ZUSAMMENFASSUNG DES SKRIPTUMS ZU *EINFÜHRUNG IN DIE WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK* VON PROF. FELSENSTEIN

PHILIPP DÖRSEK

Der Autor übernimmt keinerlei Garantie für die Richtigkeit. Die meisten Beiträge wurden zwar ausschließlich aus dem angeführten Skriptum übernommen, doch sind Tippfehler nie auszuschließen. Daher ist der Autor immer erfreut über Hinweise, die er unter `pdoersek@fsmat.at` gerne entgegen nimmt.

(1) **Additionssatz:** Für Ereignisse A und B gilt
$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$
(2) **Additionstheorem – Gammaverteilung:** Für zwei stochastische Größen X_1 und X_2 gilt
$$X_1 \sim \gamma(\alpha, \beta),\ X_2 \sim \gamma(\tilde{\alpha}, \beta) \Rightarrow X_1 + X_2 \sim \gamma(\alpha + \tilde{\alpha}, \beta)$$
(3) **Additionstheorem – Normalverteilung:** Für zwei stochastische Größen X_1 und X_2 gilt
$$X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2),\ X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2) \Rightarrow X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma^2)$$
wobei $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho$ mit ρ dem Korrelationskoeffizient.

(4) **Additionstheorem – Poissonverteilung:** Für zwei stochastische Größen X_1 und X_2 gilt
$$X_1 \sim \mathrm{P}_{\mu_1},\ X_2 \sim \mathrm{P}_{\mu_2} \Rightarrow X_1 + X_2 \sim \mathrm{P}_{\mu_1+\mu_2}$$
(5) **Aktualisierung:** Der Vorgang, Beobachtungen in die Einschätzung des Parameters einfließen zu lassen (engl. "updating").

(6) **Bayes'sche Formel:** Sei H_i eine Zerlegung des Merkmalsraums und A ein Ereignis mit $\mathbf{P}(A) > 0$. Dann gilt
$$\mathbf{P}(H_i|A) = \frac{\mathbf{P}(A \cap H_i)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|H_i)\,\mathbf{P}(H_i)}{\sum_i \mathbf{P}(A|H_i)\,\mathbf{P}(H_i)}$$
(7) **Bayes'sches Rückschlussprinzip:** Von der bedingten Wahrscheinlichkeit $\mathbf{P}(E|H_i)$, dass E unter der Hypothese H_i eintritt, wird, nachdem E eingetreten ist, auf die Wahrscheinlichkeit der Hypothesen $\mathbf{P}(H_i|E)$ *zurückgeschlossen*.

(8) **Bayes'sches Theorem:** Sei $\pi(\theta)$ eine Dichte oder Punktwahrscheinlichkeit auf $\Theta \subseteq \mathbb{R}^n$ und die reellwertige stochastische Größe X habe die bedingte Dichte $\mathrm{f}(x|\theta)$. Dann gilt für die Dichte der bedingten Verteilung von $\theta|X$
$$\pi(\theta|x) = \frac{\mathrm{f}(x|\theta)\,\pi(\theta)}{\int_\Theta \mathrm{f}(x|\theta)\,\pi(\theta)d\theta}$$
(9) **Bayes-Schätzer:** Der *Bayes-Schätzer* θ_B ist der Erwartungswert der a-posteriori-Verteilung, i. Z.
$$\theta_B = \mathbb{E}_{\pi(\theta|x)}\,\theta$$
(10) **Beobachtung:** Das Resultat eines Experiments, das nach dessen Durchführung vorliegt.

(11) **Beobachtungsgruppe*:** Andere Bezeichnung für normalverteilte Stichproben im Rahmen einer Varianzanalyse.

(12) **Bernoulli-Experiment:** Experiment, bei dem nur 2 mögliche Versuchsausgänge existieren ("wahr" oder "falsch")

(13) **Bestimmtheitsmaß*:** Das Bestimmtheitsmaß B der Regression ist
$$B = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \overline{x}_i)^2} = 1 - \frac{\sum_{i=1}^n (\hat{x}_i - \overline{x}_i)^2}{\sum_{i=1}^n (x_i - \overline{x}_i)^2}$$
(14) **Beta-Verteilung:** Eine stochastische Größe ist *beta-verteilt* $X \sim \beta(a, b)$, wenn die Dichte von X durch
$$\mathrm{f}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} x^{a-1}(1-x)^{b-1} \quad \text{für } 0 < x < 1$$
gegeben ist.

(15) **Binomischer Lehrsatz:** Für jedes Binom $(x+y)^n$ gilt
$$(x+y)^n = \sum_{j=0}^n \binom{n}{j} x^j y^{n-j}$$
(16) **Chi-Quadrat-Test*:** Test, ob eine bestimmte Verteilung einer Stichprobe zu Grunde liegt.

(17) **Designmatrix*:** Matrix aus den Werten der Regressoren.

(18) **Dichte:** Eine nicht negative, integrierbare Funktion f, für die
$$\int_{-\infty}^{\infty} \mathrm{f}(x)dx = 1$$
gilt.

(19) **Dichte – a-posteriori:** Die mit Hilfe des Bayes'schen Theorems gefundene bedingte Verteilung $\pi(\theta|x)$.

(20) **Dichte – a-priori:** Die Dichte $\pi(\theta)$ vor der Anwendung des Bayes'schen Theorems.

(21) **Dichte – bedingte:** Die stochastische Größe (X, Y) sei mit der gemeinsamen Dichte $\mathrm{f}(x, y)$ verteilt. $\mathrm{f}_Y(.)$ bezeichne die Randdichte von Y. Die *bedingte Dichte* von X gegeben $Y = y$ ist
$$\mathrm{f}(x|y) = \frac{\mathrm{f}(x, y)}{\mathrm{f}_Y(y)}$$
(22) **Dichte – mehrdimensionale:** Eine integrierbare Funktion $\mathrm{f} : \mathbb{R}^n \to \mathbb{R}^+$ mit
$$\int \cdots \int_{\mathbb{R}^n} \mathrm{f}(x_1, \ldots, x_n)dx_1 \cdots dx_n = 1$$

(23) **diskretisiert:** Sei X stetig mit der Dichte f auf M verteilt, aber sei $H : M \to M_Y$ eine Abbildung derart, dass $H(M) = M_Y$ eine abzählbare Menge ist. Dann sind die Punktwahrscheinlichkeiten von $Y = H(X)$

$$\mathbf{P}[Y = y_i] = \int_{H^{-1}(y_i)} \mathrm{f}(x)dx$$

diskrete Werte; X wurde durch H *diskretisiert*.

(24) **effizient:** Ein Schätzer T^* heißt in der Klasse $\mathcal{S}$ von Schätzfunktionen *klassen-effizient*, wenn die mittlere quadratische Abweichung (MSE) in dieser Klasse bei T^* minimal wird. Besteht die Klasse aus den erwartungstreuen Schätzfunktionen, so heißt T^* *effizient*.

(25) **Elementarereignis:** Ein Versuchsausgang ω im Merkmalsraum Ω wird als *Elementarereignis* bezeichnet.

(26) **Ereignis:** Eine Teilmenge E des Merkmalsraums Ω wird als Ereignis bezeichnet.

(27) **Ereignisfeld:** Eine Menge $\mathcal{E}$ von Ereignissen des Merkmalsraums Ω heißt *Ereignisfeld*, wenn
- $\Omega \in \mathcal{E}$
- Für $E \in \mathcal{E}$ gilt auch $E^c \in \mathcal{E}$
- Für eine Folge von Ereignissen $E_i \in \mathcal{E}$ ist auch die Vereinigung ein Ereignis, i. Z.

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{E}$$

(28) **Erwartung:** Die *Erwartung* einer diskreten Verteilung mit den Punktwahrscheinlichkeiten p_i auf den Punkten x_i ist (falls die Summe endlich ist)

$$\mathbb{E}\, X = \sum_i x_i p_i = \sum_{\omega\in\Omega} X(\omega)\, \mathbf{P}(\omega).$$

Die *Erwartung* einer stetigen Verteilung mit der Dichte f(.) ist (falls das Integral existiert) definiert durch

$$\mathbb{E}\, X = \int_{-\infty}^{\infty} x\, \mathrm{f}(x)dx$$

(29) **Erwartung – bedingte:** Der *bedingte Erwartungswert* von X bei festem $Y = y$ ist

$$\mathbb{E}(X|Y = y) = \sum_i \mathbf{P}[X = x_i|Y = y],$$

wenn X diskret auf $x_1, x_2, \ldots$ und

$$\mathbb{E}(X|Y = y) = \int x\, \mathrm{f}(x|y)dx,$$

wenn X stetig verteilt ist.

(30) **erwartungstreu:** Eine Schätzfunktion T, die

$$\mathbb{E}\, T = \theta$$

für alle θ erfüllt, heißt *erwartungstreu* oder unverzerrt für θ.

