

Algorithmische Mathematik II

Dozent

PROFESSOR DR. PATRIK FERRARI

Mitschrift

MAXIMILIAN KESSLER

Version

3. Mai 2021 11:59

Zusammenfassung

Bei folgenden Vorlesungsnotizen handelt es sich um (inoffizielle) Mitschriften zur Vorlesung 'Algorithmische Mathematik II', die im Sommersemester 2021 an der Universität Bonn gehalten wird. Ich garantiere weder für Korrektheit noch Vollständigkeit dieser Notizen, und bin dankbar für jegliche Art von Korrektur, sowohl inhaltlich, als auch Tippfehler.

Bemerkungen, die nicht zum eigentlichen Vorlesungsinhalte gehören, wurden mit einem * gekennzeichnet. Sie werden nach eigenem Ermessen hinzugefügt, um weitere Details oder evtl. mündliche Anmerkungen beizufügen.

Weitere Informationen finden sich bei [GitHub](#) oder auf der [Vorlesungs-homepage](#)

Inhaltsverzeichnis

1 Bedingte Wahrscheinlichkeit und Unabhängigkeit	8
1.1 Bedingte Wahrscheinlichkeit	8
1.2 Baye'sche Regel	12
Stichwortverzeichnis	15

Vorlesung 6: Erwartungswert von Zufallsvariablen, Varianz

Sa 01 Mai 2021 09:18

Beispiel (Zufallsvariablen mit Werten in $\{0, 1\}$). Sei $A \in \mathcal{F}$ ein Ereignis, und definiere X durch

$$X(\omega) := \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{sonst} \end{cases}.$$

Dann ist

$$\mathbb{E}(X) = \mathbb{P}(A).$$

Beweis. Nach Definition ist

$$\begin{aligned} \mathbb{E}(X) &= 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) \\ &= \mathbb{P}(A) \end{aligned}$$

□

Beispiel (Binomialverteilung). Sei $X \sim \text{Bin}(n, p)$. Wir wollen zeigen, dass $\mathbb{E}(X) = p \cdot n$.

Beweis.

$$\mathbb{E}(X) = \sum_{k=0}^n \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{=\mathbb{P}(X=k)} \cdot k.$$

Wir wollen nun allgemein, in Anlehnung an die Binomialformel, den Wert von

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \cdot k.$$

berechnen. Dazu stellen wir fest, dass

$$p \cdot \frac{d}{dp} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = p \cdot \sum_{k=0}^n \binom{n}{k} p^{k-1} q^{n-k} \cdot k.$$

unser Ausdruck ist, also suchen wir

$$p \cdot \frac{d}{dp} (p+q)^n = p \cdot (p+q)^{n-1} \cdot n.$$

Nun können wir $q = 1-p$ auf beiden Seiten setzen, und somit erhalten wir

$$\mathbb{E}(X) = p \cdot (p + (1-p))^{n-1} \cdot n = p \cdot n.$$

wie gewünscht. □

Beispiel (Poisson-Verteilung). Sei $X \sim \text{Poi}(\lambda)$, dann behaupten wir, dass $\mathbb{E}(X) = \lambda$.

Beweis.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k \geq 0} \frac{e^{-\lambda} \lambda^k}{k!} \cdot k \\ &= e^{-\lambda} \cdot \lambda \cdot \underbrace{\sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!}}_{\rightarrow e^{\lambda}} \\ &= \lambda\end{aligned}$$

□

Bemerkung*. Diese Feststellung passt auch zur Konstruktion der Poisson-Verteilung.

Bemerkung. Oft kann man definieren

$$\psi(z) := \sum_{k \in S} p(k) z^k.$$

Dann werden wir uns

$$\frac{d}{dz} \psi(z) = \sum_{k \in S} p(k) k z^{k-1}$$

ansehen und bei $z = 1$ evaluieren, um $\mathbb{E}(X)$ zu berechnen.

Das ganze funktioniert, wenn X durch $\mathbb{P}(X = k) = p(k)$ verteilt ist und natürlich nur, wenn alle Objekte wohldefiniert sind.

Wir wollen auch Funktionen von X betrachten.

