1.3.4 Koppelung abhängiger Experimente: Satz von der totalen Wahrscheinlichkeit, Markovmodelle

Gerade bei komplexeren Anwendungen ist es meist bedeutend einfacher, 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.

Bsp 1.13: (Fahrmeir et al, S. 209)
Mannschaft gewinnt Viertelfinalspiel: wie groß ist die Chance, das Halbfinale zu gewinnen und ins Finale einzuziehen?
Betrachte: Ereignis $B =$ „Sieg im Halbfinale"
gesucht: $P(B)$

Satz 1.14 *Satz von der totalen Wahrscheinlichkeit*
Gegeben sei eine vollständige Zerlegung $A_1, A_2 \ldots, A_k$. Dann gilt für jedes Ereignis B

$$P(B) = \sum_{j=1}^{k} P(B|A_j) \cdot P(A_j) \tag{1.3.4}$$

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

Bem 1.15 *Koppelung abhängiger Experimente*
Gegeben seien n Experimente, beschrieben durch die Grundräume $\Omega_i = \{a_{i1}, \ldots, a_{ik_i}\}$ und die Wahrscheinlichkeiten $P_i, i = 1, \ldots, n$. 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:

$$P\left(A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{nj_n}\right) = P_1(A_{1j_1}) \cdot P_2\left(A_{2j_2}|A_{1j_1}\right) \cdot$$

$$\cdot P_3\left(A_{3j_3}|A_{1j_1} \cap A_{2j_2}\right) \cdot \ldots \cdot P_n\left(A_{nj_n}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{n-1j_{n-1}}\right) \tag{1.3.5}$$

Wieder werden häufig die Indizes bei P weggelassen.

Arbeitet man mit mehreren abhängigen Experimenten, so ist folgende Folgerung aus Satz 1.14 oft hilfreich:

Korollar 1.16 zu Satz 1.14
Sei $A_1, A_2, \ldots, A_k$ eine vollständige Zerlegung. Dann gilt für beliebige Ereignisse B und C mit $P(C) > 0$

$$P(B|C) = \sum_{j=1}^{k} P(B|(A_j \cap C)) \cdot P(A_j|C) \tag{1.3.6}$$

Beweisidee: $P(B|C)$ ist für festes C als Funktion in B eine Wahrscheinlichkeit (vgl Bem. 1.9 e)). Wende Satz 1.14 auf diese Wahrscheinlichkeit an.

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

Def 1.17 *Markovmodelle*
Gilt in der Situation von Bem 1.15 $\Omega_1 = \Omega_2 = \ldots = \Omega_n = \{a_1, \ldots, a_k\}$ und sind alle bedingten Wahrscheinlichkeiten nur vom jeweils unmittelbar vorhergehenden Zeitpunkt abhängig, d.h. gilt

$$P(A_{i+1,j_{i+1}}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{ij_i}) = P(A_{i+1,j_{i+1}}|A_{ij_i}) \tag{1.3.7}$$

so spricht man von einem Markovmodell mit *den Zuständen* $a_1, \ldots, a_k$. Sind die sog. Übergangswahrscheinlichkeiten in (1.3.7) unabhängig von der Zeit, gilt also $P(A_{i+1,j}|A_{il}) \equiv p_{jl}$ für alle i, j, l, so heißt das Markovmodell *homogen.*

Typische Anwendungen:

- Glücksspiel: $P(A_{i+1,j})$ mit $A_{i+1,j}$ „Spieler hat zum Zeitpunkt $i+1$ Kapitalbestand a_j“ hängt nur von dem Kapitalbestand zum Zeitpunkt i ab, also nur von $A_{i1}, \ldots A_{ik}$, nicht aber von früheren Ereignissen.

- BWL: Konsumentscheidungen//Produktwahl
  homogen/nicht homogen

- Suchtforschung: Ω ={abhängig, clean}
  hier Markovannahme sehr problematisch

- Demographie: Geburts- und Todesprozesse

- Epidemiologie

- Soziologie: z.B. Modelle sozialer Mobilität, Mobilität in Betrieben

  * Rapoport (1980): Mathematische Methoden in der Sozialwissenschaft, Physika
  * Bartholomew (1982^3): Stochastic Models for Social Processes, Wiley

Beispiel 1.18: Soziale Mobilität
Wie entwickelt sich der soziale Status durch die Generationen?

- Markoveigenschaft bedeutet hier:

- Homogenität bedeutet hier:

nach Bartholomew (1982[3], S. 18f.)
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

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

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

- 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?

- Kennt man die Randverteilung, so kann man die weitere Verteilung auf die Sektoren berechnen.

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

- Kritische Aspekte:

  * Markoveigenschaft nicht unproblematisch:

  * Zeitliche Homogenität nicht unproblematisch

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 1.19: *Diagnoseproblem* (auch Anwendung in Expertensystemen)
Durchführung eines Tests (im umgangssprachlichen Sinn), z.B Test auf Krankheit (auch: Beurteilung der Rückfallgefahr, Kreditwürdigkeitsprüfung,...)
Hier im medizinischen Kontext formuliert.

- Zu unterscheiden:

  * Patient ist krank $\longrightarrow$ Ereignis A

* $*$ Testergebnis ist 'positiv', d.h. der Test sagt, die Person sei krank $\longrightarrow$ Ereignis B

In der Praxis sind A und B nie identisch!
Ziel: möglichst geringe Fehlerwahrscheinlichkeiten

$P(B\|A) = 0.98$	*Sensitivität:*	Kranker wird als krank eingestuft
$P(\bar{B}\|\bar{A}) = 0.97$	*Spezifität:*	Gesunder wird als gesund erkannt

Gegenläufiger Zusammenhang: Steigerung der Sensitivität geht auf Kosten der Spezifität.
Sensitivität und Spezifität sind gewöhnlich aus langjähriger Erfahrung (und Zulassungsverfahren) bekannt.