(31) **erwartungstreu – asymptotisch:** Es sei $T_n = T(X_1, \ldots, X_n)$ die Schätzfunktion für θ, wenn n Beobachtungen vorliegen. Ein Schätzer heißt *asymptotisch erwartungstreu*, wenn der Erwartungswert der Schätzfunktion gegen den Parameter konvergiert, i. Z.

$$\mathbb{E}\, T_n \to \theta \quad \text{für } n \to \infty.$$

(32) **Exzess:** Der *Exzess* (die *Kurtosis*) ist ein Maß für die Gestalt der Verteilung. Er ist definiert als

$$\frac{\mathbb{E}(X - \mathbb{E}\, X)^4}{\mathrm{var}(X)^2}.$$

Empirisch verwendet man aufgrund der Tatsache, dass der Exzess der Normalverteilung 3 ist,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{(x_i - \overline{x})^4}{s^4} - 3.$$

(33) **Faltung:** Für die Summe von unabhängigen stochastischen Größen X, Y gilt, wenn X und Y diskret sind,

$$\mathbf{P}[X + Y = k] = \sum_i \mathbf{P}[Y = k - i]\, \mathbf{P}[X = i],$$

wenn X und Y mit den Dichten $\mathrm{f}_X(.)$ und $\mathrm{f}_Y(.)$ stetig verteilt sind,

$$h(x) = \int \mathrm{f}_X(t)\, \mathrm{f}_Y(x - t)dt.$$

(34) **Fehler – 1. Art:** Wird die Nullhypothese verworfen, obwohl sie stimmt, so bezeichnet man diesen Fehler als *Fehler 1. Art*.

(35) **Fehler – 2. Art:** Wird eine falsche Nullhypothese angenommen, so bezeichnet man diesen Fehler als *Fehler 2. Art*.

(36) **Fisher's Z*:** Ein Test, ob ein (linearer) Zusammenhang zwischen zwei stochastischen Größen vorliegt.

(37) **Fourier-Transformierte:** Die *charakteristische Funktion* oder Fourier-Transformierte der Verteilung der stochastischen Größe X ist

$$\varphi(t) = \mathbb{E}(e^{\mathrm{i}tX}) = \Psi(\mathrm{i}\,t).$$

(38) **Fraktil:** Das *Fraktil* (das *α-Quantil* bezeichnet jenen Wert x_α, für den

$$\mathrm{F}(x_\alpha) = \mathbf{P}(X \le x_\alpha) = \alpha$$

erfüllt ist. Diese Gleichung muss bei diskreten Verteilungen keine Lösung haben; dann definiert man als α-Quantil den kleinsten Wert j, für den $\mathrm{F}(j) \ge \alpha$ gilt.

(39) **Freiheitsgrad:** Eine χ^2_k-Verteilung (χ^2-Verteilung mit $k \in \mathbb{N}$ Freiheitsgraden) ist eine spezielle Gamma-Verteilung mit $a = k/2$ und $b = 1/2$. Die Dichte ist

$$\mathrm{f}(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{\frac{k}{2}-1} e^{-x/2} \qquad \text{für } x > 0.$$

(40) **Funktion – charakteristische:** Die *charakteristische Funktion* oder Fourier-Transformierte der Verteilung der stochastischen Größe X ist

$$\varphi(t) = \mathbb{E}(e^{\mathrm{i}tX}) = \Psi(\mathrm{i}\,t).$$

(41) **Funktion – erzeugende:** Die Funktion
$$\psi(s) = \Psi(\ln s) = \mathbb{E}(s^X)$$
mit $s = e^t$ für ein $t \in I \subseteq \mathbb{R}$ heißt *erzeugende Funktion.*

(42) **Funktion – momenterzeugende:** Ist für X für alle $t \in I \subseteq \mathbb{R}$ der Ausdruck
$$\Psi(t) = \mathbb{E}(e^{tX})$$
endlich, so heißt $\Psi(.)$ *momenterzeugende Funktion* (oder *Laplace-Transformierte*) von X.

(43) **Gammafunktion:** Für die analytische Behandlung der Fakultät auch für nicht ganzzahlige Werte wird die *Gammafunktion* als
$$\Gamma(z) := \int_0^\infty t^{z-1} e^- t dt$$
für $z > 0$ definiert, mit der Funktionalgleichung $\Gamma(z+1) = z\Gamma(z)$ ebenfalls für $z > 0$. Da $\Gamma(1) = 1$ gilt, folgt für $n \in \mathbb{N}$ $\Gamma(n) = (n-1)!$.

(44) **Gegenhypothese:** Neben der Hypothese H_0, die vermutet wird und als *Nullhypothese* bezeichnet wird, kommt es oft auch auf die entgegengesetzte Annahme an, die sogenannte *Gegenhypothese*, die mit H_1 bezeichnet wird.

(45) **Gegenwahrscheinlichkeit:** Ist E ein Ereignis, so ist $\mathbf{P}(E^c) = 1 - \mathbf{P}(E)$ die *Gegenwahrscheinlichkeit* von E.

(46) **Gesetz der großen Zahlen:** Das *schwache Gesetz der großen Zahlen* besagt, dass für eine Folge $(X_n)_{n\in\mathbb{N}}$ aus paarweise unkorrelierten stochastischen Größen mit identischem Erwartungswert $\mathbb{E}\, X_n = \mu$, deren Varianzen
$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1} n \operatorname{var}(X_i) = 0$$
erfüllen, das Mittel stochastisch konvergiert, i. Z.
$$\overline{X}_n \xrightarrow{P} \mu.$$
Das *starke Gesetz der großen Zahlen* besagt, dass für unabhängige Größen
$$\mathbf{P}[\lim_{n\to\infty} \overline{X}_n = \mu] = 1$$
gilt, d. h. nicht nur wird die Abweichungswahrscheinlichkeit immer kleiner, sondern mit Wahrscheinlichkeit 1 konvergiert das Stichprobenmittel wie eine Zahlenfolge gegen den Erwartungswert.

(47) **Gesetz der großen Zahlen – empirisches:** Das *empirische Gesetz der großen Zahlen* besagt, dass die relative Häufigkeit von E bei n Versuchen, $\mathrm{h}_n(E)$, gegen die Wahrscheinlichkeit von E konvergiert, i. Z.
$$\mathrm{h}_n(E) \xrightarrow{P} \mathbf{P}(E).$$
Es folgt aus dem Gesetz der großen Zahlen.

(48) **Gleichverteilung – diskrete:** Die stochastische Größe $X \sim \mathrm{D}_m$ beschreibe die zufällige Wahl eines Wertes aus $1, 2, \ldots, m$, die Wahrscheinlichkeit jedes Wertes ist gleich, i. Z.
$$\mathbf{P}(X = i) = \frac{1}{m}$$
für $i = 1, 2, \ldots, m$. Der Wurf eines Würfels ist ein Beispiel für eine diskrete Gleichverteilung. Die Verteilungsfunktion an einer beliebigen Stelle x zwischen $1 \leq x \leq m$ ist
$$\mathrm{F}(x) = \frac{\lfloor x \rfloor}{m}.$$

(49) **Gleichverteilung – stetige:** Auf dem Intervall (a, b) ist $X \sim \mathrm{U}_{a,b}$ stetig gleichverteilt, wenn
$$\mathrm{f}(x) = \frac{1}{b-a}$$
für $a \leq x \leq b$ erfüllt wird; außerhalb des Intervalls ist die Dichte 0. Die Verteilungsfunktion ist
$$\mathrm{F}(x) = \frac{x-a}{b-a},$$
der Median $x_{\text{med}} = \frac{a+b}{2}$.

(50) **Häufigkeit – absolute:** Die Anzahl $\mathrm{H}_n(E)$, wie oft ein Ereignis E bei n Versuchen aufgetreten ist, bezeichnet man als *absolute Häufigkeit.*

(51) **Häufigkeit – relative:** Den relativen Anteil der absoluten Häufigkeit an der Gesamtzahl der Versuche, $\mathrm{h}_n(E) = \frac{\mathrm{H}_n(E)}{n}$, bezeichnet man als *relative Häufigkeit.*

(52) **Histogramm:** Die Darstellung der relativen Häufigkeiten als Balkendiagramm bezeichnet man als *Histogramm.*

(53) **Histogramm – zweidimensionales:** Die graphische Darstellung einer Kontingenztafel zweier stochastischer Größen nennt man *zweidimensionales Histogramm.*

(54) **HPD-Bereich:** Die Überdeckungswahrscheinlichkeit sei $0 \leq \beta \leq 1$, $\pi(\theta|D)$ sei die a-posteriori-Dichte. Dann nennt man einen Bereich von der Form
$$B^* = \theta | \, \pi(\theta|D) \geq k_\beta,$$
wobei k_β die größte Konstante ist, sodass $\mathbf{P}(B^*|D) \geq \beta$ gilt, einen *Bereich höchster a-posteriori-Dichte* oder kurz *HPD-Bereich.*

(55) **Hyperparameter:** A-priori-Verteilungen besitzen meist selbst Parameter, die als *Hyperparameter* bezeichnet werden. Im Fall einer konjugierten Familie werden durch die Beobachtung nur die Hyperparameter verändert, die a-posteriori-Dichte gehört aber zur selben Familie wie die a-priori-Dichte. Beim Übergang von der a-priori- zur a-posteriori-Verteilung sind also nur Hyperparameter zu verändern; man spricht in diesem Zusammenhang auch von *unter den Beobachtungen abgeschlossenen Familien.*

(56) **Hypothese:** Für den Begriff der *Hypothese* gibt es zwei unabhängige Verwendungen:

- Die Elemente eines endliches Mengensystems $(H_i)_{i=1}^n$, das den gesamten Merkmalsraum disjunkt zerlegt, d. h.
$$\Omega = H_1 + \cdots H_n = \bigcup_{i=1}^n \quad \text{und} \quad H_i \bigcap \left(\bigcup_{j\neq i} H_j \right) = \emptyset,$$
werden als *Hypothesen* bezeichnet.
- Eine Annahme über einen unbekannten Parameter θ eines Modells, z. B. $\theta \geq \theta_0$, wird als *Hypothese H* bezeichnet.

