

⟨ S t a t W t h 1 7 ⟩

7. Schließende Statistik

Werner Gurker

Institut für Stochastik und Wirtschaftsmathematik

Technische Universität Wien

Schließende Statistik

Die **Grundaufgabe** der **schließenden Statistik** besteht darin, basierend auf **Stichproben** – d. h. Daten oder Beobachtungen – Rückschlüsse auf das zu Grunde liegende („datengenerierende“) **statistische Modell** zu ziehen.

Häufig sind statistische Modelle durch **Parameter** charakterisiert und die Aufgabe besteht konkreter darin:

- ▶ diese Parameter zu **schätzen**
- ▶ Aussagen über die **Genauigkeit** der Schätzungen zu treffen
- ▶ **Hypothesen** über die Parameter zu **testen**

Naturgemäß ist das nur unter Inkaufnahme von mehr oder weniger großen **Unsicherheiten** möglich.

Parametrische Statistik

Parametrische statistische Modelle sind dadurch charakterisiert, dass sie sich durch einen ein- oder mehrdimensionalen **Parameter** $\theta \in \Theta \subseteq \mathbb{R}^k$ beschreiben lassen.

Die Menge Θ aller möglichen Parameter nennt man den **Parameterraum**.

Bsp1: Ein **diskretes** parametrisches Modell ist etwa die Klasse („Familie“) aller $B(n, p)$ –Verteilungen (mit festem $n \in \mathbb{N}$):

$$\mathcal{P} = \{B(n, p) \mid p \in (0, 1)\}, \quad \Theta = (0, 1) \subset \mathbb{R}$$

Bsp2: Ein **stetiges** parametrisches Modell ist etwa die Klasse („Familie“) aller $N(\mu, \sigma^2)$ –Verteilungen:

$$\mathcal{P} = \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, 0 < \sigma^2 < \infty\}, \quad \Theta = \mathbb{R} \times \mathbb{R}^+ \subset \mathbb{R}^2$$

Nichtparametrische Statistik

Im Gegensatz zu parametrischen Modellen können **nichtparametrische statistische Modelle** nicht durch einen **endlichdimensionalen** Parameter charakterisiert werden.

Bsp: Beispielsweise definiert ein Modell der folgenden Art:

$$\mathcal{P} = \{F \mid F \text{ eine stetige Verteilungsfunktion}\}$$

ein **nichtparametrisches Modell**. Der „Parameterraum“:

Klasse aller stetigen VFn

ist nicht **endlichdimensional**.

Stichprobe

Die sGn X_1, X_2, \dots, X_n bilden eine **Stichprobe** (SP) einer sG X , wenn die X_i ua. und so wie X verteilt sind, d. h., wenn es sich um **iid-Größen** handelt. Häufig schreibt man die SP als **Vektor**: $\mathbf{X} = (X_1, X_2, \dots, X_n)'$.

Achtung: Man unterscheide genau zwischen den **sGn** X_1, X_2, \dots, X_n und ihren **Realisationen** x_1, x_2, \dots, x_n (= **konkrete** Beobachtungen).

Ist $p(x)$ (bzw. $f(x)$) die **W-Funktion** (bzw. **Dichte**) von X , so ist die **gemeinsame Verteilung** der Stichprobe gegeben durch:

$$\text{diskret: } p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

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

Statistiken

Allgemein nennt man eine **Funktion** $T(\mathbf{X}) = T(X_1, X_2, \dots, X_n)$ einer **Stichprobe** X_1, X_2, \dots, X_n eine **Statistik**.

Handelt es sich bei T um eine Abbildung in den Parameterraum Θ , d. h., gilt $T : \mathbb{R}^n \rightarrow \Theta$, nennt man T eine **Schätzfunktion** oder kurz einen **Schätzer** für den Parameter $\theta \in \Theta$. In diesem Fall schreibt man:

$$\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$$

Ebenso verfährt man bei anderen unbekannten Größen. So bezeichnet beispielsweise $\hat{F}_n(x)$ einen **Schätzer** (auf Basis von n Beobachtungen) für die **Verteilungsfunktion** $F(x)$.

Beispiele

Statistiken sind uns schon an mehreren Stellen begegnet.

Beispielsweise sind die in **Kapitel 1** diskutierten **grafischen** Darstellungen (Balkendiagramm, Histogramm, Boxplot, ...) von Daten x_1, x_2, \dots, x_n verschiedener Art **Statistiken** im obigen Sinn.

Ebenso handelt es sich bei den diversen **Kennzahlen** (Mittelwert, Median, Streuung, ...) um Beispiele für **Schätzfunktionen**.

Die zwei wichtigsten Schätzfunktionen sind der **Stichprobenmittelwert** \bar{X}_n und die **Stichprobenvarianz** S_n^2 .

Stichprobenmittelwert und -varianz

Ist X_1, \dots, X_n eine **Stichprobe** der sG X , so sind **Stichprobenmittelwert** und **Stichprobenvarianz** gegeben durch:

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

Allgemein gilt: \bar{X}_n ist ein Schätzer für den **Erwartungswert** (Mittelwert) $\mu = \mathbb{E}(X)$ und S_n^2 ist ein Schätzer für die **Varianz** $\sigma^2 = \text{Var}(X)$ der sG X .

Ein Schätzer für die **Streuung** $\sigma = \sqrt{\text{Var}(X)}$ ist die **Stichprobenstreuung**:

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

Schätzproblem

X_1, X_2, \dots, X_n sei eine **Stichprobe** aus einer sG X mit VF F .

Besteht die Vermutung, dass die VF $F(x) = F(x; \theta)$, $\theta \in \Theta$, zu einer bestimmten **parametrischen** Verteilungsfamilie (z. B., Normalverteilungen) gehört, möchte man in einem ersten Schritt herausfinden, welcher Wert des Parameters θ zur vorliegenden Stichprobe geführt hat.

M. a. W., man möchte den/die unbekannten Parameter θ – auf Basis der vorliegenden Stichprobe – **schätzen**.

Vielfach liegt aber kein bestimmtes parametrisches Modell zu Grunde, sondern man weiß z. B. nur, dass X stetig und positiv ist.

In diesem Fall möchte man zuerst – auf Basis der Stichprobe – die VF F selbst **schätzen** (oder, falls X stetig ist, auch die Dichte $f = F'$).

Empirische Verteilungsfunktion

Die **Verteilung** einer sG X ist durch ihre **Verteilungsfunktion** spezifiziert:

$$F(x) = P(X \leq x) \quad \text{für } x \in \mathbb{R}$$

Hat man eine **Stichprobe** X_1, X_2, \dots, X_n von X , so kann man F mittels der schon früher diskutierten **empirischen Verteilungsfunktion** schätzen:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \quad \text{für } x \in \mathbb{R}$$

Für ein festes $x \in \mathbb{R}$ ist $\hat{F}_n(x)$ der **Anteil der Beobachtungen** aus der Stichprobe, die kleiner oder gleich x sind.

