1.3.4 Koppelung abhängiger Experimente

Als nächster Schritt werden komplexere Experimente aus viel einfacheren, voneinander abhängigen Einzelexperimenten aufgebaut. 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)

Bsp. 1.34. [nach Fahrmeir et al., 2012, Kap. 4.6]

Eine Mannschaft gewinnt das Viertelfinalspiel. Wie groß ist die Chance, das Halbfinale zu gewinnen und ins Finale einzuziehen?

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:

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

Technisch entscheidend:

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

Es ergibt sich also

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

Merke:

Das Ergebnis lässt sich verallgemeinern auf

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

Satz 1.35. [Satz von der totalen Wahrscheinlichkeit]

Gegeben sei eine vollständige Zerlegung $A_1, A_2, \ldots, A_k$ von Ω (vgl. Bem. 1.21). Dann gilt für jedes Ereignis B

$$P(B) = \sum_{j=1}^{k} P(B|A_j) \cdot P(A_j) = \sum_{j=1}^{k} P(B \cap A_j). \tag{1.8}$$

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).

Arbeitet man mit mehreren abhängigen Experimenten, so ist folgende Folgerung aus dem Satz der totalen Wahrscheinlichkeit oft hilfreich:

Korollar 1.36.

Sei $A_1, A_2, \ldots, A_k$ eine vollständige Zerlegung. Dann gilt für beliebige Ereignisse B und C mit $P(C) > 0$

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

Beweisidee: $P(B|C)$ ist für festes C als Funktion in B eine Wahrscheinlichkeit (vgl. (1.7)). Wende den Satz der totalen Wahrscheinlichkeit auf diese Wahrscheinlichkeit an.

Koppelung abhängiger Experimente: Gegeben seien n Experimente, beschrieben durch die Grundräume $\Omega_i = \{a_{i1}, \ldots, a_{ik_i}\}$ und die Wahrscheinlichkeitsbewertungen $P_i, i = 1, \ldots, n$. Bezeichnet man für beliebiges $i = 1, \ldots, n$ und $j = 1, \ldots, k_i$, mit A_{ij} jeweils das zu a_{ij} gehörige Elementarereignis (also das Ereignis „a_{ij} tritt ein“), so gilt:

$$\begin{aligned} P\left(A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{nj_n}\right) &= P_1(A_{1j_1}) \cdot P_2\left(A_{2j_2}|A_{1j_1}\right) \cdot P_3\left(A_{3j_3}|A_{1j_1} \cap A_{2j_2}\right) \\ &\quad \cdot \ldots \cdot P_n\left(A_{nj_n}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{n-1j_{n-1}}\right) \end{aligned}$$

Häufig werden die Indizes bei P weggelassen. $i = 1, \ldots, n$ ist oft als Zeitpunkt interpretierbar.

$$P(\underbrace{A_{12} \cap A_{25} \cap A_{31}}_{C} \cap \underbrace{A_{42}}_{B})$$

(Wahl- oder Produktentscheidung)

Im Zeitpunkt 1	Produkt 2
Im Zeitpunkt 2	Produkt 5
Im Zeitpunkt 3	Produkt 1
Im Zeitpunkt 4	Produkt 2

Von hinten nach vorne: mit Formel $P(C \cap B) = P(B \cap C) = P(B|C) \cdot P(C)$:

$$P(\underbrace{A_{42}}_{B} \,|\, \underbrace{A_{31} \cap A_{25} \cap A_{12}}_{C}) \cdot P(\underbrace{A_{31} \cap A_{25} \cap A_{12}}_{C}) =$$

Satz wieder anwenden auf $P(\underbrace{A_{31}}_{B_{\text{neu}}} \cap \underbrace{A_{25} \cap A_{12}}_{C_{\text{neu}}})$:

$$P(A_{42}|A_{31} \cap A_{25} \cap A_{12}) \cdot P(A_{31}|A_{25} \cap A_{12}) \cdot P(A_{25} \cap A_{12}) =$$

$$P(A_{42}|A_{31} \cap A_{25} \cap A_{12}) \cdot P(A_{31}|A_{25} \cap A_{12}) \cdot P(A_{25}|A_{12}) \cdot P(A_{12})$$

Anwendungsbeispiele:

- Komplexere Urnenmodelle ohne Zurücklegen, Wahrscheinlichkeit im n-ten Zug ist davon abhängig, welche Kugeln vorher gezogen wurden.
- Sicherheitsstudie zu komplexen Anlagen: Wahrscheinlichkeit für komplexe Pfade praktisch nicht angebbar, aber eben bedingte Einzelwahrscheinlichkeiten (eventuell) schon.
- Mehrstufige Auswahl (z.B. Gemeinden $\rightarrow$ Schulen $\rightarrow$ Klassen $\rightarrow$ Schüler(innen))
- $i = 1, \ldots$ oft auch als Zeit interpretiert

* Betrachte die Entwicklung der Berufssektorenzugehörigkeit einer z.B. männlichen Generationenfolge A_{ij}, i-te Generation ist in Sektor j beschäftigt

* ähnlich soziale Mobilität: Sozialstatus über die Generationen

* Markovmodelle (dynamische Modelle mit „einfacher Bedingung“)

Markovmodelle:

Hier interpretiert man den Laufindex praktisch ausnahmslos als Zeit. Gilt in der Koppelung abhängiger Experiment $\Omega_1 = \Omega_2 = \ldots = \Omega_n = \{a_1, \ldots, a_k\}$ und sind alle bedingten Wahrscheinlichkeiten nur vom jeweils unmittelbar vorhergehenden Zeitpunkt abhängig, d.h. gilt

$$P(A_{i+1,j_{i+1}}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{ij_i}) = P(A_{i+1,j_{i+1}}|A_{ij_i}), \tag{1.9}$$

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

- Sind zudem die sogenannten *Übergangswahrscheinlichkeiten* in (1.9) unabhängig von der Zeit, gilt also $P(A_{i+1,j}|A_{i\ell}) \equiv p_{j\ell}$ für alle i, j, ℓ, so heißt das Markovmodell *homogen*.

Typische Anwendungen:

- Glücksspiel: Die Wahrscheinlichkeit $P(A_{i+1,j})$ mit $A_{i+1,j}$ = „Spieler hat zum Zeitpunkt $i+1$ Kapitalbestand a_j" hängt nur vom Kapitalbestand zum Zeitpunkt i ab, also nur von $A_{i1}, \ldots, A_{ik}$, nicht aber von früheren Ereignissen.
- BWL: Konsumentscheidungen / Produktwahl
- Demographie: Geburts- und Todesprozesse
- Epidemiologie
- Bildet das Wetter mit $\Omega = \{$Sonniger Tag, bewölkter Tag, regnerischer Tag, verschneiter T eine Markovkette?
- Soziologie: z.B. Modelle der Sektorenmobilität, soziale Mobilität in Betrieben
  - Rapoport (1980): Mathematische Methoden in der Sozialwissenschaft, Physika
  - Bartholomew (1982): Stochastic Models for Social Processes, Wiley

Bsp. 1.37. [Sektorenmobilität (nach Bartholomew (1982, S. 18f.)]

Wie entwickelt sich die Art der Erwerbstätigkeit über die Generationen?

- Markoveigenschaft bedeutet hier:
- Homogenität bedeutet hier:

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

Väter \ Söhne		a_1	a_2	a_3
nicht handwerkliche Tätigkeit ≈ Dienstleistung	a_1	0.594	0.396	0.009
handwerkliche Tätigkeit ≈ verarb. Gewerbe	a_2	0.211	0.782	0.007
landwirtschaftliche Tätigkeit ≈ Land- u. Forstwirtschaft	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.

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

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

- Man sieht: für feste A_{1l} ist $P(A_{2j}|A_{1l})$ als Funktion in A_{2j} eine Wahrscheinlichkeitsverteilung, d.h. die jeweiligen Zeileneinträge summieren sich (bis auf Rundungsfehler) zu 1.
- Inhaltliche Interpretation:
- Unter der Annahme, dass eine homogene Markov-Kette vorliegt, kann man mit den Daten weitere Entwicklungen prognostizieren.
- Mit Hilfe der Übergangsmatrix allein kann man Fragen der Art beantworten: Wie groß ist die Wahrscheinlichkeit, dass der Enkel eines in der Landwirtschaft Tätigen eine Tätigkeit im nicht handwerklichen Sektor ausüben wird?