(57) **Information:** *Information* ist alles, was bestehende Einschätzungen ändert.

(58) **Kleinste-Quadrate-Schätzer*:** Schätzwerte, die die Quadratsumme

$$\sum_{i=1}^{n}(x_i - \theta_1 - \theta_2 z_i)^2$$

minimieren.

(59) **Kolmogoroff-Smirnoff-Test*:** Die Kolmogoroff-Smirnoff-Verteilung $\sqrt{n}D_n = \sqrt{n}\sup_x|\mathrm{F}_n^*(x) - \mathrm{F}_0(x)|$ wird mit dem $1-\alpha$-Fraktil der Grenzverteilung verglichen. Ist der Wert der Verteilung größer, so muss die Hypothese, dass die Stichprobe die Verteilung $\mathrm{F}_0(.)$ besitzt, verworfen werden.

(60) **Kombination:** Ziehungen ohne Berücksichtigung der Reihenfolge. Ohne Wiederholung ergibt sich bei n Objekten und k Ziehungen der Binomialkoeffizient $\binom{n}{k}$, mit Wiederholung $\binom{n+k-1}{k}$.

(61) **Konfidenzbereich:** Ein *Konfidenzbereich* $\mathrm{C}(\mathbf{x})$ für $\theta \in \Theta$ ist eine Abbildung

$$\mathbf{x} \to \mathrm{C}(\mathbf{x}) \subseteq \Theta$$

für jedes $\mathbf{x}$ aus dem Stichprobenraum. $\mathrm{C}(\mathbf{x})$ ist ein Konfidenzbereich mit *Überdeckungswahrscheinlichkeit* β (oder zum *Niveau* β), wenn

$$\mathbf{P}_\theta(\theta \in \mathrm{C}(\mathbf{X})) = \beta$$

für alle θ gilt.

(62) **konjugiert:** Besitzt eine a-priori-Verteilung die gleiche Struktur wie die Likelihoodfunktion und passt sie sich daher dem Modell in natürlicher Art und Weise an, so nennt man die a-priori-Verteilung *zum Modell konjugiert*.

(63) **konsistent:** Eine Folge T_n von Schätzfunktionen ist konsistent, wenn sie asymptotisch erwartungstreu ist und die Varianz der Schätzung gegen 0 strebt, i. Z.

$$\mathrm{var}(T_n) \to 0.$$

(64) **Kontingenztafel:** Seien X und Y zwei diskrete Merkmale (sind X und Y stetig, so können sie z. B. durch eine Einteilung in Klassen diskretisiert werden) und (x_i, y_i) gemessene Daten. Dann wird eine Tabelle der absoluten Häufigkeiten *Kontingenztafel* genannt.

(65) **Kontrollgruppe*:** Soll festgestellt werden, ob ein Wirkstoff Veränderungen bewirkt, so wird ein Teil der Objekte mit dem Wirkstoff behandelt, ein anderer nicht. Der nicht behandelte Teil wird auch *Kontrollgruppe* genannt.

(66) **Korrelationskoeffizient:** Der *Korrelationskoeffizient* der stochastischen Größen X und Y ist definiert durch

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}.$$

(67) **Korrelationskoeffizient – empirischer:** Der *empirische Korrelationskoeffizient* ist der Funktionswert der Statistik $R = r$ mit

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

und daher die empirische Messung des linearen Zusammenhangs zwischen dem Merkmal X und dem Merkmal Y.

(68) **Korrelationsmatrix:** Die *empirische Korrelationsmatrix* enthält die empirischen Korrelationskoeffizienten r_{ij} zwischen $\mathbf{x}_i$ und $\mathbf{x}_j$.

(69) **Kovarianz:** Seien X, Y stochastische Größen mit endlicher Varianz. Die *Kovarianz* von X und Y ist

$$\mathrm{cov}(X,Y) = \mathbb{E}(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y).$$

(70) **Kovarianz – empirische:** Die *empirische Kovarianz* ist definiert als

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

Zur Berechnung verwendet man für gewöhnlich den Verschiebungssatz

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}.$$

(71) **Kovarianzmatrix:** Die Komponenten X_i des stochastischen Vektors $\mathbf{X} = (X_1, \ldots, X_k)$ sollen endliche Varianz besitzen. Dann heißt die Matrix $\Sigma = (\sigma_{ij})_{i,j}$ mit $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$ der Dimension $k \times k$ die *Kovarianzmatrix* von $\mathbf{X}$.

Die Matrix Σ ist immer symmetrisch und nichtnegativ definit, in der Diagonale enthält sie die Varianzen der Komponenten. Sie heißt auch *verallgemeinerte Varianz*.

Die *empirische Kovarianzmatrix* $S = (s_{ij}$ hat als Elemente die empirischen Kovarianzen, i. Z.

$$s_{ij} = \frac{1}{n}\sum_{l=1} n(x_{il} - \overline{x}_i)(x_{jl} - \overline{x}_j),$$

bei einer konkreten Stichprobe $\mathbf{x}_1, \ldots, \mathbf{x}_n$ mit $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})$ und $\overline{x}_i = \frac{1}{n}\sum_{i=1}^{n} x_{il}$.

(72) **Kurtosis:** Die *Kurtosis* (der *Exzess*) ist ein Maß für die Gestalt der Verteilung. Er ist definiert als

$$\frac{\mathbb{E}(X - \mathbb{E}\,X)^4}{\mathrm{var}(X)^2}.$$

Empirisch verwendet man aufgrund der Tatsache, dass die Kurtosis der Normalverteilung 3 ist,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \overline{x})^4}{s^4} - 3.$$

(73) **Laplace-Raum:** Sei Ω ein Merkmalsraum mit endlich vielen Elementen, d. h. $|\Omega| = n \in \mathbb{N}$. Dann ist Ω ein *Laplace-Raum*, wenn auf Ω eine Gleichverteilung definiert wird, d. h. für $E \subseteq \Omega$ gilt

$$\mathbf{P}(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{n}.$$

(74) **Laplace-Transformierte:** Ist für X für alle $t \in I \subseteq \mathbb{R}$ der Ausdruck

$$\Psi(t) = \mathbb{E}(e^{tX})$$

endlich, so heißt $\Psi(.)$ *Laplace-Transformierte* (oder *momenterzeugende Funktion*) von X.

(75) **Likelihood-Gleichung:** Wird die Ableitung der Likelihoodfunktion (oder ihres Logarithmus) gleich 0 gesetzt, um die lokalen Extrema zu bestimmen, so erhält man die *Likelihood-Gleichung* $\frac{\partial}{\partial\lambda} l = 0$.

(76) **Likelihoodfunktion:** Als die *Likelihoodfunktion* (oder *Plausibilitätsfunktion*) $\mathrm{l}(\theta)$ wird nach der Beobachtung von $\mathbf{x} = (X_1, \ldots, X_n)$

- die gemeinsame Dichte
$$\mathrm{l}(\theta) = \mathrm{l}(\theta; x_1, \ldots, x_n) = \mathrm{f}(x_1, \ldots, x_n|\theta)$$
bei einem stetigen Merkmal und
- die gemeinsame Wahrscheinlichkeit
$$\mathrm{l}(\theta) = \mathrm{l}(\theta; x_1, \ldots, x_n) = \mathbf{P}[X_1 = x_1, \ldots, X_n = x_n|\theta]$$
bei einer diskreten Verteilung

bezeichnet.

(77) **Markoff'sche Ungleichung:** Für eine stochastische Größe X mit existierendem Erwartungswert gilt für jedes $\epsilon > 0$
$$\mathbf{P}[|X| \geq \epsilon] \leq \frac{\mathbb{E}|X|}{\epsilon}.$$

(78) **Maximum-Likelihood-Schätzer:** Der *Maximum-Likelihood-Schätzer* ist jener Parameterwert, der die Likelihoodfunktion $\mathrm{l}(\theta)$ maximiert.

