Statistik Formelsammlung

Katharina Ring

8. April 2019

Inhaltsverzeichnis

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

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

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 $\in \mathbb{N}$, x $>$ nα

Minimum/Maximum

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

Arithmetisches Mittel

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

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

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

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

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

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:

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

(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

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

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

Dies sind auch die Werte für die Lorenzkurve.

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

Spearmans 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

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

Entspricht ohne Bindungen:

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

mit

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

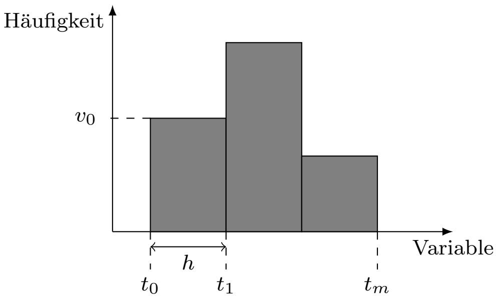
$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2 \quad \text{bzw. } s_{xy} = \frac{S_{xy}}{n}$$
$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{bzw. } s_{xx} = \frac{S_{xx}}{n}$$
$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{bzw. } s_{yy} = \frac{S_{yy}}{n}$$

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

1.2 Tabellen

1.3 Diagramme

1.3.1 Histogramm



Stichprobe: $X = \{x_1, x_2, ...; x_n\}$
k-te Klasse: $B_k = [t_k, t_{k+1})\,, k = \{0, 1, ..., m-1\}$
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		$n!$	$\frac{n!}{n_1!\cdots n_s!}$
Kombinationen:	ohne Reihenfolge	$\binom{n}{m}$	$\binom{n+m-1}{m}$
	mit Reihenfolge	$\binom{n}{m}m!$	n^m

Dabei gilt:
$n! = n \cdot (n-1) \cdot ... \cdot 1$
$\binom{n}{m} = \frac{n!}{m!(n-m)!}$

2.2 Wahrscheinlichkeitsrechnung

Laplace-Wahrscheinlichkeit

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

Axiome von Kolmogorov
mathematische Definition von Wahrscheinlichkeit

(1) $0 \leq P(A) \leq 1 \qquad \forall A \in \mathcal{A}$
(2) $P(\Omega) = 1$
(3) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
$\forall A_i \in \mathcal{A}, i = 1, ..., \infty$ 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 $A_i, ..., A_n$ vollständige Zerlegung von Ω in paarweise disjunkte Ereignisse

Mises' Wahrscheinlichtkeitsbegriff
frequentistische Definition von Wahrscheinlichkeit

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

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

Bedingte Wahrscheinlichkeit

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

Multiplikationssatz

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

Satz von der totalen Wahrscheinlichkeit

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

Satz von Bayes

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

Stochastische Unabhängigkeit

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

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

2.3 Zufallsvariablen

Definition

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

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*.

Momente

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

- **Varianz (2. zentriertes Moment)**:
  $\sigma^2 = Var(Y) = E(\{Y - E(Y)\}^2) = \int (y - E(Y))^2 f(y)dy$
  Varianzverschiebungssatz: $E(\{Y - \mu\}^2) = E(Y^2) - \mu^2$

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

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

Momenterzeugende Funktion

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

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

2.4 Zufallsvektoren

Dichte und Verteilungsfunktion

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

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

Marginale Dichte

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

Bedingte Dichte

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

Iterierter Erwartungswert

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

Beweis:

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

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

Beweis:

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

2.5 Verteilungen

2.5.1 Diskrete Verteilungen

Diskrete Gleichverteilung

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

Binomialverteilung Erfolge in unabhängigen Versuchen

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

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.

