Statistics

Collection of Formulas

Contents

1 Deskriptive Statistics

1.1 Summary Statistics: Sample

1.1.1 Location

Mode Most frequent value of x_i. Two or more modes are possible (bimodal).

Median

$$\tilde{x}_{0.5} = \begin{cases} x_{((n+1)/2)} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}) & \text{falls } n \text{ gerade} \end{cases}$$

Quantile

$$\tilde{x}_\alpha = \begin{cases} x_{(k)} & \text{falls } n\alpha \notin \mathbb{N} \\ \frac{1}{2}(x_{(n\alpha)} + x_{(n\alpha+1)}) & \text{falls } n\alpha \text{ ganzzahlig} \end{cases}$$

with

$k = \min x \in \mathbb{N}, \quad x > n\alpha$

Minimum/Maximum

$$x_{\min} = \min_{i \in \{1,\ldots,N\}}(x_i) \qquad x_{\max} = \max_{i \in \{1,\ldots,N\}}(x_i)$$

Arithmetic Mean

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

Estimates the expectation $\mu = E[X]$ (first moment).

Calculation Rules:

- ⋆ $E(a + b \cdot X) = a + b \cdot E(X)$
- ⋆ $E(X \pm Y) = E(X) \pm E(Y)$

Geometric Mean

$$\bar{x}_G = \sqrt[n]{\sum_{i=1}^{n} x_i}$$

For growth factors: $\bar{x}_G = \sqrt[n]{\frac{B_n}{B_0}}$

Harmonic Mean

$$\bar{x}_H = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}}$$

1.1.2 Dispersion

Range

$$R = x_{(n)} - x_{(1)}$$

Interquartile Range

$$d_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}$$

(Empirical) Variance

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

Estimates the second centralized moment.

Calculation Rules:

- ⋆ $Var(aX + b) = a^2 \cdot Var(X)$
- ⋆ $Var(X \pm Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

(Empirical) Standard Deviation

$$s = \sqrt{s^2}$$

Coefficient of Variation

$$\nu = \frac{s}{\bar{x}}$$

Average Absolute Deviation

$$e = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

Estimates the first absolute centralized moment.

1.1.3 Concentration

Gini Coefficient

$$G = \frac{2\sum_{i=1}^{n} i x_{(i)} - (n+1)\sum_{i=1}^{n} x_{(i)}}{n\sum_{i=1}^{n} x_{(i)}} = 1 - \frac{1}{n}\sum_{i=1}^{n}(v_{i-1} + v_i)$$

with

$$u_i = \frac{i}{n}, \quad v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{i} x_{(j)}} \qquad (u_0 = 0, \quad v_0 = 0)$$

These are also the values for the Lorenz curve.

Range: $0 \le G \le \frac{n-1}{n}$

Lorenz-Münzner Coefficient (normed G)

$$G^+ = \frac{n}{n-1}G$$

Range: $0 \le G^+ \le 1$

1.1.4 Shape

(Empirical) Skewness

$$\nu = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Estimates the third centralized moment, scaled with $(\sigma^2)^{\frac{2}{3}}$

(Empirical) Kurtosis

$$k = \left[n(n+1) \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3(n-1) \right] \cdot \frac{n-1}{(n-2)(n-3)} + 3$$

Estimates the fourth centralized moment, scaled with $(\sigma^2)^2$

Excess

$$\gamma = k - 3$$

1.1.5 Dependence

for two nominal variables

χ^2-Statistic

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \frac{n_{i+} n_{+j}}{n})^2}{\frac{n_{i+} n_{+j}}{n}} = n \left(\sum_{i=1}^{k} \sum_{j=1}^{l} \frac{n_{ij}^2}{n_{i+} n_{+j}} - 1 \right)$$

Range: $0 \leq \chi^2 \leq n(\min(k,l) - 1)$

Phi-Coefficient

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

Range: $0 \leq \Phi \leq \sqrt{\min(k,l) - 1}$

Cramér's V

$$V = \sqrt{\frac{\chi^2}{\min(k,l) - 1}}$$

Range: $0 \leq V \leq 1$

Contingency Coefficient C

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Range: $0 \leq C \leq \sqrt{\frac{\min(k,l)-1}{\min(k,l)}}$

Corrected Contingency Coefficient C_{corr}

$$C_{corr} = \sqrt{\frac{\min(k,l)}{\min(k,l) - 1}} \cdot \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Range $0 \leq C_{corr} \leq 1$

Odds-Ratio

$$OR = \frac{ad}{bc} = \frac{n_{ii} n_{jj}}{n_{ij} n_{ji}}$$

Range: $0 \leq OR < \infty$

for two ordinal variables

Gamma (Goodman and Kruskal)

$$\gamma = \frac{K - D}{K + D}$$

$K = \sum_{i<m} \sum_{j<n} n_{ij} n_{mn}$	Number of concordant pairs
$D = \sum_{i<m} \sum_{j>n} n_{ij} n_{mn}$	Number of reversed pairs

Range: $-1 \leq \gamma \leq 1$

Kendall's τ_b

$$\tau_b = \frac{K - D}{\sqrt{(K + D + T_X)(K + D + T_Y)}}$$

with

$T_X = \sum_{i=m} \sum_{j<n} n_{ij} n_{mn}$	Number of ties w.r.t. X
$T_Y = \sum_{i<m} \sum_{j=n} n_{ij} n_{mn}$	Number of ties w.r.t. Y

Range: $-1 \leq \tau_b \leq 1$

Kendall's/Stuart's τ_c

$$\tau_c = \frac{2 \min(k,l)(K - D)}{n^2(\min(k,l) - 1)}$$

Range: $-1 \leq \tau_c \leq 1$

Spearman's Rank Correlation Coefficient

$$\rho = \frac{n(n^2-1) - \frac{1}{2} \sum_{j=1}^{J} b_j(b_j^2 - 1) - \frac{1}{2} \sum_{k=1}^{K} c_k(c_k^2 - 1) - 6 \sum_{i=1}^{n} d_i^2}{\sqrt{n(n^2-1) - \sum_{j=1}^{J} b_j(b_j^2 - 1)} \sqrt{n(n^2-1) - \sum_{k=1}^{K} c_k(c_k^2 - 1)}}$$

or

$$\rho = \frac{s_{rg_x rg_y}}{\sqrt{s_{rg_x rg_x} s_{rg_y rg_y}}}$$

Without ties:

