Inhaltsverzeichnis

Kapitel VIII

Schließende Statistik

1 Statistische Modellbildung und statistisches Entscheiden

Grundform wichtiger statistischer Fragestellungen:

(i) Zufallsexperiment mit unbekannter Verteilung Q.

(ii) Verteilungsannahme: $Q \in \mathfrak{P}$ für eine Menge $\mathfrak{P}$ von Wahrscheinlichkeitsmaßen auf $\mathfrak{B}_d$.

(iii) (a) Schätzproblem. Gegeben: Abbildung $\eta : \mathfrak{P} \to \mathbb{R}$. Bestimme $\eta(Q)$, einen „Aspekt" der Verteilung Q.

(b) Testproblem. Gegeben $\emptyset \neq \mathfrak{P}_0 \subsetneq \mathfrak{P}$. Entscheide, ob $Q \in \mathfrak{P}_0$.

(iv) Verfügbar: Stichprobe $(x_1, \ldots, x_n) \in \mathbb{R}^{n \cdot d}$ aus n-maliger unabhängiger Wiederholung des Zufallsexperimentes.

Bezeichnung. Fortan bezeichnen wir mit $\mathbf{B}(n,p)$, $\mathbf{N}(\mu, \sigma^2)$, … auch die entsprechenden Wahrscheinlichkeitsmaße auf $\mathfrak{B}_1$.

Beispiel 1. Geschlecht eines Neugeborenen ($1 \triangleq$ W, $0 \triangleq$ M), siehe Beispiel I.1.13. Hier $d = 1$ und

(i) $Q = \mathbf{B}(1,p)$, wobei p die Wahrscheinlichkeit, daß Neugeborenes weiblich.

(ii) $\mathfrak{P} := \{\mathbf{B}(1,p) : p \in]0,1[\}$.

(iii) (a) $\eta(\mathbf{B}(1,p)) := p$.

(b) $\mathfrak{P}_0 := \{\mathbf{B}(1, 1/2)\}$ oder $\mathfrak{P}_0 := \{\mathbf{B}(1,p) : p < 1/2\}$.

(iv) Geschlecht bei n Lebendgeburten.

Bei einem Stichprobenumfang von $n = 25\,171\,123$ scheint eine „verläßliche“ Bestimmung von $\eta(Q)$ und eine Entscheidung, ob $\eta(Q) \in \mathfrak{P}_0$, möglich.
Das empirische Mittel

$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{12\,241\,392}{25\,171\,123} = 0.4863\ldots$$

legt nahe, daß $\eta(Q)$ ungefähr 0.48 beträgt und daß $Q \neq \mathbf{B}(1, 1/2)$ gilt. Siehe Hesse (2003, p. 23).

Bemerkung 2. Studiert werden auch Varianten dieser Grundform, z. Bsp. abhängige Beobachtungen oder $\mathbb{R}^k$-wertige Abbildungen η. Die Modellierung und Analyse solcher Fragestellungen ist Gegenstand der Mathematischen Statistik (Gütekriterien, Optimalitätsaussagen, Quantifizierung von Risiken).

Bemerkung 3. Oft ist $\mathfrak{P}$ in natürlicher Weise parametrisiert, siehe Beispiel 1.

Fortan betrachten wir der Einfachheit halber den Fall $d = 1$.

Definition 4. Ein *statistisches Experiment* mit *Parameterraum* Θ und *Stichprobenraum* $\mathbb{R}^n$ ist gegeben durch

(i) eine Familie $(\Omega, \mathfrak{A}, P^\vartheta)_{\vartheta \in \Theta}$ von Wahrscheinlichkeitsräumen,

(ii) einen Zufallsvektor $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ mit folgenden Eigenschaften. Für alle $\vartheta \in \Theta$ ist

$$(X_1, \ldots, X_n) \text{ unabhängig bzgl. } P^\vartheta$$

und

$$P^\vartheta_{X_1} = \ldots = P^\vartheta_{X_n}.$$

Ferner ist $P^\vartheta_{X_1} \neq P^{\vartheta'}_{X_1}$ für alle $\vartheta, \vartheta' \in \Theta$ mit $\vartheta \neq \vartheta'$.

Bemerkung 5. In vorliegenden Kontext beschreibt die Menge

$$\mathfrak{P} = \{P^\vartheta_{X_1} : \vartheta \in \Theta\}$$

die sogenannte *Verteilungsannahme*. Die konkrete Gestalt von Ω, $\mathfrak{A}$ und den Wahrscheinlichkeitsmaßen P^ϑ ist im folgenden irrelevant. Die in Definition 4.(ii) geforderte Injektivität dient nur der mathematischen Bequemlichkeit, da sie die Identifikation der Verteilung $P^\vartheta_{X_1} \in \mathfrak{P}$ mit dem Parameter $\vartheta \in \Theta$ erlaubt.

Wir nehmen an, daß die vorliegenden Daten $(x_1, \ldots, x_n) \in \mathbb{R}^n$ eine Realisierung von $(X_1, \ldots, X_n)$ sind, d.h.

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)) \in \mathbb{R}^n$$

für ein $\omega \in \Omega$.

Definition 6. Gegeben sei ein statistisches Experiment mit Parameterraum Θ. Ein *Schätzproblem* ist definiert durch eine Abbildung

$$\gamma : \Theta \to \mathbb{R}.$$

Definition 7. Gegeben sei ein statistisches Experiment mit Stichprobenraum $\mathbb{R}^n$. Eine *Schätzfunktion* ist eine Borel-meßbare Abbildung

$$g_n : \mathbb{R}^n \to \mathbb{R}.$$

Bemerkung 8. Die Abbildung γ und ebenso η mit $\eta(P^{\vartheta}_{X_1}) = \gamma(\vartheta)$ beschreibt einen „Aspekt“ der Verteilungen $P^{\vartheta}_{X_1}$. Wichtige Beispiele sind der Erwartungswert und die Varianz.

