

Statistik

Andrej Depperschmidt

Sommersemester 2016

Schätzen der Varianz mit Stichprobenmittel

Sei $X = (X_1, \dots, X_n)$ eine Stichprobe u.i.v. ZV mit

$$E[X_i] = \mu \in \mathbb{R}, \text{Var}[X_i] = \sigma^2 \in (0, \infty) \text{ und } \mu_4 = E[(X_i - \mu)^4] < \infty.$$

Wenn μ bekannt ist, so können wir σ^2 wie folgt schätzen:

- ▶ $Y_i = (X_i - \mu)^2$. Dann sind Y_1, \dots, Y_n u.i.v. mit $E[Y_i] = \text{Var}[X_i] = \sigma^2$ und $\text{Var}[Y_i] < \infty$
- ▶ Dann ist

$$\hat{\sigma}_1^2(X) := \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

ein Schätzer von σ^2 . Als Stichprobenmittel ist er erwartungstreu, konsistent etc.

Schätzen der Varianz mit Stichprobenvarianz

Wenn μ bekannt ist so können wir dennoch die Stichprobenvarianz

$$\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (1)$$

als Schätzer für σ^2 verwenden. Auch dieser Schätzer ist erwartungstreu, denn es ist

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n ((X_i - \bar{X}) + (\bar{X} - \mu))^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \underbrace{\sum_{i=1}^n (X_i - \bar{X})}_{=0, \text{ da } \sum_{i=1}^n X_i = n\bar{X}}. \end{aligned}$$

Es folgt

$$\begin{aligned} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^n (X_i - \mu)^2\right] - E\left[\sum_{i=1}^n (\bar{X} - \mu)^2\right] \\ &= n \operatorname{Var}[X_1] - n \operatorname{Var}[\bar{X}] = n\sigma^2 - \sigma^2 = (n-1)\sigma^2 \end{aligned}$$

und somit

$$E[\hat{\sigma}^2] = \sigma^2.$$

Man kann weiter nachrechnen, dass

$$\operatorname{Var}[\hat{\sigma}^2] = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

χ_n^2 -Verteilung

Seien die Zufallsvariablen Z_1, \dots, Z_n unabhängig und identisch verteilt mit $Z_i \sim \mathcal{N}_{0,1}$.

- ▶ Die Zufallsvariable $\sum_{i=1}^n Z_i^2$ ist χ_n^2 verteilt.
- ▶ Die Verteilung heißt (zentrale) χ^2 -Verteilung mit n Freiheitsgraden.
- ▶ Es gilt $\chi_n^2 = \Gamma_{1/2, n/2}$. Die Dichte ist gegeben durch

$$f_n(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad x > 0.$$

t_n -Verteilung

Seien $U \sim \mathcal{N}_{0,1}$ und $V \sim \chi_n^2$ stochastisch unabhängig.

- ▶ Die Zufallsvariable $\frac{U}{\sqrt{V/n}}$ ist t_n verteilt. Die Verteilung heißt **studentsche t -Verteilung (oder einfach t -Verteilung) mit n Freiheitsgraden**. Die Dichte ist gegeben durch

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}. \quad (2)$$

- ▶ **Bemerkung:** Man kann zeigen, $f_n(x) \xrightarrow{n \rightarrow \infty} \phi(x)$, $x \in \mathbb{R}$, wobei ϕ die Dichte der Standardnormalverteilung ist.

$F_{n,m}$ -Verteilung

Seien $X_1 \sim \chi_n^2$ und $X_2 \sim \chi_m^2$ unabhängig.

- Die Zufallsvariable $\frac{X_1/n}{X_2/m}$ ist $F_{n,m}$ verteilt. Die Verteilung heißt **F-Verteilung mit n Freiheitsgraden im Zähler und m Freiheitsgraden im Nenner**. Die Dichte bekommt man aus den entsprechenden Dichten der χ^2 Verteilung. Sie ist gegeben durch

$$f_{n,m}(x) = \frac{n^{n/2} m^{m/2} \Gamma((n+m)/2) x^{n/2-1}}{\Gamma(n/2) \Gamma(m/2) (m+nx)^{(n+m)/2}}, \quad x > 0. \quad (3)$$

Verteilung von $\hat{\mu}$ und $\hat{\sigma}^2$ bei Normalität

Satz

Seien X_1, \dots, X_n u.i.v. mit $X_i \sim \mathcal{N}_{\mu, \sigma^2}$, $0 < \sigma^2 < \infty$.

(i) $\hat{\mu}(X) = \frac{1}{n} \sum_{i=1}^n X_i$ und $\hat{\sigma}^2(X)$ sind unabhängig.