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

with

$d_i = R(x_i) - R(y_i)$ rank difference

Range: $-1 \leq \rho \leq 1$

for two metric variables

Correlation Coefficient (Bravais-Pearson)

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}}$$

with

$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2$ or $s_{xy} = \frac{S_{xy}}{n}$

$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$ or $s_{xx} = \frac{S_{xx}}{n}$

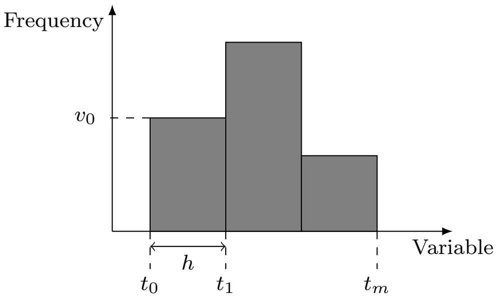
$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$ or $s_{yy} = \frac{S_{yy}}{n}$

Range: $-1 \leq r \leq 1$

1.2 Tables

1.3 Diagrams

1.3.1 Histogram



sample: $X = \{x_1, x_2, ...; x_n\}$

k-th bin: $B_k = [t_k, t_{k+1}), k = \{0, 1, ..., m-1\}$

Number of observations in the k-th bin: v_k

bin width: $h = t_{k+1} - t_k, \forall k$

Scott's Rule

$$h^* \approx 3.5 \sigma n^{-\frac{1}{3}}$$

For approximately normal distributed data (min. MSE)

1.3.2 QQ-Plot

1.3.3 Scatterplot

2 Probability

2.1 Combinatorics

		without replacement	with replacement
Permutations		$n!$	$\frac{n!}{n_1! \cdots n_s!}$
Combinations:	without order	$\binom{n}{m}$	$\binom{n+m-1}{m}$
	with order	$\binom{n}{m} m!$	n^m

with:

$n! = n \cdot (n-1) \cdot ... \cdot 1$

$\binom{n}{m} = \frac{n!}{m!(n-m)!}$

2.2 Probability Theory

Laplace

$$P(A) = \frac{|A|}{|\Omega|}$$

Kolmogorov Axioms mathematical definition of probability

(1) $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{A}$

(2) $P(\Omega) = 1$

(3) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
$\forall A_i \in \mathcal{A}, i = 1, ..., \infty$ with $A_i \cap A_j = \emptyset$ for $i \neq j$

Implications:

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$, f'ur $A_i, ..., A_n$ complete decomposition of Ω into pairwise disjoint events

Probability (Mises) frequentist definition of probability

$$P(A) = \lim_{n \to \infty} \frac{n_A(n))}{n}$$

with n repetitions of a random experiment and $n_A(n)$ events A

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{für } P(B) > 0$$

$\Rightarrow P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$

Law of Total Probability

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{for } P(A), P(B) > 0$$

Stochastic Independence

A, B independent $\Leftrightarrow P(A \cap B) = P(A) + P(B)$

X, Y independent $\Leftrightarrow f_{XY}(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall x, y$

2.3 Random Variables/Vectors

Random Variables $\in \mathbb{R}$

Definition

$$Y : \Omega \to \mathbb{R}$$

The Subset of possible values for $\mathbb{R}$ is called support.
Notation: Realisations of Y are depicted with lower case letters.
$Y = y$ means, that y is the realisation of Y.

Discrete and Continuous Random Variables

If the support is uncountably infinite, the random variable is called *continuous*, otherwise it is called *discrete*.

- **Density $f(\cdot)$:**

  For continuous variables: $P(Y \in [a,b]) = \int_a^b f_Y(y)dy$

  For discrete variables the density (and other functions) can be depicted like the corresponding function for continuous variables, if the notation is extended as follows: $\int_{-\infty}^{y} f_Y(\tilde{y})d\tilde{y} := \sum_{k:k\leq y} P(Y=k)$. This notation is used.

- **Cumulative Distribution Function $F(\cdot)$:**
  $F_Y(y) = P(Y \leq y)$

Relationship:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(\tilde{y})d\tilde{y}$$

Moments

- **Expectation (1. Moment)**: $\mu = E(Y) = \int y f_Y(y)dy$

- **Variance (2. centralized Moment)**:
  $\sigma^2 = Var(Y) = E(\{Y - E(Y)\}^2) = \int (y - E(Y))^2 f(y)dy$
  Note: $E(\{Y-\mu\}^2) = E(Y^2) - \mu^2$

  Proof:
  $E(\{Y-\mu\}^2) = E(Y^2 - 2Y\mu + \mu^2) = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2$

- **kth Moment**: $E(Y^k) = \int y^k f_Y(y)dy$,
  k. centralized Moment: $E(\{Y - E(Y)\}^k)$

Moment Generating Function

$$M_Y(t) = \mathrm{E}(e^{tY})$$

with $\left.\frac{\partial^k M_Y(t)}{\partial t^k}\right|_{t=0} = \mathrm{E}(Y^k)$

Cumulant Generating Function $K_Y(t) = \log M_Y(t)$

A random variable is uniquely defined by its moment generating function and vice versa (as long as moments and cumulants are finite).

Random Vectors $\in \mathbb{R}^q$

Density and Cumulative Distribution Function

$$F(y_1, ..., y_q) = P(Y_1 \leq y_1, ..., Y_q \leq y_q)$$

$$P(a_1 \leq Y_1 \leq b_1, ..., a_q \leq Y_q \leq b_q)$$
$$= \int_{a_1}^{b_1} ... \int_{a_q}^{b_q} f(y_1, .., y_q)dy_1...dy_q$$

Marginal Density

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(y_1, ..., y_k)dy_2...dy_k$$

Conditional Density

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{f(y_1, ..., y_2)}{f(y_2)} \text{ f'ur } f(y_2) > 0$$

Iterative Expectation

$$\mathrm{E}(Y) = \mathrm{E}_X(\mathrm{E}(Y|X))$$

Proof:

$$\mathrm{E}(Y) = \int y f(y)dy = \int\int y f(y|x)dy f_X(x)dx = \mathrm{E}_X(\mathrm{E}(Y|X))$$

$$\mathrm{Var}(Y) = \mathrm{E}_X(\mathrm{Var}(Y|X)) + \mathrm{Var}_X(\mathrm{E}(Y|X))$$