Zur Lösung des Schätzproblems wählt man eine Schätzfunktion g_n und schätzt $\gamma(\vartheta)$ durch

$$g_n(x_1, \dots, x_n) = g_n(X_1(\omega), \dots, X_n(\omega)).$$

Ziel: Für jedes $\vartheta \in \Theta$ liegen die Werte der Zufallsvariable

$$g_n(X_1, \dots, X_n)$$

auf $(\Omega, \mathfrak{A}, P^{\vartheta})$ „nahe “ bei $\gamma(\vartheta)$.

Beispiel 9. Geschlecht eines Neugeborenen, siehe Beispiel 1. Modellierung durch

$$\Theta :=]0, 1[$$

und

$$P^{p}_{X_1} := \mathbf{B}(1, p)$$

für $p \in \Theta$. Das durch

$$\gamma(p) := p$$

definierte Schätzproblem beschreibt die Bestimmung der zugrundeliegenden Verteilung. Als Schätzfunktion für γ haben wir bereits

$$g_n(x_1, \dots, x_n) := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{1}$$

betrachtet.

Definition 10. Gegeben sei ein statistisches Experiment mit Parameterraum Θ. Ein *Testproblem* ist definiert durch eine Menge

$$\emptyset \neq \Theta_0 \subsetneq \Theta,$$

genannt *Hypothese*.

Definition 11. Gegeben sei ein statistisches Experiment mit Stichprobemraum $\mathbb{R}^n$. Ein *Verwerfungsbereich* ist eine Borel-Menge

$$R_n \in \mathfrak{B}_n.$$

Bemerkung 12. Die Hypothese Θ_0 definiert eine Teilmenge

$$\emptyset \neq \mathfrak{P}_0 = \{P_{X_1}^{\vartheta} : \vartheta \in \Theta_0\} \subsetneq \mathfrak{P}.$$

Zur Lösung des Testproblems wählt man einen Verwerfungsbereich R_n und lehnt die Hypothese Θ_0 (bzw. „$\vartheta \in \Theta_0$“ oder „$P_{X_1}^{\vartheta} \in \mathfrak{P}_0$“) genau dann ab, wenn

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)) \in R_n$$

gilt.

Ziel: Für jedes $\vartheta \in \Theta_0$ ist die Wahrscheinlichkeit

$$P^{\vartheta}(\{(X_1, \ldots, X_n) \in R_n\})$$

des *Fehlers 1. Art* „klein“, und für jedes $\vartheta \in \Theta \setminus \Theta_0$ ist die Wahrscheinlichkeit

$$P^{\vartheta}(\{(X_1, \ldots, X_n) \notin R_n\})$$

des *Fehlers 2. Art* „klein“. Siehe jedoch Bemerkung 3.9.(ii).

Die in den Bemerkungen 8 und 12 genannten Ziele bei Schätz- und Testproblem entsprechen jeweils einer worst case-Analyse über alle $\vartheta \in \Theta$.

Beispiel 13. Geschlecht eines Neugeborenen, siehe Beispiel 9. Die durch

$$\Theta_0 := \{1/2\}$$

oder

$$\Theta_0 :=]0, 1/2[$$

definierten Testprobleme betreffen die Hypothesen, daß Neugeborene mit gleicher Wahrscheinlichkeit weiblich wie männlich sind bzw. daß weibliche Lebendgeburten mit kleinerer Wahrscheinlichkeit auftreten.

Im Fall $\Theta_0 := \{1/2\}$ ist die Wahl eines Verwerfungsbereiche von der Form

$$R_n := \{x \in \mathbb{R}^n : |g_n(x) - 1/2| \geq k_n\}$$

mit g_n gemäß (1) und $k_n > 0$ naheliegend.

Beispiel 14. Analog: Halbwertszeit, siehe Beispiel VII.2.8. Hier gilt

$$\Theta :=]0, \infty[$$

und

$$P_{X_1}^{\lambda} := \mathbf{Exp}(\lambda)$$

für $\lambda \in \Theta$. Die Bestimmung der Halbwertszeit entspricht einem Schätzproblem mit

$$\gamma(\lambda) := \ln(2)/\lambda.$$

Die Entscheidung, ob die Halbwertszeit h_0 vorliegt, entspricht einem Testproblem mit

$$\Theta_0 := \{\ln(2)/h_0\}.$$

Ausblick: nicht-parametrische Statistik.

2 Schätzprobleme

Wir betrachten für ein statistisches Experiment, gegeben durch $(\Omega, \mathfrak{A}, P^{\vartheta})_{\vartheta \in \Theta}$ und $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, ein Schätzproblem, gegeben durch

$$\gamma : \Theta \to \mathbb{R}.$$

Bezeichnung. Wir setzen

$$\mathfrak{L}_1^{\vartheta} = \mathfrak{L}_1(\Omega, \mathfrak{A}, P^{\vartheta}), \qquad \mathfrak{L}_2^{\vartheta} = \mathfrak{L}_2(\Omega, \mathfrak{A}, P^{\vartheta})$$

und bezeichnen mit E^{ϑ} und Var^{ϑ} den Erwartungswert bzw. die Varianz bzgl. P^{ϑ}. Eine Eigenschaft gilt P^{ϑ}-fast sicher (f.s.), falls sie für alle ω aus einer Menge $A \in \mathfrak{A}$ mit $P^{\vartheta}(A) = 1$ gilt.

Definition 1. Die *Schätzvariable* zu einer Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ ist die Zufallsvariable $g_n(X)$.