- Jetzt konkrete Beobachtung bei einem Patienten. Test zeigt 'krank'; mit welcher Wahrscheinlichkeit ist die Person tatsächlich krank?

  D.h. gesucht: $P(A|B)$ aus $P(B|A)$

Satz 1.20 *Theorem von Bayes*
Sei $A_1, \dots A_k$ eine vollständige Zerlegung von Ω (wobei $P(A_i) > 0$, $P(B|A_i) > 0$, $i = 1, \dots k$ und $P(B) > 0$ erfüllt seien.) Dann gilt

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)} . \tag{1.3.9}$$

Bsp 1.21: Fortsetzung von Bsp 1.19
Sei $P(A) = 0.001$

Vgl. etwa (sehr kritisch): Diskussionsbeitrag der Landesärztekammer Baden-Württemberg zum Mammographie-Screening. http://www.aerztekammer-bw.de/25/ressourcen/screening.pdf

Bem 1.22: (Zum Theorem von Bayes)

a) übliche Bezeichnung

   $P(A_i)$: „a priori Wahrscheinlichkeiten" (Wsk **vor** der Beobachtung des Testergebnisses, bei Krankheiten „Prävalenz")

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

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

Es dient als Grundlage der sog. *Bayesianischen Inferenz*, einer bestimmten Schule der statistischen Methodologie, die hier praktisch 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.

1.4 Zufallsvariable, Verteilungsfunktion, Dichte

1.4.1 Diskrete Zufallselemente und Zufallsvariablen

Def und Bem 1.24

a) Gegeben seien ein diskreter, d.h. höchstens abzählbarer, Ergebnisraum Ω und die Wahrscheinlichkeit P auf Ω.
Jede Abbildung

$$\begin{aligned} X : \Omega &\mapsto \Omega_X \\ \omega &\mapsto X(\omega) \end{aligned}$$

heißt Zufalls*element*. Setzt man für jede *Realisation* $x \in \Omega_X$

$$P_X(\{x\}) := P(\{X = x\}) := P(\{\omega | X(\omega) = x\}), \tag{1.4.1}$$

so erhält man eine Wahrscheinlichkeit auf Ω_X. (Oft wird auch $P(X = x)$ statt $P(\{X = x\})$ geschrieben.)

b) Ist $\Omega_X = \mathbb{R}$, so bezeichnet man das Zufallselement X als *Zufallsvariable*.

(In der Literatur Zufalls*elemente* relativ selten verwendet, gerade aber in den Sozialwissenschaften oft nicht reelle Zahlen im Sinne einer metrischen Skala: Zufallselemente entsprechen nominal skalierten Merkmalen.)

Bem und Bsp 1.25: Standardanwendung in der Statistik
Betrachtet werde die Situation von Beispiel 1.4 und 1.11 f.
Gegeben Grundgesamtheit $\widetilde{\Omega}$ (hier: alle Wähler)
reine Zufallsauswahl:

$$\text{Ergebnisraum} \quad \Omega = \widetilde{\Omega} \times \widetilde{\Omega} \times \ldots \times \widetilde{\Omega}$$

mit typischem Ergebnis $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$
(hier: ω_i zufällig beim i-ten Zug gezogener Wähler)

$$\text{Merkmal} \quad \widetilde{X} : \widetilde{\Omega} \longrightarrow \{\text{SPD, CDU/CSU,...}\}$$

individuelle Wahlentscheidung jedes Wählers $\tilde{\omega} \in \widetilde{\Omega}$, $\widetilde{X}(\tilde{\omega})$ von $\tilde{\omega}$ gewählte Partei.
Betrachtet werden die Ereignisse A_{ij}: i-te gezogene Person hat Merkmalsausprägung a_j, jetzt durch Zufallselement beschreibbar:

Sei X_i die „Auswertung des Merkmals $\widetilde{X}$ an der i-ten zufällig ausgewählten Person", d.h. an ω_i, so ist X_i ein Zufallselement.

$$\begin{array}{rcl} \text{Abb. } X_i\colon \Omega & \longrightarrow & \Omega_X = \{a_1, \ldots, a_k\} \\ \omega & \longmapsto & \widetilde{X}(\omega_i) \end{array}$$

A_{ij} läßt sich dann schreiben als

$$\{X_i = a_j\}$$

Es gilt also für jedes i und j (vgl. auch (1.2.8))

$$\begin{array}{rcl} P_{X_i}(\{a_j\}) & = & P(\{X_i = a_j\}) = P(A_{ij})\,, \\ \text{also} & & P(\{X_i = a_j\}) = f_j \end{array} \tag{1.4.2}$$

Die Wahrscheinlichkeitsverteilung des Zufallselements X_i (Stichprobe!) spiegelt also genau die Häufigkeitsverteilung des Merkmals $\widetilde{X}$ (Grundgesamtheit!) wider.
Fasst man man die einzelnen X_i zusammen, so bezeichnet man den Vektor $(X_1, X_2, \ldots, X_n)$ als *i.i.d. Stichprobe* oder *reine Zufallsstichprobe* des Merkmals $\widetilde{X}$. Die Abkürzung *i.i.d.* steht für

- **i**ndependently (die einzelnen Ziehungen sind stochastisch unabhängig)
- **i**dentically **d**istributed"(jedes X_i besitzt dieselbe Wahrscheinlichkeitsverteilung)

Nach dem Durchführen des Zufallsexperiments und der Auswertung von $\widetilde{X}$ erhält man die Realisationen $x_1 := X_1(\omega_1), x_2 := X_2(\omega_2), \ldots, x_n := X_n(\omega_n)$, also einen Vektor $(x_1, x_2, \ldots x_n)$, der formal korrekt als Realisation oder *Stichprobenrealisation* der *i.i.d.* Stichprobe $(X_1, X_2, \ldots X_n)$ bezeichnet werden würde, allgemein üblich aber einfach auch als *Stichprobe* bezeichnet wird.

