

Statistik Formelsammlung

Katharina Ring

6. Mai 2019

Inhaltsverzeichnis

1 Deskriptive Statistik	3	3.2 Verlustfunktionen	9
1.1 Kenngrößen (Parameter): Stichprobe	3	3.3 Maximum Likelihood (ML)	10
1.1.1 Lagemaße	3	3.4 Suffizienz und Konsistenz	11
1.1.2 Streuungsmaße	3	3.5 Konfidenzintervalle	11
1.1.3 Konzentrationsmaße	3		
1.1.4 Gestaltsmaße	4	4 Hypothesentests	12
1.1.5 Zusammenhangsmaße	4	4.1 Tests für Einstichprobenprobleme	12
1.2 Tabellen	5	4.1.1 Normalverteilung	12
1.3 Diagramme	5		
1.3.1 Histogramm	5	5 Regression	12
1.3.2 QQ-Plot	5	5.1 Annahmen	12
1.3.3 Plot der Realisationen	5	5.2 Verfahren	12
1.3.4 Scatterplot	5	5.2.1 Kleinste Quadrate (OLS)	12
		5.3 Modell	12
2 Wahrscheinlichkeit	5	5.3.1 lineare Einfachregression	12
2.1 Kombinatorik	5	5.3.2 Multivariate lineare Regression	12
2.2 Wahrscheinlichkeitsrechnung	5	5.4 ANOVA (Streuungszerlegung)	13
2.3 Zufallsvariablen	6	5.5 Gütemaße	13
2.4 Zufallsvektoren	6	5.5.1 Bestimmtheitsmaß	13
2.5 Verteilungen	7		
2.5.1 Diskrete Verteilungen	7	6 Klassifikation	13
2.5.2 Stetige Verteilungen	7	6.1 Diskriminanzanalyse (Bayes)	13
2.5.3 Exponentialfamilie	8		
2.6 Grenzwertsätze	9	7 Clusteranalyse	13
3 Inferenz	9	8 Bayessche Statistik	13
3.1 Methode der Momente	9	8.1 Grundlagen	13
		8.2 Markov Chain / Monte Carlo	13

1 Deskriptive Statistik

1.1 Kenngrößen (Parameter): Stichprobe

1.1.1 Lagemaße

Modus Häufigster Wert von x_i . Auch zwei oder mehr Modi sind möglich (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

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

mit

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

Minimum/Maximum

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

1.1.2 Streuungsmaße

Spannweite

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

Quartilsabstand

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

(Empirische) Varianz

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

Schätzer für das zweite zentrierte Moment, inkl.

Varianzverschiebungssatz

Rechenregeln:

$$\star \operatorname{Var}(aX + b) = a^2 \cdot \operatorname{Var}(X)$$

1.1.3 Konzentrationsmaße

Gini-Koeffizient

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

mit

Arithmetisches Mittel

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

Schätzer für den Erwartungswert $\mu = E[X]$
(erstes Verteilungsmoment)

Rechenregeln:

$$\star E(a + b \cdot X) = a + b \cdot E(X)$$

$$\star E(X \pm Y) = E(X) \pm E(Y)$$

Geometrisches Mittel

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

Für Wachstumsfaktoren: $\bar{x}_G = \sqrt[n]{\frac{B_n}{B_0}}$

Harmonisches Mittel

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

$$\star \operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$$

(Empirische) Standardabweichung

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

Variationskoeffizient

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

Mittlere absolute Abweichung

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

Schätzer für das erste absolute zentrierte Moment

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

Dies sind auch die Werte für die Lorenzkurve.

$$\text{Wertebereich: } 0 \leq G \leq \frac{n-1}{n}$$

Lorenz-Münzner-Koeffizient (G normiert)

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

Wertebereich: $0 \leq G^+ \leq 1$

1.1.4 Gestaltmaße

(Empirische) Schiefe

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

Schätzer für das dritte zentrierte Moment, normiert durch $(\sigma^2)^{\frac{2}{3}}$