Definition 2. Eine Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ heißt *erwartungstreu* für γ, falls

$$\forall \vartheta \in \Theta : \mathrm{E}^{\vartheta}(g_n(X)) = \gamma(\vartheta).$$

Eine Folge von Schätzfunktionen $g_n : \mathbb{R}^n \to \mathbb{R}$ heißt

(i) *schwach konsistent* für γ, falls

$$\forall \vartheta \in \Theta \; \forall \varepsilon > 0 : \lim_{n \to \infty} P^{\vartheta}(\{|g_n(X) - \gamma(\vartheta)| > \varepsilon\}) = 0.$$

(ii) *stark konsistent* für γ, falls

$$\forall \vartheta \in \Theta : \lim_{n\to\infty} g_n(X) = \gamma(\vartheta) \qquad P^\vartheta\text{-f.s.}$$

Bemerkung 3. Die starke Konsistenz impliziert die schwache Konsistenz, siehe Satz VII.2.12.

Wir untersuchen die Schätzung eines Erwartungswertes, siehe Beispiele 1.9 und 1.14.

Definition 4. Sei $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Dann heißt

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

empirisches Mittel der Stichprobe x.

Die zugehörige Schätzvariable ist

$$\overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^n X_i.$$

Satz 5. Gelte $X_1 \in \mathfrak{L}_1^\vartheta$ für alle $\vartheta \in \Theta$, und sei

$$\gamma(\vartheta) := \mathrm{E}^\vartheta(X_1).$$

Dann definieren die empirischen Mittel $g_n(x) := \overline{x}_n$ eine stark konsistente Folge von erwartungstreuen Schätzfunktionen für γ.

Beweis. Für jedes $\vartheta \in \Theta$ gilt

$$\mathrm{E}^\vartheta(g_n(X)) = \mathrm{E}^\vartheta(\overline{X}_n) = \frac{1}{n} \cdot \sum_{i=1}^n \mathrm{E}^\vartheta(X_i) = \gamma(\vartheta).$$

Die starke Konsistenz ist genau das starke Gesetz der großen Zahlen. □

Beispiel 6. In Beispiel 1.9 mit den Daten aus Beispiel 1.1 gilt für jedes $p \in \,]0,1[$ aufgrund der Tschebyschev-Ungleichung

$$P^p(\{|\overline{X}_n - p| \geq 10^{-2}\}) \leq 9.9 \cdot 10^{-5},$$
$$P^p(\{|\overline{X}_n - p| \geq 10^{-3}\}) \leq 9.9 \cdot 10^{-3}$$

und aufgrund der Hoeffdingschen Ungleichung, siehe Satz VII.1.5,

$$P^p(\{|\overline{X}_n - p| \geq 10^{-2}\}) \leq 9.2 \cdot 10^{-2187},$$
$$P^p(\{|\overline{X}_n - p| \geq 10^{-3}\}) \leq 2.7 \cdot 10^{-22}.$$

Wir untersuchen nun die Schätzung der Varianz.

Definition 7. Sei $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Dann heißt

$$s_n^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2$$

empirische Varianz der Stichprobe x.

Satz 8. Gelte $X_1 \in \mathfrak{L}_2^\vartheta$ für alle $\vartheta \in \Theta$, und sei

$$\gamma(\vartheta) := \mathrm{Var}^\vartheta(X_1).$$

Ferner sei $n > 1$. Dann definieren die empirischen Varianzen $v_n(x) := s_n^2$ eine stark konsistente Folge von erwartungstreuen Schätzern für γ.

Beweis. Setze $\mu(\vartheta) := \mathrm{E}^\vartheta(X_1)$. Es gilt

$$\begin{aligned} v_n(x) &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left((x_i - \mu(\vartheta)) + (\mu(\vartheta) - \overline{x}_n)\right)^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu(\vartheta))^2 - \frac{n}{n-1} \cdot (\mu(\vartheta) - \overline{x}_n)^2 \end{aligned}$$

Mit Satz VI.2.13 folgt

$$\begin{aligned} \mathrm{E}^\vartheta(v_n(X)) &= \frac{n}{n-1} \cdot \gamma(\vartheta) - \frac{n}{n-1} \cdot \mathrm{Var}^\vartheta(\overline{X}_n) \\ &= \frac{n}{n-1} \cdot \gamma(\vartheta) - \frac{1}{n-1} \cdot \gamma(\vartheta) = \gamma(\vartheta). \end{aligned}$$

Nach dem starken Gesetz der großen Zahlen gilt P^ϑ-f.s.

$$\lim_{n\to\infty} \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \mu(\vartheta))^2 = \mathrm{E}^\vartheta(X_1 - \mu(\vartheta))^2 = \mathrm{Var}^\vartheta(X_1)$$

und

$$\lim_{n\to\infty} \overline{X}_n = \mu(\vartheta).$$

Es folgt P^ϑ-f.s.

$$\lim_{n\to\infty} v_n(X) = \gamma(\vartheta).$$

□

Bemerkung 9. Für die empirische Varianz gilt

$$s_n^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}_n^2 \right),$$

wie der erste Teil des Beweises von Satz 8 mit $\mu(\vartheta) := 0$ zeigt.

Beispiel 10. In Beispiel 1.1 beträgt die empirische Varianz 0.2498....

Beispiel 11. Sei $\Theta := [0,1]$ und für $p \in \Theta$ gelte $P^p_{X_1} = \mathbf{B}(1,p)$. Ferner sei

$$\gamma(p) := \sqrt{p \cdot (1-p)}$$

die entsprechende Standardabweichung. Für jede Schätzfunktion g_1 gilt

$$\mathrm{E}^p(g_1(X)) = p \cdot g_1(1) + (1-p) \cdot g_1(0).$$

Also existiert keine erwartungstreue Schätzfunktion für γ. Analog für $n > 1$.
Stichwort: asymptotische Erwartungstreue.

Wir studieren nun folgende Fragen:

(i) Wie definiert man die „Güte" einer Schätzfunktion?