(ii) $\hat{\mu}(X) \sim \mathcal{N}_{\mu, \sigma^2/n}$, $\frac{n-1}{\sigma^2} \hat{\sigma}^2(X) \sim \chi_{n-1}^2$.

Folgerung

Für X_1, \dots, X_n wie oben gilt $\frac{\sqrt{n}(\hat{\mu}(X) - \mu)}{\hat{\sigma}} \sim t_{n-1}$.

Beweis.

$$\frac{\sqrt{n}(\hat{\mu}(X) - \mu)}{\sigma} \frac{\sigma}{\hat{\sigma}} = \underbrace{\frac{\sqrt{n}(\hat{\mu}(X) - \mu)}{\sigma}}_{\sim \mathcal{N}_{0,1}} \frac{1}{\underbrace{\left((n-1) \frac{\hat{\sigma}^2}{\sigma^2} \frac{1}{n-1} \right)^{1/2}}_{\sim \chi_{n-1}^2}}$$



Konfidenzgrenzen für μ bei bekannter Varianz Normalverteilungsmodell

Für X_1, \dots, X_n unabhängig mit $X_i \sim \mathcal{N}_{\mu, \sigma^2}$ mit **bekanntem** $\sigma^2 \in (0, \infty)$ haben wir in gezeigt, dass das Konfidenzintervall für μ zur Sicherheit $1 - \alpha$ durch

$$I(X) = \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

gegeben ist, wobei $z_{\alpha/2}$ das obere $\alpha/2$ -Quantil (= $1 - \alpha/2$ -Quantil) der Standardnormalverteilung ist, d.h.

$$z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2) \quad \text{bzw.} \quad P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}, \quad Z \sim \mathcal{N}_{0,1}.$$

Konfidenzgrenzen für μ bei geschätzter Varianz Normalverteilungsmodell

Ist $\sigma^2 \in (0, \infty)$ unbekannt und wird es mit $\hat{\sigma}^2(X)$ geschätzt dann kann man sich analog überlegen, dass

$$\tilde{I}(X) = \left(\bar{X} - t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \right)$$

ein Konfidenzintervall für μ zur Sicherheit $1 - \alpha$ ist.

Dabei bezeichnet $t_{n-1, \alpha/2}$ das obere $\alpha/2$ Quantil der t_{n-1} -Verteilung.

Konfidenzgrenzen für μ bei geschätzter Varianz allgemeine Verteilung

Sind X_1, \dots, X_n u.i.v. mit $E[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2 \in (0, \infty)$ dann hat $\tilde{I}(X)$ die asymptotische Sicherheit $1 - \alpha$, denn es gilt

$$\bar{X} - t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \iff \frac{\sqrt{n} |\bar{X}_n - \mu|}{\hat{\sigma}_n} < t_{n-1, \alpha/2}$$

und somit

$$\begin{aligned} P(\mu \in \tilde{I}(X)) &= P\left(\frac{\sqrt{n} |\bar{X}_n - \mu|}{\hat{\sigma}_n} \leq t_{n-1, \alpha/2}\right) \\ &= P\left(\underbrace{\frac{\sqrt{n} |\bar{X}_n - \mu|}{\sigma}}_{\Rightarrow \mathcal{N}_{0,1}} \underbrace{\frac{\sigma}{\hat{\sigma}_n}}_{\xrightarrow{P} 1} \underbrace{\frac{z_{\alpha/2}}{t_{n-1, \alpha/2}}}_{\rightarrow 1} \leq z_{\alpha/2}\right) \\ &\xrightarrow{n \rightarrow \infty} P(|Z| \leq z_{\alpha/2}) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha. \end{aligned}$$

Konfidenzgrenzen für μ bei geschätzter Varianz allgemeine Verteilung

$$\begin{aligned} P(\mu \in \tilde{I}(X)) &= P\left(\underbrace{\frac{\sqrt{n}|\bar{X}_n - \mu|}{\sigma}}_{\Rightarrow \mathcal{N}_{0,1}} \underbrace{\frac{\sigma}{\hat{\sigma}_n}}_{\xrightarrow{P} 1} \underbrace{\frac{z_{\alpha/2}}{t_{n-1,\alpha/2}}}_{\rightarrow 1} \leq z_{\alpha/2}\right) \\ &\xrightarrow{n \rightarrow \infty} P(|Z| \leq z_{\alpha/2}) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha. \end{aligned}$$

Bemerkung:

- ▶ Statt $t_{n-1,\alpha/2}$ kann man $z_{\alpha/2}$ (auch im allgemeinen Modell) verwenden. Liefert dieselbe asymptotische Sicherheit.
- ▶ Praxisüblich ist es mit t -Quantilen zu arbeiten. Sie sind exakt im Normalverteilungsmodell und liefern konservativere (größere) Konfidenzintervalle, da $z_{\alpha/2} < t_{n,\alpha/2} \forall n \in \mathbb{N}, \alpha < 1/2$.

