Inhaltsverzeichnis

(Ausschnitt)

Bedingte Wahrscheinlichkeiten I

- Oft interessant: Eintrittswahrscheinlichkeit eines Ereignisses B, wenn schon bekannt ist, dass ein (anderes) Ereignis A, in diesem Zusammenhang auch **Bedingung** genannt, eingetreten ist.
- Beispiele:
  - Werfen eines Würfels:
    Wahrscheinlichkeit dafür, eine 2 gewürfelt zu haben, falls bekannt ist, dass eine ungerade Zahl gewüfelt wurde $\rightsquigarrow 0$
    Wahrscheinlichkeit dafür, eine 3 oder 4 gewürfelt zu haben, falls bekannt ist, dass eine Zahl größer als 3 gewürfelt wurde $\rightsquigarrow \frac{1}{3}$
  - Wahrscheinlichkeit, aus einer Urne mit 2 schwarzen und 2 weißen Kugeln bei Ziehung ohne Zurücklegen in der zweiten Ziehung eine weiße Kugel zu ziehen, wenn bekannt ist, dass in der ersten Ziehung eine schwarze Kugel gezogen wurde $\rightsquigarrow \frac{2}{3}$
  - Wahrscheinlichkeit, dass man beim Poker-Spiel (Texas Hold'em) zum Ende des Spiels einen Vierling als höchstes Blatt hat, wenn man bereits ein Paar in der Starthand hält $\rightsquigarrow 0.8424\%$

Bedingte Wahrscheinlichkeiten II

- Offensichtlich sind diese Wahrscheinlichkeiten nur dann interessant, wenn das Ereignis A auch mit positiver Wahrscheinlichkeit eintritt.
- *Rückblick:*
  In deskriptiver Statistik: Begriff der *bedingten relativen Häufigkeiten*
  - $r(a_i|Y = b_j) := \frac{h_{ij}}{h_{\cdot j}} = \frac{r_{ij}}{r_{\cdot j}}$
  - $r(b_j|X = a_i) := \frac{h_{ij}}{h_{i\cdot}} = \frac{r_{ij}}{r_{i\cdot}}$

  für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.
- Analog zur Einschränkung der statistischen Masse bei bedingten relativen Häufigkeiten: Einschränkung von Ω auf bedingendes Ereignis A.

Bedingte Wahrscheinlichkeiten III

- Zur Ermittlung der bedingten Wahrscheinlichkeit, eine 3 oder 4 gewürfelt zu haben (Ereignis $B = \{3, 4\}$), falls bekannt ist, dass eine Zahl größer als 3 gewürfelt wurde (Ereignis $A = \{4, 5, 6\}$):
  - Berechnung der Wahrscheinlichkeit des gemeinsamen Eintretens der Bedingung A und des interessierenden Ereignisses B:
    $P(A \cap B) = P(\{3, 4\} \cap \{4, 5, 6\}) = P(\{4\}) = \frac{1}{6}$
  - Berechnung der Wahrscheinlichkeit des Eintretens der Bedingung A:
    $P(A) = P(\{4, 5, 6\}) = \frac{3}{6}$
  - Analog zum Fall relativer bedingter Häufigkeiten: Berechnung des Verhältnisses $\frac{P(A\cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$.

Bedingte Wahrscheinlichkeiten IV

Definition 7.1

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ mit $P(A) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit** *von B unter der Bedingung A.*

Satz 7.1

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$ mit $P(A) > 0$. Dann ist die Abbildung

$$P(\,\cdot\,|A) : \mathcal{F} \to \mathbb{R}; B \mapsto P(B|A)$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$ gemäß Definition 5.4, also auch $(\Omega, \mathcal{F}, P(\,\cdot\,|A))$ ein Wahrscheinlichkeitsraum.

Bedingte Wahrscheinlichkeiten V

- **Wichtig:** Satz 7.1 gilt nur dann, wenn das bedingende Ereignis A festgehalten wird. Bei der Abbildung $P(A|\,\cdot\,) : \mathcal{F} \to \mathbb{R}$ handelt es sich im Allgemeinen **nicht** (und bei strenger Auslegung von Definition 7.1 sogar **nie**) um ein Wahrscheinlichkeitsmaß.
- Aus Definition 7.1 folgt unmittelbar

$$P(A \cap B) = P(A) \cdot P(B|A) \tag{1}$$

  für $A, B \in \mathcal{F}$ mit $P(A) > 0$.
- Obwohl $P(B|A)$ nur für $P(A) > 0$ definiert ist, bietet es sich aus technischen Gründen an, Schreibweisen wie in Gleichung (1) auch für $P(A) = 0$ zuzulassen. In diesem Fall sollen Produkte, in denen neben (eigentlich) nicht definierten bedingten Wahrscheinlichkeiten mindestens ein Faktor 0 auftritt, definitionsgemäß ebenfalls den Wert 0 annehmen.