Man nimmt diese Stichprobe als Realisation der Stichprobe $X_1, \ldots X_n$ und versucht jetzt auf auf die Grundgesamtheit, genauer auf die $f_1, \ldots f_n$, zu schließen.

Koppelt man die einzelnen Zufallsexperimente, so kann man die sogenannte *gemeinsame Verteilung* der $X_1, X_2, \ldots X_n$ berechnen.

$$\begin{array}{rl} & P(\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \ldots \cap \{X_n = x_n\}) \\ = & P(\{X_1 = x_1\}) \cdot P(\{X_2 = x_2\}) \cdot \ldots \cdot P(\{X_n = x_n\}) \end{array}$$

und damit, unter Verwendung von (1.4.2) für jede potentielle Stichprobe(nrealisation) die Wahrscheinlichkeit, genau sie zu erhalten. (siehe auch Beispiel 1.12)

1.4.2 Die Verteilungsfunktion

Betrachtet werde in diesem Abschnitt eine Zufalls**variable** X, also ein Zufallselement mit reellwertigen Realisationen.

Bsp. 1.26:
Sei X die Zufallsvariable *Anzahl der Haushaltsmitglieder* mit der Verteilung
P({X=1})=0.4
P({X=2})=0.3
P({X=3})=0.2
P({X=4})=0.1
(Annahme: Nur bis zu 4-Personen-Haushalte).
Man berechne die Wahrscheinlichkeit, bei reiner Zufallsauswahl vom Umfang 1 einen Mehrpersonenhaushalt zu erhalten und die Wahrscheinlichkeit des Ereignisses „Die Zahl der Haushaltsmitglieder ist gerade".

Bem 1.27: Die Wahrscheinlichkeitsverteilung einer Zufallsvariable X kann man durch die *Verteilungsfunktion*

$$F(x) := P(X \leq x) \tag{1.4.3}$$

eineindeutig erklären.
Die Wahrscheinlichkeit anderer Ereignisse ergibt sich aus dem (geeignet verallgemeinerten) dritten Kolmogorovschen Axiom.
Es gilt dann

$$P(a < X \leq b) = F(b) - F(a), \tag{1.4.4}$$

denn (kleine Übung zum Rechnen mit Wahrscheinlichkeiten)

Bsp 1.28: Fortsetzung von Bsp 1.26

P({X=1})=0.4
P({X=2})=0.3
P({X=3})=0.2
P({X=4})=0.1

Berechne die Verteilungsfunktion und zeichne sie.

Man sieht generell:
$P(X = x)$ ist genau die Sprunghöhe der Verteilungsfunktion im Punkt x.

Bsp 1.29: Fortsetzung von Bsp 1.26

Berechne: $P(2.5 < X \leq 3.5)$
$P(1 < X \leq 3)$
$P(1 \leq X \leq 3)$

1.4.3 Stetige Zufallsvariablen

Zufallsvariable $X: \; \Omega \longrightarrow \Omega_X = \mathbb{R}$

Jetzt: Ω und Ω_X überabzählbar

Vorstellung (vgl. oben): Auswertung eines *stetigen* Merkmals $\tilde{X}$ an zufällig ausgewählter Person aus einer *unendlich* großen Grundgesamtheit.

Problem: Wahrscheinlichkeit, genau einen bestimmten Wert x (z.B. Haushaltseinkommen=179385.17) zu erhalten ist Null!

(stetiges Merkmal, beliebig große Messgenauigkeit)

$$P_X(\{x\}) = 0 \quad \text{für jedes } x \in \mathbb{R} \tag{1.4.5}$$

Hierdurch ist kein Wahrscheinlichkeitsmaß festlegbar, man muss anders vorgehen.

Idee: Verteilungsfunktion betrachten, d.h. $P(\{X \leq x\})$ spezifizieren.

In der Tat ist - unter gewissen Regularitätsbedingungen - Bem 1.27 nach wie vor gültig:

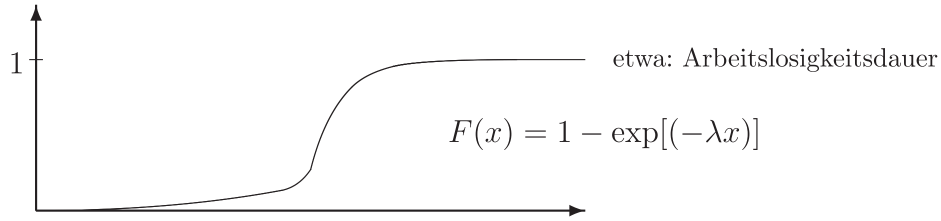
- Die Wahrscheinlichkeitsverteilung ist durch die Verteilungsfunktion

  $$F(x) = P(X \leq x)$$

  eindeutig festgelegt;
  für andere Ereignisse ergibt sich die Wahrscheinlichkeitsbewertung über das dritte Kolmogorovsche Axiom (bzw. einer Verallgemeinerung davon für abzählbar viele Ereignisse).
  Insbesondere:

Die Verteilungsfunktion einer stetigen Zufallsvariablen ist stetig

Typische Verteilungsfunktion



Die Kurve ist unterschiedlich steil; sie hat zwar in keinem Punkt eine Sprungstelle ($P(X = x) = 0$), aber in jedem kleinen Intervall um x ist:

$$P(x - h < X < x + h) = F(x + h) - F(x - h)$$

durchaus unterschiedlich. Die „Steilheit“

$$\lim_{h \to 0} \frac{F(x + h) - F(x - h)}{h}$$

enthält also wesentliche Information über $P \Longrightarrow$ Ableitung betrachten!