(79) **messbar:** Für eine *messbare Funktion* $X : \Omega \to \mathbb{R}$ von einem Merkmalsraum nach $\mathbb{R}$ gilt, dass das Urbild eines halboffenen Intervalls ein Ereignis ist, i. Z. $X^{-1}((a, b]) \in \mathcal{E}$.

(80) **Median:** Der *Median* teilt die Verteilung in gleich wahrscheinliche Bereiche, d. h. der Median ist das 0.5-Quantil, i. Z. $x_{\mathrm{med}} = x_{0.5}$, und damit gilt $\mathrm{F}(x_{\mathrm{med}}) = \frac{1}{2}$.

(81) **Merkmalsraum:** Der *Merkmalsraum* ist die Menge Ω der möglichen Resultate eines Experiments.

(82) **Mittel:** Das *Mittel* (die *Erwartung*) einer diskreten Verteilung mit den Punktwahrscheinlichkeiten p_i auf den Punkten x_i ist (falls die Summe endlich ist)
$$\mathbb{E}\, X = \sum_i x_i p_i = \sum_{\omega \in \Omega} X(\omega)\, \mathbf{P}(\omega).$$
Das *Mittel* (die *Erwartung*) einer stetigen Verteilung mit der Dichte $\mathrm{f}(.)$ ist (falls das Integral existiert) definiert durch
$$\mathbb{E}\, X = \int_{-\infty}^{\infty} x\, \mathrm{f}(x) dx$$

(83) **mittlere absolute Distanz:** Ein Maß für die Streuung, das weniger empfindlich auf große Abweichungen reagiert, ist
$$\mathbb{E}|X - \mathrm{med}(X)|,$$
wofür die *mittlere absolute Distanz* (MAD)
$$\mathrm{MAD} = \frac{1}{n} \sum_{i=1}^{n} |x_i - x_{\mathrm{med}}|$$
der Schätzwert aus der empirischen Verteilung ist.

(84) **mittlere quadratische Abweichung:** Für nicht erwartungstreue Schätzer ist nicht die Varianz, sondern die *mittlere quadratische Abweichung*
$$\mathrm{MSE}(T) = \mathbb{E}(T - \theta)^2$$
das Maß für die Ungenauigkeit.

(85) **Modell:** Als *parametrisches Modell* wird die Festlegung eines Verteilungstyps bis auf sogenannte *Modellparameter* bzw. kurz *Parameter* $\theta = (\theta_1, \ldots, \theta_r)$ bezeichnet.

(86) **Modus:** Unter dem *Modus* einer stetigen Verteilung versteht man jenen Wert x_{mod}, an dem die Dichte das Maximum erreicht.

(87) **Moment:** Das *erste Moment* ist eine andere Bezeichnung für den Erwartungswert.

Das *k-te Moment* (ein *höheres Moment*) bezeichnet den Erwartungswert von X^k mit $k > 1$, d. h.
$$\mathbb{E}\, X^k.$$

(88) **Moment – zentrales:** Ein *zentrales Moment* bezeichnet ein höheres Moment, für dessen Berechnung die Erwartung von X nach 0 verschoben wurde, d. h.
$$\mathbb{E}(X - \mathbb{E}\, X)^k.$$

(89) **Multinomial-Lehrsatz:** Der *Multinomial-Lehrsatz* liefert eine Möglichkeit, einen Ausdruck der Form $(x_1 + \cdots + x_s)^n$ zu berechnen. Es gilt
$$(x_1 + \cdots + x_s)^n = \sum_{k_1 + \cdots + k_s = n} \binom{n}{k_1, \ldots, k_s} x_1^{k_1} \cdots x_s^{k_s}.$$

(90) **Multinomialkoeffizient:** Der *Multinomialkoeffizient*, eine Verallgemeinerung des Binomialkoeffizienten, ist definiert als
$$\binom{n}{k_1, \ldots, k_s} = \frac{n!}{k_1! \cdots k_s!}.$$

(91) **Multiplikationssatz:** Für Ereignisse A_i mit positiver Wahrscheinlichkeit für den Durchschnitt gilt
$$\mathbf{P}(A_1 \cap \cdots \cap A_n) = \mathbf{P}(A_1)\, \mathbf{P}(A_2|A_1)\, \mathbf{P}(A_3|A_1 \cap A_2) \cdots \mathbf{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

(92) **multivariat:** *Multivariate Daten* entstehen, wenn nicht nur ein Merkmal, sondern mehrere (in einem stochastischen Vektor zusammengefasste) Merkmale enthält.

(93) **Normalgleichung*:** Die Nullstelle der (vektorwertigen) Ableitung der Quadrate der Residuen, die Lösung der Kleinste-Quadrate-Schätzer, erfülllt die *Normalgleichungen*
$$D^T D\, \boldsymbol{\theta} = D^T \mathbf{x}.$$

(94) **Normalverteilung:** Die stochastische Größe $X \sim \mathrm{N}(\mu, \sigma^2)$ ist nach einer *Normalverteilung* (*Gaußverteilung*) mit der Dichte
$$\mathrm{f}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right)$$
verteilt. Dabei gilt $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. $\mathrm{N}(0, 1^2)$ nennt man *Standardnormalverteilung*, deren Dichte für gewöhnlich mit $\phi(.)$ und deren Verteilungsfunktion mit $\Phi(.)$ bezeichnet wird. Für $\Phi(.)$ müssen Tabellen verwendet werden, weil sie nicht in geschlossener Form angegeben werden kann.

Eine allgemeine Normalverteilung kann durch eine lineare Transformation auf die Standardnormalverteilung zurückgeführt werden; die Dichte ist
$$\mathrm{f}(x) = \frac{1}{\sigma}\, \phi(\frac{x - \mu}{\sigma})$$

und die Verteilungsfunktion ist

$$\mathrm{F}(x) = \Phi(\frac{x-\mu}{\sigma}).$$

Weiters gilt $\mu = x_{\mathrm{mod}} = x_{\mathrm{med}}$.

(95) **Normalverteilung – bivariate:** Die *bivariate Normalverteilung* wird mit $\mathrm{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ beschrieben, wobei ρ den Korrelationskoeffizienten bezeichnet. Die Dichte hat die Form

$$\mathrm{f}(x,y) = c_N \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right],$$

wobei $c_N = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$ gilt.

(96) **Normalverteilung – logarithmische:** Die positive Größe $X \sim \mathrm{LN}(\mu, \sigma^2)$ heißt *logarithmisch normalverteilt*, wenn gilt

$$\log(X) \sim \mathrm{N}(\mu, \sigma^2).$$

(97) **Normalverteilung – multivariate:** Der Vektor $\mathbf{X} \in \mathbb{R}^k$ heißt multivariat normalverteilt $\mathrm{N}(\boldsymbol{\mu}, \Sigma)$, wenn gilt

$$\mathrm{f}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

(98) **Nullhypothese:** Die *Nullhypothese* H_0 wird vermutet (z. B. $H_0 : \mu = 0$). Weiters gibt es die Gegenhypothese.

(99) **Ordnungsstatistik:** Das Ordnen der Werte eines Vektors unabhängiger stochastischer Größen $X_1, \ldots, X_n$ führt auf die *Ordnungsstatistiken*, i. Z. $X_{(1)}, \ldots, X_{(n)}$ mit

$$X_{(1)} \leq \cdots \leq X_{(n)}.$$

Durch das Ordnen entsteht ein Vektor $X_{(1)}, \ldots, X_{(n)}$ von stochastischen Größen, die nicht mehr unabhängig sind.

(100) **Parameter:**
- In der allgemeinen Definition einer Familie von Verteilungen tauchen noch Parameter auf. Erst, wenn diese eingesetzt werden, entsteht eine konkrete Verteilung.
- In einem parametrischen Modell ist die Verteilung nur bis auf *Modellparameter* bzw. kurz *Parameter* festgelegt.

(101) **Pearsonsche Testfunktion*:** Die *Pearsonsche Testfunktion*

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(H_i - np_i)^2}{np_i}$$

wird bei χ^2-Tests verwendet, weil sie asymptotisch eine χ^2-Verteilung mit $k-1$ Freiheitsgraden besitzt.

(102) **Permutation:** Die Anordnungen einer n-elementigen Menge nennt man ihre *Permutationen*. Es gibt $n!$ dieser Anordnungen.

(103) **Pivotgröße:** Eine stochastische Größe $\mathrm{Q}(\theta, X_1, \ldots, X_n)$, deren Verteilung unabhängig vom Parameter θ ist, heißt *Pivotgröße*.

(104) **Produktraum:** Ein mehrdimensionaler Merkmalsraum kann als Produktraum über eindimensionale Merkmalsräume aufgefasst werden.

