1.4 Zufallsvariablen und ihre Verteilung

1.4.1 Diskrete Zufallsvariablen

- Ein Zufallsexperiment wird beschrieben durch einen Grundraum Ω und eine Wahrscheinlichkeit P auf Ω.
- Häufig interessieren nicht die Ergebnisse an sich, sondern bestimmte abgeleitete Eigenschaften/Konsequenzen.

Bsp. 1.41. [Würfelwurf mit fairem Würfel]

Bem. 1.42.

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

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

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

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

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

- P_X heißt *Wahrscheinlichkeitsverteilung* von X.
- X (als Variable) beschreibt den Ausgang eines Zufallsexperiments *vor der Durchführung* (Auszahlungs*regel* beim Würfelspiel: wenn 3 dann 10 Euro, wenn . . . , dann . . .).

- x (als Realisation) gibt den Wert der Variablen nach Durchführung des Zufallsexperiments an (daher „Realisation“, konkreter Auszahlungs$betrag$).
- Weiteres Beispiel:
  * X Größe der nächsten eintretenden Person (als Messvorschrift)
  * x Wert, z.B. 167 cm
- In der Verwendung analog zur Unterscheidung Merkmal / Merkmalsausprägung in Statistik I (siehe später).
- Es ist häufig üblich, bei P_X den Index wegzulassen, also $P(\{x\})$ statt $P_X(\{x\})$ zu schreiben.
- Ist $\Omega_X \subset \mathbb{R}$, so bezeichnet man das Zufallselement X als $Zufallsvariable$. In der Literatur wird der Begriff Zufalls$element$ relativ selten verwendet, gerade aber in den Sozialwissenschaften sind oft nicht reelle Zahlen im Sinne einer metrischen Skala gegeben: Zufallselemente umfassen auch nominal- und ordinalskalierte Merkmale.

Definition 1.43.

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.

Bem. 1.44.

- Die $Wahrscheinlichkeitsfunktion$ $f(x)$ einer diskreten Zufallsvariable X ist für $x \in \mathbb{R}$ definiert durch

$$f(x) := \begin{cases} P(X = x_i) = p_i, & x = x_i \in \mathcal{X} = \{x_1, x_2, \ldots, x_k, \ldots\} \\ 0, & \text{sonst.} \end{cases}$$

- Man kann $\{X = x_i\}$ mit $x_i \in \mathcal{X}$ als Elementarereignisse sehen. Durch die Wahrscheinlichkeitsfunktion ist also die ganze Wahrscheinlichkeitsverteilung eindeutig bestimmt. Für beliebige Mengen A erhält man:

$$P(X_i \in A) = \sum_{x_i \in A \cap \mathcal{X}} P(X_i = x_i) = \sum_{x_i \in A \cap \mathcal{X}} f(x_i)$$

Bsp. 1.45. [Zum Rechnen mit Zufallsvariablen]

Sei X die Zufallsvariable $Anzahl\ der\ Haushaltsmitglieder$ mit der Verteilung

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

(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“.

$$\begin{aligned} P(\{X>1\}) &= P(\{X=2\})+P(\{X=3\})+P(\{X=4\}) \\ &= f(2)+f(3)+f(4) \\ &= 0.3+0.2+0.1 \\ &= 0.6 \end{aligned}$$

alternativ:

$$
\begin{aligned}
P(\{X > 1\}) &= 1 - P(\{X \leq 1\}) \\
&= 1 - P(\{X = 1\}) \\
&= 0.6
\end{aligned}
$$

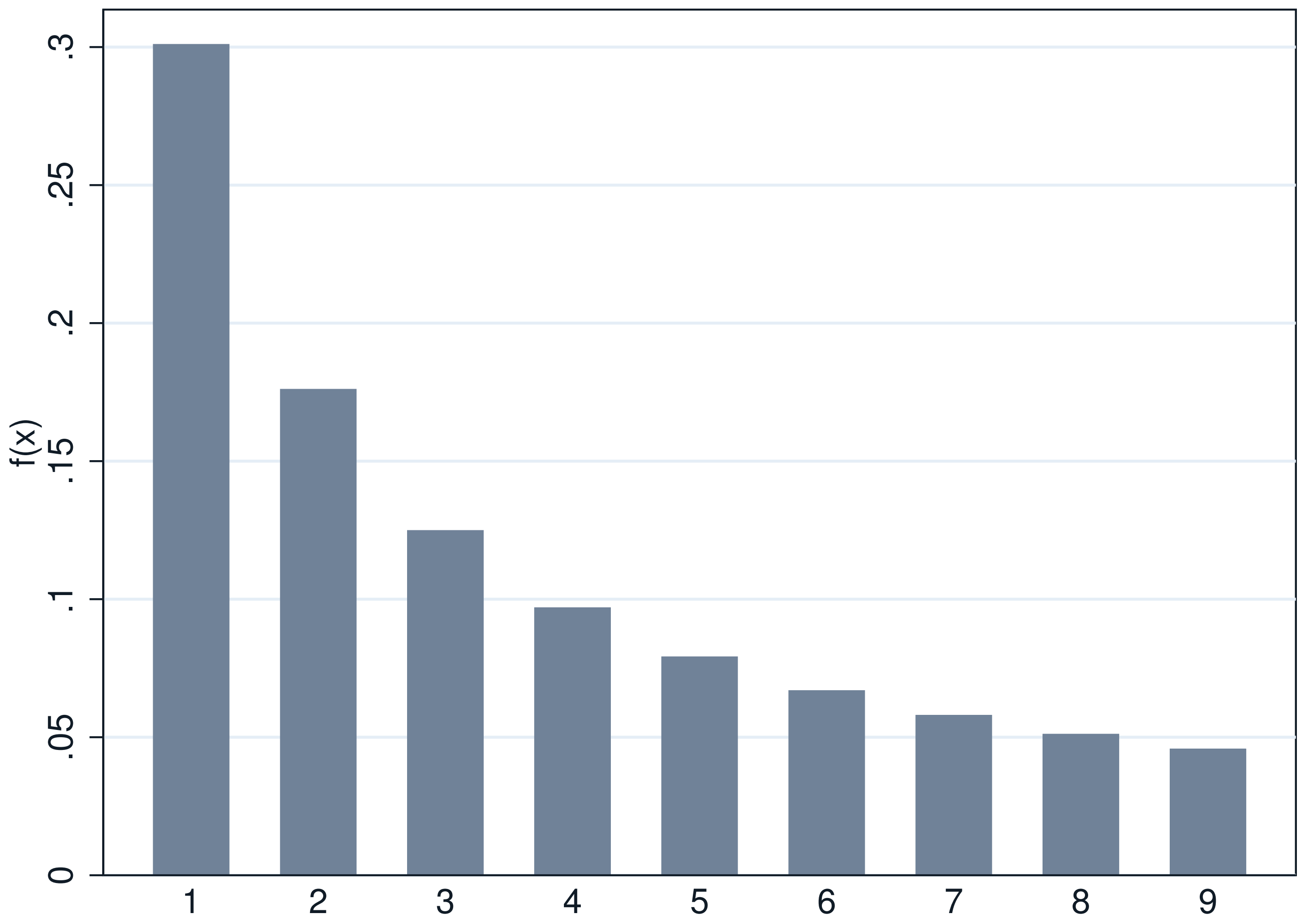
$$
\begin{aligned}
P(\{X = 2\} \cup \{X = 4\}) &\overset{\text{disjunkt}}{=} P(\{X = 2\}) + P(\{X = 4\}) \\
&= f(2) + f(4) \\
&= 0.3 + 0.1 \\
&= 0.4
\end{aligned}
$$

Bsp. 1.46. [Benfords Gesetz]

Simon Newcomb (1835–1909) und später Frank Benford (1883–1948) machten die (zunächst) verblüffende Entdeckung, dass die Anfangsziffern 1–9 von ganzen Zahlen in vielen Fällen nicht gleich häufig vorkommen. Am häufigsten ist die Anfangsziffer 1, am zweithäufigsten die Anfangsziffer 2 usw.

