

1 Der Begriff der Wahrscheinlichkeit

Stochastik befasst sich mit *Zufallsexperimenten*. Deren Ergebnisse sind unter „Versuchsbedingungen“ verschieden.

Bsp.

- *Kartenziehen, Würfeln, Roulette*
- *Simulation*
- *Komplexe Phänomene (zumindest approximativ): Börse, Data-Mining, Genetik, Wetter*

Ergebnisse von Zufallsexperimenten werden in *Ereignisse* zusammengefasst.

- *Ereignisraum* (Grundraum) Ω : Menge aller möglichen Ergebnisse des Zufallsexperiments
- *Elementarereignisse* ω : Elemente von Ω , also die möglichen Ergebnisse des Zufallsexperiments
- *Ereignis*: Teilmenge von Ω
- Operationen der Mengenlehre haben natürliche Interpretation in der Sprache der Ereignisse:

Durchschnitt	$A \cap B$	A und B
Vereinigung	$A \cup B$	A oder B
Komplement	A^c	Nicht A
Differenz	$A \setminus B$	A ohne B

Das Vorgehen der Stochastik zur Lösung eines Problems kann in drei Schritte unterteilt werden:

1. Man bestimmt die Wahrscheinlichkeiten gewisser Ereignisse A_i . Dabei sind Expertenwissen, Daten und Plausibilitäten wichtig.
2. Man berechnet aus den Wahrscheinlichkeiten $\mathcal{P}(B_j)$ die Wahrscheinlichkeiten von gewissen anderen Ereignissen B_j gemäss den Gesetzen der Wahrscheinlichkeitstheorie (oft vereinfachend unter Unabhängigkeitsannahme).
3. Man interpretiert die Wahrscheinlichkeiten $\mathcal{P}(B_j)$ im Hinblick auf die Problemstellung.

Das *Bestimmen von Wahrscheinlichkeiten* (siehe Schritt 1) wird oft konkreter formalisiert.

Bsp. (Laplace-Modell)

Die *Wahrscheinlichkeit von einem Ereignis A* ist gegeben durch

$$\mathcal{P}(A) = \frac{|A|}{|\Omega|} \quad \left(= \frac{\# \text{ günstige Fälle}}{\# \text{ mögliche Fälle}} \right). \quad (1.1)$$

Dem *Laplace-Modell* liegt die *uniforme Verteilung von Elementarereignissen* ω zugrunde:

$$\mathcal{P}(\{\omega\}) = \frac{1}{|\Omega|} \quad (1.2)$$

Andere Wahrscheinlichkeitsverteilungen werden mit Hilfe des Konzepts von *Zufallsvariablen* (siehe Kapitel 2) eingeführt. Es sei aber bereits hier festgehalten: die Stochastik geht weit über das Laplace-Modell hinaus. Für viele Anwendungen ist das Laplace-Modell ungeeignet.

1.1 Rechenregeln für Wahrscheinlichkeiten

Die drei Axiome sind:

- (A1) $\mathcal{P}(A) \geq 0$: Wahrscheinlichkeiten sind immer nicht-negativ.
- (A2) $\mathcal{P}(\Omega) = 1$: sicheres Ereignis Ω hat Wahrscheinlichkeit eins.
- (A3) $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) \quad \forall$ Ereignisse A, B , die sich gegenseitig ausschliessen (d.h. $A \cup B = \emptyset$).

Weitere (abgeleitete) Regeln:

$$\mathcal{P}(A^c) = 1 - \mathcal{P}(A) \quad (1.3)$$

für jedes Ereignis A ,

$$\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B) \quad (1.4)$$

für je zwei Ereignisse A und B ,

$$\mathcal{P}(A_1 \cup \dots \cup A_n) \leq \mathcal{P}(A_1) + \dots + \mathcal{P}(A_n) \quad (1.5)$$

für je n Ereignisse A_1, \dots, A

$$\mathcal{P}(B \setminus A) = \mathcal{P}(B) - \mathcal{P}(A) \quad (1.6)$$

für je zwei Ereignisse A und B mit $A \subset B$.