Def 1.30 Gegeben sei eine stetige Zufallsvariable X mit differenzierbarer Verteilungsfunktion $F_X(x)$.
Dann heißt die Ableitung von $F(x)$ nach x, also

$$f(x) = \frac{dF(x)}{dx} \tag{1.4.7}$$

Dichte der Zufallsvariablen X.

Umkehrung der Differentiation: Integration:

Satz 1.31 Es gilt dann in der Situation von Def 1.30

$$F(x) = \int_{-\infty}^{x} f(u)\, du \tag{1.4.8}$$

und damit für beliebige reelle Zahlen a und b mit $a < b$

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) = P(a < X < b) \\ &= P(a \leq X < b) = \int_a^b f(x)\, dx. \end{aligned} \tag{1.4.9}$$

Bsp 1.32
Gegeben sei eine Zufallsvariable X mit Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{40} \cdot x & x \in [0, 40] \\ 1 & x > 40 \end{cases}$$

Bestimmen Sie die Dichte $f(x)$ von X, skizzieren Sie $f(x)$ und interpretieren Sie $f(x)$ anschaulich!

Bei der Modellbildung geht man auch häufig umgekehrt vor:
Gib Dichte an, damit Verteilungsfunktion (fast) eindeutig bestimmt!

Jede Funktion f auf $\mathbb{R}$ mit $f(x) \geq 0$ für alle x und

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

kann als Dichte verwendet werden. Man erhält die Verteilungsfunktion gemäß Satz 1.31 durch

$$F(x) = \int_{-\infty}^{x} f(u)\,du$$

und das Wahrscheinlichkeitsmaß P über

$$P(a < X \leq b) = \int_{a}^{b} f(x)\,dx.$$

Bsp 1.33

Gegeben sei die Funktion

$$f_c(x) \quad = \quad \begin{cases} c \cdot x & x \in [0,1] \\ 0 & \text{sonst,} \end{cases}$$

abhängig von einem Parameter c.

a) Wie ist c zu wählen, dass f_c eine Dichte ist?

b) Berechnen Sie die Verteilungsfunktion und $P(X \in [\frac{1}{4}, \frac{3}{4}])$!

1.4.4 Lebensdauern; Hazardrate und Survivorfunktion

moderner Zweig vieler empirischer Untersuchungen: Lebensdaueranalyse, Ereignisanalyse $\longrightarrow$ Lehrempfehlung der DGS

Hier nur kurz. Weiterführend:

- Rohwer und Pötter (2001): *Grundzüge der sozialen Statistik*, Teil III. Juventa, Soziologische Grundlagentexte.
- Blossfeld, Hamerle, Mayer (1986): *Ereignisanalyse: statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften.* Campus.
- Diekmann und Mitter (1984): *Methoden zur Analyse von Zeitverläufen.* Teubner.
- Blossfeld und Rohwer (1995): *Techniques of Event History Modelling.* Erlbaur.

Betrachtet wird die Zufallsgröße „Zeit bis zu einem Ereignis“: Tod, Rückkehr aus Arbeitslosigkeit, Konkurs

<u>Bem 1.34</u> Die Verteilung einer nicht negativen stetigen Zufallsvariable X wird auch eineindeutig durch die sog. *Überlebensfunktion* (Survivorfunktion)

$$S(x) := P(X \geq x) = 1 - F(x) \tag{1.4.10}$$

und durch die *Hazardrate*

$$\lambda(x) := \lim_{h \to 0} \frac{P(x \leq X \leq x + h | X \geq x)}{h} \tag{1.4.11}$$

beschrieben.

Es gilt:

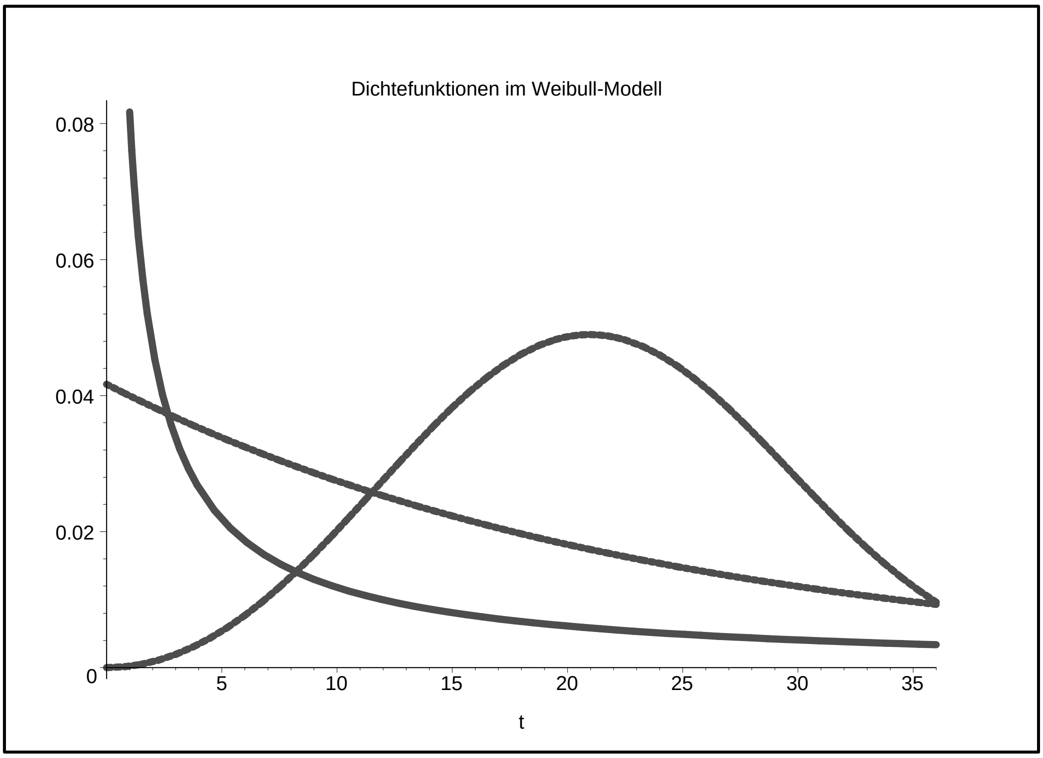
$$S(x) = \exp(-\int_0^x \lambda(u) du) \tag{1.4.12}$$