Beispiele sind

- die Häufigkeit der Anfangsziffern von Zahlen in Zeitungsartikeln
- die Häufigkeit der Anfangsziffern von in Steuererklärungen angebenen Beträgen
- die Häufigkeit der ersten Ziffer der Dateigröße von gespeicherten Dateien.
- in regionalen, nicht bevölkerungsproportional organisierten Stimmkreisen die Anzahl der abgegebenen Stimmen bei einer Wahl


f(x)
0
.05
.1
.15
.2
.25
.3
1
2
3
4
5
6
7
8
9

Benford postulierte für die Zufallsvariable

$$X = \text{„Anfangsziffer von Zahlen“}$$

die Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \log_{10}\left(\dfrac{x+1}{x}\right), & x = 1, \dots, 9 \\ 0, & \text{sonst} \end{cases}$$

Bem. 1.47. [Induktive Brücke II] :

Gegeben sei eine Grundgesamtheit $\mathcal{G}$ (z.B. alle Wähler(innen)). Wir betrachten eine reine Zufallsauswahl mit Ergebnisraum

$$\mathcal{S} = \mathcal{G} \times \mathcal{G} \times \ldots \times \mathcal{G}$$

mit Ergebnissen s $=$ (s$_1$, s$_2$, . . . , $_n$), wobei s$_i$ der beim i-ten Zug gezogenen Einheit, also z.B. dem i-ten gezogenen Wähler entspricht.

Betrachtet man nun ein Merkmal

$$X : \mathcal{G} \longrightarrow \{\mathsf{a}_1, \ldots \mathsf{a}_\mathsf{k}\}$$

und die Ereignisse $A_{ij} =$„i-te gezogene Person hat Merkmalsausprägung a_j“.
Stehen z.B. a$_1$, a$_2$, . . . für bestimmte Parteien, so gibt $X : \mathcal{G} \longrightarrow$ {CDU/CSU, SPD, . . . } für jede(n) Wähler(in) g$\in\mathcal{G}$ ihre/seine Wahlentscheidung X(g) an. Die Ereignisse A$_{ij}$ sind nun durch Zufallselemente beschreibbar:

Sei X_i die „Auswertung des Merkmals X an der i-ten zufällig ausgewählten Person“, d.h. an s$_i$, so ist X_i ein Zufallselement

$$\begin{aligned} X_i\colon \mathcal{G} &\longrightarrow \Omega_X = \{a_1, \ldots, a_k\} \\ \omega &\longmapsto X(\omega_i) \end{aligned}$$

Das Ereignis A_{ij} lässt sich dann schreiben als

$$\{X_i = a_j\}\,.$$

Es gilt also für jedes i (z.B. Nummer des Wählenden in der Stichprobe) und j (z.B. Name der Partei)

$$P_{X_i}(\{a_j\}) = P(\{X_i = a_j\}) = P(A_{ij}) = f_j.$$

Die Wahrscheinlichkeitsverteilung des Zufallselements X_i (Stichprobe!) spiegelt also genau die Häufigkeitsverteilung des Merkmals 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

- independently (die einzelnen Ziehungen sind stochastisch unabhängig)
- identically distributed (jedes X_i besitzt dieselbe Wahrscheinlichkeitsverteilung)

Nach dem Durchführen des Zufallsexperiments und der Auswertung von 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 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.

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

und damit für jede potentielle Stichprobe(nrealisation) die Wahrscheinlichkeit, genau sie zu erhalten.

1.4.2 Verteilungsfunktion

Jetzt Zufallsvariablen betrachten, also reellwertige Realisationen.
Viele interessierende Ereignisse besitzen folgende Form:

$$\{X \leq a\} \quad \text{oder} \quad \{X \epsilon [a, b]\} = \{a \leq X \leq b\},$$

wobei a und b feste reelle Zahlen sind.

$P(\{X \leq a\})$ für variables a entspricht der empirischen Verteilungsfunktion. In der Tat definiert man:

Definition 1.48.

Sei X eine diskrete Zufallsvariable. Die Funktion

$$\begin{aligned} F : \mathbb{R} &\to [0; 1] \\ x &\mapsto F(x) \end{aligned}$$

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

heißt *Verteilungsfunktion* von x.

Satz 1.49.

Die Wahrscheinlichkeitsverteilung einer (diskreten) Zufallsvariablen X kann man durch die Verteilungsfunktion eineindeutig erklären.

Die Wahrscheinlichkeit anderer Ereignisse ergibt sich aus dem dritten Kolmogorovschen Axiom. Es gilt zum Beispiel

$$P(a < X \leq b) = F(b) - F(a).$$

Die Ereignisse $\{X \leq a\} = \{\omega | X(\omega) \leq a\}$, $\{a < X \leq b\}$ und $\{X > b\}$ sind disjunkt und ergeben in ihrer Vereinigung Ω. Also gilt

$$\begin{aligned} 1 &= P(\Omega) = P(X \leq a) + P(a < X \leq b) + P(X > b) \\ &\Leftrightarrow \quad 1 - P(X \leq a) - P(X > b) = P(a < X \leq b) \\ &\Leftrightarrow \quad P(X \leq b) - P(X \leq a) = P(a < X \leq b) \end{aligned}$$

Bsp. 1.50. [Fortsetzung von Bsp. 1.45]

Berechne die Verteilungsfunktion und zeichne sie.

Allgemein gilt: $F(x)$ ist eine stückweise konstante Treppenfunktion und $P(X = x)$ ist genau die Sprunghöhe der Verteilungsfunktion im Punkt x.

Bsp. 1.51. [Fortsetzung von Bsp. 1.50]

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

$P(1 < X \leq 3)$

$P(1 \leq X \leq 3)$

1.4.3 Stetige Zufallsvariablen

Im Folgenden nur die Grundidee, keine „mathematisch saubere“ Herleitung.

- Eine stetige Zufallsvariable
$$X:\ \Omega \longrightarrow \Omega_X = \mathbb{R}$$
besitzt überabzählbaren Ergebnisraum Ω_X, d.h. jeder Wert innerhalb eines Intervalls $[a, b]$ ist ein mögliches Ergebnis.

- Vorstellung: Auswertung eines $stetigen$ Merkmals $\tilde{X}$ an zufällig ausgewählter Person aus einer Grundgesamtheit. (in Erweiterung der Situation aus 1.47)

- Problem: Den einzelnen Ereignissen kann keine positive Wahrscheinlichkeit mehr zugeordnet werden: „Unendlich oft“ positive Werte aufsummieren ergibt immer eine Zahl $> 1 = P(\Omega)$

Idee zur Erläuterung: Versuche eine stetige Gleichverteilung auf dem Intervall $[0, 1]$ zu konstruieren. Dazu geben wir uns ein diskretes, gleichabständiges Gitter bestehend aus n Werten aus $[0, 1]$ vor, also die Werte

$$\mathcal{X} = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}.$$

Die diskrete Gleichverteilung auf diesem Gitter ergibt die Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \frac{1}{n+1}, \quad x \in \mathcal{X}.$$

$\Rightarrow$ Lässt man nun $n \to \infty$, macht also das Gitter beliebig fein, so geht die Wahrscheinlichkeit für jeden einzelnen Wertm gegen Null.

