Bedingte Wahrscheinlichkeit I

„Herzoperation in Krankenhaus“

Überleben der Operation

Alle Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	500	500	0.5
Krankenhaus K	900	100	0.1

Frage: „In welchem Krankenhaus würden Sie sich behandeln lassen?“

Bedingte Wahrscheinlichkeit II

Schwere der behandelten Fälle

	schwere Fälle	leichte Fälle
Krankenhaus U	900	100
Krankenhaus K	100	900

Frage: „Bleiben Sie bei Ihrer Entscheidung?“

Bedingte Wahrscheinlichkeit III

Überleben der Operation aufgeteilt nach der Schwere der behandelten Fälle

Schwere Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	400	500	0.56
Krankenhaus K	30	70	0.7

Leichte Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	100	0	0
Krankenhaus K	870	30	0.033

Definition der bedingten Wahrscheinlichkeit

In dem Beispiel betrachten wir das Risiko gegeben „schwerer Fall“.
Das Risiko wird berechnet durch

$$\frac{\text{Anzahl (schwere Fälle und nicht überlebt)}}{\text{Anzahl(schwere Fälle)}}$$

Allgemein definieren wir die Wahrscheinlichkeit von „Ereignis B gegeben A“

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

Bedinte Wahrscheinlichkeit: Beispiel

B: Nicht überleben
A: Schwerer Fall

Krankenhaus U

$$\begin{aligned} P(B) &= 500/1000 = 0.5 \\ P(A) &= 900/1000 = 0.9 \\ P(A \cap B) &= 500/1000 = 0.5 \\ P(B|A) &= 0.5/0.9 = 0.56 \end{aligned}$$

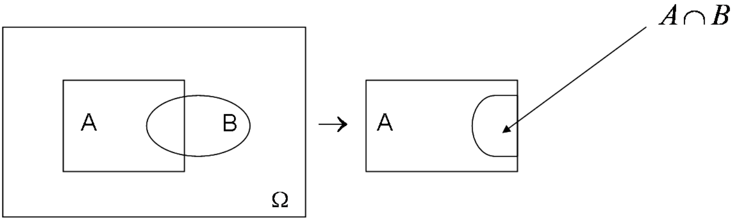
Krankenhaus K

$$\begin{aligned} P(B) &= 100/1000 = 0.1 \\ P(A) &= 100/1000 = 0.1 \\ P(A \cap B) &= 70/1000 = 0.07 \\ P(B|A) &= 0.07/0.1 = 0.7 = 70\% \end{aligned}$$

Schwere Fälle	OP überlebt	OP nicht überl.	P(nicht ü) „Risiko“
Krankenh U	400	500	0.56
Krankenh K	30	70	0.7

Leichte Fälle	OP überlebt	OP nicht überl.	P(nicht ü) „Risiko“
Krankenh U	100	0	0
Krankenh K	870	30	0.033

Einschränkung des Ergebnisraumes und bedingte Wahrscheinlichkeit

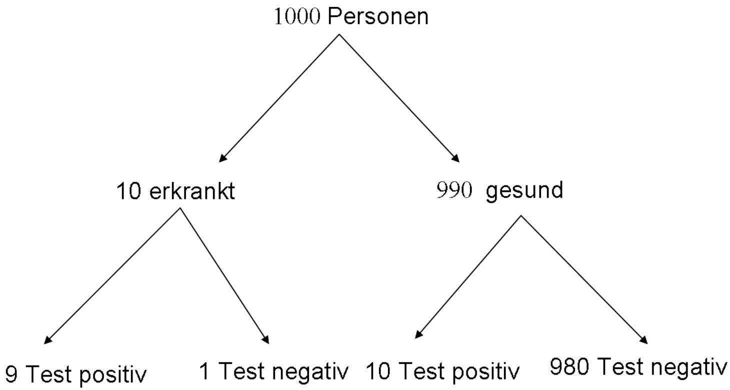


Beispiel: Würfeln

$$
\begin{aligned}
\Omega &= \{1,2,3,4,5,6\} \\
A &= \{2,4,6\} \quad \text{„gerade“} \\
B &= \{4,5,6\} \quad \text{„groß“} \\
A \cap B &= \{4,6\} \\
\\
P(A) &= 3/6 \\
P(A \cap B) &= 2/6 \\
P(B|A) &= P(A \cap B)/P(A) = (2/6)/(3/6) = 2/3
\end{aligned}
$$

Interpretation:
Wenn bekannt ist, dass die gewürfelte Zahl gerade ist, steigt die Wahrscheinlichkeit für „groß“ auf $2/3$.