(Empirische) Wölbung/Kurtosis

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

Schätzer für das vierte zentrierte Moment, normiert durch $(\sigma^2)^2$

Exzess

$$\gamma = k - 3$$

1.1.5 Zusammenhangsmaße

Für zwei nominale Variablen

χ^2 -Statistik

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

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

Phi-Koeffizient

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

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

Cramérs V

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

Wertebereich: $0 \leq V \leq 1$

Kontingenzkoeffizient C

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

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

Korrigierter Kontingenzkoeffizient C_{korr}

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

Wertebereich: $0 \leq C_{korr} \leq 1$

Odds-Ratio

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

Wertebereich: $0 \leq OR < \infty$

Für zwei ordinale Variablen

Gamma nach Goodman und Kruskal

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

$K = \sum_{i < m} \sum_{j < n} n_{ij}n_{mn}$ Anzahl konkordanter Paare

$D = \sum_{i < m} \sum_{j > n} n_{ij}n_{mn}$ Anzahl diskordanter Paare

Wertebereich: $-1 \leq \gamma \leq 1$

Kendalls τ_b

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

mit

$T_X = \sum_{i=m} \sum_{j < n} n_{ij}n_{mn}$ Anzahl Bindungen bzgl. X

$T_Y = \sum_{i < m} \sum_{j=n} n_{ij}n_{mn}$ Anzahl Bindungen bzgl. Y

Wertebereich: $-1 \leq \tau_b \leq 1$

Kendalls/Stuarts τ_c

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

Wertebereich: $-1 \leq \tau_c \leq 1$

Spearman's Rangkorrelationskoeffizient

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

oder

$$\rho = \frac{s_{rg_x} r_{g_y}}{\sqrt{s_{rg_x} r_{g_x} s_{rg_y} r_{g_y}}}$$

Entspricht ohne Bindungen:

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

mit

$$d_i = R(x_i) - R(y_i) \quad \text{Rangdifferenz}$$

Wertebereich: $-1 \leq \rho \leq 1$

Für zwei metrische Variablen

Korrelationskoeffizient nach Bravais-Pearson

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

mit

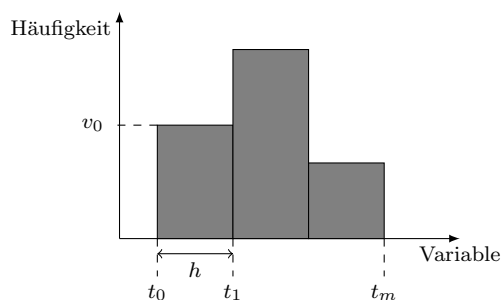
$$\begin{aligned} S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) & \text{bzw. } s_{xy} &= \frac{S_{xy}}{n} \\ S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 & \text{bzw. } s_{xx} &= \frac{S_{xx}}{n} \\ S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2 & \text{bzw. } s_{yy} &= \frac{S_{yy}}{n} \end{aligned}$$

Wertebereich: $-1 \leq r \leq 1$

1.2 Tabellen

1.3 Diagramme

1.3.1 Histogramm



Stichprobe: $X = \{x_1, x_2, \dots, x_n\}$

k -te Klasse: $B_k = [t_k, t_{k+1})$, $k = \{0, 1, \dots, m-1\}$

Anzahl Beobachtungen in der k -ten Klasse: v_k

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

Scotts Regel

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

Für annähernd normalverteilte Daten (min MSE)

1.3.2 QQ-Plot

1.3.3 Plot der Realisationen

1.3.4 Scatterplot

2 Wahrscheinlichkeit

2.1 Kombinatorik

	ohne Wiederholung	mit Wiederholung
Permutationen	$n!$	$\frac{n!}{n_1! \dots n_s!}$
Kombinationen: ohne Reihenfolge	$\binom{n}{m}$	$\binom{n+m-1}{m}$
mit Reihenfolge	$\binom{n}{m} m!$	n^m