- Beachte bisher ist nichts über die (unbedingte) Verteilung auf die einzelnen Sektoren ausgesagt. Kennt man zusätzlich die Startverteilung $P(A_{11}), P(A_{12}), P(A_{13})$, so kann man die weitere Verteilung auf die Sektoren berechnen.

- Für Nebenfachstudierende:

  * Diese Rechnungen lassen sich sehr einfach durch Matrixmultiplikation bewerkstelligen. Für die Verteilung nach n Schritten gilt:
    Startverteilung · (Übergangsmatrix)^n

  * Man kann auch eine Gleichgewichtsverteilung bestimmen. Unter Regularitätsbedingungen kann der asymptotische Gleichgewichtszustand über den zum größten Eigenwert gehörenden Eigenvektor bestimmt werden

- **Kritische Aspekte: Nicht nur für Nebenfachstudierende**

  + 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 absolut problematisch (in der Tat gute 30 Jahre später hat sich die Wahrscheinlichkeit, in der Landwirtschaft zu bleiben, nochmals mehr als halbiert)

1.3.5 Das Theorem von Bayes

Bei der Anwendung bedingter Wahrscheinlichkeiten ist es häufig von Interesse, „Bedingung und Ereignis“ zu vertauschen.

Also: gegeben $P(B|A)$, gesucht $P(A|B)$

Bsp. 1.38. [Diagnoseproblem]

- Bei der Durchführung eines Tests auf eine bestimmte Krankheit ist zu unterscheiden:

  * Patient ist krank $\longrightarrow$ Ereignis K

  * Testergebnis ist positiv, d.h. der Test sagt, die Person sei krank $\longrightarrow$ Ereignis T^+

  In der Praxis sind K und T^+ nie für alle Personen identisch!

  Ziel bei der Konstruktion eines Tests: möglichst geringe Fehlerwahrscheinlichkeiten

$P(T^+\|K)$	*Sensitivität:*	Wsk, dass Kranker als krank eingestuft wird
$P(\overline{T}^+\|\bar{K})$	*Spezifität:*	Wsk, dass Gesunder als gesund eingestuft wird
P(K)	*Prävalenz:*	Wsk, dass eine zufällig ausgewählte Person krank ist

Häufiges Problem in der Praxis: Steigerung der Sensitivität geht auf Kosten der Spezifität und umgekehrt.

- Jetzt konkrete Beobachtung bei einem Patienten: Testergebnis 'krank'. Mit welcher Wahrscheinlichkeit ist die Person tatsächlich krank?

Allgemeiner nicht nur bei Dichotomie K und $\bar{K}$, sondern bei beliebiger vollständiger Zerlegung $K_1, \ldots, K_k$ (vgl. Bem. 1.21) anwendbar:

Satz 1.39. [Theorem von Bayes]

Sei $A_1, \ldots A_k$ eine vollständige Zerlegung von Ω (wobei $P(A_i) > 0$, $P(B|A_i) > 0$, $i = 1, \ldots k$ und $P(B) > 0$ erfüllt seien.) Dann gilt

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)}.$$

Für beliebige Ereignisse A und B mit
$P(A) > 0, P(\bar{A}) > 0, P(B|A) > 0, P(B|\bar{A}) > 0, P(B) > 0, P(\bar{B}) > 0$ gilt also:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} \tag{1.10}$$

Bsp. 1.40. [Fortsetzung von Bsp. 1.38]

Gegeben seien

$$\begin{array}{lcll} P(T^+|K) & = & 0.98 & \text{(Sensitivität)} \\ P(\overline{T}^+|\bar{K}) & = & 0.97 & \text{(Spezifität)} \\ P(K) & = & 0.001 & \text{(Prävalenz)} \end{array}$$

Bem. 1.41. [Inhaltliche Bemerkungen zu Bsp. 1.38 und Bsp. 1.40]

- Die in solchen Situationen natürlicherweise (!) enstehende Zahl von falsch positiven Testergebnissen bewirkt ein substantielles Dilemma, das sehr kontrovers diskutiert wird. Problematik: Flächendeckendes Screening nicht unumstritten, da viele falsch-positive Ergebnisse. Also (?): Anwendung nur auf Risikopatienten.
- z.B. bei Mammographie oder PSA-Test auf Prostatakrebs teilweise sogar noch viel geringere Spezifität $\rightarrow$ Effekte noch drastischer.
- Wert der mathematischen Theorie: Wenn es etwas komplexer wird, verlässt einen sofort der „gesunde Menschenverstand“. Untersuchungen (u.a. von Gigerenzer [1]) haben gezeigt, dass viele Ärzte sich der hier ermittelten Problematik nicht bewusst sind.