(105) **Quadratische Konvergenz:** Unter *quadratischer Konvergenz* (auch *L_2-Konvergenz*) versteht man, dass der quadratische Abstand gegen 0 konvergiert, i. Z.

$$\lim_{n\to\infty} \mathbb{E}(X_n - X)^2 = 0.$$

(106) **Quantil:** Das *α-Quantil* (das *Fraktil* bezeichnet jenen Wert x_α, für den

$$\mathrm{F}(x_\alpha) = \mathbf{P}(X \leq x_\alpha) = \alpha$$

erfüllt ist. Diese Gleichung muss bei diskreten Verteilungen keine Lösung haben; dann definiert man als α-Quantil den kleinsten Wert j, für den $\mathrm{F}(j) \geq \alpha$ gilt.

(107) **Quantilsfunktion:** Weil bei diskreten Verteilungen die Gleichung

$$\mathrm{F}(x_\alpha) = \mathbf{P}(X \leq x_\alpha) = \alpha$$

keine Lösung x_α besitzen muss, verwendet man als *Quantilsfunktion* die verallgemeinerte Inverse

$$\mathrm{F}^{-1}(y) = \inf\{x | \mathrm{F}(x) \geq y\}.$$

(108) **Randdichte:** Bei stetigen Verteilungen mehrerer stochastischer Variabler $X_1, \ldots, X_n$ mit Verteilungsfunktion $f(x_1, \ldots, x_n)$ bezeichnet

$$\mathrm{f}(z) = \int \cdots \int \mathrm{f}(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

die Dichte der Einzelverteilung von X_i und wird *Randdichte* genannt.

(109) **Randverteilung:**
- Ist $(X_i)_{i=1}^n$ ein stochastischer Vektor diskreter Größen über einem Merkmalsraum M, so gilt für die Einzelwahrscheinlichkeit von X_i

$$\mathbf{P}[X_i = z] = \sum_{\mathbf{x}\in E_i} \mathbf{P}[\mathbf{X} = \mathbf{x}],$$

  wobei $E_i = \{\mathbf{x} \in M | x_i = z\}$ gilt. Diese Verteilung wird *Randverteilung* genannt.
- Ist $(X_i)_{i=1}^n$ ein stochastischer Vektor stetiger Größen über $\mathbb{R}^n$, so gilt für die Randverteilung der stochastischen Größe X_i

$$\mathrm{F}_{X_i}(z) = \mathbf{P}[X_i \leq z] = \mathbf{P}[X_1 < \infty, \ldots, X_{i-1} < \infty, X_i < z, X_{i+1} < \infty, \ldots, X_n < \infty] = \mathrm{F}(\infty, \ldots, \infty, z, \infty, \ldots, \infty),$$

  wobei $T(\infty) = \lim_{x\to\infty} T(x)$ bezeichnen soll.

(110) **Realisierung:** Nach der Durchführung eines Experiments, das durch die stochastische Größe X beschrieben wird, liegt die *Realisierung* (das *Ergebnis*) x des Experiments vor. x ist *keine* stochastische Größe mehr.

(111) **Regression – erster Art:** Die Funktion $y \mapsto \mathbb{E}(X|Y = y)$ wird als *Regressionsfunktion erster Art* bezeichnet. Es gilt, dass der bedingte Erwartungswert unter $Y = y$

$$\mathbb{E}(X - V(y))^2$$

für $V(y) = \mathbb{E}(X|Y = y)$ minimal wird.

(112) **Regression – lineare*:** Das Verfahren der *linearen Regression* stellt eine Möglichkeit dar, einen Zusammenhang zwischen zwei stochastischen Größen zu modellieren.

(113) **Regression – multiple*:** Wird bei einer Regression nicht nur ein Regressor, sondern mehrere verwendet, so spricht man von einem *multiplen Regressionsmodell*.

(114) **Regressor*:** Wenn im Rahmen einer Regression für die stochastischen Größen X und Z ein Zusammenhang der Form $X = Q(z) + \epsilon$ modelliert wird, d. h. man mit Hilfe des Werts $Z = z$ den Wert von X prognostizieren möchte, dann bezeichnet man Z als den *Regressor*.

(115) **Residuum*:** Man bezeichnet die Differenzen
$$\mathrm{res}_i = x_i - \theta_1 - \theta_2 z_i$$
als *Residuen*.

(116) **Satz von – der vollständigen Wahrscheinlichkeit:** Für Hypothesen H_i mit $\mathbf{P}(H_i) > 0$ gilt für jedes Ereignis A
$$\mathbf{P}(A) = \sum_i \mathbf{P}(A|H_i)\,\mathbf{P}(H_i).$$

(117) **Satz von – Glivenko-Cantelli:** Die empirische Verteilungsfunktion werde aus der Stichprobe $X_1, \ldots, X_n$ mit $X_i \sim \mathrm{F}(.)$ berechnet. Dann konvergiert der Abstand
$$D_n := \sup_{x \in \mathbb{R}} |\mathrm{F}_n^*(x) - \mathrm{F}(x)|$$
fast sicher gegen 0, i. Z.
$$\mathbf{P}(\lim_{n \to \infty} D_n = 0) = 1.$$

(118) **Schätzung:** Als *Schätzung* (kurz für *Parameterschätzung*) bezeichnet man das Berechnen fehlender Parameter aus Messdaten.

(119) **Schiefe:** Als *Schiefe* der Verteilung wird
$$\frac{\mathbb{E}(X - \mathbb{E}\,X)^3}{\mathrm{var}(X)^{3/2}}$$
bezeichnet. Als *empirische Schiefe* berechnet man daher
$$\frac{1}{n}\sum_{i=1}^{n} \frac{(x_i - \overline{x})^3}{s^3}.$$

(120) **Schwarzsche Ungleichung:** Die *Schwarzsche Ungleichung* besagt, dass allgemein für stochastische Größen X, Y
$$|\mathbb{E}(XY)| \le \sqrt{\mathbb{E}\,X^2}\sqrt{\mathbb{E}\,Y^2}$$
gilt; Gleichheit genau bei linear abhängigen Größen.

(121) **Spannweite:** Die *Spannweite* S einer Stichprobe ist die Differenz zwischen ihrem größten und kleinsten Wert; mit Hilfe der Ordnungsstatistik gilt
$$S = X_{(n)} - X_{(1)}.$$

(122) **Standardabweichung – empirische:** Die *empirische Standardabweichung* (oder *empirische Streuung*) ist definiert durch
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.}$$
Hier gilt wieder der Verschiebungssatz, und so kann die Summe durch
$$\sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - n\overline{x}$$
oft einfacher berechnet werden.

(123) **Standardnormalverteilung:** Die *Standardnormalverteilung* $\mathrm{N}(0, 1^2)$ nimmt eine Sonderrolle unter den Normalverteilungen ein. Ihre Dichte wird mit $\phi(.)$, ihre Verteilungsfunktion mit $\Phi(.)$ bezeichnet. Die Werte der Verteilungsfunktion sind tabelliert. Die allgemeine Dichte $\mathrm{f}(.)$ einer Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ kann über die lineare Transformation
$$\mathrm{f}(x) = \frac{1}{\sigma}\,\phi(\frac{x-\mu}{\sigma}),$$
ihre allgemeine Verteilungsfunktion $\mathrm{F}(.)$ über die lineare Transformation
$$\mathrm{F}(x) = \Phi(\frac{x-\mu}{\sigma})$$
auf die Standardnormalverteilung zurückgeführt werden.

(124) **Statistik:** Eine messbare Funktion $\mathrm{T} : \Omega^n \to \mathbb{R}^k, (X_1, \ldots, X_n) \mapsto \mathrm{T}(X_1, \ldots, X_n)$ mit $k \ge 1$ heißt *Statistik*.

(125) **Stetigkeitskorrektur:** Wird eine Binomialverteilung aufgrund des Zentralen Grenzverteilungssatzes mit Hilfe einer Normalverteilung genähert, so können die Integralgrenzen noch am Rand um 0.5 angepasst werden, um der Tatsache, dass ansonsten eine Einheit weniger beachtet werden würde, Rechnung zu tragen. Dadurch verschwindet der Unterschied zwischen exakter Berechnung und Näherung fast gänzlich.

(126) **Stichprobe:** Der stochastische Vektor $(X_1, \ldots, X_n)$ heißt *Stichprobe* der stochastischen Größe X, wenn alle X_i dieselbe Verteilung wie X besitzen und unabhängig sind.