Das Maximum-Likelihood Prinzip

Betrachte eine einzige Beobachtung einer ZV mit Werten in $\{0, 1, 2\}$ mit Verteilung P_θ , $\theta \in \{\theta_0, \theta_1\}$, gegeben wie folgt:

	$x = 0$	$x = 1$	$x = 2$
$\theta = \theta_0$	0.7	0.2	0.1
$\theta = \theta_1$	0.2	0.3	0.5

- ▶ $X = 0$ beobachtet $\Rightarrow \theta = \theta_0$ plausibler.
- ▶ $X = 1$ oder $X = 2$ beobachtet $\Rightarrow \theta = \theta_1$ plausibler.

Folgende Schätzfunktion bietet sich also an

$$\hat{\theta}(X) = \begin{cases} \theta_0 & : X = 0, \\ \theta_1 & : X \neq 0. \end{cases}$$

Das Maximum-Likelihood Prinzip

Beispiel verallgemeinerbar auf den Fall diskreter Verteilungen P_θ mit $\theta \in \Theta \subset \mathbb{R}^k$.

Wenn $X = x$ beobachtet wird, dann ist θ_1 plausibler als θ_2 genau dann wenn $P_{\theta_1}(\{x\}) > P_{\theta_2}(\{x\})$.

Als Schätzer $\hat{\theta}$ würde man also das $\theta \in \Theta$ wählen, das $\theta \mapsto P_\theta(\{x\})$ maximiert, wenn es existiert.

Likelihood Funktion, Maximum-Likelihood Schätzer

$X = (X_1, \dots, X_n) \in \mathcal{X}$, $X \sim P \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

f_θ sind Dichten bzw. Wahrscheinlichkeitsgewichte von P_θ .

Definition

(i) Für jedes $x \in \mathcal{X}$ heißt die Funktion

$$L(\cdot, x) \mapsto L(\theta, x) = f_\theta(x) \in \mathbb{R}$$

Likelihood-Funktion bei Beobachtung $x \in \mathcal{X}$.

(ii) Wir nennen $\hat{\theta} \in \Theta$ die **Maximum-Likelihood Schätzung (ML-Schätzung)** für θ falls gilt

$$L(\hat{\theta}, x) = \sup_{\theta \in \Theta} L(\theta, x).$$

Fasst man $\hat{\theta}$ als Funktion von X auf, dann nennt man $\hat{\theta} = \hat{\theta}(X)$ **ML-Schätzfunktion** bzw. **ML-Schätzer** für θ .

Konkrete Bestimmung einer ML-Schätzung

- ▶ Da Logarithmus streng monoton wachsend ist, kann man anstelle von $L(\cdot, x)$ äquivalent auch $\log L(\cdot, x)$, also **Log-Likelihood-Funktion** maximieren.
Besonders hilfreich, wenn $L(\cdot, x)$ Produktform hat. Dies gilt wenn X_1, \dots, X_n unabhängig sind.
- ▶ Ist $L(\theta, x)$ bzw. $\log L(\theta, x)$ differenzierbar in θ auf $\Theta_0 \subset \Theta$, Θ_0 offen, so kann man oft (aber nicht immer) die ML-Schätzung durch explizites Lösen der **Likelihood-Gleichung**

$$\frac{\partial}{\partial \theta} L(\theta, x) = 0$$

bzw. der **Log-Likelihood-Gleichung**

$$\frac{\partial}{\partial \theta} \log L(\theta, x) = 0$$

bestimmen.

ML-Schätzung für Bernoulli Verteilung

Seien X_1, \dots, X_n u.i.v. mit $X_i \sim \text{Ber}_\theta$, $\theta \in (0, 1)$. Es gilt

$$L(\theta, x) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i},$$

$$\log L(\theta, x) = \sum_{i=1}^n x_i \log \theta + (n - \sum_{i=1}^n x_i) \log(1 - \theta).$$

Die Log-Likelihood-Gleichung ist

$$0 = \frac{\partial}{\partial \theta} \log L(\theta, x) \Big|_{\theta=\hat{\theta}} = \frac{1}{\hat{\theta}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\theta}} (n - \sum_{i=1}^n x_i).$$

\Leftrightarrow

$$\sum_{i=1}^n x_i - \hat{\theta} \sum_{i=1}^n x_i = \hat{\theta} n - \hat{\theta} \sum_{i=1}^n x_i$$

und wir erhalten

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

ML-Schätzung für Bernoulli Verteilung

Für die zweite Ableitung gilt

$$\frac{\partial^2}{\partial^2 \theta} \log L(\theta, x) \Big|_{\theta=\hat{\theta}} = -\frac{1}{\hat{\theta}^2} \sum_{i=1}^n x_i - \frac{1}{(1-\hat{\theta})^2} \left(n - \sum_{i=1}^n x_i \right) < 0.$$

Also ist $\hat{\theta} = \bar{x}$ das Maximum wenn $\bar{x} \in (0, 1)$.