Medizinische Tests



Bedingte Wahrscheinlichkeiten

Beachte: Die Bedingung entspricht der Bezugspopulation:

- 9 von 10 Kranken werden als solche erkannt:
  $P(\text{Test OK (positiv)} \mid \text{Patient krank}) = 9/10$
- 980 von 990 Gesunden werden als solche erkannt:
  $P(\text{Test OK (negativ)} \mid \text{Patient gesund}) = 98/99$
- 9 von 19 Patienten mit positivem Test sind tatsächlich krank:
  $P(\text{Diagnose OK (richtig)} \mid \text{Test positiv}) = 9/19$

Ohne Test: $P(\text{Patient krank}) = 1/100$

Bezugspopulation von zentraler Bedeutung

Beispiel: Wahlverhalten und Schicht

B:	zufällig gezogene Person wählt Partei X
A:	zufällig gezogene Person gehört der Oberschicht an
$P(B)$:	W'keit, dass zufällig gezogene Person Partei X wählt
$P(B \cap A)$:	W'keit, dass zufällig gewählte Person Partei X wählt *und* der Oberschicht angehört.

$P(B|A)$: W'keit, dass zufällig gewählte Person Partei X wählt, wenn sie der Oberschicht angehört.

Unterschied:
$P(B|A)$: Man betrachtet nur Angehörige der Oberschicht (A ist *sicher eingetreten*)

Interpretation analog zu bedingten Häufigkeiten in Statistik I: $P(B|A)$ ist die Wahrscheinlichkeit von B wenn man bereits weiß, dass A gilt.

Bezug zu Statistik I

- Die Beziehung zu Statistik I und den bedingten relativen Häufigkeiten ergibt sich, wenn man wieder die durch A und B erzeugte $(2 \times 2)-$Tafel betrachtet.
- An Stichprobenmodell denken: Grundgesamtheit Ω

$$P(B) \hat{=} f_{\bullet 1}, \quad P(A \cap B) \hat{=} f_{11}$$

$\hat{=}$ „entspricht“

	1	2	
1	f_{11}	f_{12}	$f_{1\bullet}$
2	f_{21}	f_{22}	$f_{2\bullet}$
	$f_{\bullet 1}$	$f_{\bullet 2}$	

Beispiel: Hörerbefragung

- 1000 HörerInnen
- 600 männlich (M), davon 450 positiv (Pos)
- 400 weiblich (W), davon 300 positiv (Pos)

Wir ziehen zufällig 1 Hörer.

$$
\begin{aligned}
P(M) &= 0.6 \\
P(W) &= 0.4 \\
P(Pos) &= 0.75 \\
P(M \cap Pos) &= 0.45 \\
P(Pos|M) &= P(M \cap Pos)/P(M) = 0.45/0.6 = 0.75 = P(Pos) \\
P(M|Pos) &= P(M \cap Pos)/P(Pos) = 0.45/0.75 = 0.6 = P(M)
\end{aligned}
$$

Interpretation: Die Ereignisse „Männlich“ und „Positiv“ sind unabhängig.

Definition stochastische Unabhängigkeit

Äquivalent sind unter $P(A) > 0$ und $P(B) > 0$

1. $P(B|A) = P(B)$
2. $P(A|B) = P(A)$
3. $P(A \cap B) = P(A) \cdot P(B)$

Definition

Zwei Ereignisse A und B heißen *stochastisch unabhängig* wenn gilt:

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

Diese Definition der Unabhängigkeit besitzt den Vorteil, dass man nicht $P(A) = 0$, $P(B) = 0$ ausschließen muss; für die Interpretation ist aber die Bezugnahme auf bedingte Wahrscheinlichkeiten viel anschaulicher.