(ii) Kennt man „optimale" Schätzfunktionen?

(iii) Wie „verläßlich" ist ein Schätzwert?

Zunächst zur Frage (i).

Definition 12. Der *Quadratmittel-Fehler* einer Schätzfunktion g_n für γ ist definiert als

$$R^\vartheta(g_n) := \mathrm{E}^\vartheta(g_n(X) - \gamma(\vartheta))^2,$$

falls $g_n(X) \in \mathfrak{L}_2^\vartheta$.

Definition 13. Der *Bias* einer Schätzfunktion g_n für γ ist definiert als

$$B^\vartheta(g_n) := \mathrm{E}^\vartheta(g_n(X)) - \gamma(\vartheta),$$

falls $g_n(X) \in \mathfrak{L}_1^\vartheta$.

Bemerkung 14.

(i) Die Erwartungstreue ist äquivalent zu

$$\forall \vartheta \in \Theta : B^\vartheta(g_n) = 0.$$

(ii) Der Quadratmittel-Fehler wird auch als Risiko bei quadratischer Verlustfunktion bezeichnet. Es gilt

$$R^\vartheta(g_n) = \mathrm{Var}^\vartheta(g_n(X)) + (B^\vartheta(g_n))^2,$$

siehe $\boxed{\text{Übung}}$.

Im folgenden sei
$$g_n^*(x) := \overline{x}_n.$$

Beispiel 15. Für die Schätzung des Erwartungswertes gilt $B^\vartheta(g_n^*) = 0$ und
$$R^\vartheta(g_n^*) = \frac{1}{n} \cdot \mathrm{Var}^\vartheta(X_1),$$
falls $X_1 \in \mathfrak{L}_2^\vartheta$. In der Situation von Beispiel 1.9 folgt
$$R^p(g_n^*) = \frac{p \cdot (1-p)}{n}$$
und weiter
$$\sup_{p\in\Theta} R^p(g_n^*) = \frac{1}{4 \cdot n}.$$
In der Situation von Beispiel 1.14 gilt $\gamma(\lambda) = \ln(2) \cdot \mathrm{E}^\lambda(X_1)$, und es folgt
$$R^\lambda(\ln(2) \cdot g_n^*) = \frac{(\ln(2))^2}{\lambda^2 \cdot n},$$
so daß der Quadratmittelfehler zwar für jeden Parameter $\lambda > 0$ mit $n \to \infty$ gegen null strebt, aber für jedes $n \in \mathbb{N}$ unbeschränkt auf $\Theta =]0, \infty[$ ist.

Wir untersuchen die Frage (ii) der Optimalität exemplarisch für die Problemstellung aus Beispiel 1.9. Im folgenden sei
$$\gamma(p) := \mathrm{E}^p(X_1) = p.$$
Für $p \in]0, 1[$ und $x \in D := \{0, 1\}^n$ sei
$$L_x(p) := P^p(\{X = x\}) = p^{k(x)} \cdot (1-p)^{n-k(x)}$$
mit
$$k(x) := |\{i \in \{1, \ldots, n\} : x_i = 1\}|$$
die sogenannte Likelihood-Funktion, vgl. $\boxed{\text{Übung}}$, sowie
$$\ell_x(p) := \ln(L_x(p))$$
die sogenannte Log-Likelihood-Funktion.

Satz 16 (Ungleichung von Fréchet, Cramér, Rao)**.** Jede erwartungstreue Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ für γ erfüllt
$$\forall p \in]0, 1[: \ R^p(g_n) \geq \frac{1}{\mathrm{E}^p\big((\ell'_X(p))^2\big)}.$$

Beweis. Für g_n wie oben gilt

$$p = \mathrm{E}^p(g_n(X)) = \sum_{x \in D} g_n(x) \cdot L_x(p)$$

und somit

$$1 = \sum_{x \in D} g_n(x) \cdot L'_x(p) = \sum_{x \in D} g_n(x) \cdot \ell'_x(p) \cdot L_x(p) = \mathrm{E}^p(g_n(X) \cdot \ell'_X(p)).$$

Aus $\sum_{x \in D} L_x(p) = 1$ folgt

$$0 = \sum_{x \in D} L'_x(p) = \sum_{x \in D} \ell'_x(p) \cdot L_x(p) = \mathrm{E}^p(\ell'_X(p)).$$

Also liefert die Cauchy-Schwarzsche Ungleichung

$$1 = \Big(\mathrm{E}^p((g_n(X) - \gamma(p)) \cdot \ell'_X(p))\Big)^2 \leq \mathrm{Var}^p(g_n(X)) \cdot \mathrm{E}^p\big((\ell'_X(p))^2\big).$$

□

Bemerkung 17. Im Beweis von Satz 16 wurde die konkrete Verteilungsannahme nicht wesentlich genutzt. Die untere Schranke des Satzes gilt deshalb unter viel allgemeineren Voraussetzungen. Siehe Krengel (2000, §4.5) und Irle (2001, p. 308).

Wir erhalten die Optimalität des empirischen Mittels in einem sehr starken Sinn.

Satz 18. Es gilt

$$\forall p \in \,]0,1[\, : \;\; R^p(g_n^*) = \inf \{R^p(g_n) : g_n \text{ erwartungstreue Schätzfunktion für } \gamma\}\,.$$

Beweis. Man verifiziert

$$\mathrm{E}^p\big((\ell'_X(p))^2\big) = \frac{n}{p \cdot (1-p)},$$

siehe Krengel (2000, p. 68), und wendet Satz 16 und Beispiel 15 an. □

Satz 18 besagt, daß g_n^* eine gleichmäßig beste erwartungstreue Schätzfunktion in der vorliegenden Situation ist. Letzteres ist wesentlich, wie das folgende Beispiel zeigt.