Ist $\bar{x} = 0$, so gilt $\sup_{\theta \in (0,1)} \log L(\theta, \underline{0}) = \sup_{\theta \in (0,1)} n \log(1 - \theta)$. Da $\log(1 - \theta)$ fallend in θ ist, ist $\hat{\theta} = 0 = \frac{1}{n} \sum_{i=1}^n 0$.

Ist $\bar{x} = 1$, so gilt $\sup_{\theta \in (0,1)} \log L(\theta, \underline{1}) = \sup_{\theta \in (0,1)} n \log \theta$. Da $\log \theta$ wachsend in θ ist, ist $\hat{\theta} = 1 = \frac{1}{n} \sum_{i=1}^n 1$.

Insgesamt ist also $\hat{\theta}(x) = \bar{x}$ die ML-Schätzung von θ .

ML für Gleichverteilung mit Mittelpunkt θ

X_1, \dots, X_n u.i.v. mit $X_i \sim \mathcal{U}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \Theta = \mathbb{R}$.

Es gilt

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^n f_{\theta}(x_i) = \prod_{i=1}^n \mathbb{1}_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x_i) \\ &= \begin{cases} 1 & : \theta - \frac{1}{2} < x_i < \theta + \frac{1}{2} \quad \forall i \\ 0 & : \text{sonst,} \end{cases} \\ &= \mathbb{1}_{(x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})}(\theta). \end{aligned}$$

Dabei ist $x_{(1)} = \min\{x_1, \dots, x_n\}$ und $x_{(n)} = \max\{x_1, \dots, x_n\}$.

Für jede Schätzung $\hat{\theta} \in (x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})$ gilt

$$L(\hat{\theta}, x) = 1 = \sup_{\theta \in \mathbb{R}} L(\theta, x).$$

Jedes $\hat{\theta} \in (x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})$ ist ML Schätzung; **keine Eindeutigkeit.**

ML für Normalverteilung

X_1, \dots, X_n u.i.v. mit $X_i \sim \mathcal{N}_{\mu, \sigma^2}$, $\theta = (\mu, \sigma^2)$

Es gilt

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

und

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

Die log-Likelihood Gleichung besteht aus zwei Gleichungen, die man lösen kann. Durch Betrachten der Hessematrix sowie des Verhaltens an den Rändern bekommt man die ML-Schätzer

$$\hat{\mu}(X) = \bar{X} \quad \text{und} \quad \hat{\sigma}^2(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Ungleichung von Cramér Rao

$X = (X_1, \dots, X_n) \in \mathcal{X}$, $X \sim P \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

f_θ sind Dichten bzw. Wahrscheinlichkeitsgewichte von P_θ .

Für $\vartheta = h(\theta)$ sei $\hat{\vartheta}(X)$ ein Schätzer mit

$$m(\theta) = E_\theta[\hat{\vartheta}(X)].$$

Definiere **Fisher Information** durch

$$I(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2 \right]$$

Unter bestimmten Regularitätsvoraussetzungen und falls $0 < I(\theta) < \infty$ gilt die **Cramér-Rao Ungleichung**

$$\text{Var}_\theta[\hat{\vartheta}(X)] \geq \frac{(m'(\theta))^2}{I(\theta)}.$$

Schätzer, für die asymptotisch die Gleichung gilt, heißen **effizient**. ML Schätzer sind unter relativ milden Bedingungen an das Modell effizient.

Elemente der Testtheorie

Statistische Tests

- ▶ Beobachtungsmodell $\{(\mathcal{X}, \mathcal{B}, P_\theta) : \theta \in \Theta\}$.
- ▶ $\Theta = \Theta_0 \cup \Theta_1$ eine disjunkte Vereinigung. Ist das wahre θ in Θ_0 oder in Θ_1 ?