Eigenschaften der empirischen VF

Für festes $x \in \mathbb{R}$ gilt:

$$(1) \mathbb{E}[\hat{F}_n(x)] = F(x)$$

$$(2) \text{Var}[\hat{F}_n(x)] = \frac{F(x)[1 - F(x)]}{n} \longrightarrow 0 \quad \text{für } n \longrightarrow \infty$$

Bem: Die Varianz von $\hat{F}_n(x)$ ist also für den **Median** \tilde{x} , d. h. für $F(\tilde{x}) = 0.5$, am größten, in den Ausläufern von F am kleinsten.

$$(3) \hat{F}_n(x) \xrightarrow{P} F(x) \quad \text{für } n \longrightarrow \infty$$

Beweis: Skriptum S. 253, Buch S. 259

Fundamentalsatz der Statistik (Gliwenko-Cantelli)

Für eine Stichprobe X_1, X_2, \dots, X_n von $X \sim F$ gilt:

$$P \left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = 0 \right) = 1$$

D. h., mit **Wahrscheinlichkeit 1** konvergiert die empirische VF \hat{F}_n **gleichmäßig** gegen die zu Grunde liegende VF F .

Beispiel zum Fundamentalsatz

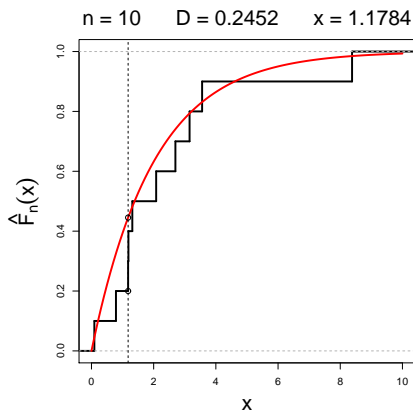
Wir simulieren Beobachtungen einer $\text{Exp}(\tau = 2)$ -Verteilung, zeichnen die empirische VF \hat{F}_n und bestimmen Stelle und Wert des größten Abstands D_n zur (theoretischen) VF F :

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - (1 - e^{-x/2})|$$

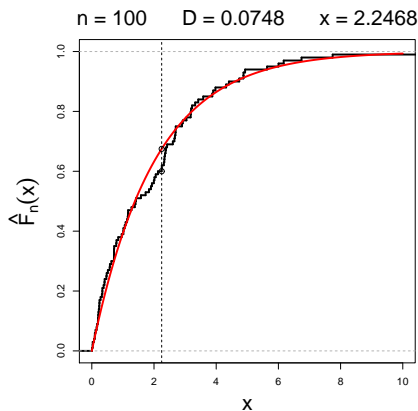
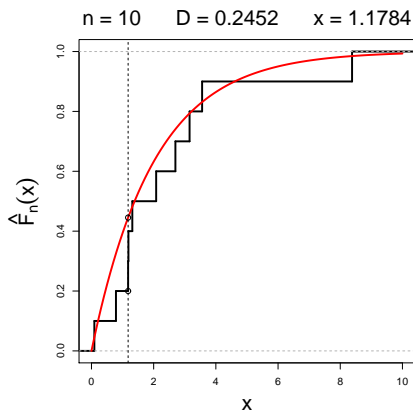
Die folgenden Abbildungen zeigen das Ergebnis für eine **kleine** Stichprobe (z. B. $n = 10$) und für eine **große** Stichprobe (z. B. $n = 100$).

Deutlich zeigt sich der über ganz \mathbb{R}^+ **gleichmäßig kleinere Abstand** von \hat{F}_n und F für die größere Stichprobe.

Beispiel zum Fundamentalsatz (Forts.)



Beispiel zum Fundamentalsatz (Forts.)



Momentenmethode

k -tes Moment einer **sG** X : $\mathbb{E}(X^k)$

k -tes Moment einer **Stichprobe** X_1, X_2, \dots, X_n : $\frac{1}{n} \sum_{i=1}^n X_i^k$

Momentenmethode: Gibt es in einem Verteilungsmodell m unbekannte Parameter, $\theta_1, \theta_2, \dots, \theta_m$, so lassen sich durch Auflösen des folgenden Gleichungssystems:

$$\mathbb{E}(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots, m$$

Schätzer (die **Momentenschätzer**) für die Parameter gewinnen.

Bsp zur Momentenmethode im Skriptum S. 256, Buch S. 262.

Maximum Likelihood (ML) Schätzer: diskreter Fall

Ist X eine **diskrete** sG mit der W-Funktion $p(x; \theta)$, wobei $\theta \in \Theta$ ein einzelner unbekannter Parameter ist, und sind x_1, x_2, \dots, x_n (konkrete) Beobachtungen einer Stichprobe X_1, X_2, \dots, X_n von X , so ist die **Likelihood-Funktion** (kurz **Likelihood**) der Stichprobe definiert durch:

$$L(\theta) = \prod_{i=1}^n p(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der **Maximum-Likelihood-Schätzer** (kurz **ML-Schätzer**) von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Interpretation: Der ML-Schätzer ist jener θ -Wert, der die Beobachtung der (konkreten) Stichprobe x_1, x_2, \dots, x_n am wahrscheinlichsten (oder plausibelsten) macht.

Beispiel

Bei einer Überprüfung von $n = 100$ zufällig ausgewählten ICs aus einer bestimmten Produktion sind 13 ICs defekt und 87 ICs intakt. Wie lautet der **ML-Schätzwert** des Defektanteils θ bei dieser Produktion?

Likelihood: $L(\theta) = \theta^{13}(1 - \theta)^{87}$

Log-Likelihood: $l(\theta) = \ln L(\theta) = (13) \ln \theta + (87) \ln(1 - \theta)$

Ableiten und Nullsetzen:

$$\frac{d l(\theta)}{d \theta} = \frac{13}{\theta} - \frac{87}{1 - \theta} = 0 \quad \Longrightarrow \quad \hat{\theta} = \frac{13}{100} = 0.13$$

Bem: Eine ausführlichere Diskussion dieses Beispiels findet man im Skriptum auf S. 258, im Buch auf S. 264.

Schätzer / Schätzwert

Man unterscheide genau zwischen dem **Schätzer** (oder **Schätzfunktion**) und dem **Schätzwert**. Schätzer sind **sGn**, Schätzwerte aber (konkrete) **Zahlenwerte**.

Z. B. ist der **ML-Schätzer** von θ im vorigen Beispiel gegeben durch:

$$\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{eine sG!}$$

wobei $X_i = 1$, falls der i -te überprüfte IC defekt ist, und $X_i = 0$ im anderen Fall. Hingegen ist der **ML-Schätzwert** von θ gegeben durch:

$$\hat{\theta} = \bar{x} = \frac{13}{100} = 0.13 \quad \text{ein Zahlenwert!}$$

Maximum Likelihood (ML) Schätzer: stetiger Fall

Ist X eine **stetige** sG mit der Dichte $f(x; \theta)$, wobei $\theta \in \Theta$ ein einzelner unbekannter Parameter ist, und sind x_1, x_2, \dots, x_n Beobachtungen einer Stichprobe X_1, X_2, \dots, X_n von X , so ist die **Likelihood-Funktion** (kurz **Likelihood**) der Stichprobe definiert durch:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der **ML-Schätzer** von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Interpretation: Die Interpretation, die für den diskreten Fall gegeben wurde, lässt sich sinngemäß auf den stetigen Fall übertragen.

