deshalb ist mit

$$\underline{\theta}(X_1,\ldots,X_n) = \overline{X}_{1n_1} - \overline{X}_{2n_2} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_{1n_1}}{n_1} + \frac{\overline{X}_{2n_2}}{n_2}}$$

bzw.

$$\overline{\theta}(X_1,\ldots,X_n) = \overline{X}_{1n_1} - \overline{X}_{2n_2} + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_{1n_1}}{n_1} + \frac{\overline{X}_{2n_2}}{n_2}}$$

ein asymptotisches Konfidenzintervall $\bigl(\underline{\theta}(X_1,\ldots,X_n), \overline{\theta}(X_1,\ldots,X_n)\bigr)$ für $\lambda_1-\lambda_2$ zum Niveau $\gamma = 1-\alpha$ gegeben.

2. *Konfidenzintervall für den Quotienten von Erwartungswerten (bei Poisson-Verteilung)*

- So wie in dem vorhergehenden Beispiel nehmen wir an, dass $X_{1i} \sim \mathrm{Poi}(\lambda_1)$ und $X_{2i} \sim \mathrm{Poi}(\lambda_2)$ für (unbekannte) $\lambda_1,\lambda_2 > 0$ gilt, wobei jedoch der Quotient $\rho = n_1/n_2$ der beiden Stichprobenumfänge $n_1,n_2 \in \mathbb{N}$ nun eine Konstante sei.
- Wir betrachten den Quotienten

$$p = \frac{\lambda_2}{\rho\lambda_1 + \lambda_2} \qquad \left(= \frac{n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2}\right), \tag{41}$$

wobei $n_2\lambda_2$ der Erwartungswert der Summe $X_{21} + \ldots + X_{2n_2}$ der Stichprobenvariablen in der zweiten (Teil-) Stichprobe und $n_1\lambda_1 + n_2\lambda_2$ der Erwartungswert der Summe der insgesamt beobachteten Stichprobenvariablen ist.
- Um ein asymptotisches Konfidenzintervall für p herzuleiten, wenden wir den zentralen Grenzwertsatz für Summen mit einer *zufälligen Anzahl* von (unabhängigen und identisch verteilten) Summanden an; vgl. Beispiel 3 in Abschnitt WR–5.3.2.
- *Zur Erinnerung*: Seien $Y_1, Y_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit $\mathbb{E}\,(Y^2) < \infty$ und $\sigma^2 = \mathrm{Var}\, Y_n > 0$.
- Sei $N_1, N_2, \ldots$ eine Folge von nichtnegativen und ganzzahligen Zufallvariablen, so dass $N_1 \le N_2 \le \ldots$ und $N_n \to \infty$ mit Wahrscheinlichkeit 1.
- Außerdem gebe es Konstanten $a_1, a_2, \ldots > 0$ und $c > 0$ mit $a_1 \le a_2 \le \ldots$ und $a_n \to \infty$ gibt, so dass für $n \to \infty$

$$\frac{N_n}{a_n} \xrightarrow{\mathrm{P}} c\,. \tag{42}$$

- In Beispiel 3 des Abschnittes WR–5.3.2 hatten wir gezeigt, dass dann für $n \to \infty$

$$\frac{S'_{N_n}}{\sqrt{N_n}} \xrightarrow{\mathrm{d}} Y \sim \mathrm{N}(0,1)\,, \tag{43}$$

wobei $S'_n = Y'_1 + \ldots + Y'_n$ und $Y'_i = (Y_i - \mathbb{E}\,Y_i)/\sigma$.
- Hieraus ergibt sich der folgende Hilfssatz.

Lemma 3.2 *Für die Stichprobenumfänge n_1, n_2 gelte $n_1, n_2 \to \infty$, so dass der Quotient $\rho = n_1/n_2 \in (0,\infty)$ eine Konstante ist, die nicht von n_1, n_2 abhängt. Für die Partialsummen $S_{in_i} = X_{i1} + \ldots + X_{in_i}$, wobei $i = 1,2$, gilt dann*

$$\sqrt{S_{1n_1} + S_{2n_2}}\Bigl(\frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} - p\Bigr) \xrightarrow{\mathrm{d}} Y \sim \mathrm{N}(0, p(1-p)), \tag{44}$$

wobei p in (41) gegeben ist.

Beweis

- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR–5.15) ergibt sich, dass

$$\begin{aligned}\frac{S_{1n_1} + S_{2n_2}}{n_1\lambda_1 + n_2\lambda_2} &= \frac{S_{1n_1}}{n_1\lambda_1 + n_2\lambda_2} + \frac{S_{2n_2}}{n_1\lambda_1 + n_2\lambda_2}\\ &= \frac{\overline{X}_{1n_1}}{\lambda_1 + \rho^{-1}\lambda_2} + \frac{\overline{X}_{2n_2}}{\rho\lambda_1 + \lambda_2}\\ &\xrightarrow{\text{f.s.}} \frac{\lambda_1}{\lambda_1 + \rho^{-1}\lambda_2} + \frac{\lambda_2}{\rho\lambda_1 + \lambda_2}\\ &= \frac{\rho\lambda_1}{\rho\lambda_1 + \lambda_2} + \frac{\lambda_2}{\rho\lambda_1 + \lambda_2} = 1\,.\end{aligned}$$

- Für $N_{n_2} = S_{1n_1} + S_{2n_2}$ ist somit die Bedingung (42) erfüllt.
- Für beliebige $n \in \mathbb{N}$ und $k \in \{0, 1, \ldots, n\}$ gilt außerdem

$$\begin{aligned}P_{(\lambda_1,\lambda_2)}(S_{2n_2} = k \mid S_{1n_1} + S_{2n_2} = n) &= \frac{P_{(\lambda_1,\lambda_2)}(S_{2n_2} = k,\, S_{1n_1} + S_{2n_2} = n)}{P_{(\lambda_1,\lambda_2)}(S_{1n_1} + S_{2n_2} = n)}\\ &= \frac{P_{(\lambda_1,\lambda_2)}(S_{2n_2} = k,\, S_{1n_1} = n-k)}{P_{(\lambda_1,\lambda_2)}(S_{1n_1} + S_{2n_2} = n)}\\ &= \frac{e^{-n_2\lambda_2}\,\dfrac{(n_2\lambda_2)^k}{k!}\; e^{-n_1\lambda_1}\,\dfrac{(n_1\lambda_1)^{n-k}}{(n-k)!}}{e^{-(n_1\lambda_1 + n_2\lambda_2)}\,\dfrac{(n_1\lambda_1 + n_2\lambda_2)^n}{n!}}\\ &= \frac{n!}{k!(n-k)!}\Bigl(\frac{n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2}\Bigr)^k\Bigl(\frac{n_1\lambda_1}{n_1\lambda_1 + n_2\lambda_2}\Bigr)^{n-k}\\ &= \binom{n}{k} p^k (1-p)^{n-k}\,,\end{aligned}$$

  wobei p in (41) gegeben ist.
- Die bedingte Verteilung von S_{2n_2} unter der Bedingung $\{S_{1n_1} + S_{2n_2} = n\}$ ist also die Binomialverteilung $\mathrm{Bin}(n,p)$.
- Hieraus folgt, dass

$$P_{(\lambda_1,\lambda_2)}\Bigl(\frac{S_{2n_2} - pn}{\sqrt{p(1-p)n}} \le x \mid S_{1n_1} + S_{2n_2} = n\Bigr) = P\Bigl(\frac{S'_n}{\sqrt{n}} \le x\Bigr)\,, \qquad \forall\, x \in \mathbb{R}\,,$$

  falls $Y_i \sim \mathrm{Bin}(1,p)$ bzw. $Y'_i = (Y_i - p)/\sqrt{p(1-p)}$.
- Durch beidseitiges Multiplizieren dieser Gleichung mit $P_{(\lambda_1,\lambda_2)}\bigl(N_{n_2} = n\bigr)$ und anschließendes Summieren über n ergibt sich aus der Formel der totalen Wahrscheinlichkeit, dass

$$\frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{p(1-p)(S_{1n_1} + S_{2n_2})}} \stackrel{\mathrm{d}}{=} \frac{S'_{N_{n_2}}}{\sqrt{N_{n_2}}}\,,$$

  wobei die Zufallsvariablen $Y_1, Y_2, \ldots$ unabhängig von $N_{n_2} = S_{1n_1} + S_{2n_2}$ sind.

- Aus (43) ergibt sich nun, dass für $n_1, n_2 \to \infty$

$$\frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{p(1-p)(S_{1n_1} + S_{2n_2})}} \xrightarrow{\mathrm{d}} Y \sim \mathrm{N}(0,1).$$

- Hieraus ergibt sich die Behauptung (44). $\square$

Beachte

- Genauso wie im Beweis von Lemma 3.2 ergibt sich, dass für $n_1, n_2 \to \infty$

$$\frac{n_1 \overline{X}_{1n_1}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} \xrightarrow{\text{f.s.}} 1 - p\,, \qquad \frac{n_2 \overline{X}_{2n_2}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} \xrightarrow{\text{f.s.}} p\,.$$

- Mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem WR–5.11) ergibt sich somit aus Lemma 3.2, dass für jedes $\alpha \in (0,1)$

$$\lim_{n_1, n_2 \to \infty} P_{(\lambda_1, \lambda_2)}\Biggl(-z_{1-\alpha/2} \le \frac{\sqrt{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}}\,\Bigl(\dfrac{n_2 \overline{X}_{2n_2}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} - p\Bigr)}{\sqrt{\dfrac{n_1 \overline{X}_{1n_1}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} \cdot \dfrac{n_2 \overline{X}_{2n_2}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}}}} \le z_{1-\alpha/2}\Biggr) = 1 - \alpha\,.$$

- Also ist mit

$$\underline{\theta}(X_1, \ldots, X_n) = \frac{n_2 \overline{X}_{2n_2}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} - z_{1-\alpha/2} \sqrt{\frac{n_1 \overline{X}_{1n_1} \cdot n_2 \overline{X}_{2n_2}}{(n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2})^3}}\,,$$

$$\overline{\theta}(X_1, \ldots, X_n) = \frac{n_2 \overline{X}_{2n_2}}{n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2}} + z_{1-\alpha/2} \sqrt{\frac{n_1 \overline{X}_{1n_1} \cdot n_2 \overline{X}_{2n_2}}{(n_1 \overline{X}_{1n_1} + n_2 \overline{X}_{2n_2})^3}}$$

  ein asymptotisches Konfidenzintervall $\bigl(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\bigr)$ für p zum Niveau $\gamma = 1 - \alpha$ gegeben.

- Weil $0 < p < 1$ vorausgesetzt wird, ist auch $\bigl(\underline{\theta}'(X_1, \ldots, X_n), \overline{\theta}'(X_1, \ldots, X_n)\bigr)$ mit

$$\underline{\theta}'(X_1, \ldots, X_n) = \max\{0, \underline{\theta}(X_1, \ldots, X_n)\} \qquad \text{bzw.} \qquad \overline{\theta}'(X_1, \ldots, X_n) = \min\{1, \overline{\theta}(X_1, \ldots, X_n)\}$$

  ein asymptotisches Konfidenzintervall für p zum Niveau $\gamma = 1 - \alpha$.

4 Tests statistischer Hypothesen

4.1 Problemstellung und Modellbeschreibung

Genauso wie in Kapitel 1–3 nehmen wir auch in diesem Kapitel an, dass

- die Stichprobenvariablen $X_1, X_2, \ldots$ unabhängig und identisch verteilt sind und dass
- der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, über dem die Stichprobenvariablen $X_1, X_2, \ldots$ definiert sind, der kanonische Wahrscheinlichkeitsraum dieser Zufallsvariablen ist.
- Dabei prüfen wir nun Hypothesen (d.h. Vermutungen) über die Beschaffenheit der unbekannten Verteilung bzw. der Verteilungsfunktion der Stichprobenvariablen $X_1, \ldots, X_n$.

In dieser einführenden Vorlesung untersuchen wir vor allem

- *parametrische Tests*, bei denen vorausgesetzt wird, dass die Verteilung der Stichprobenvariablen $X_1, \ldots, X_n$ zu einer vorgegebenen *parametrischen Familie* von Verteilungen $\{P_\theta,\, \theta \in \Theta\}$ gehört; $\Theta \subset \mathbb{R}^m$ mit $m < \infty$.

Darüber hinaus gibt es aber auch

- *nichtparametrische Tests*, bei denen *nicht* vorausgesetzt wird, dass die Verteilung der Stichprobenvariablen $X_1, \ldots, X_n$ zu einer vorgegebenen parametrischen Familie von Verteilungen gehört, vgl. die Vorlesung „Statistik II".

Beachte

- Wir betrachten zunächst lediglich sogenannte *nichtrandomisierte Tests*.
- Die Entscheidung, ob eine Hypothese verworfen wird (bzw. ob sie nicht verworfen wird), hängt dabei nur von den beobachten Daten $x_1, \ldots, x_n$ ab, d.h. von der beobachteten Realisierung $(x_1, \ldots, x_n)$ der Zufallsstichprobe $(X_1, \ldots, X_n)$.
- Die Entscheidungsregel wird so konstruiert, dass die Wahrscheinlichkeiten von Fehlentscheidungen möglichst klein sind (bzw. vorgegebene Schwellenwerte nicht überschreiten).
- Dabei kann man ähnlich wie bei der Konstruktion von Konfidenzintervallen zu einem (vorgegebenen) Konfidenzniveau $\gamma = 1 - \alpha$ vorgehen, die in Abschnitt 3 diskutiert worden sind,
- denn bei einem Konfidenzintervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ für einen unbekannten Parameterwert $g(\theta)$ kann man α als die „Fehlerwahrscheinlichkeit" interpretieren, dass der Wert $g(\theta)$ nicht von dem Konfidenzintervall $\big(\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\big)$ überdeckt wird, d.h., dass $g(\theta)$ nicht innerhalb der Schranken $\underline{\theta}(X_1, \ldots, X_n)$ bzw. $\overline{\theta}(X_1, \ldots, X_n)$ liegt.

4.1.1 Null-Hypothese und Alternativ-Hypothese; kritischer Bereich

Sei Δ die Familie der insgesamt in Betracht gezogenen (d.h. potentiell möglichen bzw. zulässigen) Verteilungsfunktionen F der Stichprobenvariablen $X_1, \ldots, X_n$.

- Zerlegen Δ in zwei (nichtleere) Teilmengen Δ_0 und Δ_1, d.h.
$$\Delta = \Delta_0 \cup \Delta_1\,, \qquad \text{wobei} \qquad \Delta_0 \cap \Delta_1 = \emptyset\,.$$
- Betrachten die Hypothesen
$$H_0 : F \in \Delta_0 \qquad \text{bzw.} \qquad H_1 : F \in \Delta_1\,,$$
dass die unbekannte Verteilungsfunktion F der Stichprobenvariablen $X_1, \ldots, X_n$ zu der Teilmenge $\Delta_0 \subset \Delta$ bzw. $\Delta_1 \subset \Delta$ innerhalb der Familie Δ der insgesamt in Betracht gezogenen Verteilungsfunktionen gehört.

Dabei ist die folgende Sprechweise üblich:

- Die Hypothese $H_0 : F \in \Delta_0$ heißt *Nullhypothese*, während $H_1 : F \in \Delta_1$ *Alternativhypothese* genannt wird.
- Die Nullhypothese H_0 bzw. die Alternativhypothese H_1 heißen *einfach*, falls die Teilmenge Δ_0 bzw. Δ_1 nur aus einem Element besteht. Ansonsten sagt man, dass H_0 bzw. H_1 *zusammengesetzte Hypothesen* sind.

Um zwischen der Nullhypothese H_0 und der Alternativhypothese H_1 abwägen zu können, wird ein *Test*, d.h. eine *Entscheidungsregel* nach dem folgenden Prinzip konstruiert:

- Der Stichprobenraum $\mathbb{R}^n$ wird in zwei Borel-Mengen K und $K^c = \mathbb{R}^n \setminus K$ zerlegt.
- Dabei heißt $K \subset \mathbb{R}^n$ der *kritische Bereich* (d.h. der Ablehnungsbereich der Nullhypothese H_0).
- Die Nullhypothese H_0 wird verworfen (d.h. abgelehnt), falls $(x_1, \ldots, x_n) \in K$.
- Ansonsten, d.h., falls $(x_1, \ldots, x_n) \notin K$, wird H_0 nicht verworfen.

Mit anderen Worten:

- Betrachten die Stichprobenfunktion $\varphi : \mathbb{R}^n \to \{0, 1\}$ mit

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1\,, & \text{falls } (x_1, \ldots, x_n) \in K, \\ 0\,, & \text{falls } (x_1, \ldots, x_n) \notin K. \end{cases} \tag{1}$$

- Die Nullhypothese H_0 wird also verworfen, falls $\varphi(x_1, \ldots, x_n) = 1$.
- Ansonsten, d.h., falls $\varphi(x_1, \ldots, x_n) = 0$, wird H_0 nicht verworfen.

Dies führt zu der folgenden

Definition Sei $\alpha \in (0, 1)$ eine beliebige, jedoch vorgegebene Zahl. Man sagt, dass die in (1) eingeführte Stichprobenfunktion $\varphi : \mathbb{R}^n \to \{0, 1\}$ ein

- (nichtrandomisierter) Test zum *Niveau* α ist, falls

$$P_F(\varphi(X_1, \ldots, X_n) = 1) \le \alpha\,, \qquad \forall\, F \in \Delta_0\,, \tag{2}$$

  wobei P_F das kanonische Wahrscheinlichkeitsmaß, d.h. die (gemeinsame) Verteilung der Zufallsstichprobe $(X_1, \ldots, X_n)$ bezeichnet, die als Produkt-Maß aus der Verteilungsfunktion F der (einzelnen) Stichprobenvariablen $X_1, \ldots, X_n$ gebildet wird.
- Falls außerdem noch

$$P_F(\varphi(X_1, \ldots, X_n) = 1) \ge \alpha\,, \qquad \forall\, F \in \Delta_1\,, \tag{3}$$

  dann heißt $\varphi : \mathbb{R}^n \to \{0, 1\}$ ein *unverfälschter Test* zum Niveau α.

4.1.2 Fehlerwahrscheinlichkeiten

Beim Testen von Hypothesen können zwei Arten von Fehlern auftreten:

- Die Nullhypothese kann abgelehnt werden, obwohl sie „richtig" ist, und es ist möglich, dass
- die Nullhypothese nicht abgelehnt wird, obwohl sie „falsch" ist.

Dies führt zu den folgenden Begriffsbildungen.

Definition

- Für jedes $F \in \Delta_0$ heißt
$$\alpha_n(F) = P_F(\varphi(X_1, \ldots, X_n) = 1) \qquad \Big(= P_F((X_1, \ldots, X_n) \in K)\Big)$$
die *Wahrscheinlichkeit für Fehler erster Art*, und
- für jedes $F \in \Delta_1$ heißt
$$\beta_n(F) = P_F(\varphi(X_1, \ldots, X_n) = 0) \qquad \Big(= P_F((X_1, \ldots, X_n) \notin K)\Big)$$
die *Wahrscheinlichkeit für Fehler zweiter Art.*

Beachte

- Die Wahrscheinlichkeit $\alpha_n(F)$ für Fehler erster Art, ist die Wahrscheinlichkeit, dass Daten $x_1, \ldots, x_n$ beobachtet werden, die zur Ablehnung der Nullhypothese $H_0 : F \in \Delta_0$ führen, obwohl die Daten $x_1, \ldots, x_n$ gemäß der Verteilungsfunktion $F \in \Delta_0$ (beispielsweise mittels Computer-Simulation) erzeugt wurden.
- Analog hierzu ist $\beta_n(F)$ die Wahrscheinlichkeit, dass Daten $x_1, \ldots, x_n$ beobachtet werden, die *nicht* zur Ablehnung der Nullhypothese $H_0 : F \in \Delta_0$ führen, obwohl die Daten $x_1, \ldots, x_n$ gemäß der Verteilungsfunktion $F \in \Delta_1$ generiert worden sind.