1.2 Unabhängigkeit von Ereignissen

Wenn zwischen zwei Ereignissen A und B kein kausaler Zusammenhang besteht (d.h. es gibt

keine gemeinsamen Ursachen oder Ausschlüssen), dann werden sie *unabhängig* genannt, genauer: Zwei Ereignisse A und B heissen (*stochastisch*) *unabhängig* wenn für jedes $k \leq n$ und all $1 \leq i_1 < \dots < i_k \leq n$ gilt

$$\mathcal{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathcal{P}(A_{i_1}) \cdot \dots \cdot \mathcal{P}(A_{i_k}).$$

Achtung: Zwei *unabhängige* Ereignisse A und B sind *nicht disjunkt* (und umgekehrt), vorausgesetzt die W'keiten $\mathcal{P}(A), \mathcal{P}(B) \neq 0$.

1.3 Interpretation von Wahrscheinlichkeiten

Die beiden wichtigsten Interpretationen sind:

- frequentistisch: „Idealisierung der relative Häufigkeiten bei vielen unabhängigen Wiederholungen“
- subjektive: „Mass für den Glauben, dass ein Ereignis eintreten wird“ (Bayes'sch)

1.4 Bedingte Wahrscheinlichkeiten

Die *bedingte Wahrscheinlichkeit von A* gegeben B ist definiert als:

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)} \quad (1.7)$$

„ $\mathcal{P}(A|B)$ ist die Wahrscheinlichkeit für das Ereignis A , wenn wir *wissen*, dass das Ereignis B schon eingetroffen ist.“

Rechenregeln:

$$0 \leq \mathcal{P}(A|B) \leq 1 \quad (1.8)$$

für jedes Ereignis A

$$\mathcal{P}(B|B) = 1 \quad (1.9)$$

$$\mathcal{P}(A_1 \cup A_2|B) = \mathcal{P}(A_1|B) + \mathcal{P}(A_2|B) \quad (1.10)$$

für A_1, A_2 disjunkt

$$\mathcal{P}(A^c|B) = 1 - \mathcal{P}(A|B) \quad (1.11)$$

Allgemein gilt: $\mathcal{P}(\bullet|B)$ ist W'keit mit allen Regeln, $\mathcal{P}(A|\bullet)$ nicht.

Weiterhin gilt:

$$\mathcal{P}(A \cap B) = \mathcal{P}(A|B)\mathcal{P}(B) = \mathcal{P}(B|A)\mathcal{P}(A) \quad (1.12)$$

Deshalb können wir *Unabhängigkeit* auch definieren als:

$$A, B \text{ unabhängig} \Leftrightarrow \mathcal{P}(A|B) = \mathcal{P}(A) \Leftrightarrow \mathcal{P}(B|A) = \mathcal{P}(B) \quad (1.13)$$

D.h. die Wahrscheinlichkeiten ändern sich nicht, wenn das andere Ereignis schon eingetreten ist. Achtung, im Allgemeinfall gilt:

$$\mathcal{P}(A|B) \neq \mathcal{P}(B|A) \quad (1.14)$$

$$\mathcal{P}(A|B^c) \neq 1 - \mathcal{P}(A|B) \quad (1.15)$$

1.5 Satz der totalen Wahrscheinlichkeit

Haben k disjunkte Ereignisse B_1, \dots, B_k mit $B_1 \cup \dots \cup B_k = \Omega$, d.h. „alle möglichen Fälle sind abgedeckt“, dann gilt:

$$\mathcal{P}(A) \stackrel{(A3)}{=} \sum_{i=1}^k \mathcal{P}(A \cap B_i) \stackrel{(?)}{=} \sum_{i=1}^k \mathcal{P}(A|B_i)\mathcal{P}(B_i) \quad (1.16)$$

1.6 Satz von Bayes

Haben wieder k disjunkte Ereignisse B_1, \dots, B_k mit $B_1 \cup \dots \cup B_k = \Omega$, d.h. „alle möglichen Fälle sind abgedeckt“, dann gilt:

$$\mathcal{P}(B_i|A) = \frac{\mathcal{P}(A \cap B_i)}{\mathcal{P}(A)} \quad (1.17)$$

$$= \frac{\mathcal{P}(A|B_i)\mathcal{P}(B_i)}{\mathcal{P}(A)} \quad (1.18)$$

$$\stackrel{1.17}{=} \frac{\mathcal{P}(A|B_i)\mathcal{P}(B_i)}{\sum_{i=1}^k \mathcal{P}(A|B_i)\mathcal{P}(B_i)} \quad (1.19)$$

2 Zufallsvariable und Wahrscheinlichkeitsverteilung

Ergebnisse eines physikalischen Versuchs (Zufallsexperiment) sind oft Zahlen (Messungen). Diese werden als Beobachtung von so genannten Zufallsvariablen interpretiert, d.h. beobachtet wird nicht das ω welches bei einem Zufallsexperiment herauskommt, sondern die Werte aller beobachteten Zufallsvariablen.