Maximum Likelihood Schätzung: mehrere Parameter

Die **ML-Methode** lässt sich auch für die Schätzung von mehreren Parametern anwenden, im Folgenden formuliert nur für den stetigen Fall (diskreter Fall analog).

Ist X eine stetige sG mit der Dichte $f(x; \theta)$, wobei $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$ ein k -dimensionaler Parameter aus $\Theta \subseteq \mathbb{R}^k$ ist, so ist für eine (konkrete) Stichprobe x_1, x_2, \dots, x_n von X die **Likelihood (-Funktion)** gegeben durch:

$$L(\theta) = L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der **ML-Schätzer** von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Beispiel: $X \sim N(\mu, \sigma^2)$

Wie lauten die **ML-Schätzer** für μ und σ^2 ?

Für eine (konkrete) **Stichprobe** x_1, x_2, \dots, x_n von $X \sim N(\mu, \sigma^2)$ ist die **Likelihood** gegeben durch:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right], \quad (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ \end{aligned}$$

Die Funktion $L(\mu, \sigma^2)$ ist nach μ und σ^2 zu **maximieren**.

Es ist einfacher, dafür die **Log-Likelihood** $\ln L(\mu, \sigma^2)$ zu nehmen.

Beispiel: $X \sim N(\mu, \sigma^2)$ (Forts.)

Die **Log-Likelihood** lautet wie folgt:

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

Partiell ableiten und **Nullsetzen**:

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \implies \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \implies \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Invarianzeigenschaft der ML-Schätzung

Die **ML-Schätzmethode** hat eine sehr nützliche Eigenschaft unter **Transformationen** des Parameter(vektor)s $\theta \in \Theta$.

X_1, X_2, \dots, X_n sei eine **Stichprobe** von $X \sim f(x; \theta)$ (oder $X \sim p(x; \theta)$) und $\eta = g(\theta)$ sei eine **Funktion** des Parameters. Ist $\hat{\theta}$ der **ML-Schätzer** von θ , so ist der **ML-Schätzer** von η gegeben durch:

$$\hat{\eta} = g(\hat{\theta}) = g(\hat{\theta})$$

Bsp: Im vorigen Beispiel ist etwa der ML-Schätzer für die **Streuung** $\sigma = \sqrt{\sigma^2} = g(\mu, \sigma^2)$ ohne weitere Rechnung gegeben durch:

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

Erwartungstreue

Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ für einen Parameter $\theta \in \Theta$ heißt **erwartungstreu** (oder **unverzerrt**), wenn:

$$\mathbb{E}_{\theta}(\hat{\theta}_n) = \theta \quad \text{für alle } \theta \in \Theta$$

Gilt für $n \rightarrow \infty$, dass:

$$\mathbb{E}_{\theta}(\hat{\theta}_n) \rightarrow \theta \quad \text{für alle } \theta \in \Theta$$

nennt man $\hat{\theta}_n$ **asymptotisch erwartungstreu** (oder **unverzerrt**).

Interpretation: Bei Verwendung eines erwartungstreuer Schätzers macht man keinen **systematischen** Fehler, sondern liegt **im Durchschnitt** (d. h. im Erwartungswert) an der richtigen Stelle.

Beispiel: Stichprobenmittelwert und –varianz

Der **Stichprobenmittelwert** \bar{X}_n ist ein **erwartungstreuer** Schätzer für den **Erwartungswert** $\mu = \mathbb{E}(X)$ einer sG X :

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{\mu} = \frac{n\mu}{n} = \mu$$

Ebenso ist die **Stichprobenvarianz** S_n^2 ein **erwartungstreuer** Schätzer für die **Varianz** $\sigma^2 = \text{Var}(X)$ einer sG X :

$$\mathbb{E}(S_n^2) = \mathbb{E}\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right] = \sigma^2$$

Beweis: Skriptum S. 265, Buch S. 271

Effizienz

Neben dem Erwartungswert spielt auch die Varianz eine wesentliche Rolle bei der Beurteilung von Schätzern. Man sagt, dass ein erwartungstreuer Schätzer $\hat{\theta}_1$ des Parameters θ effizienter als ein anderer erwartungstreuer Schätzer $\hat{\theta}_2$ desselben Parameters ist, wenn:

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$

Ist ein erwartungstreuer Schätzer des Parameters θ effizienter als jeder andere erwartungstreue Schätzer desselben Parameters, nennt man ihn effizient.

Bem: Abgesehen von Spezialfällen (s. nächste Folie) erfordert der Nachweis der Effizienz eines Schätzers meist tieferliegende Aussagen der (mathematischen) Statistik.

Lineare effiziente Schätzer

Beschränkt man sich auf **lineare Schätzer**, d. h. auf Schätzer der Form:

$$T_n = \sum_{i=1}^n a_i X_i, \quad a_i \in \mathbb{R}$$

kann man zeigen (Skriptum S. 266, Buch S. 272), dass der **Stichprobenmittelwert** \bar{X}_n der (eindeutig bestimmte) **linear effiziente Schätzer** des **Erwartungswerts** $\mu = \mathbb{E}(X)$ einer sG X ist.

Bsp: Für $X \sim \text{Exp}(\tau)$ gilt $\tau = \mathbb{E}(X)$. Der Stichprobenmittelwert \bar{X}_n ist nicht nur **erwartungstreu**, sondern im obigen Sinn auch **linear effizient** für die Schätzung von τ . D. h., es gibt keinen anderen linearen (erwartungstreuen) Schätzer für τ , der eine kleinere Varianz als \bar{X}_n hat.

Konsistenz

Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ heißt (schwach) **konsistent** für den Parameter $\theta \in \Theta$, wenn:

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{für} \quad n \longrightarrow \infty$$

Interpretation: Anschaulich bedeutet **Konsistenz**, dass sich ein Schätzer mit dieser Eigenschaft für wachsendes n mit **hoher Wahrscheinlichkeit** in der **Nähe** des zu schätzenden Parameters aufhält. Letzteres ist eine sehr wünschenswerte Eigenschaft von „guten“ Schätzern.

Beispiel: Stichprobenmittelwert und –varianz

Der **Stichprobenmittelwert** \bar{X}_n und die **Stichprobenvarianz** S_n^2 sind **konsistente** Schätzer für $\mu = \mathbb{E}(X)$ bzw. $\sigma^2 = \text{Var}(X)$:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}(X) = \mu, \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \xrightarrow{P} \sigma^2$$

Bem: Man beachte, dass die **Konsistenz** von \bar{X}_n für μ nur eine andere Formulierung für das **schGGZ** ist.

