

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)*
- 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. Erlbaum.*

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 1.58.

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

$$S(x) := P(X \geq x) = 1 - F(x)$$

als auch durch die *Hazardrate*

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

beschrieben.

ii) Es gelten folgende Zusammenhänge

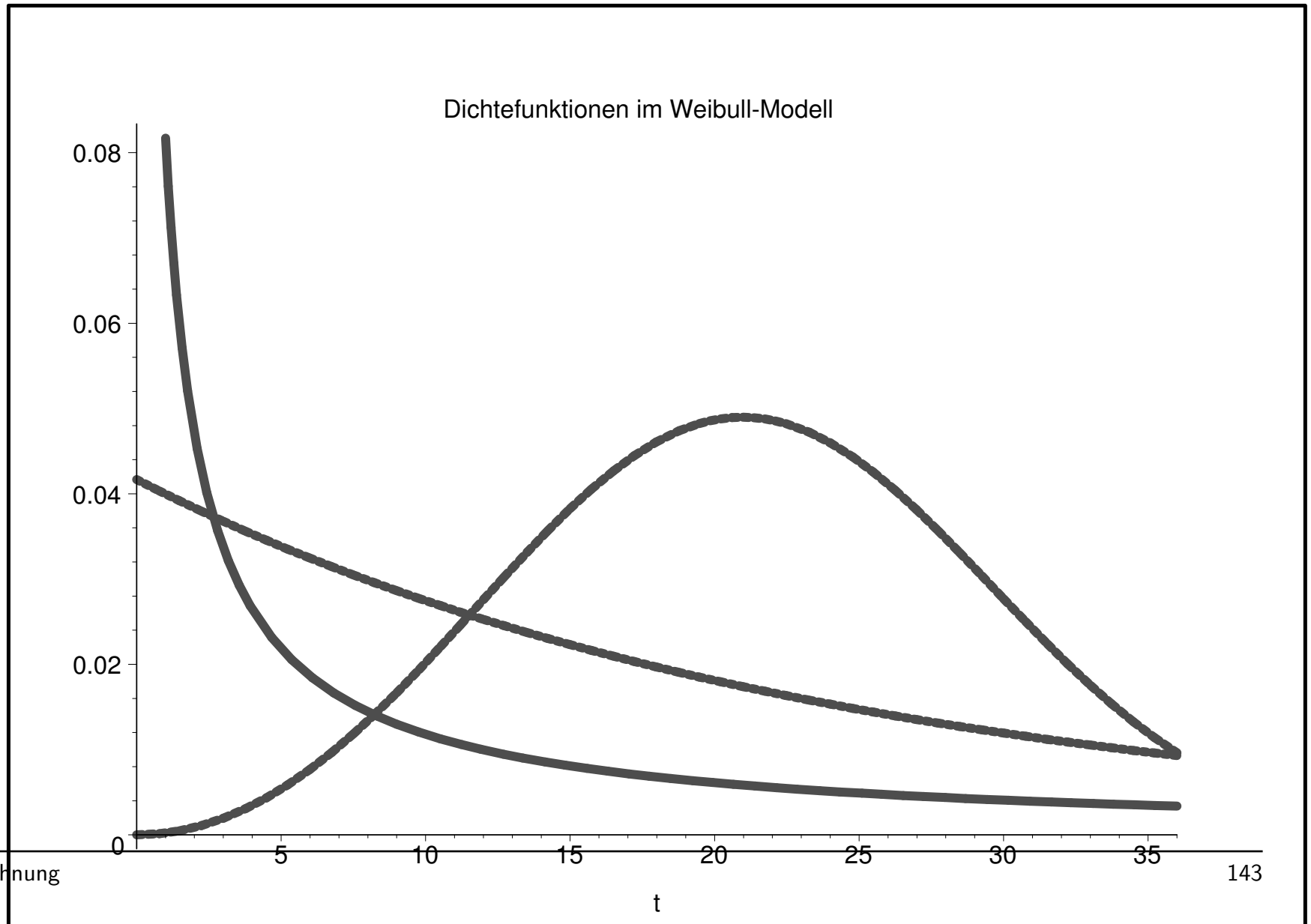
$$S(x) = \exp\left(-\int_0^x \lambda(u) du\right)$$