Koppelung von unabhängigen Experimenten

Mit dem Begriff der Unabhängigkeit (und bedingten Wahrscheinlichkeiten) kann man komplexere Situationen aus „Einzelbausteinen“ zusammensetzen:

Bisher: Unabhängigkeit als zu überprüfende Eigenschaft

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) \quad \Longrightarrow \text{ unabhängig.}$$

Jetzt: Unabhängigkeit inhaltlich postulieren. Gegeben $P(A_1)$, $P(A_2)$ und A_1 und A_2 unabhängig. Dann gilt:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2).$$

Formale Darstellung

Experiment I (Ω_1, P_1) — Experiment II (Ω_2, P_2)

Unabhängigkeit

$\Downarrow$

Gesamtexperiment $(\Omega_1 \times \Omega_2, P)$ mit

$$P(A_1 \cap A_2)\,[= P(A_1 \times A_2)] := P_1(A_1) \cdot P_1(A_2)$$

Beispiel

Werfen eines Würfels ($\Omega_1 = \{1, \dots, 6\}$) und eines Oktaeders ($\Omega_2 = \{1, \dots, 8\}$) unabhängig voneinander.

$$A_1 \subset \Omega_1 : \quad A_1 = \{5, 6\}\ , \quad A_2 \subset \Omega_2 : \ A_2 = \{7, 8\}$$

$A_1 \cap A_2$: „5 oder 6 mit Würfel und 7 oder 8 mit Oktaeder“

Dann definiert man

$$P(A_1 \cap A_2)\ [:= P(A_1 \times A_2) =]\ = P_1(A_1) \cdot P_2(A_2);$$

also erhält man bei einem fairem Würfel und einem fairem Oktaeder mit

$$P_1(\{j\}) = \frac{1}{6},\ i = 1, \dots, 6, \quad \text{und} \quad P_2(\{j\}) = \frac{1}{8},\ i = 1, \dots, 8,$$

$$P(A_1 \cap A_2) \quad = \quad \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

Diese Konstruktion führt man für alle möglichen $A_1 \subset \Omega_1,\ A_2 \subset \Omega_2$ durch.

Wiederholungen eines Experiments

Von besonderer Bedeutung ist der Fall *unabhängiger und identischer Wiederholungen*, bei dem dasselbe Experiment wiederholt durchgeführt wird.

Zufallsstichprobe vom Umfang n mit Zurücklegen
Das Experiment „Ziehen einer Person und Ermittlung ihrer Parteipräferenz“ wird n-mal unabhängig (Befragte dürfen sich nicht gegenseitig beeinflussen!) durchgeführt.

Berechnungen und Beispiele folgen später mit Hilfe der Binomialverteilung

Koppelung abhängiger Experimente

Als nächster Schritt werden komplexere Experimente aus viel einfacheren, voneinander abhängigen Einzelexperimenten gebaut. Gerade bei komplexeren Anwendungen ist es meist bedeutend einfacher, (und auch sicherer, da sich die Chance erhöht, korrektes Expertenwissen zu erhalten) bedingte statt unbedingte Wahrscheinlichkeiten anzugeben. Beispielsweise kann man versuchen, die Wahrscheinlichkeit eines Ereignisses dadurch zu bestimmen, dass man als Zwischenschritt „auf alle Eventualitäten bedingt“ und zunächst die entsprechenden bedingten Wahrscheinlichkeiten bestimmt. ($\rightarrow$ Baumstruktur)

Fußball Beispiel

Wie groß ist die Wahrscheinlichkeit, das Halbfinale zu gewinnen ?
Gesucht: $P(B)$ mit $B =$ „Sieg im Halbfinale“
Siegchancen sind abhängig vom jeweiligen Gegner!
$\Longrightarrow$ bedingte Wahrscheinlichkeiten.