Aus den Eigenschaften der **stochastischen Konvergenz** folgt auch die **Konsistenz** der **Stichprobenstreuung** S_n für die Streuung σ :

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

Asymptotische Normalverteilung

Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ ist **asymptotisch normalverteilt**, wenn er **in Verteilung** gegen eine **normalverteilte** sG konvergiert:

$$\lim_{n \rightarrow \infty} P \left(\frac{\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)}{\sqrt{\text{Var}(\hat{\theta}_n)}} \leq z \right) = \Phi(z) \quad \text{für alle } z \in \mathbb{R}$$

Das lässt sich auch wie folgt ausdrücken:

$$\hat{\theta}_n \stackrel{\text{asympt.}}{\sim} N(\mathbb{E}(\hat{\theta}_n), \text{Var}(\hat{\theta}_n))$$

Bsp: Der **Stichprobenmittelwert** \bar{X}_n ist **asymptotisch normalverteilt**:

$$\bar{X}_n \stackrel{\text{asympt.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\hat{=} \text{ZGVS!})$$

Eigenschaften von ML-Schätzern

ML-Schätzer haben – unter bestimmten **Bedingungen** – eine Reihe von „guten“, für **große Stichproben** sogar „optimale“, Eigenschaften :

- (1) invariant
- (2) asymptotisch erwartungstreu
- (3) asymptotisch effizient
- (4) konsistent
- (5) asymptotisch normalverteilt

Konfidenzintervalle

Wie **genau** sind Schätzer? Sog. **Intervallschätzer** bieten eine präzise Möglichkeit zur Beschreibung der (Un-) **Genauigkeit** von Schätzern.

Ist $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ eine Stichprobe der sG X , deren Verteilung von einem unbekannten Parameter $\theta \in \Theta$ abhängt, und sind $T_1(\mathbf{X}) < T_2(\mathbf{X})$ zwei Funktionen der Stichprobe, so nennt man das Zufallsintervall $(T_1(\mathbf{X}), T_2(\mathbf{X}))$ ein **Konfidenzintervall** (kurz **KI**) für θ mit **Konfidenz-koeffizient** $1 - \alpha$, wenn:

$$P_{\theta}(T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})) \geq 1 - \alpha \quad \text{für alle } \theta \in \Theta$$

Vielfach gilt die obige Aussage nur **approximativ**; dann spricht man von einem **approximativen Konfidenzintervall**.

Die – i. A. von θ abhängige – Wahrscheinlichkeit $P_{\theta}(T_1(\mathbf{X}) < \theta < T_2(\mathbf{X}))$ nennt man die **Überdeckungswahrscheinlichkeit** (kurz **ÜW**).

Konfidenzintervalle (Forts.)

Für ein **exaktes** $(1 - \alpha)$ -Konfidenzintervall beträgt die ÜW mindestens $1 - \alpha$ für alle $\theta \in \Theta$; für **approximative** KIe kann die tatsächliche ÜW für bestimmte θ auch kleiner als $1 - \alpha$ sein.

Es gibt mehrere **Methoden** zur Konstruktion von Konfidenzintervallen. Im Folgenden betrachten wir nur die sog. **Pivotmethode** etwas genauer.

Eine kurze Beschreibung einer auf **Simulation** basierenden Computer-methode (das **Bootstrapping**) findet man im Skriptum auf S. 284 bzw. im Buch auf S. 290.

Pivotgrößen

Eine **Pivotgröße** (kurz **Pivot**) ist eine sG $T = T(\mathbf{X}, \theta)$, die eine Funktion der Stichprobe $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ **und** des Parameters θ ist, deren **Verteilung** aber **bekannt** ist, und **nicht** von θ abhängt.

Bsp: X_1, X_2, \dots, X_n sei eine Stichprobe von $X \sim N(\mu, \sigma_0^2)$ (d. h., die Varianz sei bekannt gleich σ_0^2). Nach dem **Additionstheorem** gilt:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma_0^2}{n}\right) \implies T = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma_0} \sim N(0, 1)$$

D. h., T ist ein **Pivot** für μ . Damit gilt:

$$P_\mu(-z_{1-\alpha/2} < T < z_{1-\alpha/2}) = 1 - \alpha \quad \text{für alle } \mu \in \mathbb{R}$$

Pivotgrößen (Forts.)

Der Ausdruck in Klammern lässt sich wie folgt umformen:

$$\underbrace{\bar{X}_n - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}}_{T_1} < \mu < \underbrace{\bar{X}_n + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}}_{T_2}$$

Das Zufallsintervall (T_1, T_2) ist also ein **Konfidenzintervall** für μ mit **ÜW** $1 - \alpha$. Auf Grund der Symmetrie dieses Intervalls um \bar{X}_n schreibt man auch kürzer:

$$\bar{X}_n \pm z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

Pivotmethode

In Verallgemeinerung des vorhin betrachteten Beispiels lässt sich die **Pivotmethode** zur Konstruktion von KIn wie folgt beschreiben:

- (1) Formuliere ein statistisches **Modell** für die Stichprobe \mathbf{X} .
- (2) Bestimme einen geeigneten **Pivot** $T(\mathbf{X}, \theta)$.
- (3) Bestimme die **Verteilung** des Pivots.
- (4) Bestimme zwei **Quantile** q_1 und q_2 der **Pivotverteilung**, sodass:

$$P(q_1 < T(\mathbf{X}, \theta) < q_2) = 1 - \alpha$$

- (5) Bringe den Ausdruck in Klammern in die **äquivalente Form** $T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})$. (D. h., bringe θ „in die Mitte“.)
- (6) $(T_1(\mathbf{X}), T_2(\mathbf{X}))$ ist ein **$100(1 - \alpha)\%$ -Konfidenzintervall** für θ .

Bemerkungen zur Pivotmethode

- (a) Meist wählt man für q_1 das $(\alpha/2)$ - und für q_2 das $(1 - \alpha/2)$ -Quantil der Pivotverteilung. Kle dieser Art nennt man **Equal-Tails-Konfidenzintervalle**.
- (b) Je kleiner α umso breiter das KI; sehr breite Kle sind aber nur von geringer praktischer Relevanz. Übliche Werte für α sind 0.01, 0.05 oder 0.1.
- (c) Vielfach findet man keine exakten sondern nur **approximative Pivots** (etwa auf Basis des ZGVS). Dabei ist zu beachten, dass bei kleinen (oder mittleren) Stichprobengrößen die **tatsächliche** ÜW von mit approximativen Pivots konstruierten KIn u. U. erheblich vom **nominellen** $1 - \alpha$ abweichen kann.

Approximatives KI für den Mittelwert

X_1, X_2, \dots, X_n sei eine Stichprobe von einer sG X mit Mittelwert μ und Varianz $\sigma^2 < \infty$, wobei wir annehmen, dass beide Parameter **unbekannt** sind. Nach dem ZGVS gilt:

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) \implies \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx N(0, 1)$$

Ersetzt man das unbekannte σ durch S_n (konsistenter Schätzer für σ), erhält man den folgenden **approximativen Pivot** für μ :

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \approx N(0, 1)$$

Approximatives KI für den Mittelwert (Forts.)

Damit gilt für **große Stichproben**:

$$P\left(\bar{X}_n - z_{1-\alpha/2} \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + z_{1-\alpha/2} \frac{S_n}{\sqrt{n}}\right) \approx 1 - \alpha$$

Ein **approximatives** $(1 - \alpha)$ -**KI** für μ ist also gegeben durch:

$$\bar{X}_n \pm z_{1-\alpha/2} \frac{S_n}{\sqrt{n}}$$

(Vgl. für ein **Anwendungsbeispiel** – „Monte Carlo Integration“ – Skriptum S. 273, Buch S. 279.)