2.1 Definition einer Zufallsvariablen

Eine Zufallsvariable X ist ein Zufallsexperiment mit möglichen Werten in \mathbb{R} , bzw. in einer Teilmenge von \mathbb{R} , z.B. $\mathbb{N}_0 = \{0, 1, \dots\}$. Deren Wert ist im Voraus nicht bekannt, sondern hängt vom Ergebnis eines Zufallsexperiments ab. Mathematisch ist eine Zufallsvariable einfach nur eine Abbildung von Ω nach \mathbb{R} :

$$X : \Omega \rightarrow \mathbb{R}, \\ \omega \mapsto X(\omega).$$

Das heisst wenn das Ergebnis ω herauskommt, nimmt die Zufallsvariable den Wert $X(\omega)$ an.

Bsp. (Wert einer zuf. gez. Jasskarte.)

Sei $\Omega = \{\text{Jasskarten}\}$; ein $\omega \in \Omega$ ist z.B. ein Schulten-As; Zufallsvariable X :

$$\begin{aligned} \text{As irgendeiner Farbe} &\mapsto 11 \\ \text{König irgendeiner Farbe} &\mapsto 4 \\ \text{„Brettchen“ irgendeiner Farbe} &\mapsto 0. \end{aligned}$$

2.2 Wahrscheinlichkeitsverteilung auf \mathbb{R}

Eine Zufallsvariable X legt eine Wahrscheinlichkeit Q auf \mathbb{R} fest, die sogenannte *Verteilung* von X :

$$\begin{aligned} Q(B) &= P(\{\omega; X(\omega) \in B\}) \\ &= P(X \in B) \end{aligned}$$

Bsp. (Wert einer zuf. gez. Jk. Forts.)

In obigem Beispiel ist beispielsweise

$$\begin{aligned} Q(11) &= P(\text{As irgendeiner Farbe}) \\ &= \frac{4}{36}. \end{aligned}$$

Die *kumulative Verteilungsfunktion* ist definiert als

$$\begin{aligned} F(b) &= P(X \leq b) \\ &= Q((-\infty, b]). \end{aligned}$$

Sie enthält dieselbe Information wie die Verteilung $Q(\cdot)$, ist aber einfacher darzustellen. Die Umkehrung der Verteilungsfunktion stellen die sogenannten Quantile dar, für $\alpha \in (0, 1)$ ist das α -Quantil von X definiert als das kleinste $x \in \mathbb{R}$ für welches $F(x) \geq \alpha$ gilt, also

$$\begin{aligned} q_\alpha &:= q(\alpha) \\ &:= \min\{x \in \mathbb{R} \mid F(x) \geq \alpha\} \end{aligned}$$

Es gilt

$$F(q_\alpha) = \alpha$$

bzw. äquivalent dazu

$$q_\alpha = F^{-1}(\alpha).$$

Das $\frac{1}{2}$ -Quantil von X heisst auch *Median* von X .

2.3 Diskrete und stetige Zufallsvariablen

Eine Zufallsvariable X heisst *diskret*, falls die Menge W der möglichen Werte von X endlich oder abzählbar ist. Zum Beispiel $W = \{0, 1, 2, \dots, 100\}$ oder $W = \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Die Verteilung einer diskreten Zufallsvariablen ist festgelegt durch die Angabe der sogenannten *Wahrscheinlichkeitsfunktion*:

$$p(x) := P(X = x), \quad x \in W.$$

Offensichtlich ist die kumulative Verteilungsfunktion eine Treppenfunktion mit Sprüngen an den Stellen $x \in W$ mit Sprunghöhen $p(x)$, also nicht stetig. Ferner gilt

$$Q(B) = \sum_{x \in B} p(x).$$

Eine Zufallsvariable X heisst *stetig* falls die Menge der möglichen Werte W ein Intervall enthält. Zum Beispiel $W = [0, 1]$ oder $W = \mathbb{R}$.