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

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

Geometrische Verteilung

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

Negative Binomialverteilung

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

2.5.2 Stetige Verteilungen

Stetige Gleichverteilung

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

Univariate Normalverteilung symmetrisch mit μ und σ^2

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

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

$$Y \sim \mathrm{t}_\nu(\mu,\sigma) \text{ mit } \mu \in \mathbb{R}, \sigma^2, \nu > 0, \; y \in \mathbb{R}$$
$$p(y|\mu,\sigma^2,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\Gamma(\sqrt{\nu\pi}\sigma)}\left(1 + \frac{(y-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$
$$\mathrm{E}(Y|\mu,\sigma^2,\nu) = \mu \text{ für } \nu > 1,$$
$$\mathrm{Var}(Y|\mu,\sigma^2,\nu) = \sigma^2\frac{\nu}{\nu-2} \text{ für } \nu > 2$$

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

$$Y \sim \mathrm{Be}(a,b) \text{ mit } a,b > 0, \; y \in [0,1]$$
$$p(y|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}$$
$$\mathrm{E}(Y|a,b) = \frac{a}{a+b},$$
$$\mathrm{Var}(Y|a,b) = \frac{ab}{(a+b)^2(a+b+1)},$$
$$\mathrm{mod}(Y|a,b) = \frac{a-1}{a+b-2} \text{ für } a,b > 1$$

Gammaverteilung

$$Y \sim \mathrm{Ga}(a,b) \text{ mit } a,b > 0, \; y > 0$$
$$p(y|a,b) = \frac{b^a}{\Gamma(a)} y^{a-1}\exp(-by)$$
$$\mathrm{E}(Y|a,b) = \frac{a}{b},$$
$$\mathrm{Var}(Y|a,b) = \frac{a}{b^a},$$
$$\mathrm{mod}(Y|a,b) = \frac{a-1}{b} \text{ für } a \geq 1$$

Invers-Gammaverteilung

$$Y \sim \mathrm{IG}(a,b) \text{ mit } a,b > 0, \; y > 0$$
$$p(y|a,b) = \frac{b^a}{\Gamma(a)} y^{-a-1}\exp(-\frac{b}{y})$$
$$\mathrm{E}(Y|a,b) = \frac{b}{a-1} \text{ für } a > 1,$$
$$\mathrm{Var}(Y|a,b) = \frac{b^2}{(a-1)^2(a-2)} \text{ für } a \geq 2,$$
$$\mathrm{mod}(Y|a,b) = \frac{b}{a+1}$$

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

Exponentialverteilung Zeit zwischen Poisson-Ereignissen

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

Chi-Quadrat-Verteilung quadrierte standardnormalverteilte Zufallsvariablen mit ν Freiheitsgraden

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

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 $\frac{\partial\kappa(\theta)}{\partial\theta} = \mathrm{E}(t(Y))$ 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**

2.6 Grenzwertsätze

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

Beweis:
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 $\frac{\partial^k K_Z(t)}{\partial t^k}\Big|_{t=0}$ 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: $K_{Z_n}(t) = nK_Y(t/\sqrt{n})$.

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

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

3 Hypothesentests

3.1 Tests für Einstichprobenprobleme

3.1.1 Normalverteilung

4 Regression

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

4.1 Annahmen

4.2 Verfahren

4.2.1 Kleinste Quadrate (OLS)

KQ-Schätzer (Einfachregression)

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

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

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

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

4.3 Modell

4.3.1 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

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

4.3.2 Multivariate lineare Regression

4.4 ANOVA (Streuungszerlegung)

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

mit

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

4.5 Gütemaße

4.5.1 Bestimmtheitsmaß

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

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

5 Inferenz

5.1 Methode der Momente

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: $\hat{\theta}_{MM} = \hat{\theta}_{ML}$

5.2 Verlustfunktionen

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

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

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

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

hier auch Kullback Leibler Distanz?

5.3 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** $I(\theta)$: $-\mathrm{E}_\theta\left(\frac{\partial s(\theta;Y_1,...,Y_n)}{\partial\theta}\right)$

Eigenschaften der Score-Funktion

erste Bartlett Gleichung:

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

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:

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

Beweis:

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

$\Leftrightarrow \mathrm{E}_\theta\left(s(\theta;Y)s(\theta;Y)\right) = \mathrm{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(Y;\theta)\right)$
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: $s\left(\hat{\theta}_{ML};y_1,...,y_n\right) = 0$
Der ML-Schätzer ist invariant.

5.4 Suffizienz, Konstistenz und Effizienz

5.5 Konfidenzintervalle

6 Klassifikation

6.1 Diskriminanzanalyse (Bayes)

7 Clusteranalyse

8 Bayessche Statistik

8.1 Grundlagen

Bayes-Formel

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

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