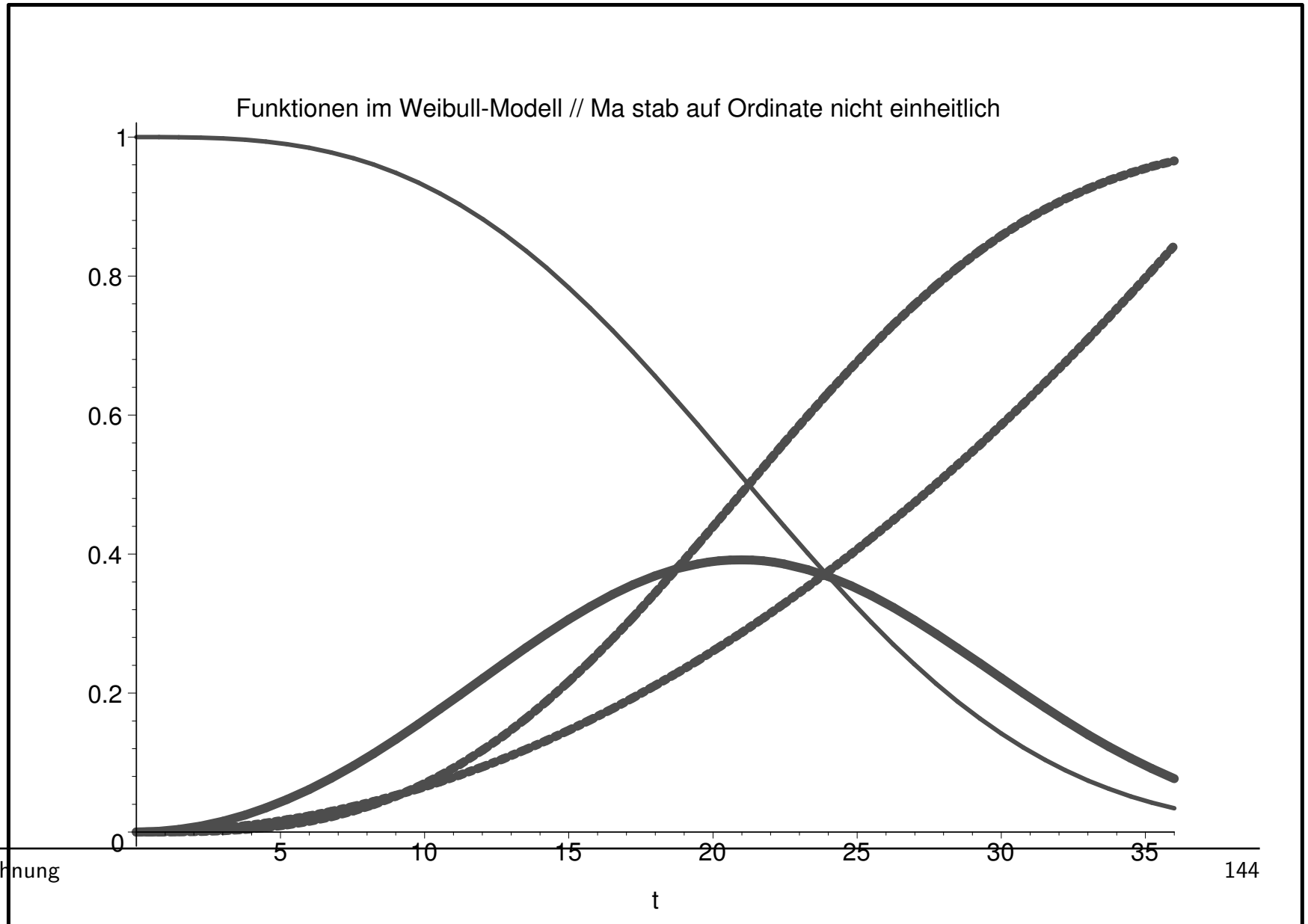
$$F(x) = 1 - \exp\left(-\int_0^x \lambda(u) du\right)$$