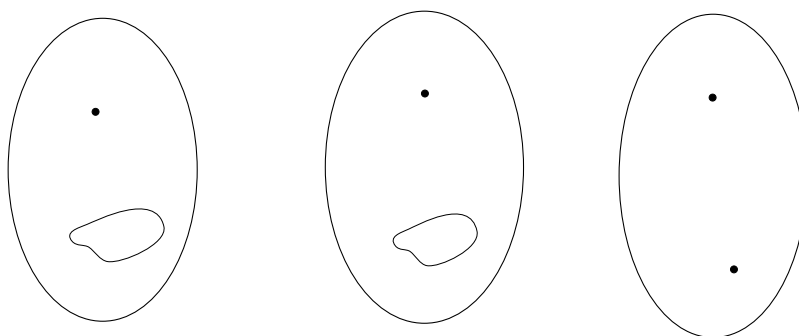


Abbildung 1: Funktionen von Zufallsvariablen

Satz 0.1 (Transformationssatz). Sei $X : \Omega \rightarrow \mathcal{S}$ eine diskrete Zufallsvariable und $f : \mathcal{S} \rightarrow \mathbb{R}$ eine Funktion. Dann ist $f(X) := f \circ X : \Omega \rightarrow \mathbb{R}$ auch eine Zufallsvariable und

$$\mathbb{E}(f(X)) = \sum_{s \in \mathcal{S}} f(s) \mathbb{P}(X = s).$$

falls die Summe wohldefiniert ist.

Beweis. Messbarkeit: Es ist

$$\{f(X) = a\} = \bigcup_{s \in f^{-1}(a)} \{X = s\} \in \mathcal{F}.$$

weil

$$\{\omega \mid X(\omega) = s\} \in \mathcal{F}.$$

da X eine Zufallsvariable ist. Nach Definition ist nun

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{a \in f(\mathcal{S})} a \cdot \mathbb{P}(f(X) = a) \\ &= \sum_{a \in f(\mathcal{S})} a \cdot \mathbb{P}\left(\bigcup_{s \in f^{-1}(a)} \{X = s\}\right) \\ &= \sum_{a \in f(\mathcal{S})} a \cdot \sum_{s \in f^{-1}(a)} \mathbb{P}(X = s) \\ &= \sum_{a \in f(\mathcal{S})} \sum_{s \in f^{-1}(a)} f(s) \mathbb{P}(X = s) \\ &= \sum_{s \in \mathcal{S}} f(s) \mathbb{P}(X = s) \end{aligned}$$

□

Der Erwartungswert ist LINEAR und MONOTON:

Satz 0.2 (Linearität des Erwartungswerts). Seien $X_1 : \Omega \rightarrow \mathcal{S}_1$ und $X_2 : \Omega \rightarrow \mathcal{S}_2$ zwei diskrete Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Falls $\mathbb{E}(|X_1|) < \infty$ und $\mathbb{E}(|X_2|) < \infty$, dann ist $\forall \lambda_1, \lambda_2 \in \mathbb{R}$:

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 \mathbb{E}(X_1) + \lambda_2 \mathbb{E}(X_2).$$

Beweis. Es ist

$$|\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2)| \leq |\lambda_1| \mathbb{E}(|X_1|) + |\lambda_2| \mathbb{E}(|X_2|) < \infty.$$

(nach Dreiecksungleichung). Nun rechnen wir aus:

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \mathbb{E}(f(X_1, X_2)).$$

wobei $f(x_1, x_2) = \lambda_1 x_1 + \lambda_2 x_2$, also

$$\begin{aligned} &= \sum_{\substack{x_1 \in \mathcal{S}_1 \\ x_2 \in \mathcal{S}_2}} f(x_1, x_2) \mathbb{P}(X_1 = x_1 \cap X_2 = x_2) \\ &= \lambda_1 \sum_{x_1 \in \mathcal{S}_1} x_1 \sum_{x_2 \in \mathcal{S}_2} \mathbb{P}(X_1 = x_1 \cap X_2 = x_2) + \text{sym.} \\ &= \lambda_1 \sum_{x_1 \in \mathcal{S}_1} x_1 \mathbb{P}(X_1 = x_1) + \text{sym.} \\ &= \lambda_1 \mathbb{E}(X_1) + \lambda_2 \mathbb{E}(X_2) \end{aligned}$$

