STATISTIK

Grundgesamtheit : Querschnittstudien, Longitudinalstudien, Zeitreihen

$\downarrow$

Merkmal : Nominalskala, Ordinalskala, Intervallskala, Verhältnisskalen

Quantile -> Median -> Range -> Fences -> Whiskers

Momente: $\frac{1}{n}\sum_{i=1}^{n} x_i^r$ Zentrale Momente: $\frac{1}{n}\sum (x_i-\bar{x})^r$

Merkmalsraum Ω -> Ereignis -> σ-Algebra (System)
↳ $P(\Omega)$
-> Borelmengen

} Wahrscheinlichkeitsmaß

=> Messraum (Ω, A) => W-Raum (Ω, A, P)

- Additionstheorem
- Bedingte Wahrscheinlichkeit
- Multiplikationstheorem
- Vollständige Wahrscheinlichkeit
- Bayessche Formel (Odds)
- Unabhängigkeit: $P(AB) = P(A|B) \cdot P(B) = P(A) \cdot P(B)$
- Mehrstufige Experimente -> Wahrscheinlichkeitsraum

Merkmalsraum $\rightarrow$ Stochastische Größe $\leftarrow$ Ereignissystem (Urbild $\subseteq G$)

Verteilungsfunktion: $0 \leq F(x) \leq 1$, monoton wachsend, $\lim F(x) = 0/1$; rechtsstetig

Verallgemeinerte Inverse: Quantile

Transformationen:

- Methode der Verteilungsfunktion
- Transformationssatz für Dichten $\quad f_Y(y) = f_X(g^{-1}(y)) \cdot \overbrace{\left| \frac{d\, g^{-1}(y)}{dy} \right|}^{\text{Jacobian}}$
  - ↳ affine Transformationen

Erwartungswert: $\mathbb{E}(X) = \sum x \cdot p(x) = \int_{-\infty}^{\infty} x \cdot p(x)$ $\qquad \mathbb{E}(a) = a$, $\mathbb{E}(aX+b) = a \cdot \mathbb{E}(X) + b$

↳ LotUS: $\mathbb{E}(Y) = g(x) \cdot p(x)$

↳ Alternative: Transformation

Varianz: $Var(X) = \mathbb{E}[(X - \mu_X)^2] = \sum (x - \mathbb{E}(X))^2 \, p_X(x)$ $\qquad Var(a) = 0$, $Var(aX+b) = a^2 Var(X)$

↳ Verschiebungssatz: $Var(X) = \mathbb{E}(X^2) - \mu^2$

Inversionsmethode

Bivariate Verteilung: Merkmalsraum $\rightarrow$ 2 stochastische Größen

- Randverteilung
- Bedingte Verteilung: $\frac{p(x_1, x_2)}{p(x_1)} = P(X_2 = x_2 \mid X_1 = x_1)$

Korrelation: $Cov(X,Y) = \mathbb{E}((X - \mu_x)(Y - \mu_y)) = \mathbb{E}(XY) - \mu_x \mu_y$

$\rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

Unabhängigkeit: $f(x_1, x_2) = f(x_2 \mid x_1) \cdot f(x_1) = f(x_2) \cdot f(x_1)$

↳ $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$

↳ $Cov(X,Y) = 0$

Transformationen: Serien: $Y = \min\{X_1, X_2 \ldots\}$

$F_Y(y) = 1 - P(X_1 > y, X_2 > y, \ldots)$

$= 1 - \prod P(X > y)$

$= 1 - \prod (1 - F_X(y))$

Parallel: $Y = \max\{X_1, X_2, \ldots\}$

$F_Y(y) = P(X_1 < y, X_2 < y, \ldots)$

$= \prod F_X(y)$

k-aus-n: $\sum_{j=0}^{k} \binom{n}{j} [1-F(y)]^j \cdot F(y)^{n-j}$

Stichprobenmittelwert: $\mathbb{E}(X_n) = \mu$

$\mathrm{Var}(X_n) = \frac{\sigma^2}{n}$

Faltung

↳ Additionstheoreme

Konvergenz: Markow'sche Ungleichung: $P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$

Tschebyschew'sche Ungleichung: $P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$

- Konvergenz in der Wahrscheinlichkeit:

$$X_n \xrightarrow{P} X: \quad \lim_{n\to\infty} P(|X_n - X| \ngeq \varepsilon) = 0 \qquad \varepsilon > 0$$

- Konvergenz in der Verteilung

$$X_n \xrightarrow{D} X: \quad \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad x \in C(F_X)$$

- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$

Schwaches Gesetz der großen Zahlen: $\overline{X}_n \xrightarrow{P} \mu$

Starkes Gesetz der großen Zahlen: $P(\lim \overline{X}_n = \mu) = 1$ (fast sichere Konvergenz)

Zentraler Grenzverteilungssatz

$$Y_n = \frac{\sum X_i - n\mu}{\sqrt{\sigma^2 n}} = \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2 n}{n^2}}} = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n} \cdot (\overline{X}_n - \mu)}{\sigma}$$

$$F(Y_n \leq z) \to \Phi(z) \Rightarrow Y_n \xrightarrow{D} N(0,1)$$

$$\Rightarrow X_n \xrightarrow{D} N(\mu_{X_n}, \sigma^2_{X_n}) \qquad \mu_{X_n} = n \cdot \mu_X \qquad \sigma^2_{X_n} = n \cdot \sigma^2_X$$

Normalapproximation: $\sum_{i=1}^{n} X_i \approx N(n \cdot \mu_X, n\, \sigma^2_X)$

↳ Stetigkeitskorrektur

Stichprobe: $X = (X_1, X_2, \dots, X_n)$ X_i iid

Statistik: Funktion einer Stichprobe
<-> Schätzfunktionen

Stichprobenmittelwert: $\bar{X}_n = \frac{1}{n} \sum X_i$

Stichprobenvarianz: $S_n = \frac{1}{n-1} \sum (X_i - X_n)^2$

Empirische Verteilungsfunktion: $\hat{F}_n(x) = \frac{1}{n} \sum I_{(-\infty, x]}(X_i)$

Fundamentalsatz der Statistik: $P\left(\lim_{n \to \infty} \sup_{x \in R} |\hat{F}_n(x) - F(x)| = 0\right) = 1$

Momentschätzer

Maximum Likelihood Schätzer -> Invarianz

- Erwartungstreue: $E_\Theta(\hat{\Theta}_n) = \Theta$

  $\Downarrow$
- Unverzerrt (asympt. Erwartungstr.): $E_\Theta(\hat{\Theta}_n) \to \Theta$ $n \to \infty$
- Effizienz: $Var(\hat{\Theta}_1) < Var(\hat{\Theta}_2) \Rightarrow \hat{\Theta}$ effizienter
- Konsistenz: $\hat{\Theta} \xrightarrow{P} \Theta$ für $n \to \infty$
- Asymptotische Normalverteilung: $\hat{\Theta}_n \overset{asympt.}{\sim} N(E(\hat{\Theta}_n), Var(\hat{\Theta}_n))$

Konvergenz in Wahrscheinlichkeit: GGZ $\bar{X}_n \xrightarrow{P} \mu$
Konsistenz

Fast sichere Konvergenz: GGZ
Fundamentalsatz

Konvergenz in Verteilung: ZGVS

$(T_1(X), T_2(X))$: Konfidenzintervall für Θ mit Konfidenzkoeffizient $1-\alpha$
wenn $P_\Theta(T_1(X) < \Theta < T_2(X)) \geq 1-\alpha$

Pivotmethode: Formuliere statistisches Modell für X
Wähle Pivotgröße $T(X, \Theta)$
Bestimme Verteilung des Pivots
Bestimme Quantile der Pivotverteilung
...

Resampling und Bootstrapping

Parametertests: $H_0: \Theta \in \Theta_0$ gegen $H_1: \Theta \in \Theta_1$

Typ 1 Fehler: Verwerfe H_0 obwohl sie zutrifft
Typ 2 Fehler: Akzeptiere H_0 obwohl sie nicht zutrifft

↳ $\alpha = P(\text{Typ 1 Fehler}) = P_{\Theta_0}(X \in C)$ -> Signifikanzniveau
$\beta = P(\text{Typ 2 Fehler}) = P_\Theta(X \in C^c)$

C ... kritischer Bereich (bei Test definiert)

Power: Wahrscheinlichkeit Verwerfung H_0 wenn H_1 zutrifft
=> $1 - \beta(\Theta)$

p-Wert: beobachtetes Signifikanzniveau
Wahrscheinlichkeit bei Zutreffen von H_0 eine extremere Beobachtung zu bekommen

... Teststatistiken
↳ gepoolte Varianzschätzer (gewichteter Mittelwert)

Verteilungshypothese

QQ-Plot
Chiquadrat Anpassungstest

Diskrete Verteilungen

Diskrete uniforme Verteilung

$p(x_i) = 1/n \qquad \mathbb{E}(X) = \bar{x} \qquad Var(X) = 1/n \sum (x_i - \bar{x})^2$

