

### 3 Statistische Inferenz

#### 3.1 Stichprobentheorie

##### 3.1.1 Stichproben und Statistiken

- **Ausgangspunkt:** Zufallsvariable  $X : \Omega \rightarrow E$  mit unbekannter Verteilung  $P_X$ , die über einem (i.a. nicht näher spezifizierten) Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  definiert ist. Das Tupel  $(\Omega, \mathfrak{A}, P, X)$  wird auch als „statistisches Modell“ bezeichnet.
- **Ziel:** Schätzung der unbekanntes Verteilung  $P_X$  bzw. Schlußfolgerungen über deren Eigenschaften.
- **Ansatz:** Datenerhebung (reale oder synthetische Daten) und Speicherung in Vektor  $\mathbf{x} = (x_1, \dots, x_n)$ . Die Daten  $(x_1, \dots, x_n)$  entsprechen Realisierungen eines stochastischen Modells

$$\underbrace{X_i}_{\text{Zufallsvariable}} \rightarrow \underbrace{x_i}_{\text{Realisierung}}, \quad i = 1, \dots, n; \quad X_i : \Omega \rightarrow E.$$

- allgemeine Annahme: Die  $n$  Zufallsvariablen  $X_1, \dots, X_n$  sind unabhängig und identisch verteilt. Man zieht die  $n$  Daten unabhängig voneinander (d.h. mit Zurücklegen) unter gleichen Bedingungen, d.h.

$$f(x_1, \dots, x_n) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^n f(x_i)$$

Eine solche Stichprobe heißt iid (independently, identically distributed).

**Definition 3.1.** (i) Der Zufallsvektor  $(X_1, \dots, X_n)$  heißt *Zufallsstichprobe*.

(ii) Der Vektor der Realisierungen  $(x_1, \dots, x_n)$  heißt (konkrete) *Stichprobe* (Stichprobenrealisation).

(iii) Für jedes  $i = 1, \dots, n$  heißt  $x_i$  *Stichprobenwert* von  $(x_1, \dots, x_n)$ . Analog nennt man  $X_i$  *Stichprobenvariable* von  $(X_1, \dots, X_n)$ .

(iv) Die Dimension  $n$  von  $(X_1, \dots, X_n)$  bzw.  $(x_1, \dots, x_n)$  heißt *Stichprobenumfang*.

(v) Die Menge  $X \subset E^n$  aller (potentiell möglichen) Stichproben  $(X_1, \dots, X_n)$  heißt *Stichprobenraum*.

- häufige Annahme: Die Dichte bzw. Wahrscheinlichkeitsfunktion von  $X$  entstammt einer parametrischen Verteilungsfamilie, d.h. man kann sie schreiben als  $f(x|\theta)$ , wobei  $\theta$  ein Parameter(vektor) ist. Die gemeinsame Dichte bzw. Wahrscheinlichkeitsfunktion der Stichprobe lautet dann

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

**Definition 3.2.** Sei  $X_1, \dots, X_n$  eine Zufallsstichprobe der Dimension  $n$ . Eine Funktion  $T(X_1, \dots, X_n)$ ,  $T : X \rightarrow \mathbb{R}$  heißt *Statistik*.

**Beispiel 3.1.** Die Zufallsvariable  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  heißt *Stichprobenmittel* der Zufallsstichprobe  $(X_1, \dots, X_n)$ . Die Realisation lautet  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ , basierend auf den Stichprobenrealisationen  $(x_1, \dots, x_n)$ .

**Satz 3.3.** Gegeben seien iid Stichprobenvariablen  $X_1, \dots, X_n$  mit  $\mu = E(X_i)$  und  $\sigma^2 = \text{Var}(X_i)$ . Dann gilt

$$(i) E(\bar{X}_n) = \mu,$$

$$(ii) \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad (\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = 0),$$

- (iii)  $P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$ ,  
 (iv)  $\lim_{n \rightarrow \infty} P(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x) = \Phi(x)$ .

*Beweis.* (i)  $E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$ ,  
 (ii)  $Var(\bar{X}_n) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$ ,  
 (iii) starkes Gesetz der großen Zahlen,  
 (iv) zentraler Grenzwertsatz für Summen von iid Zufallsvariablen. □

- Die Stichprobenvarianz lautet

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

**Satz 3.4.** Sei  $\mu_4 = E[(X_i - \mu)^4]$  und  $\sigma^4 = [Var(X_i)]^2$  das vierte zentrale Moment bzw. die quadrierte Varianz der Stichprobenvariablen  $X_1, \dots, X_n$ . Dann gilt

- (i)  $E[S_n^2] = \sigma^2$ ,  
 (ii)  $Var[S_n^2] = \frac{1}{n} \left( \mu_4 - \frac{n-3}{n-1} \sigma^4 \right)$ , (falls  $E(X_i^4) < \infty$ ),  
 (iii)  $P(\lim_{n \rightarrow \infty} S_n^2 = \sigma^2) = 1$ ,  
 (iv)  $\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} \leq x\right) = \Phi(x)$ .

**Satz 3.5.** Sei  $X_1, \dots, X_n$  eine Zufallsstichprobe einer Grundgesamtheit mit momentenerzeugender Funktion  $M_X(t)$ . Dann gilt

$$M_{\bar{X}}(t) = \left[ M_X\left(\frac{t}{n}\right) \right]^n.$$

*Beweis.*  $M_{\bar{X}}(t) = E[e^{t\bar{X}}] = E[e^{t(X_1 + \dots + X_n)/n}] = E\left[e^{\frac{t}{n}X_1} e^{\frac{t}{n}X_2} \dots e^{\frac{t}{n}X_n}\right] = E\left[e^{\frac{t}{n}X_i}\right]^n = [M_X\left(\frac{t}{n}\right)]^n$ . □

### 3.1.2 Normalverteilte Stichprobenvariablen

- nun zusätzliche Annahme:  $X_i \sim N(\mu, \sigma^2)$ ,  $i = 1, \dots, n$ ,  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ .

**Satz 3.6.** Sei  $X_1, \dots, X_n$  eine Zufallsstichprobe aus  $N(\mu, \sigma^2)$ -verteilter Grundgesamtheit. Seien  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  und  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ . Dann gilt

- (i)  $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ,  
 (ii)  $\bar{X}_n$  und  $S_n^2$  sind unabhängige Zufallsvariablen,  
 (iii)  $(n-1) \frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2$ .

#### Die t-Verteilung

- Ist  $X_i \sim N(\mu, \sigma^2)$ , so gilt

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Frage: Wie lautet die Verteilung von

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \tag{3.1}$$

- Ansatz: Erweiterung mit  $\frac{1}{\sigma}/\frac{1}{\sigma}$  ergibt

$$\frac{(\bar{X}_n - \mu)/(\sigma/\sqrt{n})}{\sqrt{S_n^2/\sigma^2}} \quad (3.2)$$

Der Zähler ist  $N(0,1)$ -verteilt (standardnormalverteilt), der Nenner ist  $\sqrt{\chi_{n-1}^2/(n-1)}$ -verteilt. Zähler und Nenner sind unabhängig!

**Definition 3.7.** Eine Zufallsvariable  $T$  heißt (Student)  $t$ -verteilt mit  $p$  Freiheitsgraden, d.h.  $T \sim t_p$ , falls sie die Dichte

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty \quad (3.3)$$

besitzt.

- Bemerkung: Für  $p = 1$  entspricht (3.3) der Dichte einer Cauchy-Verteilung. Dieser Fall tritt bei einem Stichprobenumfang von 2 ( $n = 2$ ) auf.

**Satz 3.8.** Sei  $X_1, \dots, X_n$  eine Zufallsstichprobe aus einer  $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit. Dann gilt

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}. \quad (3.4)$$

- Eigenschaften der  $t$ -Verteilung. Erwartungswert und Varianz lauten

$$E(T_p) = 0, \text{ falls } p > 1 \quad \text{und} \quad \text{Var}(T_p) = \frac{p}{p-2}, \text{ falls } p > 2. \quad (3.5)$$

- Sei  $T \sim t_p$ . Für  $\alpha \in (0, 1)$  wird dann die (eindeutige) Lösung  $t_{p,\alpha}$  der Gleichung  $F_T(t_{p,\alpha}) = \alpha$  das  $\alpha$ -Quantil der  $t$ -Verteilung mit  $p$  Freiheitsgraden genannt. Die Werte liegen tabelliert vor.

### Die $F$ -Verteilung

- Seien  $X_1, \dots, X_n$  aus einer  $N(\mu_X, \sigma_X)$ -verteilten Grundgesamtheit und  $Y_1, \dots, Y_m$  aus einer  $N(\mu_Y, \sigma_Y)$ -verteilten Grundgesamtheit
- Beim Vergleich der Streuungen beider Stichproben stellt sich die Frage: Wie ist  $\sigma_X/\sigma_Y$  verteilt?
- Informationen darüber sind in  $S_X^2/S_Y^2$  enthalten.
- Die  $F$ -Verteilung gibt die Verteilung folgender Größe an

$$\frac{S_X^2/S_Y^2}{\sigma_X/\sigma_Y} = \frac{S_X^2/\sigma_X}{S_Y^2/\sigma_Y}. \quad (3.6)$$

Bemerkung: (3.6) ist der Quotient zweier skalierten  $\chi^2$ -Verteilungen.

**Definition 3.9.** Eine Zufallsvariable  $F$  heißt (Snedecor)  $F$ -verteilt mit  $p$  und  $q$  Freiheitsgraden, d.h.  $F \sim F_{p,q}$ , falls sie die Dichte

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{(1+(p/q)x)^{(p+q)/2}}, \quad 0 < x < \infty \quad (3.7)$$

besitzt.

**Satz 3.10.** Seien  $X_1, \dots, X_n$  aus einer  $N(\mu_X, \sigma_X)$ -verteilten Grundgesamtheit und  $Y_1, \dots, Y_m$  aus einer  $N(\mu_Y, \sigma_Y)$ -verteilten Grundgesamtheit. Dann gilt

$$F = \frac{S_X^2/\sigma_X}{S_Y^2/\sigma_Y} \sim F_{n-1, m-1}. \quad (3.8)$$

- Eigenschaften der  $F$ -Verteilung:

**Satz 3.11.** Es gilt

- (i) Ist  $X \sim F_{p,q}$ , so ist  $1/X \sim F_{q,p}$ , d.h. das Reziproke einer  $F$ -verteilten Zufallsvariable ist wieder  $F$ -verteilt.
- (ii) Ist  $X \sim t_q$ , so ist  $X^2 \sim F_{1,q}$ .
- (iii) Ist  $X \sim F_{p,q}$ , dann ist

$$\frac{p}{q} \frac{X}{1 + (p/q)X} \sim \beta(p/2, q/2). \quad (3.9)$$

### 3.1.3 Ordnungsstatistiken

- Idee: Statistiken wie der kleinste, größte oder mittlere Wert einer Stichprobe können zusätzliche Informationen beinhalten.