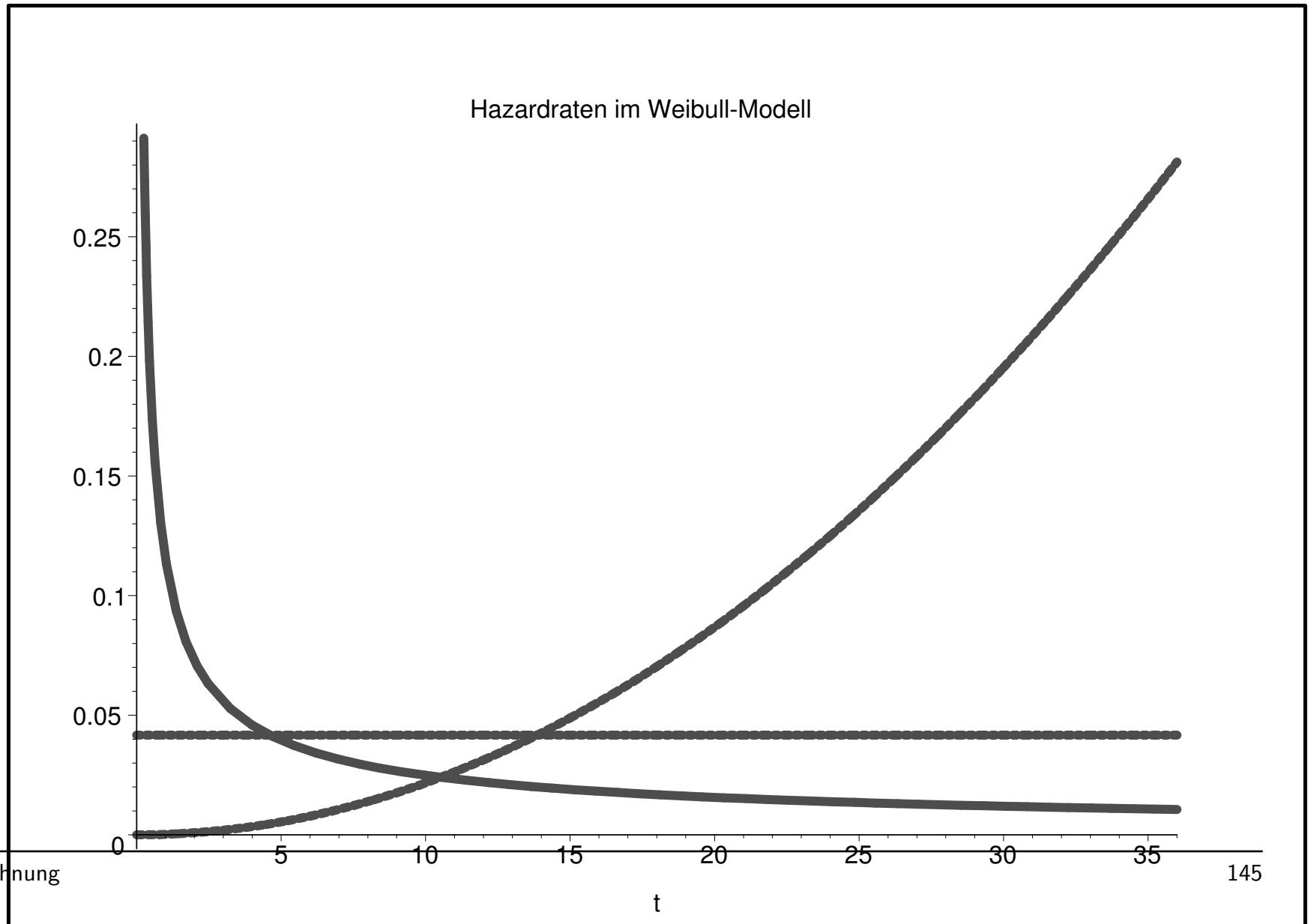
$$f(x) = \lambda(x) \cdot S(x)$$

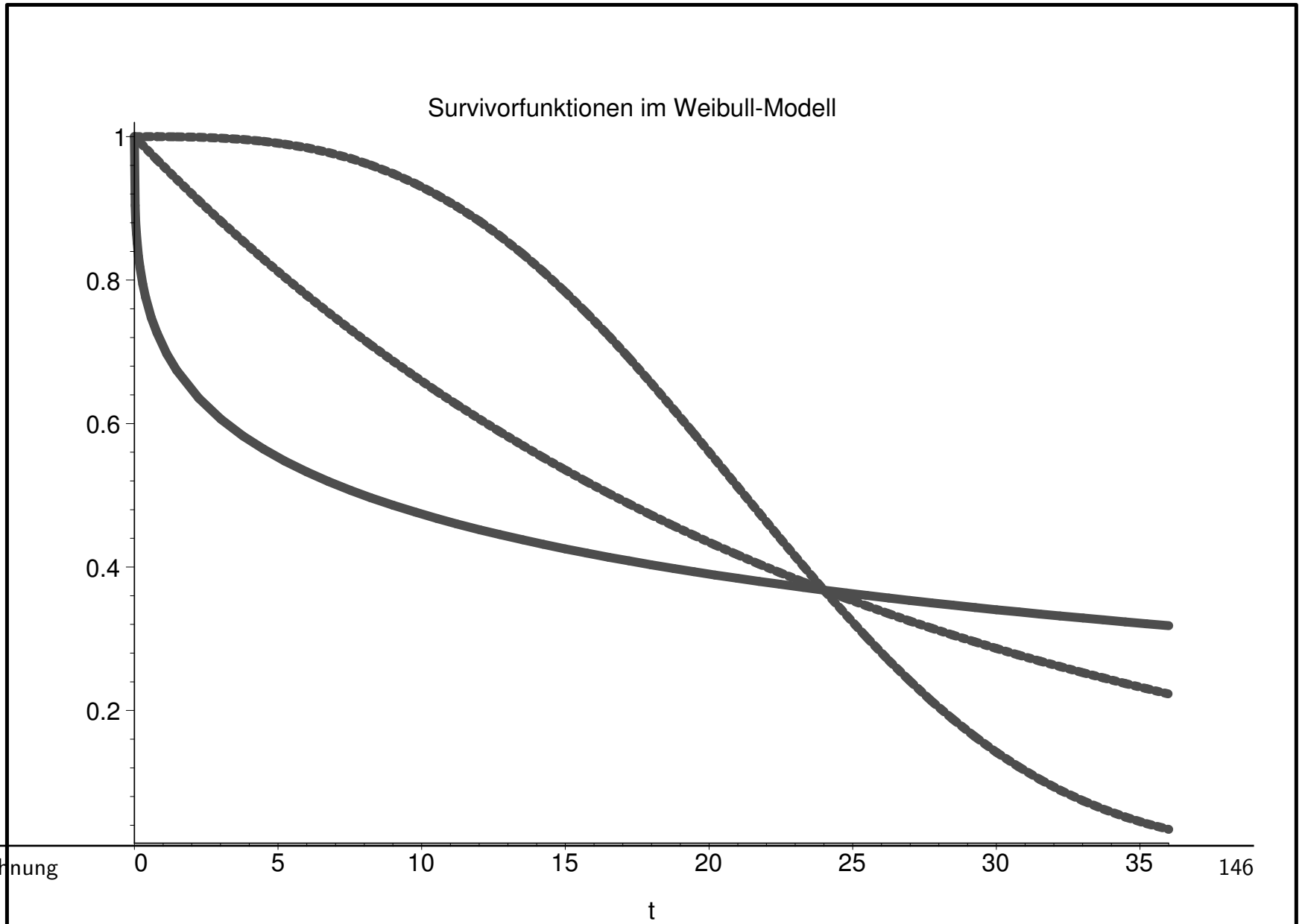
Zur Interpretation der Hazardrate:

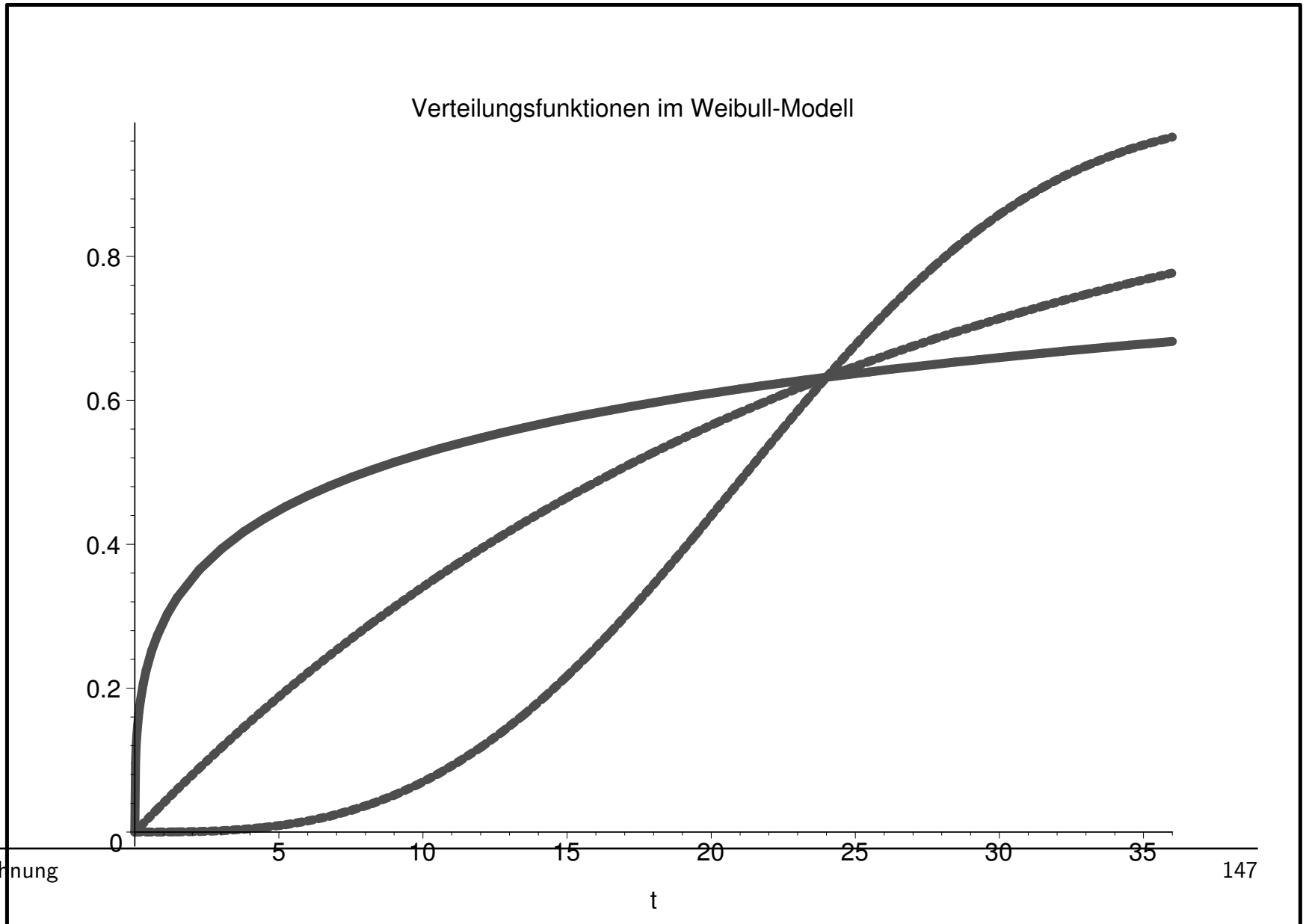
- Beachte: $\lambda(\cdot)$ ist keine Wahrscheinlichkeit, kann Werte zwischen 0 und unendlich annehmen.
- Sehr anschauliches Instrument zur Beschreibung von Lebensdauerverteilungen.

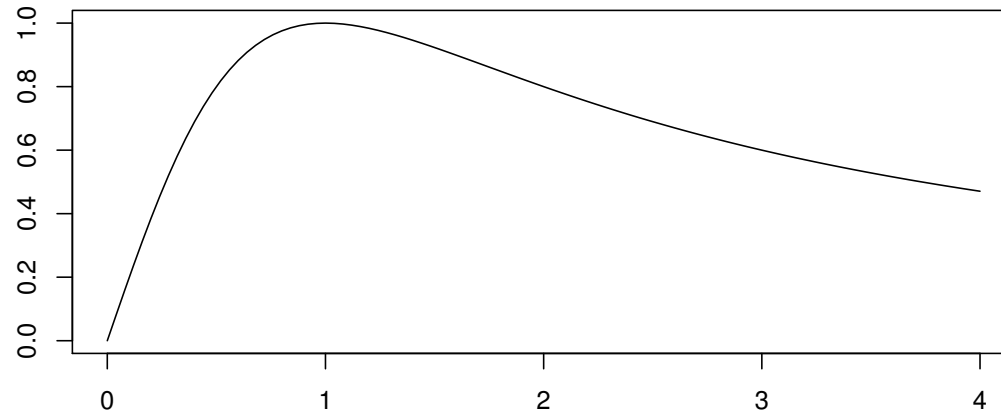










Hazardrate einer beispielhaften log-logistischen Verteilung

1.4.5 Unabhängigkeit von Zufallsvariablen

Definition 1.59.

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. 1.60.

- Entspricht der Definition der Unabhängigkeit für die Ereignisse

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

(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 1.61.

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

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

Erwartungswert von X ,

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

Varianz von X und

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

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 *Zufallsvariable* (im Unterschied zu *Zufallselement*) stets implizit unterstellt.
- Allgemein bezeichnet man $E(X^k)$ als *k-tes Moment*.
- Zur Berechnung der Varianz ist der sogenannte *Verschiebungssatz* sehr praktisch:

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

Bsp. 1.62.

Sei X eine Zufallsvariable mit der Wahrscheinlichkeitsverteilung

$$P(\{X = 1\}) = 0.4$$

$$P(\{X = 2\}) = 0.3$$

$$P(\{X = 3\}) = 0.2$$

$$P(\{X = 4\}) = 0.1$$

Berechne Erwartungswert
und Varianz von X !

Bemerkungen zur Interpretation:

- Man kann zeigen (\longrightarrow Gesetz der großen Zahl, 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 $\mathbb{E}(X)$ 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.
- Es gilt auch wieder die induktive Brücke: Betrachtet man die Grundgesamtheit $\tilde{\Omega}$, das Merkmal \tilde{X} und versteht X_i als Auswertung von \tilde{X} an der i -ten durch reine Zufallsauswahl gewonnenen Einheit ω_i dann gilt:

Ist $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N$ die Urliste von \tilde{X} ;

$\mu := \bar{\tilde{x}}$ das arithmetische Mittel und

$\sigma^2 := \tilde{s}_x^2$ die empirische Varianz, so ist für jedes i :

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

1.5.2 Stetige Zufallsvariablen

Definition 1.63.