Bernoulli Verteilung: $A(p) \quad = B(1,p)$

$p(x) = p^x \cdot (1-p)^{1-x} \qquad x \in \{0,1\} \qquad \mathbb{E}(X) = p \qquad Var(X) = p \cdot (1-p)$

Binomial Verteilung: $B(n,p)$

$p(x) = \binom{n}{x} p^x \cdot (1-p)^{n-x} \qquad \mathbb{E}(X) = n \cdot p \qquad Var(X) = n \cdot p \cdot (1-p)$

Negative Binomial Verteilung: $NB(n,p)$

$p(x) = \binom{x-1}{n-1} p^n \cdot (1-p)^{x-n} \qquad \mathbb{E}(X) = \frac{n}{p} \qquad Var(X) = \frac{n \cdot (1-p)}{p^2}$

Geometrische Verteilung: $G(p) \quad = NB(1,p)$

$p(x) = p \cdot (1-p)^{x-1} \qquad \mathbb{E}(X) = 1/p \qquad Var(X) = \frac{1-p}{p^2}$

Hypergeometrische Verteilung: $H(N,A,n)$

$p(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} \qquad \mathbb{E}(X) = n \cdot \frac{A}{N} \qquad Var(X) = n \cdot \frac{A}{N}\left(1 - \frac{A}{N}\right)\left(\frac{N-n}{N-1}\right)$

Poisson Verteilung $P(\lambda)$

$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \mathbb{E}(X) = \lambda \qquad Var(X) = \lambda$

Stetige Verteilungen

Stetige uniforme Verteilung : $\mathcal{U}(a,b)$

$$f(x) = \begin{cases} 1/b-a & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases} \qquad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Exponentialverteilung : $\mathrm{Exp}(\lambda)$, $\mathrm{Exp}(\tau)$ $\quad \lambda = 1/\tau \quad = \mathrm{Gam}(1,\lambda)$

$$f(x) = \lambda e^{-\lambda x} \qquad F(x) = 1 - e^{-\lambda x}$$

$$\mathbb{E}(X) = 1/\lambda \qquad \mathrm{Var}(X) = 1/\lambda^2$$

Gammaverteilung : $\mathrm{Gam}(\alpha, \lambda)$

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \qquad \mathbb{E}(x) = \frac{\alpha}{\lambda} \qquad \mathrm{Var}(x) = \frac{\alpha}{\lambda^2}$$

Chiquadratverteilung : $\chi^2(n)$ $\quad = \mathrm{Gam}(n/2, 1/2)$

$$f(x) = \frac{(1/2)^{n/2} x^{n/2-1} e^{-1/2 x}}{\Gamma(n/2)} \qquad \mathbb{E}(x) = n \qquad \mathrm{Var}(x) = 2n$$

Quantile : $\chi_{n;p}$

Normalverteilung : $N(\mu, \sigma^2)$

→ Standardnormalverteilung : $N(0,1)$

Verteilungsfkt: Φ, $\Phi(-x) = 1 - \Phi(x)$ $\qquad$ Quantile : z_p, $z_p = -z_{1-p}$

Standardisierung: $X \sim N(\mu, \sigma^2) \Rightarrow Z = (X-\mu)/\sigma \sim N(0,1)$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) \qquad x_p = \mu + \sigma \cdot z_p$$

$$\mathbb{E}(X) = \mu \qquad \mathrm{Var}(X) = \sigma^2$$

F-Verteilung : $F(m,n)$

$f(x)$ = Skriptum S174 $\qquad \mathbb{E}(X) = \frac{n}{n-2} \qquad Var(X)$ = Skriptum

$X \sim F(m,n) = \frac{1}{X} \sim F(n,m) \qquad$ Quantile: $F_{m;n;p} = \frac{1}{F_{m;n;1-p}}$

t-Verteilung : $t(n)$

$f(x)$ = Skriptum S176 $\qquad \mathbb{E}(X) = 0 \qquad Var(X) = \frac{n}{n-2}$

Quantile: $t_{n;p} = -t_{n;1-p}$

Betaverteilung : $Be(a,b)$

$f(x)$ = Skriptum S177 $\qquad \mathbb{E}(X) = \frac{a}{a+b} \qquad Var(X) = \frac{a \cdot b}{(a+b+1)(a+b)^2}$

$X \sim Be(a,b) \Rightarrow 1-X \sim Be(b,a)$