Dabei gilt:

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

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

2.2 Wahrscheinlichkeitsrechnung

Laplace-Wahrscheinlichkeit

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

Axiome von Kolmogorov

mathematische Definition von Wahrscheinlichkeit

- (1) $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{A}$
- (2) $P(\Omega) = 1$
- (3) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
 $\forall A_i \in \mathcal{A}, i = 1, \dots, \infty$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$

Folgerungen:

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(B) = \sum_{i=1}^n P(B \cap A_i)$, für A_i, \dots, A_n vollständige Zerlegung von Ω in paarweise disjunkte Ereignisse

Mises' Wahrscheinlichkeitsbegriff

frequentistische Definition von Wahrscheinlichkeit

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

mit n Anzahl der Wiederholungen eines Zufallsexperiments und $n_A(n)$ Anzahl an Ereignissen A

Bedingte Wahrscheinlichkeit

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

Multiplikationssatz

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

2.3 Zufallsvariablen

Definition

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

Die Untermenge möglicher Werte von \mathbb{R} heißt Träger
Notation: Realisationen von Y werden als Kleinbuchstaben dargestellt. $Y = y$ bedeutet, dass Y die Realisation y angenommen hat.

Stetige und diskrete Zufallsvariablen

Ist der Träger überabzählbar unendlich, so heißt die Zufallsvariable *stetig*, sonst heißt sie *diskret*.

• Dichte $f(\cdot)$:

Für stetige Variablen: $P(Y \in [a, b]) = \int_a^b f_Y(y) dy$

Für diskrete Variablen lässt sich die Dichte (und andere Funktionen) wie die gleichen Funktionen für den stetigen Fall aufschreiben, wenn man

$\int_{-\infty}^y f_Y(\tilde{y}) d\tilde{y} := \sum_{k: k \leq y} P(Y = k)$ definiert. Diese Notation wird hier verwendet.

• Verteilungsfunktion $F(\cdot)$: $F_Y(y) = P(Y \leq y)$

Zusammenhang:

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

Ist der Träger endlich oder abzählbar unendlich, so heißt die Zufallsvariable *diskret*.

2.4 Zufallsvektoren

Dichte und Verteilungsfunktion

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

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

Marginale Dichte

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

Satz von der totalen Wahrscheinlichkeit

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

Satz von Bayes

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

Stochastische Unabhängigkeit

$$A, B \text{ unabhängig} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

$$X, Y \text{ unabhängig} \Leftrightarrow f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \quad \forall x, y$$

Momente

• Erwartungswert (1. Moment): $\mu = E(Y) = \int y f_Y(y) dy$

• Varianz (2. zentriertes Moment):

$$\sigma^2 = \text{Var}(Y) = E(\{Y - E(Y)\}^2) = \int (y - E(Y))^2 f(y) dy$$

$$\text{Varianzverschiebungssatz: } E(\{Y - \mu\}^2) = E(Y^2) - \mu^2$$

Beweis:

$$\begin{aligned} E(\{Y - \mu\}^2) &= E(Y^2 - 2Y\mu + \mu^2) = E(Y^2) - 2\mu^2 + \mu^2 = \\ &= E(Y^2) - \mu^2 \end{aligned}$$

• k. Moment: $E(Y^k) = \int y^k f_Y(y) dy$,

k. zentrales Moment: $E(\{Y - E(Y)\}^k)$

Momenterzeugende Funktion

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

$$\text{mit } \frac{\partial^k M_Y(t)}{\partial t^k} \Big|_{t=0} = E(Y^k)$$

kumulanterzeugende Funktion $K_Y(t) = \log M_Y(t)$

Eine Zufallsvariable ist durch ihre momenterzeugende Funktion eindeutig definiert und andersherum (solange die Momente und Kumulanten endlich sind).