**Definition 3.12.** Die *Ordnungsstatistik* einer Zufallsstichprobe  $X_1, \dots, X_n$  ist die Anordnung der Stichprobenvariablen in aufsteigender Ordnung. Die resultierenden Zufallsvariablen werden mit  $X_{(1)}, \dots, X_{(n)}$  bezeichnet.

- Folgerung: Die Zufallsvariablen einer Ordnungsstatistik erfüllen die Eigenschaft

$$X_{(1)} \leq \dots \leq X_{(n)}$$

- Wichtigste Ordnungsstatistik ist der *Median*

$$M = \begin{cases} X_{((n+1)/2)}, & \text{falls } n \text{ ungerade ist,} \\ (X_{(n/2)} + X_{(n/2+1)})/2, & \text{falls } n \text{ gerade ist.} \end{cases} \quad (3.10)$$

- Ordnungsstatistiken spielen in der nichtparametrischen Statistik eine Rolle.

### 3.1.4 Empirische Verteilungsfunktion

- Wir betrachten nun eine Klasse von Stichprobenfunktionen, mit deren Hilfe die Verteilungsfunktion  $F$  der Stichprobenvariablen  $X_1, \dots, X_n$  aus den vorliegenden Daten  $x_1, \dots, x_n$  bestimmt werden kann.
- Hierfür betrachten wir die Statistik

$$T(x_1, \dots, x_n) = \frac{\#\{i : 1 \leq i \leq n, x_i \leq x\}}{n}, \quad (3.11)$$

Interpretation:  $T(\cdot)$  ist die *relative Häufigkeit* derjenigen Stichprobenwerte, die den Schwellenwert  $x$  nicht überschreiten.

- Man kann sich leicht überlegen, dass für jeden Vektor  $(x_1, \dots, x_n) \in \mathbb{R}^n$  die in (3.11) definierte Abbildung die Eigenschaften einer Verteilungsfunktion besitzt.
- Daher Bezeichnung empirische Verteilungsfunktion der (konkreten) Stichprobe  $(x_1, \dots, x_n)$

**Definition 3.13.** Die Abbildung  $\hat{F}_n : \mathbb{R} \times \Omega \rightarrow [0, 1]$  mit

$$\hat{F}_n(x, \omega) = \frac{\#\{i : 1 \leq i \leq n, X_i(\omega) \leq x\}}{n} \quad (3.12)$$

heißt empirische Verteilungsfunktion der Zufallsstichprobe  $(X_1, \dots, X_n)$ .

- Betrachte die Abweichung

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \quad (3.13)$$

der empirischen Verteilungsfunktion  $\hat{F}_n$  von der zu schätzenden Verteilungsfunktion  $F$  der Stichprobenvariablen  $X_1, \dots, X_n$

- Die in (3.13) definierte Zufallsvariable  $D_n$  heißt Kolmogorow-Abstand von  $\hat{F}_n$  und  $F$ .
- Bemerkung:  $D_n$  kann als Schätzfehler bei der Schätzung der Funktionswerte  $F(x)$  durch  $\hat{F}_n(x)$  interpretiert werden.

**Satz 3.14. (Gliwienko-Castelli, Hauptsatz der Statistik)** Es gilt

$$P(\lim_{n \rightarrow \infty} D_n = 0) = 1. \quad (3.14)$$

**Interpretation:** Die wahre Verteilungsfunktion der Grundgesamtheit kann durch die Verteilungsfunktion der Stichprobe mit wachsendem Stichprobenumfang beliebig genau approximiert werden.

### 3.1.5 Simulation von Zufallsstichproben

- Frage: Wie simuliert man eine Zufallsstichprobe  $X_1, \dots, X_n$  aus einer gegebenen Verteilung  $f(x|\theta)$ ?
- Es existieren mehrere zuverlässige Zufallszahlen-Generatoren, die standardgleichverteilte Zufallsvariablen produzieren können.

**Satz 3.15. (Inversionsverfahren)** Gegeben sei eine stetige univariate Zufallsvariable  $Y \sim F_Y(y)$  mit streng monotoner Verteilungsfunktion auf  $\mathcal{T}$ . Sei  $U \sim U([0, 1])$ . Dann gilt  $F_Y^{-1}(u) \sim F_Y(y)$ .

- Beispiel:  $Y \sim \text{Exp}(\lambda)$ , d.h.  $F_Y(y) = 1 - \exp^{-\lambda y}$  für  $y \geq 0$ . Als Umkehrfunktion erhalten wir  $F_Y^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$ . Fazit: Generieren wir uns  $n$  standardgleichverteilte Zufallsvariablen  $U_1, \dots, U_n$ , so können wir diese mit Hilfe von  $F_Y^{-1}(u)$  in exponentialverteilte Zufallsvariablen transformieren.
- Problem: Dieses Verfahren basiert auf der Lösung einer Integraltransformation

$$F_Y^{-1}(u) = y \iff u = \int_{-\infty}^y f_y(t) dt.$$

Diese Lösung ist nicht für alle Verteilungen (z.B. für die Normalverteilung) angebar.

## 3.2 Punktschätzer

### 3.2.1 Parametrische Modelle

- Annahme: Die Verteilungsfunktion  $F$  der Stichprobenvariablen  $X_1, \dots, X_n$  gehört zu einer vorgegebenen (d.h. bekannten) parametrischen Familie von Verteilungsfunktionen  $\{F(x|\theta), \theta \in \Theta\}$
- Die Menge  $\Theta \subset \mathbb{R}^p$  heißt Parameterraum,  $p \in \mathbb{N}$  ist die Dimension des Parametervektors.

- Beispiel: Eine der wichtigsten parametrischen Verteilungsfamilien von Stichprobenvariablen ist die Normalverteilungsfamilie  $\{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ . In diesem Fall ist  $p = 2$ ,  $\Theta = \mathbb{R} \times (0, \infty)$  und  $\theta = (\theta_1, \theta_2)^\top = (\mu, \sigma^2)^\top$ .

**Definition 3.16.** Eine Schätzfunktion oder Schätzstatistik ist eine Funktion

$$T = g(X_1, \dots, X_n)$$

der Stichprobenvariablen  $X_1, \dots, X_n$ . Der aus den Realisierungen  $x_1, \dots, x_n$  resultierende numerische Wert

$$g(x_1, \dots, x_n)$$

ist der zugehörige Schätzwert.

### 3.2.2 Eigenschaften von Punktschätzern

**Finite Gütemaße für Schätzfunktionen:**

- **Erwartungstreue:** Eine Schätzstatistik  $T = g(X_1, \dots, X_n)$  für den Parameter  $\theta$  heißt *erwartungstreu* oder *unverzerrt*, wenn

$$E_\theta(T) = \theta,$$

d.h. die Schätzstatistik liefert tendenziell den richtigen Wert (keine systematische Unter- oder Überschätzung). Systematische Über- oder Unterschätzung einer Schätzstatistik wird erfaßt in der Verzerrung, auch Bias genannt. Sie wird bestimmt durch

$$Bias_\theta(T) = E_\theta(T) - \theta.$$

- **Erwartete mittlere quadratische Abweichung (MSE):** Die erwartete mittlere quadratische Abweichung (mean squared error) ist bestimmt durch

$$MSE = E((T - \theta)^2) = Var(T) + Bias(T)^2.$$

- **(MSE-)Effizienz von Schätzstatistiken:** Von zwei Schätzstatistiken  $T_1$  und  $T_2$  heißt  $T_1$  effizient, wenn

$$MSE(T_1) \leq MSE(T_2)$$

für alle zugelassenen Verteilungen gilt. Eine Statistik heißt MSE-effizient, wenn ihre mittlere quadratische Abweichung für alle zugelassenen Verteilungen den kleinsten möglichen Wert annimmt.

**Asymptotische Gütemaße für Schätzfunktionen:**

- **Asymptotische Erwartungstreue:** Eine Schätzfunktion  $T$  heißt asymptotisch erwartungstreu für  $\theta$ , wenn gilt

$$\lim_{n \rightarrow \infty} E_\theta(T) = \theta.$$

- **MSE-Konsistenz (Konsistenz im quadratischen Mittel):** Eine Schätzfunktion heißt MSE-konsistent, wenn gilt

$$MSE \xrightarrow{n \rightarrow \infty} 0.$$

- **Schwache Konsistenz:** Die Schätzstatistik  $T = g(X_1, \dots, X_n)$  heißt schwach konsistent, wenn zu beliebigem  $\epsilon > 0$  gilt

$$\lim_{n \rightarrow \infty} P(|T - \theta| < \epsilon) = 1$$

bzw.

$$\lim_{n \rightarrow \infty} P(|T - \theta| \geq \epsilon) = 0.$$

- Konvergenz in Verteilung, Asymptotische Effizienz

### 3.2.3 Methoden zur Gewinnung von Punktschätzern

#### Momentenmethode

#### Maximum-Likelihood-Schätzung

- Ausgangspunkt ist eine Stichprobe  $X_1, \dots, X_n$  unabhängig und identisch verteilter Stichprobenvariablen aus einer Grundgesamtheit mit parametrisierter Dichte oder Wahrscheinlichkeitsfunktion  $f(x|\theta)$ , wobei  $\theta \in \Theta \subset \mathbb{R}^p$  ein unbekannter Parameter(vektor) ist.
- Die Likelihood-Funktion ist definiert als

$$L(\theta|\mathbf{x}) = L(\theta_1, \dots, \theta_p | x_1, \dots, x_n) = \prod_{i=1}^n f(x_i | \theta_1, \dots, \theta_p). \quad (3.15)$$

**Definition 3.17.** Der Maximum-Likelihood-Schätzer (ML-Schätzer) für  $\theta$  ist definiert als

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} L(\theta|\mathbf{x}). \quad (3.16)$$

**Satz 3.18. (Invarianz der Likelihood)** Ist  $\hat{\theta}_{ML}$  der ML-Schätzer von  $\theta$ , dann gilt für jede eindeutige Funktion  $\tau(\theta)$

$$(\widehat{\tau(\theta)})_{ML} = \tau(\hat{\theta}_{ML}).$$

Interpretation: Der Wert der Likelihood wird durch diese Transformation nicht geändert. Insbesondere gilt

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} \tau(L(\theta|\mathbf{x})).$$

- Die direkte Maximierung der Likelihood-Funktion ist häufig numerisch schwierig. Da die Logarithmusfunktion monoton ist, kann aber auch die sogenannte **log-Likelihood-Funktion**

$$l(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x}) \quad (3.17)$$

verwendet werden.