2.4 Erwartungswert und Varianz

Eine Verteilung einer Zufallsvariablen X kann durch mindestens zwei Kennzahlen zusammengefasst werden, eine für die Lage (der Erwartungswert $E(X) = \mu_X$) und eine für die Streuung (die Standardabweichung σ_X). Der *Erwartungswert* einer diskreten Zufallsvariable X ist definiert durch

$$\mu_X = E(X) := \sum_{x \in W} xp(x) \quad (2.1)$$

Für den Erwartungswert einer transformierten diskreten Zufallsvariable $Y = f(X)$ ergibt sich daraus:

$$E(Y) = E(f(X)) = \sum_{x \in W} f(x)p(x) \quad (2.2)$$

Die *Varianz* einer diskreten Zufallsvariable X ist definiert durch:

$$V(X) := E((X - E(X))^2) = \sum_{x \in W} (x - \mu_X)^2 p(x) \quad (2.3)$$

Die *Standardabweichung* ist die Wurzel aus der Varianz, d.h. $\sigma_X = \sqrt{V(X)}$. Folgende Rechenregeln sind nützlich:

$$E(a + bX) = a + bE(X), \quad a, b \in \mathbb{R} \quad (2.4)$$

$$V(X) = E(X^2) - E(X)^2 \quad (2.5)$$

$$V(a + bX) = b^2 V(X) \quad (2.6)$$

In der frequentistischen Interpretation ist der Erwartungswert eine Idealisierung des arithmetischen Mittels der Werte einer Zufallsvariablen bei vielen Wiederholungen.

2.5 Die wichtigsten diskreten Verteilungen

2.5.1 Binomialverteilung

Die *Binomialverteilung* $\text{Bin}(n, p)$ ist die Verteilung der Anzahl Erfolge bei n unabhängigen Wiederholungen eines Experiments mit Erfolgswahrscheinlichkeit p .

$$W = 0, 1, \dots, n \quad (2.7)$$

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (2.8)$$

$$E(X) = np \quad (2.9)$$

$$\sigma_X = \sqrt{np(1-p)} \quad (2.10)$$

2.5.2 Poissonverteilung

Die *Poissonverteilung* $\text{Poi}(\lambda)$ ist eine Approximation der Binomialverteilung für grosses n und kleines p , mit $np = \lambda$. Die Anzahl Ausfälle einer Komponente oder eines Systems in einem Intervall der Länge t ist oft in erster Näherung Poisson-verteilt mit Parameter λt .

$$W = 0, 1, \dots \quad (2.11)$$

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad (2.12)$$

$$E(X) = \lambda \quad (2.13)$$

$$\sigma_X = \sqrt{\lambda} \quad (2.14)$$

2.5.3 Geometrische Verteilung

Die *geometrische Verteilung* $\text{Geo}(p)$ ist die Verteilung der Anzahl Wiederholungen bis ein Ereignis mit Wahrscheinlichkeit p eintritt.

$$W = 1, 2, \dots \quad (2.15)$$

$$p(x) = p(1-p)^{x-1} \quad (2.16)$$

$$E(X) = \frac{1}{p} \quad (2.17)$$

$$\sigma_X = \frac{\sqrt{1-p}}{p} \quad (2.18)$$

3 Stetige Wahrscheinlichkeitsverteilung

Bei einer stetigen Zufallsvariablen X ist $\mathcal{P}(X = x) = 0$ für jedes feste x . Wir betrachten nur Fälle, wo $\mathcal{P}(x \leq X \leq x+h)$ für kleine h ungefähr proportional zu h ist. Die Proportionalitätskonstante heisst die *Dichte* f von X .

3.1 Wahrscheinlichkeitsdichte

Die Dichte von einer stetigen Verteilung P ist definiert als

$$f(x) = \lim_{h \downarrow 0} \frac{\mathcal{P}(x \leq X \leq x+h)}{h} \quad (3.1)$$

Zwischen der Dichte f und der kumulativen Verteilungsfunktion F bestehen die folgenden Beziehungen:

$$f(x) = F'(x), \quad F(x) = \int_{-\infty}^x f(x) dx \quad (3.2)$$

Erwartungswert und Varianz berechnen sich gemäss

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x) dx \quad (3.3)$$

$$V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx \quad (3.4)$$

und es gelten die gleichen Rechenregeln wie im diskreten Fall (TODO: Referenz hier?).