Von besonderem Interesse sind Tests, deren

- Wahrscheinlichkeiten für Fehler erster Art eine vorgegebene „Irrtumswahrscheinlichkeit" α nicht überschreiten und für die gleichzeitig
- die Wahrscheinlichkeiten für Fehler zweiter Art möglichst klein sind.

4.1.3 Parametrische Tests

Falls die Familie Δ der insgesamt in Betracht gezogenen Verteilungen der Stichprobenvariablen $X_1, \ldots, X_n$ eine parametrische Familie $\Delta = \{P_\theta, \, \theta \in \Theta\}$ von Verteilungen ist, dann zerlegen wir Δ auf die folgende Weise in zwei Teilmengen Δ_0 und Δ_1:

- Betrachten eine Zerlegung
$$\Theta = \Theta_0 \cup \Theta_1 \, , \qquad \Theta_0 \cap \Theta_1 = \emptyset$$
des Parameterraumes Θ in zwei (nichtleere) disjunkte Teilmengen Θ_0, Θ_1, so dass
$$\Delta_0 = \{P_\theta, \, \theta \in \Theta_0\} \qquad \text{bzw.} \qquad \Delta_1 = \{P_\theta, \, \theta \in \Theta_1\} \, ,$$
- wobei die Hypothesen H_0 und H_1 dann wie folgt formuliert werden:
$$H_0 : \theta \in \Theta_0 \qquad \text{bzw.} \qquad H_1 : \theta \in \Theta_1 \, .$$

Beispiel Sei $\Delta = \{\mathrm{N}(\mu, \sigma^2), \, \mu \in \mathbb{R}\}$ die Familie der (eindimensionalen) Normalverteilungen mit einer vorgegebenen (bekannten) Varianz σ^2. Dann ist $H_0 : \mu = 0$ eine einfache Hypothese, während $H_1 : \mu \neq 0$ eine zusammengesetzte Hypothese ist.

Für parametrische Verteilungsfamilien können die Begriffe, die bisher in Abschnitt 4 eingeführt worden sind, wie folgt spezifiziert bzw. durch weitere Begriffe ergänzt werden.

Definition

- Im Fall einer parametrischen Familie $\Delta = \{P_\theta,\, \theta \in \Theta\}$ von Verteilungen heißt die in (1) eingeführte Stichprobenfunktion $\varphi : \mathbb{R}^n \to \{0,1\}$ ein *Parametertest zum Niveau* α, falls
$$P_\theta(\varphi(X_1,\ldots,X_n)=1) \le \alpha\,, \quad \forall\, \theta \in \Theta_0\,. \tag{4}$$
- Falls außerdem noch
$$P_\theta(\varphi(X_1,\ldots,X_n)=1) \ge \alpha\,, \quad \forall\, \theta \in \Theta_1\,, \tag{5}$$
dann heißt $\varphi : \mathbb{R}^n \to \{0,1\}$ ein *unverfälschter Test* zum Niveau α.
- Die Abbildung $\alpha_n : \Theta \to [0,1]$ mit $\alpha_n(\theta) = P_\theta(\varphi(X_1,\ldots,X_n)=1)$ heißt *Gütefunktion* des Tests φ.
- Die Einschränkung $\alpha_n : \Theta_1 \to [0,1]$ dieser Abbildung auf die Teilmenge Θ_1 des Parameterraumes heißt die *Macht* (power) von φ. Für jedes $\theta \in \Theta_1$ heißt $\alpha_n(\theta)$ die *Macht* von φ bei θ.
- Der Test φ zum Niveau α heißt *konsistent*, falls $\lim\limits_{n\to\infty} \alpha_n(\theta) = 1$ für jedes $\theta \in \Theta_1$ gilt.
- Seien $\varphi : \mathbb{R}^n \to \{0,1\}$ und $\varphi' : \mathbb{R}^n \to \{0,1\}$ zwei Parametertests zum Niveau α. Falls die Macht von φ' größer ist als die Macht von φ, d.h.
$$P_\theta(\varphi(X_1,\ldots,X_n)=1) \le P_\theta(\varphi'(X_1,\ldots,X_n)=1)\,, \quad \forall\, \theta \in \Theta_1\,, \tag{6}$$
dann sagt man, dass der Test φ' *besser* als der Test φ ist.

Beachte

1. Bei der praktischen Konstruktion des kritischen Bereiches K eines Parametertests zum Niveau α kann oft wie folgt vorgegangen werden:
   - Bestimme eine Stichprobenfunktion $T : \mathbb{R}^n \to \mathbb{R}$ (genannt *Testgröße*), so dass
   - die Zufallsvariable $T(X_1,\ldots,X_n)$ für $\theta \in \Theta_0$ eine bekannte (Prüf-) Verteilung hat, und
   - bestimme einen *Schwellenwert* $c > 0$, so dass $P_\theta\big(|T(X_1,\ldots,X_n)| > c\big) \le \alpha$ für $\theta \in \Theta_0$ gilt.
   - Dann ist mit $K = \{(x_1,\ldots,x_n) : |T(x_1,\ldots,x_n)| > c\}$ der kritische Bereich eines Parametertests zum Niveau α gegeben.
2. Um einen Test zum Niveau α mit einer möglichst großen Macht zu erhalten, ist es jedoch manchmal zweckmäßiger,
   - *zwei* Schwellenwerte $c_1, c_2 \in \mathbb{R} \cup \{\pm\infty\}$ mit $c_1 < c_2$ zu betrachten, so dass für $\theta \in \Theta_0$
   $$P_\theta\big(c_1 \le T(X_1,\ldots,X_n) \le c_2\big) \ge 1-\alpha\,.$$
   - Dann ist mit $K = \mathbb{R}^n \setminus \{(x_1,\ldots,x_n) : c_1 \le T(x_1,\ldots,x_n) \le c_2\}$ der kritische Bereich eines (asymmetrischen) Parametertests zum Niveau α gegeben.

In den folgenden Abschnitten wird dieses Konstruktionsprinzip anhand einer Reihe von Beispielen näher diskutiert.

4.2 Parametertests bei Normalverteilung

Wir nehmen in diesem Abschnitt an, dass die Stichprobenvariablen $X_1,\ldots,X_n$ normalverteilt sind, d.h., es gelte $\Delta \subset \{\mathrm{N}(\mu,\sigma^2),\, \mu \in \mathbb{R}, \sigma^2 > 0\}$. Wir betrachten also die Familie der (eindimensionalen) Normalverteilungen bzw. Teilmengen hiervon.

4.2.1 Test des Erwartungswertes

Wir konstruieren zunächst Tests für den Erwartungswert μ. Dabei gibt es Ähnlichkeiten zur Konstruktion der Konfidenzintervalle für μ, die in Abschnitt 3.2.1 diskutiert worden sind.

Test des Erwartungswertes μ (bei bekannter Varianz σ^2)

- Wir nehmen an, dass $X_i \sim \mathrm{N}(\mu, \sigma^2)$ für ein (unbekanntes) $\mu \in \mathbb{R}$ und ein (bekanntes) $\sigma^2 > 0$.
- Für einen vorgegebenen (hypothetischen) Zahlenwert $\mu_0 \in \mathbb{R}$ soll die Hypothese $H_0 : \mu = \mu_0$ (gegen die Alternative $H_1 : \mu \neq \mu_0$) getestet werden, d.h. $\Theta = \{\mu : \mu \in \mathbb{R}\}$ mit
$$\Theta_0 = \{\mu_0\} \qquad \text{bzw.} \qquad \Theta_1 = \{\mu : \mu \in \mathbb{R}, \mu \neq \mu_0\}$$
- Wir betrachten die Testgröße $T : \mathbb{R}^n \to \mathbb{R}$ mit
$$T(x_1, \ldots, x_n) = \sqrt{n}\, \frac{\overline{x}_n - \mu_0}{\sigma} \tag{7}$$
- und den Schwellenwert $c = z_{1-\alpha/2}$, d.h. das $(1-\alpha/2)$-Quantil der Standardnormalverteilung.
- Weil $T(X_1, \ldots, X_n) \sim \mathrm{N}(0,1)$, gilt dann $P_{\mu_0}\big(|T(X_1, \ldots, X_n)| > z_{1-\alpha/2}\big) = \alpha$.
- Die Hypothese $H_0 : \mu = \mu_0$ wird abgelehnt, falls $|T(x_1, \ldots, x_n)| > z_{1-\alpha/2}$.

Für $\alpha = 0.05$ und $\sigma = 0.10$ wollen wir nun prüfen, ob

- die Hypothese $H_0 : \mu = 42.0$ mit der bereits in Abschnitt 3.2.1 betrachteten (konkreten) Stichprobe
$$(x_1, \ldots, x_{10}) = (41.60,\ 41.48,\ 42.34,\ 41.95,\ 41.86,\ 42.18,\ 41.72,\ 42.26,\ 41.81,\ 42.04)$$
vereinbar ist.
- In diesem Fall gilt $\overline{x}_{10} = 41.924$, und aus (7) ergibt sich somit
$$\begin{aligned} \big|T(x_1, \ldots, x_{10})\big| &= \Big|\sqrt{10}\, \frac{\overline{x}_{10} - \mu_0}{\sigma}\Big| \\ &= \Big|\sqrt{10}\, \frac{41.924 - 42.0}{0.10}\Big| \\ &= 2.40 > z_{0.975} = 1.96\,, \end{aligned}$$
wobei das Quantil $z_{0.975} = 1.96$ aus Tabelle 1 entnommen wurde.
- Die Hypothese $H_0 : \mu = 42.0$ wird also verworfen.

Beachte

1. Wir betrachten die Gütefunktion $\alpha_n : \mathbb{R} \to [0,1]$ dieses Tests.
   - Für beliebige $n \in \mathbb{N}$ und $\mu \in \mathbb{R}$ gilt
$$\begin{aligned} \alpha_n(\mu) &= P_\mu\big(|T(X_1, \ldots, X_n)| > z_{1-\alpha/2}\big) \\ &= P_\mu\Big(\Big|\sqrt{n}\, \frac{\overline{X}_n - \mu}{\sigma} - \sqrt{n}\, \frac{\mu_0 - \mu}{\sigma}\Big| > z_{1-\alpha/2}\Big) \\ &= 1 - P_\mu\Big(-z_{1-\alpha/2} + \sqrt{n}\, \frac{\mu_0 - \mu}{\sigma} \le \sqrt{n}\, \frac{\overline{X}_n - \mu}{\sigma} \le z_{1-\alpha/2} + \sqrt{n}\, \frac{\mu_0 - \mu}{\sigma}\Big) \\ &= 1 - \Phi\Big(z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\, \sqrt{n}\Big) + \Phi\Big(-z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\, \sqrt{n}\Big) \\ &= \Phi\Big(-z_{1-\alpha/2} - \frac{\mu_0 - \mu}{\sigma}\, \sqrt{n}\Big) + \Phi\Big(-z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\, \sqrt{n}\Big)\,. \end{aligned}$$

- Hieraus ergibt sich, dass die Ableitung $\alpha_n^{(1)}(\mu)$ für jedes $\mu > \mu_0$ positiv ist und dass $\alpha_n(\mu_0 - x) = \alpha_n(\mu_0 + x)$ für jedes $x \in \mathbb{R}$.
- Somit gilt $\alpha_n(\mu) \geq \alpha$ für beliebige $n \in \mathbb{N}$ und $\mu \in \mathbb{R}$, und $\lim_{n\to\infty} \alpha_n(\mu) = 1$ für jedes $\mu \neq \mu_0$.
- Der Test ist also unverfälscht und konsistent.

2. Falls die Hypothese $H_0 : \mu = \mu_0$ gegen die (einseitige) Alternative $H_1 : \mu > \mu_0$ getestet werden soll, d.h. $\Theta = \{\mu : \mu \in \mathbb{R}, \mu \geq \mu_0\}$ mit

$$\Theta_0 = \{\mu_0\} \qquad \text{bzw.} \qquad \Theta_1 = \{\mu : \mu \in \mathbb{R}, \mu > \mu_0\},$$

dann könnte man zwar so wie bisher vorgehen und für die in (7) definierte Testgröße T

- den kritischen Bereich $K = \{(x_1, \ldots, x_n) : |T(x_1, \ldots, x_n)| > z_{1-\alpha/2}\}$ betrachten.
- Ein besserer Test ergibt sich jedoch, wenn der folgende (einseitige) kritische Bereich

$$K' = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > z_{1-\alpha}\}$$

betrachtet wird, denn es gilt

$$\alpha_n(\mu) \leq \alpha'_n(\mu), \qquad \forall\, \mu \geq \mu_0, \tag{8}$$

wobei $\alpha_n(\mu)$ bzw. $\alpha'_n(\mu)$ die Macht des Tests mit dem kritischen Bereich K bzw. K' bezeichnet, d.h.

$$\alpha_n(\mu) = 1 - \Phi\Bigl(z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n}\Bigr) + \Phi\Bigl(-z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n}\Bigr)$$

bzw.

$$\alpha'_n(\mu) = 1 - \Phi\Bigl(z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n}\Bigr).$$

- Um die Gültigkeit der Ungleichung (8) einzusehen, betrachten wir die Funktion $g : [\mu_0, \infty) \to \mathbb{R}$ mit $g(\mu) = \alpha'_n(\mu) - \alpha_n(\mu)$.
  - Offenbar gilt $g(\mu_0) = 0$ und $\lim_{\mu\to\infty} g(\mu) = 0$.
  - Außerdem gilt $g^{(1)}(\mu_0) > 0$ sowie $\lim_{\mu\to\infty} g^{(1)}(\mu) = 0$ gilt, wobei $g^{(1)}$ die Ableitung von g bezeichnet.
  - Darüber hinaus kann man sich leicht davon überzeugen, dass $g^{(1)}(\mu) = 0$ für genau ein $\mu > \mu_0$.
  - Hieraus folgt, dass $g(\mu) \geq 0$ für jedes $\mu \geq \mu_0$.

3. Falls die Hypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu < \mu_0$ getestet werden soll, dann wird der kritische Bereich $K'' = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) < -z_{1-\alpha}\}$ betrachtet.

Test des Erwartungswertes μ (bei unbekannter Varianz σ^2)

- Im Unterschied zu dem vorhergehenden Fall nehmen wir nun an, dass $X_i \sim \mathrm{N}(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ *und* $\sigma^2 > 0$ unbekannt seien.
- Für einen vorgegebenen (hypothetischen) Zahlenwert $\mu_0 \in \mathbb{R}$ soll erneut die Hypothese $H_0 : \mu = \mu_0$ (gegen die zweiseitige Alternative $H_1 : \mu \neq \mu_0$) getestet werden.
- Dies bedeutet, dass jetzt $\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ mit

$$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu, \sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\},$$

d.h., sowohl H_0 als auch H_1 sind jetzt zusammengesetzte Hypothesen.

- Damit hängt auch zusammen, dass die in (7) betrachtete Testgröße T nun nicht mehr geeignet ist, weil in (7) die unbekannte Größe σ vorkommt.
- Wir betrachten vielmehr die Testgröße $T : \mathbb{R}^n \to \mathbb{R}$ mit

$$T(x_1, \ldots, x_n) = \sqrt{n}\,\frac{\overline{x}_n - \mu_0}{s_n} \tag{9}$$

- und den Schwellenwert $c = t_{n-1,1-\alpha/2}$, d.h. das $(1-\alpha/2)$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden.
- Aus Theorem 1.13 ergibt sich dann, dass $P_{(\mu_0,\sigma^2)}\bigl(|T(X_1,\ldots,X_n)| > t_{n-1,1-\alpha/2}\bigr) = \alpha$ für jedes $\sigma^2 > 0$ gilt.
- Die Hypothese $H_0 : \mu = \mu_0$ wird abgelehnt, falls $|T(x_1,\ldots,x_n)| > t_{n-1,1-\alpha/2}$.

Für $\alpha = 0.05$ wollen wir nun prüfen, ob

- die Hypothese $H_0 : \mu = 42.0$ mit der schon vorher betrachteten (konkreten) Stichprobe

  $$(x_1,\ldots,x_{10}) = (41.60,\ 41.48,\ 42.34,\ 41.95,\ 41.86,\ 42.18,\ 41.72,\ 42.26,\ 41.81,\ 42.04)$$

  vereinbar ist, wobei jetzt jedoch angenommen wird, dass die Varianz σ^2 unbekannt ist.
- Es gilt $\overline{x}_{10} = 41.924$ und $s_{10}^2 = 0.0807$ bzw. $s_{10} = 0.284$.
- Aus (9) ergibt sich somit, dass

  $$\begin{aligned} \bigl|T(x_1,\ldots,x_{10})\bigr| &= \left|\sqrt{10}\,\frac{\overline{x}_{10}-\mu_0}{s_{10}}\right| \\ &= \left|\sqrt{10}\,\frac{41.924-42.0}{0.284}\right| \\ &= 0.846 < t_{9,0.975} = 2.262\,, \end{aligned}$$

  wobei das Quantil $t_{9,0.975} = 2.262$ aus Tabelle 3 entnommen wurde.
- Unter den jetzt gemachten Modellannahmen wird also die Hypothese $H_0 : \mu = 42.0$ nicht verworfen.
- Der Hauptgrund hierfür ist, dass die Stichprobenvarianz $s_{10}^2 = 0.0807$ deutlich größer ist als der bei dem vorhergehenden Beispiel betrachtete Wert $\sigma^2 = 0.01$ der (z.B. aufgrund von Vorinformationen) als bekannt angenommenen Varianz σ^2.
- Mit anderen Worten: Wegen der relativ großen Stichprobenvarianz s_{10}^2 wird jetzt die Abweichung $0.076 = |41.924 - 42.0|$ des Stichprobenmittels $\overline{x}_{10} = 41.924$ von dem hypothetischen Erwartungswert $\mu_0 = 42.0$ toleriert, während sie bei dem vorhergehenden Beispiel (vor allem) wegen der kleineren Varianz $\sigma^2 = 0.01$ nicht toleriert wurde.

Beachte

1. Falls die Hypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ getestet werden soll, dann wird (ähnlich wie in dem vorhergehenden Beispiel) für die in (9) definierte Testgröße T der kritische Bereich K' betrachtet mit
   $$K' = \{(x_1,\ldots,x_n) : T(x_1,\ldots,x_n) > t_{n-1,1-\alpha}\}\,.$$
2. Analog wird für den Test der Hypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu < \mu_0$ der kritische Bereich K'' betrachtet mit
   $$K'' = \{(x_1,\ldots,x_n) : T(x_1,\ldots,x_n) < -t_{n-1,1-\alpha}\}\,.$$

4.2.2 Tests der Varianz

Wenn die Varianz σ^2 der normalverteilten Stichprobenvariablen $X_1, \ldots, X_n$ getestet werden soll, dann kann man ähnlich wie in Abschnitt 4.2.1 vorgehen. Außerdem gibt es Ähnlichkeiten zur Konstruktion der Konfidenzintervalle für σ^2, die in Abschnitt 3.2.2 diskutiert worden sind.