Wir betrachten folgende Hypothesen

$$H_0 : \theta \in \Theta_0 \quad (\text{Nullhypothese}), \quad H_1 : \theta \in \Theta_1 \quad (\text{Alternative}).$$

Eine Funktion d mit $d : \mathcal{X} \rightarrow \{0, 1\}$ heißt **Test für H_0 gegen H_1** wenn sie messbar ist, und wie folgt interpretiert werden kann:

$$\begin{aligned} d(x) = 1 & \iff \text{Verwerfen der Nullhypothese } H_0, \\ d(x) = 0 & \iff \text{Annahme der Alternative von } H_0. \end{aligned}$$

Insbesondere gibt es einen **kritischen Bereich** $K \in \mathcal{B}$ mit $d = \mathbb{1}_K$.

Bemerkung: es gibt **randomisierte** Tests mit $d : \mathcal{X} \rightarrow [0, 1]$.

Fehler, Gütefunktion und Fehlerwahrscheinlichkeiten

Beim Testen können im Prinzip folgende Fehler entstehen:

	$d(x) = 0$	$d(x) = 1$
$\theta \in \Theta_0$	kein Fehler	Fehler 1. Art
$\theta \in \Theta_1$	Fehler 2. Art	kein Fehler

Wir bezeichnen die entsprechenden Irrtumswahrscheinlichkeiten mit

$$\alpha_d(\theta) := P_\theta(d(X) = 1), \theta \in \Theta_0 \quad (\text{W'keit für Fehler 1. Art})$$

$$\beta_d(\theta) := P_\theta(d(X) = 0), \theta \in \Theta_1 \quad (\text{W'keit für Fehler 2. Art})$$

Definition (Gütefunktion)

Die Funktion $G_d : \Theta \rightarrow [0, 1]$, $\theta \mapsto G_d(\theta) = P_\theta(d(X) = 1) = E_\theta[d(X)]$ heißt **Gütefunktion der Tests d (auch Schärfe, Trennschärfe, Power)**.

Dabei gilt

$$\alpha_d(\theta) = G_d(\theta) \quad \text{für } \theta \in \Theta_0$$

$$\beta_d(\theta) = 1 - G_d(\theta) \quad \text{für } \theta \in \Theta_1.$$

Typisches Vorgehen beim Testen

- ▶ Im Allgemeinen ist es nicht möglich Fehler beider Arten klein zu halten (→ Beispiel: Einseitiger Gaußtest).
- ▶ Typischerweise wird die Hypothese H_0 so gewählt, dass man sie durch Experimente ablehnen will.
- ▶ Fehler 1. Art ist mit diesem Ansatz der „schlimmere“ der beiden Fehler. Diesen möchte man kontrollieren und unter einer vorgegebenen Schranke halten.
- ▶ Unter solchen Tests sucht man nach Tests mit möglichst kleiner Fehlerwahrscheinlichkeit 2. Art.

Beispiel

- ▶ Eine Gruppe von Umweltschützer wollen beweisen, dass der Gehalt eines Schadstoffs in einem Lebensmittel zu hoch ist . Der Hersteller will das Gegenteil beweisen.
- ▶ Sei μ der wahre Inhalt und μ_0 der zugelassene Grenzwert.
- ▶ Die Testansätze der beiden Seiten werden lauten:
 - ▶ Umweltschützer:

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

- ▶ Hersteller:

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

Trennschärfste und unverfälschte Tests

Definition

- (i) Ein Test d für H_0 gegen H_1 heißt **Test zum Signifikanzniveau oder einfach Niveau α** , wenn $\alpha_d \leq \alpha$, also $G_d(\theta) \leq \alpha$ für alle $\theta \in \Theta_0$.
- (ii) Ein Test d^* zum Niveau α heißt **trennschärfster Test** (engl. uniformly most powerfull (UMP)) für H_0 gegen H_1 wenn für jedes $\theta \in \Theta_1$ gilt

$$G_{d^*}(\theta) = \sup\{G_d(\theta) : d \text{ Niveau } \alpha \text{ Test für } H_0 \text{ gegen } H_1\}.$$

- (iii) Ein Niveau α Test d heißt **unverfälscht** zum Niveau α , wenn $\beta_d \leq 1 - \alpha$, d.h. wenn $G_d(\theta) \geq \alpha$ für alle $\theta \in \Theta_1$.

Trennschärfste und unverfälschte Tests II

Lemma

Ein trennschärfster Niveau α Test d ist unverfälscht.

Beweis.

Sei d' eine Ber_α verteilte Zufallsvariable (unabhängig vom Rest).

$\Rightarrow G_d(\theta) = E_\theta[d'] = \alpha$ für alle $\theta \in \Theta$, d.h. d' ist ein Niveau α Test.

