Grundlegende Definitionen in der Wahrscheinlichkeitstheorie

Begriff	Definition / Bezeichnung	Zusammenhang / Eigenschaften / Beispiele
Ergebnismenge Ω	(Endlich) abzählbare oder überabzählbare Menge aller (Elementar-) Ergebnisse $\omega \in \Omega$	z. B. $\Omega = [0,1], \Omega = \{0,1\}, \Omega = \mathbb{R}, \Omega = \mathbb{R}^+$, $\Omega = \{rot, gelb, blau, schwarz\}$
Ereignis A	Beliebige Teilmenge $A \subset \Omega$ bzw. Element der Potenzmenge von Ω: $A \in \mathcal{P}(\Omega)$	z. B. $A = \{\omega\}, A = \emptyset, A = \Omega, A = \{rot, gelb, blau\} \subset \Omega$
Zufallsvariable (ZV) X	Eine Zufallsvariable ist eine Funktion $X : \Omega \mapsto \mathbb{R}$; $X(\omega) \in \mathbb{R}$ heißt Realisation; Ist diese Funktion stetig (diskret), so nennt man die Zufallsvariable stetig (diskret)	z. B. Münzwurf: $X(\text{Kopf}) = 1, X(\text{Zahl}) = 0$ ist diskret; Temperatur: $X(\text{Temperatur in Celsius}) \in \mathbb{R}$ wäre stetig (indem man jeder Grad Celsius Zahl deren Skalar zuordnet)
Wahrscheinlichkeitsfunktion P (probability function)	Funktion $P : \mathcal{F} \mapsto [0,1]$ wobei $\mathcal{F}$ eine „geeignete" Menge von Ereignissen ist und P folgende Eigenschaften erfüllt: 1. $P(A) \geq 0 \quad \forall A \in \mathcal{F}$ (ungenauer $A \subset \Omega$) 2. $P(\Omega) = 1$ und $P(\emptyset) = 0$ 3. Wenn $A_1, A_2, \ldots$ paarweise disjunkt sind, dann gilt: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$	Statt $P(\{\omega \in \Omega \mid X(\omega) = x\})$ kann man auch $P(X(\omega) = x)$ oder $P(X = x)$ schreiben. Im Fall einer stetigen ZV gilt $P(X = x) = 0$ $\mathcal{F}$ nennt man auch Sigmaalgebra. Dies ist grob gesagt eine Teilmenge der Potenzmenge mit besonderen Eigenschaften.
Verteilungsfunktion F_X (cumulative density function CDF) einer Zufallsvariable X (Univariater Fall)	Funktion $F : \mathcal{F} \mapsto [0,1]$ definiert durch $F_X(x) := P(X \leq x)$ mit den Eigenschaften: 1. $\lim_{x\to-\infty} F_X(x) = 0$ und $\lim_{x\to\infty} F_X(x) = 1$ 2. $F_X(x)$ ist monoton steigend und rechtsseitig stetig 3. $P(a < X \leq b) = F_X(b) - F_X(a)$ 4. Falls X stetig ist, gilt: $F_X(x) = P(X \leq x) = P(X < x)$	Diskreter Fall: $F_X(x) = \sum_{x_i \leq x} P(X = x_i)$ Da im stetigen Fall $P(X = x) = 0 \quad \forall x \in \mathbb{R}$ gilt, kann man den Begriff der Dichte einführen und die Wahrscheinlichkeit als Fläche unter dieser Dichte $f_X(x)$ betrachten: Stetiger Fall: $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ (in gewisser Weise verallgemeinert das Integral die unendliche Summe von abzählbare Mengen auf überabzählbare)
Dichtefunktion und Verteilungsfunktion (nur im stetigen Fall!)	Die Dichtefunktion ist die Ableitung der Verteilung, d. h. $f_X(x) := \frac{dF_X(x)}{dx}$ bzw. umgekehrt ist die Verteilung das Integral der Dichte (eindeutig mit $F_X(„\infty") = 1$), also: $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$	Beispiel Normalverteilung $X \sim N(\mu, \sigma^2)$: Dichte $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ Verteilung: $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt$
Multivariate Verteilungen (joint probability distribution - JDF) Verteilung ist nun von mehreren ZVen $X_1, \ldots, X_n$ abhängig	Will man untersuchen, wie Wahrscheinlichkeiten $P(\{X_1 = x_1\} \cap \ldots \cap \{X_m = x_m\})$ für den gleichzeitigen Eintritt mehrerer Ereignisse aussehen, so definiert man analog die gemeinsame Verteilungsfunktion $F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) := P(X_1 \leq x_1, \ldots, X_m \leq x_m)$ bzw. $F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = P(\{X_1 \leq x_1\} \cap \ldots \cap \{X_m \leq x_m\})$	Man erhält die Randverteilung (bzw. marginale Wahrscheinlichkeitsverteilung) aus der multivarianten auf folgende Weise: $F_{X_j}(x) = F_{X_1,\ldots,X_m}(\infty, \ldots, x, \ldots, \infty)$ und analog die Randdichte $f_{X_j}(x) := \frac{dF_{X_j}(x)}{dx}$ Ist die gem. Verteilung stetig differenzierbar, so kann man auch die gemeinsame Dichte bilden: $f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) := \frac{\partial^m F_{X_1,\ldots,X_m}(x_1,\ldots,x_m)}{\partial x_1 \ldots \partial x_m}$