□

Korollar 0.3 (Monotonie des Erwartungswerts). Seien X_1, X_2 reellwertige Zufallsvariablen mit $X_1(\omega) \leq X_2(\omega)$ für alle $\omega \in \Omega$. Dann ist

$$\mathbb{E}(X_1) \leq \mathbb{E}(X_2).$$

Beweis. Da $X_2(\omega) - X_1(\omega) \geq 0$, also ist trivialerweise $\mathbb{E}(X_2 - X_1) \geq 0$. Wegen der Linearität ist nun $\mathbb{E}(X_2 - X_1) = \mathbb{E}(X_2) - \mathbb{E}(X_1)$ und somit sind wir fertig. □

Beispiel. Seien $A_1, A_2, \dots, A_n \in \mathcal{F}$ mit $\mathbb{P}(A_i) = p$ für alle i . Sei $X_i := \mathbb{1}_{A_i}$. Dann ist $X_i \sim \text{Ber}(p)$ und $\mathbb{E}(X_i) = p$. Sei

$$S_n := \sum_{i=1}^n X_i.$$

Dann ist

$$\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n p = np.$$

Das ist eine Verallgemeinerung des Falles, in dem A_1, \dots, A_n unabhängig sind, weil dann $S_n \sim \text{Bin}(n, p)$.

Oft interessieren wir uns auch dafür, wie weit eine Zufallsvariable von ihrem Ursprungswert entfernt ist. Ist z.B. $\mathbb{P}(X = k) = \mathbb{P}(X = -k)$, so ist $\mathbb{E}(X) = 0$, so sind immer noch folgende Fälle denkbar:

Plot einfügen

Frage. Wie weit sind die Werte von X Mittelwert ($\mathbb{E}(X)$) entfernt?

Die Antwort liefert die sogenannte Varianz $\text{Var}(X)$:

Definition 0.4 (Varianz). Sei X eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{E}(X^2) < \infty$. Die **Varianz** von X ist durch

$$\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$$

definiert.

Lemma* (Eigenschaften der Varianz). Es gilt folgendes:

- i) $\text{Var}(X) \geq 0$, und $\text{Var}(X) = 0 \Leftrightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1$, d.h. $X(\omega)$ ist in diesem Fall eine Konstante.
- ii) Es ist

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad \text{Var}(\lambda \cdot X) = \lambda^2 \text{Var}(X).$$

- iii) Die Varianz hängt nicht vom Erwartungswert ab, d.h.

$$\text{Var}(X) = \text{Var}(X + a) \quad \forall a \in \mathbb{R}.$$

Beweis.* i) $\text{Var}(X) \geq 0$ ist klar, weil Quadrate nichtnegativ sind, Gleichheit gilt, wenn $X = \mathbb{E}(X)$ für jedes X , also wenn $\mathbb{P}(X =$

$\mathbb{E}(X) = 1$, und dann ist $X(\omega)$ konstant $\mathbb{E}(X)$.

ii) Wir rechnen nach:

$$\begin{aligned}\mathbb{E}((X - \mathbb{E}(X))^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X) \cdot \mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$

Für die 2. Behauptung ist

$$\text{Var}(\lambda X) = \mathbb{E}((\lambda X)^2) - \mathbb{E}(\lambda X)^2 = \mathbb{E}(\lambda^2 X^2) - (\lambda \mathbb{E}(X))^2 = \lambda^2 \text{Var}(X).$$

iii) Es ist

$$\text{Var}(X + a) = \mathbb{E}((X + a - \mathbb{E}(X + a))^2) = \mathbb{E}((X - \mathbb{E}(X))^2) = \text{Var}(X).$$

□

Beispiel. (a) Ist $X \sim \text{Ber}(p)$, so ist $\text{Var}(X) = p(1 - p)$.

(b) Ist $X \sim \text{Bin}(n, p)$, so ist $\text{Var}(X) = n \cdot p(1 - p)$.