Tatsächlich gilt für jede stetige Zufallsvariable

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

d.h. die Wahrscheinlichkeitsbewertung von Elementarereignissen enthält keine Information mehr. Andererseits bleibt die Wahrscheinlichkeit

$$P(X \leq 0.5) = 0.5$$

im Wesentlichen konstant, unabhängig vom Feinheitsgrad des Gitters. Allgemeiner gilt hier (im Grenzfall) sogar

$$P(X \leq x) = x, \quad \text{für alle } x \in [0, 1]$$

$\Rightarrow$ Verteilungsfunktion betrachten, sie enthält weiterhin die Information über die Wahrscheinlichkeitsverteilung. Im Gegensatz zu diskreten Zufallsvariablen ist die Verteilungsfunktion jetzt allerdings stetig.

Bem. 1.52.

Die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen ist (nicht mehr durch die Wahrscheinlichkeitsbewertung der Elementarereignisse mittels der Wahrscheinlichkeitsfunktion, sondern nur mehr) durch die Verteilungsfunktion

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

eindeutig festgelegt.

Allgemeiner als zuvor gilt hier

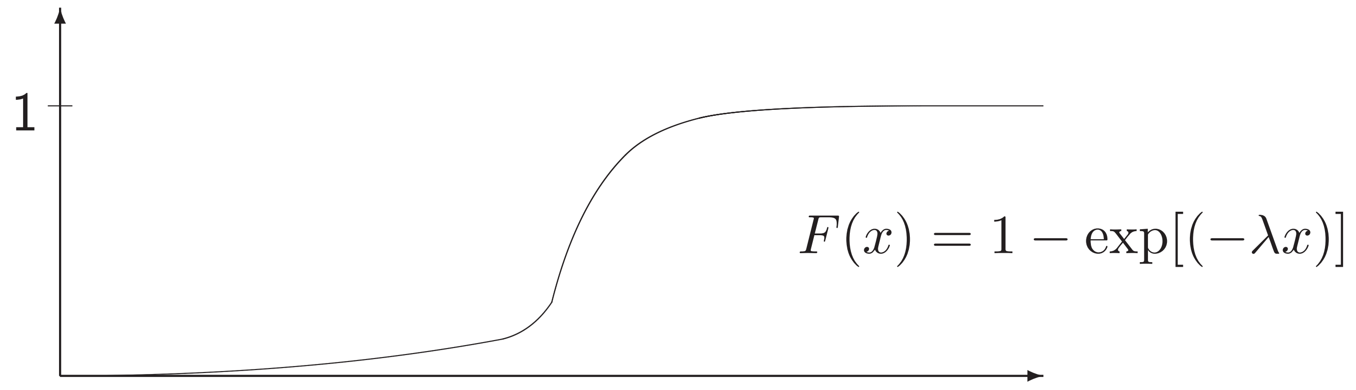
$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) \\ &= P(a < X < b) = F(b) - F(a) \end{aligned}$$

da $P(X = a) = P(X = b) = 0$.

Bem. 1.53.

- Stetige Zufallsvariablen sind in der Wahrscheinlichkeitsrechnung sehr wichtig. Später wird fast ausschließlich mit stetigen Zufallsvariablen gerechnet.
- Insbesondere ergeben sich Approximationsmöglichkeiten für diskrete durch stetige Zufallsvariablen bei größeren Stichprobenumfängen. Damit lassen sich zahlreiche Berechnungen vereinfachen (auch wenn die stetige Formulierung zunächst komplizierter wirkt).

Typische Verteilungsfunktion: (z.B. zur Beschreibung der Dauer von Arbeitslosigkeit)



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 Steigung

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

enthält also wesentliche Information über P. Dies führt zu folgender Definition:

Definition 1.54.

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

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

Dichte der Zufallsvariablen X.

Bem. 1.55.

Man kann zeigen, dass man die Definition allgemein auch dort anwenden kann, wenn die Funktion „nur stückweise differenzierbar“ ist.

Umgekehrt erhält man aus der Dichte die Verteilungsfunktion durch Integration:

Satz 1.56.

In der Situation der obigen Definition gilt

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

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

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

Bem. 1.57.

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.

Alternative Definition stetiger Zufallsvariblen: Eine Zufallsvariable X heißt stetig, wenn es eine Funktion $f(x) \geq 0$ gibt, so dass für jedes Intervall $[a, b]$

$$P(a \leq X \leq b) = \int_a^b f(x)dx \quad (\widehat{=} \text{ Fläche zwischen } a \text{ und } b \text{ unter der Funktion})$$

gilt.

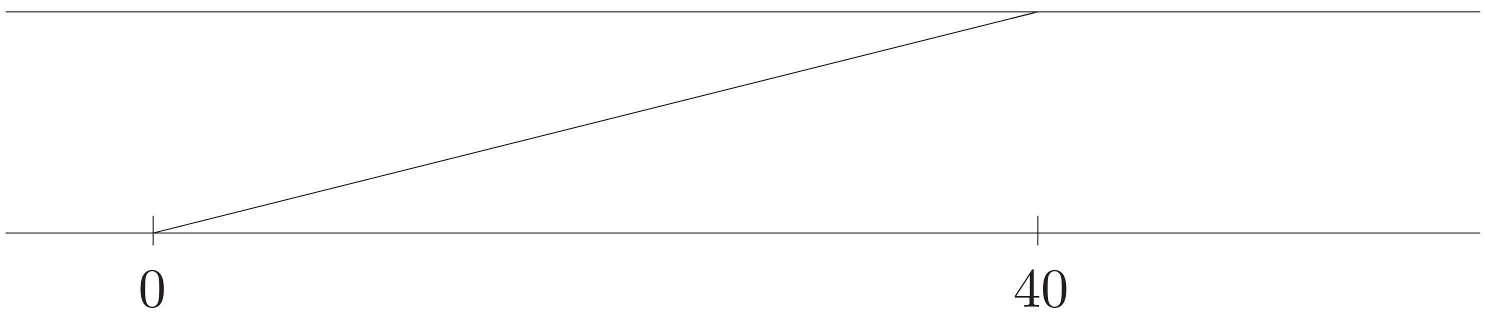
Bsp. 1.58.

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!

Skizze zu $F(x)$:



Bei der Modellbildung geht man auch häufig umgekehrt vor: Gegeben ist eine Dichte, die die Verteilungsfunktion eindeutig bestimmt.

Man erhält die Verteilungsfunktion durch

$$F(x) = P(X \leq x) = \int\limits_{-\infty}^{x} f(u)\ du$$

und das Wahrscheinlichkeitsmaß P über

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

Bsp. 1.59.

Gegeben sei die Funktion

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

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 bzw. allgemeiner Ereignisanalyse. Im Folgenden nur eine kurze Einführung, weiterführende Texte sind z.B. mit einem Schwergewicht auf sozialwissenschaftlichen Anwendungen:

- Rohwer und Pötter (2001): *Grundzüge der sozialwissenschaftlichen Statistik*, Soziologische Grundlagentexte (Teil III). Beltz Juventa
- 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“, z.B. Tod, Rückkehr aus Arbeitslosigkeit, Konkurs. Um den zeitlichen Aspekt (time) zu betonen, wird die interessierende Zufallsvariable häufig mit T statt mit X bezeichnet.

