4. Grundzüge der Wahrscheinlichkeitsrechnung

Andreas Eberle
Institut für angewandte Mathematik

Oktober 2008

- In der beschreibenden Statistik haben wir verschiedene Kennzahlen (Statistiken) für Stichproben aus einer Grundgesamtheit kennengelernt.
- In der Regel enthält die Stichprobe aber viel weniger statistische Einheiten als die Grundgesamtheit. Beispielsweise besteht die Grundgesamtheit bei einer Wählerbefragung aus allen Wahlberechtigten - in einer Stichprobe werden aber in der Regel nur 1000 bis 5000 Wahlberechtigte befragt.
- Wieso und wie können wir trotzdem Rückschlüsse von der Stichprobe auf die Grundgesamtheit ziehen ? Wenn die relative Häufigkeit einer Partei in der Stichprobe bei 40 Prozent liegt, können wir dann davon ausgehen, daß die relative Häufigkeit unter allen Wahlberechtigten auch ungefähr bei 40 Prozent liegt ? Wie groß sind die zu erwartenden Fluktuationen ?

- Das Entnehmen der Stichprobe aus der Grundgesamtheit ist (im Idealfall) ein **zufälliger Vorgang ("Zufallsexperiment").** Die beobachteten Merkmalsausprägungen $x_1, x_{2,...,}x_n$ können wir als Realisierungen von **"Zufallsgrößen"** $X_1, X_{2,...,}X_n$ betrachten.
- Um die Fragen von oben zu beantworten, benötigen wir deshalb mathematische Strukturen, mit deren Hilfe wir Zufallsvorgänge modellieren, und Wahrscheinlichkeiten berechnen können. Diese Strukturen stellt die **Wahrscheinlichkeitstheorie** bereit.

- Eine **Zufallsgröße** X ist eine Größe, deren Wert wir nicht exakt kennen bzw. vorhersagen können (aufgrund unbekannter Einflüsse, mangelnder Information, oder echten Zufalls). Wir können den möglichen Werten nur **Wahrscheinlichkeiten** zuordnen. Genau wie Merkmale, können Zufallsgrößen auch sowohl Werte mit quantitativer als auch mit qualitativer Ausprägung annehmen (z.B. könnte der Wert einer Zufallsgröße X, die das Geschlecht einer Versuchsperson beschreibt mit Wahrscheinlichkeit $1/2$ "weiblich" und mit Wahrscheinlichkeit $1/2$ "männlich" sein).
- Wenn wir festlegen, welcher Wert mit welcher Wahrscheinlichkeit auftritt, dann stellen wir ein **mathematisches Modell** für den Zufallsvorgang auf (aufgrund von logischen Überlegungen, wissenschaftlichen Erkenntnissen, Einschätzungen, Erfahrungen, Vorwissen und Intuition). Ob das Modell das Zufallsexperiment wirklich korrekt beschreibt, kann man im nachhinein testen (z.B. mithilfe statistischer Verfahren, s.u.).

Beispiel. (Entnahme einer Zufallsstichprobe $X_1, X_2, \ldots, X_n$ aus der Grundgesamtheit, z.B. bei der Wählerbefragung)

Hier könnte man ansetzen, daß die Zufallsgrößen X_i unabhängig sind, und mit einer Wahrscheinlichkeit p eine bestimmte Partei als Wert annehmen. Dies liefert für jeden Wert von p ein mathematisches Modell. Welches dieser Modelle das beste ist, versuchen wir im nachhinein - ausgehend von den berechneten Wahrscheinlichkeiten der Beobachtungsdaten in den verschiedenen möglichen Modellen - herauszubekommen.

- Mathematisch präzise beschreibt man eine Zufallsgröße X durch eine Funktion $X(\omega)$, die von einem "Zufallsparameter" ω abhängt, und spricht daher auch von einer **Zufallsvariable**. Wir werden den Begriff meist intuitiv verwenden, und nicht im Detail auf den mathematischen Formalismus eingehen.

4.1. Ereignisse und ihre Wahrscheinlichkeit

Ereignisse

In einem stochastischen (=wahrscheinlichkeitstheoretischen) Modell ordnen wir **"Ereignissen"** Wahrscheinlichkeiten zu. Sind X, Y, Z, etc. die uns interessierenden Zufallsgrößen, dann sind typische Ereignisse denen wir Wahrscheinlichkeiten zuordnen wollen, z.B.