Test der Varianz σ^2 (bei bekanntem Erwartungswert μ)

- Wir nehmen an, dass $X_i \sim \mathrm{N}(\mu, \sigma^2)$ für ein (bekanntes) $\mu \in \mathbb{R}$ und ein (unbekanntes) $\sigma^2 > 0$.
- Für einen vorgegebenen (hypothetischen) Zahlenwert $\sigma_0^2 > 0$ soll die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ (gegen die Alternative $H_1 : \sigma^2 \neq \sigma_0^2$) getestet werden, d.h. $\Theta = \{\sigma^2 : \sigma^2 > 0\}$ mit
$$\Theta_0 = \{\sigma_0^2\} \qquad \text{bzw.} \qquad \Theta_1 = \{\sigma^2 : \sigma^2 \neq \sigma_0^2\}\,.$$
- Wir betrachten die Testgröße $T : \mathbb{R}^n \to [0, \infty)$ mit
$$T(x_1, \ldots, x_n) = \frac{n\widetilde{s}_n^2}{\sigma_0^2} \tag{10}$$
und die Schwellenwerte $c_1 = \chi^2_{n,\alpha/2}$ bzw. $c_2 = \chi^2_{n,1-\alpha/2}$. Dabei ist $\widetilde{s}_n^2$ die bereits in Abschnitt 3.2.2 eingeführte Größe
$$\widetilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2\,.$$
- Weil $T(X_1, \ldots, X_n) \sim \chi^2_n$, gilt dann $P_{\sigma_0^2}\big(\chi^2_{n,\alpha/2} \le T(X_1, \ldots, X_n) \le \chi^2_{n,1-\alpha/2}\big) = 1 - \alpha$.
- Die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ wird abgelehnt, falls $T(x_1, \ldots, x_n) < \chi^2_{n,\alpha/2}$ oder $T(x_1, \ldots, x_n) > \chi^2_{n,1-\alpha/2}$.

Beachte

1. Die Gütefunktion $\alpha_n : \Theta \to [0, 1]$ dieses Tests mit
$$\begin{aligned}\alpha_n(\sigma^2) &= 1 - P_{\sigma^2}\big(\chi^2_{n,\alpha/2} \le T(X_1, \ldots, X_n) \le \chi^2_{n,1-\alpha/2}\big)\\ &= 1 - P_{\sigma^2}\Big(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,\alpha/2} \le \frac{n\widetilde{S}_n^2}{\sigma^2} \le \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha/2}\Big)\end{aligned}$$
hat kein Minimum im Punkt $\sigma^2 = \sigma_0^2$,
   - weil $n\widetilde{S}_n^2/\sigma^2 \sim \chi^2_n$ gilt, d.h., die Verteilung der Zufallsvariablen $n\widetilde{S}_n^2/\sigma^2$ hängt nicht von σ^2 ab,
   - und weil die Funktionswerte $f_n(\chi^2_{n,\alpha/2})$ und $f_n(\chi^2_{n,1-\alpha/2})$ der in Theorem 1.6 gegebenen Dichte f_n der χ^2_n-Verteilung nicht übereinstimmen.

   Der Test ist also *nicht* unverfälscht.
2. Wenn jedoch beispielsweise die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ gegen die (einseitige) Alternative $H_1 : \sigma^2 > \sigma_0^2$ getestet werden soll, d.h. $\Theta = \{\sigma^2 : \sigma^2 \ge \sigma_0^2\}$ mit
$$\Theta_0 = \{\sigma_0^2\} \qquad \text{bzw.} \qquad \Theta_1 = \{\sigma^2 : \sigma^2 > \sigma_0^2\}\,,$$
dann wird der kritische Bereich
$$K' = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > \chi^2_{n,1-\alpha}\} \tag{11}$$
betrachtet, und die Gütefunktion $\alpha_n : \Theta \to [0, 1]$ mit
$$\begin{aligned}\alpha_n(\sigma^2) &= P_{\sigma^2}\big(T(X_1, \ldots, X_n) > \chi^2_{n,1-\alpha}\big)\\ &= P_{\sigma^2}\Big(\frac{n\widetilde{S}_n^2}{\sigma^2} > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha}\Big)\end{aligned}$$
ist monoton wachsend für $\sigma^2 \ge \sigma_0^2$. Der einseitige Test ist somit unverfälscht.

3. Wegen dieser Monotonieeigenschaft ist durch den in (11) betrachteten kritischen Bereich K' auch ein (unverfälschter) Test zum Niveau α der Hypothese $H_0 : \sigma^2 \le \sigma_0^2$ gegen die Alternativhypothese $H_1 : \sigma^2 > \sigma_0^2$ gegeben.
4. Analog liefert der kritische Bereich
$$K'' = \{(x_1,\ldots,x_n) : T(x_1,\ldots,x_n) < \chi^2_{n,\alpha}\}$$
einen (unverfälschten) Test zum Niveau α der Hypothesen $H_0 : \sigma^2 = \sigma_0^2$ bzw. $H_0 : \sigma^2 \ge \sigma_0^2$ gegen die Alternativhypothese $H_1 : \sigma^2 < \sigma_0^2$.

Test der Varianz σ^2 (bei unbekanntem Erwartungswert μ)

- Wir nehmen nun an, dass $X_i \sim \mathrm{N}(\mu,\sigma^2)$, wobei $\mu \in \mathbb{R}$ *und* $\sigma^2 > 0$ unbekannt seien.
- Für einen vorgegebenen (hypothetischen) Zahlenwert $\sigma_0^2 \in \mathbb{R}$ soll die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ (gegen die Alternative $H_1 : \sigma^2 \ne \sigma_0^2$) getestet werden, d.h. $\Theta = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ mit
$$\Theta_0 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 \ne \sigma_0^2\}\,.$$
- Wir betrachten die Testgröße $T : \mathbb{R}^n \to [0,\infty)$ mit
$$T(x_1,\ldots,x_n) = \frac{(n-1)s_n^2}{\sigma_0^2} \tag{12}$$
und die Schwellenwerte $c_1 = \chi^2_{n-1,\alpha/2}$ bzw. $c_2 = \chi^2_{n-1,1-\alpha/2}$.
- Aus Theorem 1.11 ergibt sich, dass $T(X_1,\ldots,X_n) \sim \chi^2_{n-1}$, und somit gilt
$$P_{(\mu,\sigma_0^2)}\bigl(\chi^2_{n-1,\alpha/2} \le T(X_1,\ldots,X_n) \le \chi^2_{n-1,1-\alpha/2}\bigr) = 1-\alpha$$
für jedes $\mu \in \mathbb{R}$.
- Die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ wird abgelehnt, falls
$$T(x_1,\ldots,x_n) < \chi^2_{n-1,\alpha/2} \qquad \text{oder} \qquad T(x_1,\ldots,x_n) > \chi^2_{n-1,1-\alpha/2}\,.$$

Beachte

1. Die Gütefunktion $\alpha_n : \Theta \to [0,1]$ dieses Tests mit
$$\alpha_n(\mu,\sigma^2) = 1 - P_{(\mu,\sigma^2)}\bigl(\chi^2_{n-1,\alpha/2} \le T(X_1,\ldots,X_n) \le \chi^2_{n-1,1-\alpha/2}\bigr)$$
hängt nicht von μ ab. Sie hat jedoch (bei fixiertem μ) kein Minimum im Punkt $\sigma^2 = \sigma_0^2$, wobei dies genauso wie im vorhergehenden Fall (μ bekannt) begründet werden kann. Der Test ist also *nicht* unverfälscht.
2. Wenn jedoch beispielsweise die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ gegen die (einseitige) Alternative $H_1 : \sigma^2 > \sigma_0^2$ getestet werden soll, d.h. $\Theta = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 \ge \sigma_0^2\}$ mit
$$\Theta_0 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > \sigma_0^2\}\,,$$
dann wird der kritische Bereich
$$K' = \{(x_1,\ldots,x_n) : T(x_1,\ldots,x_n) > \chi^2_{n-1,1-\alpha}\} \tag{13}$$
betrachtet, und die Gütefunktion $\alpha_n : \Theta \to [0,1]$ mit
$$\begin{aligned}\alpha_n(\mu,\sigma^2) &= P_{(\mu,\sigma^2)}\bigl(T(X_1,\ldots,X_n) > \chi^2_{n-1,1-\alpha}\bigr)\\ &= P_{(\mu,\sigma^2)}\Bigl(\frac{(n-1)S_n^2}{\sigma^2} > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,1-\alpha}\Bigr)\end{aligned}$$
hängt nicht von μ ab und ist monoton wachsend für $\sigma \ge \sigma_0^2$. Der einseitige Test ist somit unverfälscht.

3. Wegen dieser Monotonieeigenschaft ist durch den in (13) betrachteten kritischen Bereich K' auch ein (unverfälschter) Test zum Niveau α der Hypothese $H_0 : \sigma^2 \leq \sigma_0^2$ gegen die Alternativhypothese $H_1 : \sigma^2 > \sigma_0^2$ gegeben.
4. Analog liefert der kritische Bereich

$$K'' = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) < \chi^2_{n-1,\alpha}\}$$

einen (unverfälschten) Test zum Niveau α der Hypothesen $H_0 : \sigma^2 = \sigma_0^2$ bzw. $H_0 : \sigma^2 \geq \sigma_0^2$ gegen die Alternativhypothese $H_1 : \sigma^2 < \sigma_0^2$.

4.3 Zwei-Stichproben-Tests

So wie in Abschnitt 3.3 nehmen wir nun an, dass die Stichprobenvariablen $X_1, X_2, \ldots$ (zweidimensionale) normalverteilte Zufallsvektoren sind mit $X_i = (X_{1i}, X_{2i})$. Die Realisierungen von X_i bezeichnen wir mit $x_i = (x_{1i}, x_{2i})$ für $i = 1, 2, \ldots$.

Dabei setzen wir zunächst voraus, dass die Komponenten X_{1i} und X_{2i} von X_i unabhängige (jedoch im allgemeinen nicht identisch verteilte) Zufallsvariablen sind.

Für zwei beliebige, jedoch vorgegebene Zahlen $n_1, n_2 \in \mathbb{N}$ betrachten wir also zwei unabhängige Zufallsstichproben $(X_{11}, \ldots, X_{1n_1})$ und $(X_{21}, \ldots, X_{2n_2})$, wobei wir

- die beobachteten Daten $(x_{11}, \ldots, x_{1n_1})$ und $(x_{21}, \ldots, x_{2n_2})$ als Realisierungen von $(X_{11}, \ldots, X_{1n_1})$ bzw. $(X_{21}, \ldots, X_{2n_2})$ auffassen und
- zur Vereinfachung der Schreibweise $n = \max\{n_1, n_2\}$ setzen.

Wir diskutieren in diesem Abschnitt nur Beispiele von zweiseitigen Tests. Wie in Abschnitt 4.2 gezeigt wurde, lassen sich dann leicht die entsprechenden einseitigen Tests konstruieren.

4.3.1 Tests der Gleichheit von Erwartungswerten

1. *Bekannte Varianzen*
   - Wir nehmen an, dass $X_{1i} \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_{2i} \sim \mathrm{N}(\mu_2, \sigma_2^2)$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$ und (bekannte) $\sigma_1^2, \sigma_2^2 > 0$.
   - Dabei soll die Hypothese $H_0 : \mu_1 = \mu_2$ (gegen die Alternative $H_1 : \mu_1 \neq \mu_2$) getestet werden, d.h. $\Theta = \{(\mu_1, \mu_2) : \mu_1, \mu_2 \in \mathbb{R}\}$ mit

   $$\Theta_0 = \{(\mu_1, \mu_2) : \mu_1 = \mu_2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu_1, \mu_2) : \mu_1 \neq \mu_2\}$$

   - Wir betrachten die Testgröße $T : \mathbb{R}^{2n} \to \mathbb{R}$ mit

   $$T(x_1, \ldots, x_n) = \frac{\overline{x}_{1n_1} - \overline{x}_{2n_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{14}$$

   und den Schwellenwert $c = z_{1-\alpha/2}$.
   - Weil $\overline{X}_{1n_1} \sim \mathrm{N}(\mu_1, \sigma_1^2/n_1)$ bzw. $\overline{X}_{2n_2} \sim \mathrm{N}(\mu_2, \sigma_2^2/n_2)$, gilt dann $T(X_1, \ldots, X_n) \sim \mathrm{N}(0,1)$ und somit

   $$P_{(\mu_1,\mu_2)}\big(|T(X_1, \ldots, X_n)| > z_{1-\alpha/2}\big) = \alpha$$

   für alle $\mu_1, \mu_2 \in \mathbb{R}$ mit $\mu_1 = \mu_2$.

- Die Hypothese $H_0 : \mu_1 = \mu_2$ wird abgelehnt, falls $|T(x_1, \ldots, x_n)| > z_{1-\alpha/2}$.

2. *Gleiche, jedoch unbekannte Varianzen*

- Wir nehmen an, dass $X_{1i} \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_{2i} \sim \mathrm{N}(\mu_2, \sigma_2^2)$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$ und (unbekannte) $\sigma_1^2, \sigma_2^2 > 0$ mit $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
- Dabei soll erneut die Hypothese $H_0 : \mu_1 = \mu_2$ (gegen die Alternative $H_1 : \mu_1 \neq \mu_2$) getestet werden, d.h. $\Theta = \{(\mu_1, \mu_2, \sigma^2) : \mu_1, \mu_2 \in \mathbb{R}, \sigma^2 > 0\}$ mit
$$\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) : \mu_1 = \mu_2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu_1, \mu_2, \sigma^2) : \mu_1 \neq \mu_2\}$$
- Wir betrachten die Testgröße $T : \mathbb{R}^{2n} \to \mathbb{R}$ mit
$$T(x_1, \ldots, x_n) = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \; \frac{\overline{x}_{1n_1} - \overline{x}_{2n_2}}{s_{n_1 n_2}}\,, \tag{15}$$
wobei
$$s_{n_1 n_2}^2 = \frac{1}{n_1 + n_2 - 2} \Bigl(\sum_{i=1}^{n_1} (x_{1i} - \overline{x}_{1n_1})^2 + \sum_{i=1}^{n_2} (x_{2i} - \overline{x}_{2n_2})^2 \Bigr)\,.$$
- Man kann zeigen, dass dann $T(X_1, \ldots, X_n) \sim \mathrm{t}_{n_1+n_2-2}$ für $\mu_1 = \mu_2$.
- Mit dem Schwellenwert $c = t_{n_1+n_2-2,1-\alpha/2}$ ergibt sich also
$$P_{(\mu_1,\mu_2,\sigma^2)}\bigl(|T(X_1, \ldots, X_n)| > t_{n_1+n_2-2,1-\alpha/2}\bigr) = \alpha$$
für alle $\mu_1, \mu_2 \in \mathbb{R}$ mit $\mu_1 = \mu_2$ und für alle $\sigma^2 > 0$.
- Die Hypothese $H_0 : \mu_1 = \mu_2$ wird somit abgelehnt, falls $|T(x_1, \ldots, x_n)| > t_{n_1+n_2-2,1-\alpha/2}$.

3. *Verbundene Stichproben*

Wir testen nun die Gleichheit von Erwartungswerten verbundener Stichproben, vgl. auch Abschnitt 3.3.3. Dabei setzen wir voraus, dass $n_1 = n_2$ $(= n)$. Die Zufallstichproben $(X_{11}, \ldots, X_{1n})$ und $(X_{21}, \ldots, X_{2n})$ müssen jedoch jetzt *nicht* unabhängig sein.

- Wir nehmen an, dass $\mathbb{E}\,X_{1i} = \mu_1$ und $\mathbb{E}\,X_{2i} = \mu_2$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$.
- Außerdem seien die Differenzen $Y_1, \ldots, Y_n$ mit $Y_i = X_{1i} - X_{2i}$ unabhängige und identisch (normal-) verteilte Zufallsvariablen mit $Y_i \sim \mathrm{N}(\mu_1 - \mu_2, \sigma^2)$ für ein (unbekanntes) $\sigma^2 > 0$.
- Es soll die Hypothese $H_0 : \mu_1 = \mu_2$ (gegen die Alternative $H_1 : \mu_1 \neq \mu_2$) getestet werden, d.h. $\Theta = \{(\mu_1, \mu_2, \sigma^2) : \mu_1, \mu_2 \in \mathbb{R}, \sigma^2 > 0\}$ mit
$$\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) : \mu_1 = \mu_2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu_1, \mu_2, \sigma^2) : \mu_1 \neq \mu_2\}$$
- Wir betrachten die Testgröße $T : \mathbb{R}^{2n} \to \mathbb{R}$ mit
$$T(x_1, \ldots, x_n) = \sqrt{n}\; \frac{\overline{y}_n}{s_{y,n}}\,, \tag{16}$$
wobei $\overline{y}_n = \frac{1}{n} \sum\limits_{i=1}^{n} y_i$ und $s_{y,n}^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (y_i - \overline{y}_n)^2$.
- Dann gilt $T(X_1, \ldots, X_n) \sim \mathrm{t}_{n-1}$ für $\mu_1 = \mu_2$.
- Mit dem Schwellenwert $c = t_{n-1,1-\alpha/2}$ ergibt sich also
$$P_{(\mu_1,\mu_2,\sigma^2)}\bigl(|T(X_1, \ldots, X_n)| > t_{n-1,1-\alpha/2}\bigr) = \alpha$$
für alle $\mu_1, \mu_2 \in \mathbb{R}$ mit $\mu_1 = \mu_2$ und für alle $\sigma^2 > 0$.
- Die Hypothese $H_0 : \mu_1 = \mu_2$ wird somit abgelehnt, falls $|T(x_1, \ldots, x_n)| > t_{n-1,1-\alpha/2}$.

4.3.2 Test der Gleichheit von Varianzen

In diesem Abschnitt testen wir die Gleichheit zweier Varianzen bei unbekannten Erwartungswerten.

- Wir nehmen an, dass $X_{1i} \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_{2i} \sim \mathrm{N}(\mu_2, \sigma_2^2)$ für (unbekannte) $\mu_1, \mu_2 \in \mathbb{R}$ und (unbekannte) $\sigma_1^2, \sigma_2^2 > 0$.
- Dabei soll die Hypothese $H_0 : \sigma_1^2 = \sigma_2^2$ (gegen die Alternative $H_1 : \sigma_1^2 \neq \sigma_2^2$) getestet werden, d.h. $\Theta = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : \mu_1, \mu_2 \in \mathbb{R},\ \sigma_1^2, \sigma_2^2 > 0\}$ mit

$$\Theta_0 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : \sigma_1^2 = \sigma_2^2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : \sigma_1^2 \neq \sigma_2^2\}$$

- Wir betrachten die Testgröße $T : \mathbb{R}^{2n} \to [0, \infty)$ mit

$$T(x_1, \ldots, x_n) = \frac{s_{2n_2}^2}{s_{1n_1}^2}\,, \tag{17}$$

  wobei

$$s_{i\,n_i}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_{in_i})^2 \qquad i = 1, 2\,.$$

- Für $\sigma_1^2 = \sigma_2^2$ gilt dann

$$T(X_1, \ldots, X_n) \sim \mathrm{F}_{n_2-1, n_1-1}\,.$$

- Mit den Schwellenwerten $c_1 = F_{n_2-1, n_1-1, \alpha/2}$ und $c_2 = F_{n_2-1, n_1-1, 1-\alpha/2}$ gilt also

$$P_{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)}\bigl(F_{n_2-1, n_1-1, \alpha/2} \le T(X_1, \ldots, X_n) \le F_{n_2-1, n_1-1, 1-\alpha/2}\bigr) = 1 - \alpha$$

  für alle $\mu_1, \mu_2 \in \mathbb{R}$ und für alle $\sigma_1^2, \sigma_2^2 > 0$ mit $\sigma_1^2 = \sigma_2^2$.
- Die Hypothese $H_0 : \sigma_1^2 = \sigma_2^2$ wird somit abgelehnt, falls

$$T(x_1, \ldots, x_n) < F_{n_2-1, n_1-1, \alpha/2} \qquad \text{oder} \qquad T(x_1, \ldots, x_n) > F_{n_2-1, n_1-1, 1-\alpha/2}\,.$$

4.4 Asymptotische Parametertests

4.4.1 Ein-Stichproben-Probleme

Wir kehren zunächst zur Betrachtung des *Ein-Stichproben-Falles* zurück. D.h., wir nehmen an, dass

- *ein* Datensatz $(x_1, \ldots, x_n)$ beobachtet wird, den wir als Realisierung *einer* Zufallsstichprobe $(X_1, \ldots, X_n)$ auffassen.
- Dabei setzen wir jetzt nicht mehr voraus, dass die Stichprobenvariablen $X_1, X_2, \ldots$ normalverteilt sind, sondern lediglich, dass
- die Verteilung von X_i zu einer (beliebigen) parametrischen Familie $\Delta = \{P_\theta,\ \theta \in \Theta\}$ von Verteilungen gehört.