also

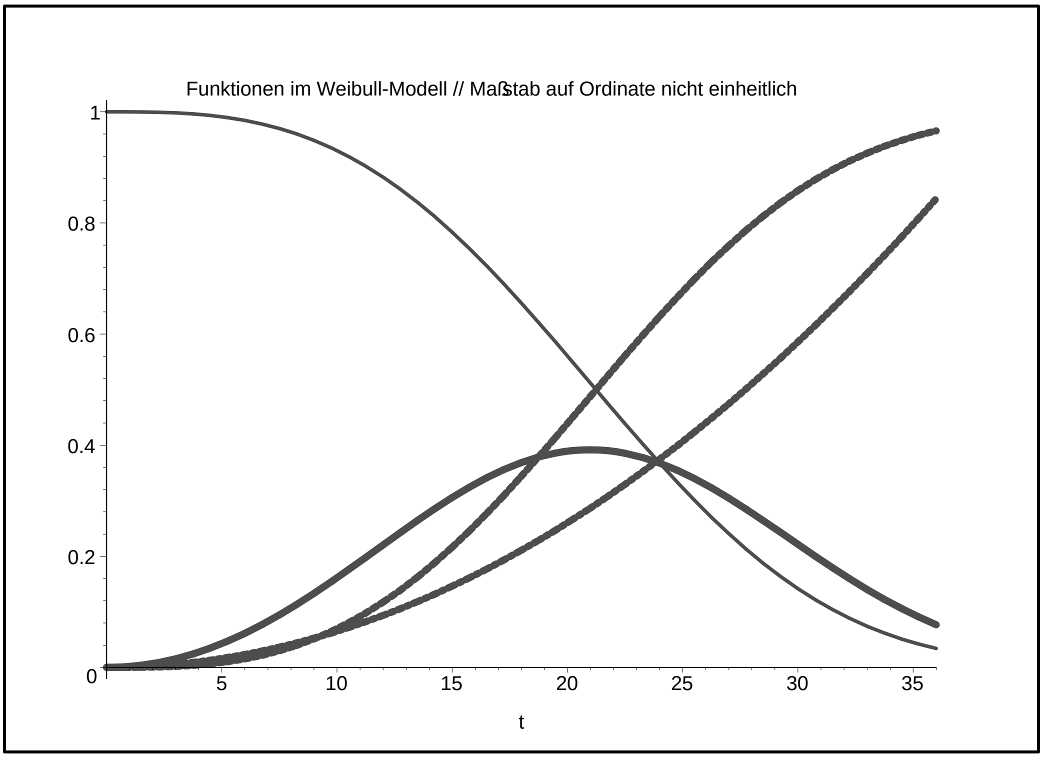
$$F(x) = 1 - \exp(-\int_0^x \lambda(u) du) \tag{1.4.13}$$