- $X = a$: *"Die Zufallsgröße X hat den Wert a"*
- $X \leq 7$: *"Der Wert von X ist ≤ 7"*
- $X = 3$ oder $X > 7$: *"Der Wert von X ist 3 oder größer als 7"*
- $X < 5$ und $Y =$"Elefant"
- $\frac{1}{n}(X_1 + \cdots + X_n) > 1$: *"Der Mittelwert der Zufallsgrößen X_i ist größer als 1"*
- $N_{X_1,X_2,\ldots,X_n}(a) = 3$: *"Die Häufigkeit des Wertes a unter den Zufallsstichproben $X_1, X_{2,\ldots,}X_n$ beträgt 3"*

Bemerkung. Allgemein gilt: Sind wir nur an den Zufallsgrößen $X_1, X_{2,\ldots,}X_n$ interessiert, dann können wir jedes für uns relevante Ereignis in der Form

$$(X_1, X_{2,\ldots,}X_n) \in B : \qquad \text{"Der Wert von } (X_{1\ldots,}X_n) \text{ liegt im Bereich } B\text{"}$$

schreiben, z.B.

- $X \leq 7 \iff X \in (-\infty, 7]$
- $X = 3$ oder $X > 7 \iff X \in \{3\} \cup (7, \infty)$
- $Y > 5X + 3 \iff$ Der Punkt (X, Y) liegt oberhalb der Geraden $x \mapsto 5x + 3$

Die **Wahrscheinlichkeit** eines Ereignisses A (z.B. $A = "X > 7"$) bezeichnen wir meistens mit $P[A]$. Dabei steht P für "*probability*". Die Wahrscheinlichkeit liegt stets zwischen 0 und 1.

- Tritt das Ereignis nie ein, dann gilt $P[A] = 0$.
- Tritt das Ereignis mit Sicherheit ein, dann gilt $P[A] = 1$.
- *Allgemeiner:*
  Gilt $P[A] = p$ mit $0 \leq p \leq 1$, dann tritt das Ereignis bei einer großen Anzahl unabhängiger Wiederholungen in ca. $p \times 100\%$ der Fälle ein
  –> **Frequentistische Interpretation der Wahrscheinlichkeit**
  –> **Gesetz der großen Zahl** (s.u.)

Ereignisse und ihre Wahrscheinlichkeit

Regeln/Axiome

Wahrscheinlichkeiten erfüllen folgende Regeln, die sofort plausibel sind, oder mit elementaren Überlegungen hergeleitet werden können:

- **Additivität:** Sind A und B Ereignisse, die **nicht** (oder nur mit Wahrscheinlichkeit 0) **gleichzeitig** eintreten können, dann gilt

$$P[A \text{ tritt ein oder } B \text{ tritt ein}] = P[A \text{ tritt ein}] + P[B \text{ tritt ein}]$$

- **Gegenereignis:**

$$P[A \text{ tritt nicht ein}] = 1 - P[A \text{ tritt ein}]$$

Beweis: Das Ereignis A kann nicht gleichzeitig eintreten und nicht eintreten, also folgt wegen der Additivität:

$$\begin{aligned} & P[A \text{ tritt ein}] + P[A \text{ tritt nicht ein}] \\ = \quad & P[A \text{ tritt ein oder } A \text{ tritt nicht ein}] = 1 \end{aligned}$$

Einschluss-/Ausschlussprinzip: Für *beliebige* Ereignisse A, B gilt

$$\begin{aligned} & P[A \text{ tritt ein oder } B \text{ tritt ein}] \\ = \; & P[A \text{ tritt ein}] + P[B \text{ tritt ein}] - P[A \text{ tritt ein und } B \text{ tritt ein}] \end{aligned}$$