Bedingte Wahrscheinlichkeiten, Verteilungen und Dichten	Falls man wissen will, wie die Wahrscheinlichkeit eines Ereignisses unter der Bedingung, dass ein anderes Ereignis eingetreten ist, aussieht, so definiert man: $P(X = x \mid Y = y) := \frac{P((X=x)\cap(Y=y))}{P(Y=y)}$ Die Verteilungsfunktion wird dann wieder definiert als: $F_{X\mid Y}(x) := P(X \leq x \mid Y)$ Da im stetigen Fall $P(X\mid Y) = 0$ gilt, benötigt man die Definition einer bedingten Dichte: $f_{X\mid Y}(x\mid Y = y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$	Diskreter Fall: $F_{X\mid Y}(x\mid y) = \sum_{x_i \leq x} P(X = x_i \mid Y = y)$ Stetiger Fall: $F_{X\mid Y}(x\mid y) = \int_{-\infty}^{x} f_{X\mid Y}(X = t\mid Y = y)dt$ Zusammenhang zur Verteilung: Diskreter Fall: $F_X(x) = \sum_{y_i} F_{X\mid Y}(x\mid y_i)P(y_i)$ Stetiger Fall: $F_X(x) = \int_{-\infty}^{\infty} F_{X\mid Y}(x\mid y)f_Y(y)dy$ (Man erhält die Verteilung, indem man über alle Bedingungen die gemeinsame Verteilung multipliziert mit der Wahrscheinlichkeit bzw. Dichte summiert bzw. integriert)
Kennzahlen von ZVen: Momente (allgemeine und zentrierte)	Für $k \in \mathbb{N}$ heißt $m_k(X) := E[X^k]$ $k-$tes Moment der ZV X und $\tilde{m}_k := E\left[(X - m_1(X))^k\right]$ $k-$tes zentriertes Moment (= Verallgemeinerungen von Erwartungswert und Varianz)	Diskreter Fall: $m_k(X) := \sum_{i=1}^{\infty} x_i^k \cdot P(X = x_i)$ bzw. $\tilde{m}_k(X) := \sum_{i=1}^{\infty}(x_i - m_1(X))^k \cdot P(X = x_i)$ Stetiger Fall: $m_k(X) := \int_{-\infty}^{+\infty} x^k \cdot f_X(x)dx$ bzw. $\tilde{m}_k(X) := \int_{-\infty}^{+\infty}(x - m_1(X))^k \cdot f_X(x)dx$
Erwartungswert $\mu = E[X] = m_1(X)$	Diskret: $E[X] = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i)$	Stetig: $E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx$
Varianz (Variance) $\sigma^2 = Var(X) = \tilde{m}_2(X)$	Diskret: $Var(X) = \sum_{i=1}^{\infty}(x_i - \mu)^2 \cdot P(X = x_i)$	Stetig: $Var(X) = \int_{-\infty}^{+\infty}(x - \mu)^2 \cdot f_X(x)dx$
Schiefe (Skewness) $\nu = v(X) = \tilde{m}_3(X)/\sigma^3$	Diskret: $v(X) = \left(\sum_{i=1}^{\infty}(x_i - \mu)^3 \cdot P(X = x_i)\right)/\sigma^3$	Stetig: $v(X) = \left(\int_{-\infty}^{+\infty}(x - \mu)^3 \cdot f_X(x)dx\right)/\sigma^3$
Wölbung (Kurtosis) $\kappa = w(X) = \tilde{m}_4(X)/\sigma^4$	Diskret: $w(X) = \left(\sum_{i=1}^{\infty}(x_i - \mu)^4 \cdot P(X = x_i)\right)/\sigma^4$	Stetig: $w(X) = \left(\int_{-\infty}^{+\infty}(x - \mu)^4 \cdot f_X(x)dx\right)/\sigma^4$
Bedingte Momente	Analog zu den vorher eingeführten Momenten kann man auch bedingte Momente definieren, in dem man statt der Wahrscheinlichkeit die bedingte Wahrscheinlichkeit und statt der Dichte die bedingte Dichte benutzt	Diskret: $m_k(X\mid Y = y) := \sum_{i=1}^{\infty} x_i^k \cdot P(X = x_i \mid Y = y)$ Stetig: $m_k(X\mid Y = y) := \int_{-\infty}^{+\infty} x^k \cdot f_{X\mid Y}(x\mid Y = y)dx$
Bedingter Erwartungswert $E[X\mid Y]$	Entsprechend dem Zusammenhang zwischen unbedingten und bedingten Wahrscheinlichkeiten existiert ein ähnlicher Zusammenhang auch bei den Erwartungswerten: $E[X] = E[E[X\mid Y]]$	Diskret: $E[X\mid Y] = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i \mid Y = y)$ Stetig: $E[X\mid Y] = \int_{-\infty}^{+\infty} x \cdot f_{X\mid Y}(x\mid Y = y)dx$
$q-$Quantile	Sei $q \in [0,1]$, dann ist das $q-$Quantil ein Wert einer Zufallsvariablen, der die Menge aller Merkmalswerte in zwei Abschnitte unterteilt. Links liegen $q \cdot 100\%$ aller Beobachtungswerte bzw. der Fläche der Verteilungskurve, rechts der restliche Anteil.	Wichtige Quantile sind Median (= $Q_{0,5}$), Quartile (= $Q_{0,25}, Q_{0,5}, Q_{0,75}$), Quintile (= $Q_{0,2}, Q_{0,4}, \ldots$), Dezile (= $Q_{0,1}, Q_{0,2}, \ldots$) und Perzentile (= $Q_{0,01}, Q_{0,02}, \ldots$)