Beispiel 19. Sei $\Theta := \mathbb{R}$ und

$$P^{\vartheta}_{X_1} := \mathbf{U}([\vartheta - 1/2, \vartheta + 1/2])$$

für $\vartheta \in \Theta$. Zu schätzen ist

$$\gamma(\vartheta) := \mathrm{E}^{\vartheta}(X_1) = \vartheta.$$

Setze

$$g_3(x_1, x_2, x_3) := \frac{\max(x_1, x_2, x_3) + \min(x_1, x_2, x_3)}{2}.$$

Dann ist g_n erwartungstreu für γ mit

$$\forall \vartheta \in \Theta : \;\; R^{\vartheta}(g_n) < R^{\vartheta}(g_n^*).$$

Siehe $\boxed{\text{ÜBUNG}}$.

Schließlich behandeln wir Frage (iii).

Definition 20. Sei $\alpha \in \,]0,1[$. Zwei Borel-meßbare Abbildungen $\ell_n, r_n : \mathbb{R}^n \to \mathbb{R}$ bilden ein *Konfidenzintervall* für γ zum Niveau $1-\alpha$, falls

$$\forall \vartheta \in \Theta: \; P^\vartheta(\{\gamma(\vartheta) \in [\ell_n(X), r_n(X)]\}) \geq 1-\alpha.$$

Bemerkung 21. Gesucht sind „möglichst kleine“ Konfidenzintervalle. Man beachte, daß nicht $\gamma(\vartheta)$ sondern $[\ell_n(X), r_n(X)]$ zufällig ist.

Im folgenden behandeln wir den Spezialfall

$$\gamma(\vartheta) := \mathrm{E}^\vartheta(X_1)$$

und Intervalle der Form

$$[\overline{x}_n - b_n(x), \overline{x}_n + b_n(x)]$$

mit Borel-meßbaren Abbildungen

$$b_n : \mathbb{R}^n \to [0, \infty[\,.$$

Bemerkung 22. Die Abbildung b_n definiert genau dann Konfidenzintervall für den Erwartungswert zum Niveau $1-\alpha$, wenn

$$\forall \vartheta \in \Theta: \; P^\vartheta(\{|\overline{X}_n - \mathrm{E}^\vartheta(X_1)| \leq b_n(X)\}) \geq 1-\alpha.$$

Zur Festlegung von b_n ist deshalb die Tschebyschev-Ungleichung anwendbar, falls (eine Schranke für) $\sup_{\vartheta\in\Theta} \mathrm{Var}^\vartheta(X_1)$ bekannt ist. Auf diese Weise erhält man jedoch oft zu große Konfidenzintervalle.

Beispiel 23. Fortsetzung von Beispiel 6. Konfidenzintervalle deterministischer Breite $2 \cdot b_n$ mittels der

(i) Tschebyschev-Ungleichung: Es gilt

$$\frac{1}{4 \cdot b_n^2 \cdot n} = \alpha \Leftrightarrow b_n = \frac{1}{\sqrt{4 \cdot \alpha \cdot n}}.$$

(ii) Hoeffdingschen Ungleichung: Es gilt

$$2 \cdot \exp(-2 \cdot b_n^2 \cdot n) = \alpha \Leftrightarrow b_n = \sqrt{\frac{\ln(2/\alpha)}{2 \cdot n}}.$$

Auf diese Weise erhält man

α	b_n per T-Ungl.	b_n per H-Ungl.
0.05	$4.5 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$
0.01	$1.0 \cdot 10^{-3}$	$3.3 \cdot 10^{-4}$
0.001	$3.2 \cdot 10^{-3}$	$3.8 \cdot 10^{-4}$

Nun bestimmen wir Konfidenzintervalle für den Erwartungswert unter Normalverteilungsannahmen. Wir unterscheiden dabei zwei Fälle:

(i) Die Varianz $\sigma^2 > 0$ ist bekannt. Also $\Theta := \mathbb{R}$ und $P^{\mu}_{X_1} := \mathbf{N}(\mu, \sigma^2)$ für $\mu \in \Theta$.

(ii) Die Varianz ist unbekannt. Also $\Theta := \mathbb{R} \times]0, \infty[$ und $P^{(\mu,\sigma)}_{X_1} := \mathbf{N}(\mu, \sigma^2)$ für $(\mu, \sigma) \in \Theta$.

Zuächst stellen wir einige Eigenschaften der Normalverteilung zusammen.

Satz 24. Seien $a, b, \mu, \mu_i \in \mathbb{R}$ mit $a \neq 0$ und $\sigma, \sigma_i \in]0, \infty[$.

(i) Falls $X \sim \mathbf{N}(\mu, \sigma^2)$, dann $a \cdot X + b \sim \mathbf{N}(a \cdot \mu + b, a^2 \cdot \sigma^2)$.

(ii) Falls $X_1, \ldots, X_n$ unabhängig und $X_i \sim \mathbf{N}(\mu_i, \sigma_i^2)$, dann $\sum_{i=1}^n X_i \sim \mathbf{N}(\mu, \sigma^2)$ mit $\mu := \sum_{i=1}^n \mu_i$ und $\sigma^2 := \sum_{i=1}^n \sigma_i^2$.

Beweis. $\boxed{\text{Übung}}$. □

Satz 25. Unter der Normalverteilungsannahme definiert

$$b_n := \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha/2)$$

bei bekannter Varianz σ^2 ein Konfidenzintervall für den Erwartungswert zum Niveau $1 - \alpha$.