So wie bisher in Abschnitt 4 betrachten wir

- eine Zerlegung $\Theta = \Theta_0 \cup \Theta_1$ des Parameterraumes in zwei (nichtleere) disjunkte Teilmengen Θ_0, Θ_1 und
- die zugehörigen Hypothesen $H_0 : \theta \in \Theta_0$ bzw. $H_1 : \theta \in \Theta_1$.

- Außerdem betrachten wir für jedes $n = 1, 2, \ldots$ eine Borel-Menge $K_n \subset \mathbb{R}^n$ und
- eine Stichprobenfunktion $\varphi_n : \mathbb{R}^n \to \{0,1\}$ mit

$$\varphi_n(x_1,\ldots,x_n) = \begin{cases} 1\,, & \text{falls } (x_1,\ldots,x_n) \in K_n, \\ 0\,, & \text{falls } (x_1,\ldots,x_n) \not\in K_n. \end{cases}$$

Definition Sei $\alpha \in (0,1)$ ein beliebige (vorgegebene) Zahl. Dann sagt man, dass durch die Folge von Stichprobenfunktionen $\{\varphi_n\}$ ein *asymptotischer Parametertest* zum Niveau α gegeben ist, falls

$$\lim_{n\to\infty} P_\theta(\varphi_n(X_1,\ldots,X_n) = 1) \le \alpha$$

für jedes $\theta \in \Theta_0$ gilt.

Ähnlich wie bei der Konstruktion asymptotischer Konfidenzintervalle in Abschnitt 3.4 vorgehend, konstruieren wir nun asymptotische Tests bei Poisson- bzw. Bernoulli-Verteilung.

So wie in Abschnitt 4.3 diskutieren wir lediglich Beispiele von (symmetrischen) zweiseitigen asymptotischen Tests. Auf die gleiche Weise lassen sich dann entsprechende asymmetrische bzw. einseitige asymptotische Tests konstruieren.

1. *Test des Erwartungswertes λ bei Poisson-Verteilung*

   - Wir nehmen an, dass $X_i \sim \mathrm{Poi}(\lambda)$ für ein (unbekanntes) $\lambda > 0$.
   - Für einen vorgegebenen (hypothetischen) Zahlenwert $\lambda_0 > 0$ soll die Hypothese $H_0 : \lambda = \lambda_0$ (gegen die Alternative $H_1 : \lambda \ne \lambda_0$) getestet werden, d.h. $\Theta = \{\lambda : \lambda > 0\}$ mit

   $$\Theta_0 = \{\lambda_0\} \qquad \text{bzw.} \qquad \Theta_1 = \{\lambda : \lambda \ne \lambda_0\}$$

   - Wir betrachten die Testgröße $T_n : \mathbb{R}^n \to \mathbb{R}$ mit

   $$T_n(x_1,\ldots,x_n) = \begin{cases} \sqrt{n}\,\dfrac{\overline{x}_n - \lambda_0}{\sqrt{\overline{x}_n}}\,, & \text{falls } \overline{x}_n > 0, \\ 0\,, & \text{falls } \overline{x}_n = 0, \end{cases} \tag{18}$$

   und den Schwellenwert $c = z_{1-\alpha/2}$.
   - Dann gilt (vgl. Beispiel 1 des Abschnittes 3.4.1)

   $$\lim_{n\to\infty} P_{\lambda_0}\big(|T_n(X_1,\ldots,X_n)| > z_{1-\alpha/2}\big) = \alpha\,.$$

   - Durch die in (18) eingeführte Folge $\{T_n\}$ von Testfunktionen ist also ein asymptotischer Test des Erwartungswertes λ zum Niveau α gegeben.
   - Dabei wird bei der praktischen Anwendung dieses Tests (bei großem Stichprobenumfang n) die Hypothese $H_0 : \lambda = \lambda_0$ abgelehnt, falls $|T_n(x_1,\ldots,x_n)| > z_{1-\alpha/2}$.

2. *Test der „Erfolgswahrscheinlichkeit" p bei Bernoulli-Verteilung*

   - Es gelte nun $X_i \sim \mathrm{Bin}(1,p)$ für ein (unbekanntes) $p \in (0,1)$, wobei

- für einen vorgegebenen (hypothetischen) Zahlenwert $p_0 \in (0,1)$ die Hypothese $H_0 : p = p_0$ (gegen die Alternative $H_1 : p \neq p_0$) getestet werden soll, d.h. $\Theta = \{p : p \in (0,1)\}$ mit

$$\Theta_0 = \{p_0\} \qquad \text{bzw.} \qquad \Theta_1 = \{p : p \neq p_0\}$$

- Wir betrachten die Testgröße $T_n : \mathbb{R}^n \to \mathbb{R}$ mit

$$T_n(x_1,\ldots,x_n) = \begin{cases} \sqrt{n}\,\dfrac{\overline{x}_n - p_0}{\sqrt{\overline{x}_n(1-\overline{x}_n)}}\,, & \text{falls } 0 < \overline{x}_n < 1, \\ 0\,, & \text{sonst} \end{cases} \tag{19}$$

  und den Schwellenwert $c = z_{1-\alpha/2}$.
- Dann gilt (vgl. Beispiel 2 des Abschnittes 3.4.1)

$$\lim_{n\to\infty} P_{p_0}\bigl(|T_n(X_1,\ldots,X_n)| > z_{1-\alpha/2}\bigr) = \alpha\,.$$

- Durch die in (19) eingeführte Folge $\{T_n\}$ von Testfunktionen ist also ein asymptotischer Test des Parameters p zum Niveau α gegeben.
- Bei großem n wird die Hypothese $H_0 : p = p_0$ abgelehnt, falls $|T_n(x_1,\ldots,x_n)| > z_{1-\alpha/2}$.

4.4.2 Zwei-Stichproben-Probleme

Wir betrachten nun noch zwei Beispiele von asymptotischen Tests für den *Zwei-Stichproben-Fall.*

- Seien $n_1, n_2 \in \mathbb{N}$ zwei (große) vorgegebene Zahlen.
- So wie in Abschnitt 3.4.2 fassen wir die (konkreten) Stichproben $(x_{11},\ldots,x_{1n_1})$ und $(x_{21},\ldots,x_{2n_2})$ als Realisierungen von zwei unabhängigen (d.h. nicht verbundenen) Zufallsstichproben $(X_{11},\ldots,X_{1n_1})$ bzw. $(X_{21},\ldots,X_{2n_2})$ auf.
- Zur Vereinfachung der Schreibweise setzen wir $n = \max\{n_1, n_2\}$ bzw. $x_i = (x_{1i}, x_{2i})$ für $i = 1,\ldots,n$.

1. *Test der Gleichheit zweier Erwartungswerte (bei Poisson-Verteilung)*
   - Wir nehmen an, dass $X_{1i} \sim \text{Poi}(\lambda_1)$ und $X_{2i} \sim \text{Poi}(\lambda_2)$ für (unbekannte) $\lambda_1, \lambda_2 > 0$.
   - Dabei soll die Hypothese $H_0 : \lambda_1 = \lambda_2$ (gegen die Alternative $H_1 : \lambda_1 \neq \lambda_2$) getestet werden, d.h. $\Theta = \{(\lambda_1,\lambda_2) : \lambda_1, \lambda_2 > 0\}$ mit

$$\Theta_0 = \{(\lambda_1,\lambda_2) : \lambda_1 = \lambda_2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(\lambda_1,\lambda_2) : \lambda_1 \neq \lambda_2\}$$

   - Wir betrachten die Testgröße $T_n : \mathbb{R}^{2n} \to \mathbb{R}$ mit

$$T_n(x_1,\ldots,x_n) = \begin{cases} \dfrac{\overline{x}_{1n_1} - \overline{x}_{2n_2}}{\sqrt{\dfrac{\overline{x}_{1n_1}}{n_1} + \dfrac{\overline{x}_{2n_2}}{n_2}}}\,, & \text{falls } \max\{\overline{x}_{1n_1}, \overline{x}_{2n_2}\} > 0, \\ 0\,, & \text{sonst} \end{cases} \tag{20}$$

     und den Schwellenwert $c = z_{1-\alpha/2}$.
   - Dann gilt (vgl. Beispiel 1 des Abschnittes 3.4.2)

$$\lim_{n_1,n_2\to\infty} P_{(\lambda_1,\lambda_2)}\bigl(|T_n(X_1,\ldots,X_n)| > z_{1-\alpha/2}\bigr) = \alpha$$

     für beliebige $\lambda_1, \lambda_2 > 0$ mit $\lambda_1 = \lambda_2$.

- Durch die in (20) eingeführte Folge $\{T_n\}$ von Testfunktionen ist also ein asymptotischer Test der Hypothese $H_0 : \lambda_1 = \lambda_2$ zum Niveau α gegeben.
- Bei großem n wird die Nullhypothese H_0 abgelehnt, falls $|T_n(x_1, \ldots, x_n)| > z_{1-\alpha/2}$.

2. *Test der Gleichheit zweier „Erfolgswahrscheinlichkeiten" (bei Bernoulli-Verteilung)*

- Es gelte nun $X_{1i} \sim \mathrm{Bin}(1, p_1)$ und $X_{2i} \sim \mathrm{Bin}(1, p_2)$ für (unbekannte) $p_1, p_2 \in (0, 1)$, wobei
- die Hypothese $H_0 : p_1 = p_2$ (gegen die Alternative $H_1 : p_1 \neq p_2$) getestet werden soll, d.h. $\Theta = \{(p_1, p_2) : p_1, p_2 \in (0, 1)\}$ mit

$$\Theta_0 = \{(p_1, p_2) : p_1 = p_2\} \qquad \text{bzw.} \qquad \Theta_1 = \{(p_1, p_2) : p_1 \neq p_2\}$$

- Wir betrachten die Testgröße $T_n : \mathbb{R}^{2n} \to \mathbb{R}$ mit

$$T_n(x_1, \ldots, x_n) = \begin{cases} \dfrac{\overline{x}_{1n_1} - \overline{x}_{2n_2}}{\sqrt{\dfrac{\overline{x}_{1n_1}}{n_1}(1 - \overline{x}_{1n_1}) + \dfrac{\overline{x}_{2n_2}}{n_2}(1 - \overline{x}_{2n_2})}}\,, & \text{falls } 0 < \overline{x}_{1n_1} < 1 \text{ oder } 0 < \overline{x}_{2n_2} < 1, \\ 0\,, & \text{sonst} \end{cases} \tag{21}$$

  und den Schwellenwert $c = z_{1-\alpha/2}$.
- Dann ergibt sich (auf die gleiche Weise wie in dem vorhergehenden Beispiel), dass

$$\lim_{n_1, n_2 \to \infty} P_{(p_1, p_2)}\big(|T_n(X_1, \ldots, X_n)| > z_{1-\alpha/2}\big) = \alpha$$

  für beliebige $p_1, p_2 \in (0, 1)$ mit $p_1 = p_2$.
- Durch die in (21) eingeführte Folge $\{T_n\}$ von Testfunktionen ist also ein asymptotischer Test der Hypothese $H_0 : p_1 = p_2$ zum Niveau α gegeben.
- Bei großem n wird die Nullhypothese H_0 abgelehnt, falls $|T_n(x_1, \ldots, x_n)| > z_{1-\alpha/2}$.

5 Einfache lineare Regression

- So wie bei den 2–Stichproben-Problemen, die wir in den Kapiteln 3 und 4 betrachtet haben, gehen wir bei der *einfachen Regression* von zwei Datensätzen $(x_1,\ldots,x_n)\in\mathbb{R}^n$ und $(y_1,\ldots,y_n)\in\mathbb{R}^n$ aus, die stochastisch modelliert werden sollen.
- Dabei fassen wir die Vektoren $(x_1,y_1),\ldots,(x_n,y_n)$ als Realisierungen von n Zufallsvektoren $(X_1,Y_1),\ldots,(X_n,Y_n)$ auf, die typischerweise *nicht* identisch verteilt sind.
- Wir deuten die Zufallsvariablen $Y_1,\ldots,Y_n$ als *Zielvariablen* und nehmen an, dass sie auf die folgende Weise von den *Ausgangsvariablen* $X_1,\ldots,X_n$ abhängen:

$$Y_i=\varphi(X_i)+\varepsilon_i\,,\qquad \forall\, i=1,\ldots,n\,, \tag{1}$$

  wobei

  - $\varphi:\mathbb{R}\to\mathbb{R}$ eine beliebige (Borel-messbare) Funktion, die sogenannte *Regressionsfunktion* ist und
  - $\varepsilon_1,\ldots,\varepsilon_n:\Omega\to\mathbb{R}$ Zufallsvariablen, sogenannte *Störgrößen* sind, durch die beispielsweise zufällige Messfehler modelliert werden können.

Beachte

- Ein wichtiger Spezialfall liegt vor, wenn die Regressionsfunktion $\varphi:\mathbb{R}\to\mathbb{R}$ eine lineare Funktion ist, die sogenannte *Regressionsgerade*, d.h., wenn es reelle Zahlen $\alpha,\beta\in\mathbb{R}$ gibt mit

$$\varphi(x)=\alpha+\beta x\,,\qquad \forall\, x\in\mathbb{R}, \tag{2}$$

  wobei α die *Regressionskonstante* und β der *Regressionskoeffizient* genannt wird.
- Die Größen $\alpha,\beta\in\mathbb{R}$ sind unbekannte Modellparameter, die aus den beobachteten Daten $(x_1,\ldots,x_n)\in\mathbb{R}^n$ und $(y_1,\ldots,y_n)\in\mathbb{R}^n$ geschätzt werden sollen.

5.1 Schätzung der Modellparameter

5.1.1 Methode der kleinsten Quadrate

- Wir diskutieren zunächst die folgende *Methode der kleinsten Quadrate* (MKQ) zur Bestimmung von Schätzwerten $\widehat{\alpha}$, $\widehat{\beta}$ für die unbekannten Parameter α,β, bei der keine zusätzlichen Voraussetzungen über die Störgrößen $\varepsilon_1,\ldots,\varepsilon_n:\Omega\to\mathbb{R}$ benötigt werden.
- Und zwar sollen $\widehat{\alpha}$, $\widehat{\beta}$ so bestimmt werden, dass der *mittlere quadratische Fehler*

$$e(\alpha,\beta)=\frac{1}{n}\sum_{i=1}^{n}\bigl(y_i-(\alpha+\beta x_i)\bigr)^2 \tag{3}$$

  für $(\alpha,\beta)=(\widehat{\alpha},\widehat{\beta})$ minimal wird, wobei wir voraussetzen, dass $n\ge 2$ und dass nicht alle $x_1,\ldots,x_n$ gleich sind.

Theorem 5.1 *Der Vektor* $(\widehat{\alpha},\widehat{\beta})$ *mit*

$$\widehat{\beta}=\frac{s_{xy}^2}{s_{xx}^2}\,,\qquad \widehat{\alpha}=\overline{y}_n-\widehat{\beta}\,\overline{x}_n\,, \tag{4}$$

minimiert den mittleren quadratischen Fehler $e(\alpha,\beta)$, *wobei* $\overline{x}_n$, $\overline{y}_n$ *die Stichprobenmittel bezeichnen, d.h.*

$$\overline{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i\,,\qquad \overline{y}_n=\frac{1}{n}\sum_{i=1}^{n}y_i\,,$$

und die Stichprobenvarianzen s^2_{xx}, s^2_{yy} bzw. die Stichprobenkovarianz s^2_{xy} gegeben sind durch

$$s^2_{xx} = \frac{1}{n-1}\sum_{i=1}^n \bigl(x_i - \overline{x}_n\bigr)^2\,, \quad s^2_{xy} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)(y_i - \overline{y}_n)\,, \quad s^2_{yy} = \frac{1}{n-1}\sum_{i=1}^n \bigl(y_i - \overline{y}_n\bigr)^2\,.$$

Beweis

- Die Minimierung kann wie folgt durchgeführt werden: Durch Differenzieren nach α erkennt man, dass für jedes fest vorgegebene $\beta \in \mathbb{R}$ die Zahl

$$\alpha = \frac{1}{n}\sum_{i=1}^n (y_i - \beta x_i) = \overline{y}_n - \beta\overline{x}_n$$

den Wert des Ausdruckes

$$\sum_{i=1}^n \bigl(y_i - (\alpha + \beta x_i)\bigr)^2 = \sum_{i=1}^n \bigl((y_i - \beta x_i) - \alpha\bigr)^2$$

minimiert.
- Mit anderen Worten: Für jedes fest vorgegebene $\beta \in \mathbb{R}$ ist

$$\begin{aligned}\sum_{i=1}^n \bigl((y_i - \beta x_i) - (\overline{y}_n - \beta\overline{x}_n)\bigr)^2 &= \sum_{i=1}^n \bigl((y_i - \overline{y}_n) - \beta(x_i - \overline{x}_n)\bigr)^2 \\ &= (n-1)\bigl(s^2_{yy} - 2\beta s^2_{xy} + \beta^2 s^2_{xx}\bigr)\end{aligned}$$

der kleinste Wert des mittleren quadratischen Fehlers.
- Durch Differenzieren dieses Ausdruckes nach β ergibt sich nun, dass das globale Minimum an der Stelle

$$\beta = \frac{s^2_{xy}}{s^2_{xx}}$$

angenommen wird. □

Beachte

- Der in (3) definierte mittlere quadratische Fehler $e(\alpha,\beta)$ ist der mittlere quadratische *vertikale* Abstand zwischen den Punkten (x_i, y_i) und den Werten $(x_i, \varphi(x_i))$ der Regressionsgeraden $\varphi(x) = \alpha + \beta x$ an den Stellen $x_1, \ldots, x_n$.
- Anstelle der vertikalen Abstände kann man beispielsweise auch die horizontalen Abstände betrachten. Durch Vertauschen der Rollen von x und y ergibt sich dann der MKQ-Ansatz

$$\widehat{\beta'}(x,y) = \frac{s^2_{xy}}{s^2_{yy}}\,, \qquad \widehat{\alpha'}(x,y) = \overline{x}_n - \widehat{\beta'}\overline{y}_n$$

zur Schätzung der Parameter $\alpha', \beta' \in \mathbb{R}$ der (inversen) Regressionsgeraden $\varphi'(y) = x = \alpha' + \beta' y$.
- Wenn wir diese Geradengleichung nach y auflösen, dann ergibt sich die Gleichung

$$y = -\frac{\alpha'}{\beta'} + \frac{1}{\beta'}\,x\,,$$

wobei allerdings die Schätzer $-\widehat{\alpha'}/\widehat{\beta'}$ und $\widehat{\beta'}^{-1}$ für Regressionskonstante bzw. Regressionskoeffizient im allgemeinen verschieden von den Schätzern $\widehat{\alpha}$ bzw. $\widehat{\beta}$ sind, die in (4) hergeleitet worden sind.