3.2 Die wichtigsten stetigen Verteilungen

3.2.1 Uniforme Verteilung

Die *uniforme Verteilung* $\text{Uni}[a, b]$ tritt auf bei Rundungsfehlern und als Formalisierung der

völligen "Ignoranz".

$$W = [a, b] \quad (3.5)$$

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (3.6)$$

$$E(X) = \frac{a+b}{2} \quad (3.7)$$

$$\sigma_X = \frac{b-a}{\sqrt{12}} \quad (3.8)$$

3.2.2 Exponentialverteilung

Die *Exponentialverteilung* $\text{Exp}(\lambda)$ ist das einfachste Modell für Wartezeiten auf Ausfälle und eine stetige Version der geometrischen Verteilung.

$$W = [0, \infty) \quad (3.9)$$

$$f(x) = \lambda e^{-\lambda x} \text{ für } x > 0 \quad (3.10)$$

$$F(x) = 1 - e^{-\lambda x} \quad (3.11)$$

$$E(X) = \sigma_X = \frac{1}{\lambda} \quad (3.12)$$

Wenn die Zeiten zwischen den Ausfällen eines Systems $\text{Exponential}(\lambda)$ -verteilt sind, dann ist die Anzahl Ausfälle in einem Intervall der Länge t $\text{Poi}(\lambda t)$ -verteilt.

3.2.3 Normal- oder Gaussverteilung

Die *Normal- oder Gaussverteilung* $\mathcal{N}(\mu, \sigma^2)$ ist die häufigste Verteilung für Messwerte.

$$W = \mathbb{R} \quad (3.13)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (3.14)$$

$$E(X) = \mu \quad (3.15)$$

$$\sigma_X = \sigma \quad (3.16)$$

Die $\mathcal{N}(0, 1)$ Verteilung, auch *Standardverteilung* bezeichnet, ist ein wichtiger Sonderfall, weshalb es für dessen Verteilungsfunktion sogar ein eigenes Symbol gibt: $\Phi(x) := F_{\mathcal{N}(0,1)}$, $x \in \mathbb{R}$ und deren Umkehrfunktionen (d.h. die Quantile) kürzen wir ab mit $z_\alpha := \Phi^{-1}(\alpha)$, $\alpha \in [0, 1]$. Die Verteilungsfunktion F einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariable ist nicht geschlossen darstellbar, sie wird aus der Verteilungsfunktion Φ der Standardnormalverteilung (welche tabel-

liert ist) berechnet mittels der Formel:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R} \quad (3.17)$$

3.3 Transformationen

Bei stetigen Verteilungen spielen Transformationen $Y = g(X)$ eine wichtige Rolle. Falls g linear ist: $g(x) = a + bx$ mit $b > 0$, dann gilt $E(Y) = a + bE(X)$, $\sigma_Y = b\sigma_X$, $F_Y(x) = F_X((x-a)/b)$ und $f_Y(x) = f_X((x-a)/b)/b$. Durch Skalenänderungen kann man also alle Exponentialverteilungen ineinander überführen, und ebenso durch lineare Transformationen alle Normalverteilungen ineinander. Für beliebiges g gilt

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx \quad (3.18)$$

Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$ normalverteilt ist, dann heisst $Y = e^X$ *lognormal-verteilt*. Es gilt z.B. $E(Y) = \exp(\mu + \sigma^2/2)$.

3.4 Simulation von Zufallsvariablen

Wenn U uniform auf $[0, 1]$ verteilt ist und F eine beliebige kumulative Verteilungsfunktion, dann ist die Verteilungsfunktion von $X = F^{-1}(U)$ gleich F . Dies ist ein wichtiges Faktum um Verteilungen, respektive Realisierungen von Zufallsvariablen, zu *simulieren*:

1. Erzeuge Realisation u von uniform verteilter Zufallsvariable $U \sim \text{Uni}[0, 1]$. Dies wird mittels einem "Standard-Paket" gemacht.
2. Berechne $x = F^{-1}(u)$. Gemäss obigem Faktum ist dann x eine Realisation einer Zufallsvariablen X mit kumulativer Verteilungsfunktion F .

Diese Methode ist nicht immer rechentechnisch effizient.