(c) Ist $X \sim \text{Poi}(\lambda)$, so ist $\text{Var}(X) = \lambda$

(d) Ist $X \sim \text{Geo}(q)$, so ist $\text{Var}(X) = \frac{q}{(1-q)^2}$

Beweis. (a) Wir benutzen $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Nun ist

$$\mathbb{E}(X^2) = 0^2 \mathbb{P}(X = 0) + 1^2 \mathbb{P}(X = 1) = p.$$

und somit $\text{Var}(X) = p - p^2 = p(1 - p)$

(b) Sei $X \sim \text{Bin}(n, p)$, wir wissen bereits $\mathbb{E}(X) = p \cdot n$. Nun ist

$$\mathbb{E}(X^2) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} k^2.$$

Wir benutzen den gleichen Trick wie vorher nochmal, indem wir feststellen:

$$\begin{aligned}\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k^2 &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k(k-1) + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k \\ &= p^2 \frac{d^2}{dp^2} (p+q)^n + p \frac{d}{dp} (p+q)^n \\ &= p^2 (p+q)^{n-1} n(n-1) + p(p+q)^{n-1} \cdot n\end{aligned}$$

Einsetzen von $q = 1 - p$ liefert nun:

$$\mathbb{E}(X^2) = p^2 n(n-1) + p \cdot n = p^2 n^2 - p^2 n + pn.$$

Damit erhalten wir schlussendlich

$$\text{Var}(X) = p^2 n^2 - p^2 n + pn - (pn)^2 = np(1 - p).$$

© Ist $X \sim \text{Poi}(\lambda)$, so wissen wir bereits $\mathbb{E}(X) = \lambda$. Nun ist

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k \geq 0} k^2 \cdot \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k \geq 0} k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} + \sum_{k \geq 0} k \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} \sum_{k \geq 2} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda} \lambda \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda^2 + \lambda\end{aligned}$$

Und damit ergibt sich für die Varianz:

$$\text{Var}(X) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

□

Bemerkung. Sei $X : \Omega \rightarrow \mathcal{S}$ eine Zufallsvariable. Wir beobachten X n mal. Der Erwartungswert des Empirischen Masses ist genau der empirische Mittelwert, also für Beobachtungen x_1, \dots, x_n genau

$$m_n := \frac{1}{n} \sum_{i=1}^n x_i.$$

Beispiel (Anwendung). Ein alternativer Beweis des ??s lässt sich mit der Linearität des Erwartungswerts führen:

Es ist

$$\begin{aligned}\mathbb{P}((A_1 \cup \dots \cup A_n)^c) &= \mathbb{P}(A_1^c \cap \dots \cap A_n^c) \\ &= E(1_{A_1^c \cap \dots \cap A_n^c}) = (1_{A_1^c} \cdot \dots \cdot 1_{A_n^c}) \\ &= E((1 - 1_{A_1}) \cdot \dots \cdot (1 - 1_{A_n})) \\ &\stackrel{\text{Linearität}}{=} \sum_{k=0}^n (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{E}(1_{A_{i_1}} \cdot \dots \cdot 1_{A_{i_k}}) \\ &= \sum_{k=0}^n (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})\end{aligned}$$

Komplementbildung liefert nun das gewünschte Ergebnis:

$$\begin{aligned}\mathbb{P}(A_1 \cup \dots \cup A_n) &= 1 - \mathbb{P}((A_1 \cap \dots \cap A_n)^c) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})\end{aligned}$$

Vorlesung 7: Bedingte Wahrscheinlichkeit

Mo 03 Mai 2021 10:17

1 Bedingte Wahrscheinlichkeit und Unabhängigkeit

1.1 Bedingte Wahrscheinlichkeit

Beispiel. Es werden statistische Daten über Kleinkinder gemessen: Wann sie krabbeln und wann sie laufen. Betrachten wir die Ereignisse

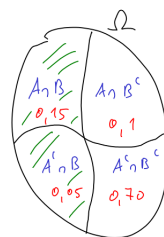
$$A = \{\text{Kind läuft vor dem 10. Monat}\} \quad B = \{\text{Kind krabbelt vor dem 6. Monat}\}.$$

Aus den Daten geht hervor, dass $\mathbb{P}(A) = 25\%$ und $\mathbb{P}(B) = 20\%$.

Frage. Sei ein Kind, das mit 6 Monaten krabbelt, gegeben. Wie hoch ist die Wahrscheinlichkeit, dass es mit 10 Monaten schon läuft?