Satz von Student (Hauptsatz der Statistik)

Für eine Stichprobe X_1, X_2, \dots, X_n von $X \sim N(\mu, \sigma^2)$ gilt:

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

$$(2) \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1) \quad (\text{Pivot!})$$

(3) \bar{X}_n und S_n^2 sind (stochastisch) **unabhängig**.

$$(4) \quad \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1) \quad (\text{Pivot!})$$

Kle für die Parameter einer Normalverteilung

X_1, X_2, \dots, X_n sei eine Stichprobe aus $X \sim N(\mu, \sigma^2)$

Beide Parameter μ und σ^2 seien **unbekannt**

Aus dem **Satz von Student** (**Hauptsatz**) ergeben sich die folgenden **$100(1 - \alpha)\%$ -Kle**:

für μ : $\bar{X}_n \pm t_{n-1; 1-\alpha/2} \frac{S_n}{\sqrt{n}}$

für σ^2 : $\left(\frac{(n-1)S_n^2}{\chi_{n-1; 1-\alpha/2}^2}, \frac{(n-1)S_n^2}{\chi_{n-1; \alpha/2}^2} \right)$

für σ : $\left(\sqrt{\frac{(n-1)S_n^2}{\chi_{n-1; 1-\alpha/2}^2}}, \sqrt{\frac{(n-1)S_n^2}{\chi_{n-1; \alpha/2}^2}} \right)$

KI für die Differenz der Mittelwerte für zwei unabhängige Normalverteilungen

$$\left. \begin{array}{l} X_1, X_2, \dots, X_m \\ Y_1, Y_2, \dots, Y_n \end{array} \right\} \text{unabhängige Stichproben aus} \left\{ \begin{array}{l} X \sim N(\mu_X, \sigma_X^2) \\ Y \sim N(\mu_Y, \sigma_Y^2) \end{array} \right.$$

Alle Parameter $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ seien **unbekannt**.

Annahme: $\sigma_X^2 = \sigma_Y^2$

Gepoolter Varianzschätzer: $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$

$$100(1-\alpha)\% \text{ KI für } \mu_X - \mu_Y : \bar{X} - \bar{Y} \pm t_{m+n-2; 1-\alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$$

KI für den Parameter einer Bernoulli-Verteilung

X_1, X_2, \dots, X_n sei eine Stichprobe von $X \sim A(p)$ (Bernoulli)

$\hat{p} = \bar{X}_n$ ist der **ML-Schätzer** von p

Für **große** Stichproben findet man einen **approximativen Pivot** für p mit Hilfe des **ZGVS**:

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right) \implies \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1)$$

Ersetzt man im Wurzelausdruck p durch \hat{p} bekommt man ein (approximatives) **100(1 - α)%-KI für p** wie folgt:

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Das nennt man das **Standardintervall** (oder **Wald-Intervall**) für p .

KI für den Parameter einer Poisson–Verteilung

X_1, X_2, \dots, X_n sei eine Stichprobe von $X \sim P(\lambda)$ (Poisson)

$\hat{\lambda} = \bar{X}_n$ ist der **ML–Schätzer** von λ

Für **große** Stichproben findet man einen **approximativen Pivot** für λ mit Hilfe des **ZGVS**:

$$\hat{\lambda} \approx N\left(\lambda, \frac{\lambda}{n}\right) \implies \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \approx N(0, 1)$$

Ersetzt man im Wurzelausdruck λ durch $\hat{\lambda}$ bekommt man ein (approximatives) **100(1 – α)%–KI für λ** wie folgt:

$$\hat{\lambda} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{\lambda}}{n}}$$

Das nennt man das **Standardintervall** (oder **Wald–Intervall**) für λ .

Statistische Tests

Neben **Punktschätzungen** und **Konfidenzintervallen** betrachtet man in der schließenden Statistik auch das **Testen** von **statistischen Hypothesen**.

Man unterscheidet zwei Arten von Hypothesen:

- ▶ **Parameterhypothesen**: Behauptung über den (oder die) Parameter von einer (oder mehreren) Verteilung(en). Bei dieser Art von Hypothesen nimmt man an, dass der **Verteilungstyp** bekannt ist (z. B., Normalverteilung).
- ▶ **Verteilungshypothesen**: In vielen Fällen ist der Verteilungstyp nicht bekannt und man möchte testen, ob eine bestimmte **Verteilung** oder **Verteilungsfamilie** (z. B., Normalverteilungen) ein mögliches Modell für die vorliegenden Beobachtungen darstellt.

Parametertests: Null– / Gegenhypothese

Wir beobachten eine sG X mit Dichte $f(x; \theta)$ (oder W–Funktion $p(x; \theta)$), wobei der **Parameter** $\theta \in \Theta$ **unbekannt** ist.

Null–/Gegenhypothese: Auf Grund einer Theorie (einer Vermutung, ...) gelte $\theta \in \Theta_0$ oder $\theta \in \Theta_1$, wobei $\Theta_0 \cap \Theta_1 = \emptyset$ und $\Theta_0 \cup \Theta_1 \subseteq \Theta$. Die erste Behauptung nennt man die **Nullhypothese**, die zweite die **Gegen–** oder **Alternativhypothese** und schreibt das **Testproblem** wie folgt:

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta \in \Theta_1$$

Parametertests: Null– / Gegenhypothese (Forts.)

Achtung: Es ist nicht gleichgültig, welche Behauptung die Null– und welche die Gegenhypothese ist!

Nullhypothese: In der Regel diejenige Behauptung, die die bisherige Situation, den „Normalfall“ (oder „Status quo“) repräsentiert.

Gegenhypothese: Meist einfach das Komplement zur Nullhypothese, oder diejenige Behauptung, deren Zutreffen ein bestimmtes Handeln erfordert oder die gravierenderen Konsequenzen (positive / negative) nach sich zieht.

Bsp: Ein bekanntes Medikament habe eine Heilungschance von beispielsweise $p_0 = 0.67$. Nun kommt ein neues Medikament auf den Markt und man möchte prüfen, ob dessen Heilungschance größer ist. In diesem Fall testet man:

$$\mathcal{H}_0 : p = p_0 \quad \text{gegen} \quad \mathcal{H}_1 : p > p_0$$

Parametertests: Ein- / Zweiseitige Gegenhypothesen

Wir betrachten nur **einfache** Nullhypothesen der Form $\mathcal{H}_0 : \theta = \theta_0$.

Lautet die Gegenhypothese $\theta \neq \theta_0$ nennt man sie **zweiseitig**:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta \neq \theta_0$$

In den beiden folgenden Fällen ist die Gegenhypothese **einseitig**:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta < \theta_0 \quad \text{oder} \quad \mathcal{H}_1 : \theta > \theta_0$$

Bem: Diese Unterscheidung ist wichtig für die Wahl des korrekten **Schwellenwerts** (vgl. später).

Testentscheidung

Eine auf einer **Stichprobe** $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ von X basierende **Entscheidungsregel** über die beiden Hypothesen nennt man einen (statistischen) **Test**.