Beweis. Satz 24 zeigt

$$Z := \frac{\sqrt{n}}{\sigma} \cdot \big(\overline{X}_n - \mathrm{E}^{\mu}(X_1)\big) \sim \mathbf{N}(0, 1)$$

bzgl. P^{μ}. Für alle $b_n > 0$ folgt

$$\begin{aligned} P^{\mu}(\{|\overline{X}_n - \mathrm{E}^{\mu}(X_1)| \leq b_n\}) &= P^{\mu}(\{|Z| \leq \sqrt{n}/\sigma \cdot b_n\}) \\ &= 2 \cdot \Phi(\sqrt{n}/\sigma \cdot b_n) - 1. \end{aligned}$$

Schließlich gilt

$$2 \cdot \Phi(\sqrt{n}/\sigma \cdot b_n) - 1 = 1 - \alpha \Leftrightarrow b_n = \sigma/\sqrt{n} \cdot \Phi^{-1}(1 - \alpha/2).$$

□

Beispiel 26. Die Abbildungen VIII.1 und VIII.2 zeigen 5 bzw. 50 Konfidenzintervalle nach der Konstruktion aus Satz 25 mit

$$\alpha := 0.05, \qquad \sigma := 2, \qquad n := 15$$

(und $\mu := 3$).

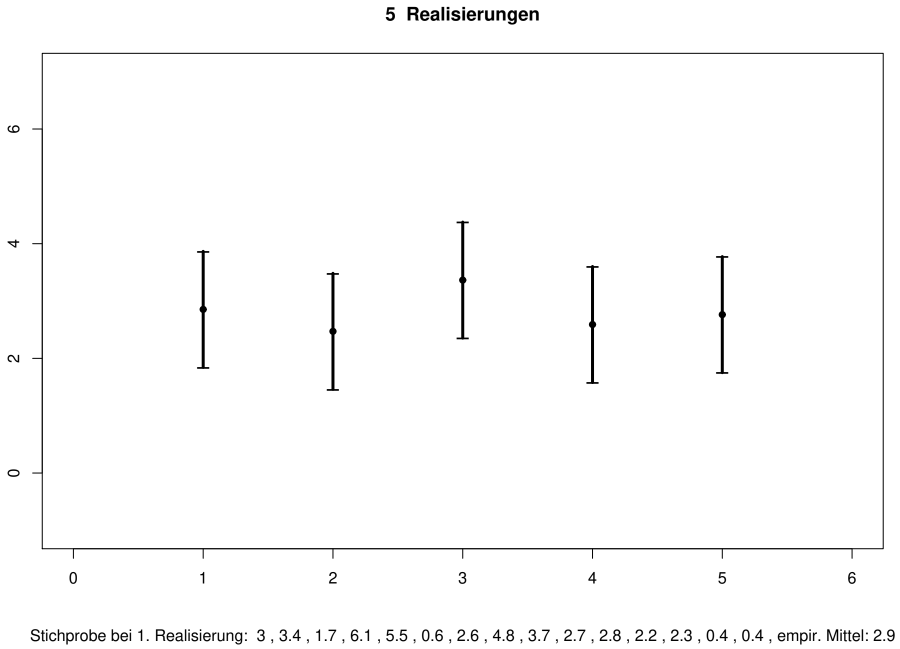

Abbildung VIII.1: Konfidenzintervalle unter Normalverteilungsannahme bei bekannter Varianz

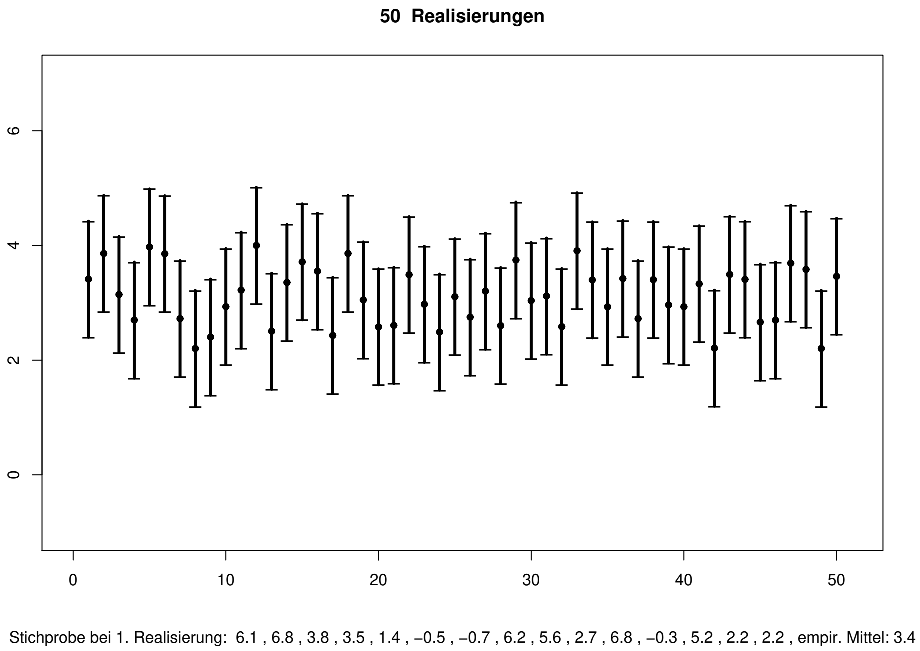

Abbildung VIII.2: Konfidenzintervalle unter Normalverteilungsannahme bei bekannter Varianz

Bei unbekannter Varianz ist es naheliegend σ^2 durch die empirische Varianz $v_n(x) := s_n^2$ zu ersetzen. Im folgenden sei $n > 1$ und $X'_1, \dots, X'_n$ iid mit $X'_1 \sim \mathbf{N}(0,1)$. Setze

$$\overline{X'}_n := \sum_{i=1}^{n} X'_i/n,$$
$$X' := (X'_1, \dots, X'_n)$$

und

$$T'_n := \frac{\overline{X'}_n}{\sqrt{v_n(X')/n}}.$$

Beachte: in der Definition von T'_n ist der Nenner fast sicher ungleich Null.