- Man bestimmt den Maximum-Likelihood-Schätzer typischerweise über den Score-Vektor, d.h. den Vektor der partiellen Ableitungen

$$S(\theta) = \frac{\partial}{\partial \theta} l(\theta|\mathbf{x}) = \left( \frac{\partial l(\theta|\mathbf{x})}{\partial \theta_1}, \dots, \frac{\partial l(\theta|\mathbf{x})}{\partial \theta_p} \right)^\top \quad (3.18)$$

- Der ML-Schätzer  $\hat{\theta}_{ML}$  ergibt sich dann als Lösung des Gleichungssystems

$$S(\theta) = \mathbf{0}. \quad (3.19)$$

Es handelt sich dabei meistens um ein nichtlineares Gleichungssystem, das iterativ gelöst werden muss (Newton-Raphson-Methode, Scoring-Methode)

- Die Fisher-Informationsmatrix ist eine  $(p \times p)$ -Matrix der partiellen Ableitungen zweiter Ordnung der log-Likelihood, multipliziert mit minus eins, d.h.

$$I(\theta) = \left( -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta_i \partial \theta_j} \right) \quad (3.20)$$

- Problem: Die Fisher-Informationsmatrix hängt in vielen Fällen von den beobachteten Daten  $\mathbf{x}$  ab. Um sich von dieser Abhängigkeit zu lösen, betrachtet man häufig die erwartete Fisher-Informationsmatrix  $J(\theta)$ , also den Erwartungswert

$$J(\theta) = E(I(\theta)).$$

**Satz 3.19.** *Unter Regularitätsvoraussetzungen (Vertauschung von Differentiation und Integration ist möglich) gilt*

- (i)  $E(S(\theta)) = \mathbf{0}$ ,
- (ii)  $V(S(\theta)) = J(\theta)$ ,
- (iii)  $J(\theta) = E(S(\theta)S(\theta)^\top)$ .

Interpretation: Im Mittel ist der ML-Schätzer erwartungstreu. Dies ist im allgemeinen nicht richtig, gilt jedoch zumindest asymptotisch.

**Satz 3.20. (Cramer-Rao-Schranke)** *Sei  $T(X_1, \dots, X_n)$  eine Schätzstatistik für  $g(\theta)$  basierend auf der Stichprobe  $X_1, \dots, X_n$  mit*

$$E(T(X_1, \dots, X_n)) = g(\theta), \quad \forall \theta \in \Theta.$$

*Ferner sei  $J(\theta)$  die erwartete Fisher-Informationsmatrix von  $X_1, \dots, X_n$  bezüglich  $\theta$ . Unter Regularitätsvoraussetzungen gilt dann*

$$\text{Var}(T(X_1, \dots, X_n)) \geq \frac{\left[\frac{\partial}{\partial \theta} g(\theta)\right]^2}{J(\theta)}. \quad (3.21)$$

*Im Spezialfall  $g(\theta) = \theta$  gilt*

$$\text{Var}(T(X_1, \dots, X_n)) \geq J(\theta)^{-1}. \quad (3.22)$$

Bemerkung: Die Cramer-Rao-Schranke gibt eine untere Schranke für die Varianz eines erwartungstreuen Schätzers an. Die Varianz wird üblicherweise als Gütekriterium verwendet. Falls die Varianz eines erwartungstreuen Schätzers die Cramer-Rao-Schranke annimmt, weiß man somit, dass es keinen besseren erwartungstreuen Schätzer geben kann.

### Eigenschaften des ML-Schätzers:

**Satz 3.21. (Schwache Konsistenz des ML-Schätzers)** *Seien  $X_1, \dots, X_n \sim f(x|\theta_0)$  eine iid Zufallsstichprobe und  $\theta_0$  der wahre Parameter(vektor). Dann existiert für  $n \rightarrow \infty$  eine schwach konsistente Folge von ML-Schätzern  $\hat{\theta}_n$  (die als lokales Maximum der log-Likelihood definiert sind).*

**Satz 3.22. (Asymptotische Normalverteilung)** *Seien  $X_1, \dots, X_n \sim f(x|\theta_0)$  eine iid Zufallsstichprobe und  $\hat{\theta}_{ML}$  ein konsistenter Schätzer für  $\theta_0$ . Dann gilt (unter Regularitätsvoraussetzungen):*

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{D} N_p(\mathbf{0}, [J_1(\theta_0)]^{-1}),$$

wobei  $J_1$  die erwartete Fisher-Informationsmatrix einer Beobachtung  $X_i \sim f(x|\theta_0)$  ist. Die erwartete Fisher-Informationsmatrix der gesamten Zufallsstichprobe ist damit  $n \cdot J_1(\theta_0) = J(\theta_0)$ . Damit gilt

$$\hat{\theta}_{ML} \stackrel{a}{\sim} N_p(\theta_0, [J(\theta_0)]^{-1}),$$

d.h.

- (i) der ML-Schätzer ist asymptotisch erwartungstreu,
- (ii) der ML-Schätzer ist asymptotisch effizient, d.h. er nimmt für  $n \rightarrow \infty$  die Cramer-Rao-Schranke an.

### ML-Schätzung bei Normalverteilung

$X_i$  seien unabhängig und identisch normalverteilt mit Parameter  $\mu$  und  $\sigma^2$ , d.h. für die Wahrscheinlichkeitsfunktion gilt

$$f(x_i|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}.$$

- (i) Für  $\mathbf{x} = (x_1, \dots, x_n)$  ergibt sich zunächst die Likelihood-Funktion als Produkt der einzelnen Wahrscheinlichkeitsfunktionen gemäß

$$L(\mu, \sigma^2 | \mathbf{x}) = \prod_{i=1}^n f(x_i | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^n \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}$$

und damit die log-Likelihood-Funktion

$$l(\mu, \sigma^2 | x) = \ln L(\cdot) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

**Bestimmung von  $\hat{\mu}_{ML}$ :**

Partielles Ableiten von  $l(\mu, \sigma^2 | x)$  nach  $\mu$  und Nullsetzen liefert

$$\frac{\partial l(\mu, \sigma^2 | x)}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2 \cdot (x_i - \mu) \cdot (-1) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{1}{\sigma^2} \left( \sum_{i=1}^n x_i - n\mu \right) \stackrel{!}{=} 0$$

und damit

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

**Bestimmung von  $\hat{\sigma}_{ML}^2$ :**

Partielles Ableiten von  $l(\mu, \sigma^2 | x)$  nach  $\sigma^2$  und Nullsetzen liefert

$$\frac{\partial l(\mu, \sigma^2 | x)}{\partial \sigma^2} = -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} - \frac{1}{2} \frac{-1}{(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{2\sigma^2} \left( -n + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \stackrel{!}{=} 0$$

und damit

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

bzw. falls  $\mu$  unbekannt und durch  $\hat{\mu} = \bar{x}$  geschätzt wird

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- (ii) Zum Nachweis der schwachen Konsistenz lässt sich verwenden, dass aus MSE-Konsistenz stets die schwache Konsistenz folgt, d.h. also es ist

$$\lim_{n \rightarrow \infty} MSE(\hat{\mu}_{ML}) = \lim_{n \rightarrow \infty} \{Var(\hat{\mu}_{ML}) + Bias(\hat{\mu}_{ML})^2\} = 0$$

zu zeigen.

Dabei gilt im vorliegenden Fall

$$\begin{aligned} E(\hat{\mu}_{ML}) &= E(\bar{X}) = \mu \\ Var(\hat{\mu}_{ML}) &= Var(\bar{X}) = \frac{\sigma^2}{n} \end{aligned}$$

sowie

$$\lim_{n \rightarrow \infty} MSE(\hat{\mu}_{ML}) = \lim_{n \rightarrow \infty} \left\{ \frac{\sigma^2}{n} + 0^2 \right\} = 0$$

und damit die MSE- bzw. schwache Konsistenz von  $\hat{\mu}_{ML}$ .

## Bayes-Schätzung

- bisher: Bei Schätzung des Parametervektors  $\theta$  sind nur Informationen in den Daten berücksichtigt worden.
- Gelegentlich besitzt man noch zusätzliche Informationen (sog. non-sample information), die dann meist aus wissenschaftlichen Theorien stammen.
- Diese zusätzlichen Informationen können deterministischer Natur sein (z.B. Parameter liegen in bestimmtem Intervall, Summe der Parameter muss eins ergeben,...) aber auch stochastischer Natur sein (z.B. Parameter liegt innerhalb eines bestimmten Intervalls aber mit bestimmter Verteilung).
- Sind die zusätzlichen Informationen deterministisch, so können wir diese als Nebenbedingungen in die ML-Schätzung einbauen (Stichwort: Lagrange-Ansatz)
- Sind die zusätzlichen Informationen stochastisch, so verwenden wir die sog. Bayes-Schätzung (Bayes Inferenz).
- In der Bayes-Inferenz ist  $\theta$  nun eine Zufallsvariable mit priori-Verteilung  $f(\theta)$  (Zusammenfassung der Informationen, bevor man die Daten betrachtet)
- Man betrachtet nach der Beobachtung der Daten  $\mathbf{X} = \mathbf{x}$  die posteriori-Verteilung  $f(\theta|\mathbf{x})$ .
- Nach dem Satz von Bayes erhalten wir für die posteriori-Verteilung

$$f(\theta|\mathbf{x}) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} = \frac{f(x|\theta)f(\theta)}{f(x)} \propto f(x|\theta)f(\theta).$$

- Inferenz für  $\theta$  basiert dann ausschließlich auf der posteriori-Verteilung. Mögliche Punktschätzer für  $\theta$  sind beispielsweise der Erwartungswert, der Median oder der Modus der posteriori-Verteilung.