Bedingte Wahrscheinlichkeiten VI

- Mit dieser Konvention für Bedingungen mit Eintrittswahrscheinlichkeit 0 lässt sich durch wiederholtes Einsetzen von Gleichung (1) leicht der **Multiplikationssatz**

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

  für $A_1, \ldots, A_n \in \mathcal{F}$ herleiten.
- Mit dem Multiplikationssatz können insbesondere Wahrscheinlichkeiten in sogenannten Baumdiagrammen (für „mehrstufige" Zufallsexperimente) ausgewertet werden.
- In vielen Anwendungen sind häufig vor allem bedingte Wahrscheinlichkeiten bekannt (bzw. werden als bekannt angenommen).
- Es ist nicht immer einfach, die Angabe bedingter Wahrscheinlichkeiten auch als solche zu erkennen!

Bedingte Wahrscheinlichkeiten VII

- Beispiel:
  In einer kleinen Druckerei stehen 3 Druckmaschinen unterschiedlicher Kapazität zur Verfügung:
  - Maschine 1, mit der 50% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.1% ein fehlerhaftes Ergebnis,
  - Maschine 2, mit der 30% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.25% ein fehlerhaftes Ergebnis,
  - Maschine 3, mit der 20% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.5% ein fehlerhaftes Ergebnis.

  Sind die interessierenden Ereignisse gegeben durch
  - M_i: Maschine i wird zur Produktion des Druckjobs eingesetzt ($i \in \{1, 2, 3\}$),
  - F: Die Produktion des Druckjobs ist fehlerbehaftet,

  so sind insgesamt bekannt:

$$\begin{array}{ccc} P(M_1) = 0.5 & P(M_2) = 0.3 & P(M_3) = 0.2 \\ P(F|M_1) = 0.001 & P(F|M_2) = 0.0025 & P(F|M_3) = 0.005 \end{array}$$

Bedingte Wahrscheinlichkeiten VIII

- In der Situation des Druckmaschinen-Beispiels interessiert man sich häufig für die *unbedingte* Wahrscheinlichkeit, einen fehlerhaften Druckjob zu erhalten, also für $P(F)$.
- Diese Wahrscheinlichkeit kann mit einer Kombination von Gleichung (1) auf Folie 174 und der letzten Rechenregel von Folie 148 berechnet werden:

Satz 7.2 (Satz der totalen Wahrscheinlichkeit)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathcal{F}$ (für $n \in \mathbb{N}$) eine Zerlegung von Ω, es gelte also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{j=1}^{n} A_j = \Omega$. Für beliebige Ereignisse $B \in \mathcal{F}$ gilt dann

$$P(B) = \sum_{j=1}^{n} P(B|A_j) \cdot P(A_j) \ .$$

Bedingte Wahrscheinlichkeiten IX

- Im Druckmaschinen-Beispiel erhält man so:

$$\begin{aligned} P(F) &= P(F|M_1) \cdot P(M_1) + P(F|M_2) \cdot P(M_2) + P(F|M_3) \cdot P(M_3) \\ &= 0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2 \\ &= 0.00225 = 0.225\% \end{aligned}$$

- Die Wahrscheinlichkeit von $A \cap B$ kann mit Hilfe bedingter Wahrscheinlichkeiten in zwei Varianten berechnet werden:

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

- Dies kann (bei Kenntnis der restlichen beteiligten Wahrscheinlichkeiten!) dazu benutzt werden, Bedingung und interessierendes Ereignis umzudrehen. Ist z.B. $P(B|A)$ (sowie $P(A)$ und $P(B)$) gegeben, so erhält man $P(A|B)$ durch

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

Bedingte Wahrscheinlichkeiten X

- Wird dabei die unbedingte Wahrscheinlichkeit $P(B)$ mit dem Satz der totalen Wahrscheinlichkeit (Satz 7.2) berechnet, so erhält man:

Satz 7.3 (Formel von Bayes)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathcal{F}$ (für $n \in \mathbb{N}$) eine Zerlegung von Ω, es gelte also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{j=1}^n A_j = \Omega$. Sei $B \in \mathcal{F}$ ein weiteres Ereignis mit $P(B) > 0$. Dann gilt:

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{j=1}^n P(B|A_j) \cdot P(A_j)}, \qquad k \in \{1, \ldots, n\}$$

- Anwendung der Formel von Bayes im Druckmaschinen-Beispiel:
  Frage: Wenn eine Fehlproduktion aufgetreten ist, mit welchen Wahrscheinlichkeiten sind dann die verschiedenen Druckmaschinen für den Fehldruck „verantwortlich"?