Proof:

$$\begin{aligned}
\mathrm{Var}(Y) &= \int (y - \mu_Y)^2 f(y)dy \\
&= \int (y - \mu_Y)^2 f(y|x)f(x)dydx \\
&= \int (y - \mu_{Y|x} + \mu_{Y|x} - \mu_Y)^2 f(y|x)f(x)dydx \\
&= \int (y - \mu_{Y|x})^2 f(y|x)f(x)dydx + \\
&\quad \int (\mu_{Y|x} - \mu_Y)^2 f(y|x)f(x)dydx + \\
&\quad 2\int (y - \mu_{Y|x})(\mu_{Y|x} - \mu_Y)f(y|x)f(x)dydx \\
&= \int \mathrm{Var}(Y|x)f(x)dx + \int (\mu_{Y|x} - \mu_Y)^2 f(x)dx \\
&= \mathrm{E}_X(\mathrm{Var}(Y|X)) + \mathrm{Var}_X(\mathrm{E}(Y|X))
\end{aligned}$$

2.4 Probability Distributions

2.4.1 Discrete Distributions

Diskrete Gleichverteilung

$$Y \sim \mathrm{U}(\{y_1, ..., y_k\}),\ y \in \{y_1, ..., y_k\}$$
$$P(Y = y_i) = \frac{1}{k},\ i = 1, ..., k$$
$$\mathrm{E}(Y) = \frac{k+1}{2},\ \mathrm{Var}(Y) = \frac{k^2-1}{12}$$

Binomialverteilung Erfolge in unabhängigen Versuchen

$$Y \sim \mathrm{Bin}(n, \pi) \text{ mit } n \in \mathbb{N}, \pi \in [0,1]\,,\ y \in \{0, ..., n\}$$
$$P(Y = y|\lambda) = \binom{n}{y}\pi^k(1-\pi)^{n-y}$$
$$\mathrm{E}(Y|\pi, n) = n\pi,\ \mathrm{Var}(Y|\pi, n) = n\pi(1-\pi)$$

Poissonverteilung Zählmodelle für seltene Ereignisse

Immer nur ein Ereignis pro Zeitpunkt, Eintreten der Ereignisse ist unabhängig von bisheriger Geschichte, mittlere Anzahl der Ereignisse pro Zeit ist konstant und proportional zur Länge des betrachteten Zeitintervalls.

$$Y \sim \mathrm{Po}(\lambda) \text{ mit } \lambda \in [0, +\infty]\,,\ y \in \mathbb{N}_0$$
$$P(Y = y|\lambda) = \frac{\lambda^y exp^{-\lambda}}{y!}$$
$$\mathrm{E}(Y|p) = \lambda,\ \mathrm{Var}(Y|p) = \lambda$$

Häufig wird die Varianz durchdas Poisson-Modell unterschätzt, es liegt Überdispersion vor.
Approximation der Binomialverteilung für kleine p

Geometrische Verteilung

$$Y \sim \mathrm{Geom}(\pi) \text{ mit } \pi \in [0,1]\,,\ y \in \mathbb{N}_0$$
$$P(Y = y|\pi) = \pi(1-\pi)^{y-1}$$
$$\mathrm{E}(Y|\pi) = \frac{1}{\pi},\ \mathrm{Var}(Y|\pi) = \frac{1-\pi}{\pi^2}$$

Negative Binomialverteilung

$$Y \sim \mathrm{NegBin}(\alpha, \beta) \text{ mit } \alpha, \beta \geq 0,\ y \in \mathbb{N}_0$$
$$P(Y = y|\alpha, \beta) = \binom{\alpha+y-1}{\alpha-1}\left(\frac{\beta}{\beta-1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y}$$
$$\mathrm{E}(Y|\alpha, \beta) = \frac{\alpha}{\beta},\ \mathrm{Var}(Y|\alpha, \beta) = \frac{\alpha}{\beta^2}(\beta+1)$$

2.4.2 Continuous Distributions

Stetige Gleichverteilung

$$Y \sim \mathrm{U}(a, b) \text{ mit } \alpha, \beta \in \mathbb{R}, a \leq b,\ y \in [a, b]$$
$$p(y|a, b) = \frac{1}{b-a}$$
$$\mathrm{E}(Y|a, b) = \frac{a+b}{2},\ \mathrm{Var}(Y|a, b) = \frac{(b-a)^2}{12}$$

Univariate Normalverteilung symmetrisch mit μ und σ^2

$$Y \sim \mathrm{N}(\mu, \sigma^2) \text{ mit } \mu \in \mathbb{R}, \sigma^2 > 0,\ y \in \mathbb{R}$$
$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$
$$\mathrm{E}(Y|\mu, \sigma^2) = \mu,\ \mathrm{Var}(Y|\mu, \sigma^2) = \sigma^2$$

Multivariate Normalverteilung symmetrisch mit μ und Σ

$$Y \sim \mathrm{N}(\mu, \Sigma) \text{ mit } \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d} s.p.d.,\ y \in \mathbb{R}^d$$
$$p(y|\mu, \Sigma) = (2\pi)^{-\frac{d}{2}}\det(\Sigma)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right)$$
$$\mathrm{E}(Y|\mu, \Sigma) = \mu,\ \mathrm{Var}(Y|\mu, \Sigma) = \Sigma$$

Log-Normalverteilung

$$Y \sim \mathrm{LogN}(\mu, \sigma^2) \text{ mit } \mu \in \mathbb{R}, \sigma^2 > 0,\ y > 0$$
$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right)$$
$$\mathrm{E}(Y|\mu, \sigma^2) = \exp(\mu + \frac{\sigma^2}{2}),$$
$$\mathrm{Var}(Y|\mu, \sigma^2) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

Zusammenhang: $\log(Y) \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Y \sim \mathrm{LogN}(\mu, \sigma^2)$