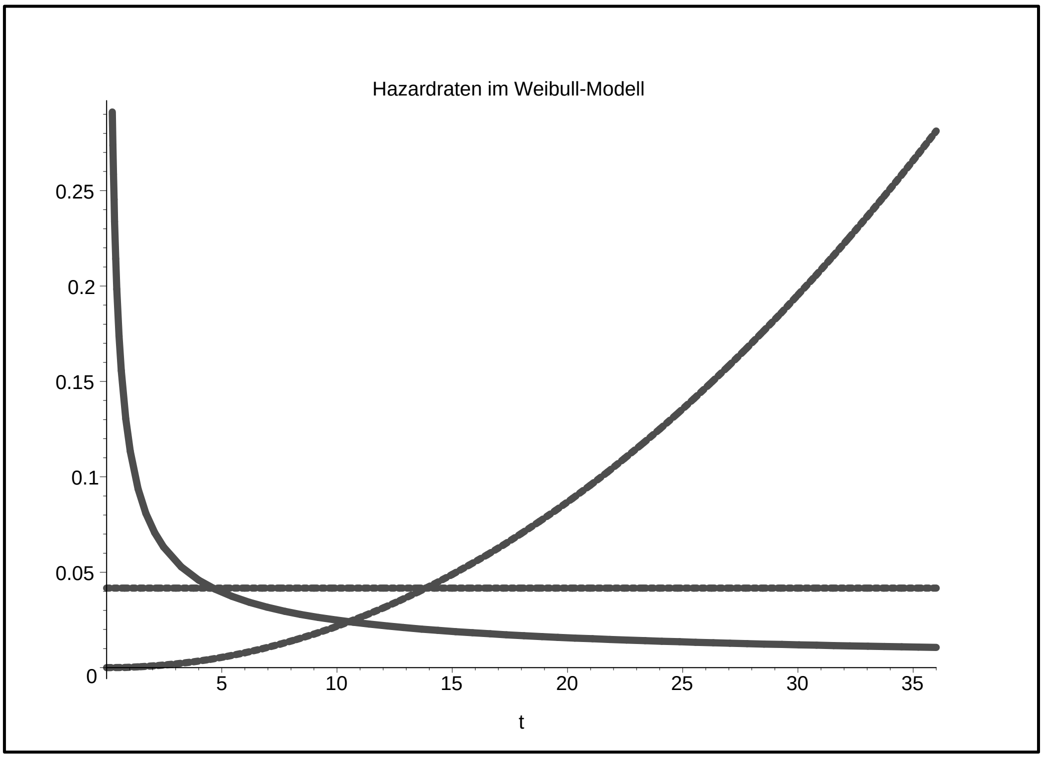
und

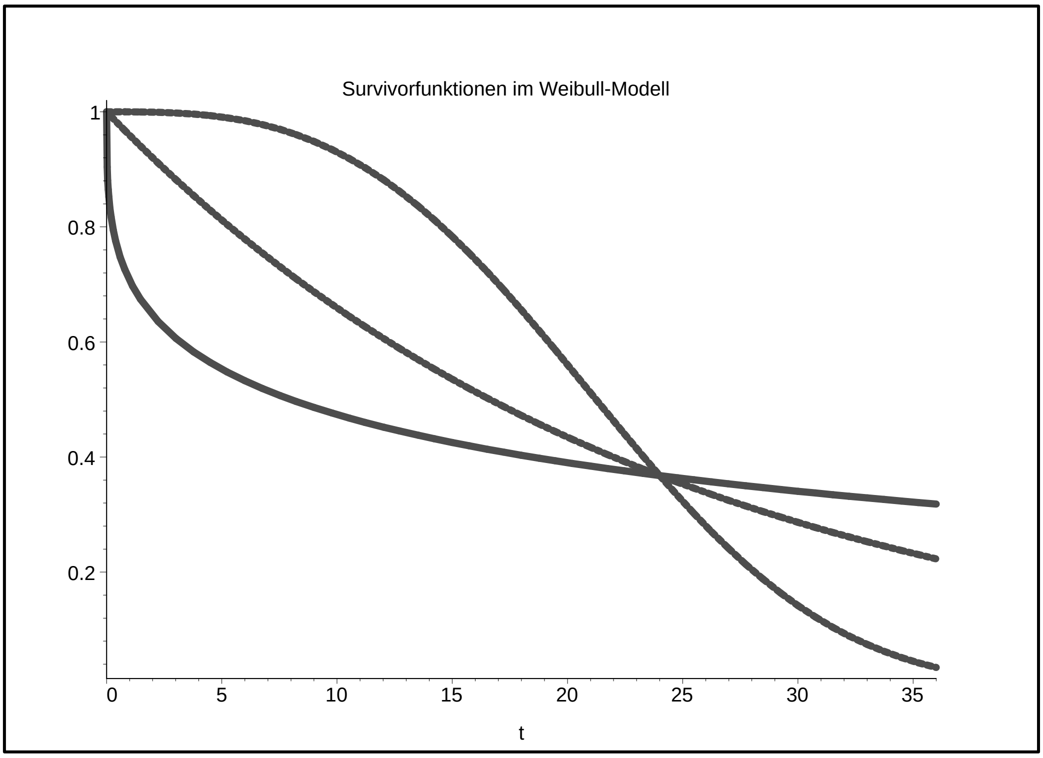
$$f(x) = \lambda(x) \cdot S(x) \tag{1.4.14}$$