Ist d trennschärfster Niveau α Test, so gilt für alle $\theta \in \Theta_1$

$$G_d(\theta) \geq G_{d'}(\theta) = \alpha \quad \text{d.h. } d \text{ unverfälscht.}$$



Trennschärfste und unverfälschte Tests II

Lemma

Ein trennschärfster Niveau α Test d ist unverfälscht.

Beweis.

Sei d' eine Ber_α verteilte Zufallsvariable (unabhängig vom Rest).

$\Rightarrow G_{d'}(\theta) = E_\theta[d'] = \alpha$ für alle $\theta \in \Theta$, d.h. d' ist ein Niveau α Test.

Ist d trennschärfster Niveau α Test, so gilt für alle $\theta \in \Theta_1$

$$G_d(\theta) \geq G_{d'}(\theta) = \alpha \quad \text{d.h. } d \text{ unverfälscht.}$$



Klassisches Ziel:

- ▶ Finde trennschärfsten Niveau α Test unter allen Test zu finden.
- ▶ Falls kein trennschärfster existiert, dann finde trennschärfsten unter den unverfälschten Niveau α Tests.

Beispiel: einseitiger Gaußtest

$X = (X_1, \dots, X_n)$ u.i.v. mit $X_i \sim \mathcal{N}_{\mu, \sigma^2}$, $\theta = \mu \in \mathbb{R}$, $\sigma^2 > 0$ bekannt.
Für $\mu_0 \in \mathbb{R}$ betrachte

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

Plausibler Ansatz: Für zu bestimmendes c

$$d(X) = \mathbb{1}_{(c, \infty)}(\bar{X}).$$

Es gilt

$$\begin{aligned} G_d(\mu) &= P_\mu(d(X) = 1) = P_\mu(\bar{X} > c) = P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{c - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{c - \mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu - c}{\sigma/\sqrt{n}}\right) \quad (\text{wachsend in } \mu \text{ und fallend in } c). \end{aligned}$$

Für ein Test zum Niveau α muss gelten

$$\alpha \geq \sup_{\mu \leq \mu_0} G_d(\mu) = G_d(\mu_0) = \Phi\left(\frac{\mu_0 - c}{\sigma/\sqrt{n}}\right).$$

Beispiel: einseitiger Gaußtest II

$$\begin{aligned}\alpha \geq \Phi\left(\frac{\mu_0 - c}{\sigma/\sqrt{n}}\right) &\iff 1 - \alpha \leq \Phi\left(\frac{c - \mu_0}{\sigma/\sqrt{n}}\right) \\ &\iff z_\alpha = \Phi^{-1}(1 - \alpha) \leq \frac{c - \mu_0}{\sigma/\sqrt{n}}.\end{aligned}$$

Für jedes $c \geq \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$ ist

$$d(X) = \mathbb{1}_{(c, \infty)}(\bar{X}) \quad \text{ein Niveau } \alpha \text{ Test.}$$

Für trennschärfsten Test d^* (unter solchen Tests) muss gelten

$$G_{d^*}(\mu) = \sup\left\{G_d(\mu) : d(X) = \mathbb{1}_{(c, \infty)}(\bar{X}), c \geq \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}}\right\}.$$

Dies ist für $d^*(X) = \mathbb{1}_{(c^*, \infty)}(\bar{X})$ mit $c^* = \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$ der Fall, denn für $c > c^*$ gilt

$$G_{d^*}(\mu) = \Phi\left(\frac{\mu - c^*}{\sigma/\sqrt{n}}\right) > \Phi\left(\frac{\mu - c}{\sigma/\sqrt{n}}\right) = G_d(\mu)$$

für alle μ und insbesondere für alle $\mu > \mu_0$.

Beispiel: einseitiger Gaußtest III

Die Gütefunktion ist gegeben durch

$$G_{d^*}(\mu) = \Phi\left(\frac{\mu - c^*}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right).$$

Die maximalen Fehlerwahrscheinlichkeiten sind

$$\sup_{\mu \leq \mu_0} \alpha_{d^*}(\mu) = \sup_{\mu \leq \mu_0} G_{d^*}(\mu) = \sup_{\mu \leq \mu_0} \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right) = \Phi(-z_\alpha) = \alpha$$

und

$$\begin{aligned} \sup_{\mu \leq \mu_0} \beta_{d^*}(\mu) &= \sup_{\mu > \mu_0} (1 - G_{d^*}(\mu)) = 1 - \inf_{\mu > \mu_0} G_{d^*}(\mu) \\ &= 1 - G_{d^*}(\mu_0) = 1 - \Phi(-z_\alpha) = 1 - \alpha. \end{aligned}$$