Nichtzentrale Studentverteilung statistische Tests für μ mit unbekannter (geschätzter) Varianz und ν Freiheitsgraden

$$Y \sim \mathrm{t}_\nu(\mu, \sigma) \text{ mit } \mu \in \mathbb{R}, \sigma^2, \nu > 0,\ y \in \mathbb{R}$$
$$p(y|\mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\Gamma(\sqrt{\nu\pi}\sigma)}\left(1 + \frac{(y-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$
$$\mathrm{E}(Y|\mu, \sigma^2, \nu) = \mu \text{ f'ur } \nu > 1,$$
$$\mathrm{Var}(Y|\mu, \sigma^2, \nu) = \sigma^2\frac{\nu}{\nu-2} \text{ f'ur } \nu > 2$$

Zusammenhang: $Y|\theta \sim \mathrm{N}(\mu, \frac{\sigma^2}{\theta}),\ \theta \sim \mathrm{Ga}(\frac{\nu}{2}, \frac{\nu}{2}) \Rightarrow Y \sim \mathrm{t}_\nu(\mu, \sigma)$

Betaverteilung

$$
\begin{aligned}
&Y \sim \mathrm{Be}(a,b) \text{ mit } a,b>0,\ y\in[0,1]\\
&p(y|a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}\\
&\mathrm{E}(Y|a,b)=\frac{a}{a+b},\\
&\mathrm{Var}(Y|a,b)=\frac{ab}{(a+b)^2(a+b+1)},\\
&\mathrm{mod}(Y|a,b)=\frac{a-1}{a+b-2} \text{ f'ur } a,b>1
\end{aligned}
$$

Gammaverteilung

$$
\begin{aligned}
&Y \sim \mathrm{Ga}(a,b) \text{ mit } a,b>0,\ y>0\\
&p(y|a,b)=\frac{b^a}{\Gamma(a)}y^{a-1}\exp(-by)\\
&\mathrm{E}(Y|a,b)=\frac{a}{b},\\
&\mathrm{Var}(Y|a,b)=\frac{a}{b^a},\\
&\mathrm{mod}(Y|a,b)=\frac{a-1}{b} \text{ f'ur } a\geq 1
\end{aligned}
$$

Invers-Gammaverteilung

$$
\begin{aligned}
&Y \sim \mathrm{IG}(a,b) \text{ mit } a,b>0,\ y>0\\
&p(y|a,b)=\frac{b^a}{\Gamma(a)}y^{-a-1}\exp(-\frac{b}{y})\\
&\mathrm{E}(Y|a,b)=\frac{b}{a-1} \text{ f'ur } a>1,\\
&\mathrm{Var}(Y|a,b)=\frac{b^2}{(a-1)^2(a-2)} \text{ f'ur } a\geq 2,\\
&\mathrm{mod}(Y|a,b)=\frac{b}{a+1}
\end{aligned}
$$

Zusammenhang: $Y^{-1}\sim \mathrm{Ga}(a,b) \Leftrightarrow Y\sim \mathrm{IG}(a,b)$

Exponentialverteilung Zeit zwischen Poisson-Ereignissen

$$
\begin{aligned}
&Y\sim \mathrm{Exp}(\lambda) \text{ mit } \lambda>0,\ y\geq 0\\
&p(y|\lambda)=\lambda\exp(-\lambda y)\\
&\mathrm{E}(Y|\lambda)=\frac{1}{\lambda},\ \mathrm{Var}(Y|\lambda)=\frac{1}{\lambda^2}
\end{aligned}
$$

Chi-Quadrat-Verteilung quadrierte standardnormalverteilte Zufallsvariablen mit ν Freiheitsgraden

$$
\begin{aligned}
&Y\sim \chi^2(\nu) \text{ mit } \nu>0,,\ y\in\mathbb{R}\\
&p(y|\nu)=\frac{y^{\frac{\nu}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\\
&\mathrm{E}(Y|\nu)=\nu,\ \mathrm{Var}(Y|\nu)=2\nu
\end{aligned}
$$

2.4.3 Exponential Family

Definition
Zur Exponentialfamilie gehören alle Verteilungen, deren Dichte wie folgt geschrieben werden kann:

$$f_Y(y,\theta)=\exp^{t^T(y)\theta-\kappa(\theta)}h(y)$$

mit $h(y)\geq 0$, $t(y)$ Vektor der kanonischen Statistiken, θ Parametervektor und $\kappa(\theta)$ Normalisationskonstante.

Normalisierungskonstante

$$
\begin{aligned}
1&=\int \exp^{t^T(y)\theta}h(y)dy\exp^{-\kappa(\theta)}\\
\Leftrightarrow \kappa(\theta)&=\log\int\exp^{t^T(y)\theta}h(y)dy
\end{aligned}
$$

$\kappa(\theta)$ ist die kumulanterzeugende Funktion, somit $\frac{\partial\kappa(\theta)}{\partial\theta}=\mathrm{E}(t(Y))$ und $\frac{\partial^2\kappa(\theta)}{\partial\theta^2}=\mathrm{Var}(t(Y))$

Mitglieder

- **Poissonverteilung**
- **Geometrische Verteilung**
- **Exponentialverteilung**
- **Normalverteilung** $t(y)=\left(-\frac{y^2}{2},y\right)^T$, $\theta=\left(\frac{1}{\sigma^2},\frac{\mu}{\sigma^2}\right)^T$, $h(y)=\frac{1}{\sqrt{2\pi}}$, $\kappa(\theta)=\frac{1}{2}\left(-\log\frac{1}{\sigma^2}+\frac{\mu^2}{\sigma^2}\right)$
- **Gammaverteilung**
- **Chi-Quadrat-Verteilung**
- **Betaverteilung**

2.5 Limit Theorems

Gesetz der großen Zahlen

Zentraler Grenzwertsatz

$$Z_n\xrightarrow{d}\mathrm{N}(0,\sigma^2)$$

mit $Z_n=\sum_{i=1}^n\frac{Y_i}{\sqrt{n}}$ und Y_i i.i.d. mit $\mu=0$ und Varianz σ^2