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

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

Erwartungswert von X ,

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

Varianz von X und

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

Standardabweichung von X .

Anmerkungen:

- Der Verschiebungssatz zur Berechnung der Varianz gilt nach wie vor (vgl. 1.5.1).

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

- 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 1.64.

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

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

$$E(a) = a,$$

$$E(aX) = a \cdot E(X)$$

$$E(X + Y) = E(X) + E(Y)$$

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

Sind X und Y zusätzlich unabhängig, so gilt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Bem. 1.65.

- Der Erwartungswert ist immer additiv aufspaltbar, die Varianz dagegen nur bei Unabhängigkeit!
- Die Additivität der Varianz unter Unabhängigkeit gilt nicht für die Standardabweichung σ :

$$\sqrt{\text{Var}(X + Y)} \neq \sqrt{\text{Var}(X)} + \sqrt{\text{Var}(Y)}$$

- Man beachte explizit, dass wegen b) gilt $\text{Var}(-X) = \text{Var}(X)$ und damit unter Unabhängigkeit

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y).$$

- Im Allgemeinen gilt:

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

also z.B.

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

und

$$E(X^2) \neq (E(X))^2.$$

Definition 1.66.

Die Zufallsvariable

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

heißt *standardisierte Zufallsvariable*. Es gilt

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

Bsp. 1.67. [Abschließendes Beispiel zu Erwartungswert und Varianz: Chuck-a-Luck]

- Beim Spiel Chuck-a-Luck werden drei Würfel geworfen. Der Spieler setzt 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 (z.B. "6") gesetzt wurde:

G = Gewinn	Würfelnkombinationen	Anzahl	Wahrscheinlichkeit
3	666	1	1/216
2	66a, 6a6, a66 mit a=1,2,3,4,5	15	15/216
1	6ab, a6b, ab6, mit a,b=1,2,3,4,5	75	75/216
-1	abc mit a,b,c=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, \dots, X_6 Gewinn, wenn beim ersten Wurf ein Einsatz auf
1, 2, ..., 6 gesetzt wird.

Y_1, Y_2, \dots, Y_6 Gewinn, wenn beim zweiten Wurf ein Einsatz auf
1, 2, ..., 6 gesetzt wird.

- Mögliche Spielstrategien 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:

$$E(2X_6) = 2E(X_6) = -\frac{34}{216}$$

$$E(X_1 + X_6) = E(X_1) + E(X_6) = -\frac{34}{216}$$

$$E(X_6 + Y_6) = E(X_6) + E(Y_6) = -\frac{34}{216}$$

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

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

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

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

und, wegen der Unabhängigkeit von X_6 und Y_6 ,

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

- Da X_1 und X_6 nicht unabhängig sind, muss hier die Varianz explizit berechnet 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

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

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:

$$P(X_6 + Y_6 = 6) = P(X_6 = 3 \cap Y_6 = 3) = P(X_6 = 4) \cdot P(Y_6 = 3) = \left(\frac{1}{216}\right)^2$$

- 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.
- * Am wenigsten riskant ist Strategie 2.
- * Typische Situation bei Portfolio Optimierung (außer, dass Erwartungswert < 0)

1.6 Wichtige Verteilungsmodelle

Wir behandeln hier nur die Binomial-, die Poisson- und die Normalverteilung. Einige weitere Verteilungsmodelle werden direkt dort eingeführt, wo sie benötigt werden.

1.6.1 Binomialverteilung

Konstruktionsprinzip:

- Ein Zufallsexperiment wird n mal unabhängig durchgeführt.
- Wir interessieren uns jeweils nur, ob ein bestimmtes Ereignis A eintritt oder nicht.
- $X =$ „absolute Häufigkeit, mit der Ereignis A bei n unabhängigen Versuchen eintritt“.
- Träger von X : $\mathcal{X} = \{0, 1, 2, \dots, n\}$.

Herleitung der Wahrscheinlichkeitsfunktion:

- Bezeichne $\pi = P(A)$ die Wahrscheinlichkeit für A in *einem* Experiment.
- Das Ereignis $\{X = x\}$ tritt z.B. auf, wenn in den ersten x Versuchen A eintritt und anschließend nicht mehr. Die Wahrscheinlichkeit dafür ist

$$\begin{aligned} P(A_1 \cap \dots \cap A_x \cap \bar{A}_{x+1} \cap \dots \cap \bar{A}_n) &= \underbrace{\pi \cdot \dots \cdot \pi}_{x \text{ mal}} \cdot \underbrace{(1 - \pi) \cdot \dots \cdot (1 - \pi)}_{n-x \text{ mal}} \\ &= \pi^x (1 - \pi)^{n-x}. \end{aligned}$$

- Insgesamt gibt es $\binom{n}{x}$ Möglichkeiten für die Verteilung der x Erfolge (Auftreten von A) auf n Plätze. Damit gilt:

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

Definition 1.68.