**Definition 3.23.** Sei  $\mathcal{F} = \{f(\mathbf{x}|\theta) : \theta \in \Theta\}$  eine Familie von Wahrscheinlichkeits- bzw. Dichtefunktionen der Daten  $\mathbf{X}$ . Man nennt eine Klasse  $\mathcal{G}$  von priori-Verteilungen konjugiert bezüglich  $\mathcal{F}$ , falls für alle priori-Verteilungen  $g(\theta) \in \mathcal{G}$  gilt

$$g(\theta|\mathbf{x}) \in \mathcal{G}.$$

**Definition 3.24.** Eine priori-Verteilung  $f(\theta)$  mit

$$\int f(\theta)d\theta = \infty \quad \text{bzw.} \quad \sum_{\theta:f(\theta)>0} f(\theta) = \infty$$

heißt uneigentliche priori-Verteilung (*improper prior*).

Häufig steht man vor der Frage, möglichst wenig Vorwissen in die priori-Verteilung einfließen zu lassen. Ein naiver Ansatz wäre die Verwendung der stetigen Gleichverteilung für  $\theta$ . Problem: die priori-Verteilung  $f(\theta)$  ist dann möglicherweise nicht integrierbar, also uneigentlich.

**Definition 3.25. (Jeffreys priori)** Sei  $X$  eine Zufallsvariable mit Wahrscheinlichkeits- bzw. Dichtefunktion  $f(x|\theta)$  und  $\theta$  der unbekannte skalare Parameter. Jeffrey's priori wird definiert als

$$g(\theta) \propto [J(\theta)]^{1/2},$$

wobei  $J(\theta)$  die erwartete Fisher-Information von  $\theta$  bezeichnet.

Jeffreys priori ist invariant gegenüber eindeutigen Transformationen von  $X$ . Prinzipiell lässt sich Jeffrey's priori auch für mehrdimensionale Zufallsvariablen definieren.

Die Wahl der Punktschätzer hängt von der verwendeten Verlustfunktion ab.

**Definition 3.26. (Verlustfunktion)** Eine Verlustfunktion  $l(a, \theta)$  gibt den Verlust an, den man beim Schätzen von  $\theta$  durch  $a$  erleidet.

Beispiele für Verlustfunktionen:

- (i) quadratische Verlustfunktion:  $l(a, \theta) = (a - \theta)^2$ ,
- (ii) lineare Verlustfunktion:  $l(a, \theta) = |a - \theta|$ ,
- (iii) Null-Eins-Verlustfunktion:

$$l_\epsilon(a, \theta) = \begin{cases} 0, & |a - \theta| \leq \epsilon, \\ 1, & |a - \theta| > \epsilon. \end{cases}$$

**Definition 3.27.** Ein Bayes-Schätzer von  $\theta$  bezüglich einer Verlustfunktion  $l(a, \theta)$  minimiert den erwarteten Verlust bzgl. der posteriori-Verteilung, d.h. er minimiert

$$E(l(a, \theta) | \mathbf{x}) = \int_{\Theta} l(a, \theta) f(\theta | \mathbf{x}) d\theta.$$

Der Punktschätzer  $a$  wird so gewählt, dass er den erwarteten Verlust minimiert.

**Satz 3.28.** Folgende Verlustfunktionen determinieren den Bayes-Schätzer:

- (i) quadratische Verlustfunktion  $\Rightarrow$  posteriori-Erwartungswert,
- (ii) lineare Verlustfunktion  $\Rightarrow$  posteriori-Median,
- (iii) Null-Eins-Verlustfunktion  $\xrightarrow{\text{für } \epsilon \rightarrow 0}$  posteriori-Modus.

Bemerkungen:

- (i) In konjugierten Verteilungsklassen sind posteriori-Erwartungswert und posteriori-Modus meist in geschlossener Form bekannt. Der posteriori-Median muss meistens numerisch bestimmt werden.
- (ii) In nicht-konjugierten Verteilungsklassen muss der posteriori-Modus meistens durch Optimierung, der posteriori-Erwartungswert meistens durch numerische Integration berechnet werden. (Alternative: Monte-Carlo-Methoden)

## 3.3 Intervallschätzung

### 3.3.1 Theorie

- Sei  $X = (X_1, \dots, X_n)$  das  $n$ -Tupel der Stichprobenvariablen und  $\mathcal{X}$  der Stichprobenraum.
- **Definition:** Ein Intervallschätzer eines Parameters  $\theta \in \mathbb{R}$  wird gebildet aus einem Paar von Schätzfunktionen  $G_u = g_u(X_1, \dots, X_n)$  und  $G_o = g_o(X_1, \dots, X_n)$  mit  $G_u \leq G_o$  für alle  $X \in \mathcal{X}$ . Wenn  $X = \mathbf{x}$  beobachtet wird, schließt man daraus  $g_u(\mathbf{x}) \leq \theta \leq g_o(\mathbf{x})$ . Das zufällige Intervall  $[G_u, G_o]$  heißt *Intervallschätzer*.
- **Frage:** Warum ist es sinnvoll, von Punktschätzern zu Intervallschätzern überzugehen?
- **Beispiel:** Sei  $X_1, \dots, X_4$  eine Stichprobe aus einer  $N(\mu, 1)$  verteilten Grundgesamtheit. Wenn wir  $\mu$  durch  $\bar{X}$  schätzen, beträgt die Wahrscheinlichkeit, den richtigen Wert zu treffen  $P(\bar{X} = \mu) = 0$ . Ein Intervallschätzer des unbekanntes  $\mu$  ist z.B.  $[\bar{X} - 1, \bar{X} + 1]$ . Durch den Übergang zu einem Intervallschätzer verlieren wir an Präzision, dafür erhalten wir mit positiver Wahrscheinlichkeit den korrekten Parameterwert.

- Für das obige Beispiel erhalten wir

$$P(\mu \in [\bar{X} - 1, \bar{X} + 1]) = 0.9544,$$

d.h. mit einer Wahrscheinlichkeit von 0.9544 liegt der wahre Parameter in diesem Intervall.

- Quantifizierung durch Irrtumswahrscheinlichkeit  $\alpha$  (=Wahrscheinlichkeit, mit der der Intervallschätzer den wahren Parameter **nicht** überdeckt).
- **Definition:** Zu einer vorgegebenen Irrtumswahrscheinlichkeit  $\alpha$  liefern die Schätzfunktionen  $G_u$  und  $G_o$  ein  $(1 - \alpha)$ -Konfidenzintervall, wenn gilt

$$P(G_u \leq G_o) = 1 \quad \text{und} \quad P(G_u \leq \theta \leq G_o) = 1 - \alpha.$$

- Seien  $X_1, \dots, X_n$  unabhängig und identisch verteilte Zufallsvariablen mit  $E(X_i) = \mu$  ( $\mu$  unbekannt) und  $Var(X_i) = \sigma^2 > 0$ . Gesucht ist ein Konfidenzintervall für  $\mu$  unter Verwendung der Schätzfunktion  $\bar{X}$ .
- **Fall 1:  $\sigma^2$  bekannt**

Das  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  bei normalverteiltem Merkmal bzw. bei beliebiger Verteilung und  $n > 30$  (zentraler Grenzwertsatz) lautet

$$\left[ \bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

- **Fall 2:  $\sigma^2$  unbekannt**

Das  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  bei normalverteiltem Merkmal lautet

$$\left[ \bar{X} - t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}} \right]$$

mit  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ .

Bei beliebiger Verteilung und  $n > 30$  lautet das  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$

$$\left[ \bar{X} - z_{1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{S}{\sqrt{n}} \right].$$

### 3.3.2 Konfidenzintervalle für die Varianz normalverteilter Grundgesamtheiten

Ansatz für ein Konfidenzintervall für  $\sigma^2$  bei normalverteilter Grundgesamtheit: Es gilt

$$\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2.$$

Sei  $q_{\alpha/2} = \chi_{\alpha/2, n-1}^2$  bzw.  $q_{1-\alpha/2} = \chi_{1-\alpha/2, n-1}^2$  das  $(\alpha/2)$  bzw.  $(1 - \alpha/2)$ -Quantil der  $\chi^2$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Dann ergibt sich das Konfidenzintervall folgendermaßen:

$$\begin{aligned} P\left(q_{\alpha/2} \leq \frac{n-1}{\sigma^2} S^2 \leq q_{1-\alpha/2}\right) &= 1 - \alpha \\ \iff P\left(\frac{q_{\alpha/2}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{q_{1-\alpha/2}}{(n-1)S^2}\right) &= 1 - \alpha \\ \iff P\left(\frac{(n-1)S^2}{q_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{q_{\alpha/2}}\right) &= 1 - \alpha \\ \implies KI = \left[\frac{(n-1)S^2}{q_{1-\alpha/2}}; \frac{(n-1)S^2}{q_{\alpha/2}}\right] \end{aligned}$$

Beispiel: Tinas monatliche Ausgaben für Kaffee und Kuchen sind normalverteilt mit unbekanntem Parametern  $\mu, \sigma^2$ . Aus den 20 Quittungen für den Monat April hat Tina einen Mittelwert von 5.50 Euro und eine Stichprobenvarianz von 0.25 errechnet. Geben Sie ein 95%-Konfidenzintervall für  $\sigma^2$  an.

Das zweiseitige Konfidenzintervall für  $\sigma^2$ , bei normalverteiltem Merkmal, ist bestimmt durch die Grenzen:

$$\left[ \frac{(n-1)S^2}{q_{1-\alpha/2}}, \frac{(n-1)S^2}{q_{\alpha/2}} \right].$$

Dabei sind  $q_{\alpha/2}$  und  $q_{1-\alpha/2}$  die  $(\alpha/2)$ - und  $(1 - \alpha/2)$ -Quantile der  $\chi_{n-1}^2$  Verteilung. Für  $n = 20$  und  $\alpha = 0.05$  ergeben sich die folgenden Werte:

$$\chi_{0.025,19}^2 = 8.9065$$

und

$$\chi_{0.975,19}^2 = 32.852.$$

Daraus ergibt sich das 95%-KI für die Varianz von Tinas Kaffee-und-Kuchen Ausgaben:

$$\begin{aligned} KI &= \left[ \frac{19 \cdot 0.25}{32.852}, \frac{19 \cdot 0.25}{8.9065} \right] \\ &= [0.1446, 0.5333]. \end{aligned}$$

### 3.3.3 Konfidenzintervalle für den Anteilswert

- Gesucht: Anteilswert bzw. Auftretenswahrscheinlichkeit

$$\pi = P(X = 1)$$

in einer dichotomen Grundgesamtheit, wobei  $X \in \{0, 1\}$ .