Bedingte Dichte

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{f(y_1, \dots, y_2)}{f(y_2)} \quad \text{für } f(y_2) > 0$$

Iterierter Erwartungswert

$$E(Y) = E_X(E(Y|X))$$

Beweis:

$$E(Y) = \int y f(y) dy = \int \int y f(y|x) dy f_X(x) dx = E_X(E(Y|X))$$

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

Beweis:

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

2.5 Verteilungen

2.5.1 Diskrete Verteilungen

Diskrete Gleichverteilung

$$\begin{aligned}Y &\sim U(\{y_1, \dots, y_k\}), y \in \{y_1, \dots, y_k\} \\ P(Y = y_i) &= \frac{1}{k}, i = 1, \dots, k \\ E(Y) &= \frac{k+1}{2}, \text{Var}(Y) = \frac{k^2-1}{12}\end{aligned}$$

$$Y \sim \text{Po}(\lambda) \text{ mit } \lambda \in [0, +\infty], y \in \mathbb{N}_0$$

$$P(Y = y|\lambda) = \frac{\lambda^y \exp^{-\lambda}}{y!}$$

$$E(Y|p) = \lambda, \text{Var}(Y|p) = \lambda$$

Häufig wird die Varianz durch das Poisson-Modell unterschätzt, es liegt Überdispersion vor.

Approximation der Binomialverteilung für kleine p

Binomialverteilung Erfolge in unabhängigen Versuchen

$$\begin{aligned}Y &\sim \text{Bin}(n, \pi) \text{ mit } n \in \mathbb{N}, \pi \in [0, 1], y \in \{0, \dots, n\} \\ P(Y = y|\pi) &= \binom{n}{y} \pi^y (1-\pi)^{n-y} \\ E(Y|\pi, n) &= n\pi, \text{Var}(Y|\pi, n) = n\pi(1-\pi)\end{aligned}$$

Geometrische Verteilung

$$Y \sim \text{Geom}(\pi) \text{ mit } \pi \in [0, 1], y \in \mathbb{N}_0$$

$$P(Y = y|\pi) = \pi(1-\pi)^{y-1}$$

$$E(Y|\pi) = \frac{1}{\pi}, \text{Var}(Y|\pi) = \frac{1-\pi}{\pi^2}$$

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.

Negative Binomialverteilung

$$Y \sim \text{NegBin}(\alpha, \beta) \text{ mit } \alpha, \beta \geq 0, y \in \mathbb{N}_0$$

$$P(Y = y|\alpha, \beta) = \binom{\alpha+y-1}{\alpha-1} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^y$$

$$E(Y|\alpha, \beta) = \frac{\alpha}{\beta}, \text{Var}(Y|\alpha, \beta) = \frac{\alpha}{\beta^2}(\beta+1)$$

2.5.2 Stetige Verteilungen

Stetige Gleichverteilung

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

$$Y \sim N(\mu, \sigma^2) \text{ mit } \mu \in \mathbb{R}, \sigma^2 > 0, y \in \mathbb{R}$$

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

$$E(Y|\mu, \sigma^2) = \mu, \text{Var}(Y|\mu, \sigma^2) = \sigma^2$$

Univariate Normalverteilung symmetrisch mit μ und σ^2

Multivariate Normalverteilung symmetrisch mit μ und Σ

$$Y \sim N(\mu, \Sigma) \text{ mit } \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d} \text{ s.p.d., } y \in \mathbb{R}^d$$

$$p(y|\mu, \Sigma) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1}(y - \mu)\right)$$

$$E(Y|\mu, \Sigma) = \mu, \text{ Var}(Y|\mu, \Sigma) = \Sigma$$

Log-Normalverteilung

$$Y \sim \text{LogN}(\mu, \sigma^2) \text{ mit } \mu \in \mathbb{R}, \sigma^2 > 0, y > 0$$