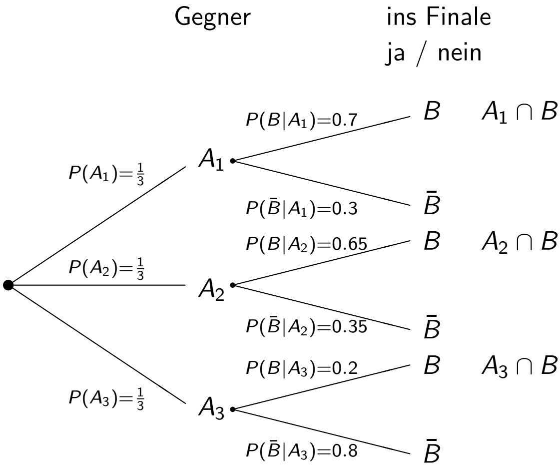
A_1	Gegner ist Mannschaft	1
A_2	"	2
A_3	"	3

Bedingte Wahrscheinlichkeiten leicht(er) anzugeben:

$$
\begin{aligned}
P(B|A_1) &= 0.7 \\
P(B|A_2) &= 0.65 \\
P(B|A_3) &= 0.2
\end{aligned}
$$

Gegner wird ausgelost $\Longrightarrow$ Annahme: $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$

Wahrscheinlichkeitsbaum (Fußball Beispiel)



Fußball Beispiel(2)

Welche „Wege“ im Wahrscheinlichkeitsbaum führen zu B?

$$\left.\begin{array}{rcl} P(A_1 \cap B) & = & P(A_1) \cdot P(B|A_1) = \frac{1}{3} \cdot 0.7 \\ P(A_2 \cap B) & = & P(A_2) \cdot P(B|A_2) = \frac{1}{3} \cdot 0.65 \\ P(A_3 \cap B) & = & P(A_3) \cdot P(B|A_3) = \frac{1}{3} \cdot 0.2 \end{array}\right\} \text{insgesamt: } 0.52$$

Verallgemeinerung: Vollständige Zerlegung

- A_1, A_2, A_3 bilden eine vollständige Zerlegung.
- $(A_1 \cap B)$, $(A_2 \cap B)$ und $(A_3 \cap B)$ sind disjunkt und ergeben in der Vereinigung B

Damit ergibt sich

$$\begin{aligned}
P(B) &= P\big((A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)\big) \\
&= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\
&= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) \;=\; 0.52
\end{aligned}$$

Entlang der Äste multiplizieren, dann summieren
Das Ergebnis lässt sich verallgemeinern auf:

- Beliebige Ereignisse B
- und vollständige Zerlegungen $(A_i)_{i=1,\ldots k}$.

Satz von der totalen Wahrscheinlichkeit

Gegeben sei eine vollständige Zerlegung $A_1, A_2, \ldots, A_k$. Dann gilt für jedes Ereignis B

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

Allgemeiner erlauben bedingte Wahrscheinlichkeiten die Modellierung komplexer „Experimente“, welche aus sukzessiven „Einzelexperimenten“ bestehen, bei denen die Ergebnisse jeweils von den vorherigen Experimenten abhängen dürfen (insb. dynamische stochastische Modelle).

Anwendungsbeispiele

- Komplexere Urnenmodelle ohne Zurücklegen, Wahrscheinlichkeit im n-ten Zug ist davon abhängig, welche Kugeln vorher gezogen wurden.
- Sicherheitsstudie zu Kernkraftwerken: Wahrscheinlichkeit für komplexe Pfade praktisch nicht angebbar, aber eben bedingte Einzelwahrscheinlichkeiten.
- Markovmodelle (dynamische Modelle mit „einfacher Bedingung“)

Markovmodelle I

- Koppelung abhängiger Experimente
  $\Omega_1 = \Omega_2 = \ldots = \Omega_T = \{a_1, \ldots, a_k\}$.
- Der Index $t = 1, \ldots T$ bezeichnet die Zeit.
- A_{tj} ist das Ereignis : *Zustand j zum Zeitpunkt t*