Bedingt durch die spezielle Anwendung, werden in der Lebensdaueranalyse meist nicht die Dichte oder die Verteilungsfunktion betrachtet, sondern alternative Charakterisierungen einer Wahrscheinlichkeitsverteilung.

Satz 1.60.

i) Die Verteilung einer nicht negativen, stetigen Zufallsvariable X wird eineindeutig sowohl durch die *Überlebensfunktion* (Survivorfunktion)

$$S(x) := P(X \geq x)$$

als auch durch die *Hazardrate*

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

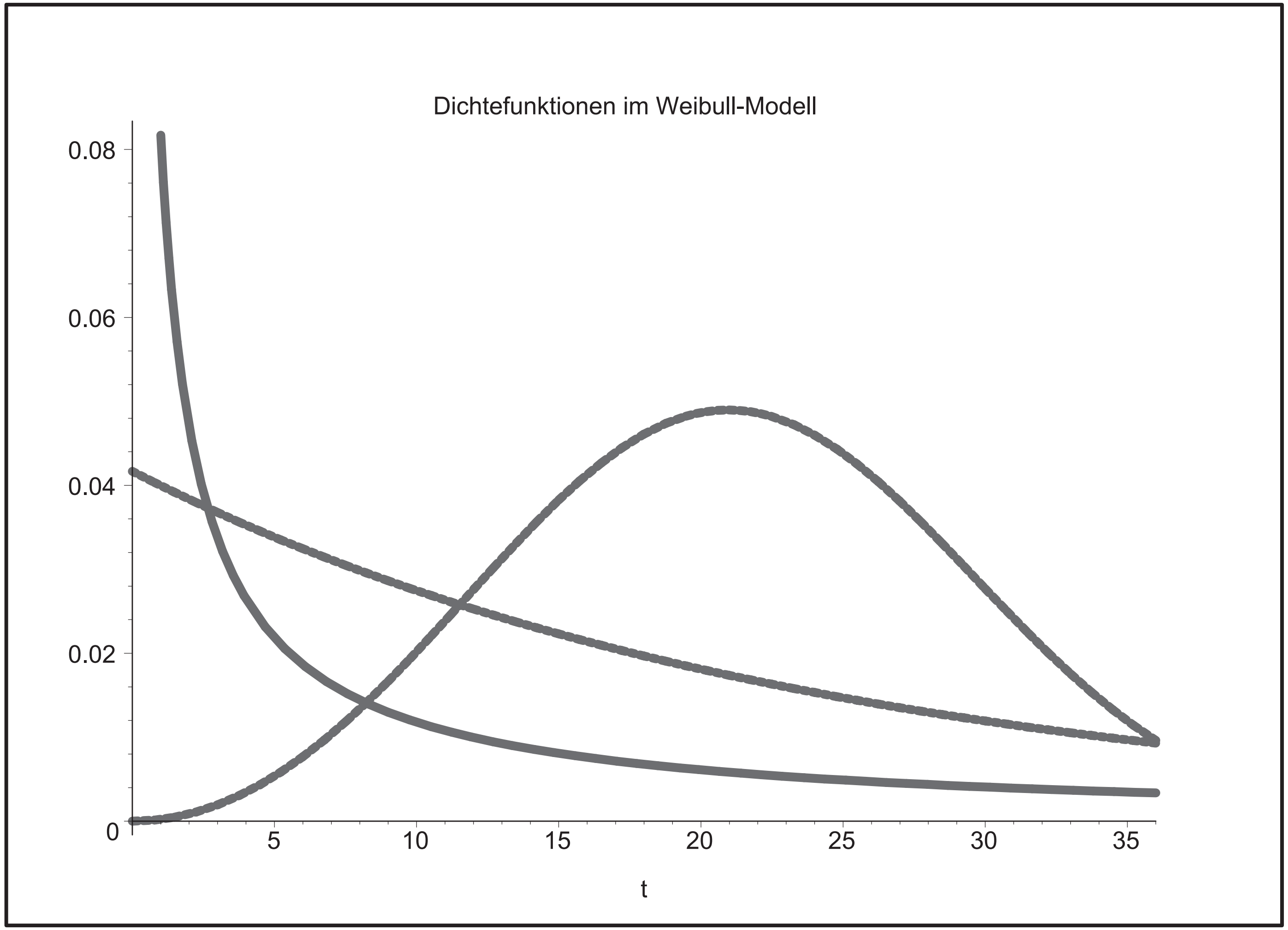
beschrieben.