Lemma 27. Die Zufallsvariable T'_n besitzt die Dichte

$$f_n(x) := \frac{\Gamma(n/2)}{\Gamma((n-1)/2) \cdot \sqrt{\pi \cdot (n-1)}} \cdot \left(1 + x^2/(n-1)\right)^{-n/2}.$$

Beweis. Siehe Irle (2001, Kapitel 20). □

Bemerkung 28. Die Dichte f_n ist symmetrisch, und für alle $x \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} f_n(x) = 1/\sqrt{2\pi} \cdot \exp(-x^2/2).$$

Definition 29. Die Verteilung der Zufallsvariable T'_n heißt *t-Verteilung* mit $n-1$ Freiheitsgraden. Bez.: $\mathbf{t}_{n-1}$.

Die Abbildungen VIII.3 und VIII.4 zeigen die Dichten und Verteilungsfunktion von $\mathbf{t}_5$ bzw. $\mathbf{t}_{20}$. Zum Vergleich sind auch die Dichte und die Verteilungsfunktion von $\mathbf{N}(0,1)$ angegeben.

Bemerkung 30. Zur Berechnung der Verteilungsfunktion von $\mathbf{t}_n$ und entsprechender Quantile: Numerik, Tabellen, Plots.

Lemma 31. Bzgl. $P^{(\mu,\sigma)}$ gilt

$$\frac{\overline{X}_n - \mu}{\sqrt{v_n(X)/n}} \sim \mathbf{t}_{n-1}.$$

Beweis. Setze

$$X'_i := \frac{X_i - \mu}{\sigma}.$$

Bzgl. $P^{(\mu,\sigma)}$ gilt $X'_1, \dots, X'_n$ iid und $X'_1 \sim \mathbf{N}(0,1)$. Ferner gilt

$$\overline{X'_n} = \sum_{i=1}^{n} \frac{X'_i}{n} = \frac{\overline{X_n} - \mu}{\sigma},$$

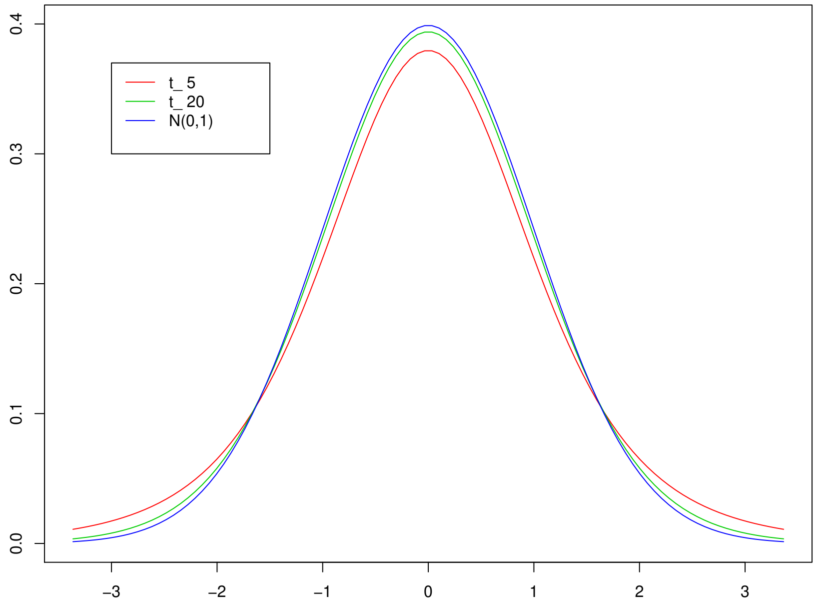

Abbildung VIII.3: Dichten von t-Verteilungen

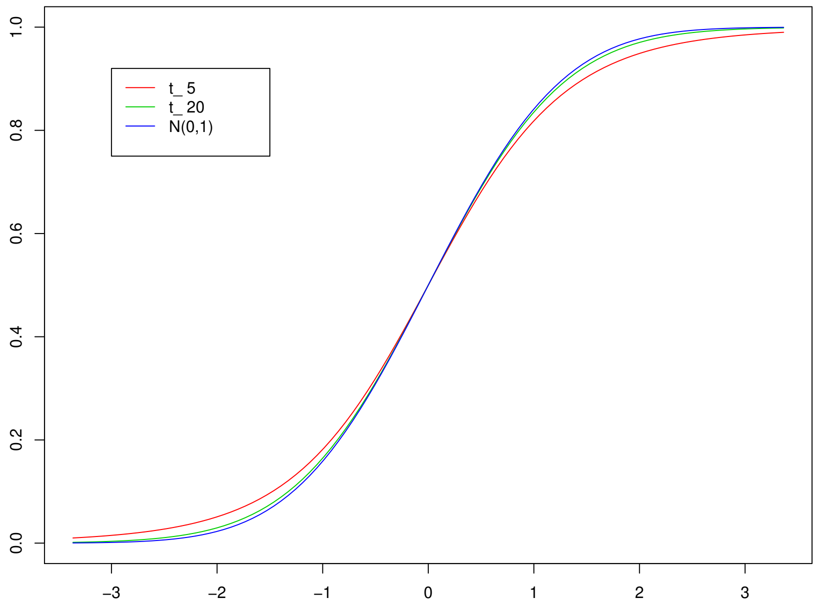

Abbildung VIII.4: Verteilungsfunktionen von t-Verteilungen

$$X_i' - \overline{X_n'} = \frac{X_i - \overline{X_n}}{\sigma}$$

und

$$(n-1)\cdot v_n(X') = \sum_{i=1}^{n}(X_i' - \overline{X_n'})^2 = \frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(X_i - \overline{X_n})^2 = \frac{(n-1)\cdot v_n(X)}{\sigma^2}.$$

Fazit

$$\frac{\overline{X_n'}}{\sqrt{v_n(X')}/\sqrt{n}} = \frac{\overline{X_n}-\mu}{\sigma}\cdot\frac{\sigma}{\sqrt{v_n(X)}/\sqrt{n}} = \frac{\overline{X_n}-\mu}{\sqrt{v_n(X)}/\sqrt{n}}.$$

□

Satz 32. Sei $t_{n-1;1-\alpha/2}$ das $(1-\alpha/2)$-Quantil von $\mathbf{t}_{n-1}$. Unter der Normalverteilungsannahme definiert

$$b_n(x) := \sqrt{\frac{v_n(x)}{n}}\cdot t_{n-1;1-\alpha/2}$$

bei unbekannter Varianz ein Konfidenzintervall für den Erwartungswert zum Niveau $1-\alpha$.