Beispiel (vgl. Casella/Berger (2002) *Statistical Inference*, Duxbury, S. 540ff.)

- Im Weinanbau werden die jeweils im Herbst geernteten Erträge in Tonnen je 100 m^2 (t/ar) gemessen.
- Es ist bekannt, dass der Jahresertrag bereits im Juli ziemlich gut prognostiziert werden kann, und zwar durch die Bestimmung der mittleren Anzahl von Beeren, die je Traube gebildet worden sind.
- Mit Hilfe des folgenden Zahlenbeispiels soll illustriert werden, wie die einfache lineare Regression zur Vorhersage des Jahresertrages dienen kann.
- Dabei fassen wir den Jahresertrag als Zielvariable (Y_i) auf, und die mittlere Clusterzahl je Traube (X_i) als Ausgangsvariable.

Jahr	Ertrag (y_i)	Clusterzahl (x_i)
1971	5.6	116.37
1973	3.2	82.77
1974	4.5	110.68
1975	4.2	97.50
1976	5.2	115.88
1977	2.7	80.19
1978	4.8	125.24
1979	4.9	116.15
1980	4.7	117.36
1981	4.1	93.31
1982	4.4	107.46
1983	5.4	122.30

Die Daten des Jahres 1972 fehlen, weil in diesem Jahr das untersuchte Weinanbaugebiet von einem Wirbelsturm verwüstet worden war.

Übungsaufgabe

- Zeichnen Sie ein Streuungsdiagramm (Punktwolke) für die beobachteten Daten.
- Bestimmen Sie für dieses Zahlenbeispiel die Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ sowie $\widehat{\alpha}'$ und $\widehat{\beta}'$ und zeichnen Sie die geschätzte Regressionsgerade in das Streuungsdiagramm ein.
- Prognostizieren Sie mit Hilfe der geschätzten Regressionsgerade $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ den Jahresertrag, der einer mittleren Clusterzahl von 100 Beeren je Traube entsprechen würde.

5.1.2 Beste lineare erwartungstreue Schätzer

- In Abschnitt 5.1.1 wurden keine spezifischen Modellannahmen über die Störgrößen $\varepsilon_1, \ldots, \varepsilon_n$ benötigt.
- Allerdings konnten deshalb in Abschnitt 5.1.1 auch keine Güteeigenschaften der MKQ-Schätzer $\widehat{\alpha}$, $\widehat{\beta}$ (außer der Minimierung des mittleren quadratischen Fehlers $e(\alpha, \beta)$) hergeleitet werden.
- Nun setzen wir zusätzlich voraus, dass die Zufallsvariablen $\varepsilon_1, \ldots, \varepsilon_n$ unkorreliert sind und dass
$$\mathbb{E}\,\varepsilon_i = 0\,, \qquad \operatorname{Var}\varepsilon_i = \sigma^2 \tag{5}$$
für jedes $i = 1, \ldots, n$ gilt, wobei $\sigma^2 > 0$ eine gewisse (im allgemeinen unbekannte) Konstante ist.

- Die Ausgangsvariablen $X_1, \ldots, X_n$ seien deterministisch, d.h., es gelte $X_1 = x_1, \ldots, X_n = x_n$ für fest vorgegebene (d.h. bekannte) Zahlen $x_1, \ldots, x_n \in \mathbb{R}$, wobei wir stets voraussetzen, dass $n \geq 2$ und dass nicht alle $x_1, \ldots, x_n$ gleich sind.
- Für die Zielvariablen $Y_1, \ldots, Y_n$ gilt dann für jedes $i = 1, \ldots, n$

$$Y_i = \alpha + \beta x_i + \varepsilon_i \,, \tag{6}$$

und aus den allgemeinen Rechenregeln für Erwartungswert bzw. Varianz (vgl. die Abschnitte WR-4.1 bzw. WR-4.2 der Vorlesung „Wahrscheinlichkeitsrechnung") ergibt sich für jedes $i = 1, \ldots, n$

$$\mathbb{E}\, Y_i = \alpha + \beta x_i \,, \qquad \operatorname{Var} Y_i = \sigma^2 \,. \tag{7}$$

- Die Regressionskonstante α und der Regressionskoeffizient β sollen nun mit einem linearen Schätzer aus den beobachteten Realisierungen $y_1, \ldots, y_n$ der Zielvariablen $Y_1, \ldots, Y_n$ geschätzt werden.
- Dabei verstehen wir unter einem *linearen Schätzer* eine Linearkombination $d_1Y_1 + \ldots + d_nY_n$ der Zielvariablen, wobei $d_1, \ldots, d_n \in \mathbb{R}$ fest vorgegebene (d.h. bekannte) Konstanten sind.

Theorem 5.2 *Der lineare Schätzer $\widehat{\beta} = d_1Y_1 + \ldots + d_nY_n$ ist genau dann ein erwartungstreuer Schätzer für β, wenn*

$$\sum_{i=1}^n d_i = 0 \qquad \textit{und} \qquad \sum_{i=1}^n d_i x_i = 1 \,. \tag{8}$$

Beweis

- Der Schätzer $\widehat{\beta} = d_1Y_1 + \ldots + d_nY_n$ ist erwartungstreu für β, wenn

$$\mathbb{E}\, \widehat{\beta} = \mathbb{E} \sum_{i=1}^n d_i Y_i = \beta$$

für beliebige Parameterwerte $\alpha, \beta \in \mathbb{R}$ gilt.
- Wegen (6) und (7) bedeutet dies, dass

$$\begin{aligned} \beta &= \mathbb{E} \sum_{i=1}^n d_i Y_i = \sum_{i=1}^n d_i \mathbb{E}\, Y_i \\ &= \sum_{i=1}^n d_i(\alpha + \beta x_i) \\ &= \alpha\Bigl(\sum_{i=1}^n d_i\Bigr) + \beta\Bigl(\sum_{i=1}^n d_i x_i\Bigr) . \end{aligned}$$

- Es ist klar, dass diese Gleichung genau dann für alle $\alpha, \beta \in \mathbb{R}$ gilt, wenn die Bedingungen (8) erfüllt sind. □

Beachte

- In Abschnitt 2.3.5 hatten wir einen erwartungstreuen Schätzer besten erwartungstreuen Schätzer genannt, wenn seine Varianz minimal ist.
- Analog hierzu wird ein linearer erwartungstreuer Schätzer *bester linearer erwartungstreuer Schätzer* genannt, wenn es keinen linearen erwartungstreuen Schätzer gibt, dessen Varianz kleiner ist.
- In der englischsprachigen Literatur ist es üblich, solche Schätzer BLUE zu nennen (BLUE = best linear unbiased estimator).

Theorem 5.3 *Der lineare Schätzer $\widehat{\beta} = d_1 Y_1 + \ldots + d_n Y_n$ ist genau dann ein BLUE–Schätzer für β, wenn die Konstanten $d_1, \ldots, d_n \in \mathbb{R}$ gegeben sind durch*

$$d_i = \frac{x_i - \overline{x}_n}{(n-1)s_{xx}^2}, \qquad \forall\, i = 1, \ldots, n. \tag{9}$$

Beweis

- Weil mit den Störgrößen $\varepsilon_1, \ldots, \varepsilon_n$ auch die Zielvariablen $Y_1, \ldots, Y_n$ unkorreliert sind, ergibt sich aus der allgemeinen Formel für die Varianz der Summe von unkorrelierten Zufallsvariablen (vgl. Theorem WR-4.13), dass

$$\begin{aligned} \operatorname{Var} \sum_{i=1}^n d_i Y_i &= \sum_{i=1}^n \operatorname{Var}(d_i Y_i) \\ &= \sum_{i=1}^n d_i^2 \operatorname{Var} Y_i \\ &= \sigma^2 \sum_{i=1}^n d_i^2 \end{aligned}$$

  für beliebige Konstanten $d_1, \ldots, d_n \in \mathbb{R}$.
- Der Schätzer $\widehat{\beta} = d_1 Y_1 + \ldots + d_n Y_n$ ist also genau dann ein BLUE–Schätzer für β, wenn die Konstanten $d_1, \ldots, d_n \in \mathbb{R}$ so gewählt sind, dass die Bedingung (8) erfüllt ist und dass die Summe $\sum_{i=1}^n d_i^2$ minimal ist.
- Wegen (8) gilt $\sum\limits_{i=1}^n d_i x_i = 1$ und somit

$$\frac{\Bigl(\sum\limits_{i=1}^n d_i x_i\Bigr)^2}{\sum\limits_{i=1}^n d_i^2} = \frac{1}{\sum\limits_{i=1}^n d_i^2}. \tag{10}$$

- Die Summe $\sum_{i=1}^n d_i^2$ ist also genau dann minimal, wenn die linke Seite von (10) maximal ist.
- Weil darüber hinaus $\overline{d}_n = n^{-1} \sum_{i=1}^n d_i = 0$, gilt außerdem

$$\begin{aligned} \frac{\Bigl(\sum\limits_{i=1}^n d_i x_i\Bigr)^2}{\sum\limits_{i=1}^n d_i^2} &= \frac{\Bigl(\sum\limits_{i=1}^n (d_i - \overline{d}_n)(x_i - \overline{x}_n)\Bigr)^2}{\sum\limits_{i=1}^n \bigl(d_i - \overline{d}_n\bigr)^2} \\ &= n \sum_{i=1}^n \frac{\bigl(x_i - \overline{x}_n\bigr)^2}{n} \left(\frac{\sum\limits_{i=1}^n \frac{1}{n} \bigl(d_i - \overline{d}_n\bigr)\bigl(x_i - \overline{x}_n\bigr)}{\sqrt{\sum\limits_{i=1}^n \frac{\bigl(d_i - \overline{d}_n\bigr)^2}{n} \sum\limits_{i=1}^n \frac{\bigl(x_i - \overline{x}_n\bigr)^2}{n}}} \right)^2. \end{aligned}$$

- Der quadrierte Ausdruck kann nun als Korrelationskoeffizient aufgefasst werden bei (diskreter) Gleichverteilung der natürlichen Zahlen $1, \ldots, n$ und zwar
  - für den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\Omega = \{1, \ldots, n\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{i\}) = 1/n$, sowie
  - für die Zufallsvariablen $D, X : \Omega \to \mathbb{R}$ mit $D(i) = d_i$ und $X(i) = x_i$.
- Gemäß Theorem WR-4.12 ist also dieser quadrierte Ausdruck genau dann maximal, wenn es Konstanten $a, b \in \mathbb{R}$ gibt, so dass

$$d_i = a x_i + b, \qquad \forall\, i = 1, \ldots, n. \tag{11}$$

- Wegen (8) müssen die Konstanten $a, b \in \mathbb{R}$ den folgenden beiden Gleichungen genügen:

$$\sum_{i=1}^n (ax_i + b) = 0\,, \qquad \sum_{i=1}^n (ax_i + b)x_i = 1\,.$$

- Hieraus folgt, dass

$$b = -a\overline{x}_n\,, \qquad a = \frac{1}{\sum\limits_{i=1}^n (x_i - \overline{x}_n)^2}\,.$$

- Wird dies in (11) eingesetzt, so ergibt sich die Behauptung (9). □

Beachte

- Der in Theorem 5.3 hergeleitete BLUE-Schätzer

$$\widehat{\beta} = \sum_{i=1}^n \frac{x_i - \overline{x}_n}{(n-1)s_{xx}^2}\, Y_i = \sum_{i=1}^n \frac{(x_i - \overline{x}_n)(Y_i - \overline{Y}_n)}{(n-1)s_{xx}^2} \qquad \Bigl(= s_{xY}^2 / s_{xx}^2\Bigr) \tag{12}$$

  für β stimmt mit dem in Theorem 5.1 hergeleiteten MKQ-Schätzansatz überein.
- Aus dem Beweis von Theorem 5.3 ist ersichtlich, dass die Varianz des BLUE-Schätzers $\widehat{\beta}$ in (12) gegeben ist durch:

$$\operatorname{Var}\widehat{\beta} = \sigma^2 \sum_{i=1}^n d_i^2 = \frac{\sigma^2}{(n-1)s_{xx}^2} = \frac{\sigma^2}{\sum\limits_{i=1}^n (x_i - \overline{x}_n)^2}\,. \tag{13}$$

Übungsaufgabe

- Auf ähnliche Weise wie in den Theoremen 5.2 und 5.3 lässt sich ein BLUE-Schätzer $\widehat{\alpha} = c_1 Y_1 + \ldots + c_n Y_n$ für die Regressionskonstante α bestimmen.
- Und zwar lässt sich wie im Beweis von Theorem 5.2 zeigen, dass

$$\sum_{i=1}^n c_i = 1 \qquad \text{und} \qquad \sum_{i=1}^n c_i x_i = 0 \tag{14}$$

  gelten muss, damit $\widehat{\alpha}$ ein erwartungstreuer Schätzer für α ist.
- Außerdem lässt sich wie im Beweis von Theorem 5.3 zeigen, dass die Varianz von $\widehat{\alpha} = c_1 Y_1 + \ldots + c_n Y_n$ minimal ist, wenn

$$c_i = \frac{1}{n} - \frac{\overline{x}_n(x_i - \overline{x}_n)}{(n-1)s_{xx}^2}\,, \qquad \forall\, i = 1, \ldots, n\,. \tag{15}$$

5.1.3 Normalverteilte Störgrößen

- Zusätzlich zu den in Abschnitt 5.1.2 gemachten Modellannahmen setzen wir von nun an voraus, dass die Störgrößen $\varepsilon_1, \ldots, \varepsilon_n$ unabhängig und normalverteilt sind.
- Wegen (5) gilt dann $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ bzw. $Y_i \sim \mathrm{N}(\alpha + \beta x_i, \sigma^2)$ für jedes $i = 1, \ldots, n$.
- Außerdem ergibt sich aus dem Satz über die Unabhängigkeit zusammengesetzter Abbildungen (vgl. Theorem WR-3.18), dass die Zielvariablen $Y_1, \ldots, Y_n$ unabhängig sind.
- Für jeden Vektor $(x_1, \ldots, x_n) \in \mathbb{R}$ betrachten wir die Loglikelihood-Funktion der unabhängigen (jedoch im allgemeinen *nicht* identisch verteilten) Stichprobenvariablen $Y_1, \ldots, Y_n$:

$$\log L(y_1, \ldots, y_n; \alpha, \beta, \sigma^2) = -\,\frac{n}{2}\,\log(2\pi) - \frac{n}{2}\,\log\sigma^2 - \frac{\sum\limits_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\,.$$

- Für jedes $\sigma^2 > 0$ und für jeden Vektor $(y_1, \ldots, y_n) \in \mathbb{R}$ nimmt die Loglikelihood-Funktion als Funktion von (α, β) ihr Maximum für denjenigen Vektor $(\widehat{\alpha}, \widehat{\beta})$ an, der den Ausdruck $\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$ minimiert.
- Dieses Minimierungsproblem wurde bereits in Theorem 5.1 betrachtet: Die Lösung lautet

$$\widehat{\beta} = \frac{s_{xy}^2}{s_{xx}^2}\,, \qquad \widehat{\alpha} = \overline{y}_n - \widehat{\beta}\overline{x}_n\,, \tag{16}$$

- Mit anderen Worten: Bei normalverteilten Störgrößen stimmt der MKQ-Schätzer mit dem ML-Schätzer für (α, β) überein.

Beachte

- Weil $(\widehat{\alpha}, \widehat{\beta})$ die Loglikelihood-Funktion für jedes $\sigma^2 > 0$ maximiert, ergibt sich der ML-Schätzer $\widehat{\sigma}^2$ für σ^2 als Maximum von

$$\log L(y_1, \ldots, y_n; \widehat{\alpha}, \widehat{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{\sum\limits_{i=1}^n (y_i - \widehat{\alpha} - \widehat{\beta}x_i)^2}{2\sigma^2}\,.$$

- Ähnlich wie im Fall von unabhängigen und *identisch* normalverteilten Stichprobenvariablen (vgl. Beispiel 5 in Abschnitt 2.2.2) ergibt sich die Lösung dieses Maximierungsproblems durch zweimaliges Differenzieren nach σ^2:

$$\widehat{\sigma}^2(y_1, \ldots, y_n) = \frac{1}{n}\sum_{i=1}^n (y_i - \widehat{\alpha} - \widehat{\beta}x_i)^2\,. \tag{17}$$

- Der in (17) gegebene ML-Schätzer $\widehat{\sigma}^2$ für σ^2 ist nicht erwartungstreu.
- Um dies zu zeigen, betrachten wir zunächst die folgenden (Abweichungs-) *Residuen*

$$\widehat{\varepsilon}_i = Y_i - \widehat{\alpha} - \widehat{\beta}x_i\,, \qquad \forall\, i = 1, \ldots, n\,. \tag{18}$$

- Offenbar gilt

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i^2\,. \tag{19}$$

- Um den Erwartungswert $\mathbb{E}\,\widehat{\sigma}^2$ zu bestimmen, genügt es also, die zweiten Momente $\mathbb{E}\,(\widehat{\varepsilon}_i^2)$ der Residuen $\widehat{\varepsilon}_i$ für jedes $i = 1, \ldots, n$ zu bestimmen.
- Hierfür ist der folgende Hilfssatz nützlich.