Wir brauchen mehr Information als die obige, um die Frage beantworten zu können! Wir gehen also davon aus, dass wir sogar folgende Daten zur Verfügung haben:

\cap	A	A^c
B	0,15	0,05
B^c	0,10	0,70



Wir wissen, dass B eintritt, also befinden wir uns bereits im Zustandsraum $\Omega_B := \{\omega \in \Omega \mid \omega \in B\}$. Ziel ist es also, eine neue Massenfunktion $\mathbb{P}(\cdot|B)$ (auf Ω) zu definieren, die die Information ' $\omega \in B$ ' berücksichtigt. Insbesondere muss also gelten:

$$\mathbb{P}(\omega|B) = 0 \quad \forall \omega \in \Omega_B^c.$$

Zudem wollen wir, dass die Information ' $\omega \in B$ ' dieselbe ist für alle $\omega \in \Omega_B$, d.h.

$$p(\omega|B) = c \cdot p(\omega) \quad \forall \omega \in \Omega_B.$$

wobei $p(\omega)$ die Massenfunktion von \mathbb{P} ist. Wegen Normierung ergibt sich also bereits

$$1 = \sum_{\omega \in \Omega} \mathbb{P}(\omega|B) = c \cdot \sum_{\omega \in \Omega_B} p(\omega) \quad \Leftrightarrow \quad c = \frac{1}{\mathbb{P}(B)}.$$

Also ergibt sich, dass

$$p(\omega | B) = \begin{cases} \frac{p(\omega)}{\mathbb{P}(B)} & \text{falls } \omega \in B \\ 0 & \text{sonst} \end{cases}.$$

Wir können das ganze so darstellen:

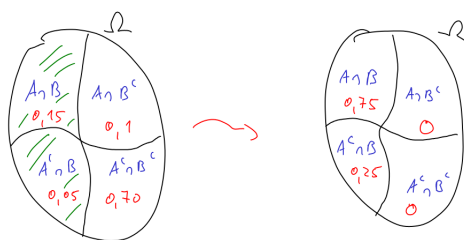


Abbildung 2: Änderung des Zustandsraums bei bedingten Wahrscheinlichkeiten

Wir erhalten also:

Antwort. Ein Kind, das mit 6 Monaten krabbelt, wird mit einer Wahrscheinlichkeit von 75% mit 10 Monaten laufen können.

Besser
Skizzen
machen

Definition 1.1 (Bedingte Wahrscheinlichkeit). Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $A, B \in \mathcal{F}$ Ereigniss mit $\mathbb{P}(B) \neq 0$. Dann definieren wir

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

und nennen diese die **bedingte Wahrscheinlichkeit von A gegeben B**.

Bemerkung. Die Abbildung

$$\mathbb{P}(\cdot | B) : \begin{cases} \mathcal{F} & \longrightarrow \mathbb{R}_+ \\ A & \longmapsto \mathbb{P}(A | B) \end{cases}$$

ist eine Wahrscheinlichkeitsverteilung auf (Ω, \mathcal{F}) , die wir auch die **bedingte Wahrscheinlichkeitsverteilung gegeben B** nennen.

Definition 1.2 (Bedingter Erwartungswert). Sei $X : \Omega \rightarrow \mathcal{S} \subseteq \mathbb{R}$ eine (diskrete) Zufallsvariable mit Verteilung $\mathbb{P}(\cdot | B)$. Dann hat X den Erwartungswert

$$\sum_{s \in \mathcal{S}} s \cdot \mathbb{P}(X = s | B) =: \mathbb{E}(X | B).$$

Dieser heißt **bedingter Erwartungswert von X gegeben B**.

Beispiel. Wir werfen eine faire Münze N mal, dabei beobachten wir n mal das Ergebnis 'Zahl'.

Frage. Was ist die Wahrscheinlichkeit, dass bei den ersten m Würfeln immer 'Zahl' gefallen ist?

Ohne Weitere Informationen (dass insgesamt n mal Zahl gefallen

ist) würden wir hier $\mathbb{P} \equiv \frac{1}{2^m}$ erhalten.

Betrachte nun den Zustandsraum

$$\Omega = \{\omega = (x_1, \dots, x_N) \mid x_i \in \{0, 1\}, 1 \leq i \leq N\}.$$

wobei

$$x_k := \begin{cases} 1 & \text{falls } k\text{-ter Wurf ist 'Zahl'} \\ 0 & \text{falls } k\text{-ter Wurf ist 'Kopf'} \end{cases}.$$

und versehe ihn mit $\mathcal{F} = \mathcal{P}(\Omega)$ sowie \mathbb{P} als Gleichverteilung auf Ω .