Beispiel: einseitiger Gaußtest III

Die Gütefunktion ist gegeben durch

$$G_{d^*}(\mu) = \Phi\left(\frac{\mu - c^*}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right).$$

Die maximalen Fehlerwahrscheinlichkeiten sind

$$\sup_{\mu \leq \mu_0} \alpha_{d^*}(\mu) = \sup_{\mu \leq \mu_0} G_{d^*}(\mu) = \sup_{\mu \leq \mu_0} \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right) = \Phi(-z_\alpha) = \alpha$$

und

$$\begin{aligned} \sup_{\mu > \mu_0} \beta_{d^*}(\mu) &= \sup_{\mu > \mu_0} (1 - G_{d^*}(\mu)) = 1 - \inf_{\mu > \mu_0} G_{d^*}(\mu) \\ &= 1 - G_{d^*}(\mu_0) = 1 - \Phi(-z_\alpha) = 1 - \alpha. \end{aligned}$$

Bemerkung:

1. Einseitiger Gaußtest mit $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$ analog.
2. Zweiseitiger Gaußtest mit $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0 \rightarrow$ Übung.

t-Test

$X = (X_1, \dots, X_n)$ u.i.v. mit unbekanntem $\mu = E[X_i] \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$. Ähnlich wie beim zweiseitigen Test wollen wir

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$$

testen. Wir haben gesehen, dass

$$\tilde{I}(X) = \left(\bar{X} - t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \right)$$

ein Konfidenzintervall für μ zur Sicherheit $1 - \alpha$ ist. Wobei

$$\hat{\sigma}^2 = \hat{\sigma}^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{Stichprobenvarianz})$$

$$t_{n-1, \alpha/2} = F_{n-1}^{-1}(1 - \alpha/2), \quad \text{oberes } \alpha/2\text{-Quantil der } t_{n-1}\text{-Verteilung}$$

Das Intervall $\tilde{I}(X)$ ist „exakt“ wenn $X_i \sim \mathcal{N}_{\mu, \sigma^2}$ und approximativ sonst.

t -Test II

Wir machen folgenden Ansatz

$$d(X) = 1 \iff \mu_0 \notin \tilde{I}(X)$$

Dann gilt für den Fehler 1. Art

$$P_{\mu_0}(d(X) = 1) = P_{\mu_0}(\mu_0 \notin \tilde{I}(X)) \leq \alpha$$

Also ist d ein Test für $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ zum Niveau α .

Fishers exakter Test

Seien A und B zwei Ereignisse auf einem W 'Raum. Wir nehmen an, dass wir n Versuche beobachten und zählen wie oft das Ereignis jeweils in den folgenden Mengen landet

$$A \cap B, A \cap B^C, A^C \cap B, A^C \cap B^C$$

Man kann die Ergebnisse in einer **Kontingenztafel** zusammenfassen

	A	A^C	total
B	X_{11}	X_{12}	n_1
B^C	X_{21}	X_{22}	n_2
total	m_1	m_2	n

Es soll getestet werden:

H_0 : A und B sind unabhängig, gegen H_1 : A und B sind abhängig

Fishers exakter Test II

Bestimme die Verteilung von X_{11} gegeben

- ▶ $X_{11} + X_{12}$ = Gesamtanzahl der Versuche in B
- ▶ $X_{11} + X_{21}$ = Gesamtanzahl der Versuche in A

Interpretation als Urnenexperiment

- ▶ n = Gesamtanzahl der Kugeln in der Urne
- ▶ $n_1 = X_{11} + X_{12}$ sind markiert
- ▶ Es werden $m_1 = X_{11} + X_{21}$ Kugeln gezogen, wovon X_{11} markiert sind.

⇒ X_{11} ist unter H_0 hypergeometrisch verteilt:

$$P_0(X_{11} = k) = \binom{n_1}{k} \binom{n - n_1}{m_1 - k} / \binom{n}{m_1}.$$

Liegen konkrete Beobachtungen vor so kann man obige W 'keit ausrechnen¹. Ist diese zu klein, dann lehnt man die Nullhypothese ab.

¹Diese W 'keit wird als **p -Wert** bzw. **beobachtetes Signifikanzniveau** bezeichnet. Es ist das kleinste Signifikanzniveau zu dem die Hypothese noch abgelehnt wird.