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)$$

$$E(Y|\mu, \sigma^2) = \exp\left(\mu + \frac{\sigma^2}{2}\right),$$

$$\text{Var}(Y|\mu, \sigma^2) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

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

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

$$Y \sim t_\nu(\mu, \sigma) \text{ mit } \mu \in \mathbb{R}, \sigma^2, \nu > 0, y \in \mathbb{R}$$

$$p(y|\mu, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\sqrt{\nu\pi\sigma})} \left(1 + \frac{(y - \mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

$$E(Y|\mu, \sigma^2, \nu) = \mu \text{ für } \nu > 1,$$

$$\text{Var}(Y|\mu, \sigma^2, \nu) = \sigma^2 \frac{\nu}{\nu - 2} \text{ für } \nu > 2$$

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

Betaverteilung

$$Y \sim \text{Be}(a, b) \text{ mit } a, b > 0, y \in [0, 1]$$

$$p(y|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}$$

$$E(Y|a, b) = \frac{a}{a+b},$$

$$\text{Var}(Y|a, b) = \frac{ab}{(a+b)^2(a+b+1)},$$

$$\text{mod}(Y|a, b) = \frac{a-1}{a+b-2} \text{ für } a, b > 1$$

2.5.3 Exponentialfamilie

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

$$1 = \int \exp^{t^T(y)\theta} h(y) dy \exp^{-\kappa(\theta)}$$

$$\Leftrightarrow \kappa(\theta) = \log \int \exp^{t^T(y)\theta} h(y) dy$$

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

Gammaverteilung

$$Y \sim \text{Ga}(a, b) \text{ mit } a, b > 0, y > 0$$

$$p(y|a, b) = \frac{b^a}{\Gamma(a)} y^{a-1} \exp(-by)$$

$$E(Y|a, b) = \frac{a}{b},$$

$$\text{Var}(Y|a, b) = \frac{a}{b^2},$$

$$\text{mod}(Y|a, b) = \frac{a-1}{b} \text{ für } a \geq 1$$

Invers-Gammaverteilung

$$Y \sim \text{IG}(a, b) \text{ mit } a, b > 0, y > 0$$

$$p(y|a, b) = \frac{b^a}{\Gamma(a)} y^{-a-1} \exp(-\frac{b}{y})$$

$$E(Y|a, b) = \frac{b}{a-1} \text{ für } a > 1,$$

$$\text{Var}(Y|a, b) = \frac{b^2}{(a-1)^2(a-2)} \text{ für } a \geq 2,$$

$$\text{mod}(Y|a, b) = \frac{b}{a+1}$$

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

Exponentialverteilung Zeit zwischen Poisson-Ereignissen

$$Y \sim \text{Exp}(\lambda) \text{ mit } \lambda > 0, y \geq 0$$

$$p(y|\lambda) = \lambda \exp(-\lambda y)$$

$$E(Y|\lambda) = \frac{1}{\lambda}, \text{ Var}(Y|\lambda) = \frac{1}{\lambda^2}$$

Chi-Quadrat-Verteilung quadrierte standardnormalverteilte Zufallsvariablen mit ν Freiheitsgraden

$$Y \sim \chi^2(\nu) \text{ mit } \nu > 0, y \in \mathbb{R}$$

$$p(y|\nu) = \frac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$$

$$E(Y|\nu) = \nu, \text{ Var}(Y|\nu) = 2\nu$$

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**

2.6 Grenzwertsätze

Gesetz der großen Zahlen

Zentraler Grenzwertsatz

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

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

Beweis:

Für eine normalverteilte Zufallsvariable $Z \sim N(\mu, \sigma^2)$ 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 $Z_n = (Y_1 + Y_2 + \dots + Y_n)/\sqrt{n}$ gilt:

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

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 $K_{Z_n}(t) = 0 + \sqrt{n}\mu t + \frac{1}{2}\sigma^2 t^2 + \dots$ schreiben, wobei die Terme in \dots alle für $n \rightarrow \infty$ gegen 0 gehen.