Mit $X_k(\omega) := x_k$ interessieren wir uns also für

$$\mathbb{P}\left(X_1 = X_2 = \dots = X_m = 1 \mid \sum_{k=1}^N X_k = n\right).$$

Nach Definition ist dies

$$\begin{aligned} & \mathbb{P}\left((X_1 = \dots = X_m = 1) \cap \left(\sum_{k=m+1}^N X_k = n - m\right)\right) \\ &= \frac{\mathbb{P}\left((X_1 = \dots = X_m = 1) \cap \left(\sum_{k=m+1}^N X_k = n - m\right)\right)}{\mathbb{P}\left(\sum_{k=1}^N X_k = n\right)} \\ &= \frac{\frac{1}{2^N} \binom{N-m}{n-m}}{\frac{1}{2^N} \binom{N}{n}} \\ &= \frac{(N-m)!}{(n-m)!(N-n)!} \cdot \frac{(N-n)!n!}{N!} \\ &= \frac{(N-m)!n!}{N!(n-m)!} \end{aligned}$$

Notation. Zur Vereinfachung der Notation schreiben wir oft

$$\mathbb{P}(X_1 = X_2 = a) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = a\}).$$

sowie

$$\mathbb{P}(X_1 = a, X_2 = b) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = b\}).$$

Wir haben gerade aus einer Wahrscheinlichkeitsverteilung \mathbb{P} die Verteilung $\mathbb{P}(\cdot \mid B)$ gewonnen. Das ganze geht auch umgekehrt:

Satz 1.3. Sei $\Omega = \bigcup_{k \in I} H_k$ eine **disjunkte** Zerlegung von Ω in (abzählbar viele) Ereignisse $H_k, k \in I$, wobei $\mathbb{P}(H_k) \neq 0$. Dann ist $\forall A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{k \in I} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k).$$

Beweis. $\forall A \in \mathcal{F}$ ist

$$A = A \cap \bigcup_{k \in I} H_k = \bigsqcup_{k \in I} (A \cap H_k).$$

eine disjunkte Vereinigung. Also folgt aus σ -Additivität, dass

$$\begin{aligned}\mathbb{P}(A) &= \sum_{k \in I} \mathbb{P}(A \cap H_k) \\ &= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \cap H_k) \\ &= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)\end{aligned}$$

□

Beispiel. Eine Urne A enthält **2 rote** und **3 blaue** Kugeln. In Urne B liegen umgekehrt **3 rote** und nur **2 blaue** Kugeln. Wir gehen davon aus, dass die Urnen immer gut gemischt sind. Nun machen wir Folgendes:

- ① Wir ziehen eine Kugel K_1 aus A und legen sie in B
- ② Wir ziehen eine Kugel K_2 aus B und lg

Frage. Was ist die Wahrscheinlichkeit, dass K_2 **rot** ist?

Wir erhalten nun

$$\begin{aligned}\mathbb{P}(K_2 \text{ ist rot}) &= \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ rot}) \cdot \mathbb{P}(K_1 \text{ rot}) \\ &\quad + \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ blau}) \cdot \mathbb{P}(K_1 \text{ blau}) \\ &= \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{6} \cdot \frac{3}{5} = \frac{17}{30}\end{aligned}$$

Graphik
einfügen

1.2 Baye'sche Regel

In der Baye'schen Statistik ist $\mathbb{P}(H_k)$ auch die **a-priori-Einschätzung** der Wahrscheinlichkeit einer Hypothese H_k , das könnte z.B. sein

$H_k = \{\text{Die Unfallkosten pro Jahr liegen im Bereich } [100k, 100(k+1))\}.$

Aus statistischen Daten weiß man, dass ein Ereignis $A \in \mathcal{F}$ mit einer Wahrscheinlichkeit $\mathbb{P}(A) \neq 0$ eintritt, also z.B.

$A = \text{'Es handelt sich um einen Auffahrunfall'}.$

Dazu kennt man $\mathbb{P}(A \mid H_k)$. Falls A eintritt, werden die Versicherungskosten neu berechnet, auf der Basis von

$$\mathbb{P}(H_k \mid A).$$

Dies nennt man dann auch **a-posteriori-Verteilung** von H_k .