- *Beweis:* Dafür, daß A oder B eintritt, gibt es drei Möglichkeiten, die nicht gleichzeitig auftreten können:
  1. A tritt ein und B tritt nicht ein.
  2. B tritt ein und A tritt nicht ein.
  3. A und B treten beide ein.
- Also folgt aus der Additivität:

$$\begin{aligned} & P[A \text{ tritt ein oder } B \text{ tritt ein}] \\ = \; & \underbrace{P[A \text{ und nicht } B]}_{P[A]-P[A \text{ und } B]} + \underbrace{P[B \text{ und nicht } A]}_{P[B]-P[A \text{ und } B]} + P[A \text{ und } B] \\ = \; & P[A] + P[B] - P[A \text{ und } B] \end{aligned}$$

Das Einschluss-/Ausschlussprinzip und die anderen Regeln kann man sich auch gut mit Mengendiagrammen veranschaulichen (siehe Vorlesung).

Beispiel. (Wetter)

- *Die Wahrscheinichkeit, daß es morgen regnet, betrage 50 %, und die Wahrscheinlichkeit, daß morgen die Sonne scheint 60 %. Die Wahrscheinlichkeit, daß es regnet und die Sonne scheint, sei 30 %. Wie groß ist dann die Wahrscheinlichkeit, daß es weder regnet noch die Sonne scheint ?*
- *Nach dem Einschluss-/Ausschlussprinzip gilt:*

$$\begin{aligned} P[\text{Regen oder Sonne}] &= P[\text{Regen}] + P[\text{Sonne}] - P[\text{Regen und Sonne}] \\ &= 0.5 + 0.6 - 0.3 = 0.8 \end{aligned}$$

- *Also erhalten wir für das Gegenereignis:*

$$\begin{aligned} P[\text{kein Regen und keine Sonne}] &= 1 - P[\text{Regen oder Sonne}] \\ &= 0.2 = 20\% \end{aligned}$$

4.2. Bedingte Wahrscheinlichkeiten

- Angenommen, wir wollen die Wahrscheinlichkeit eines Ereignisses A bestimmen unter der Prämisse, daß wir schon wissen, daß ein anderes Ereignis B eintritt.
- Zum Beispiel wollen wir die Wahrscheinlichkeit bestimmen, daß ein aus allen Wahlberechtigten in den USA zufällig herausgegriffener Wähler für Obama stimmt unter der Prämisse, daß der Wähler jünger als 30 ist. Dann dürfen wir nur diejenigen Einheiten aus der Grundgesamtheit berücksichtigen, für die das Ereignis B="Alter<30" eintritt.

Definition

Die **bedingte Wahrscheinlichkeit** eines Ereignisses A gegeben ein Ereignis B ist

$$P[A|B] = \frac{P[A \text{ und } B]}{P[B]} = \frac{P[A \text{ und } B \text{ treten beide ein}]}{P[B \text{ tritt ein}]}$$

$$P[A|B] = \frac{P[A \text{ und } B]}{P[B]} = \frac{P[A \text{ und } B \text{ treten beide ein}]}{P[B \text{ tritt ein}]}$$

- Die bedingte Wahrscheinlichkeit ist also der Anteil der W'keit des Ereignisses "A und B" an der Gesamtwahrscheinlichkeit der Bedingung B.
- Gilt $P[B] = 0$, dann ist die bedingte W'keit nicht definiert.
-

Beispiel.

Die Wahrscheinlichkeit, daß es heute regnet, betrage 50 %, die Wahrscheinlichkeit, daß es heute und morgen regnet 40 %. Die bedingte W'keit, daß es morgen regnet, wenn es heute regnet, ist dann $0.4/0.5 = 0.8 = 80\%$.

- $$P[A|B] = \frac{P[A \text{ und } B]}{P[B]} = \frac{P[A \text{ und } B \text{ treten beide ein}]}{P[B \text{ tritt ein}]}$$

Beispiel.

Eine Familie hat zwei Kinder. Wie groß ist die Wahrscheinlichkeit,daß beides Mädchen sind, wenn mindestens eines ein Mädchen ist ?