Beweis. Vgl. Beweis von Satz 25. Sei $F_n := F_{Z_n}$ die Verteilungsfunktion von

$$Z_n := \frac{\overline{X}_n - \mu}{\sqrt{v_n(X)/n}}.$$

Für alle $\mu \in \mathbb{R}$ und $\sigma > 0$ gilt gemäß Lemma 31

$$\begin{aligned}
P^{(\mu,\sigma)}(\{|\overline{X_n}-\mu| \le b_n(X)\}) &= P^{(\mu,\sigma)}(\{|Z_n| \le t_{n-1;1-\alpha/2}\}) \\
&= F(t_{n-1;1-\alpha/2}) - F(-t_{n-1;1-\alpha/2}) \\
&= 2\cdot F(t_{n-1;1-\alpha/2}) - 1 \\
&= 2(1-\alpha/2) - 1 = 1-\alpha.
\end{aligned}$$

□

Beispiel 33. Für $\alpha := 0.05$ ergibt sich

n	21	51	101
$t_{n-1;1-\alpha/2}$	$2.09\dots$	$2.01\dots$	$1.98\dots$

Zum Vergleich: $\Phi^{-1}(1-\alpha/2) = 1.96\dots$.

Beispiel 34. Die Abbildung VIII.5 zeigt 50 Konfidenzintervalle nach der Konstruktion aus Satz 32 mit

$$\alpha := 0.05, \qquad n := 15$$

(und $\mu := 3$ sowie $\sigma := 2$).

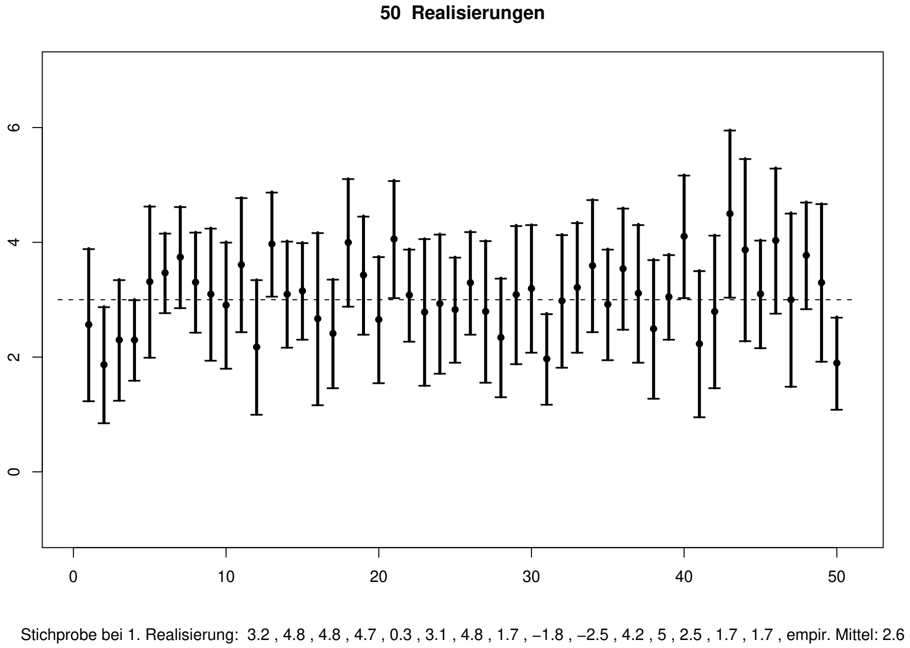

Abbildung VIII.5: Konfidenzintervalle unter Normalverteilungsannahme bei unbekannter Varianz

Ausblick: asymptotische Konfidenzintervalle. Gelte

$$\forall \vartheta \in \Theta : X_1 \in \mathfrak{L}_2^\vartheta \wedge \mathrm{Var}^\vartheta(X_1) > 0.$$

Setze

$$b_n(x) := \sqrt{\frac{v_n(x)}{n}} \cdot \Phi^{-1}(1 - \alpha/2).$$

Satz 35. Für jedes $\alpha \in]0,1[$ gilt

$$\forall \vartheta \in \Theta : \lim_{n\to\infty} P^\vartheta(\{|\overline{X}_n - \mathrm{E}^\vartheta(X_1)| \leq b_n(X)\}) = 1 - \alpha.$$

Beweis. Beruht auf dem Zentralen Grenzwertsatz. Siehe MC-Buch. □

Beispiel 36. Wir ergänzen die Konfidenzintervalle aus Beispiel 23 um asymptotische Konfidenzintervalle nach der Konstruktion aus Satz 35. Siehe Beispiel 10 zum entsprechenden Wert der empirischen Varianz.

α	b_n per T-Ungl.	b_n per H-Ungl.	b_n asymp. per ZGS
0.05	$4.5 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$1.95 \cdot 10^{-4}$
0.01	$1.0 \cdot 10^{-3}$	$3.3 \cdot 10^{-4}$	$2.57 \cdot 10^{-4}$
0.001	$3.2 \cdot 10^{-3}$	$3.8 \cdot 10^{-4}$	$3.27 \cdot 10^{-4}$