Proof:
Für eine normalverteilte Zufallsvariable $Z \sim \mathrm{N}(\mu, \sigma^2)$ gilt $K_Z(t) = \mu t + \frac{1}{2}\sigma^2 t^2$. Die ersten beiden Ableitungen $\left.\frac{\partial^k K_Z(t)}{\partial t^k}\right|_{t=0}$ entsprechen μ und σ. Alle anderen Momente sind null.
Für $Z_n = (Y_1 + Y_2 + ... + Y_n)/\sqrt{n}$ gilt:

$$\begin{aligned} M_{Z_n}(t) &= \mathrm{E}\left(e^{t(Y_1+Y_2+...+Y_n)/\sqrt{n}}\right) \\ &= \mathrm{E}\left(e^{tY_1/\sqrt{n}} \cdot e^{tY_2/\sqrt{n}} \cdot ... \cdot e^{tY_n/\sqrt{n}}\right) \\ &= \mathrm{E}\left(e^{tY_1/\sqrt{n}}\right) \mathrm{E}\left(e^{tY_2/\sqrt{n}}\right) ... \mathrm{E}\left(e^{tY_n/\sqrt{n}}\right) \\ &= M_Y^n(t/\sqrt{n}) \end{aligned}$$

Analog gilt: $K_{Z_n}(t) = nK_Y(t/\sqrt{n})$.

$$\begin{aligned} \left.\frac{\partial K_{Z_n}(t)}{\partial t}\right|_{t=0} &= \frac{n}{\sqrt{n}} \left.\frac{\partial K_Y(t)}{\partial t}\right|_{t=0} = \sqrt{n}\mu \\ \left.\frac{\partial^2 K_{Z_n}(t)}{\partial t^2}\right|_{t=0} &= \frac{n}{n} \left.\frac{\partial^2 K_Y(t)}{\partial t^2}\right|_{t=0} = \sigma^2 \end{aligned}$$

Mithilfe der Taylorreihe können wir $K_{Z_n}(t) = 0 + \sqrt{n}\mu t + \frac{1}{2}\sigma^2 t^2 + ...$ schreiben, wobei die Terme in ... alle für $n \to \infty$ gegen 0 gehen.
Damit gilt $K_{Z_n}(t) \overset{n\to\infty}{\longrightarrow} K_Z(t)$ mit $Z \sim \mathrm{N}(\sqrt{n}\mu, \sigma^2)$.

3 Inference

3.1 Method of Moments

Die theoretischen Momente werden durch die empirischen geschätzt:

$$\mathrm{E}_{\hat{\theta}_{MM}}(Y^k) = m_k(y_1, ..., y_n)$$

Für die Exponentialfamilie gilt: $\hat{\theta}_{MM} = \hat{\theta}_{ML}$

3.2 Loss Functions

Verlust

$$\mathcal{L} : \mathcal{T} \times \Theta \to \mathbb{R}^+$$

mit Parameterraum $\Theta \subset \mathbb{R}$, $t \in \mathcal{T}$ mit $t : \mathbb{R}^n \to \mathbb{R}$ eine Statistik, die den Parameter θ schätzt. Es gilt: $\mathcal{L}(\theta, \theta) = 0$

- **absoluter Verlust (L1)**: $\mathcal{L}(t, \theta) = |t - \theta|$
- **quadratischer Verlust (L2)**: $\mathcal{L}(t, \theta) = (t - \theta)^2$

Da θ unbekannt ist, ist der Verlust eine theoretische Größe. Zudem ist er die Realisation einer Zufallsvariable, da er von einer konkreten Stichprobe abhängt.

Risiko

$$\begin{aligned} R(t(.), \theta) &= \mathrm{E}_\theta\left(\mathcal{L}(t(Y_1, ..., Y_n), \theta)\right) \\ &= \int_{-\infty}^{\infty} \mathcal{L}(t(Y_1, ..., Y_n), \theta) \prod_{i=1}^{n} f(y_i; \theta) dy_i \end{aligned}$$

Minimax-Regel
Das Risiko beruht immer noch auf dem wahren Parameter θ. Vorsichtige Schätzung: Wähle θ so, dass das Risiko maximal wird, und danach $t(.)$ so, dass das Risiko minimiert wird:

$$\hat{\theta}_{minimax} = \arg\min_{t(.)} \left(\max_{\theta\in\Theta} R(t(.); \theta)\right)$$

Es wird der Worst Case minimiert.

Mean Squared Error (MSE)

$$\begin{aligned} MSE(t(.), \theta) &= \mathrm{E}_\theta\left(\{t(Y) - \theta\}^2\right) \\ &= \mathrm{Var}_\theta\left(t(Y_1, ..., Y_n)\right) + Bias^2((t(.); \theta) \end{aligned}$$

mit $Bias(t(.); \theta) = \mathrm{E}_\theta\left(t(Y_1, ..., Y_n)\right) - \theta$

Proof:
Sei $\mathcal{L}(t, \theta) = (t - \theta)^2$

$$\begin{aligned} R(t(.), \theta) &= \mathrm{E}_\theta(\{t(Y) - \theta\}^2) \\ &= \mathrm{E}_\theta(\{t(Y) - \mathrm{E}_\theta(t(Y)) + \mathrm{E}_\theta(t(Y)) - \theta\}^2) \\ &= \mathrm{E}_\theta(\{t(Y) - \mathrm{E}_\theta(t(Y))\}^2) + \mathrm{E}_\theta(\{\mathrm{E}_\theta(t(Y)) - \theta\}^2) \\ &\quad + 2\mathrm{E}_\theta(\{t(Y) - \mathrm{E}_\theta\,(t(Y))\}\{\mathrm{E}_\theta(t(Y)) - \theta\}) \\ &= \mathrm{Var}_\theta(t(Y_1, ..., Y_n)) + Bias^2((t(.); \theta) + 0 \end{aligned}$$

Cramér-Rao-Ungleichung

$$MSE(\hat{\theta},\theta) \geq Bias^2(\hat{\theta},\theta) + \frac{\left(1+\frac{\partial Bias(\hat{\theta},\theta)}{partial\theta}\right)^2}{I(\theta)}$$

Proof:
Für ungebiaste Schätzer: $\theta = \mathrm{E}_\theta(\hat{\theta}) = \int t(y)f(y;\theta)dy$

$$\begin{aligned}
1 &= \int t(y)\frac{\partial f(y;\theta)}{\partial\theta}dy \\
&= \int t(y)\frac{\partial \log f(y;\theta)}{\partial\theta}f(y;\theta)dy \\
&= \int t(y)s(y;\theta)f(y;\theta)dy \\
&= \int (t(y)-\theta)\left(s(\theta;y)-0\right)f(y;\theta)dy && \text{1. Bartlett-Gleichung } \mathrm{E}_\theta\left(s(\theta;y)\right)=0 \\
&= \mathrm{Cov}_\theta\left(t(Y);s(\theta;Y)\right) \\
&\geq \sqrt{\mathrm{Var}_\theta(t(Y))}\sqrt{\mathrm{Var}_\theta(s(\theta;Y))} && \text{Cauchy-Schwarz} \\
&= \sqrt{MSE(t(Y);\theta)}\sqrt{I(\theta)}
\end{aligned}$$