Lemma 5.1 *Seien $Y_1, \ldots, Y_n : \Omega \to \mathbb{R}$ beliebige unkorrelierte Zufallsvariablen mit $\mathbb{E}\,(Y_i^2) < \infty$ und $\operatorname{Var} Y_i = \sigma^2$ für jedes $i = 1, \ldots, n$. Für beliebige Konstanten $c_1, \ldots, c_n \in \mathbb{R}$ und $d_1, \ldots, d_n \in \mathbb{R}$ gilt dann*

$$\operatorname{Cov}\Big(\sum_{i=1}^n c_iY_i, \sum_{j=1}^n d_jY_j\Big) = \sigma^2\sum_{i=1}^n c_id_i\,. \tag{20}$$

Beweis Es gilt

$$\begin{aligned}
\operatorname{Cov}\Big(\sum_{i=1}^n c_iY_i, \sum_{j=1}^n d_jY_j\Big) &= \sum_{i=1}^n c_i\operatorname{Cov}\Big(Y_i, \sum_{j=1}^n d_jY_j\Big)\\
&= \sum_{i=1}^n\sum_{j=1}^n c_id_j\operatorname{Cov}(Y_i, Y_j)\\
&= \sum_{i=1}^n c_id_i\operatorname{Cov}(Y_i, Y_i)\\
&= \sigma^2\sum_{i=1}^n c_id_i\,.
\end{aligned}$$

□

Theorem 5.4 *Für den Erwartungswert* $\mathbb{E}\,\widehat{\varepsilon}_i$ *und die Varianz* $\operatorname{Var}\widehat{\varepsilon}_i$ *der Residuen* $\widehat{\varepsilon}_i$ *gilt für jedes* $i = 1, \ldots, n$

$$\mathbb{E}\,\widehat{\varepsilon}_i = 0 \tag{21}$$

und

$$\operatorname{Var}\widehat{\varepsilon}_i = \mathbb{E}\,(\widehat{\varepsilon}_i^2) = \sigma^2\Bigl(\frac{n-2}{n} + \frac{1}{(n-1)s_{xx}^2}\Bigl(\frac{1}{n}\sum_{j=1}^n x_j^2 + x_i^2 - 2(x_i - \overline{x}_n)^2 - 2x_i\overline{x}_n\Bigr)\Bigr)\,. \tag{22}$$

Beweis

- Weil $\mathbb{E}\,\varepsilon_i = 0$, d.h. $\mathbb{E}\,Y_i = \alpha + \beta x_i$, und weil $\widehat{\alpha}$ bzw. $\widehat{\beta}$ erwartungstreue Schätzer für α bzw. β sind, gilt

$$\begin{aligned}\mathbb{E}\,\widehat{\varepsilon}_i &= \mathbb{E}\,(Y_i - \widehat{\alpha} - \widehat{\beta}x_i)\\ &= \mathbb{E}\,Y_i - \mathbb{E}\,\widehat{\alpha} - x_i\mathbb{E}\,\widehat{\beta}\\ &= \alpha + \beta x_i - \alpha - x_i\beta = 0\,.\end{aligned}$$

- Außerdem ergibt sich aus den allgemeinen Rechenregeln für die Varianz von Summen beliebiger (nicht notwendig unabhängiger) Zufallsvariablen (vgl. Theorem WR-4.13), dass

$$\operatorname{Var}\widehat{\varepsilon}_i = \operatorname{Var}Y_i + \operatorname{Var}\widehat{\alpha} + x_i^2\operatorname{Var}\widehat{\beta} - 2\operatorname{Cov}(Y_i,\widehat{\alpha}) - 2x_i\operatorname{Cov}(Y_i,\widehat{\beta}) + 2x_i\operatorname{Cov}(\widehat{\alpha},\widehat{\beta})\,.$$

- Für die Kovarianzen ergibt sich nun aus (9) und (15) mit Hilfe von Lemma 5.1, dass

$$\operatorname{Cov}(Y_i,\widehat{\alpha}) = \sigma^2\Bigl(\frac{1}{n} - \frac{(x_i - \overline{x}_n)\overline{x}_n}{(n-1)s_{xx}^2}\Bigr)\,,\qquad \operatorname{Cov}(Y_i,\widehat{\beta}) = \sigma^2\,\frac{x_i - \overline{x}_n}{(n-1)s_{xx}^2}$$

  und

$$\operatorname{Cov}(\widehat{\alpha},\widehat{\beta}) = -\,\frac{\sigma^2\overline{x}_n}{(n-1)s_{xx}^2}\,. \tag{23}$$

- Auf ähnliche Weise ergibt sich aus (15), dass

$$\operatorname{Var}\widehat{\alpha} = \frac{\sigma^2}{n(n-1)s_{xx}^2}\sum_{i=1}^n x_i^2\,. \tag{24}$$

- Aus diesen Formeln und aus (13) ergibt sich nun die Behauptung. □

Korollar 5.1 *Für den Erwartungswert des in* (17) *gegebenen ML-Schätzers* $\widehat{\sigma}^2$ *gilt*

$$\mathbb{E}\,\widehat{\sigma}^2 = \frac{n-2}{n}\,\sigma^2\,. \tag{25}$$

Beweis

- Aus (19) und aus Theorem 5.4 ergibt sich, dass

$$\begin{aligned}\mathbb{E}\,\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\,\widehat{\varepsilon}_i^2\\ &= \frac{\sigma^2}{n}\sum_{i=1}^n\Bigl(\frac{n-2}{n} + \frac{1}{(n-1)s_{xx}^2}\Bigl(\frac{1}{n}\sum_{j=1}^n x_j^2 + x_i^2 - 2(x_i-\overline{x}_n)^2 - 2x_i\overline{x}_n\Bigr)\Bigr)\\ &= \sigma^2\Bigl(\frac{n-2}{n} + \frac{1}{n(n-1)s_{xx}^2}\Bigl(\sum_{j=1}^n x_j^2 + \sum_{i=1}^n x_i^2 - 2(n-1)s_{xx}^2 - 2\,\frac{1}{n}\,\Bigl(\sum_{i=1}^n x_i\Bigr)^2\Bigr)\Bigr)\,.\end{aligned}$$

- Weil
$$\sum_{i=1}^n x_i^2 - \frac{1}{n}\Bigl(\sum_{i=1}^n x_i\Bigr)^2 = (n-1)s_{xx}^2\,,$$
ergibt sich hieraus die Behauptung. □

Beachte

- Wegen (25) ist es üblich, anstelle des ML-Schätzers $\widehat{\sigma}^2$ den folgenden (erwartungstreuen) Schätzer S^2 für σ^2 zu verwenden:
$$S^2 = \frac{n}{n-2}\,\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n \widehat{\varepsilon}_i^2\,, \tag{26}$$
wobei vorausgesetzt wird, dass $n > 2$.
- Um Hypothesen über die Modellparameter α, β bzw. σ^2 testen zu können, die auf den Schätzern $\widehat{\alpha}$, $\widehat{\beta}$ bzw. S^2 beruhen, müssen wir die Verteilungen dieser Zufallsvariablen bzw. die (stochastischen) Zusammenhänge, die gegebenenfalls zwischen ihnen bestehen, kennen.
- In diesem Zusammenhang sind die folgenden Eigenschaften der χ^2-Verteilung bzw. der Normalverteilung nützlich.

Lemma 5.2

- *Die Zufallsvariablen $U,V : \Omega \to \mathbb{R}$ seien unabhängig.*
- *Falls außerdem*
$$V \sim \chi_m^2 \quad \textit{und} \quad U + V \sim \chi_n^2\,, \tag{27}$$
wobei $n, m \in \mathbb{N}$ beliebige natürliche Zahlen sind mit $m < n$, dann gilt $U \sim \chi_{n-m}^2$.

Beweis

- Aus der Unabhängigkeit von U und V folgt, dass
$$\varphi_{U+V}(t) = \varphi_U(t)\,\varphi_V(t)\,, \qquad \forall t \in \mathbb{R}\,,$$
wobei $\varphi_U(t)$, $\varphi_V(t)$ und $\varphi_{U+V}(t)$ die charakteristischen Funktionen von U, V bzw. $U+V$ sind (vgl. Theorem WR-5.18).
- Falls außerdem (27) gilt, dann ergibt sich hieraus und aus Theorem 1.5, dass für jedes $t \in \mathbb{R}$
$$\begin{aligned}\varphi_U(t) &= \frac{\varphi_{U+V}(t)}{\varphi_V(t)}\\ &= \frac{1}{(1-2\,\mathrm{i}\,t)^{(n-m)/2}}\,.\end{aligned}$$
- Die erneute Anwendung von Theorem 1.5 und des Eindeutigkeitssatzes für charakteristische Funktionen (vgl. Korollar WR-5.5) ergibt nun, dass $U \sim \chi_{n-m}^2$. □

Beachte

- Bei der Herleitung des folgenden Lemmas 5.3 benötigen wir eine *vektorielle Version* des Eindeutigkeitssatzes für charakteristische Funktionen (vgl. Korollar WR-5.5), die wir hier ohne Beweis angeben.

- Seien $X, Y : \Omega \to \mathbb{R}^n$ beliebige Zufallsvektoren; $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$. Dann gilt

$$X \stackrel{\mathrm{d}}{=} Y \qquad \text{genau dann, wenn} \qquad \varphi_X(t) = \varphi_Y(t) \qquad \forall\, t \in \mathbb{R}^n\,, \tag{28}$$

wobei

$$\varphi_X(t) = \mathbb{E}\,\exp\Bigl(\mathrm{i}\sum_{j=1}^n t_j X_j\Bigr)\,, \qquad \varphi_Y(t) = \mathbb{E}\,\exp\Bigl(\mathrm{i}\sum_{j=1}^n t_j Y_j\Bigr)$$

die charakteristischen Funktionen von X bzw. Y sind; $t = (t_1, \ldots, t_n)$.

Lemma 5.3

- *Die Zufallsvariablen $Y_1, \ldots, Y_n$ seien unabhängig und normalverteilt mit $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ für $i = 1, \ldots, n$.*
- *Für beliebige Konstanten $a_{ij}, b_{ik} \in \mathbb{R}$ mit $j = 1, \ldots, l$ und $k = 1, \ldots, m$ seien die Zufallsvariablen $U_1, \ldots, U_l$ und $V_1, \ldots, V_m$ gegeben durch*

$$U_j = \sum_{i=1}^n a_{ij} Y_i\,, \qquad \forall\, j = 1, \ldots, l$$

und

$$V_k = \sum_{i=1}^n b_{ik} Y_i\,, \qquad \forall\, k = 1, \ldots, m\,.$$

- *Dann gilt:*
  1. *Die Zufallsvariablen U_j und V_k sind normalverteilt mit*

  $$U_j \sim \mathrm{N}\Bigl(\sum_{i=1}^n a_{ij}\mu_i, \sum_{i=1}^n a_{ij}^2\sigma_i^2\Bigr)$$

  und

  $$V_k \sim \mathrm{N}\Bigl(\sum_{i=1}^n b_{ik}\mu_i, \sum_{i=1}^n b_{ik}^2\sigma_i^2\Bigr)\,,$$

  wobei $\mathrm{Cov}\,(U_j, V_k) = \sum_{i=1}^n a_{ij} b_{ik} \sigma_i^2$.
  2. *Die Zufallsvariablen U_j und V_k sind genau dann unabhängig, wenn* $\mathrm{Cov}\,(U_j, V_k) = 0$.
  3. *Die Zufallsvektoren $(U_1, \ldots, U_l)$ und $(V_1, \ldots, V_m)$ sind genau dann unabhängig, wenn die Komponenten U_j und V_k für beliebige $j = 1, \ldots, l$ und $k = 1, \ldots, m$ unabhängig sind.*

Beweis

- Die Normalverteilheit der Zufallsvariablen ergibt sich unmittelbar aus der Faltungsstabilität der Normalverteilung, vgl. Korollar WR-3.2.
- Die Formel für die Kovarianz von U_j und V_k ergibt sich aus Lemma 5.1.
- Die Notwendigkeit der Bedingung in Teilaussage 2 ergibt sich aus der Multiplikationsformel für den Erwartungswert des Produktes von unabhängigen Zufallsvariablen, vgl. Korollar WR-4.5.
- Um die Hinlänglichkeit der Bedingung in Teilaussage 2 zu beweisen, können (und werden) wir o.b.d.A. voraussetzen, dass $Y_i \sim \mathrm{N}(0, 1)$ für jedes $i = 1, \ldots, n$.

- Für die charakteristische Funktion $\varphi_{U_j,V_k}(t,s)$ des Zufallsvektors (U_j, V_k) gilt dann für beliebige $t, s \in \mathbb{R}$

$$\begin{aligned}
\varphi_{U_j,V_k}(t,s) &= \mathbb{E}\exp\Bigl(\mathrm{i}\,\Bigl(t\sum_{i=1}^n a_{ij}Y_i + s\sum_{i=1}^n b_{ik}Y_i\Bigr)\Bigr)\\
&= \mathbb{E}\exp\Bigl(\mathrm{i}\,\sum_{i=1}^n (ta_{ij} + sb_{ik})Y_i\Bigr)\\
&= \prod_{i=1}^n \mathbb{E}\exp\Bigl(\mathrm{i}\,(ta_{ij} + sb_{ik})Y_i\Bigr)\\
&= \prod_{i=1}^n \exp\Bigl(-\frac{(ta_{ij} + sb_{ik})^2}{2}\Bigr)\\
&= \exp\Bigl(-\frac{\sum\limits_{i=1}^n (ta_{ij})^2 + \sum\limits_{i=1}^n (sb_{ik})^2}{2}\Bigr)\\
&= \exp\Bigl(-\frac{\sum\limits_{i=1}^n (ta_{ij})^2}{2}\Bigr)\,\exp\Bigl(-\frac{\sum\limits_{i=1}^n (sb_{ik})^2}{2}\Bigr)\\
&= \varphi_{U_j}(t)\,\varphi_{V_k}(s)\,,
\end{aligned}$$

  wobei sich die drittletzte Gleichheit aus der Annahme ergibt, dass Zufallsvariablen U_j und V_k unkorreliert sind und dass deshalb

$$\operatorname{Cov}(U_j, V_k) = \sum_{i=1}^n a_{ij}b_{ik} = 0\,.$$

- Die Hinlänglichkeit der Bedingung in Teilaussage 2 ergibt sich nun aus dem Eindeutigkeitssatz für charakteristische Funktionen von Zufallsvektoren, vgl. (28), weil das Produkt der charakteristischen Funktionen von unabhängigen Zufallsvektoren Z und Z' gleich der (gemeinsamen) charakteristischen Funktion des Zufallsvektors (Z, Z') ist.
- Die Notwendigkeit der Bedingung in Teilaussage 3 ergibt sich unmittelbar aus der Definition der Unabhängigkeit von Zufallsvektoren.
- Die Hinlänglichkeit der Bedingung in Teilaussage 3 lässt sich auf auf ähnliche Weise wie die Hinlänglichkeit der Bedingung in Teilaussage 2 zeigen. □

Theorem 5.5

1. *Für das einfache lineare Regressionsmodell mit normalverteilten Störgrößen $\varepsilon_1, \ldots, \varepsilon_n$ gilt*

$$\widehat{\alpha} \sim \mathrm{N}\Bigl(\alpha, \frac{\sigma^2}{n(n-1)s_{xx}^2}\sum_{i=1}^n x_i^2\Bigr), \qquad \widehat{\beta} \sim \mathrm{N}\Bigl(\beta\,,\ \frac{\sigma^2}{(n-1)s_{xx}^2}\Bigr), \tag{29}$$

   wobei

$$\operatorname{Cov}(\widehat{\alpha}, \widehat{\beta}) = -\frac{\sigma^2\overline{x}_n}{(n-1)s_{xx}^2}\,. \tag{30}$$

2. *Die Zufallsvariablen $(\widehat{\alpha}, \widehat{\beta})$ und S^2 sind unabhängig, und es gilt*

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}\,. \tag{31}$$

Beweis

- Weil die Stichprobenvariablen $Y_1, \ldots, Y_n$ unabhängig und normalverteilt sind und weil die Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ jeweils Linearkombinationen der Stichprobenvariablen $Y_1, \ldots, Y_n$ sind, ergibt sich aus Teilaussage 1 von Lemma 5.3, dass die Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ ebenfalls normalverteilt sind.
- Die Erwartungstreue von $\widehat{\alpha}$ und $\widehat{\beta}$ wurde bereits in Abschnitt 5.1.2 diskutiert.
- Die Varianzen von $\widehat{\alpha}$ und $\widehat{\beta}$ bzw. die Kovarianz $\operatorname{Cov}(\widehat{\alpha}, \widehat{\beta})$ wurden in (13), (23) bzw. (24) bestimmt.
- Die Unabhängigkeit der Zufallsvariablen $(\widehat{\alpha}, \widehat{\beta})$ und S^2 ergibt sich aus den folgenden Überlegungen.
- Aus der Definitionsgleichung (18) der Residuen $\widehat{\varepsilon}_i$ folgt, dass

$$\widehat{\varepsilon}_i = \sum_{j=1}^{n} \bigl(\delta_{ij} - (c_j + d_j x_i)\bigr) Y_j\,, \tag{32}$$

wobei die Konstanten c_j, d_j in (9) bzw. (15) gegeben sind und

$$\delta_{ij} = \begin{cases} 1\,, & \text{falls } i = j,\\ 0\,, & \text{falls } i \neq j. \end{cases}$$

- Aus Lemma 5.1 ergibt sich nun, dass für jedes $i = 1, \ldots, n$

$$\begin{aligned} \operatorname{Cov}(\widehat{\varepsilon}_i, \widehat{\alpha}) &= \operatorname{Cov}\Bigl(\sum_{j=1}^{n} \bigl(\delta_{ij} - (c_j + d_j x_i)\bigr) Y_j, \sum_{k=1}^{n} c_k Y_k\Bigr)\\ &= \sigma^2 \sum_{j=1}^{n} \bigl(\delta_{ij} - (c_j + d_j x_i)\bigr) c_j\\ &= \sigma^2 \Bigl(c_i - \sum_{j=1}^{n} c_j^2 - x_i \sum_{j=1}^{n} c_j d_j\Bigr)\\ &= 0\,. \end{aligned}$$

- Dabei ergibt sich die letzte Gleichheit aus den Darstellungsformeln (9) und (15) für c_i und d_i, d.h.

$$c_i = \frac{1}{n} - \frac{\overline{x}_n (x_i - \overline{x}_n)}{(n-1) s_{xx}^2}\,, \qquad d_i = \frac{x_i - \overline{x}_n}{(n-1) s_{xx}^2}\,, \qquad \forall\, i = 1, \ldots, n\,,$$

denn aus diesen beiden Formeln folgt, dass

$$\sum_{j=1}^{n} c_j^2 = \frac{1}{n} + \frac{\overline{x}_n^2}{(n-1) s_{xx}^2}\,, \qquad \sum_{j=1}^{n} c_j d_j = -\frac{\overline{x}_n}{(n-1) s_{xx}^2}\,.$$

- Auf die gleiche Weise ergibt sich aus Lemma 5.1, dass $\operatorname{Cov}(\widehat{\varepsilon}_i, \widehat{\beta}) = 0$ für jedes $i = 1, \ldots, n$.
- Aus den Teilaussagen 2 und 3 von Lemma 5.3 folgt nun, dass die Zufallsvektoren $(\widehat{\alpha}, \widehat{\beta})$ und $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n)$ unabhängig sind.
- Aus dem Transformationssatz für unabhängige Zufallsvektoren (vgl. Theorem 1.8) ergibt sich somit, dass auch die Zufallsvariablen $(\widehat{\alpha}, \widehat{\beta})$ und S^2 unabhängig sind.
- Um den Beweis zu beenden, bleibt also noch die Gültigkeit der Verteilungseigenschaft (31) zu zeigen.
- Aus (16) und (17) ergibt sich, dass sich die Summe der Abweichungsquadrate $\sum_{i=1}^{n} \widehat{\varepsilon}_i^2$ bei der Skalenverschiebung

$$x_i' = x_i - \overline{x}_n\,, \qquad \forall\, i = 1, \ldots, n$$

nicht ändert. Wir können (und werden) deshalb o.B.d.A. voraussetzen, dass $\overline{x}_n = 0$.