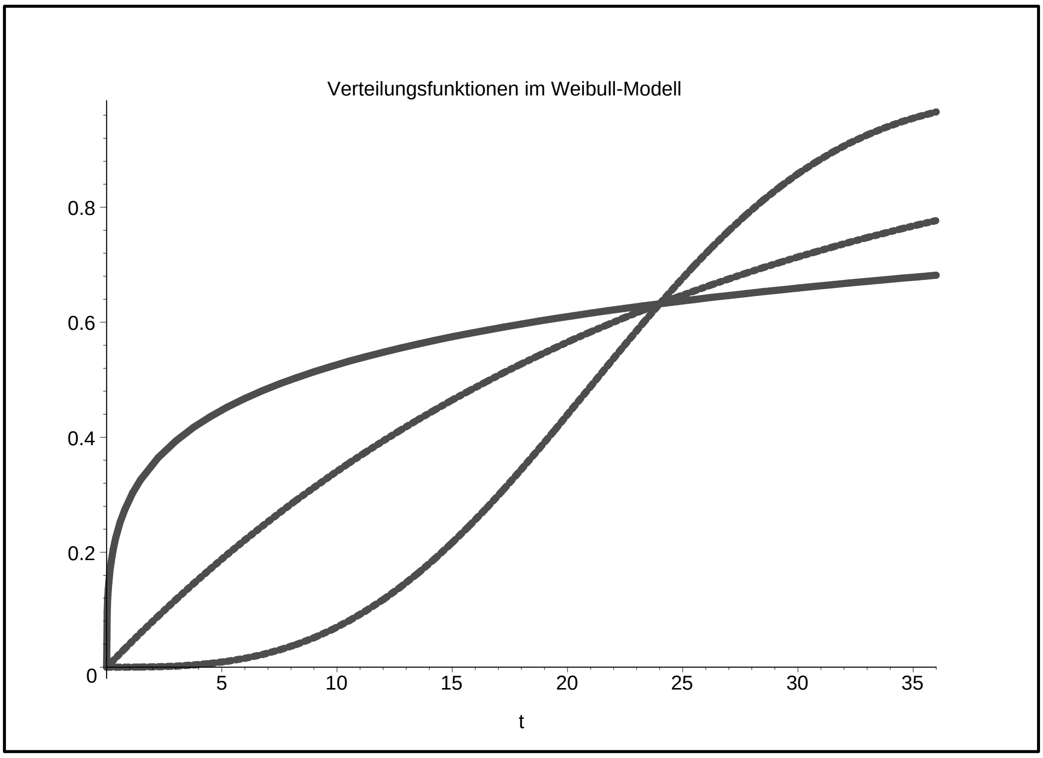


Dichtefunktionen im Weibull-Modell
0.08
0.06
0.04
0.02
0
5
10
15
20
25
30
35
t


Funktionen im Weibull-Modell // Maßstab auf Ordinate nicht einheitlich
1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
t


Hazardraten im Weibull-Modell
0.25
0.2
0.15
0.1
0.05
0
5
10
15
20
25
30
35
t


Survivorfunktionen im Weibull-Modell
1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
t


Verteilungsfunktionen im Weibull-Modell
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
t

1.4.5 Unabhängigkeit von Zufallsvariablen

Def 1.35 Zwei Zufallsvariablen X und Y mit den Verteilungsfunktionen F_X und F_Y heißen *stochastisch unabhängig*, falls für alle x und y gilt

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\}) \cdot P(\{Y \leq y\}) = F_X(x) \cdot F_Y(y)\,,$$

widrigenfalls stochastisch abhängig.

1.5 Erwartungswert und Varianz

Literatur: z.B. Fahrmeir et al., 2004, Kap 5.2, 6.2

Ziel: Charakterisiere Verteilungen von Zufallsvariablen durch Kenngrößen, insbesondere

i) „durchschnittlicher Wert“ $\longrightarrow$ Erwartungswert (Lage), z.B.

   - „mittleres“ Einkommen
   - „durchschnittliche“ Körpergröße
   - fairer Preis eines Spiels

ii) Streuung (Dispersion)

   z.B. wie stark schwankt das Einkommen, die Körpergröße etc.

1.5.1 Diskrete Zufallsvariablen

Def 1.34 Gegeben sei eine diskrete Zufallsvariable X mit der Wahrscheinlichkeitsverteilung P.
Die Menge

$$\mathcal{X} := \{x \in \mathbb{R} | P(\{x\}) > 0\}$$

heißt *Träger von X*.

Def 1.35 Gegeben sei eine diskrete Zufallsvariable X mit Träger $\mathcal{X}$. Dann heißt

$$\mathbb{E}X := \mathbb{E}(X) := \sum_{x \in \mathcal{X}} x \cdot P(X = x) \tag{1.5.1}$$

Erwartungswert von X,

$$VarX := Var(X) := \mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x \in \mathcal{X}} (x - \mathbb{E}(X))^2 \cdot P(X = x) \tag{1.5.2}$$

Varianz von X und

$$\sigma_X := \sqrt{VarX}$$

Standardabweichung von X.

Anmerkungen:

a) Die Varianz gibt die mittlere quadratische Abweichung vom Erwartungswert an. Durch das Quadrieren werden Abweichungen nach unten (negative Werte) auch positiv gezählt.

b) Damit Erwartungswert und Varianz sinnvoll interpretiert werden können, muss eine metrische Skala zugrundeliegen. Dies sei im Folgenden bei der Verwendung des Begriffs Zufalls*variable* (im Unterschied zu Zufalls*element*) stets implizit unterstellt.

c) Zur Berechnung der Varianz ist meistens der sogenannte Verschiebungssatz sehr praktisch:

$$VarX = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \qquad (1.5.3)$$

Bsp. 1.36: Fortsetzung von Bsp. 1.26 und 1.28

$$\begin{aligned} P(\{X=1\}) &= 0.4 \\ P(\{X=2\}) &= 0.3 \\ P(\{X=3\}) &= 0.2 \\ P(\{X=4\}) &= 0.1 \end{aligned}$$

Berechne Erwartungswert und Varianz von X !

Bem. 1.37 (Zur Interpretation)

a) Man kann zeigen ($\longrightarrow$ Gesetz der großen Zahl, Kap. 1.7): $\mathbb{E}X$ ist

b) Im Kontext von Beispiel 1.25:
Grundgesamtheit $\widetilde{\Omega}$, Merkmal $\widetilde{X}$
X_i Auswertung von $\widetilde{X}$ an der i-ten durch reine Zufallsauswahl gewonnenen Einheit ω_i
Jetzt:
Sei $\tilde{x}_1, \widetilde{x}_2, \ldots, \tilde{x}_N$ die Urliste von $\widetilde{X}$;
$\mu := \bar{\tilde{x}}$ das arithmetische Mittel und
$\sigma^2 := \tilde{s}^2_{\tilde{x}}$ die empirische Varianz, dann folgt aus (1.4.2) für jedes i:

(In induktiver Statistik üblich: unbekannte Kenngrößen der Grundgesamtheit (Parameter) mit griechischen Buchstaben zu bezeichnen)

1.5.2 Stetige Zufallsvariablen

Def 1.38 Sei X eine stetige Zufallsvariable mit Dichte $f(x)$. Dann heißt

$$\mathbb{E}X := \mathbb{E}(X) := \int_{-\infty}^{\infty} x \cdot f(x)\,dx \tag{1.5.4}$$

Erwartungswert von X,

$$VarX := Var(X) := \mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2 = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 \cdot f(x)\,dx \tag{1.5.5}$$

Varianz von X und

$$\sigma_X := \sqrt{VarX}$$

Standardabweichung von X.