Kullback-Leibler-Divergenz Vergleich von Verteilungen

$$KL(t,\theta) = \int_{-\infty}^{\infty} \log\frac{f(\tilde{y};\theta)}{f(\tilde{y};t)}f(\tilde{y};\theta)d\tilde{y}$$

Die KL-Divergenz ist keine Distanz, da sie nicht symmetrisch ist. Sie ist 0 für $t=\theta$ und größer/gleich 0 sonst.

Proof:
Folgt aus $\log(x) \leq x-1 \forall x \geq 0$, mit Gleichheit für $x=1$.

$R_{KL}(t(.),\theta)$ wird durch den MSE approximiert.

Proof:

$$\begin{aligned}
&R_{KL}(t(.),\theta) = \\
&= \int_{-\infty}^{\infty}\mathcal{L}_{KL}(t(Y_1,...,Y_n),\theta)\prod_{i=1}^n f(y_i;\theta)dy_i \\
&= \int\int \log\frac{f(\tilde{y};\theta)}{f(\tilde{y};t)}f(\tilde{y};\theta)d\tilde{y}\prod_{i=1}^n f(y_i;\theta)dy_i \\
&= \int\int\left(\log f(\tilde{y};\theta)-\log f(\tilde{y};t)\right)f(\tilde{y};\theta)d\tilde{y} - \prod_{i=1}^n f(y_i;\theta)dy_i \\
&\approx -\int\underbrace{\left(\int\frac{\partial\log f(\tilde{y};\theta)}{\partial\theta}f(\tilde{y};\theta)d\tilde{y}\right)}_{0}(t-\theta)\prod_{i=1}^n f(y_i;\theta)dy_i \\
&+\frac{1}{2}\int\underbrace{\left(-\int\frac{\partial^2\log f(\tilde{y};\theta)}{\partial\theta^2}f(\tilde{y};\theta)d\tilde{y}\right)}_{I(\theta)}(t-\theta)^2\prod_{i=1}^n f(y_i;\theta)dy_i
\end{aligned}$$

Wobei der letzte Schritt durch die Taylorreihe approximiert wurde: $\log f(\tilde{y},t) \approx \log f(\tilde{y},\theta) + \frac{\partial\log f(\tilde{y},\theta)}{\partial\theta}(t-\theta) + \frac{1}{2}\frac{\partial^2\log f(\tilde{y},\theta)}{\partial\theta^2}(t-\theta)^2$

3.3 Maximum Likelihood (ML)

Voraussetzungen

- $Y_i \sim f(y;\theta)$ *i.i.d.*
- $\theta \in \mathbb{R}^p$
- $f(.;\theta)$ Fisher-regulär:
  - $\{y : f(y;\theta > 0)\}$ unabhängig von θ
  - Möglicher Parameterraum Θ ist offen
  - $f(y;\theta)$ zweimal differenzierbar
  - $\int\frac{\partial}{\partial\theta}f(y;\theta)dy = \frac{\partial}{\partial\theta}\int f(y;\theta)dy$

Zentrale Funktionen

- **Likelihood** $L(\theta;y_1,...,y_n)$: $\prod_{i=1}^n f(y_i;\theta)$
- **log-Likelihood** $l(\theta;y_1,..y_n)$: $\log L(\theta;y_1,...,y_n) = \sum_{i=1}^n \log f(y_i;\theta)$
- **Score** $s(\theta;y_1,...,y_n)$: $\frac{\partial l(\theta;y_1,..y_n)}{\partial\theta}$
- **Fisher-Information** $I(\theta)$: $-\mathrm{E}_\theta\left(\frac{\partial s(\theta;Y)}{\partial\theta}\right)$
- **beobachtete Fisher-Information** $I_{obs}(\theta)$: $-\mathrm{E}_\theta\left(\frac{\partial s(\theta;y)}{\partial\theta}\right)$

Eigenschaften der Score-Funktion
erste Bartlett-Gleichung:

$$\mathrm{E}\left(s(\theta;Y)\right) = 0$$

Proof:

$$\begin{aligned}
1 &= \int f(y;\theta)dy \\
0 = \frac{\partial 1}{\partial\theta} &= \int\frac{\partial f(y;\theta)}{\partial\theta}dy = \int\frac{\partial f(y;\theta)/\partial\theta}{f(y;\theta)}f(y;\theta)dy \\
&= \int\frac{\partial}{\partial\theta}\log f(y;\theta)f(y;\theta)dy = \int s(\theta;y)f(y;\theta)dy
\end{aligned}$$

zweite Bartlett-Gleichung:

$$\mathrm{Var}_\theta\left(s(Y;\theta)\right) = \mathrm{E}_\theta\left(-\frac{\partial^2 logf(Y;\theta)}{\partial\theta^2}\right) = I(\theta)$$

Proof:

$$\begin{aligned}
0 = \frac{\partial 0}{\partial\theta} &= \frac{\partial}{\partial\theta}\int\frac{\partial}{\partial\theta}\log f(y;\theta)f(y;\theta)dy && \text{siehe oben} \\
&= \int\left(\frac{\partial^2}{\partial\theta^2}\log f(y;\theta)\right)f(y;\theta)dy \\
&\quad+\int\frac{\partial\log f(y;\theta)}{\partial\theta}\frac{\partial f(y;\theta)}{\partial\theta}dy \\
&= \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(Y;\theta)\right) \\
&\quad+\int\frac{\partial\log f(y;\theta)}{\partial\theta}\frac{\partial\log f(y;\theta)}{\partial\theta}f(y;\theta)dy
\end{aligned}$$

$$\Leftrightarrow \mathrm{E}_\theta\left(s(\theta;Y)s(\theta;Y)\right) = \mathrm{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(Y;\theta)\right)$$

Bartletts zweite Gleichung gilt dann, weil $\mathrm{E}\left(s(\theta;Y)\right)=0$

ML-Schätzer

$$\hat{\theta}_{ML} = \arg\max\, l(\theta; y_1, ...y_n)$$

für Fisher-reguläre Verteilungen: $\hat{\theta}_{ML}$ hat asymptotisch die kleinstmögliche Varianz, gegeben durch die Cramér-Rao-Ungleichung, $s\left(\hat{\theta}_{ML}; y_1, ..., y_n\right) = 0$

$\hat{\theta} \overset{a}{\sim} \mathrm{N}\left(\theta, I^{-1}(\theta)\right)$

Der ML-Schätzer ist invariant: $\hat{\gamma} = g(\hat{\theta})$ wenn $\gamma = g(\theta)$.