- Die Konstanten c_i, d_i in (32) haben dann die Form

$$c_i = \frac{1}{n}\,, \qquad d_i = \frac{x_i}{\sum\limits_{j=1}^{n} x_j^2}\,. \tag{33}$$

- Hieraus und aus (26) bzw. (32) ergibt sich nun, dass

$$\begin{aligned}(n-2)S^2 &= \sum_{i=1}^{n}\widehat{\varepsilon}_i^2 = \sum_{i=1}^{n}\bigl(Y_i-\widehat{\alpha}-\widehat{\beta}x_i\bigr)^2\\ &= \sum_{i=1}^{n}\bigl(Y_i-\alpha-\beta x_i+(\alpha-\widehat{\alpha})+(\beta-\widehat{\beta})x_i\bigr)^2\\ &= \sum_{i=1}^{n}\bigl(Y_i-\alpha-\beta x_i\bigr)^2 - n\bigl(\widehat{\alpha}-\alpha\bigr)^2 - \sum_{j=1}^{n}x_j^2\bigl(\widehat{\beta}-\beta\bigr)^2\,,\end{aligned}$$

  wobei sich die letzte Gleichheit durch Ausmultiplizieren der Klammern bzw. durch Einsetzen von (33) in die Definitionsgleichungen $\widehat{\alpha} = c_1Y_1 + \ldots + c_nY_n$ und $\widehat{\beta} = d_1Y_1 + \ldots + d_nY_n$ von $\widehat{\alpha}$ bzw. $\widehat{\beta}$ ergibt, wenn dabei berücksichtigt wird, dass $n\,\overline{x}_n = x_1 + \ldots + x_n = 0$.
- Mit anderen Worten: Es gilt

$$(n-2)S^2 + Z^2 = \sum_{i=1}^{n}\bigl(Y_i-\alpha-\beta x_i\bigr)^2\,, \tag{34}$$

  wobei

$$Z^2 = n\bigl(\widehat{\alpha}-\alpha\bigr)^2 + \sum_{j=1}^{n}x_j^2\bigl(\widehat{\beta}-\beta\bigr)^2$$

  und die Zufallsvariablen $Y_i' = Y_i - \alpha - \beta x_i$ für jedes $i = 1,\ldots,n$ unabhängig und identisch $\mathrm{N}(0,\sigma^2)$-verteilt sind.
- Aus (34) und aus der Definition der χ^2-Verteilung ergibt sich somit, dass

$$\frac{(n-2)S^2 + Z^2}{\sigma^2} \sim \chi^2_n\,. \tag{35}$$

- Weil bereits gezeigt wurde, dass die Zufallsvariablen $(\widehat{\alpha},\widehat{\beta})$ und S^2 unabhängig sind, sind somit auch die Zufallsvariablen $(n-2)S^2$ und Z^2 unabhängig.
- Außerdem gilt $Z^2 = Z_1^2 + Z_2^2$, wobei sich aus (29) und (30) bzw. aus Lemma 5.3 ergibt, dass die Zufallsvariablen

$$Z_1 = \sqrt{n}\,\bigl(\widehat{\alpha}-\alpha\bigr)\,, \qquad Z_2 = \sqrt{\sum_{j=1}^{n}x_j^2}\,\bigl(\widehat{\beta}-\beta\bigr)$$

  unabhängig und identisch $\mathrm{N}(0,\sigma^2)$-verteilt sind.
- Aus der Definition der χ^2-Verteilung ergibt sich nun, dass Z^2/σ^2 eine χ^2_2-verteilte Zufallsvariable ist.
- Die Gültigkeit von (31) folgt somit aus Lemma 5.2. $\square$

5.2 Tests und Konfidenzbereiche

5.2.1 t-Tests für Regressionskonstante und Regressionskoeffizient

- Für das einfache lineare Regressionsmodell mit normalverteilten Störgrößen kann man nun t-Tests zur Verifizierung von Hypothesen über die Regressionskonstante bzw. den Regressionskoeffizienten mit Hilfe von Theorem 5.5 herleiten.

- Die Schätzer $\widehat{\alpha}$, $\widehat{\beta}$ und S^2 seien so wie bisher durch (16) bzw. (26) gegeben, d.h.

$$\widehat{\beta} = \frac{s^2_{xY}}{s^2_{xx}}\,, \qquad \widehat{\alpha} = \overline{Y}_n - \widehat{\beta}\overline{x}_n\,, \qquad S^2 = \frac{1}{n-2}\sum_{i=1}^n \bigl(Y_i - \widehat{\alpha} - \widehat{\beta}x_i\bigr)^2 .$$

- Aus den Verteilungs- und Unabhängigkeitseigenschaften, die in Theorem 5.5 hergeleitet worden sind, ergibt sich nun unmittelbar, dass

$$\frac{\widehat{\alpha}-\alpha}{S\sqrt{\Bigl(\sum\limits_{i=1}^n x_i^2\Bigr)\Big/\bigl(n(n-1)s^2_{xx}\bigr)}} \sim \mathrm{t}_{n-2} \tag{36}$$

und

$$\frac{\widehat{\beta}-\beta}{S\big/\sqrt{(n-1)s^2_{xx}}} \sim \mathrm{t}_{n-2}\,. \tag{37}$$

- Beim Test der Hypothese $H_0 : \alpha = \alpha_0$ zum Niveau $1-\gamma \in (0,1)$ (gegen die Alternative $H_1 : \alpha \neq \alpha_0$) wird die Nullhypothese H_0 abgelehnt, falls

$$\frac{|\,\widehat{\alpha}-\alpha_0\,|}{S\sqrt{\Bigl(\sum\limits_{i=1}^n x_i^2\Bigr)\Big/\bigl(n(n-1)s^2_{xx}\bigr)}} > \mathrm{t}_{n-2,1-(1-\gamma)/2}\,, \tag{38}$$

wobei $\mathrm{t}_{n-2,1-(1-\gamma)/2}$ das $(1-(1-\gamma)/2)$-Quantil der t-Verteilung mit $n-2$ Freiheitsgraden bezeichnet.

- Analog wird beim Test der Hypothese $H_0 : \beta = \beta_0$ zum Niveau $1-\gamma \in (0,1)$ (gegen die Alternative $H_1 : \beta \neq \beta_0$) die Nullhypothese H_0 abgelehnt, falls

$$\frac{|\,\widehat{\beta}-\beta_0\,|}{S\big/\sqrt{(n-1)s^2_{xx}}} > \mathrm{t}_{n-2,1-(1-\gamma)/2}\,. \tag{39}$$

Beachte

- Von besonderem Interesse ist der Test der Hypothese $H_0 : \beta = 0$ (gegen die Alternative $H_1 : \beta \neq 0$).
- Hierbei wird die Nullhypothese H_0 abgelehnt, falls

$$\frac{|\,\widehat{\beta}\,|}{S\big/\sqrt{(n-1)s^2_{xx}}} > \mathrm{t}_{n-2,1-(1-\gamma)/2}\,. \tag{40}$$

Beispiel (vgl. L.J. Kazmier (1999) *Wirtschaftsstatistik.* McGraw-Hill, S. 256ff.)

- Eine Speditionsfirma will anhand von 10 zufällig ausgewählten Lkw-Lieferungen untersuchen, ob ein bzw. welcher Zusammenhang zwischen der Länge des Transportweges (in km) und der Lieferzeit (in Tagen) von der Abholbereitstellung bis zum Eintreffen der Lieferung beim Empfänger besteht.
- Es wurden die folgenden Daten erhoben:

Nummer der Lieferung	1	2	3	4	5	6	7	8	9	10
Weglänge (in km)	825	215	1070	550	480	920	1350	325	670	1215
Lieferzeit (in Tagen)	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

- Dabei werden die Weglängen als Ausgangsvariablen (X) und die Lieferzeiten als Zielvariablen (Y) aufgefasst.

- Aus diesen Daten ergeben sich gemäß (16) die folgenden Schätzwerte für den Regressionskoeffizienten β bzw. Regressionskonstante α:

$$\widehat{\beta} = \frac{s^2_{xy}}{s^2_{xx}} = 0.0036\,, \qquad \widehat{\alpha} = \overline{y}_{10} - \widehat{\beta}\overline{x}_{10} = 0.11\,.$$

- Die geschätzte Regressionsgerade hat somit die Gestalt: $\widehat{y} = 0.11 + 0.0036x$.
- Hieraus ergeben sich die geschätzten Lieferzeiten $\widehat{y}_i$ bzw. die Residuen $\widehat{\varepsilon}_i$ für die beobachteten Weglängen x_i:

Nummer der Lieferung	1	2	3	4	5	6	7	8	9	10
beobachtete Lieferzeit	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0
geschätzte Lieferzeit	3.08	0.88	3.96	2.09	1.84	3.42	4.97	1.28	2.52	4.48
Residuum	0.42	0.12	0.04	-0.09	-0.84	-0.42	-0.47	0.22	0.48	0.52

- Als Schätzwert der Varianz σ^2 der Störgrößen $\varepsilon_1, \ldots, \varepsilon_{10}$ ergibt sich somit gemäß (26)

$$S^2 = \frac{1}{8}\sum_{i=1}^{10} \widehat{\varepsilon}_i^2 \approx 0.48^2\,.$$

- Um zu prüfen, ob überhaupt ein (signifikanter) Zusammenhang zwischen der Länge des Transportweges und der Lieferzeit besteht, soll nun die Hypothese $H_0 : \beta = 0$ (gegen die Alternative $H_1 : \beta \neq 0$) zum Niveau $1 - \gamma = 0.05$ getestet werden.
- Aus den beobachteten Daten ergibt sich, dass

$$\overline{x}_{10} = 762\,, \quad \sum_{i=1}^{10} x_i^2 = 7104300\,, \quad \sqrt{\sum_{i=1}^{10} x_i^2 - 10\,\overline{x}^2_{10}} = 1139.24$$

  und somit

$$\frac{|\,\widehat{\beta}\,|}{S/\sqrt{\sum_{i=1}^{10} x_i^2 - 10\,\overline{x}^2_{10}}} = \frac{0.0036}{0.48/1139.24} = \frac{0.0036}{0.0004} = 9.00\,.$$

- Andererseits gilt $\mathrm{t}_{8,0.975} = 2.306$.
- Gemäß (40) wird also die Hypothese $H_0 : \beta = 0$ abgelehnt, d.h., es besteht ein signifikanter Zusammenhang zwischen der Länge des Transportweges und der Lieferzeit.

5.2.2 Konfidenzintervalle; Prognose von Zielwerten

- Aus (38) und (39) ergeben sich ohne weiteres die folgenden Konfindenzintervalle zum Niveau $\gamma \in (0,1)$ für die Modellparameter α und β.
- Und zwar gilt jeweils mit Wahrscheinlichkeit γ

$$\widehat{\alpha} - \mathrm{t}_{n-2,1-(1-\gamma)/2} S\sqrt{\Bigl(\sum_{i=1}^{n} x_i^2\Bigr)\Big/\bigl(n(n-1)s^2_{xx}\bigr)} < \alpha < \widehat{\alpha} + \mathrm{t}_{n-2,1-(1-\gamma)/2} S\sqrt{\Bigl(\sum_{i=1}^{n} x_i^2\Bigr)\Big/\bigl(n(n-1)s^2_{xx}\bigr)} \tag{41}$$

und

$$\widehat{\beta} - \mathrm{t}_{n-2,1-(1-\gamma)/2} S\big/\sqrt{(n-1)s^2_{xx}} < \beta < \widehat{\beta} + \mathrm{t}_{n-2,1-(1-\gamma)/2} S\big/\sqrt{(n-1)s^2_{xx}}\,. \tag{42}$$

Beachte

- Auf ähnliche Weise kann man auch ein Konfidenzintervall zum Niveau $\gamma \in (0,1)$ für den erwarteten Zielwert $\alpha + \beta x_0$ herleiten, der einem vorgegebenen Wert $x_0 \in \mathbb{R}$ der Ausgangsvariable X entspricht.
- Von besonderem Interesse ist dabei natürlich der Fall, dass $x_0 \notin \{x_1, \ldots, x_n\}$, d.h., wenn an der Stelle x_0 *keine* Daten erhoben werden.

Theorem 5.6 *Sei $x_0 \in \mathbb{R}$ eine beliebige reelle Zahl.*

1. *Durch den Ansatz $\widehat{\alpha} + \widehat{\beta} x_0$ ist ein erwartungstreuer Schätzer für $\alpha + \beta x_0$ gegeben mit*

$$\widehat{\alpha} + \widehat{\beta} x_0 \sim \mathrm{N}\Bigl(\alpha + \beta x_0 \, , \, \sigma^2 \Bigl(\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Bigr)\Bigr) . \tag{43}$$

2. *Mit Wahrscheinlichkeit γ gilt*

$$\widehat{\alpha} + \widehat{\beta} x_0 - \mathrm{t}_{n-2,1-(1-\gamma)/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}} < \alpha + \beta x_0 < \widehat{\alpha} + \widehat{\beta} x_0 + \mathrm{t}_{n-2,1-(1-\gamma)/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}} . \tag{44}$$

Beweis

- Weil $\widehat{\alpha}$ und $\widehat{\beta}$ erwartungstreue Schätzer für α bzw. β sind, ergibt sich aus der Linearität des Erwartungswertes, dass $\mathbb{E}\bigl(\widehat{\alpha} + \widehat{\beta} x_0\bigr) = \mathbb{E}\,\widehat{\alpha} + x_0 \mathbb{E}\,\widehat{\beta} = \alpha + \beta x_0$.
- Außerdem ergibt sich aus (29)–(30) und aus den allgemeinen Rechenregeln der Varianz, dass

$$\begin{aligned}
\mathrm{Var}\bigl(\widehat{\alpha} + \widehat{\beta} x_0\bigr) &= \mathrm{Var}\,\widehat{\alpha} + x_0^2 \mathrm{Var}\,\widehat{\beta} + 2x_0 \mathrm{Cov}\bigl(\widehat{\alpha}, \widehat{\beta}\bigr) \\
&= \frac{\sigma^2}{n(n-1)s_{xx}^2} \sum_{i=1}^n x_i^2 + \frac{\sigma^2 x_0^2}{(n-1)s_{xx}^2} - \frac{2\sigma^2 x_0 \overline{x}_n}{(n-1)s_{xx}^2} \\
&= \frac{\sigma^2}{(n-1)s_{xx}^2} \Bigl(\frac{1}{n} \sum_{i=1}^n x_i^2 - \overline{x}_n^2 + \overline{x}_n^2 - 2x_0 \overline{x}_n + x_0^2\Bigr) \\
&= \frac{\sigma^2}{(n-1)s_{xx}^2} \Bigl(\frac{1}{n} \sum_{i=1}^n (x_i^2 - \overline{x}_n^2) + (x_0 - \overline{x}_n)^2\Bigr) \\
&= \sigma^2 \Bigl(\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Bigr) .
\end{aligned}$$

- Weil $\widehat{\alpha} + \widehat{\beta} x_0$ eine Linearkombination der unabhängigen und normalverteilten Zielvariablen $Y_1, \ldots, Y_n$ ist, folgt aus Korollar WR-3.2, dass $\widehat{\alpha} + \widehat{\beta} x_0$ ebenfalls normalverteilt ist.
- Damit ist die Gültigkeit von Teilaussage 1 bewiesen.
- In Theorem 5.5 hatten wir gezeigt, dass die Zufallsvariablen $(\widehat{\alpha}, \widehat{\beta})$ und S^2 unabhängig sind.
- Folglich sind auch die Zufallsvariablen $\widehat{\alpha} + \widehat{\beta} x_0$ und S^2 unabhängig, vgl. Theorem WR-3.18.
- Aus Teilaussage 1, aus (31) und aus der Definition der t-Verteilung ergibt sich nun, dass

$$\frac{\widehat{\alpha} + \widehat{\beta} x_0 - (\alpha + \beta x_0)}{S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}}} \sim \mathrm{t}_{n-2} .$$

- Hieraus ergibt sich die Gültigkeit von (44). □

Beachte

- Bei gegebenen (Ausgangs- und Ziel-) Daten $(x_1, y_1), \ldots, (x_n, y_n)$ ist die Länge des in (44) hergeleiteten Konfidenzintervalls für $\alpha + \beta x_0$ eine monoton nichtfallende Funktion des Abstandes $|x_0 - \overline{x}_n|$, die ihr Minimum im Punkt $x_0 = \overline{x}_n$ annimmt.
- Die in (44) hergeleitete Intervallschätzung für $\alpha + \beta x_0$ ist also dann am genauesten, wenn der Wert x_0 im „Zentrum" der (Ausgangs-) Werte $x_1, \ldots, x_n$ liegt.
- *Mit anderen Worten*: In (44) wurde ein zufälliges Intervall hergeleitet, in dem
  - der (unbekannte) Erwartungswert $\mathbb{E}\, Y_0 = \alpha + \beta x_0$ der Zufallsvariable $Y_0 = \alpha + \beta x_0 + \varepsilon_0$ mit der (vorgegebenen) Wahrscheinlichkeit γ liegt, wobei
  - die Störgröße ε_0 normalverteilt und unabhängig von den Störgrößen $\varepsilon_1, \ldots, \varepsilon_n$ ist; $\varepsilon_0 \sim \mathrm{N}(0, \sigma^2)$.

Wir bestimmen nun ein zufälliges Intervall, in dem *nicht* der Erwartungswert $\mathbb{E}\, Y_0$, sondern die Zielgröße Y_0 selbst mit der Wahrscheinlichkeit γ liegt.

Definition

- Seien $L, U : \Omega \to \mathbb{R}$ zwei Zufallsvariablen, so dass $\mathbb{P}(L \le U) = 1$ und $\mathbb{P}(L < Y_0 < U) \ge \gamma$.
- Dann sagt man, dass das zufällige Intervall (L, U) ein *Prognoseintervall* für die Zielvariable Y_0 zum Niveau γ ist.

Theorem 5.7 *Mit Wahrscheinlichkeit γ gilt*

$$\widehat{\alpha} + \widehat{\beta} x_0 - \mathrm{t}_{n-2,1-(1-\gamma)/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}} < Y_0 < \widehat{\alpha} + \widehat{\beta} x_0 + \mathrm{t}_{n-2,1-(1-\gamma)/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}} \,. \tag{45}$$

Beweis

- Offenbar gilt

$$\begin{aligned} \mathbb{E}\big(Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)\big) &= \mathbb{E}\, Y_0 - \mathbb{E}\big(\widehat{\alpha} + \widehat{\beta} x_0\big) \\ &= \alpha + \beta x_0 - (\alpha + \beta x_0) = 0 \,. \end{aligned}$$

- Aus der Unabhängigkeit der Störgrößen $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ folgt, dass auch die Zufallsvariablen Y_0 und $\big(\widehat{\alpha}, \widehat{\beta}, S^2\big)$ unabhängig sind.
- Somit gilt

$$\begin{aligned} \mathrm{Var}\big(Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)\big) &= \mathrm{Var}\, Y_0 + \mathrm{Var}\big(\widehat{\alpha} + \widehat{\beta} x_0\big) \\ &= \sigma^2 + \sigma^2 \Big(\frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Big) \,. \end{aligned}$$

- Weil außerdem $Y_0 \sim \mathrm{N}(\alpha + \beta x_0, \sigma^2)$, ergibt sich hieraus und aus (43), dass

$$Y_0 - \big(\widehat{\alpha} + \widehat{\beta} x_0\big) \sim \mathrm{N}\Big(0, \sigma^2 \Big(1 + \frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Big)\Big) \,.$$

- Aus (31) und aus der Definition der t-Verteilung ergibt sich nun, dass

$$\frac{Y_0 - \big(\widehat{\alpha} + \widehat{\beta} x_0\big)}{S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x}_n)^2}{(n-1)s_{xx}^2}}} \sim \mathrm{t}_{n-2} \,.$$

- Damit ist die Behauptung bewiesen. $\square$

5.2.3 Simultane Konfidenzbereiche; Konfidenzbänder

- Auf ähnliche Weise wie in Theorem 5.6 kann man sogenannte simultane Konfidenzbereiche zum Niveau $\gamma \in (0,1)$ *gleichzeitig für meherere* erwartete Zielwerte $\alpha + \beta x_{0i}$, $i = 1, \ldots, m$ herleiten, die vorgegebenen (Ausgangs-) Werten $x_{01}, \ldots, x_{0m} \in \mathbb{R}$ entsprechen.
- Dabei ist erneut der Fall $x_{01}, \ldots, x_{0m} \not\in \{x_1, \ldots, x_n\}$ von besonderem Interesse, d.h., wenn an den Stellen $x_{01}, \ldots, x_{0m}$ keine Daten erhoben werden.

Hierfür ist die folgende *Bonferroni-Ungleichung* nützlich.