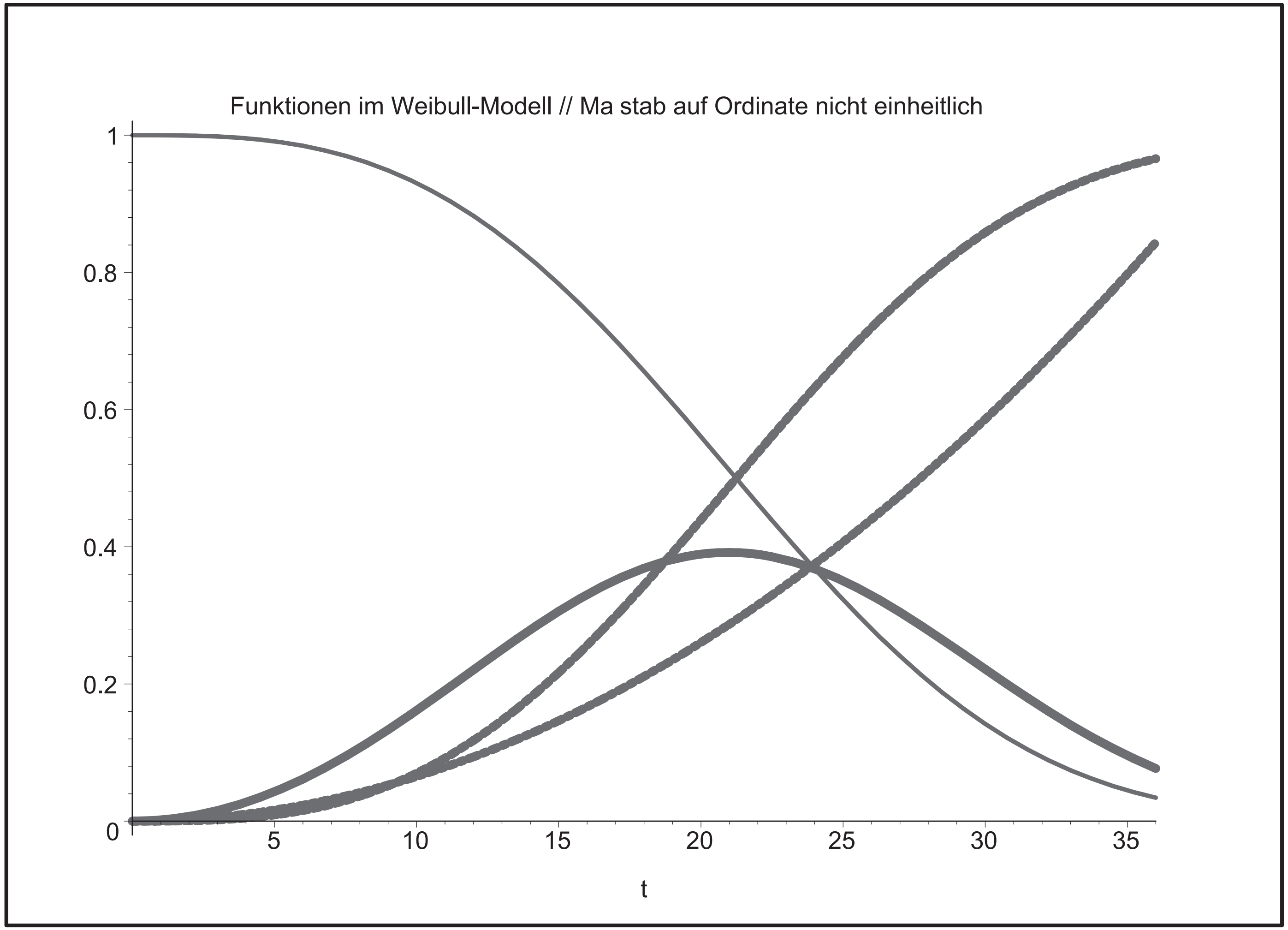
ii) Es gelten folgende Zusammenhänge

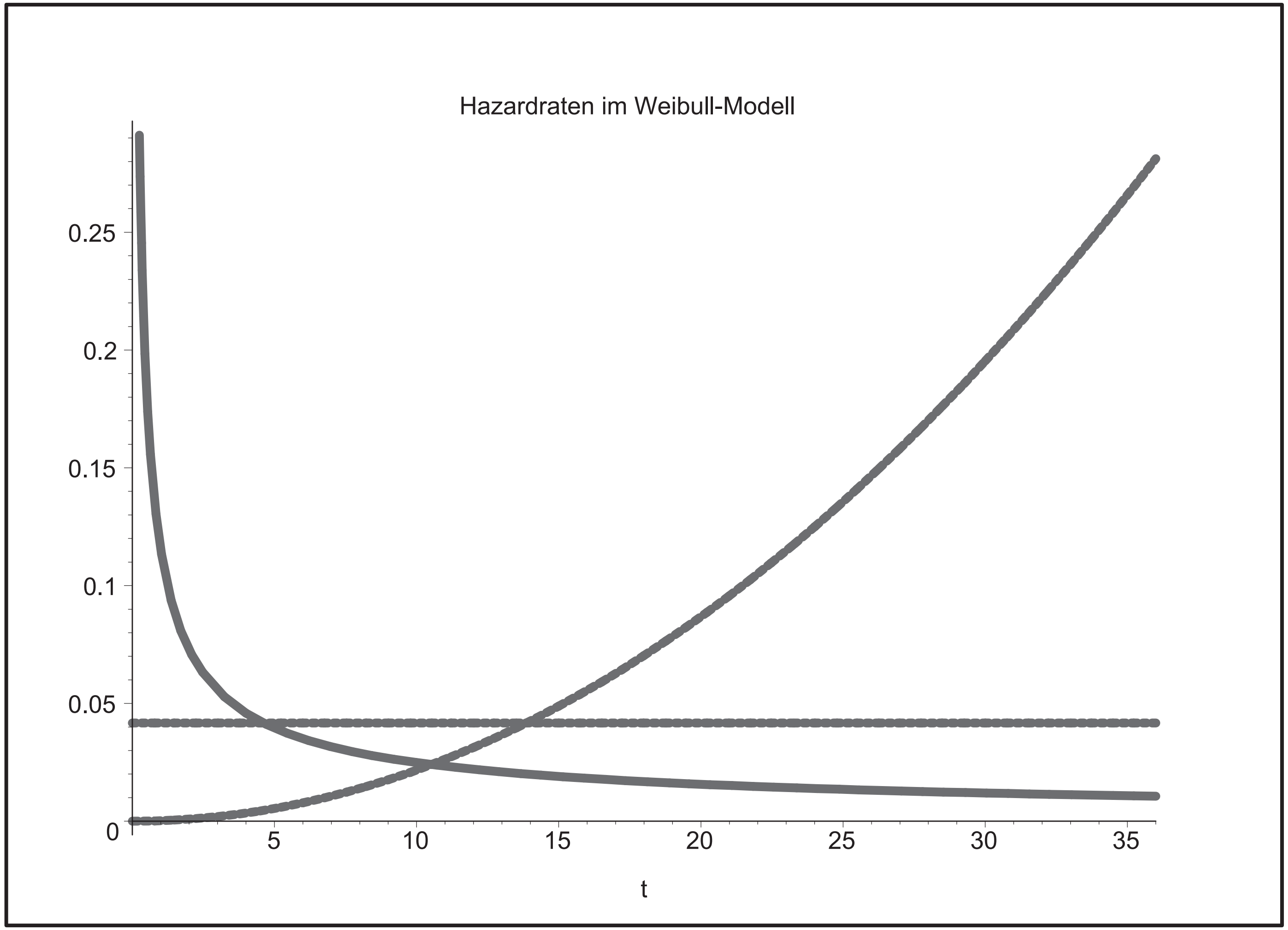
$$\begin{aligned}
S(x) &= 1 - F(x) \\
S(x) &= \exp\left(-\int_0^x \lambda(u)du\right) \\
F(x) &= 1 - \exp\left(-\int_0^x \lambda(u)du\right) \\
f(x) &= \lambda(x) \cdot S(x)
\end{aligned}$$

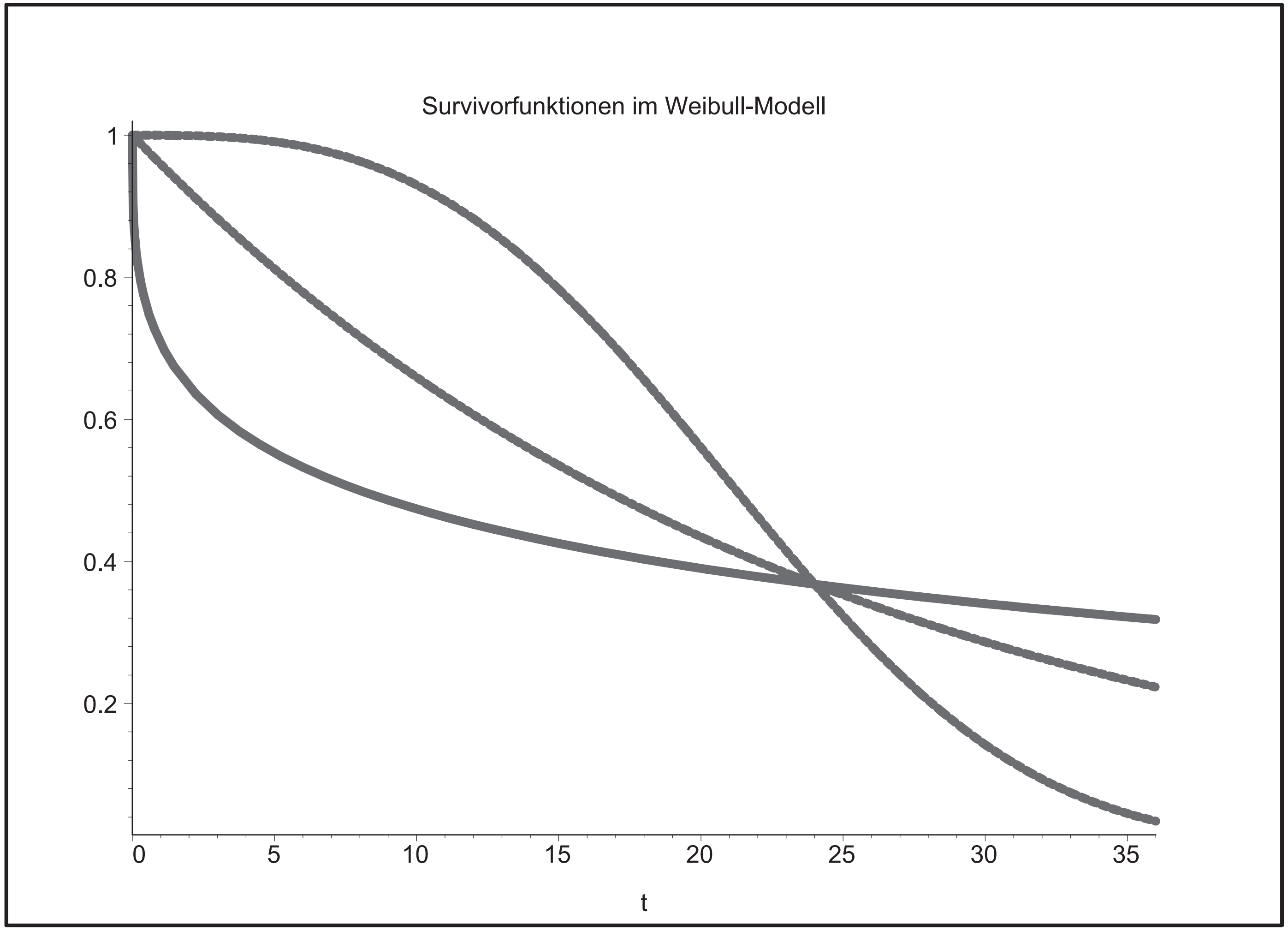
Zur Interpretation der Hazardrate:

- Beachte: $\lambda(x)$ kann Werte zwischen 0 und unendlich annehmen, ist also insbesondere keine Wahrscheinlichkeit.
- Sehr anschauliches Instrument zur Beschreibung von Lebensdauerverteilungen.

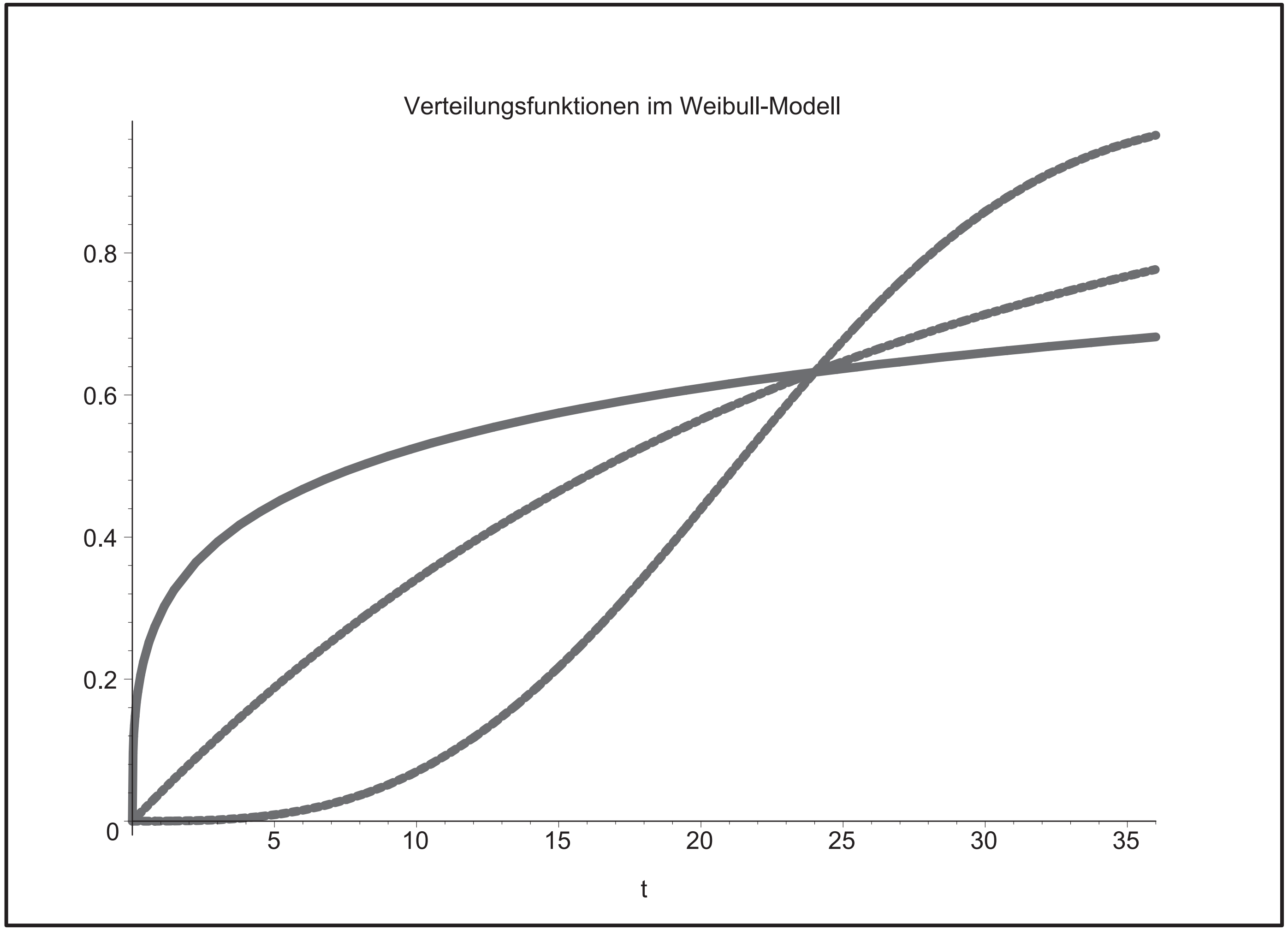

Dichtefunktionen im Weibull-Modell
0.08
0.06
0.04
0.02
0
5
10
15
20
25
30
35
t

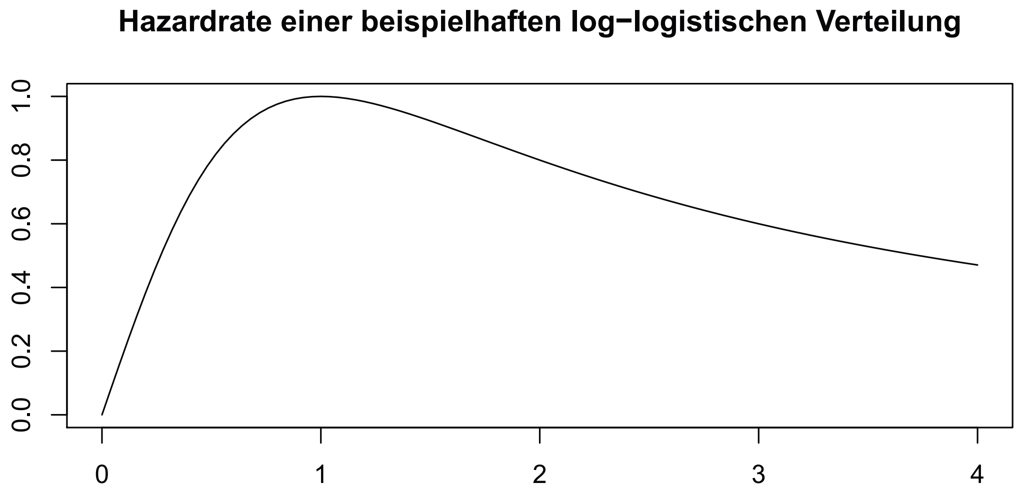

Funktionen im Weibull-Modell // Ma stab auf Ordinate nicht einheitlich
1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
t


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


Hazardrate einer beispielhaften log−logistischen Verteilung
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

1.4.5 Unabhängigkeit von Zufallsvariablen

Definition 1.61.