Ein Test wird durch seinen **kritischen Bereich** C charakterisiert, d. h. durch eine Teilmenge des **Stichprobenraumes** M_X^n (= Menge aller möglichen Stichproben von X) mit:

Verwerfe \mathcal{H}_0 falls $\mathbf{X} \in C$

Akzeptiere \mathcal{H}_0 falls $\mathbf{X} \notin C$

Typ I / Typ II–Fehler

Beim Testen von Hypothesen gibt es zwei Fehlermöglichkeiten.

Typ I–Fehler (Fehler 1. Art): Tritt auf, wenn die \mathcal{H}_0 verworfen wird, obwohl sie richtig ist.

Typ II–Fehler (Fehler 2. Art): Tritt auf, wenn die \mathcal{H}_0 nicht verworfen wird, obwohl sie falsch ist.

Die folgende Tabelle zeigt die möglichen (Fehl–) Entscheidungen:

Entscheidung	Wahrer Zustand	
	\mathcal{H}_0 trifft zu	\mathcal{H}_1 trifft zu
Verwerfe \mathcal{H}_0	Typ I–Fehler	Korrekte Entscheidung
Akzeptiere \mathcal{H}_0	Korrekte Entscheidung	Typ II–Fehler

Typ I / Typ II-Fehler (Forts.)

Die **Wahrscheinlichkeit eines Typ I-Fehlers** bezeichnet man mit α :

$$\alpha = P(\text{Typ I-Fehler}) = P_{\theta_0}(\mathbf{X} \in C)$$

Die **Wahrscheinlichkeit eines Typ II-Fehlers** bezeichnet man mit β :

$$\beta(\theta) = P(\text{Typ II-Fehler}) = P_{\theta}(\mathbf{X} \notin C), \quad \theta \in \Theta_1(\mathcal{H}_1)$$

Die Wahrscheinlichkeit eines Typ II-Fehlers ist i. A. keine Konstante, sondern hängt vom wahren Wert des Parameters ab.

Bem: Die Wahrscheinlichkeit α eines Typ I-Fehlers nennt man auch das **(Signifikanz-) Niveau** des Tests.

Starke / Schwache Schlussfolgerungen

In der Praxis verwendet man Tests, die eine vorgegebene (kleine) W. für einen **Typ I–Fehler** nicht überschreiten.

Wird \mathcal{H}_0 **verworfen**, spricht man daher von einer **starken Schlussfolgerung**.

Wird aber \mathcal{H}_0 **nicht verworfen**, hat man möglicherweise einen Typ II–Fehler begangen, und über seine Größe (d. h., Wahrscheinlichkeit) ist meist nur wenig bekannt. In diesem Fall spricht man daher von einer **schwachen Schlussfolgerung** und sagt meist vorsichtiger, dass man \mathcal{H}_0 **nicht verwerfen kann** (und nicht, dass man \mathcal{H}_0 „akzeptiert“).

M. a. W.: Das eigentliche **Ziel** des (Parameter-) Testens ist die **Verwerfung** der Nullhypothese \mathcal{H}_0 .

Powerfunktion

Die **Power** (oder **Schärfe**) eines Tests ist die W. der Verwerfung der Nullhypothese, wenn die Gegenhypothese zutrifft (d. h., die richtige Entscheidung zu treffen, wenn \mathcal{H}_0 falsch ist).

Betrachtet man die Power als Funktion von θ , spricht man von der **Powerfunktion** (oder **Schärfefunktion**):

$$\gamma_C(\theta) = 1 - \beta(\theta) \quad \text{für} \quad \theta \in \Theta_1$$

Was bedeutet es, wenn ein Test „signifikant“ ist?

Achtung: Die Sprechweise von der „Signifikanz“ eines Tests ist weit verbreitet aber mit Vorsicht zu gebrauchen!

Man sagt, dass ein Test „signifikant“ ist, wenn er die Nullhypothese verwirft.

Das ist eine formale Aussage, die von den Hypothesen (Was ist \mathcal{H}_0 , was \mathcal{H}_1 ?), vom Test, von der Stichprobengröße und von α abhängt.

Überdies sollte die statistische Signifikanz nicht mit der praktischen (oder wissenschaftlichen) Signifikanz verwechselt werden!

Möglicherweise ist ein formal signifikantes Ergebnis nur von geringer praktischer Bedeutung.

p-Wert

Verwendet man zum Testen entsprechende Statistiksoftware bekommt man i. A. keine **Testentscheidung**, sondern es wird statt dessen ein **Wahrscheinlichkeitswert** berechnet.

Beim Testen von \mathcal{H}_0 gegen \mathcal{H}_1 entspricht der **p-Wert** (engl. **p-value**) der Wahrscheinlichkeit – bei **Zutreffen von \mathcal{H}_0** – den **beobachteten** Wert der **Teststatistik** oder einen **extremere**n zu bekommen.

Was unter „extremer“ zu verstehen ist, hängt von der Gegenhypothese (oder vom kritischen Bereich) ab.

Bezug zum klassischen Testen: Ein klassischer Test ergibt sich dadurch, dass eine Nullhypothese, deren p-Wert kleiner als α ist, auf dem Niveau α verworfen wird.

Anders ausgedrückt: Der p-Wert ist der **größte** Wert von α , für den die \mathcal{H}_0 **nicht** verworfen wird.

p-Wert: Beispiel

Ein Hersteller von Computerchips behauptet, dass nicht mehr als 2% seiner Chips defekt sind. Ein Abnehmer testet dazu die folgenden Hypothesen:

$$\mathcal{H}_0 : p = 0.02 \quad \text{gegen} \quad \mathcal{H}_1 : p > 0.02$$

Angenommen eine Prüfung von 300 Chips ergibt, dass 10 defekt sind. Ist Y die Zahl der defekten Chips (von 300), berechnet sich der p -Wert bei Zutreffen von \mathcal{H}_0 wie folgt:

$$p\text{-Wert} = \sum_{i=10}^{300} \binom{300}{i} (0.02)^i (0.98)^{300-i} = 0.0818$$

Testet man zum (üblichen) Niveau von 5%, wird \mathcal{H}_0 nicht verworfen, wohl aber zum Niveau von 10%.

Interpretation des p -Werts

Meist hält man sich an das folgende **Beurteilungsschema**:

p -Wert	Signifikanz	
< 0.01	sehr hoch	(sehr starke Einwände gegen \mathcal{H}_0)
$0.01 - 0.05$	hoch	(starke Einwände gegen \mathcal{H}_0)
$0.05 - 0.10$	schwach	(schwache Einwände gegen \mathcal{H}_0)
> 0.10	keine	(sehr schwache/keine Einwände gegen \mathcal{H}_0)

Achtung: Man verwechsle den p -Wert einer Nullhypothese nicht mit $P(\mathcal{H}_0|\text{Daten})$! Derartige Aussagen sind nur im Rahmen der **Bayes'schen Statistik** (vgl. **Kapitel 8**) möglich und sinnvoll. Der p -Wert ist **nicht** die Wahrscheinlichkeit für das Zutreffen der \mathcal{H}_0 !