(127) **Stichprobenkorrelationskoeffizient:** Der *Stichprobenkorrelationskoeffizient* ist definiert als
$$\hat{\rho} = \frac{s_{12}}{\sqrt{s_1}\sqrt{s_2}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y}_n)^2}}.$$

(128) **Stichprobenmedian:** Für das 0.5-Quantil (den Median) wird üblicherweise die Schätzung $\alpha_{0.5}^*$ etwas verändert. Anstelle von $\mathrm{F}_n^*(x)$ wird eine strikt monoton wachsende Schätzung der Verteilungsfunktion (Polygonzug) verwendet. Diese Statistik heißt *Stichprobenmedian* und ist gegeben durch
$$X_{\mathrm{med}} = \begin{cases} \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} & \text{für } n \text{ gerade} \\ X_{(\frac{n+1}{2})} & \text{für } n \text{ ungerade} \end{cases}$$

(129) **Stichprobenmittel:** Das *Stichprobenmittel* $\overline{X}_n$ ist definiert durch
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(130) **Stichprobenmoment:** Das *k-te Stichprobenmoment* m_k ist definiert durch
$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k.$$

(131) **Stichprobenvarianz:** Die *Stichprobenvarianz* s^2 ist definiert durch
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$
(132) **Stirlingsche Formel:** Die *Stirlingsche Formel* besagt, dass
$$n! \sim \sqrt{2\pi n}\, n^n e^{-n},$$
d. h.
$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} = 1.$$
(133) **stochastisch unabhängig:**
- Zwei Ereignisse A, B sind genau dann *stochastisch unabhängig*, wenn gilt
$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B).$$
- Eine Familie von stochastischen Größen $(X_i)_{i\in I}$ heißt *stochastisch unabhängig*, wenn jede Familie von Ereignissen der Form $(\{X_i \in E_i\})_{i\in I}$ für beliebige Ereignisse E_i unabhängig ist.

(134) **stochastische Größe:** Eine Funktion $X : \Omega \to \mathbb{R}$, die einen Merkmalsraum auf die reellen Zahlen abbildet, heißt *stochastische Größe.*
(135) **stochastische Größe – standardisierte:** Ist X eine stochastische Größe mit Erwartung μ und Varianz σ^2, dann heißt
$$X_0 = \frac{X-\mu}{\sigma}$$
die *standardisierte stochastische Größe* von X. Es gilt $\mathbb{E}\,X_0 = 0$ und $\operatorname{var}(X_0) = 1$.
(136) **stochastische Größe – stetige:** Wird die Verteilung einer stochastischen Größe X durch eine nichtnegative, integrierbare Funktion $\mathrm{f}(.)$ mit $\int_{-\infty}^{\infty} \mathrm{f}(x)dx = 1$ beschrieben, so nennt man $\mathrm{f}(.)$ Dichte der Verteilung und X eine *stetige stochastische Größe.*
(137) **stochastische Konvergenz:** Die Folge stochastischer Größen $(X_n)_{n\in\mathbb{N}}$ *konvergiert stochastisch* (*in der Wahrscheinlichkeit*) gegen die stochastische Größe X, wenn
$$\lim_{n\to\infty} \mathbf{P}[|X_n - X| > \epsilon] = 0$$
für jedes $\epsilon > 0$ gilt. Man schreibt dann auch $X_n \xrightarrow{P} X$.
(138) **stochastischer Vektor:** Ein *stochastischer Vektor* (eine *multivariate stochastische Größe*) $\mathbf{X} = (X_1, \dots, X_n)$ ist eine Abbildung
$$\mathbf{X} : \Omega \to \mathbb{R}^n,$$
wobei das Urbild jedes n-dimensionalen Intervalls
$$I = (a_1, b_1] \times \cdots \times (a_n, b_n]$$
ein Ereignis im ursprünglichen Wahrscheinlichkeitsraum sein soll, i. Z. $\mathbf{X}^{-1}(I) \in \mathcal{E}$.
(139) **Streuung:** Die *Streuung* (oder *Standardabweichung*) ist definiert als die Wurzel der Varianz, i. Z.
$$\text{Streuung}(X) = \sqrt{\operatorname{var}(X)}.$$
(140) **Streuung – empirische:** Die *empirische Streuung* (oder *empirische Standardabweichung*) ist definiert durch
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$
Hier gilt wieder der Verschiebungssatz, und so kann die Summe durch
$$\sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - n\overline{x}$$
oft einfacher berechnet werden.
(141) **Test:** Ein (nichtrandomisierter) statistischer *Test* für H_0 und H_1 ist eine Statistik $\mathrm{T} : \mathbf{x} \mapsto \mathrm{T}(\mathbf{x}) \in \{0, 1\}$. Der *Verwerfungsraum* V ist eine Teilmenge des Stichprobenraums, $V = \{\mathbf{x}|\, \mathrm{T}(\mathbf{x}) = 1\}$. $V^c = A = \{\mathbf{x}|\, \mathrm{T}(\mathbf{x}) = 0\}$ heißt *Annahmeraum.*
(142) **Test – Bayes-:** In einem *Bayes-Test* wird für die Hypothese mit der höheren a-posteriori-Wahrscheinlichkeit entschieden.
(143) **Test – beidseitiger:** Ein Test der Form $H_0 : \theta = \theta_0$ gegen $H_1 : \theta \neq \theta_0$ wird als *beidseitiger Test* bezeichnet.
(144) **Test – einseitiger:** Ein Test der Form $H_0 : \theta \leq \theta_0$ gegen $H_1 : \theta > \theta_0$ wird als *einseitiger Test* bezeichnet.
(145) **Test – F-*:** Beim F-*Test* wird die Hypothese gleicher Varianzen zweier normalverteilter stochastischer Größen dann abgelehnt, wenn der Quotient der Stichprobenvarianzen außerhalb des Intervalls der F-Fraktile liegt.
(146) **Test – t-*:** Bei unbekannter Varianz kann nach einer normalverteilten Stichprobe entschieden werden, ob die Hypothese $H_0 : \mu \leq \mu_0$ anzunehmen ist.
(147) **Test – Zweistichproben-:** Bei einem *Zweistichprobentest* werden Modellannahmen, die sich auf zwei Stichproben gemeinsam beziehen, geprüft.
(148) **Teststatistik:** Eine im Rahmen eines statistischen Tests berechnete Datenkennzahl nennt man *Teststatistik.*
(149) **Transformation:** Eine *Transformation* einer stochastischen stochastischen Größe X führt auf eine neue stochastische Größe Y. Die Wahrscheinlichkeitsverteilung von Y folgt aus der von X. Zu ihrer Zurückführung muss die Inverse der Transformation verwendet werden.
(150) **Tschebyscheffsche Ungleichung:** Für eine stochastische Größe X mit existierender Varianz gilt für jedes $\epsilon > 0$
$$\mathbf{P}[|X - \mathbb{E}\,X| \geq \epsilon] \leq \frac{\operatorname{var}(X)}{\epsilon^2}.$$
(151) **unverzerrt:** Eine Schätzfunktion T, die
$$\mathbb{E}\,T = \theta$$
für alle θ erfüllt, heißt *unverzerrt* (*erwartungstreu*) für θ.
(152) **Varianz:** Unter der *Varianz* einer stochastischen Größe X mit $\mathbb{E}\,X^2 < \infty$ versteht man den Erwartungswert
$$\operatorname{var}(X) = \mathbb{E}(X - \mathbb{E}\,X)^2.$$
(153) **Varianz – verallgemeinerte:** Die Determinante der Kovarianzmatrix ist eine Maßzahl dafür, wieviel "Unsicherheit" im gesamten Vektor $\mathbf{X}$ enthalten ist, sie heißt daher *verallgemeinerte Varianz.*
(154) **Varianzanalyse*:** Bei einer *Varianzanalyse* werden mehrere Datenreihen auf gemeinsame Mittelwerte überprüft.

(155) **Variation:** Eine *Variation ohne Wiederholung* ist die Anzahl der Möglichkeiten, k Elemente von n auszuwählen und sie beliebig anzuordnen. Die Anzahl ist $\frac{n!}{(n-k)!}$.

(156) **Variationskoeffizient:** Der *Variationskoeffizient* v bezeichnet das Verhältnis $v = \frac{s}{\overline{x}}$, wobei s die empirische Standardabweichung und $\overline{x} \neq 0$ das Stichprobenmittel bezeichnet.

(157) **Verschiebungssatz:** Der *Steinersche Verschiebungssatz* besagt, dass für die Varianz

$$\operatorname{var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2$$

gilt. Er folgt direkt aus der Tatsache, dass der Erwartungswert linear ist.

(158) **Verteilung – Alternativ-:** Hat ein Versuch nur zwei mögliche Ausgänge, $X = 0$ und $X = 1$, wobei $\mathbf{P}(X = 1) = p$ gilt, so nennt man $X \sim \mathrm{A}$ *alternativverteilt.* Für den Erwartungswert gilt $\mathbb{E}\, X = p$, für die Varianz $\operatorname{var}(X) = p(1-p)$.

(159) **Verteilung – bedingte:** Bei diskreten Verteilungen wird die *bedingte Dichte* definiert als

$$\mathbf{P}[X = i | Y = j] = \frac{\mathbf{P}[X = i, Y = j]}{\mathbf{P}[Y = j]}.$$

Bei einer stetigen Verteilung einer stochastischen Größe (X, Y) mit gemeinsamer Dichte $\mathrm{f}(x, y)$ und $\mathrm{f}_Y(.)$ der Randdichte von Y definiert man als *bedingte Dichte* von X gegeben $Y = y$

$$\mathrm{f}(x|y) = \frac{\mathrm{f}(x, y)}{\mathrm{f}_Y(y)},$$

vorausgesetzt $f_Y(y) > 0$. Diese Dichte erklärt die bedingte Verteilung $X|Y = y$.