- Für  $n$  unabhängige Wiederholungen ist die Summe  $\sum_{i=1}^n X_i \sim B(n, \pi)$ .
- Schätzfunktion für  $\pi$ :  $\bar{X}$  mit  $E(\bar{X}) = \pi$  und  $Var(\bar{X}) = \pi(1 - \pi)/n$ .
- Nach zentralem Grenzwertsatz gilt approximativ

$$\frac{\bar{X} - \pi}{\sqrt{\pi(1 - \pi)/n}} \stackrel{a}{\approx} N(0, 1).$$

- Approximieren des Nenners, indem man  $\pi$  durch  $\hat{\pi} = \bar{X}$  ersetzt, dann erhält man

$$\begin{aligned} P\left(-z_{1-\alpha/2} \leq \frac{\hat{\pi} - \pi}{\sqrt{\hat{\pi}(1 - \hat{\pi})/n}} \leq z_{1-\alpha/2}\right) &= 1 - \alpha \\ \implies KI &= \left[ \hat{\pi} - z_{1-\alpha/2} \sqrt{\hat{\pi}(1 - \hat{\pi})/n}; \hat{\pi} + z_{1-\alpha/2} \sqrt{\hat{\pi}(1 - \hat{\pi})/n} \right] \end{aligned}$$

**Beispiel:** Von den 53 Studienanfängern des Wintersemesters 2001 im Studiengang Statistik der LMU München kamen 42 aus Bayern und 4 aus Sachsen.

- Berechnen Sie jeweils ein Konfidenzintervall zum Sicherheitsgrad 0.95 für den Anteil der Studienanfänger aus Bayern bzw. Sachsen. Warum sind die Breiten unterschiedlich?
- Wie groß hätte der Stichprobenumfang jeweils sein müssen, damit der geschätzte Anteil der Studienanfänger aus Bayern bzw. Sachsen mit 95 % Sicherheitswahrscheinlichkeit um weniger als 3 Prozentpunkte vom wahren Wert abweicht? Dabei können Sie zunächst davon ausgehen, dass die Schätzwerte  $\hat{\pi}$  aus (a) bekannt sind.

- (c) Wie groß hätte der Stichprobenumfang sein müssen, wenn Sie die in (b) geforderte Genauigkeit bei der Schätzung der unbekanntem Anteile erreichen wollen, allerdings nicht mehr davon ausgehen können, dass die Schätzwerte  $\hat{\pi}$  bekannt sind?

Aufgrund des großen Stichprobenumfangs von  $n = 53 > 30$  erhält man ein approximatives  $(1 - \alpha)$ -Konfidenzintervall für die jeweiligen Anteilswerte allgemein durch

$$\left[ \hat{\pi} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right]$$

- (a) Zur konkreten Berechnung für die Anteile der Studienanfänger aus Bayern  $\pi_B$  und Sachsen  $\pi_S$  ermittelt man zunächst die Schätzwerte

$$\hat{\pi}_B = \frac{42}{53} \approx 0.7925 \quad \hat{\pi}_S = \frac{4}{53} \approx 0.0755$$

und erhält mit  $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$  für Bayern

$$\begin{aligned} KI_B &= \left[ \hat{\pi}_B - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}_B(1-\hat{\pi}_B)}{n}}, \hat{\pi}_B + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}_B(1-\hat{\pi}_B)}{n}} \right] \\ &= \left[ 0.7925 - 1.96 \sqrt{\frac{0.7925 \cdot 0.2075}{53}}, 0.7925 + 1.96 \sqrt{\frac{0.7925 \cdot 0.2075}{53}} \right] \\ &= [0.6833, 0.9017] \end{aligned}$$

bzw. für Sachsen

$$\begin{aligned} KI_S &= \left[ \hat{\pi}_S - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}_S(1-\hat{\pi}_S)}{n}}, \hat{\pi}_S + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}_S(1-\hat{\pi}_S)}{n}} \right] \\ &= \left[ 0.0755 - 1.96 \sqrt{\frac{0.0755 \cdot 0.9245}{53}}, 0.0755 + 1.96 \sqrt{\frac{0.0755 \cdot 0.9245}{53}} \right] \\ &= [0.0044, 0.1466] \end{aligned}$$

Für die Breiten der Konfidenzintervalle gilt dann

$$d_B = 0.9017 - 0.6833 = 0.2184 \quad d_S = 0.1466 - 0.0044 = 0.1422$$

Allgemein gilt für die Breite eines  $(1 - \alpha)$ -Konfidenzintervalls

$$d = \hat{\pi} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} - \left( \hat{\pi} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right) = 2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Damit hängt die Breite bei festem  $n$  und  $\alpha$  lediglich vom Produkt  $\hat{\pi}(1 - \hat{\pi})$  und damit von der Schätzung  $\hat{\pi}$  ab. In Teilaufgabe (a) gilt dabei

$$\hat{\pi}_B(1 - \hat{\pi}_B) = 0.7925 \cdot 0.2075 = 0.1644 \quad \hat{\pi}_S(1 - \hat{\pi}_S) = 0.0755 \cdot 0.9245 = 0.0698$$

was zu den unterschiedlichen Breiten führt.

- (b) Löst man die allgemeine Formel für die Breite  $d$  aus Teilaufgabe (b) nach  $n$  auf, so erhält man allgemein für den benötigten Stichprobenumfang  $\tilde{n}$  bei vorgegebener Genauigkeit  $d$  die Relation

$$\tilde{n} \geq \frac{4 \cdot z_{1-\frac{\alpha}{2}}^2 \cdot \hat{\pi}(1-\hat{\pi})}{d^2}$$

Für die Aufgabenstellung gilt dabei

$$z_{1-\frac{\alpha}{2}} = 1.96 \quad \hat{\pi}_B(1 - \hat{\pi}_B) = 0.1644 \quad \hat{\pi}_S(1 - \hat{\pi}_S) = 0.0698 \quad d^2 = 0.06^2 = 0.0036$$

und man erhält für Bayern

$$\tilde{n}_B \geq \frac{4 \cdot 1.96^2 \cdot 0.1644}{0.0036} \approx 701.7$$

bzw. für Sachsen

$$\tilde{n}_S \geq \frac{4 \cdot 1.96^2 \cdot 0.0698}{0.0036} \approx 297.9$$

- (c) Besitzt man wie im Folgenden keine Vorkenntnis über  $\hat{\pi}$ , so muss in der Formel für  $\tilde{n}$  das Produkt  $\hat{\pi}(1 - \hat{\pi})$  abgeschätzt werden. Wie man sich leicht klar macht, gilt dabei stets  $\hat{\pi}(1 - \hat{\pi}) \leq 0.25$  und damit für den benötigten Stichprobenumfang ohne Vorkenntnis über  $\hat{\pi}$

$$\tilde{n} \geq \frac{z_{1-\frac{\alpha}{2}}^2}{d^2}$$

Damit folgt für die Aufgabenstellung

$$\tilde{n}_B = \tilde{n}_S \geq \frac{1.96^2}{0.0036} \approx 1067.1$$

*Bemerkung:* Halbierung von  $d$  bewirkt allgemein Vervierfachung von  $\tilde{n}$  bei festem  $\alpha$ .

### 3.3.4 Die Delta-Methode

- Seien  $X_1, \dots, X_n$  unabhängige Zufallsvariablen mit  $E(X_i) = \mu_i, i = 1, \dots, n$ . Sei  $\mathbf{X} := (X_1, \dots, X_n)$  und  $\boldsymbol{\mu} := (\mu_1, \dots, \mu_n)$ .  $g(\mathbf{X})$  sei eine Schätzfunktion für  $g(\boldsymbol{\mu})$ .
- Gesucht ist  $Var(g(\mathbf{X}))$ .
- Lösung: Wir verwenden eine Taylorreihenentwicklung 1. Ordnung zur Bestimmung der Varianz von  $g(\mathbf{X})$ . Dazu bezeichne

$$g'_i(\boldsymbol{\mu}) = \left. \frac{\partial}{\partial X_i} g(\mathbf{X}) \right|_{X_1=\mu_1, \dots, X_n=\mu_n}$$

die  $i$ -te partielle Ableitung an der Stelle  $X_1 = \mu_1, \dots, X_n = \mu_n$ . Dann ergibt sich die Taylorreihenentwicklung 1. Ordnung

$$\begin{aligned} g(\mathbf{X}) &= g(\boldsymbol{\mu}) + \sum_{i=1}^n g'_i(\boldsymbol{\mu})(X_i - \mu_i) + \text{Rest} \\ &\approx g(\boldsymbol{\mu}) + \sum_{i=1}^n g'_i(\boldsymbol{\mu})(X_i - \mu_i) \\ \implies E_{\boldsymbol{\mu}}[g(\mathbf{X})] &\approx g(\boldsymbol{\mu}) + \sum_{i=1}^n g'_i(\boldsymbol{\mu}) \underbrace{E_{\boldsymbol{\mu}}(X_i - \mu_i)}_{=0} \\ &= g(\boldsymbol{\mu}). \end{aligned}$$

Die Varianz leitet sich folgendermaßen her:

$$\begin{aligned}
 \text{Var}_{\boldsymbol{\mu}}[g(\mathbf{X})] &= E_{\boldsymbol{\mu}}[(g(\mathbf{X}) - E_{\boldsymbol{\mu}}[g(\mathbf{X})])^2] \\
 &\approx E_{\boldsymbol{\mu}}[(g(\mathbf{X}) - g(\boldsymbol{\mu}))^2] \\
 &\approx E_{\boldsymbol{\mu}} \left[ \left( \sum_{i=1}^n g'_i(\boldsymbol{\mu})(X_i - \mu_i) \right)^2 \right] \\
 &= E_{\boldsymbol{\mu}} \left[ \sum_{i=1}^n [g'_i(\boldsymbol{\mu})]^2 (X_i - \mu_i)^2 + \sum_{i \neq j} g'_i(\boldsymbol{\mu}) g'_j(\boldsymbol{\mu}) (X_i - \mu_i)(X_j - \mu_j) \right] \\
 &= \sum_{i=1}^n [g'_i(\boldsymbol{\mu})]^2 \underbrace{E_{\boldsymbol{\mu}}[(X_i - \mu_i)^2]}_{=\text{Var}(X_i)} + \sum_{i \neq j} g'_i(\boldsymbol{\mu}) g'_j(\boldsymbol{\mu}) \underbrace{E_{\boldsymbol{\mu}}[(X_i - \mu_i)(X_j - \mu_j)]}_{=\text{Cov}_{\boldsymbol{\mu}}(X_i, X_j)=0} \\
 \implies \text{Var}_{\boldsymbol{\mu}}[g(\mathbf{X})] &\approx \sum_{i=1}^n [g'_i(\boldsymbol{\mu})]^2 \text{Var}(X_i)
 \end{aligned}$$

**Satz 3.29. (Delta-Methode)** Sei  $Y_n$  eine Folge von Zufallsvariablen, so dass  $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ . Für eine gegebene Funktion  $g(Y_n)$  und einen spezifischen Parameterwert  $\theta$  existiere  $g'(\theta) \neq 0$ . Dann gilt

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{d} N(0, \sigma^2 [g'(\theta)]^2).$$

**Anmerkung:** Gegeben ist also, dass  $Y_n$  bereits ein erwartungstreuer Schätzer für  $\theta$  ist. Im folgenden Beispiel ist  $Y_n = \bar{X} = \hat{p}$ .