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

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

andernfalls heißen sie stochastisch abhängig.

Bem. 1.62.

- Entspricht der Definition der Unabhängigkeit für die Ereignisse

$$\{X \leq x\} \qquad \text{und} \qquad \{Y \leq y\}$$

  (wird hier allerdings für alle möglichen Werte von x und y gefordert!).

- Für diskrete Zufallsvariablen kann man alternativ fordern, dass

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

  für alle x und y gilt.

1.5 Erwartungswert und Varianz

Ziel: Charakterisiere Verteilungen von Zufallsvariablen (Bildbereich also reelle Zahlen, metrische Skala) durch Kenngrößen (in Analogie zu Lage- und Streuungsmaßen der deskriptiven Statistik). Insbesondere:

a) „durchschnittlicher Wert“ $\longrightarrow$ Erwartungswert, z.B.

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

b) Streuung (Dispersion), z.B. wie stark schwankt das Einkommen, die Körpergröße etc.

1.5.1 Diskrete Zufallsvariablen

Definition 1.63.

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

$$\mathrm{E}\,X := \mathrm{E}(X) := \sum_{x \in \mathcal{X}} x \cdot P(X = x)$$

Erwartungswert von X,

$$\begin{aligned} \mathrm{Var}\,X := \mathrm{Var}(X) := \mathbb{V}(X) \quad &:= \quad \mathrm{E}((X - \mathrm{E}(X))^2) \\ &= \quad \sum_{x \in \mathcal{X}} (x - \mathrm{E}(X))^2 \cdot P(X = x) \end{aligned}$$

Varianz von X und

$$\sigma_X := \sqrt{\mathrm{Var}(X)}$$

Standardabweichung von X.

Anmerkungen:

- Die Varianz gibt die mittlere quadratische Abweichung vom Erwartungswert an. Durch das Quadrieren werden Abweichungen nach unten (negative Werte) auch positiv gezählt.
- 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.
- Allgemein bezeichnet man $\mathrm{E}(X^k)$ als *k-tes Moment*.
- Zur Berechnung der Varianz ist der sogenannte *Verschiebungssatz* sehr praktisch:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\, X)^2 \tag{1.8}$$

Bsp. 1.64.

Sei X eine Zufallsvariable mit der Wahrscheinlichkeitsverteilung

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

Berechne Erwartungswert und Varianz von X !

Träger der Verteilung: $\mathcal{X} = \{1, 2, 3, 4\}$

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x \in \mathcal{X}} x \cdot P(X = x) \\
&= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + 4 \cdot P(X = 4) \\
&= 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.1 \\
&= 0.4 + 0.6 + 0.6 + 0.4 \\
&= 2
\end{aligned}
$$

Zur Berechnung der Varianz:

X	$(X - \mathbb{E}(X))$	$(X - \mathbb{E}(X))^2$	$P(X = x)$
1	-1	1	0.4
2	0	0	0.3
3	1	1	0.2
4	2	4	0.1

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{x \in \mathcal{X}} (X - \mathrm{E}(X))^2 \cdot P(X = x) \\
&= 1 \cdot 0.4 + 0 \cdot 0.3 + 1 \cdot 0.2 + 4 \cdot 0.1 \\
&= 0.4 + 0 + 0.2 + 0.4 \\
&= 1
\end{aligned}
$$

Alternative Berechnung über den Verschiebungssatz:

$$
\begin{aligned}
\mathrm{E}(X^2) &= \sum_{x \in \mathcal{X}} x^2 \cdot P(X = x) \\
&= 1 \cdot 0.4 + 2^2 \cdot 0.3 + 3^2 \cdot 0.2 + 4^2 \cdot 0.1 \\
&= 0.4 + 1.2 + 1.8 + 1.6 \\
&= 5
\end{aligned}
$$

Damit ergibt sich

$$
\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\, X)^2 = 5 - 2^2 = 1 \checkmark .
$$

Bemerkungen zur Interpretation:

- Man kann zeigen ($\longrightarrow$ Gesetz der großen Zahl, vgl. Kap. 1.7): $\mathbb{E}(X)$ ist der durchschnittswertliche Wert, wenn das durch X beschriebene Zufallsexperiment unendlich oft unabhängig wiederholt wird (Häufigkeitsinterpretation).
- Eine andere Interpretation, die auch mit dem subjektivistischen Wahrscheinlichkeitsbegriff verträglich ist, versteht $\mathrm{E}(X)$ als erwarteten Gewinn – und damit als fairen Einsatz – eines Spieles mit zufälliger Auszahlung X („Erwartungswert“).
- Man kann auch wieder einen direkten Bezug zu den Momenten einer Grundgesamtheit herstellen. Auch hier greift also die induktive Brücke:
  Betrachtet man die Grundgesamtheit Ω, das Merkmal X und versteht X_i als Auswertung von X an der i-ten durch reine Zufallsauswahl gewonnenen Einheit ω_i dann gilt: Ist $x_1, x_2, \ldots, x_N$ die Urliste von X; $\mu := \bar{x}$ das arithmetische Mittel und $\sigma^2 := \tilde{s}_x^2$ die empirische Varianz, so ist für jedes i:

$$\mathrm{E}\,X_i = \mu \quad \text{und} \quad \mathrm{Var}(X_i) = \sigma^2.$$

1.5.2 Stetige Zufallsvariablen

Definition 1.65.

Sei X eine stetige Zufallsvariable mit Dichte $f(x)$. Dann heißt (sofern wohldefiniert)

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

Erwartungswert von X,

$$\begin{aligned} \operatorname{Var} X := \operatorname{Var}(X) := \mathbb{V}(X) &:= \mathrm{E}((X - \mathrm{E}(X))^2 \\ &= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \cdot f(x)\;dx \end{aligned}$$

Varianz von X und

$$\sigma_X := \sqrt{\operatorname{Var}(X)}$$

Standardabweichung von X.

Anmerkungen:

- Der Verschiebungssatz zur Berechnung der Varianz gilt nach wie vor (vgl. 1.8).

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2$$

- Es gibt Verteilungen, bei denen der Erwartungswert und damit auch die Varianz nicht existiert (z.B. Cauchy-Verteilung, Anwendung etwa in der Finanzmathematik).
- Die eben gegebenen Bemerkungen zur Interpretation behalten ihre Gültigkeit.

1.5.3 Allgemeine Rechenregeln für Erwartungswert und Varianz

Satz 1.66.

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

a) $\mathrm{E}(aX + bY) = a \cdot \mathrm{E}(X) + b \cdot \mathrm{E}(Y)$ und insbesondere auch

$$\begin{aligned} \mathrm{E}(a) &= a, \\ \mathrm{E}(aX) &= a \cdot \mathrm{E}(X) \\ \mathrm{E}(X + Y) &= \mathrm{E}(X) + \mathrm{E}(Y) \end{aligned}$$

b) $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$.

c) Sind X und Y zusätzlich unabhängig, so gilt

$$\begin{aligned} \mathrm{E}(X \cdot Y) &= \mathbb{E}(X) \cdot \mathbb{E}(Y) \\ \mathrm{Var}(X + Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \end{aligned}$$

Bem. 1.67.