Beziehung zwischen Tests und Konfidenzintervallen

Es gibt enge Beziehungen zwischen **Tests** und **Konfidenzintervallen**.

Angenommen, $(T_1(\mathbf{x}), T_2(\mathbf{x}))$ ist ein **$100(1 - \alpha)\%$ -KI** für den Parameter $\theta \in \Theta$ auf Basis einer (konkreten) Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ von X .

Dann ist ein **Test zum Niveau α** für die Hypothesen:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta \neq \theta_0$$

gegeben durch:

$$\theta_0 \in (T_1(\mathbf{x}), T_2(\mathbf{x})) \quad \longrightarrow \quad \mathcal{H}_0 \text{ nicht verwerfen}$$

$$\theta_0 \notin (T_1(\mathbf{x}), T_2(\mathbf{x})) \quad \longrightarrow \quad \mathcal{H}_0 \text{ verwerfen}$$

Tests für den Mittelwert einer $N(\mu, \sigma^2)$ -Verteilung

Gegeben sei eine Stichprobe X_1, X_2, \dots, X_n von $X \sim N(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ **unbekannt** seien.

μ lässt sich durch den **Stichprobenmittelwert** \bar{X}_n **erwartungstreu** und **konsistent** schätzen.

σ^2 lässt sich durch die **Stichprobenvarianz** S_n^2 **erwartungstreu** und **konsistent** schätzen.

Nach Punkt (4) aus dem **Satz von Student (Hauptsatz)** gilt:

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

Damit lassen sich **exakte t-Tests** zum Niveau α für μ konstruieren.

Tests für den Mittelwert einer $N(\mu, \sigma^2)$ -Verteilung (Forts.)

Nullhypothese: $\mathcal{H}_0 : \mu = \mu_0$

Teststatistik: $T_0 = \frac{\bar{X}_n - \mu_0}{S_n / \sqrt{n}}$

Gegenhypothese \mathcal{H}_1

\mathcal{H}_0 verwerfen, falls

$$\mu \neq \mu_0$$

$$|T_0| > t_{n-1; 1-\alpha/2}$$

$$\mu > \mu_0$$

$$T_0 > t_{n-1; 1-\alpha}$$

$$\mu < \mu_0$$

$$T_0 < t_{n-1; \alpha} (= -t_{n-1; 1-\alpha})$$

Tests für die Varianz einer $N(\mu, \sigma^2)$ -Verteilung

Gegeben sei eine Stichprobe X_1, X_2, \dots, X_n von $X \sim N(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ **unbekannt** seien.

Nach Punkt (2) aus dem **Satz von Student (Hauptsatz)** gilt:

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

Damit lassen sich **exakte Tests** zum Niveau α für σ^2 konstruieren.

Tests für die Varianz einer $N(\mu, \sigma^2)$ -Verteilung (Forts.)

Nullhypothese: $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$

Teststatistik: $\chi_0^2 = \frac{(n-1)S_n^2}{\sigma_0^2}$

Gegenhypothese \mathcal{H}_1

\mathcal{H}_0 verwerfen, falls

$$\sigma^2 \neq \sigma_0^2$$

$$\chi_0^2 < \chi_{n-1; \alpha/2}^2 \quad \text{oder} \quad \chi_0^2 > \chi_{n-1; 1-\alpha/2}^2$$

$$\sigma^2 > \sigma_0^2$$

$$\chi_0^2 > \chi_{n-1; 1-\alpha}^2$$

$$\sigma^2 < \sigma_0^2$$

$$\chi_0^2 < \chi_{n-1; \alpha}^2$$

Approximative Tests für einen Anteil

Gegeben sei eine Stichprobe X_1, X_2, \dots, X_n von $X \sim A(p)$ (**Bernoulli**).

Für **großes** n gilt: $Y = \sum_{i=1}^n X_i \approx N(np, np(1-p))$ (**ZGVS**)

Nullhypothese: $\mathcal{H}_0 : p = p_0$

Teststatistik: $Z_0 = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}}$

Gegenhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen , falls
$p \neq p_0$	$ Z_0 > z_{1-\alpha/2}$
$p > p_0$	$Z_0 > z_{1-\alpha}$
$p < p_0$	$Z_0 < z_\alpha (= -z_{1-\alpha})$

Tests für die Mittelwerte zweier ua. Normalverteilungen

Gegeben seien zwei Stichproben X_1, X_2, \dots, X_m und Y_1, Y_2, \dots, Y_n von $X \sim N(\mu_X, \sigma_X^2)$ bzw. $Y \sim N(\mu_Y, \sigma_Y^2)$, wobei **X und Y ua.** sind.

Annahme: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (unbekannt)

Gepoolter Varianzschätzer: $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$

Mit Punkt (4) aus dem **Satz von Student (Hauptsatz)** gilt:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{1/m + 1/n}} \sim t(m+n-2)$$

Damit lassen sich **exakte (gepoolte) t-Tests** zum Niveau α für die Differenz $\mu_X - \mu_Y$ konstruieren.

Tests für die Mittelwerte zweier ua. Normalverteilungen (Forts.)

Nullhypothese: $\mathcal{H}_0 : \mu_X - \mu_Y = \Delta_0$ (Bem: Meist gilt $\Delta_0 = 0$.)

Teststatistik:
$$T_0 = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{S_p \sqrt{1/m + 1/n}}$$

Gegenhypothese \mathcal{H}_1

\mathcal{H}_0 verwerfen, falls

$$\mu_X - \mu_Y \neq \Delta_0$$

$$|T_0| > t_{m+n-2; 1-\alpha/2}$$

$$\mu_X - \mu_Y > \Delta_0$$

$$T_0 > t_{m+n-2; 1-\alpha}$$

$$\mu_X - \mu_Y < \Delta_0$$

$$T_0 < t_{m+n-2; \alpha} (= -t_{m+n-2; 1-\alpha})$$

Tests für die Varianzen zweier ua. Normalverteilungen

Für Stichproben X_1, X_2, \dots, X_m und Y_1, Y_2, \dots, Y_n von $X \sim N(\mu_X, \sigma_X^2)$ bzw. $Y \sim N(\mu_Y, \sigma_Y^2)$, wobei **X und Y ua.** sind, gilt:

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

Damit lassen sich **exakte F-Tests** zum Niveau α für den Quotienten σ_X^2/σ_Y^2 konstruieren.

Bem: Man beachte, dass für die Quantile der F-Verteilung die folgende Beziehung gilt:

$$F_{m-1, n-1; p} = \frac{1}{F_{n-1, m-1; 1-p}}$$

Tests für die Varianzen zweier ua. Normalverteilungen (Forts.)