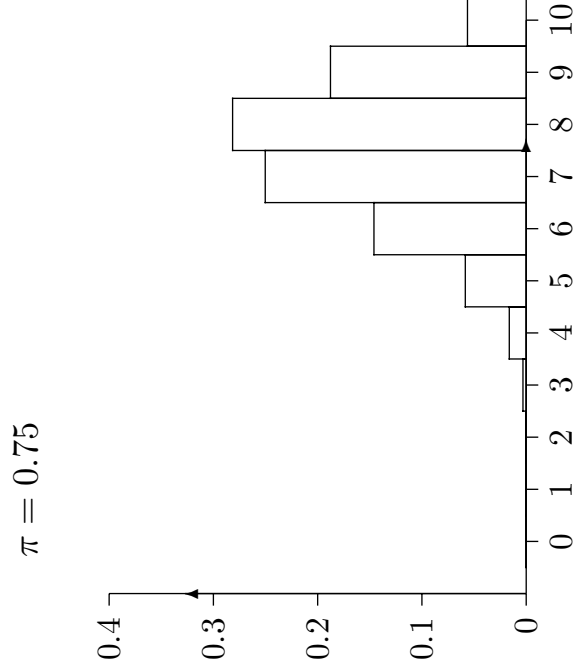
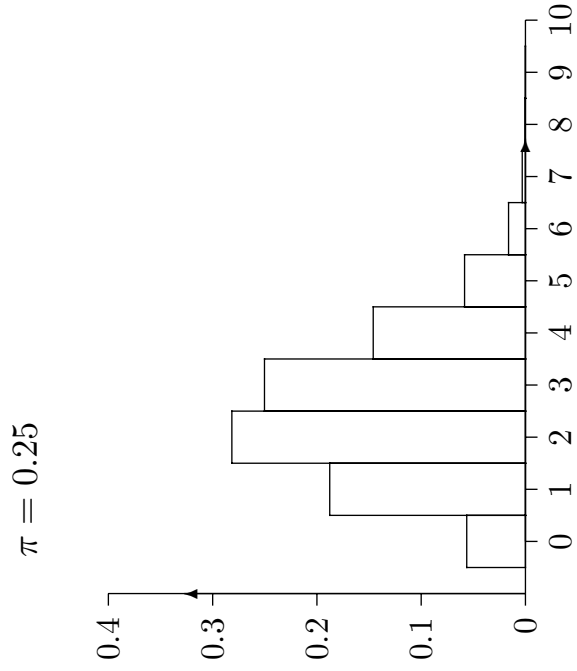
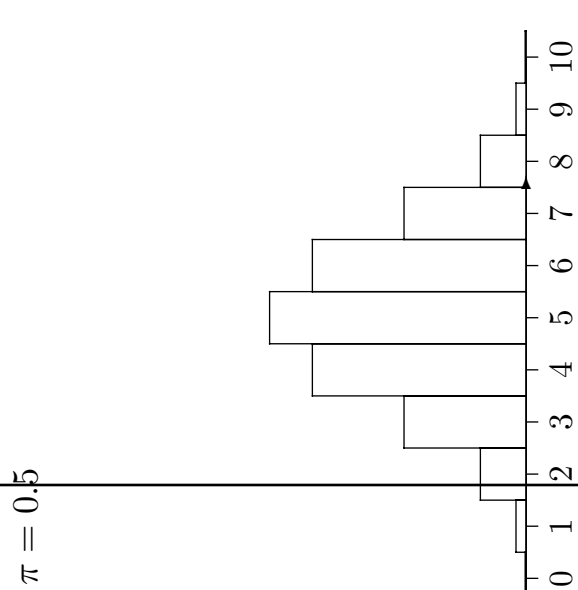
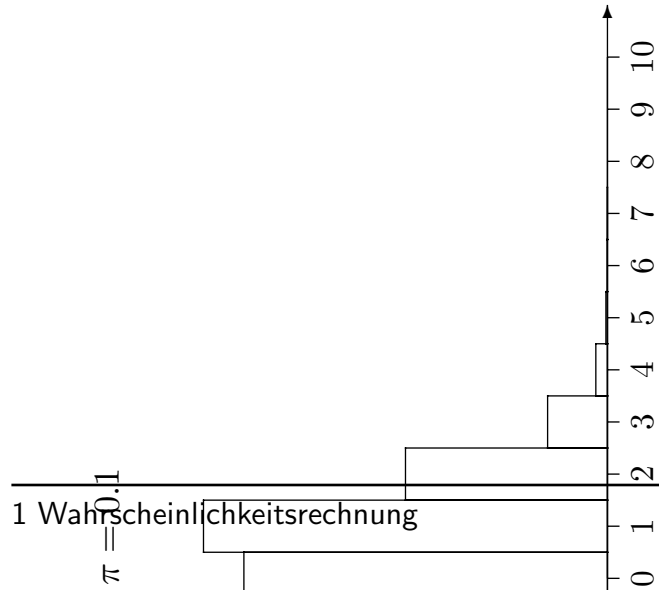
Eine Zufallsvariable heißt *binomialverteilt mit dem Parameter π bei n Versuchen*, kurz $X \sim B(n, \pi)$, wenn sie die Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} \binom{n}{x} \pi^x (1 - \pi)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Die $B(1, \pi)$ -Verteilung heißt auch *Bernoulliverteilung*.

Wahrscheinlichkeitshistogramme von Binomialverteilungen mit $n = 10$



Erwartungswert und Varianz:

- Zur Berechnung von Erwartungswert und Varianz der Binomialverteilung ist folgende Darstellung hilfreich:

$$X = X_1 + \dots + X_n$$

mit den binären Variablen

$$X_i = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst.} \end{cases}$$

- Die X_i sind stochastisch unabhängig mit

$$E(X_i) = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = \pi$$

$$\text{Var}(X_i) = E(X_i^2) - (E(X_i))^2 = 1 \cdot P(X_i = 1) - \pi^2 = \pi - \pi^2 = \pi(1 - \pi).$$

- Erwartungswert der Binomialverteilung:

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = n\pi$$

Die direkte Berechnung über

$$E(X) = \sum_{i=1}^n i \cdot P(\{X = i\}) = \sum_{i=1}^n i \binom{n}{i} \pi^i (1 - \pi)^{n-i} = \dots = n\pi$$

ist deutlich komplizierter!

- Varianz der Binomialverteilung:

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = n\pi(1 - \pi)$$

Bsp. 1.69.

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

- a) Schlagen Sie ein Modell für diese Situation vor und diskutieren Sie kurz die zugrunde gelegten Annahmen.
- b) Wie groß sind jeweils die Wahrscheinlichkeiten, dass von 20 Fahrern mindestens einer stürzt?
- c) Vergleichen Sie die durchschnittlich zu erwartende Anzahl von Stürzen von je 100 Rennläufern!