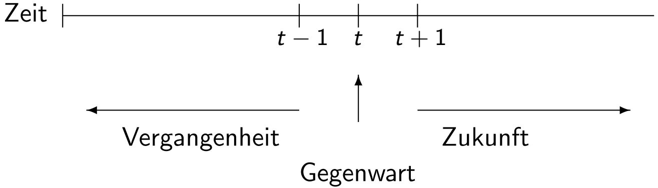
Markov–Eigenschaft:

$$P(A_{t+1,j_{t+1}} | A_{t,j_t} \cap A_{t-1,j_{t-1}} \cap \ldots) = P(A_{t+1,j_{t+1}} | A_{t,j_t}), \qquad (1.1)$$

Man spricht von einem *Markovmodell mit den Zuständen* $a_1, \ldots, a_k$.

Sind die so genannten Übergangswahrscheinlichkeiten in (1.1) unabhängig von der Zeit, gilt also $P(A_{t+1,j} | A_{tl}) \equiv p_{jl}$ für alle t, j, l, so heißt das Markovmodell *homogen*.

Markovmodelle II



Markov-Eigenschaft: „Gegeben den Zustand in der Gegenwart sind Vergangenheit und Zukunft unabhängig, d.h. die Zukunft hängt nur von der Gegenwart ab, aber nicht von der Vergangenheit“.

Für die Prognose der weiteren Entwicklung zählt also nur der aktuelle Stand, nicht aber, wie man dorthin gelangt ist.

Bei sozialen Prozessen ist die Markoveigenschaft und die Homogenität immer kritisch zu hinterfragen!

Markovmodelle: Typische Anwendungen:

- Glücksspiel: Die Wahrscheinlichkeit $P(A_{t+1,j})$ mit $A_{t+1,j}$ = „Spieler hat zum Zeitpunkt $t+1$ Kapitalbestand a_j“ hängt nur vom Kapitalbestand zum Zeitpunkt t ab.
- BWL: Konsumentscheidungen / Produktwahl
- Demographie: Geburts- und Todesprozesse
- Bildet das Wetter mit $\Omega =$ {Sonniger Tag, bewölkter Tag, regnerischer Tag, verschneiter Tag} eine Markovkette?
- Soziologie: z.B. Modelle sozialer Mobilität, Mobilität in Betrieben
    - Rapoport (1980): Mathematische Methoden in der Sozialwissenschaft, Physika
    - Bartholomew (1982): Stochastic Models for Social Processes, Wiley

Beispiel Markovmodelle: Soziale Mobilität

(nach Bartholomew (1982), S. 18f.)

Wie entwickelt sich der soziale Status durch die Generationen?

- Markoveigenschaft bedeutet hier: Status der Kinder nur abhängig vom Status der Eltern, aber nicht mehr zusätzlich vom Status der Großeltern oder vorheriger Generationen
- Homogenität bedeutet hier: Wahrscheinlichkeit für alle denkbaren Statuswechsel zeitlich konstant: z.B. Wechselhäufigkeit von Landwirtschaftssektor in Dienstleistungssektor?

Datengrundlage: männliche Generationenfolge in Marion County, Indiana (1905 – 1912)

Übergangsmatrix

Väter \ Söhne		a_1	a_2	a_3
nicht handwerkliche Tätigkeit	a_1	0.594	0.396	0.009
handwerkliche Tätigkeit	a_2	0.211	0.782	0.007
landwirtschaftliche Tätigkeit	a_3	0.252	0.641	0.108

Die obige Matrix enthält die (geschätzten) Übergangswahrscheinlichkeiten

$$i\text{-te Zeile, } j\text{-te Spalte:} \qquad P\left(A_{\mathbf{2j}} | A_{\mathbf{1i}}\right)$$

Wahrscheinlichkeit, dass die **zweite** Generation in Zustand **j** ist unter der Bedingung, dass die **erste** Generation im Zustand **i** ist.