Damit gilt $K_{Z_n}(t) \xrightarrow{n \rightarrow \infty} K_Z(t)$ mit $Z \sim N(\sqrt{n}\mu, \sigma^2)$.

3 Inferenz

3.1 Methode der Momente

Die theoretischen Momente werden durch die empirischen geschätzt:

$$E_{\hat{\theta}_{MM}}(Y^k) = m_k(y_1, \dots, y_n)$$

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

3.2 Verlustfunktionen

Verlust

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

mit Parameterraum $\Theta \subset \mathbb{R}$, $t \in \mathcal{T}$ mit $t : \mathbb{R}^n \rightarrow \mathbb{R}$ 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

$$\begin{aligned} R(t(\cdot), \theta) &= E_{\theta}(\mathcal{L}(t(Y_1, \dots, Y_n), \theta)) \\ &= \int_{-\infty}^{\infty} \mathcal{L}(t(Y_1, \dots, Y_n), \theta) \prod_{i=1}^n f(y_i; \theta) dy_i \end{aligned}$$

Minimax-Regel

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

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

Es wird der Worst Case minimiert.

Mean Squared Error (MSE)

$$\begin{aligned} MSE(t(\cdot), \theta) &= E_{\theta}(\{t(Y) - \theta\}^2) \\ &= \text{Var}_{\theta}(t(Y_1, \dots, Y_n)) + \text{Bias}^2(t(\cdot); \theta) \end{aligned}$$

mit $\text{Bias}(t(\cdot); \theta) = E_{\theta}(t(Y_1, \dots, Y_n)) - \theta$

Beweis:

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

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

Cramér-Rao-Ungleichung

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

Beweis:

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

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

Kullback-Leibler-Divergenz Vergleich von Verteilungen

3.3 Maximum Likelihood (ML)

Voraussetzungen

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

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

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

Beweis:

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

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

Beweis:

$$\begin{aligned} R_{KL}(t(\cdot), \theta) &= \int_{-\infty}^{\infty} \mathcal{L}_{KL}(t(Y_1, \dots, Y_n), \theta) \prod_{i=1}^n f(y_i; \theta) dy_i \\ &= \int \int \log \frac{f(\tilde{y}; \theta)}{f(\tilde{y}; t)} f(\tilde{y}; \theta) d\tilde{y} \prod_{i=1}^n f(y_i; \theta) dy_i \\ &= \int \int (\log f(\tilde{y}; \theta) - \log f(\tilde{y}; t)) f(\tilde{y}; \theta) d\tilde{y} - \prod_{i=1}^n f(y_i; \theta) dy_i \\ &\approx - \int \underbrace{\left(\int \frac{\partial \log f(\tilde{y}; \theta)}{\partial \theta} f(\tilde{y}; \theta) d\tilde{y} \right)}_0 (t - \theta) \prod_{i=1}^n f(y_i; \theta) dy_i \\ &\quad + \frac{1}{2} \int \underbrace{\left(- \int \frac{\partial^2 \log f(\tilde{y}; \theta)}{\partial \theta^2} f(\tilde{y}; \theta) d\tilde{y} \right)}_{I(\theta)} (t - \theta)^2 \prod_{i=1}^n f(y_i; \theta) dy_i \end{aligned}$$

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

Eigenschaften der Score-Funktion

erste Bartlett-Gleichung:

$$\mathbb{E}(s(\theta; Y)) = 0$$

Beweis:

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

zweite Bartlett-Gleichung:

$$\text{Var}_\theta(s(Y; \theta)) = \mathbb{E}_\theta\left(-\frac{\partial^2 \log f(Y; \theta)}{\partial \theta^2}\right) = I(\theta)$$

Beweis:

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

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

Bartlett's zweite Gleichung gilt dann, weil $E(s(\theta; Y)) = 0$

ML-Schätzer

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

3.4 Suffizienz und Konsistenz

Suffizienz

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

Neyman-Kriterium:

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

3.5 Konfidenzintervalle

Definition

$$\begin{aligned} [t_l(Y), t_r(Y)] &\text{ Konfidenzintervall} \\ &\Leftrightarrow \\ P_{\theta} ((t_l(Y) \leq \theta \leq t_r(Y))) &\geq 1 - \alpha \end{aligned}$$

mit $1 - \alpha$ Konfidenzlevel und α Signifikanzlevel

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

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

Beweis:

$$\gamma = g(\theta) \Leftrightarrow \theta = g^{-1}(\gamma)$$

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 $\frac{\partial \theta}{\partial \gamma} I(\theta) \frac{\partial \theta}{\partial \gamma}$

Beweis:

$$I(\gamma) = \frac{\partial^2 l(g^{-1}(\hat{\gamma}))}{\partial \gamma^2} = \frac{\partial}{\partial \gamma} \left(\frac{\partial g^{-1}(\gamma)}{\partial \gamma} \frac{\partial l(\theta)}{\partial \theta} \right)$$

Beweis:

“ \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:

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

Minimalsuffizienz:

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

(schwache) Konsistenz

$$MSE(\hat{\theta}, \theta) \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \hat{\theta} \text{ konsistent}$$

Pivotal Statistik

$$g(Y; \theta) \text{ pivotal}$$

\Leftrightarrow

Verteilung von $g(Y; \theta)$ unabhängig von θ

Approximativ pivotal Statistik

$$g(\hat{\theta}; \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}} \approx N(0, 1)$$

mit $\hat{\theta} = t(Y) \approx N(\theta, \text{Var}(\hat{\theta}))$

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

Beweis:

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

4 Hypothesentests

Exakte binomiale Konfidenzintervalle

4.1 Tests für Einstichprobenprobleme

4.1.1 Normalverteilung

5 Regression

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

5.1 Annahmen

5.2 Verfahren

5.2.1 Kleinste Quadrate (OLS)

KQ-Schätzer (Einfachregression)

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

Beweis:

$$\begin{aligned} \text{Cov}(x, y) &= \text{Cov}(x, \hat{\beta}_0 + \hat{\beta}_1 x) = \hat{\beta}_1 \text{Var}(x) \\ &\iff \hat{\beta}_1 = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \end{aligned}$$

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

Beweis:

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

5.3 Modell

5.3.1 lineare Einfachregression

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

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$$

$$\hat{e}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{x}))$$

$$\sum_{i=1}^n \hat{e}_i = \sum_{i=1}^n y_i - \sum_{i=1}^n \bar{y} - \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})$$

$$= n\bar{y} - n\bar{y} - \hat{\beta}_1 (n\bar{x} - n\bar{x}) = 0$$

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} (n\bar{y} + \hat{\beta}_1 (n\bar{x} - n\bar{x})) = \bar{y}$$

5.3.2 Multivariate lineare Regression

5.4 ANOVA (Streuungszerlegung)

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

mit

$$SS_{Total} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SS_{Explained} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SS_{Residual} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = S_{yy} - \hat{\beta}^2 S_{xx}$$

5.5 Gütemaße

5.5.1 Bestimmtheitsmaß

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

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

6 Klassifikation

6.1 Diskriminanzanalyse (Bayes)

7 Clusteranalyse

8 Bayessche Statistik

8.1 Grundlagen

Bayes-Formel

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

oder allgemeiner:

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

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

$f(\theta) = \text{const.}$ 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 $\sim \text{Be}(\alpha, \beta)$
- $X \sim \text{Binom}(n, p, k)$
- Posteriori $\sim \text{Be}(\alpha + k, \beta + n - k)$

8.2 Markov Chain / Monte Carlo