(160) **Verteilung – Beta-:** Eine stochastische Größe X ist *Beta-verteilt* $\beta(a, b)$ mit $a, b > 0$, wenn die Dichte von X durch

$$\mathrm{f}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

für $0 < x < 1$ gegeben ist. Die stetige Gleichverteilung ist der Fall $a = b = 1$.

(161) **Verteilung – Binomial-:** Eine *Binomialverteilung* $X \sim \mathrm{B}_{n,p}$ entsteht, wenn n unabhängige Alternativversuche $X_i \sim \mathrm{A}_p$ durchgeführt und die Erfolge $X_i = 1$ gezählt werden. Für die Verteilung gilt

$$\mathbf{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

für $i = 0, 1, \ldots, n$; $0 \leq p \leq 1$. Der Erwartungswert ergibt sich über $\mathbb{E}\, X = \mathbb{E}(\sum_{i=1}^n X_i) = np$, die Varianz aufgrund der Unabhängigkeit der X_i durch $\operatorname{var}(X) = \operatorname{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \operatorname{var}(X_i) = np(1-p)$.

(162) **Verteilung – Cauchy-:** Die *Cauchy-Verteilung* $X \sim \mathrm{C}_{\mu,\sigma}$ mit der Dichte

$$\mathrm{f}(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x-\mu)^2}$$

mit $\sigma > 0$ besitzt keine Momente. Die t-Verteilung mit Freiheitsgrad 1 entspricht einer Cauchy-Verteilung. Die charakteristische Funktion ist

$$\varphi_X(t) = e^{\mathrm{i}\mu t - |t\sigma|}.$$

Aufgrund der Nichtexistenz der Momente gilt für die *Cauchy-Verteilung* auch nicht das Gesetz der großen Zahlen; $\overline{X}_n$ ist genauso verteilt wie X_1.

(163) **Verteilung – Chiquadrat-:** Die χ^2_k-*Verteilung*, die χ^2-*Verteilung* mit $k \in \mathbb{N}$ Freiheitsgraden, ist eine spezielle Gamma-Verteilung mit $a = k/2$ und $b = 1/2$. Die Dichte ist demnach

$$\mathrm{f}(x) = \frac{1}{\Gamma(k/2) 2^{k/2}} x^{\frac{k}{2}-1} e^{-x/2} \quad \text{für } x > 0.$$

(164) **Verteilung – Dirac-:** Die *Dirac-Verteilung* $X \sim \delta_x$ ist jener Extremfall der Alternativverteilung, bei dem $\mathbf{P}(X = x) = 1$ gilt. Sie kann verwendet werden, um deterministische Größen im Rahmen der Wahrscheinlichkeitsrechnung zu berücksichtigen.

(165) **Verteilung – empirische:** Eine diskrete Verteilung auf den verschiedenen Werten $y_1, \ldots, y_m$ des Beobachtungsvektors $x_1 \ldots, x_n$ mit den relativen Häufigkeiten als Wahrscheinlichkeiten heißt *empirische Verteilung.* Die Verteilungsfunktion dieses Wahrscheinlichkeitsmaßes ist durch

$$\mathrm{F}^*_n(x) = \frac{1}{n} \sum_{i=1}^n \mathrm{I}_{(-\infty,x]}(x_i)$$

gegeben und heißt *empirische Verteilungsfunktion.*

(166) **Verteilung – Exponential-:** Die *Exponentialverteilung* $X \sim \mathrm{Ex}_\lambda$ ist eine spezielle Gamma-Verteilung mit $a = 1$ und $b = \lambda$ und besitzt daher die Dichte

$$\mathrm{f}(x) = \lambda e^{-\lambda x} \quad \text{für } x > 0.$$

Die Verteilungsfunktion ist

$$\mathrm{F}(x) = 1 - e^{-\lambda x} \quad \text{für } x > 0.$$

Der Erwartungswert ist $\mathbb{E}\, X = \frac{1}{\lambda}$, die Varianz $\operatorname{var} X = \frac{1}{\lambda^2}$.

(167) **Verteilung** – F-**:** Die F-*Verteilung* mit $m, n \in N$ Freiheitsgraden hat die Dichte

$$\mathrm{f}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\,\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{(1 + \frac{xm}{n})^{\frac{m+n}{2}}} \quad \text{für } x > 0,$$

es gilt $\mathbb{E}\, X = \frac{n}{n-2}$ für $n > 2$ und

$$\operatorname{var}(X) = 2\left(\frac{n}{n-2}\right)^2 \frac{m+n-2}{m(n-4)} \quad \text{für } n > 4.$$

Die momenterzeugende Funktion existiert nicht, die Momente sind

$$\mathbb{E}\, X^k = \frac{\Gamma(\frac{m+2k}{2})\,\Gamma(\frac{n-2k}{2})}{\Gamma(\frac{m}{2})\,\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^k \quad \text{für } k < \frac{n}{2}.$$

(168) **Verteilung – Gamma-:** Eine stochastische Größe X ist *Gamma-verteilt* $X \sim \gamma(a,b)$ mit $a,b>0$, wenn die Dichte von X durch

$$\mathrm{f}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \quad \text{für } x > 0$$

gegeben ist.

Der Erwartungswert ist $\mathbb{E}\, X = \frac{a}{b}$, die Varianz $\operatorname{var}(X) = \frac{a}{b^2}$.

(169) **Verteilung – Geometrische:** Die Anzahl der Bernoulli-Versuche bis zum ersten Erfolg ist *geometrisch verteilt* $X \sim \mathrm{G}_p$ mit

$$\mathbf{P}(X = i) = p(1-p)^{i-1}$$

für $i = 1, 2, \ldots; 0 \le p \le 1$. Die Verteilungsfunktion ist

$$\mathrm{F}(x) = \sum_{i=1}^{\lfloor x \rfloor} p(1-p)^{i-1} = 1 - (1-p)^{\lfloor x \rfloor}.$$

Der Erwartungswert ist $\mathbb{E}\, X = \frac{1}{p}$, die Varianz $\operatorname{var}(X) = \frac{1-p}{p^2}$.

(170) **Verteilung – Hypergeometrische:** Bei einer Ziehung ohne Zurücklegen, wobei M von N Objekten markiert sind, ist die Anzahl der gezogenen markierten Objekte $X \sim \mathrm{H}_{N,M,n}$ *hypergeometisch verteilt* mit

$$\mathbf{P}(X = i) = \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}.$$

Der Erwartungswert ist $\mathbb{E}\, X = \frac{nM}{N}$, die Varianz $\operatorname{var}(X) = \frac{nM}{N}\frac{(N-M)(N-n)}{N(N-1)}$.

(171) **Verteilung – Multinomial-:** Gibt es bei einem Experiment n verschiedene Ausgänge, so kann bei wiederholter Ausführung im Vektor $\mathbf{X} = (X_1, \ldots, X_n)$ die Anzahl der Ergebnisse einer Stichprobe gefunden werden. Für die Punktwahrscheinlichkeiten ergibt sich, wenn p_i die Wahrscheinlichkeit des i-ten Ergebnisses bei einem einzelnen Versuch ist,

$$\mathbf{P}[X_1 = x_1, \ldots, X_n = x_n] = \binom{N}{x_1, \ldots, x_n} p_1^{x_1} \cdots p_n^{x_n} = \frac{N!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n},$$

wobei $x_1 + \cdots + x_n = N$ die Anzahl der Versuche ist. Diese Verteilung heißt *Multinominalverteilung*, i. Z. $\mathbf{X} \sim \mathrm{M}_{N;p_1,\ldots,p_n}$. X_n ergibt sich aus den vorigen Ergebnissen, die Dimension ist daher eigentlich $n-1$. Für $n=2$ liegt eine Binomialverteilung vor.

(172) **Verteilung – Negative Binomial-:** Die geometrische Verteilung lässt sich insofern verallgemeinern, dass nicht mehr nur die Anzahl der Versuche bis zum ersten, sondern bis zum k-ten Erfolg gezählt werden sollen. In dem Fall ist $X \sim \mathrm{NB}_{m,p}$ *negativ binomialverteilt* mit

$$\mathbf{P}(X = i) = \binom{i-1}{m-1} p^m (1-p)^{i-m} \quad \text{für } i \ge m.$$

(173) **Verteilung – Poisson-:** Die Grenzverteilung einer Binomial- und damit auch der hypergeometrischen Verteilung nennt man *Poisson-Verteilung.* Für eine poisson-verteilte stochastische Größe $X \sim \mathrm{P}_\lambda$ gilt

$$\mathbf{P}(X = i) = \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{für } i = 0, 1, \ldots; 0 < \lambda.$$

Die Poisson-Verteilung wird häufig als Zählverteilung verwendet, d. h. für die Zählung von Einzelereignissen innerhalb eines Intervalls.