Bedingte Wahrscheinlichkeiten XI

- Es interessiert nun also $P(M_i|F)$ für $i \in \{1, 2, 3\}$.
- Mit Formel von Bayes (ohne Verwendung des Zwischenergebnisses $P(F)$):

$$\begin{aligned}
P(M_1|F) &= \frac{P(F|M_1) \cdot P(M_1)}{\sum_{i=1}^3 P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.001 \cdot 0.5}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{2}{9} \\
P(M_2|F) &= \frac{P(F|M_2) \cdot P(M_2)}{\sum_{i=1}^3 P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.0025 \cdot 0.3}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{3}{9} \\
P(M_3|F) &= \frac{P(F|M_3) \cdot P(M_3)}{\sum_{i=1}^3 P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.005 \cdot 0.2}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{4}{9}
\end{aligned}$$

Beispiel: Fehler bei bedingten Wahrscheinlichkeiten I

(aus Walter Krämer: Denkste!, Piper, München, 2000)

- Häufig (auch in den Medien!) ist der Fehler zu beobachten, dass das bedingende und das eigentlich interessierende Ereignis vertauscht werden.
- Beispiel (Schlagzeile in einer Ausgabe der ADAC-Motorwelt):

Der Tod fährt mit!
Vier von zehn tödlich verunglückten Autofahrern trugen keinen Sicherheitsgurt!

- Bezeichnet S das Ereignis „Sicherheitsgurt angelegt“ und T das Ereignis „Tödlicher Unfall“, so ist hier die Wahrscheinlichkeit $P(\overline{S}|T) = 0.4$, also die Wahrscheinlichkeit, keinen Sicherheitsgurt angelegt zu haben, falls man bei einen Unfall tödlich verunglückt ist, mit 40% angegeben.
- Was soll die Schlagzeile vermitteln?
  - Unwahrscheinlich: Anschnallen ist gefährlich, da 6 von 10 verunglückten Autofahren einen Sicherheitsgurt trugen.
  - Wohl eher: Anschnallen ist nützlich!
- **Aber:** Zitierte Wahrscheinlichkeit ist absolut ungeeignet, um Nutzen des Anschnallens zu untermauern!

Beispiel: Fehler bei bedingten Wahrscheinlichkeiten II

(aus Walter Krämer: Denkste!, Piper, München, 2000)

- Stattdessen interessant: $P(T|\overline{S})$ vs. $P(T|S)$
  (*Wie stehen Wahrscheinlichkeiten, bei nicht angelegtem bzw. angelegtem Sicherheitsgurt tödlich zu verunglücken, zueinander?*)
- Aus $P(\overline{S}|T)$ kann Verhältnis $\frac{P(T|\overline{S})}{P(T|S)}$ nur berechnet werden, falls die Wahrscheinlichkeit bekannt ist, mit der ein Autofahrer angeschnallt ist:

$$\begin{aligned} P(T|S) &= \frac{P(T\cap S)}{P(S)} = \frac{P(S|T)\cdot P(T)}{P(S)} \\ P(T|\overline{S}) &= \frac{P(T\cap \overline{S})}{P(\overline{S})} = \frac{P(\overline{S}|T)\cdot P(T)}{P(\overline{S})} \\ \Rightarrow \frac{P(T|\overline{S})}{P(T|S)} &= \frac{P(\overline{S}|T)\cdot P(S)}{P(\overline{S})\cdot P(S|T)} \end{aligned}$$

- Für $P(S) = 0.9$: $\frac{P(T|\overline{S})}{P(T|S)} = 6$, also Risiko ohne Gurt 6 mal höher als mit Gurt.
- Für $P(S) = 0.2$: $\frac{P(T|\overline{S})}{P(T|S)} = \frac{1}{6}$, also Risiko ohne Gurt 6 mal niedriger als mit Gurt.

Stochastische Unabhängigkeit

- Analog zur Unabhängigkeit von Merkmalen in deskriptiver Statistik:

Definition 7.2 (Stochastische Unabhängigkeit)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. A und B heißen **stochastisch unabhängig** *bezüglich P, wenn gilt:*

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

- Ebenfalls analog zur deskriptiven Statistik sind Ereignisse genau dann unabhängig, wenn sich ihre unbedingten Wahrscheinlichkeiten nicht von den bedingten Wahrscheinlichkeiten — sofern sie definiert sind — unterscheiden:

Satz 7.4

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. Dann gilt

- *falls $P(A) > 0$:*
  A und B sind stochastisch unabhängig bzgl. $P \Leftrightarrow P(B|A) = P(B)$
- *falls $P(B) > 0$:*
  A und B sind stochastisch unabhängig bzgl. $P \Leftrightarrow P(A|B) = P(A)$

- Der Begriff „Stochastische Unabhängigkeit" ist auch für Familien mit mehr als zwei Ereignissen von Bedeutung:

Definition 7.3

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $(A_i)_{i \in I}$ eine Familie von Ereignissen aus $\mathcal{F}$. Die Familie heißt **stochastisch unabhängig**, *wenn für jede endliche Teilmenge $K \subseteq I$ gilt:*

$$P\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} P(A_i)$$

- Besteht die Familie $(A_i)_{i \in I}$ in Definition 7.3 nur aus den drei Ereignissen A_1, A_2, A_3, so ist die Familie also stochastisch unabhängig, falls
  1. $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$
  2. $P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$
  3. $P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$
  4. $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$

  gilt.
  Insbesondere genügt **nicht** die paarweise stochastische Unabhängigkeit der Ereignisse A_1, A_2, A_3!