Interpretation Übergangsmatrix

Beispiel: Sohn „nicht handwerklich“ unter der Bedingung Vater „landwirtschaftlich“

$$P(A_{21}|A_{13}) = 0.252$$

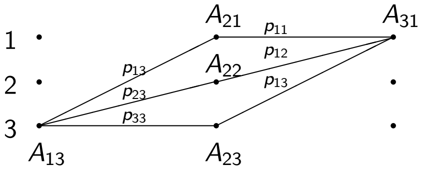
Für feste A_{1l} ist $P(A_{2j}|A_{1l})$ als Funktion in A_{2j} eine Wahrscheinlichkeits–verteilung, d.h. die Zeileneinträge summieren sich zu 1.

Inhaltliche Interpretation:

- Man sieht bei der handwerklichen Tätigkeit eine starke Tendenz zur Statuskonstanz ($P(A_{22}|A_{12}) = 0.782$)
- ähnliches gilt abgeschwächt für die nicht handwerkliche Tätigkeit ($P(A_{21}|A_{11}) = 0.594$),
- während sich der landwirtschaftliche Sektor deutlich auflöst; hier bleibt nur etwa jeder Zehnte ($P(A_{23}|A_{13}) = 0.108$), und ein „Zugewinn“ aus anderen Sektoren findet praktisch nicht statt ($P(A_{23}|A_{11}) = 0.009$, $P(A_{23}|A_{12}) = 0.007$ liegen jeweils unter einem Prozent).

Weitere Berechnungen

Wie groß ist die Wahrscheinlichkeit, dass der Enkel eines in der Landwirtschaft Tätigen eine Tätigkeit im nicht handwerklichen Sektor ausüben wird? Wahrscheinlichkeit, dass die 3. Generation nicht handwerklich unter der Bedingung, dass 1. Generation in Landwirtschaft



3. Generation
nicht handwerklich

erste Generation
Landwirtschaft

Formale Rechnung

$$
\begin{aligned}
P(A_{31}|A_{13}) &= \sum_{l=1}^{3} P(A_{31} \cap A_{2l}|A_{13}) \\
&= \sum_{l=1}^{3} P(A_{31}|(A_{2l} \cap A_{13})) \cdot P(A_{2l}|A_{13}) \\
&= \sum_{l=1}^{3} P(A_{31}|A_{2l}) \cdot P(A_{2l}|A_{13}) \\
&= \sum_{l=1}^{3} p_{l3} \cdot p_{1l} = p_{13} \cdot p_{11} + p_{23} \cdot p_{12} + p_{33} \cdot p_{13} \\
&= 0.252 \cdot 0.594 + 0.641 \cdot 0.211 + 0.108 \cdot 0.252 = 0.312
\end{aligned}
$$

Prognosen

Kennt man die Startverteilung $P(A_{11}), P(A_{12}), P(A_{13})$, so kann man die weitere Verteilung auf die Sektoren berechnen.

$$P(A_{2j}) = \sum_{m=1}^{3} P(A_{2j}|A_{1m}) \cdot P(A_{1m})$$

$$\begin{aligned}
P(A_{3j}) &= \sum_{l=1}^{3} P(A_{3j}|A_{2l}) \cdot P(A_{2l}) = \\
&= \sum_{l=1}^{3} P(A_{3j}|A_{2l}) \cdot \sum_{m=1}^{3} P(A_{2l}|A_{1m}) \cdot P(A_{1m}) = \\
&= \sum_{l=1}^{3} p_{jl} \cdot \sum_{m=1}^{3} p_{lm} \cdot P(A_{1m}) = \sum_{l=1}^{3} \sum_{m=1}^{3} p_{jl} p_{lm} \cdot P(A_{1m})
\end{aligned}$$

usw.

Man kann auch (mit weiterführenden Methoden) eine Gleichgewichtsverteilung bestimmen.

Kritische Aspekte

+ interessantes und sehr leistungsfähiges Modellierungsinstrument aber nicht in Ehrfurcht vor Methode erstarren, sondern Annahme kritisch hinterfragen

- Markoveigenschaft nicht unproblematisch: zusätzliche Rolle der Großväter!
- Zeitliche Homogenität problematisch (in der Tat gute 30 Jahre später p_{33} nochmals mehr als halbiert)