### Beispiel:

In einer Studie der Bundesregierung sollen die Chancen neu gegründeter Unternehmen untersucht werden. In einer einfachen Stichprobe von 200 neu gegründeten Unternehmen befinden sich 180 erfolgreiche Unternehmen (d.h. die innerhalb von drei Jahren einen Gewinn erwirtschaftet haben).

- (i) Berechnen Sie ein 95% Konfidenzintervall für den Anteil  $p$  der erfolgreichen Neugründungen in der Grundgesamtheit.
- (ii) Berechnen Sie ein 95% Konfidenzintervall für die Chance  $(\frac{p}{1-p})$ , ein erfolgreiches Unternehmen zu gründen.
- (i) Der Schätzer für den Anteil erfolgreicher Unternehmen in der Grundgesamtheit lautet

$$\hat{p} = \frac{180}{200} = 0.9.$$

Mit  $\hat{p}(1 - \hat{p}) = 0.9 \cdot 0.1 = 0.09$  ergibt sich für das Konfidenzintervall

$$\begin{aligned}
 KI &= \left[ \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \\
 &= [0.9 - 1.96 \cdot 0.0212, 0.9 + 1.96 \cdot 0.0212] \\
 &= [0.8584, 0.9416].
 \end{aligned}$$

Mit 95%iger Wahrscheinlichkeit liegt der Anteil erfolgreicher Unternehmen in der Grundgesamtheit im Intervall  $[0.8584, 0.9416]$ .

- (ii) Ein Schätzer für die Chance  $p/(1-p)$  ist  $\hat{p}/(1-\hat{p})$  mit Erwartungswert

$$E \left[ \frac{\hat{p}}{1-\hat{p}} \right] = \frac{p}{1-p}.$$

Für die Varianz ergibt sich nach der Delta-Methode mit

$$g(p) = \frac{p}{1-p} \quad \text{und} \quad g'(p) = \frac{1}{((1-p)^2)}$$

$$\begin{aligned} \text{Var} \left[ \frac{\hat{p}}{1-\hat{p}} \right] &\stackrel{\text{Satz 3.29}}{=} [g'(p)]^2 \text{Var}(\hat{p}) \\ &= \frac{1}{(1-p)^4} \frac{p(1-p)}{n} \\ &= \frac{p}{(1-p)^3 n}. \end{aligned}$$

Das approximative Konfidenzintervall bestimmt man durch Auflösen nach  $p/(1-p)$

$$\begin{aligned} &P \left( -z_{1-\alpha/2} \leq \frac{\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}}{\sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}}} \leq z_{1-\alpha/2} \right) = 1 - \alpha \\ \Rightarrow \quad KI &= \left[ \frac{\hat{p}}{1-\hat{p}} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}}; \frac{\hat{p}}{1-\hat{p}} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}} \right]. \end{aligned}$$

Für die gegebenen Daten erhalten wir

$$\frac{\hat{p}}{1-\hat{p}} = 9,$$

d.h. die Chance, ein erfolgreiches Unternehmen zu gründen, ist 9mal so hoch wie für ein nicht-erfolgreiches Unternehmen. Mit

$$\frac{\hat{p}}{(1-\hat{p})^3 n} = 4.5$$

erhalten wir für das approximative Konfidenzintervall

$$[9 - 1.96 \cdot 2.1213; 9 + 1.96 \cdot 2.1213] = [4.8423; 13.1577],$$

die Chance auf eine erfolgreiche Unternehmensgründung liegt mit 95%iger Wahrscheinlichkeit im Intervall  $[4.8423; 13.1577]$ .

### 3.4 Testtheorie

- Zielstellungen statistischer Fragestellungen:

- (i) Wie lautet die zugrundeliegende Verteilung in der Grundgesamtheit?  $\Rightarrow$  **Schätztheorie**
- (ii) Erfüllt die zugrundeliegende Verteilung eine gewisse Eigenschaft?  $\Rightarrow$  **Testtheorie**

- Beispiel:

- Entwicklung einer neuen Behandlungsmethode II.
- **Frage:** Kann man aufgrund der Beobachtung  $\mathbf{x} = (x_1, \dots, x_n)$  sagen, dass Behandlungsmethode II besser ist als die bisher benutzte Methode I?
- **Ziel:** Entscheidung zwischen den beiden Hypothesen „II ist *nicht* besser als I“ und „II *ist* besser als I“.

- Eine **Hypothese**  $H$  ist eine Aussage über den zugrundeliegenden Parameter(vektor) bzw. über die zugrundeliegende Verteilung.

- **Definition:** Ein *statistisches Testproblem* (*statistischer Test*) besteht aus einer Nullhypothese  $H_0$  und einer Alternativhypothese  $H_1$ , die sich gegenseitig ausschließen und Aussagen über die gesamte Verteilung oder über bestimmte Parameter des interessierenden Merkmals in der Grundgesamtheit beinhalten.
- Ein statistischer Test basiert auf einer geeignet gewählten Prüfgröße (Teststatistik  $T(\mathbf{X}) = T(X_1, \dots, X_n)$  als Funktion der Stichprobenvariablen) und liefert eine formale Entscheidungsregel, die aufgrund der Stichprobe darüber entscheidet, ob eher  $H_0$  oder  $H_1$  für die Grundgesamtheit zutrifft.
- Einfachste Testprobleme sind solche mit Hypothesen der Form

$$H_0 : \theta \leq \theta_0 \quad \text{gegen} \quad H_1 : \theta > \theta_0 \quad \text{bzw.} \quad H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta \neq \theta_0,$$

bei denen  $\theta$  ein eindimensionaler Parameter und  $\theta_0$  ein bekannter Wert ist.

- Schätzung des unbekanntes Parameters  $\theta$  durch eine Statistik  $T(\mathbf{X}) = T(X_1, \dots, X_n)$
- Entscheidung für  $H_0$  oder  $H_1$ , je nachdem ob mit geeignet reellwertigem  $c$  bzw.  $c_1 < c_2$  gilt

$$T(\mathbf{x}) \leq c \quad \text{oder} \quad T(\mathbf{x}) > c \quad \text{bzw.} \quad T(\mathbf{x}) \in [c_1, c_2] \quad \text{oder} \quad T(\mathbf{x}) \notin [c_1, c_2].$$

- Partitionierung des Parameterraumes  $\Theta$  in zwei Teilmengen  $C$  und  $C^c$ .
  - $C$  ist Teilmenge derjenigen Punkte  $\theta \in \Theta$ , bei denen man sich für  $H_1$  entscheidet (sog. **Ab-  
lehnungsbereich**)
  - $C^c$  ist Teilmenge derjenigen Punkte  $\theta \in \Theta$ , bei denen man sich für  $H_0$  entscheidet (sog. **Annahmebereich**)
  - Für obige Testprobleme erhalten wir

$$C = \{\mathbf{x} : T(\mathbf{x}) > c\} \quad \text{bzw.} \quad C = \{\mathbf{x} : T(\mathbf{x}) < c_1\} \cup \{\mathbf{x} : T(\mathbf{x}) > c_2\}.$$

- **Bemerkungen:**
  - Unterscheidung zwischen zweiseitigen und einseitigen Tests
  - einfache vs. zusammengesetzte Hypothese
- **Fehlentscheidungen:** Bei einem statistischen Testproblem  $H_0$  gegen  $H_1$  und einem geeigneten statistischen Test spricht man von einem
  - *Fehler 1. Art* ( $\alpha$ -Fehler), wenn  $H_0$  verworfen wird, obwohl  $H_0$  wahr ist,
  - *Fehler 2. Art* ( $\beta$ -Fehler), wenn  $H_0$  beibehalten wird, obwohl  $H_1$  wahr ist.
- Von besonderem Interesse sind Tests,
  - deren Wahrscheinlichkeit für Fehler 1. Art eine vorgegebene „kleine“ *Irrtumswahrscheinlichkeit*  $\alpha \in (0, 1)$  nicht überschreiten und
  - unter dieser Nebenbedingung die Wahrscheinlichkeit für Fehler 2. Art minimieren.
- **Definition:** Ein statistischer Test heißt *Signifikanztest* zum Signifikanzniveau  $\alpha$ ,  $0 < \alpha < 1$ , falls

$$P(H_1 \text{ annehmen} \mid H_0 \text{ wahr}) \leq \alpha,$$

d.h

$$P(\text{Fehler 1. Art}) \leq \alpha.$$

- **Bemerkungen:**

- (i) Typische Werte für das Signifikanzniveau  $\alpha$  sind 0.1, 0.05, 0.01.
- (ii) Ungleichbehandlung von Fehlerwahrscheinlichkeiten 1. und 2. Art  $\implies$  Es kommt bei der Formulierung derartiger Testprobleme darauf an, welche der beiden Hypothesen als  $H_0$  und welche als  $H_1$  gewählt werden.
- (iii) Zusammenhang zu Konfidenzintervallen: Ein (zweiseitiges)  $(1 - \alpha)$ -Konfidenzintervall entspricht dem Annahmehereich des zugehörigen (zweiseitigen) Signifikanztests.  $\alpha$  kann als „Fehlerwahrscheinlichkeit“ interpretiert werden, dass der unbekannte Parameter  $\theta$  **nicht** vom Konfidenzintervall überdeckt wird.