Markovketten

Irrfahrt

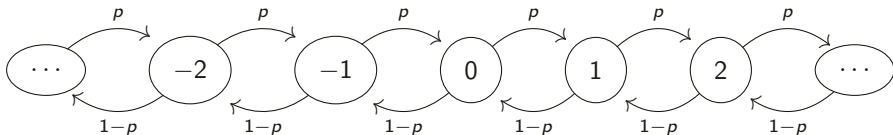
Es ξ_1, ξ_2, \dots eine Folge u.i.v. Zufallsvariablen mit $P(\xi_i = 1) = p$ und $P(\xi_i = -1) = 1 - p$ für ein $p \in [0, 1]$

Für $x \in \mathbb{Z}$ definieren wir $(X_n)_{n \geq 0}$ durch

$$X_n = x + \sum_{i=1}^n \xi_i.$$

Wegen $X_{n+1} = X_n + \xi_{n+1}$ und wegen Unabhängigkeit von ξ_{n+1} von ξ_1, \dots, ξ_n und somit auch von X_0, \dots, X_n ist klar, dass X_{n+1} von X_0, \dots, X_n nur durch X_n abhängt.

$(X_n)_{n=0,1,\dots}$ heißt **Irrfahrt (engl. random walk) auf \mathbb{Z}** .



Markovketten und Markoveigenschaft

I endliche Menge (Verallgemeinerung auf abzählbar unendlich möglich).

- ▶ $P = (p_{ij})_{i,j \in I}$ heißt **stochastische Matrix**, wenn

$$p_{ij} \geq 0 \quad \text{für alle } i, j \in I \quad \text{und} \quad \sum_{j \in I} p_{ij} = 1 \quad \text{für alle } i \in I.$$

- ▶ Sei $\nu = (\nu(i))_{i \in I}$ eine Verteilung auf I und P wie oben. Stochastischer Prozess $(X_n)_{n \geq 0}$ ist eine **Markovkette** mit **Anfangsverteilung** ν und **Übergangsmatrix** P , wenn für alle $n \in \mathbb{N}$ und alle $i_0, \dots, i_{n-1}, i, j \in I$

$$P(X_0 = i_0) = \nu(i_0),$$

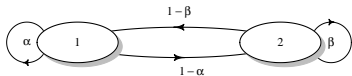
$$P(X_{n+1} = j | X_n = i, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = p_{ij}.$$

Die Eigenschaft letzte Eigenschaft wird **Markoveigenschaft** genannt.

- ▶ $P^n = (p_{ij}^{(n)})_{i,j \in I}$ heißt **n -Schritte Übergangsmatrix**

$$P(X_n = j | X_0 = i) = p_{ij}^{(n)}.$$

Beispiel



Übergangendiagramm einer allgemeinen Markovkette mit Zustandsraum $I = \{1, 2\}$, wobei $\alpha, \beta \in [0, 1]$. Die Übergangsmatrix ist

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix}.$$

Aperiodizität und Irreduzibilität

- ▶ Eine Markovkette heißt **irreduzibel**, wenn jeder Zustand von jedem Zustand aus mit positiver Wahrscheinlichkeit erreichbar ist.
- ▶ **Periode** eines Zustandes $i \in I$ ist definiert als

$$d_i = \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\},$$

mit $d_i = +\infty$, falls $p_{ii}^{(n)} = 0$ für alle $n \geq 1$ ist. Ist $d_i = 1$, so heißt der Zustand i **aperiodisch**.

- ▶ Eine Markovkette heißt **aperiodisch**, wenn alle Zustände aperiodisch sind.

Invariante Verteilungen

- ▶ Wir sagen, dass ein W-Maß $(\pi_i)_{i \in I}$ eine *invariante Verteilung* für die Markovkette mit Übergangsmatrix P ist, wenn

$$\pi P = \pi$$

- ▶ Wird eine MK in der invarianten Verteilung gestartet, so bleibt sie in der invarianten Verteilung.

Satz

Sei $X = (X_n)_{n=0,1,\dots}$ eine MK auf I (endlich) mit Übergangsmatrix P .
Dann gilt:

- X besitzt eine invariante Verteilung (i.A. mehr als eine möglich!).
- Ist X irreduzibel, dann ist invariante Verteilung π eindeutig.
- Ist X irreduzibel und aperiodisch, so gilt auch

$$p_{ij}^{(n)} \xrightarrow{n \rightarrow \infty} \pi_j > 0, \quad \text{für alle } i, j \in I.$$

Ergodensatz - Verallgemeinerung des GGZ

Satz

Es sei $(X_n)_{n \geq 0}$ eine irreduzible Markovkette mit Übergangsmatrix P und invarianter Verteilung π .

Für jede beschränkte Funktion $f : I \rightarrow \mathbb{R}$ gilt

$$P \left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{n \rightarrow \infty} \langle f, \pi \rangle \right) = 1,$$

wobei

$$\langle f, \pi \rangle := \sum_{i \in I} f(i) \pi_i$$

das Integral von f bezüglich π ist.