Korollar 1.4 (Baye'sche Regel). Für $A \in \mathcal{F}$ mit $\mathbb{P}(A) \neq 0$ gilt

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)}{\sum_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l)}.$$

Beweis. Es ist

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)}{\mathbb{P}(A)}.$$

Aus Satz 1.3 erhalten wir nun aber genau

$$\mathbb{P}(A) = \sum_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l).$$

und wir sind fertig. \square

Beispiel. Eine Krankheit K tritt selten auf, mit einer Häufigkeit von 10^{-4} , also bei 10 von 100.000 Menschen. Ein Test zur Erkennung der Krankheit ist positiv (+) bei 96% der Kranken und 0,1% der Gesunden. Der Test liefert also 0,1% falsch positive und 4% falsch negative Ergebnisse.

Frage. Was ist die Wahrscheinlichkeit, dass jemand krank ist, sofern er positiv getestet wurde?

- Die **a-priori-Wahrscheinlichkeit** beträgt $\mathbb{P}(K) = 10^{-4}$ sowie $\mathbb{P}(K^c) = 1 - 10^{-4}$.
- Wir kennen die **bedingten Wahrscheinlichkeiten** $\mathbb{P}(+ \mid K) = 0,96$ sowie $\mathbb{P}(+ \mid K^c) = 0,001$.
- Als **A-posteriori Wahrscheinlichkeit** erhalten wir nun:

$$\begin{aligned} \mathbb{P}(K \mid +) &= \frac{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(+ \mid K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,96 \cdot 10^{-4}}{0,96 \cdot 10^{-4} + 0,001 \cdot (1 - 10^{-4})} \\ &\approx 9,6\% \end{aligned}$$

Antwort. Die Wahrscheinlichkeit, dass man krank ist, wenn man positiv getestet ist, beträgt also (nur) 9,6%.

Bemerkung*. Ein Test hat üblicherweise eine Sensitivität und eine Spezifität. Die Sensitivität gibt an, welcher Anteil der tatsächlich infizierten positiv getestet werden. Die Spezifität gibt an, welcher Anteil der gesunden Menschen auch negativ getestet wird.

Beispiel (Aktuelle Corona-Zahlen). Bei den aktuellen Schnelltests gibt es (in etwa) eine falsch-positiven Rate von 2%, also $\mathbb{P}(B \mid K^c) = 2\%$, und eine falsch-negativen Rate von 20%, also $\mathbb{P}(- \mid K) = 20\%$. Bei einer Inzidenz von 150-200 pro 100.000 Einwohner pro Woche, einer Dunkelziffer nah bei 2 ergibt sich eine Schätzung der aktuell infizierten von

$$\mathbb{P}(K) \in [0.005, 0.01].$$

(Zum Vergleich: Die aktuell gemeldeten positiven Fälle liegen bei 0,0035).

Nun können wir wieder berechnen:

$$\begin{aligned}\mathbb{P}(K \mid +) &= \frac{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(+ \mid K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,8 \cdot \mathbb{P}(K)}{0,8\mathbb{P}(K) + 0,02(1 - \mathbb{P}(K))}\end{aligned}$$

Wir erhalten

- (a) Mit $\mathbb{P}(K) = 0,005$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 17\%$
- (b) Mit $\mathbb{P}(K) = 0,01$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 29\%$
- (c) Mit $\mathbb{P}(K) = 0,001$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 3,8\%$

Frage. Lohnt es sich die Schnelltests in den Schulen zu machen?

Also: Was ist $\mathbb{P}(K \mid -)$

Auch das lässt sich mit der gleichen Formel beantworten, mit $\mathbb{P}(- \mid K) = 0,2$ und $\mathbb{P}(- \mid K^c) = 0,98$ erhalten wir

- (a) Für $\mathbb{P}(K) = 0,005$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,1\%$
- (b) Für $\mathbb{P}(K) = 0,01$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,2\%$
- (c) Für $\mathbb{P}(K) = 0,001$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,02\%$

Stichwortverzeichnis

a-posteriori-Verteilung, 12
a-priori-Einschätzung, 12

Erwartungswert
bedingter, 10

Varianz, 6

Wahrscheinlichkeit
bedingte, 10

Wahrscheinlichkeitsverteilung
bedingte, 10