• **Vorgehensweise bei statistischen Testproblemen:**

- (1) Quantifizierung des inhaltlichen Problems
- (2) Formulierung der Modellannahmen
- (3) Formulierung der Null- und Alternativhypothesen
- (4) Festlegung des Signifikanzniveaus
- (5) Bestimmung von Teststatistik, sowie Annahme- und Ablehnungsbereich des Tests
- (6) Berechnung der Realisation der Teststatistik für konkrete Stichprobe
- (7) Testentscheidung

**Anmerkung 1: Warum Teststatistik unter Gültigkeit von  $\mu = \mu_0$ ? (Vgl. Abb. 1)**

- Betrachte  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$
- $X_1, \dots, X_n$  seien unabhängig und identisch verteilt mit  $X_i \sim N(\mu, \sigma^2)$  und  $\sigma^2$  bekannt
- Fall 1:  $\mu = \mu_0$

– Dann gilt

$$Z = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \sim N(0, 1).$$

– Ablehnungsbereich:  $C = \{z : z > z_{1-\alpha}\}$ , mit  $z_{1-\alpha}$  als  $(1 - \alpha)$ -Quantil der Standardnormalverteilung.

- Fall 2:  $\mu < \mu_0$

– Es gilt

$$Z = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} = \sqrt{n} \frac{\bar{X} - \mu + \mu - \mu_0}{\sigma} \sim N\left(\sqrt{n} \frac{\mu - \mu_0}{\sigma}, 1\right)$$

– Daraus folgt: Für  $\mu < \mu_0$  ist  $P(Z > z_{1-\alpha}) < \alpha$ , d.h. die Wahrscheinlichkeit  $H_0$  fälschlicherweise abzulehnen ist noch kleiner als  $\alpha$ .

**Anmerkung 2: Asymmetrische Testentscheidungen**

- Testentscheidungen sind asymmetrisch:  $H_0$  beibehalten vs.  $H_0$  verwerfen.  
 $\implies H_0$  wird durch  $\alpha$  geschützt.
- $z$  nicht im Ablehnungsbereich heißt nur, Befund spricht nicht stark genug gegen  $H_0$   
 $\implies H_0$  nicht bestätigt, nur nicht abgelehnt
- $z$  im Ablehnungsbereich heißt, trotz Schutz durch  $\alpha$  setzt sich der Datenbefund gegen  $H_0$  durch  
 $\implies H_1$  bestätigt, „signifikant nachgewiesen“

$\implies$  Nachzuweisende Hypothese in  $H_1$  !

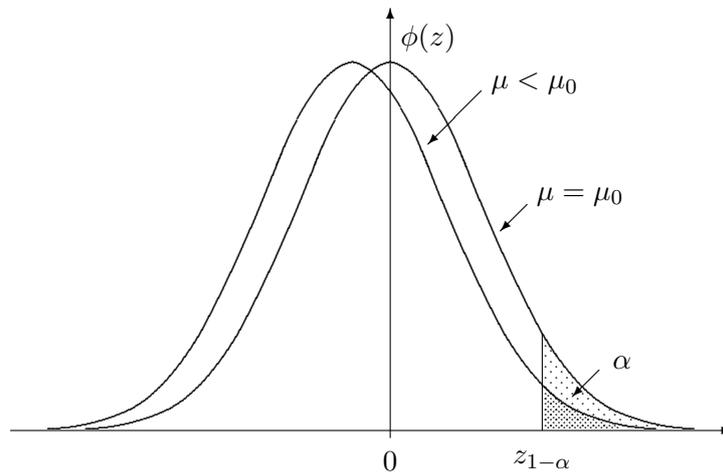


Abbildung 1: Ablehnungsbereiche für das Testproblem  $H_0 : \mu \leq \mu_0$  und  $H_1 : \mu > \mu_0$  basierend auf  $Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$

### Beispiele:

- Es soll überprüft werden, ob der Ausschussanteil  $\pi$  in einer Lieferung größer als 10 Prozent ist.
- Zwei Hypothesenpaare möglich:

$$(i) \quad H_0 : \pi \leq 0.1 \quad H_1 : \pi > 0.1 \quad (ii) \quad H_0 : \pi \geq 0.1 \quad H_1 : \pi < 0.1.$$

- Entscheidung für Hypothesenpaar (i). Begründung:
  - $P(H_1|H_0) = \alpha$  entspricht der Wahrscheinlichkeit, dass der wahre Ausschussanteil kleiner als 10 Prozent ist, obwohl der Test das nicht erkennt. In diesem Fall würde eine Nachjustierung suggeriert werden, die aber gar nicht notwendig ist.
  - Durch Vorgabe von  $\alpha$  kann dieses Risiko einer Fehlentscheidung kontrolliert werden.
  - $P(H_0|H_1) = \beta$  (unbekannt) entspricht der Wahrscheinlichkeit, dass der wahre Ausschussanteil größer als 10 Prozent ist, obwohl der Test das nicht erkennt. In diesem Fall würde eine Nachjustierung unterbleiben, obwohl sie erforderlich wäre.
  - Das Risiko einer unnötigen Nachjustierung wird höher eingeschätzt (und soll daher über die Irrtumswahrscheinlichkeit  $\alpha$  abgesichert werden) als das Risiko von Regressforderungen.
- Kontrollkarten: Aufwand einer Nachjustierung ist sehr groß, daher Absicherung, ob Prozeß tatsächlich außer Kontrolle ist, d.h. das größere Risiko steckt in  $H_1$ .

### Gütefunktion

- Ein einführendes Beispiel:
  - Annahme: Die wahre produzierte Bleistiftlänge betrage  $\mu = 17.2[cm]$ . Der Sollwert ist aber als  $\mu_0 = 17[cm]$  festgelegt.
  - Problem: Ein statistischer Test wird diesen Unterschied nur schwer erkennen und dazu tendieren,  $H_0$  nicht zu verwerfen.
  - Fazit:  $P(H_0|H_1) = \beta$  wird „groß“ sein.
  - Liegt das (wahre) Mittel der produzierten Bleistiftlängen jedoch bei  $\mu = 20[cm]$ , dann wird  $P(H_0|H_1) = \beta$  „klein“ sein. (Es ist sehr wahrscheinlich, dass die Nullhypothese verworfen wird.)

- Fazit: Offensichtlich ist die Größe des Fehlers 2. Art ( $\beta$ -Fehler) eine Funktion des wahren Parameters  $\mu$  (bzw. seiner Abweichung von  $\mu_0$ ).  $\implies$  Operationalisierung über Gütefunktion.

- **Definition:** Die *Gütefunktion*  $g$  gibt für ein vorgegebenes Signifikanzniveau  $\alpha$  und festen Stichprobenumfang  $n$  die Wahrscheinlichkeit für einen statistischen Test an, die Nullhypothese zu verwerfen, d.h.

$$g(\mu) := P(H_0 \text{ ablehnen} | \mu) = P(C | \mu).$$

- $g(\mu)$  gibt für die verschiedensten Werte des unbekanntes, aber wahren Parameters  $\mu$  die Wahrscheinlichkeit an,  $H_0$  zu verwerfen. Gilt
  - $\mu \in H_0$ , so ist  $g(\mu) \leq \alpha$ ,
  - $\mu \in H_1$ , so ist  $1 - g(\mu)$  die Wahrscheinlichkeit für einen  $\beta$ -Fehler (Operationscharakteristik (OC)).

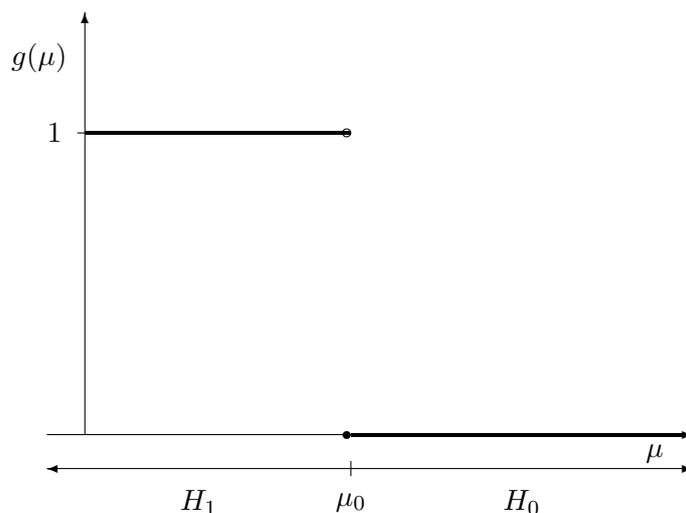


Abbildung 2: Verlauf einer idealen Gütefunktion

- Im Folgenden leiten wir die Gütefunktion für den einseitigen Gauß-Test her mit den Hypothesen

$$H_0 : \mu \geq \mu_0 \quad H_1 : \mu < \mu_0$$

$$\begin{aligned}
 g(\mu) &= P(H_0 \text{ verwerfen} | \mu) \\
 &= P\left(\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} < z_\alpha | \mu\right) \\
 &= P\left(\frac{\bar{x} - \mu + \mu - \mu_0}{\sigma} \sqrt{n} < z_\alpha | \mu\right) \\
 &= P\left(\frac{\bar{x} - \mu}{\sigma} \sqrt{n} < z_\alpha - \frac{\mu - \mu_0}{\sigma} \sqrt{n} | \mu\right) \\
 &= P\left(Z < z_\alpha - \frac{\mu - \mu_0}{\sigma} \sqrt{n} | \mu\right) \\
 &= \Phi\left(z_\alpha - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)
 \end{aligned}$$

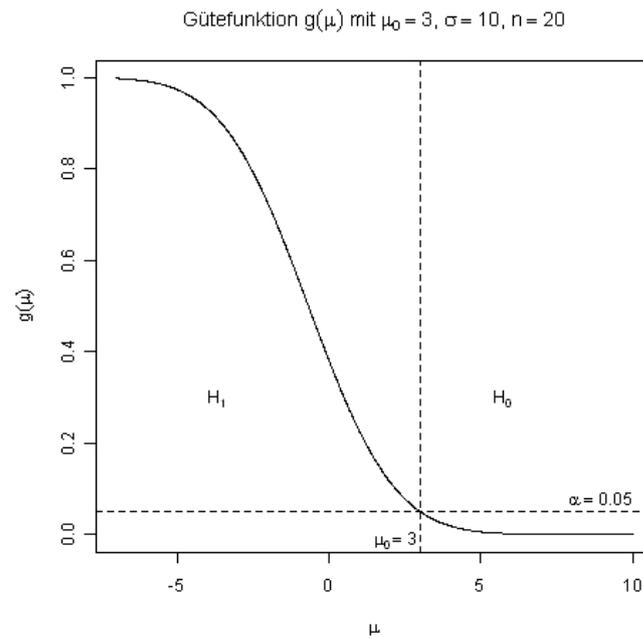


Abbildung 3: Gütefunktion zum Gauß-Test mit  $H_0 : \mu \geq \mu_0$  vs.  $H_1 : \mu < \mu_0$