[1] Gerd Gigerenzer (2002): Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken. Berlin Verlag. ISBN 9783827000798

Bem. 1.42. [Bemerkungen zu Satz 1.39 (Theorem von Bayes)]

- Übliche Bezeichnungen:

  $P(A_i)$: „a priori Wahrscheinlichkeiten" (Wahrscheinlichkeit vor der Beobachtung des Testergebnisses)

  $P(A_i|B)$: „a posteriori Wahrscheinlichkeiten" (Wahrscheinlichkeit $nach$ der Beobachtung des Testergebnisses)

  <u>Einschub:</u> Im Beispiel:

Bem. 1.43. *"Satz von Bayes in Odds-Form"*

Oft wird der durch den Satz von Bayes ausgedrückte Schlussprozess durch Beobachtungen (hier T^+) auch dadurch beschrieben, wie sich die Odds (z.B. auf "krank" versus "gesund" verändern)

∗ priori odds (ohne die aktuelle Beobachtung)

$\frac{P(K)}{P(\bar{K})}$

∗ posteriori odds (nach der Beobachtung des Testergebnisses)

$$\frac{P(K|T^+)}{P(\overline{K}|T^+}) = \frac{P(K \cap T^+)/P(T^+)}{P(\overline{K} \cap T^+)/P(T^+)} = \frac{P(K \cap T^+)}{P(\overline{K} \cap T^+)} = \frac{P(T^+|K) \cdot P(K)}{P(T^+|\overline{K}) \cdot P(\overline{K})} = \underbrace{\frac{P(T^+|K)}{P(T^+|\overline{K})}}_{\frac{\text{Sens}}{\text{1-Spec}}} \cdot \underbrace{\frac{P(K)}{P(\overline{K})}}_{\text{priori odds}}$$

analog:

$\frac{K|T^-}{\overline{K}|T^-} = \frac{T^-|K}{P(T^-|\overline{K})} \cdot \frac{P(K)}{P(\overline{K})}$

Im Bsp. priori odds: $\frac{0.001}{0.999} = 0.001$

posteriori odds: $\frac{0.98}{1-0.97} \cdot \frac{0.001}{0.999} = 0.033$

Die Odds haben sich also um den Faktor $\frac{0.98}{1-0.97} = 32.67$ erhöht.

- Im Prinzip liefert das Theorem von Bayes ein Schema für das probabilistische Lernen aus Beobachtungen („Aufdatieren von Wahrscheinlichkeiten“).

$$\left.\begin{array}{r}\text{priori}\\ +\ \text{Daten}\end{array}\right\} \longrightarrow \text{posteriori}$$

  Es dient als Grundlage der sog. *Bayesianischen Inferenz*, einer bestimmten Schule der statistischen Methodologie, die hier nicht behandelt wird. Dabei geht es darum, aus Daten zu lernen, indem man die subjektiven Wahrscheinlichkeiten $P(A_i)$ für bestimmte Modellparameter mit Hilfe der Daten (B) aufdatiert, und somit zur besseren Wahrscheinlichkeitsaussagen für die Modellparameter kommt.
  Eine elementare Erläuterung findet sich etwa bei Henk Tijms[1], der den Verdacht einer manipulierten Ziehung der Halbfinalpartien in der Champions League 2013 bayesianisch als didaktisches Beispiel untersucht.

- Hier Formulierung im medizinischen Kontext. Anwendung auch zur Beurteilung der Rückfallgefahr, Kreditwürdigkeitsprüfung, etc.

[1] Henk Tijms: Teaching Note-Was the Champions League Draw Rigged? `http://personal.vu.nl/h.c.tijms/TeachingNoteBayes.pdf`, aufgerufen am 25.04.2015