Nullhypothese: $\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2$

Teststatistik: $F_0 = \frac{S_X^2}{S_Y^2}$

Gegenhypothese \mathcal{H}_1 \mathcal{H}_0 **verwerfen**, falls

$$\sigma_X^2 \neq \sigma_Y^2 \qquad F_0 < F_{m-1, n-1; \alpha/2} \quad \text{oder} \quad F_0 > F_{m-1, n-1; 1-\alpha/2}$$

$$\sigma_X^2 > \sigma_Y^2 \qquad F_0 > F_{m-1, n-1; 1-\alpha}$$

$$\sigma_X^2 < \sigma_Y^2 \qquad F_0 < F_{m-1, n-1; \alpha}$$

Normal-QQ-Plot

Der **Normal-QQ-Plot** ist ein **grafischer Test** zur Überprüfung der (Null-) Hypothese, dass eine Stichprobe x_1, x_2, \dots, x_n aus einer **$N(\mu, \sigma^2)$ -Verteilung** stammt.

Der Plot wird nach den folgenden Schritten erstellt:

- (1) Daten der Größe nach ordnen: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$
- (2) Der kumulierte Anteil der Daten links von $x_{(i)}$ ist gegeben durch:

$$p_i = \frac{i - 0.5}{n}, \quad i = 1, 2, \dots, n$$

- (3) Bestimme die p_i -Quantile der Standardnormalverteilung:

$$z_i = \Phi^{-1}(p_i), \quad i = 1, 2, \dots, n$$

Normal-QQ-Plot (Forts.)

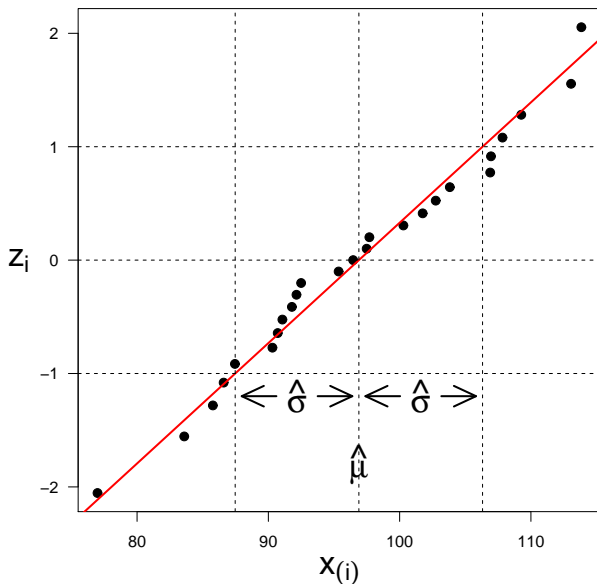
- (4) Zeichne die folgenden Punkte in ein übliches Koordinatensystem:

$$(x_{(i)}, z_i), \quad i = 1, 2, \dots, n$$

- (5) Lege eine („robuste“) Vergleichsgerade durch die Punkte.
- (6) Liegen die Punkte annähernd auf der Vergleichsgeraden, lässt sich die Hypothese „Daten aus Normalverteilung“ nicht verwerfen.
- (7) Entnehme im positiven Fall dem Plot (grobe) Schätzwerte für den Mittelwert μ und für die Streuung σ .

Begründung: Skriptum S. 310, Buch S. 316

Normal-QQ-Plot: Beispiel



Chiquadrat-Anpassungstests

Der **Chiquadrat GOF** (*goodness-of-fit*) **Test** ist ein allgemein anwendbares Testverfahren zur Überprüfung von **Verteilungshypothesen**, z. B:

$$\mathcal{H}_0 : X \sim N(\mu, \sigma^2) \quad \text{gegen} \quad \mathcal{H}_1 : X \not\sim N(\mu, \sigma^2)$$

Die Tests beruhen auf einer **Klasseneinteilung** des **Merkmalraums** (des statistischen Experiments) und da es sich um **approximative** Tests handelt, sollte die Stichprobe nicht zu klein sein (dafür gibt es „Faustregeln“; s. unten).

Einfache/Zusammengesetzte Tests: Ist die \mathcal{H}_0 **vollständig** spezifiziert (d. h., sind alle Parameter bekannt), nimmt man den **einfachen**, sonst den **zusammengesetzten** Chiquadrat GOF-Test.

Einfacher Chiquadrat-Anpassungstest

Der **Merkmalraum** M eines statistischen Experiments zerfalle in k paarweise disjunkte Teilmengen A_1, \dots, A_k , wobei $p_i = P(A_i)$ für $i = 1, \dots, k$. Das Experiment werde n Mal (unabhängig) wiederholt und X_i sei die **Anzahl der Versuchsausgänge** in A_i . Dann ist die **Teststatistik** für einen Test von:

$$\mathcal{H}_0 : p_i = p_{i0}, \quad i = 1, \dots, k \quad \text{gegen} \quad \mathcal{H}_1 : \exists i \text{ mit } p_i \neq p_{i0}$$

gegeben durch:

$$Q_{k-1} = \sum_{i=1}^k \frac{(X_i - np_{i0})^2}{np_{i0}}$$

Einfacher Chiquadrat-Anpassungstest (Forts.)

\mathcal{H}_0 wird zum **approximativen** Niveau α verworfen, falls:

$$Q_{k-1} > \chi_{k-1; 1-\alpha}^2$$

Faustregel: Die übliche (vergleichsweise „strenge“) Faustregel besagt, dass die **χ^2 -Approximation** von Q_{k-1} (unter \mathcal{H}_0) dann ausreichend gut ist, wenn man unter \mathcal{H}_0 in jeder Klasse zumindest 5 Beobachtungen **erwarten** kann, d. h., wenn

$$np_{i0} \geq 5 \quad \text{für} \quad i = 1, \dots, k$$

Sonst muss man benachbarte Klassen zusammenfassen.

Bsp im Skriptum S. 314, Buch S. 320 bzw. in der Übung.

Zusammengesetzter Chiquadrat-Anpassungstest

Sind die Klassenwahrscheinlichkeiten $p_i = P(A_i)$ durch die \mathcal{H}_0 **nicht vollständig** spezifiziert, d. h. gilt $p_i = p_i(\theta)$ für einen (nicht bekannten) s -dimensionalen Parameter(vektor) $\theta \in \Theta$, muss der Test modifiziert werden. Die **Teststatistik** eines Tests von:

$$\mathcal{H}_0 : p_i = p_i(\theta), \quad i = 1, 2, \dots, k \quad \text{gegen} \quad \mathcal{H}_1 : \exists i \text{ mit } p_i \neq p_i(\theta)$$

ist gegeben durch:

$$Q_{k-s-1} = \sum_{i=1}^k \frac{[X_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}$$

Dabei ist $\hat{\theta}$ der **ML-Schätzwert** von θ (auf Basis der *unklassierten* Daten).

Zusammengesetzter Chiquadrat-Anpassungstest (Forts.)

\mathcal{H}_0 wird zum **approximativen** Niveau α verworfen, falls:

$$Q_{k-s-1} > \chi^2_{k-s-1; 1-\alpha}$$

Faustregel: Ähnlich wie beim einfachen Test, besagt die üblicherweise verwendete Faustregel, dass die **χ^2 -Approximation** von Q_{k-s-1} (unter \mathcal{H}_0) dann ausreichend gut ist, wenn:

$$np_i(\hat{\theta}) \geq 5 \quad \text{für } i = 1, \dots, k$$

Sonst muss man benachbarte Klassen zusammenfassen.

Bsp im Skriptum S. 316, Buch S. 322 bzw. in der Übung.