- Analog: Herleitung der Gütefunktion für den einseitigen Gauß-Test mit den Hypothesen

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$$

$$\begin{aligned}
 g(\mu) &= P(H_0 \text{ verwerfen} \mid \mu) \\
 &= P\left(\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha} \mid \mu\right) \\
 &= P\left(\frac{\bar{x} - \mu + \mu - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha} \mid \mu\right) \\
 &= P\left(\frac{\bar{x} - \mu}{\sigma} \sqrt{n} > z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \mid \mu\right) \\
 &= P\left(Z > z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \mid \mu\right) \\
 &= 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)
 \end{aligned}$$

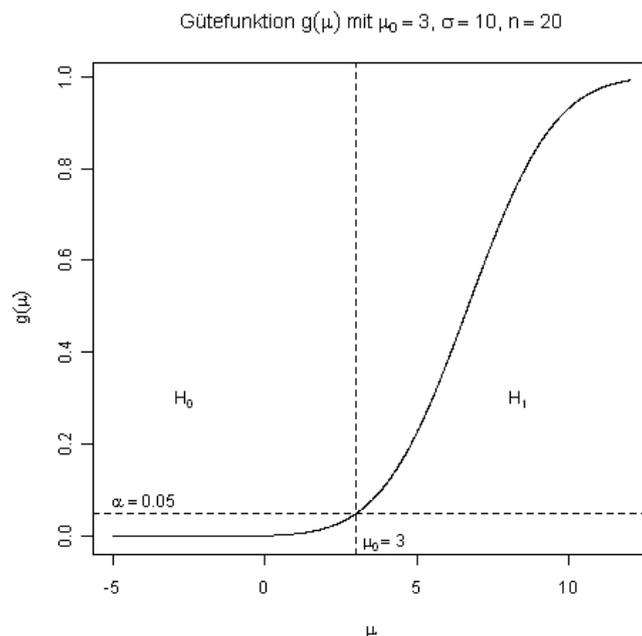


Abbildung 4: Gütefunktion zum Gauß-Test mit  $H_0 : \mu \leq \mu_0$  vs.  $H_1 : \mu > \mu_0$

#### Beispiel (Kontrollkarten):

- Testproblem:

$$H_0 : \mu \leq 17 \text{ [cm]} \quad \text{gegen} \quad H_1 : \mu > 17 \text{ [cm]}.$$

- Sei  $\alpha = 0.05$  und  $n = 10$ . Standardabweichung sei mit  $\sigma = 1.5$  als bekannt vorausgesetzt.
- Dann ist die Gütefunktion gegeben als

$$g(\mu) = 1 - \Phi \left( z_{0.95} - \frac{\mu - 17}{1.5} \sqrt{10} \right) = 1 - \Phi \left( 1.64 - \frac{\mu - 17}{1.5} \cdot 3.16 \right).$$

- Die Werte der Funktion  $g(\mu)$  können aus der Tabelle der Verteilungsfunktion der Standardnormalverteilung für verschiedene Werte von  $\mu$  abgelesen werden. Man erhält folgende Wertetabelle:

$\mu$	16	16.5	17	17.5	18	18.5	19
$g(\mu)$	0	0.003	0.05	0.279	0.68	0.936	0.995

- Als Rechenbeispiel betrachte man  $\mu = 17.5$ , wofür sich ergibt:

$$g(17.5) = 1 - \Phi \left( 1.64 - \frac{17.5 - 17}{1.5} \cdot 3.16 \right) = 1 - \Phi(0.59) = 0.279.$$

Man sieht,

- die Wahrscheinlichkeit,  $H_0$  korrekterweise zu verwerfen, ist für  $\mu = 17.5$  [cm] mit 0.279 sehr klein.
- Obwohl wir sicher wissen, daß mit  $\mu = 17.5$  cm die Alternative zutrifft, fällt die Entscheidung des Tests mit einer großen Wahrscheinlichkeit von  $1 - 0.279 = 0.721$  fälschlich für  $H_0$ .
- Je größer die Abweichung von  $\mu_0$ , also je größer der zu entdeckende Effekt ist, desto kleiner wird die Wahrscheinlichkeit für den Fehler 2. Art.

- Die Wahrscheinlichkeit für den Fehler 1. Art ist für alle Werte  $\mu$  aus  $H_0$  kleiner oder gleich  $\alpha = 0.05$ .
- Für  $\mu = \mu_0$  nimmt  $g(\mu)$  den Wert des Signifikanzniveaus  $\alpha$  an.
- Für  $\mu \in H_1$  gibt die Gütefunktion  $1 -$  die Wahrscheinlichkeit für den Fehler 2. Art an.
- Für  $\mu \in H_0$  gibt die Gütefunktion die Wahrscheinlichkeit für den Fehler 1. Art an.
- Der Fehler 1. Art ist durch  $\alpha$  an der Stelle  $\mu = \mu_0$  nach oben beschränkt.

#### Gütefunktion für den Gaußtest:

Für vorgegebenes Signifikanzniveau  $\alpha$  und festen Stichprobenumfang  $n$  gibt die Gütefunktion  $g$  die Wahrscheinlichkeit für einen statistischen Test an, die Nullhypothese zu verwerfen.

Speziell für den Gauß-Test ergibt sich die Gütefunktion  $g(\mu)$  im Fall des Testproblems

- (a)  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  als

$$g(\mu) = \Phi\left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right) + \Phi\left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$

- (b)  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$  als

$$g(\mu) = \Phi\left(z_\alpha - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$

- (c)  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  als

$$g(\mu) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right),$$

wobei  $\Phi$  die Verteilungsfunktion der  $N(0, 1)$ -Verteilung bezeichnet.

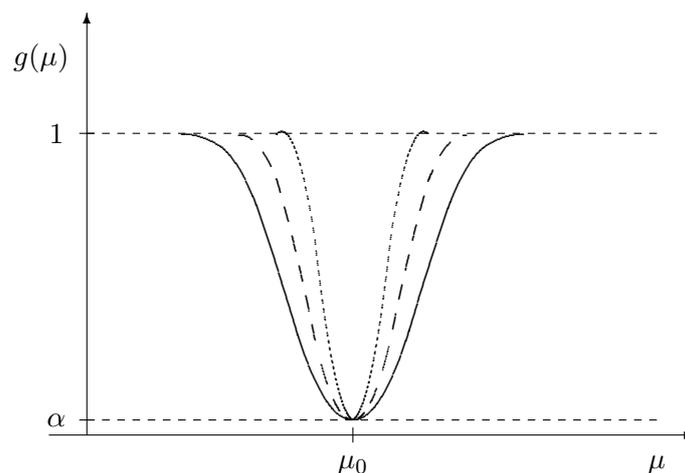


Abbildung 5: Skizze einer Gütefunktion des zweiseitigen Gauß-Tests mit  $n = 10$ ,  $n = 20$ ,  $n = 50$

- Die Gütefunktion erlaubt Aussagen über die Qualität eines statistischen Tests.
- Sie enthält nicht nur Informationen darüber, für welche Parameterwerte die Nullhypothese mit großer Wahrscheinlichkeit verworfen wird, sondern auch das Signifikanzniveau.

- Diese Zweiteilung bei der Interpretation der Gütefunktion spiegelt sich auch in ihrer Namensgebung wider. Für Werte aus der Alternative spricht man von der Gütefunktion auch als *Macht*, *Trennschärfe* oder Power eines Tests.
- Gütefunktionen werden daher zum Vergleich mehrerer konkurrierender Tests zu einem Testproblem herangezogen. Man wählt, falls möglich, den Test unter allen Niveau- $\alpha$ -Tests aus, der die größte Macht besitzt, und der somit, wie bereits zu Beginn dieses Unterkapitels formuliert, die geringste Wahrscheinlichkeit für einen Fehler 2. Art besitzt.

#### Eigenschaften der Gütefunktion eines statistischen Tests:

- Für Werte aus  $H_1$  heißt die Gütefunktion Trennschärfe oder Macht.
- Für Werte aus  $H_0$  ist die Gütefunktion kleiner gleich  $\alpha$ .
- Für wachsendes  $n$  wird die Macht eines Tests größer, d.h. die Gütefunktion wird steiler.
- Für wachsendes  $\alpha$  wird die Macht eines Tests größer.
- Für eine wachsende Abweichung zwischen Werten aus  $H_1$  und  $H_0$  wird die Macht eines Tests größer.

#### Beispiel Kontrollkarten:

- Abhängigkeit vom Stichprobenumfang:
  - $n = 10$  :  $g(17.5) = 1 - \Phi(1.64 - \frac{17.5-17}{1.5} \cdot \sqrt{10}) = 1 - \Phi(0.59) = 0.279$ ,
  - $n = 50$  :  $g(17.5) = 1 - \Phi(1.64 - \frac{17.5-17}{1.5} \cdot \sqrt{50}) = 1 - \Phi(-0.71) = 0.761$ ,
  - $n = 100$  :  $g(17.5) = 1 - \Phi(1.64 - \frac{17.5-17}{1.5} \cdot \sqrt{100}) = 1 - \Phi(-1.69) = 0.954$ ,
  - Interpretation: Die Wahrscheinlichkeit  $H_0$  für  $\mu = 17.5$  richtigerweise zu verwerfen, wächst mit steigendem Stichprobenumfang.
- Abhängigkeit der Gütefunktion vom Signifikanzniveau:
  - $\alpha = 0.01$  :  $g(17.5) = 1 - \Phi(2.3263 - \frac{17.5-17}{1.5} \sqrt{10}) = 1 - \Phi(1.27) = 0.102$ ,
  - $\alpha = 0.1$  :  $g(17.5) = 1 - \Phi(1.2816 - \frac{17.5-17}{1.5} \sqrt{10}) = 1 - \Phi(0.23) = 0.41$ .
  - Interpretation: Die Wahrscheinlichkeit  $H_0$  für  $\mu = 17.5$  richtigerweise zu verwerfen, wächst mit steigendem Signifikanzniveau  $\alpha$ . Bei wachsender Wahrscheinlichkeit für einen  $\alpha$ -Fehler, sinkt die Wahrscheinlichkeit für einen  $\beta$ -Fehler.