- Der Erwartungswert ist immer additiv aufspaltbar, die Varianz dagegen nur bei Unabhängigkeit!
- Die Additivität der Varianz unter Unabhängigkeit gilt nicht für die Standardabweichung σ:
$$\sqrt{\mathrm{Var}(X+Y)} \neq \sqrt{\mathrm{Var}(X)}+\sqrt{\mathrm{Var}(Y)}$$
- Man beachte explizit, dass wegen b) gilt $\mathrm{Var}(-X) = \mathrm{Var}(X)$ und damit unter Unabhängigkeit
$$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

- Im Allgemeinen gilt:

$$\mathrm{E}(g(X)) \neq g(\mathrm{E}(X))$$

also z.B.

$$\mathrm{E}\left(\frac{1}{X}\right) \neq \frac{1}{\mathrm{E}(X)}$$

und

$$\mathrm{E}(X^2) \neq (\mathrm{E}(X))^2.$$

Definition 1.68.

Die Zufallsvariable

$$Z := \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}}$$

heißt *standardisierte Zufallsvariable*. Es gilt

$$\mathrm{E}(Z) = 0 \text{ und } \mathrm{Var}(Z) = 1.$$

Bsp. 1.69. [Abschließendes Beispiel zu Erwartungswert und Varianz: Chuck-a-Luck]

- Beim Spiel Chuck-a-Luck werden drei Würfel geworfen. Der Spieler setzt vor dem Wurf auf eine der Zahlen $1, 2, 3, 4, 5, 6$. Zeigt keiner der Würfel die gesetzte Zahl, so ist der Einsatz verloren. Andernfalls erhält der Spieler (zusätzlich zu seinem Einsatz) für jeden Würfel, der die gesetzte Zahl zeigt, einen Betrag in Höhe des Einsatzes, hier als eine Einheit festgelegt.

- Wahrscheinlichkeitsfunktion des Gewinns nach einem Spiel, bei dem auf eine bestimmte Zahl (hier z.B. “6“) gesetzt wurde:

G = Gewinn	Würfelkombinationen	Anzahl	Wahrscheinlichkeit
3	666	1	1/216
2	66a, 6a6, a66 mit a$\in\{1,2,3,4,5\}$	15	15/216
1	6ab, a6b, ab6, mit a,b$\in\{1,2,3,4,5\}$	75	75/216
-1	abc mit a,b,c$\in\{1,2,3,4,5\}$	125	125/216
Summe		216	1

Diese Rechnung gilt genauso für jede andere Zahl.

- Für den Erwartungswert erhält man

$$E(G) = 3 \cdot \frac{1}{216} + 2 \cdot \frac{15}{216} + 1 \cdot \frac{75}{216} - 1 \cdot \frac{125}{216} = -\frac{17}{216} = -0.078$$

also einen erwarteten Verlust von 7.8% des Einsatzes.

- Betrachte die Zufallsvariablen:

$X_1, X_2, \ldots, X_6$	Gewinn, wenn beim ersten Wurf ein Einsatz auf $1, 2, \ldots, 6$ gesetzt wird.
$Y_1, Y_2, \ldots, Y_6$	Gewinn, wenn beim zweiten Wurf ein Einsatz auf $1, 2, \ldots, 6$ gesetzt wird.

- Mögliche Spielstrategien bei einem Kapitaleinsatz von zwei Einheiten und zugehörige Gewinne:

$2X_6$	Gewinn, wenn beim ersten Wurf ein zweifacher Einsatz auf 6 gesetzt wird (Strategie 1).
$X_1 + X_6$	Gewinn, wenn beim ersten Wurf jeweils ein Einsatz auf 1 und 6 gesetzt wird (Strategie 2).
$X_6 + Y_6$	Gewinn, wenn beim ersten und zweiten Wurf ein Einsatz auf 6 gesetzt wird (Strategie 3).

- Erwartungswerte: Aus $E(X_i) = E(Y_i) = -\frac{17}{216}$ folgt:

$$\begin{aligned} E(2X_6) &= 2E(X_6) = -\frac{34}{216} \\ E(X_1 + X_6) &= E(X_1) + E(X_6) = -\frac{34}{216} \\ E(X_6 + Y_6) &= E(X_6) + E(Y_6) = -\frac{34}{216}, \end{aligned}$$

d.h. bei den drei Strategien sind die Erwartungswerte alle gleich!

- Trotzdem gibt es deutliche Unterschiede in den drei Strategien:

Strategie	Wertebereich	$P(\{-2\})$
$2X_6$	-2,2,4,6	0.579
$X_1 + X_6$	-2,0,1,2,3	0.296
$X_6 + Y_6$	-2,0,1,2,3,4,5,6	0.335

- Varianz des Gewinns nach einem Spiel

$$\begin{aligned}
\mathrm{Var}(G) &= \left(3+\frac{17}{216}\right)^2 \cdot \frac{1}{216} + \left(2+\frac{17}{216}\right)^2 \cdot \frac{15}{216} + \left(1+\frac{17}{216}\right)^2 \cdot \frac{75}{216} + \\
&\quad + \left(-1+\frac{17}{216}\right)^2 \cdot \frac{125}{216} \\
&= 0.04388156 + 0.30007008 + 0.40402836 + 0.4911961 = \\
&= 1.2391761 \\
\sqrt{\mathrm{Var}(G)} &= 1.113183
\end{aligned}$$

- Nach den Rechenregeln für Varianzen erhält man für die Strategien 1 und 3:

$$\mathrm{Var}(2X_6) = 4\,\mathrm{Var}(X_6) = 4 \cdot 1.2391761 = 4.956704$$

und, wegen der Unabhängigkeit von X_6 und Y_6,

$$\mathrm{Var}(X_6 + Y_6) = \mathrm{Var}(X_6) + \mathrm{Var}(Y_6) = 1.2391761 + 1.2391761 = 2.4783522.$$

- Da X_1 und X_6 nicht unabhängig sind, muss hier die Varianz explizit berechnet (oder die später betrachteten Formeln für die Kovarianz verwendet) werden.

- Wahrscheinlichkeitsverteilung von $X_1 + X_6$:

x	-2	0	1	2	3
$P(X_1 + X_6 = x)$	0.29630	0.44444	0.11111	0.12037	0.02778

$$\begin{aligned}
\mathrm{Var}(X_1 + X_6) &= \left(-2 + \frac{34}{216}\right)^2 \cdot 0.29630 + \left(0 + \frac{34}{216}\right)^2 \cdot 0.44444 + \\
&\quad + \left(1 + \frac{34}{216}\right)^2 \cdot 0.11111 + \left(2 + \frac{34}{216}\right)^2 \cdot 0.12037 + \\
&\quad + \left(3 + \frac{34}{216}\right)^2 \cdot 0.02778 = \\
&= 2.003001
\end{aligned}$$

- Fazit:

  * Strategie 1, also $2X_6$, ist am riskantesten, sie hat die höchste Varianz. Hohes Verlustrisiko, in der Tat ist $P(\{-2\})$ am größten, andererseits ist hier z.B. die Chance, 6 Einheiten zu gewinnen am grössten, denn es gilt

    bei Strategie 1:
    $P(2X_6 = 6) = P(X_6 = 3) = \frac{1}{216}$
    bei Strategie 2:
    $P(X_1 + X_6 = 6) = P(X_1 = 3 \cap X_6 = 3) = P(\emptyset) = 0$
    bei Strategie 3:
    $P(X_6 + Y_6 = 6) = P(X_6 = 3 \cap Y_6 = 3) = P(X_6 = 4) \cdot P(Y_6 = 3) = (\frac{1}{216})^2$

  * Am wenigsten riskant ist Strategie 2.

  * Typische Situation bei Portfolio Optimierung (außer, dass Erwartungswert < 0):