Exkurs: Zur Problematik der Argumentation mittels „natürlicher Häufigkeiten“ (vgl. v.a. Kap 1.2 und Kap 1.3):

Es wurde wiederholt vorgeschlagen, Wahrscheinlichkeiten anschaulich über „natürliche Häufigkeiten“ zu kommunizieren, also $P(A) = 0.3753$ darstellen als „von 10000 Personen haben 3753 die Eigenschaft A.“

Man würde demgemäß die Wahrscheinlichkeit $\pi_r = 0.1$ kommunizieren als „von 100 stürzen 10 Rennläufer“.

Diese Darstellung läuft Gefahr, die beträchtliche Variabilität zufälliger Prozesse zu verschleiern. In der Tat ist hier die Wahrscheinlichkeit, dass genau 10 von 100 Läufern stürzen,

$$P(X = 10) = \binom{100}{10} \cdot 0.1^{10} \cdot 0.9^{90} \approx 0.13,$$

also lediglich etwa 13%. „Natürliche Häufigkeiten“ müssen also unbedingt als Durchschnittswerte bzw. Erwartungswerte begriffen werden.

Eigenschaften der Binomialverteilung:

- Symmetrieeigenschaft (vertausche Rolle von A und \bar{A}):
- Summeneigenschaft: Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$. Sind X und Y unabhängig, so gilt

$$X + Y \sim$$

Tabellierung der Binomialverteilung: Tabelliert ist oft $P(X \leq x)$

$\pi = 0.3$	$n = 11$	$n = 12$...
$x \leq 0$	0.0198	0.0138	
1	0.1130	0.0850	...
2	0.3127	0.2528	
\vdots	\vdots	\vdots	

Daraus lassen sich die interessierenden Wahrscheinlichkeiten ablesen:

$$P(X = x) = P(X \leq x) - P(X \leq x - 1), \quad x \in \mathbb{N}_0$$

Zum Beispiel:

$$\begin{aligned} P(X = 2) &= P(X \leq 2) - P(X \leq 1) = \\ &= 0.1997 \\ &= \binom{11}{2} \cdot 0.3^2 \cdot 0.7^9. \end{aligned}$$

Wegen der Symmetrieeigenschaft gibt es meist nur Tabellen für $\pi \leq 0.5$.

Für großes n verwendet man Approximationen durch die Normalverteilung (vgl. Abschnitt 1.7).

1.6.2 Poisson Verteilung

Eine weitere wichtige diskrete Verteilung ist die Poisson-Verteilung. Sie modelliert die Anzahl (eher seltener) Ereignisse in einem Zeitintervall (Unfälle, Todesfälle; Sozialkontakte, deviante Verhaltensmuster, etc.).

Definition 1.70. [Poisson-Verteilung]

Eine Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x \in \{0, 1, \dots\} \\ 0, & \text{sonst} \end{cases}$$

heißt *Poisson-verteilt* mit *Parameter (oder Rate)* $\lambda > 0$, kurz $X \sim Po(\lambda)$. Es gilt

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda$$

Bem. 1.71.

Die Poisson-Verteilung kann auch als Näherungsmodell für eine Binomialverteilung gesehen werden, wenn die Anzahl der Versuchswiederholungen n groß und die „Trefferwahrscheinlichkeit“ π sehr klein ist (seltene Ereignisse!).

Der Erwartungswert λ ist dann gleich $n \cdot \pi$.

Es gilt also abgekürzt geschrieben

$$X \sim B(n, \pi) \underset{\substack{n \text{ groß} \\ \pi \text{ klein}}}{\implies} X \approx Po(n \cdot \pi)$$

Hat man mehrere unabhängige „Poisson-Prozesse“, also dynamische Situationen, bei denen die Ereignisanzahl Poisson-verteilt ist, also z.B. verschiedene deviante Verhaltensmuster, so ist die Gesamtanzahl der einzelnen Ereignisanzahlen wieder Poisson-verteilt, genauer gilt:

Satz 1.72. [Addition von Poisson-verteilten Zufallsvariablen]

Sind $X \sim Po(\lambda_X)$, $Y \sim Po(\lambda_Y)$ voneinander unabhängig, so gilt

$$X + Y \sim Po(\lambda_X + \lambda_Y).$$

Beachte, die Unabhängigkeit (genauer die Unkorreliertheit, siehe später) ist wesentlich. Nimmt man als Extremfall zwei Ereignisse, bei denen das eine das andere voraussetzt (Scheidungen, Prozesse um das Sorgerecht für Kinder), so ist die Gesamtzahl nicht mehr Poisson-verteilt.

Da bei der Poisson-Verteilung Erwartungswert und Varianz identisch sind, müsste gelten, wenn $X + Y$ Poisson-verteilt wäre:

$$\text{Var}(X + Y) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \text{Var}(X) + \text{Var}(Y),$$

was aber bei abhängigen (korrelierten) X und Y verletzt ist.

Bsp. 1.73.

Max geht gerne auf Open-Air Festivals. Im Durchschnitt trifft er dort 6 weibliche Bekannte und 3 männliche Bekannte.

- a) Formulieren Sie ein geeignetes Modell.
- b) Wie groß ist die Wahrscheinlichkeit, dass er genau 6 weibliche Bekannte trifft?
- c) Wie groß ist die Wahrscheinlichkeit, dass er mindestens einen männlichen Bekannten trifft?
- d) Berechnen Sie die Wahrscheinlichkeit, dass er weder einen männlichen noch eine weibliche Bekannte trifft, auf 2 verschiedene Arten. Diskutieren Sie eventuell zu treffende Zusatzannahmen.