- Es gibt vier mögliche Fälle: Junge-Junge, Junge-Mädchen, Mädchen-Junge, Mädchen-Mädchen.
- Wir nehmen an, daß alle vier Fälle gleich wahrscheinlich sind (stimmt in Wirklichkeit nicht ganz). Dann folgt:

$$\begin{aligned} P[MM] &= \frac{1}{4}, \\ P[\text{"mindestens ein Mädchen"}] &= P[MM] + P[MJ] + P[JM] = \frac{3}{4}, \end{aligned}$$

also

$$P[MM|\text{"mindestens ein Mädchen"}] = \frac{1/4}{3/4} = \frac{1}{3}.$$

Bedingte Wahrscheinlichkeiten

Satz von der totalen Wahrscheinlichkeit

- Oft kennt man nur die bedingten Wahrscheinlichkeiten unter verschiedenen Hypothesen H_i, möchte aber die absolute Wahrscheinlichkeit eines Ereignisses berechnen. Dies ist möglich, wenn wir wissen mit welcher Wahrscheinlichkeit die einzelnen Fälle/Hypothesen eintreten.

Theorem

Sind $H_1, H_2, \ldots, H_n$ Ereignisse, die nicht gleichzeitig eintreten können, und von denen stets eines eintritt (disjunkte Hypothesen), dann gilt:

$$P[A] = \sum_{i=1}^{n} P[A|H_i] \cdot P[H_i]$$

- Die bedingten W'keiten sind also mit der W'keit der jeweiligen Hypothese zu gewichten.

Theorem

Sind $H_1, H_2, \ldots, H_n$ Ereignisse, die nicht gleichzeitig eintreten können, und von denen stets eines eintritt (disjunkte Hypothesen), dann gilt:

$$P[A] = \sum_{i=1}^{n} P[A|H_i] \cdot P[H_i]$$

Beispiel.

- *Heute schneit es mit W'keit 10%, es regnet mit W'keit 50%, und mit W'keit 40% fällt kein Niederschlag. Wenn es heute schneit, dann schneit es morgen mit W'keit 50%; wenn es heute regnet, dann schneit es morgen mit W'keit 10%, und wenn heute kein Niederschlag fällt, dann schneit es morgen mit W'keit 20%. Wie groß ist die W'keit, daß es morgen schneit (Ereignis A)?*

- $H_1 =$ *"heute Schnee"*, $H_2 =$ *"heute Regen"*, $H_3 =$ *"heute trocken"*

$$\begin{aligned} P[A] &= P[A|H_1] \cdot P[H_1] + P[A|H_2] \cdot P[H_2] + P[A|H_3] \cdot P[H_3] \\ &= 0.5 \cdot 0.1 + 0.1 \cdot 0.5 + 0.2 \cdot 0.4 = 0.18 = 18\% \end{aligned}$$

Beispiel (Simpson Paradoxon, Bewerbungen in Berkeley)

- *Statistik 1:*

Männer	angenommen (A)	Frauen	angenommen (A)
2083	996	1067	349

$$P[A|M] \approx 0.48, \qquad P[A|F] \approx 0.33$$

- *Statistik 2* (genauere Analyse durch Unterteilung in Fachbereiche):

	Männer	A		Frauen	A	
Bereich 1	825	511	62%	108	89	82%
Bereich 2	560	353	63%	25	17	68%
Bereich 3	325	110	34%	593	219	37%
Bereich 4	373	22	6%	341	24	7%

$P[A|M, \text{Bereich } i] < P[A|F, \text{Bereich } i]$ für jeden einzelnen Bereich !!

- **Wie ist das möglich ???**

- Satz von der totalen Wahrscheinlichkeit:

$$
\begin{aligned}
P[A|M] &= \sum_{i=1}^{4} P[A|M, \text{Bereich } i] \cdot P[\text{Bereich } i|M] \\
P[A|F] &= \sum_{i=1}^{4} P[A|F, \text{Bereich } i] \cdot P[\text{Bereich } i|F]
\end{aligned}
$$

- In unserem Beispiel war jede der bedingten Erfolgswahrscheinlichkeiten, über die summiert wird, für Frauen größer als für Männer.
- Trotzdem gilt insgesamt: $P[A|M] > P[A|F]$
- Dies ist möglich, weil sich Frauen eher in Fächern mit geringen Zulassungsquoten, und Männer eher in Fächern mit hohen Zulassungsquoten beworben haben.
- Statistik 1 vermischt verschiedene Populationen (Fachbereiche), und legt deshalb eventuell eine falsche Schlußfolgerung nahe.