(174) **Verteilung – t-:** Die *Student-* oder t-*Verteilung* mit $k \in \mathbb{N}$ Freiheitsgraden $X \sim \mathrm{t}_k$ ist durch

$$\mathrm{f}(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \left[1 + \frac{x^2}{k}\right]^{-\frac{k+1}{2}}$$

definiert. Für den Erwartungswert gilt $\mathbb{E}\, X = 0$ und für die Varianz $\operatorname{var}(X) = \frac{k}{k-2}$ für $k > 2$. Die momenterzeugende Funktion existiert nicht, die geraden Momente sind

$$\mathbb{E}\, X^{2l} = \frac{\Gamma(l + \frac{1}{2})\,\Gamma(\frac{k}{2} - l)}{\sqrt{\pi}\,\Gamma(\frac{k}{2})} k^l$$

für $l < \frac{k}{2}$, ungerade Momente verschwinden.

(175) **Verteilungsfunktion:** Die *Verteilungsfunktion* wird für jedes x als Wahrscheinlichkeit de Ereignisses $E = [X \le x]$ definiert und berechnet sich als die Summe der Punktwahrscheinlichkeiten von Punkten, die kleiner oder gleich x sind, i. Z.

$$\mathrm{F}(x) := \mathbf{P}(X \le x),$$

d. h. für diskrete Verteilungen $\mathrm{F}(x) = \sum_{i \le x} p_i$ und für stetige $\mathrm{F}(x) = \int_{-\infty}^{x} \mathrm{f}(t)dt$.

(176) **Verteilungsfunktion – empirische:** Die *empirische Verteilungsfunktion* ist die Verteilungsfunktion der empirischen Verteilung und ergibt sich als

$$\mathrm{F}_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{(-\infty,x]}(x_i).$$

(177) **Verteilungsfunktion – mehrdimensionale:** Die *n-dimensionale Verteilungsfunktion* an der Stelle $(x_1, \ldots, x_n)$ ist die Wahrscheinlichkeit

$$\mathrm{F}(x_1, \ldots, x_n) = \mathbf{P}[X_1 \le x_1, \ldots, X_n \le x_n].$$

(178) **Verteilungskonvergenz:** Eine Folge (X_n) von stochastischen Größen konvergiert gegen X *in der Verteilung* (oder *schwach*), wenn für die Verteilungsfunktionen von $X_n \sim \mathrm{F}_n$ und $X \sim \mathrm{F}$

$$\lim_{n \to \infty} \mathrm{F}_n(x) = \mathrm{F}(x)$$

für jeden Stetigkeitspunkt x von F gilt. Die Verteilungskonvergenz wird durch $X_n \xrightarrow{D} X$ bezeichnet.

(179) **Vertrauensbereich:** Ein *Vertrauensbereich* (oder *Konfidenzbereich*) $\mathrm{C}(\mathbf{x})$ für $\theta \in \Theta$ ist eine Abbildung

$$\mathbf{x} \to \mathrm{C}(\mathbf{x}) \subseteq \Theta$$

für jedes $\mathbf{x}$ aus dem Stichprobenraum. $\mathrm{C}(\mathbf{x})$ ist ein Konfidenzbereich mit *Überdeckungswahrscheinlichkeit* β (oder zum *Niveau* β), wenn

$$\mathbf{P}_\theta(\theta \in \mathrm{C}(\mathbf{X})) = \beta$$

für alle θ gilt.

(180) **Verwerfung:** Entscheidet man nach Betrachten einer Stichprobe, dass die Hypothese H_0 nicht gerechtfertigt war, so bezeichnet man das als *Verwerfung* von H_0.

(181) **Verwerfungsraum:** Der *Verwerfungsraum* V ist eine Teilmenge des Stichprobenraums

$$V = \{\mathbf{x} |\, \mathrm{T}(\mathbf{x}) = 1\}.$$

(182) **Wahrscheinlichkeit:** Intuitiv ergibt sich die Wahrscheinlichkeit $\mathbf{P}(.)$ aus der Annahme, dass die relativen Häufigkeiten $\mathrm{h}_n(.)$ gegen einen festen Wert konvergieren, der dann mit $\mathbf{P}(E) := \lim_{n\to\infty} \mathrm{h}_n(E)$ bezeichnet wird. Die Sinnhaftigkeit dieses Ansatzes ergibt sich aus dem Gesetz der großen Zahlen.

Eine Mengenfunktion $\mathbf{P}(.)$ auf einem Ereignisfeld $\mathcal{E}$ mit dem Merkmalsraum Ω heißt eine *Wahrscheinlichkeitsfunktion*, wenn $0 \leq \mathbf{P}(E) \leq 1$ für $E \in \mathcal{E}$, $\mathbf{P}(\Omega) = 1$ und für jede Folge einander ausschließender Ereignisse A_i mit $A_i \cap A_j = \emptyset$ für $i \neq j$

$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$$

gilt.

(183) **Wahrscheinlichkeit – a-posteriori-:** Mit Hilfe der Bayesschen Formel gelangt man von den a-priori-Wahrscheinlichkeiten $\mathbf{P}(H_i)$ zu den *a-posteriori-Wahrscheinlichkeiten* $\mathbf{P}(H_i|A)$.

(184) **Wahrscheinlichkeit – a-priori-:** Bevor eine Beobachtung vorliegt, gelten für die Hypothesen H_i die *a-priori-Wahrscheinlichkeiten* $\mathbf{P}(H_i)$.

(185) **Wahrscheinlichkeit – bedingte:** Die *bedingte Wahrscheinlichkeit* vom Ereignis A unter dem Ereignis B mit positiver Wahrscheinlichkeit ist

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

(186) **Wahrscheinlichkeit – geometrische:** Ein geometrisches Objekt mit Fläche (bzw. Länge oder Volumen) $\mathrm{A}(\Omega) = 1$ lässt sich sehr gut verwenden, um Wahrscheinlichkeiten zu definieren. Für Teilmengen E von Ω, die ebenfalls eine Fläche $\mathrm{A}(E)$ besitzen, wird als Wahrscheinlichkeit

$$\mathbf{P}(E) = \frac{\mathrm{A}(E)}{\mathrm{A}(\Omega)}$$

definiert. Flächen erfüllen die für Wahrscheinlichkeiten geforderte Additivität. Diese Interpretation wird als *geometrische Wahrscheinlichkeit* bezeichnet und ergibt eine Gleichverteilung auf dem Merkmalsraum. Es ermöglicht einen intuitiven Zugang zu einigen Problemen.

(187) **Wahrscheinlichkeitsfunktion:** Eine Mengenfunktion $\mathbf{P}(.)$ auf einem Ereignisfeld $\mathcal{E}$ mit dem Merkmalsraum Ω heißt eine *Wahrscheinlichkeitsfunktion*, wenn $0 \leq \mathbf{P}(E) \leq 1$ für $E \in \mathcal{E}$, $\mathbf{P}(\Omega) = 1$ und für jede Folge einander ausschließender Ereignisse A_i mit $A_i \cap A_j = \emptyset$ für $i \neq j$

$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$$

gilt.

Jede nichtnegative, normierte und additive Funktion eignet sich als Wahrscheinlichkeitsfunktion.

(188) **Wahrscheinlichkeitsraum:** Das Tripel $(\Omega, (E), \mathbf{P})$ bezeichnet man als *Wahrscheinlichkeitsraum*.

(189) **Wahrscheinlichkeitsraum – diskreter:** Ein *diskreter Wahrscheinlichkeitsraum* enthält einen höchstens abzählbaren Merkmalsraum.

(190) **Wechselwirkung*:** Bei der Varianzanalyse können *Wechselwirkungen* der einzelnen Faktoren auftreten.

(191) **Zentraler Grenzverteilungssatz:** Der *zentrale Grenzverteilungssatz* besagt, dass für X eine stochastische Größe mit dem Erwartungswert μ und der Varianz $\sigma^2 > 0$ und $X_1, \ldots, X_n$ eine Stichprobe von X

$$\sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n}\sigma} \xrightarrow{D} Z$$

mit $Z \sim \mathrm{N}(0, 1)$ gilt.

Es reichen auch schwächere Voraussetzungen; so müssen die X_i nicht dieselbe Verteilung haben, lediglich $0 < \delta < \mathrm{var}(X_i) < M < \infty$ für alle i reicht bereits aus. Eine allgemeine Formulierung besagt, dass

$$S = X_1 + \cdots + X_n$$

mit X_i und der gerade gegebenen Bedingung für $n \to \infty$ annähernd normalverteilt ist. Der Erwartungswert dieser Normalverteilung ist $\mathbb{E}\,S$ und die Varianz ist $\mathrm{var}(S)$.