Proof:
$\gamma = g(\theta) \Leftrightarrow \theta = g^{-1}(\gamma)$
Für die Loglikelihood von γ an der Stelle $\hat{\theta}$ gilt:

$$\frac{\partial l(g^{-1}(\hat{\gamma}))}{\partial \gamma} = \frac{\partial g^{-1}(\gamma)}{\partial \gamma} \underbrace{\frac{\partial l(\hat{\theta})}{\partial \theta}}_{=0} = 0$$

Die Fisher-Information ist dann $\frac{\partial \theta}{\partial \gamma} I(\theta) \frac{\partial \theta}{\partial \gamma}$

Proof:

$$I_\gamma(\gamma) = -\mathrm{E}\left(\frac{\partial^2 l(g^{-1}(\hat{\gamma}))}{\partial \gamma^2}\right) = -\mathrm{E}\left(\frac{\partial}{\partial \gamma}\left(\frac{\partial g^{-1}(\gamma)}{\partial \gamma}\frac{\partial l(\theta)}{\partial \theta}\right)\right)$$
$$= -\mathrm{E}\left(\underbrace{\frac{\partial^2 g^{-1}(\gamma)}{\partial \gamma}\frac{\partial l(\theta)}{\partial \theta}}_{\text{Erwartungswert } 0} + \frac{\partial g^{-1}(\gamma)}{\partial \gamma}\frac{\partial^2 l(\theta)}{\partial \theta^2}\frac{\partial g^{-1}(\gamma)}{\partial \gamma}\right)$$
$$= \frac{\partial g^{-1}(\gamma)}{\partial \gamma} I(\theta) \frac{\partial g^{-1}(\gamma)}{\partial \gamma} = \frac{\partial \theta}{\partial \gamma} I(\theta) \frac{\partial \theta}{\partial \gamma}$$

Delta-Regel: $\gamma \overset{a}{\sim} \mathrm{N}(\hat{\gamma}, \frac{\partial \theta}{\partial \gamma} I^{-1}(\theta) \frac{\partial \theta}{\partial \gamma}$

Numerical computation of the ML estimate Fisher-Scoring as statistical version of the Newton-Raphson procedure

1. Initialize $\theta_{(0)}$
2. Repeat: $\theta_{(t+1)} := \theta_{(t)} + I^{-1}(\theta_{(t)}) s(\theta_{(t)}; y)$
3. Stop if $\|\theta_{(t+1)} - \theta_{(t)}\| < \tau$; return $\hat{\theta}_{ML} = \theta_{(t+1)}$

Proof:

$$0 = s(\hat{\theta}_{ML}; y) \underset{Series}{\overset{Taylor}{\approx}} s(\theta; y) + \frac{\partial s(\theta; y)}{\partial \theta}(\hat{\theta}_{ML} - \theta) \Leftrightarrow$$
$$\hat{\theta}_{ML} \approx \theta - \left(\frac{\partial s(\theta; y)}{\partial \theta}\right)^{-1} s(\theta; y) \approx \theta - I^{-1}(\theta) s(\theta; y)$$

As $\frac{\partial s(\theta; y)}{\partial \theta}$ is often complicated, its expectation $I(\theta)$ is used.

The second part in 2 can be weighted with a step size δ or $\delta(t)$ $\in (0, 1)$, e. g. to ensure convergence.
If $I(\theta)$ can't be analytically derived, simulation from $f(y; \theta_{(t)})$ can be used. For the exponential family, step 2 then changes to $\theta_{(t+1)} := \theta_{(t)} + \hat{\mathrm{Var}}_{\theta_{(t)}}(t(Y))^{-1} \mathrm{E}_{\theta_{(t)}}(t(Y))$ as the ML estimate is the expectation.

Log Likelihood Ratio

$$lr(\theta, \hat{\theta}) := l(\hat{\theta}) - l(\theta) = \log \frac{L(\hat{\theta})}{L(\theta)}$$

with $2 \cdot lr(\theta, \hat{\theta}) \overset{a}{\sim} \chi^2_1$

Proof:

$$l(\theta) \underset{Series}{\overset{Taylor}{\approx}} l(\hat{\theta}) + \underbrace{\frac{\partial l(\hat{\theta})}{\partial \theta}}_{=0}(\theta - \hat{\theta}) + \frac{1}{2} \underbrace{\frac{\partial^2 l(\hat{\theta})}{\partial \theta^2}}_{\approx I^{-1}(\theta) s(\theta; Y)} (\underbrace{\theta - \hat{\theta}}_{\approx -I(\theta)})^2$$
$$\approx l(\hat{\theta}) - \frac{1}{2}\frac{s^2(\theta, Y)}{I(\theta)}$$

$s(\theta, Y)$ is asymptotically normal.

If $\theta \in \mathbb{R}^p$ the corresponding distribution is χ^2_p.

3.4 Sufficiency und Consistency

Suffizienz
Eine Statistik $t(y_1, ..., y_n)$ ist suffizient für θ, wenn die bedingte Verteilung $f(y_1, ..., y_n | t_0 = t(y_1, ..., y_n); \theta)$ unabhängig von θ ist.