Lemma 5.4 *Für jede natürliche Zahl $m = 1, 2, \ldots$ und für beliebige Ereignisse $A_1, \ldots, A_m \in \mathcal{F}$ gilt*

$$\mathbb{P}\Bigl(\bigcap_{i=1}^m A_i\Bigr) \ge \sum_{i=1}^m \mathbb{P}(A_i) - (m-1)\,. \tag{46}$$

Beweis Aus der Subadditivität von Wahrscheinlichkeitsmaßen (vgl. Theorem WR-2.2) ergibt sich, dass

$$\begin{aligned}
\mathbb{P}\Bigl(\bigcap_{i=1}^m A_i\Bigr) &= 1 - \mathbb{P}\Bigl(\bigcup_{i=1}^m A_i^c\Bigr) \\
&\ge 1 - \sum_{i=1}^m \mathbb{P}(A_i^c) \\
&= \sum_{i=1}^m \mathbb{P}(A_i) + 1 - \sum_{i=1}^m \bigl(\mathbb{P}(A_i) + \mathbb{P}(A_i^c)\bigr) \\
&= \sum_{i=1}^m \mathbb{P}(A_i) + 1 - m\,.
\end{aligned}$$

□

Theorem 5.8 *Die Wahrscheinlichkeit, dass*

$$\widehat{\alpha} + \widehat{\beta} x_{0i} - \mathrm{t}_{n-2,1-(1-\gamma)/(2m)} S \sqrt{\frac{1}{n} + \frac{(x_{0i} - \overline{x}_n)^2}{(n-1)s_{xx}^2}} < \alpha + \beta x_{0i} < \widehat{\alpha} + \widehat{\beta} x_{0i} + \mathrm{t}_{n-2,1-(1-\gamma)/(2m)} S \sqrt{\frac{1}{n} + \frac{(x_{0i} - \overline{x}_n)^2}{(n-1)s_{xx}^2}} \tag{47}$$

gleichzeitig für jedes $i = 1, \ldots, m$ gilt, ist mindestens gleich γ.

Beweis

- Gemäß Teilaussage 2 von Theorem 5.6 ist die Wahrscheinlichkeit für jedes der m Ereignisse $A_1, \ldots, A_m$ in (47) gegeben durch
$$\gamma' = 1 - \frac{1-\gamma}{m}\,.$$
- Aus Lemma 5.4 ergibt sich nun für die Wahrscheinlichkeit des Durchschnittes $A_1 \cap \ldots \cap A_m$, dass

$$\begin{aligned}
\mathbb{P}\Bigl(\bigcap_{i=1}^m A_i\Bigr) &\ge \sum_{i=1}^m \mathbb{P}(A_i) - (m-1) \\
&= m\Bigl(1 - \frac{1-\gamma}{m}\Bigr) - (m-1) \\
&= \gamma\,.
\end{aligned}$$

□

Beachte

- Das kartesische Produkt der m Intervalle in (47) heißt *simultaner Konfidenzbereich* für den Vektor $(\alpha + \beta x_{01}, \ldots, \alpha + \beta x_{0m})$.
- Wir können noch einen Schritt weiter gehen als in Theorem 5.8 und fragen, ob es eine Zahl $a_\gamma > 0$ gibt, so dass die Wahrscheinlichkeit mindestens gleich γ ist, dass
$$\widehat{\alpha} + \widehat{\beta} x - a_\gamma S \sqrt{\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}} < \alpha + \beta x < \widehat{\alpha} + \widehat{\beta} x + a_\gamma S \sqrt{\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}} \tag{48}$$
gleichzeitig für jedes $x \in \mathbb{R}$ gilt.
- Die Menge $B_\gamma \subset \mathbb{R}^2$ mit
$$B_\gamma = \left\{ (x,y) \in \mathbb{R}^2 \, : \, \widehat{\alpha} + \widehat{\beta} x - a_\gamma S \sqrt{\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}} < y < \widehat{\alpha} + \widehat{\beta} x + a_\gamma S \sqrt{\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}} \right\} \tag{49}$$
heißt dann *Konfidenzband* zum Niveau γ für die Regressionsgerade $y = \alpha + \beta x$.

Bei der Lösung dieser Fragestellung ist die Klasse der F-Verteilungen in Verbindung mit dem folgenden Hilfssatz nützlich.

Lemma 5.5 *Seien $a, b, c, d \in \mathbb{R}$ beliebige reelle Zahlen mit $a \neq 0$, $c > 0$ und $d > 0$. Dann gilt*
$$\max_{x \in \mathbb{R}} \frac{(a + bx)^2}{c + dx^2} = \frac{a^2}{c} + \frac{b^2}{d} \,. \tag{50}$$
.

Beweis

- Die Behauptung gilt genau dann, wenn für alle $x \in \mathbb{R}$
$$(a + bx)^2 \le a^2 + \frac{a^2 d x^2}{c} + b^2 x^2 + \frac{c b^2}{d}$$
und wenn es ein $x \in \mathbb{R}$ gibt, so dass beide Seiten dieser Ungleichung übereinstimmen.
- Die Ungleichung gilt genau dann, wenn
$$2abx \le \frac{a^2 d x^2}{c} + \frac{c b^2}{d} \qquad \text{bzw.} \qquad 2abcdx \le (adx)^2 + (cb)^2 \,.$$
- Die letzte Ungleichung ergibt sich unmittelbar aus der binomischen Formel, und für $x = cb/(ad)$ gilt die Gleichheit. $\square$

Theorem 5.9 *Sei $a_\gamma = \sqrt{2\,\mathrm{F}_{2,n-2,\gamma}}$. Dann ist durch die in (49) definierte Menge $B_\gamma \subset \mathbb{R}^2$ ein Konfidenzband zum Niveau γ für die Regressionsgerade $y = \alpha + \beta x$ gegeben.*

Beweis

- Die Behauptung ist bewiesen, wenn wir zeigen, dass
$$\mathbb{P}\left(\frac{\bigl((\widehat{\alpha} + \widehat{\beta} x) - (\alpha + \beta x)\bigr)^2}{S^2 \Bigl(\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Bigr)} \le a_\gamma^2 \text{ für jedes } x \in \mathbb{R} \right) = \gamma$$
bzw.
$$\mathbb{P}\left(\max_{x \in \mathbb{R}} \frac{\bigl((\widehat{\alpha} + \widehat{\beta} x) - (\alpha + \beta x)\bigr)^2}{S^2 \Bigl(\frac{1}{n} + \frac{(x - \overline{x}_n)^2}{(n-1)s_{xx}^2}\Bigr)} \le a_\gamma^2 \right) = \gamma \,. \tag{51}$$

- Aus (16) ergibt sich, dass $\widehat{\alpha}+\widehat{\beta}x = \overline{Y}_n + \widehat{\beta}(x-\overline{x}_n)$. Weil $\widehat{\alpha}$ und $\widehat{\beta}$ erwartungstreue Schätzer für α bzw. β sind, gilt außerdem, dass $\alpha+\beta x = \mathbb{E}\,\overline{Y}_n + \beta(x-\overline{x}_n)$.
- Wir führen deshalb die gleiche Skalenverschiebung durch, die wir bereits im Beweis von Theorem 5.5 betrachtet haben:
$$x_i' = x_i - \overline{x}_n\,, \qquad \forall\, i=1,\ldots,n\,,$$
d.h. wir nehmen (o.B.d.A.) an, dass $\overline{x}_n = 0$.
- Mit anderen Worten: Anstelle (51) zeigen wir, dass
$$\mathbb{P}\left(\max_{x\in\mathbb{R}} \frac{\bigl((\overline{Y}_n - \mathbb{E}\,\overline{Y}_n) + (\widehat{\beta}-\beta)x\bigr)^2}{S^2\Bigl(\dfrac{1}{n} + \dfrac{x^2}{(n-1)s_{xx}^2}\Bigr)} \le a_\gamma^2\right) = \gamma\,.$$
- Weil S^2 nicht von x abhängt, ergibt sich aus Lemma 5.5, dass
$$\begin{aligned}\max_{x\in\mathbb{R}} \frac{\bigl((\overline{Y}_n - \mathbb{E}\,\overline{Y}_n) + (\widehat{\beta}-\beta)x\bigr)^2}{S^2\Bigl(\dfrac{1}{n} + \dfrac{x^2}{(n-1)s_{xx}^2}\Bigr)} &= \frac{n\bigl(\overline{Y}_n - \mathbb{E}\,\overline{Y}_n\bigr)^2 + (n-1)s_{xx}^2\bigl(\widehat{\beta}-\beta\bigr)^2}{S^2}\\ &= \frac{\dfrac{\bigl(\overline{Y}_n - \mathbb{E}\,\overline{Y}_n\bigr)^2}{\sigma^2/n} + \dfrac{\bigl(\widehat{\beta}-\beta\bigr)^2}{\sigma^2/((n-1)s_{xx}^2)}}{S^2/\sigma^2}\,.\end{aligned}$$
- Weil $\overline{x}_n = 0$ und weil demzufolge $\overline{Y}_n = \widehat{\alpha}$ gilt, ergibt sich aus Theorem 5.5, dass die Zufallsvariablen $\overline{Y}_n$ und $\widehat{\beta}$ normalverteilt und unkorreliert (und somit gemäß Teilaussage 2 von Lemma 5.3 auch unabhängig) sind, vgl. auch Übungsaufgabe 2.3.
- Der Zähler des letzten Quotienten ist also die Summe der Quadrate von zwei unabhängigen $\mathrm{N}(0,1)$-verteilten Zufallsvariablen, d.h., der Zähler ist χ_2^2-verteilt.
- Außerdem ergibt sich aus Theorem 5.5, dass Zähler und Nenner unabhängig sind und dass $(n-2)S^2/\sigma^2 \sim \chi_{n-2}^2$.
- Hieraus und aus der Definition der F-Verteilung folgt, dass (51) gilt, falls
$$\frac{a_\gamma^2}{2} = \mathrm{F}_{2,n-2,\gamma}\,.$$
□

Beachte

- Es gilt $\mathrm{t}_{n-2,1-(1-\gamma)/(2m)} > \sqrt{2\,\mathrm{F}_{2,n-2,\gamma}}$ für jedes hinreichend große m.
- Hieraus folgt, dass der simultane Konfidenzbereich, der in Theorem 5.8 mit Hilfe der Bonferroni-Ungleichung (46) hergeleitet worden ist, für große m *größer* ist als der simultane Konfidenzbereich, der sich aus dem in Theorem 5.9 konstruierten Konfidenzband ergibt.
- Auf den ersten Blick scheint dies ein Widerspruch zu sein, weil in Theorem 5.9 die Überdeckungseigenschaft für *alle* $x \in \mathbb{R}$ gefordert wird, während diese Eigenschaft in Theorem 5.8 nur für *endlich viele* Ausgangswerte $x_{01},\ldots,x_{0m}$ betrachtet wird.
- Der Grund, dass Theorem 5.9 für große m zu kleineren (d.h. besseren) simultanen Konfidenzbereichen führt, besteht darin, dass die Bonferroni-Ungleichung (46) für große m nur eine sehr ungenaue untere Schranke für die Wahrscheinlichkeit $\mathbb{P}\Bigl(\bigcap_{i=1}^m A_i\Bigr)$ liefert.

6 Tabellen für Verteilungsfunktionen und Quantile

Tabelle 1 Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung

x	0	1	2	3	4	5	6	7	8	9
0,0	0,500000	0,503989	0,507978	0,511967	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621719	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705402	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,878999	0,881000	0,882977
1,2	0,884930	0,886860	0,888767	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903199	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935744	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959071	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965621	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971284	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995603	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,0	0,998650	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,998930	0,998965	0,998999
3,5	0,999767	0,999776	0,999784	0,999792	0,999800	0,999807	0,999815	0,999821	0,999828	0,999835
4,0	0,999968	0,999970	0,999971	0,999972	0,999973	0,999974	0,999975	0,999976	0,999977	0,999978

Tabelle 2 γ-Quantil $\chi^2_{r,\alpha}$ der χ^2-Verteilung mit r Freiheitsgraden

$r\backslash\alpha$	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
1	.00004	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.6	5.23	6.26	7.26	8.55	22.31	25	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

Tabelle 3 γ-Quantil $t_{r,\alpha}$ der t-Verteilung mit r Freiheitsgraden

r \ α	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656
2	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660
70	0,387	0,527	0,678	0,847	1,044	1,294	1,667	1,994	2,381	2,648
80	0,387	0,526	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639
90	0,387	0,526	0,677	0,846	1,042	1,291	1,662	1,987	2,368	2,632
100	0,386	0,526	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626
150	0,386	0,526	0,676	0,844	1,040	1,287	1,655	1,976	2,351	2,609
200	0,386	0,525	0,676	0,843	1,039	1,286	1,653	1,972	2,345	2,601
500	0,386	0,525	0,675	0,842	1,038	1,283	1,648	1,965	2,334	2,586
1000	0,385	0,525	0,675	0,842	1,037	1,282	1,646	1,962	2,330	2,581

Tabelle 4a γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r, s) Freiheitsgraden

$s \setminus r$	1	2	3	4	5	6	7	8	9	10	11	12	α
1	161	200	216	225	230	234	237	239	241	242	243	244	0,95
	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106	0,99
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	0,95
	98,49	99,00	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42	0,99
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	0,95
	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05	26,92	0,99
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	0,95
	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37	0,99
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	0,95
	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89	0,99
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	0,95
	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	0,99
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	0,95
	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47	0,99
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	0,95
	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67	0,99
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	0,95
	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11	0,99
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	0,95
	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71	0,99
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	0,95
	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40	0,99
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	0,95
	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16	0,99

Tabelle 4b γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r, s) Freiheitsgraden

$s \setminus r$	14	16	20	24	30	40	50	75	100	200	500	∞	α
1	245	246	248	249	250	251	252	253	253	254	254	254	0,95
	6142	6169	6208	6234	6258	6286	6302	6323	6334	6352	6361	6366	0,99
2	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	0,95
	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50	0,99
3	8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	0,95
	26,83	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12	34,12	0,99
4	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	0,95
	14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46	0,99
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	0,95
	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02	0,99
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	0,95
	7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88	0,99
7	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	0,95
	6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65	0,99
8	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	0,95
	5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86	0,99
9	3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71	0,95
	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31	0,99
10	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	0,95
	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91	0,99
11	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40	0,95
	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60	0,99
12	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30	0,95
	4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36	0,99

Tabelle 4c γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r,s) Freiheitsgraden

$s \setminus r$	1	2	3	4	5	6	7	8	9	10	11	12	α
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	0,95
	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	0,99
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	0,95
	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	0,99
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	0,95
	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	0,99
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	0,95
	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55	0,99
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	0,95
	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45	0,99
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	0,95
	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,44	3,37	0,99
19	4,38	3,51	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	0,95
	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	0,99
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	0,95
	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23	0,99
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	0,95
	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,24	3,17	0,99
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	0,95
	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	0,99
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	0,95
	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	0,99
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	0,95
	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,09	3,03	0,99
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16	0,95
	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	3,05	2,99	0,99

Tabelle 4d γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r, s) Freiheitsgraden

$s \setminus r$	14	16	20	24	30	40	50	75	100	200	500	∞	α
13	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21	0,95
	3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16	0,99
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13	0,95
	3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00	0,99
15	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07	0,95
	3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87	0,99
16	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01	0,95
	3,45	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75	0,99
17	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96	0,95
	3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65	0,99
18	2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92	0,95
	3,27	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57	0,99
19	2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88	0,95
	3,19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49	0,99
20	2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84	0,95
	3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42	0,99
21	2,20	2,15	2,09	2,05	2,00	1,96	1,93	1,89	1,87	1,84	1,82	1,81	0,95
	3,07	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36	0,99
22	2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78	0,95
	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31	0,99
23	2,14	2,10	2,05	2,00	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76	0,95
	2,97	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26	0,99
24	2,13	2,09	2,02	1,98	1,94	1,89	1,86	1,82	1,80	1,76	1,74	1,73	0,95
	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21	0,99
25	2,11	2,06	2,00	1,96	1,92	1,87	1,84	1,80	1,77	1,74	1,72	1,71	0,95
	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17	0,99

Tabelle 4e γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r, s) Freiheitsgraden

$s \setminus r$	1	2	3	4	5	6	7	8	9	10	11	12	α
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	0,95
	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	3,02	2,96	0,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	0,95
	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,98	2,93	0,99
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	0,95
	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,95	2,90	0,99
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	0,95
	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,08	3,00	2,92	2,87	0,99
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	0,95
	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,90	2,84	0,99
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	0,95
	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,73	2,66	0,99
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95	0,95
	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,62	2,56	0,99
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	0,95
	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	0,99
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89	0,95
	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,77	2,67	2,59	2,51	2,45	0,99
80	3,96	3,11	2,72	2,48	2,33	2,21	2,12	2,05	1,99	1,95	1,91	1,88	0,95
	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,41	0,99
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	0,95
	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,43	2,36	0,99
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80	0,95
	6,76	4,71	3,88	3,41	3,11	2,90	2,73	2,60	2,50	2,41	2,34	2,28	0,99
1000	3,85	3,00	2,61	2,38	2,22	2,10	2,02	1,95	1,89	1,84	1,80	1,76	0,95
	6,66	4,62	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,26	2,20	0,99

Tabelle 4f γ-Quantil $F_{r,s,\alpha}$ der F-Verteilung mit (r, s) Freiheitsgraden

$s \setminus r$	14	16	20	24	30	40	50	75	100	200	500	∞	α
26	2,10	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,72	1,70	1,69	0,95
	2,86	2,77	2,66	2,58	2,50	2,41	2,36	2,28	2,25	2,19	2,15	2,13	0,99
27	2,08	2,03	1,97	1,93	1,88	1,84	1,80	1,76	1,74	1,71	1,68	1,67	0,95
	2,83	2,74	2,63	2,55	2,47	2,38	2,33	2,25	2,21	2,16	2,12	2,10	0,99
28	2,06	2,02	1,96	1,91	1,87	1,81	1,78	1,75	1,72	1,69	1,67	1,65	0,95
	2,80	2,71	2,60	2,52	2,44	2,35	2,30	2,22	2,18	2,13	2,09	2,06	0,99
29	2,05	2,00	1,94	1,90	1,85	1,80	1,77	1,73	1,71	1,68	1,65	1,64	0,95
	2,77	2,68	2,57	2,49	2,41	2,32	2,27	2,19	2,15	2,10	2,06	2,03	0,99
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62	0,95
	2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01	0,99
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51	0,95
	2,56	2,49	2,37	2,29	2,20	2,11	2,05	1,97	1,94	1,88	1,84	1,81	0,99
50	1,90	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44	0,95
	2,46	2,39	2,26	2,18	2,10	2,00	1,94	1,86	1,82	1,76	1,71	1,68	0,99
60	1,86	1,81	1,75	1,70	1,65	1,59	1,56	1,50	1,48	1,44	1,41	1,39	0,95
	2,40	2,32	2,20	2,12	2,03	1,93	1,87	1,79	1,74	1,68	1,63	1,60	0,99
70	1,84	1,79	1,72	1,67	1,62	1,56	1,53	1,47	1,45	1,40	1,37	1,35	0,95
	2,35	2,28	2,15	2,07	1,98	1,88	1,82	1,74	1,69	1,62	1,56	1,53	0,99
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,42	1,38	1,35	1,32	0,95
	2,32	2,24	2,11	2,03	1,94	1,84	1,78	1,70	1,65	1,57	1,52	1,49	0,99
100	1,79	1,75	1,68	1,63	1,57	1,51	1,48	1,42	1,39	1,34	1,30	1,28	0,95
	2,26	2,19	2,06	1,98	1,89	1,79	1,73	1,64	1,59	1,51	1,46	1,43	0,99
200	1,74	1,69	1,62	1,57	1,52	1,45	1,42	1,35	1,32	1,26	1,22	1,19	0,95
	2,17	2,09	1,97	1,88	1,79	1,69	1,62	1,53	1,48	1,39	1,33	1,28	0,99
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08	0,95
	2,09	2,01	1,89	1,81	1,71	1,61	1,54	1,44	1,38	1,28	1,19	1,11	0,99