Bedingte Wahrscheinlichkeiten

Bayessche Regel

- Umgekehrt wollen wir nun nach der Wahrscheinlichkeit der Hypothesen H_i fragen, wenn wir wissen, daß ein bestimmtes Ereignis A eintritt.
- Sei $P[H_i]$ die Wahrscheinlichkeit der Hypothese H_i, wenn wir keine Informationen über das Eintreten anderer Ereignisse vorliegen haben. In der **Bayesschen Statistik** reflektiert $P[H_i]$ unser Vorwissen (aufgrund von vorhergehenden Untersuchungen, mathematischen Modellen, Naturgesetzen,...) oder auch einfach nur unsere subjektive Einschätzung über die vorliegende Situation **("a priori degree of belief").**
- Angenommen, wir wissen nun zusätzlich, daß ein Ereignis A eintritt, und wir kennen die bedingten Wahrscheinlichkeiten $P[A|H_i]$ für das Eintreten von A unter den verschiedenen Hypothesen H_i. Wie sieht dann unsere neue Einschätzung der Wahrscheinlichkeiten der H_i aus **("a posteriori degree of belief")**?

Bedingte Wahrscheinlichkeiten

Bayessche Regel

Theorem

Sind $H_1, H_2, \ldots, H_n$ Ereignisse, die nicht gleichzeitig eintreten können, und von denen stets eines eintritt (disjunkte Hypothesen), dann gilt:

$$P[H_i|A] = \frac{P[A|H_i] \cdot P[H_i]}{\sum_{j=1}^{n} P[A|H_j] \cdot P[H_j]} = const. \cdot P[A|H_i] \cdot P[H_i]$$

- Die *a posteriori Wahrscheinlichkeiten* der Hypothesen H_i sind also proportional zu den *a priori Wahrscheinlichkeiten* mulitpliziert mit den bedingten W'keiten $P[A|H_i]$ *(likelihood Funktion)*.
- Den Proportionalitätsfaktor erhält man, indem man das ganze so normiert, daß die Summe der a posteriori Wahrscheinlichkeiten aller Hypothesen gleich 1 ist.

Theorem

Sind $H_1, H_2, \ldots, H_n$ Ereignisse, die nicht gleichzeitig eintreten können, und von denen stets eines eintritt (disjunkte Hypothesen), dann gilt:

$$P[H_i|A] = \frac{P[A|H_i] \cdot P[H_i]}{\sum_{j=1}^{n} P[A|H_j] \cdot P[H_j]} = const. \cdot P[A|H_i] \cdot P[H_i]$$

- **Beweis:** Aus der Defition der bedingten Wahrscheinlichkeit erhalten wir durch Anwenden des Satzes von der totalen W'keit:

$$\begin{aligned} P[H_i|A] &= \frac{P[H_i \text{ und } A]}{P[A]} \\ &= \frac{P[A|H_i] \cdot P[H_i]}{\sum_{j=1}^{n} P[A|H_j] \cdot P[H_j]} \end{aligned}$$

Beispiel. (Falsche postive Tests)

Es werde mit einem Test eine bestimmte seltene Krankheit (Beispiel: Vogelgrippe) festgestellt. Die Krankheit tritt bei einem von 10.000 untersuchten Tieren auf. Der Test ist aber nicht 100% genau. Es gelte:

$P(+\|K) = 0.96$	*Der Test ist positiv bei 96% der kranken Tiere ...*
$P(+\|G) = 0.001$	*... und bei 0,1% der gesunden Tiere.*

Man sagt: Test falsch positiv: 4%, Test falsch negativ: 0,1%

- Der Test scheint recht zuverlässig zu sein.
- Mit welcher Wahrscheinlichkeit ist bei einem positiven Befund das Tier wirklich krank?