Neyman-Kriterium:

$$t(Y_1, ..., Y_n) \text{ suffizient } \Leftrightarrow f(y; \theta) = h(y) g\left(t(y); \theta\right)$$

Proof:
"$\Rightarrow$":

$$f(y; \theta) = \underbrace{f(y | t = t(y); \theta)}_{h(y)} \underbrace{f_t(t | y; \theta)}_{g(t(y); \theta)}$$

"$\Leftarrow$":

$$f_t(t; \theta) = \int_{t = t(y)} f(y; \theta) dy = \int_{t = t(y)} h(y) g(t; \theta) dy$$

Damit:

$$f\left(y | t = t(y); \theta\right) = \frac{f(y, t = t(y); \theta)}{f_t(t, \theta)} = \begin{cases} \frac{h(y) g(t; \theta)}{g(t; \theta)} & t = t(y) \\ 0 & \text{sonst} \end{cases}$$

Minimalsuffizienz:
$t(.)$ ist suffizient und $\forall\, \tilde{t}(.) \,\exists\, h(.)$ s.t. $t(y) = h(\tilde{t}(y))$

(schwache) Konsistenz

$$MSE(\hat{\theta}, \theta) \overset{n \to \infty}{\longrightarrow} 0 \Rightarrow \hat{\theta} \text{ konsistent}$$

3.5 Confidence Intervals

Definition

$$[t_l(Y), t_r(Y)] \text{ Konfidenzintervall}$$
$$\Leftrightarrow$$
$$P_\theta\left((t_l(Y) \leq \theta \leq t_r(Y)\right) \geq 1-\alpha$$

mit $1-\alpha$ Konfidenzlevel und α Signifikanzlevel

Pivotale Statistik

$$g(Y;\theta) \text{ pivotal}$$
$$\Leftrightarrow$$
$$\text{Verteilung von } g(Y;\theta) \text{ unabhängig von } \theta$$

Approximativ pivotale Statistik

$$g(\hat{\theta};\theta) = \frac{\hat{\theta}-\theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \overset{\alpha}{\sim} \mathrm{N}(0,1)$$

mit $\hat{\theta} = t(Y) \overset{\alpha}{\sim} \mathrm{N}(\theta, \mathrm{Var}(\hat{\theta})$

$$KI = \left[\hat{\theta} - z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{Var}(\hat{\theta})}, \hat{\theta} + z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{Var}(\hat{\theta})}\right]$$

Proof:
$1-\alpha \approx P\left(z_{\frac{\alpha}{2}} \leq \frac{\hat{\theta}-\theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \leq z_{1-\frac{\alpha}{2}}\right)$

4 Hypothesis Testing

Exakte binomiale Konfidenzintervalle

4.1 Tests for One Sample

4.1.1 Normal Distribution

5 Regression

μ gesucht, σ^2 bekannt (Einfacher Gauß-Test)

5.1 Assumptions

5.2 Procedure

5.2.1 Ordinary Least Squares (OLS)

KQ-Schätzer (Einfachregression)

$$\hat{\beta}_1 = \frac{Cov(x,y)}{Var(x)} = \frac{S_{xy}}{S_{xx}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \cdot \sqrt{\frac{S_{yy}}{S_{xx}}} = r\sqrt{\frac{S_{yy}}{S_{xx}}}$$

Proof:
$Cov(x,y) = Cov(x, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}) = \hat{\beta}_1 Var(x)$
$\iff \hat{\beta}_1 = \frac{Cov(x,y)}{Var(x)}$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Proof:
$E[y] = E\left[\hat{\beta}_0 + \hat{\beta}_1 x + \hat{e}\right] \iff \hat{\beta}_0 = E[y] - \hat{\beta}_1 E[x]$

5.3 Model

5.3.1 Simple Linear Regression

Theoretisches Modell

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Empirisches Modell

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

Eigenschaften der Regressionsgeraden

$$\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) \\
\hat{e}_i &= y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\
&= y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{x})) \\
\sum_{i=1}^{n} \hat{e}_i &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y} - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x}) \\
&= n\bar{y} - n\bar{y} - \hat{\beta}_1 (n\bar{x} - n\bar{x}) = 0 \\
\bar{\hat{y}} &= \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i = \frac{1}{n} (n\bar{y} + \hat{\beta}_1 (n\bar{x} - n\bar{x})) = \bar{y}
\end{aligned}$$

5.3.2 Multivariate Linear Regression

5.4 Analysis of Variances (ANOVA)

$$SS_{Total} = SS_{Explained} + SS_{Residual}$$

mit

$$\begin{aligned}
SS_{Total} &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
SS_{Explained} &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\
SS_{Residual} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = S_{yy} - \hat{\beta}^2 S_{xx}
\end{aligned}$$

5.5 Goodness of Fit

5.5.1 Bestimmtheitsmaß

$$R^2 = \frac{SS_{Explained}}{SS_{Total}} = 1 - \frac{SS_{Residual}}{SS_{Total}} = r^2$$

Range: $0 \leq R^2 \leq 1$

6 Classification

6.1 Diskriminant Analysis (Bayes)

7 Cluster Analysis

8 Bayesian Statistics

8.1 Basics

Bayes-Formel

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{für } P(A), P(B) > 0$$

oder allgemeiner:

$$\begin{aligned}
f(\theta|X) &= \frac{f(X|\theta) \cdot f(\theta)}{\int f(X|\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta}} \\
&= C \cdot f(X|\theta) \cdot f(\theta) \quad \text{wähle C so, dass } \int f(\theta|X) = 1 \\
&\propto f(X|\theta) \cdot f(\theta)
\end{aligned}$$

Punktschätzer

Kredibilitätsintervall

Sensitivitätsanalyse

Prädiktive Posteriori

$$f(x_Z|\mathbf{x}) = \int f(x_Z, \lambda|\mathbf{x})d\lambda = \int f(x_Z|\lambda)p(\lambda|\mathbf{x})$$

Uninformative Priori

$f(\theta) = const.$ für $\theta > 0$, damit: $f(\theta|X) = C \cdot f(X|\theta)$

(Da $\int f(\theta) = 1$ so nicht möglich, ist das eigentlich keine Dichte)

Konjugierte Priori

Wenn die Priori- und die Posteriori-Verteilung denselben Typ hat für eine gegebene Likelihoodfunktion, so nennt man sie konjugiert.

Binomial-Beta-Modell:

- Priori $\sim Be(\alpha, \beta)$
- $X \sim Binom(n, p, k)$
- Posteriori $\sim Be(\alpha + k, \beta + n - k)$

8.2 Markov Chain / Monte Carlo