Anmerkungen:

a) Der Verschiebungssatz (vgl. (1.5.3)) gilt nach wie vor.

b) Es gibt Verteilungen, bei denen der Erwartungswert und damit auch die Varianz nicht existiert.

c) Allgemein bezeichnet man $\mathbb{E}\,X^k$ als *k-tes Moment*.

1.5.3 Allgemeine Rechenregeln für Erwartungswert und Varianz

Satz 1.39 Seien X und Y diskrete oder stetige Zufallsvariablen (mit existierendem Erwartungswerten und Varianzen). Dann gilt:

a)
$$\mathbb{E}(aX + bY) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) \tag{1.5.6}$$
$$\text{Insb. } \mathbb{E}(a) = a$$
$$\text{und } \mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
$$\text{und } \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

b)
$$Var(aX + b) = a^2 \cdot VarX \tag{1.5.7}$$

c) sind X und Y unabhängig, so gilt
$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{1.5.8}$$
$$Var(X + Y) = Var(X) + Var(Y) \tag{1.5.9}$$

Vorsicht:

- Erwartungswert immer additiv aufspaltbar, Varianz nur bei Unabhängigkeit!

- i.A.: $\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$; z.B. $\mathbb{E}\left(\frac{1}{X}\right) \neq \frac{1}{\mathbb{E}(X)}$ und $\mathbb{E}(X^2) \neq (\mathbb{E}X)^2$

Def und Bem 1.40 Die Zufallsvariable

$$Z := \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} \tag{1.5.10}$$

heißt *standardisierte Zufallsvariable*; es gilt $\mathbb{E}(Z) = 0$ und $Var(Z) = 1$.

1.6 Wichtige Verteilungsmodelle

Literatur: Fahrmeir et. al., 2004, 5.3, 6.3.1

Hier nur Binomial- und Normalverteilung. Einige weitere Verteilungsmodelle direkt dort, wo sie benötigt werden. (Für weitere Modelle sei auf die Literatur verwiesen.)

1.6.1 Binomialverteilung

Def 1.41 Seien n und k natürliche Zahlen oder 0.

i) $$n! := n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 \tag{1.6.1}$$
$0! := 1$
$n!$ bezeichnet man als „n *Fakultät*."

ii) $\binom{n}{k} := \dfrac{n!}{k!(n-k)!}$ heißt *Binomialkoeffizient*.

Anmerkungen:

i) $n!$ gibt die Anzahl aller möglichen Anordnungen von n verschiedenen Elementen an.
$\binom{n}{k}$ ist die Anzahl aller k-elementigen Auswahlen aus n ohne Berücksichtigung der Ziehungsreihenfolge.
z.B. beim Lotto gibt es $\binom{49}{6}$ verschiedene Ergebnisse (ohne Zusatz- und Superzahl).

ii) Bei der konkreten Berechnung von Binomialkoeffizienten empfiehlt es sich häufig, „vorher zu kürzen":

Bem 1.42 *Das Grundmodell der Binomialverteilung*

- n unabhängige Wiederholungen eines Experiments, bei dem ein Ereignis A mit Wahrscheinlichkeit π eintritt.
- Betrachtet wird die Zufallsvariable X, die die Anzahl der Versuche zählt, bei denen A eintritt.
- häufig A: „Treffer", dann X: Anzahl der „Treffer".
- Es gilt für alle $x \in \{0, 1, \ldots, n\}$

  $$P(X = x) = \binom{n}{x} \cdot \pi^x \cdot (1 - \pi)^{n-x} \tag{1.6.2}$$

  (und $P(X = x) = 0$ für alle anderen x)

- X heißt *binomialverteilt*; das Wahrscheinlichkeitsmaß aus (1.6.2) heißt *Binomialverteilung*. Abkürzung: $X \sim B(n, \pi)$
- Es gilt für binomialverteiltes X

  - $\mathbb{E}X = n \cdot \pi$ (1.6.3)
  - $VarX = n \cdot \pi \cdot (1 - \pi)$ (1.6.4)

Beispiel 1.43
Risikobereite Slalomfahrer stürzen mit Wahrscheinlichkeit 10%, vorsichtigere mit 2%.

a) Wie groß sind jeweils die Wahrscheinlichkeiten, dass von je 20 Fahrern mindestens einer stürzt?

b) Vergleichen Sie die jeweils durchschnittlich zu erwartende Anzahl von Stürzen von je 100 Rennläufern!

kleiner Exkurs: Zur Problematik der Argumentation mittels „natürlicher Häufigkeiten“, wie sie Gigerenzer erfolgreich empfohlen hat. Man würde demgemäß die Wahrscheinlichkeit π_r=0,1 kommunizieren als „von 100 Rennläufern stürzen 10 “.
Diese Interpretation läuft Gefahr, die beträchtliche Variabilität zu verschleiern. In der Tat ist hier die Wahrscheinlichkeit, dass genau 10 von 100 Läufern stürzen,

$$\begin{aligned} P(X=10) &= \binom{100}{10} \cdot 0.1^{10} \cdot 0.9^{90} = \\ &= 0.13, \end{aligned}$$

also lediglich etwa 13%.

Bem 1.44 *Zwei weitere Eigenschaften der Binomialverteilung:*

- Symmetrieeigenschaft:
  (Vertausche Rolle von Treffer und Nichttreffer)

- Summeneigenschaft:

Bem 1.45 Tabellierung der Binomialverteilung