- **A priori Annahme** (basierend auf Statistik der Häufigkeit der Krankheit):

$$P[K] = \frac{1}{10000}, \qquad P[G] = 1 - P[K] = \frac{9999}{10000}.$$

- **Likelihood:**

$$P[+|K] = 0.96, \qquad P[+|G] = 0.001.$$

- **A posteriori Wahrscheinlichkeit** (nach Bayesscher Formel):

$$\begin{aligned} P[K|+] &= \frac{P[+|K] \cdot P[K]}{P[+|K] \cdot P[K] + P[+|G] \cdot P[G]} \\ &= \frac{0.96 \cdot 10^{-4}}{0.96 \cdot 10^{-4} + 10^{-3} \cdot 0.9999} = 0.088 = 8.8\% \end{aligned}$$

- Ohne zusätzliche Informationen muß man also davon ausgehen, daß über 90% der positiv getesteten Tiere in Wirklichkeit gesund sind !

Das Problem ist, dass es nur sehr wenige kranke Individuen gibt, und deswegen viel mehr gesunde Individuen getestet werden. Der "falsch positive" Ausgang des Tests schlägt deshalb stärker zu Buche als der "richtig positive" !

4.3. Unabhängigkeit von Ereignissen

- Wenn zwei Ereignisse A und B voneinander unabhängig sind, dann sollte die Wahrscheinlichkeit von A nicht davon beeinflusst werden, daß wir schon wissen, ob B eintritt (und entsprechend umgekehrt). Es sollte also gelten

$$P[A] = P[A|B] = \frac{P[A \text{ und } B]}{P[B]},$$

das heißt

$$P[A \text{ und } B] = P[A] \cdot P[B].$$

Definition

Die Ereignisse A und B heißen unabhängig, falls gilt:

$$P[A \text{ und } B \text{ treten ein}] = P[A] \cdot P[B].$$

- **"Unabhängigkeit $\Longleftrightarrow$ Multiplizieren der Einzelw'keiten"**

- Allgemeiner heißen Ereignisse $A_1, A_2, \ldots, A_n$ **unabhängig**, falls für jede Teilkollektion $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ dieser Ereignisse gilt:

$$P[A_{i_1} \text{ und } A_{i_2} \text{ und } .... \text{ und } A_{i_k} \text{ treten ein}] = P[A_{i_1}] \cdot P[A_{i_2}] \cdot \cdots \cdot P[A_{i_k}]$$

- **"Unabhängigkeit $\iff$ Multiplizieren der Einzelw'keiten"**

- Beispiel.

  Bei einer Wählerbefragung zur Präsidentenwahl in den USA nehmen wir an, daß die Angaben der Befragten unabhängig sind, und in 10% der Fälle die Auskunft verweigert wird. Wie groß ist die W'keit, daß unter 10 Befragten mindestens einer ist, der keine Auskunft gibt ?

  - Sei A_i das Ereignis

    $$A_i = \text{"der i-te Befragte gibt Auskunft"}.$$

    Wir betrachten zunächst das Gegenereignis, daß alle 10 Befragten Auskunft geben. Wegen der vorausgesetzten Unabhängigkeit gilt:

    $$\begin{aligned} P[\text{alle 10 geben Auskunft}] &= P[A_1 \text{ und } A_2 \text{ und } \dots\dots \text{ und } A_{10}] \\ &= P[A_1] \cdot P[A_2] \cdot \dots \cdot P[A_{10}] = 0.9^{10}. \end{aligned}$$

  - Also erhalten wir:

    $$\begin{aligned} P[\text{einer gibt keine Auskunft}] &= 1 - P[\text{alle 10 geben Auskunft}] \\ &= 1 - 0.9^{10} = 65\% \end{aligned}$$

- Betrachten wir den anderen Extremfall, daß sich die 10 Befragten abgesprochen haben, sich gleich zu verhalten. Dann würden mit W'keit 0.9 einer, und damit alle, Auskunft geben, und mit Wahrscheinlichkeit 0.1 würden alle die Auskunft verweigern.
- In diesem Fall wäre also

$$P[\text{einer gibt keine Auskunft}] = P[\text{alle 10 geben keine Auskunft}] = 10\%$$

- Dieses Ergebnis ist offensichtlich völlig anders als das bei Unabhängigkeit erhaltene !
- **Aussagen, die für unabhängige Ereignisse hergeleitet wurden, können in der Regel nicht auf abhängige Ereignisse übertragen werden**
  (.... es sei denn die Abhängigkeit ist so schwach, daß sie ohne größere Fehler vernachlässigt werden kann).