1.6.3 Normalverteilung

Die Normalverteilung ist wohl das wichtigste Verteilungsmodell der Statistik, denn

- viele Zufallsvariablen sind (nach Transformation) (ungefähr) normalverteilt.
- beim Zusammenwirken vieler zufälliger Einflüsse ist der geeignet aggregierte Gesamteffekt oft approximativ normalverteilt (Zentraler Grenzwertsatz, Kap. 1.7).
- die asymptotische Grenzverteilung, also die Verteilung bei unendlich großem Stichprobenumfang, typischer statistischer Größen ist die Normalverteilung.

Definition 1.74.

Eine stetige Zufallsvariable X heißt *normalverteilt* mit den Parametern μ und σ^2 , in Zeichen $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \quad x \in \mathbb{R} \quad (1.6)$$

und *standardnormalverteilt*, in Zeichen $X \sim \mathcal{N}(0; 1)$, falls $\mu = 0$ und $\sigma^2 = 1$ gilt (π ist hier die Kreiszahl $\pi = 3.14\dots$).

Grundlegende Eigenschaften:

a) Die *Dichte der Standardnormalverteilung* wird oft mit $\varphi(x)$ bezeichnet, also

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

und die zugehörige Verteilungsfunktion mit

$$\Phi(x) = \int_{-\infty}^x \varphi(u) du$$

b) $\Phi(x)$ lässt sich nicht in geschlossener Form durch elementare Funktionen beschreiben
 \implies numerische Berechnung, Tabellierung.

c) μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann

$$\mathbb{E}(X) = \mu \quad \text{und} \quad \text{Var}(X) = \sigma^2.$$

d) Die Dichte ist symmetrisch um μ , d.h.

$$f(\mu - x) = f(\mu + x).$$

Grundlegendes zum Rechnen mit Normalverteilungen:

- Es gilt:

$$\Phi(-x) = 1 - \Phi(x)$$

(folgt aus der Symmetrie der Dichte).

- Gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die zugehörige standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma}$$

standardnormalverteilt.

- Entscheidende Eigenschaft für die Tabellierung: Es reicht, die Standardnormalverteilung zu tabellieren. Normalverteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 muss man, wie unten erläutert, zuerst standardisieren, dann kann man aber auch die Standardnormalverteilungstabelle verwenden.
- Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.

Ablesebeispiel: $\Phi(1.75) =$

- Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$

	0.00	0.01	...	0.05	...	0.09
⋮						
1.5	0.9332	0.9345	.	0.9394	.	0.9441
1.6	0.9452	0.9463	.	0.9505	.	0.9545
1.7	0.9554	0.9564	.	0.9599	.	0.9633
1.8	0.9641	0.9649	.	0.9678	.	0.9706
1.9	0.9713	0.9719	.	0.9744	.	0.9767
2.0	0.9772	0.9778	.	0.9798	.	0.9817
⋮						

Berechnung bei „allgemeiner Normalverteilung“: Wie bestimmt man bei $X \sim \mathcal{N}(\mu, \sigma^2)$ die Wahrscheinlichkeiten $P(X \leq a)$ aus der Tabelle der Standardnormalverteilung?

Abgeschlossenheit gegenüber Linearkombinationen: Seien X_1 und X_2 unabhängig

und $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt:

$$Y_1 := a_1 X_1 + b \sim \mathcal{N}(a_1 \mu_1 + b; a_1^2 \sigma_1^2)$$

$$Y_2 := a_1 X_1 + a_2 X_2 \sim \mathcal{N}(a_1 \mu_1 + a_2 \mu_2; a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2).$$

Das Ergebnis lässt sich auf mehrere Summanden verallgemeinern.

Bsp. 1.75. [aus Fahrmeir et al.]

- Schultischhöhe: $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\mu_Y = 113$, $\sigma_Y^2 = 16$
Stuhlhöhe: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $\mu_X = 83$, $\sigma_X^2 = 25$
- optimale Sitzposition: Tisch zwischen 27 und 29 cm höher als Stuhl.
- Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Paar zueinander gut passt?

Differenz: $Y - X$ soll zwischen $[27, 29]$ sein.

Definiere also

$$V := Y - X = Y + (-X)$$

Wegen $-X \sim \mathcal{N}(-83, 25)$ gilt dann

$$V \sim \mathcal{N}(113 - 83, 16 + 25) = \mathcal{N}(30, 41).$$

Außerdem ergibt sich durch Standardisieren:

$$\begin{aligned}27 \leq V \leq 29 &\iff 27 - 30 \leq V - 30 \leq 29 - 30 \\ &\iff \frac{27 - 30}{\sqrt{41}} \leq \frac{V - 30}{\sqrt{41}} \leq \frac{29 - 30}{\sqrt{41}}\end{aligned}$$

Damit lässt sich die gesuchte Wahrscheinlichkeit bestimmen:

$$\begin{aligned}P(27 \leq V \leq 29) &= P\left(-0.469 \leq \frac{V - 30}{\sqrt{41}} \leq -0.156\right) = \\ &= \Phi(-0.156) - \Phi(-0.469) = \\ &= (1 - \Phi(0.156)) - (1 - \Phi(0.469)) = \\ &= -0.5636 + 0.6808 = 0.1172\end{aligned}$$

1.7 Grenzwertsätze und Approximationen

Gerade in der Soziologie beobachtet man häufig *große* Stichprobenumfänge.

- Was ist aus der Sicht der Wahrscheinlichkeitsrechnung das Besondere daran?
- Vereinfacht sich etwas und wenn ja was?
- Kann man „Wahrscheinlichkeitsgesetzmäßigkeiten“ durch Betrachten vielfacher Wiederholungen erkennen?

1.7.1 Das i.i.d.-Modell

Betrachtet werden diskrete oder stetige Zufallsvariablen X_1, \dots, X_n , die *i.i.d.* (independently, identically distributed) sind, d.h. die

- 1) unabhängig sind und
- 2) die gleiche Verteilung besitzen.

Ferner sollen der Erwartungswert μ und die Varianz σ^2 existieren. Die Verteilungsfunktion werde mit F bezeichnet.

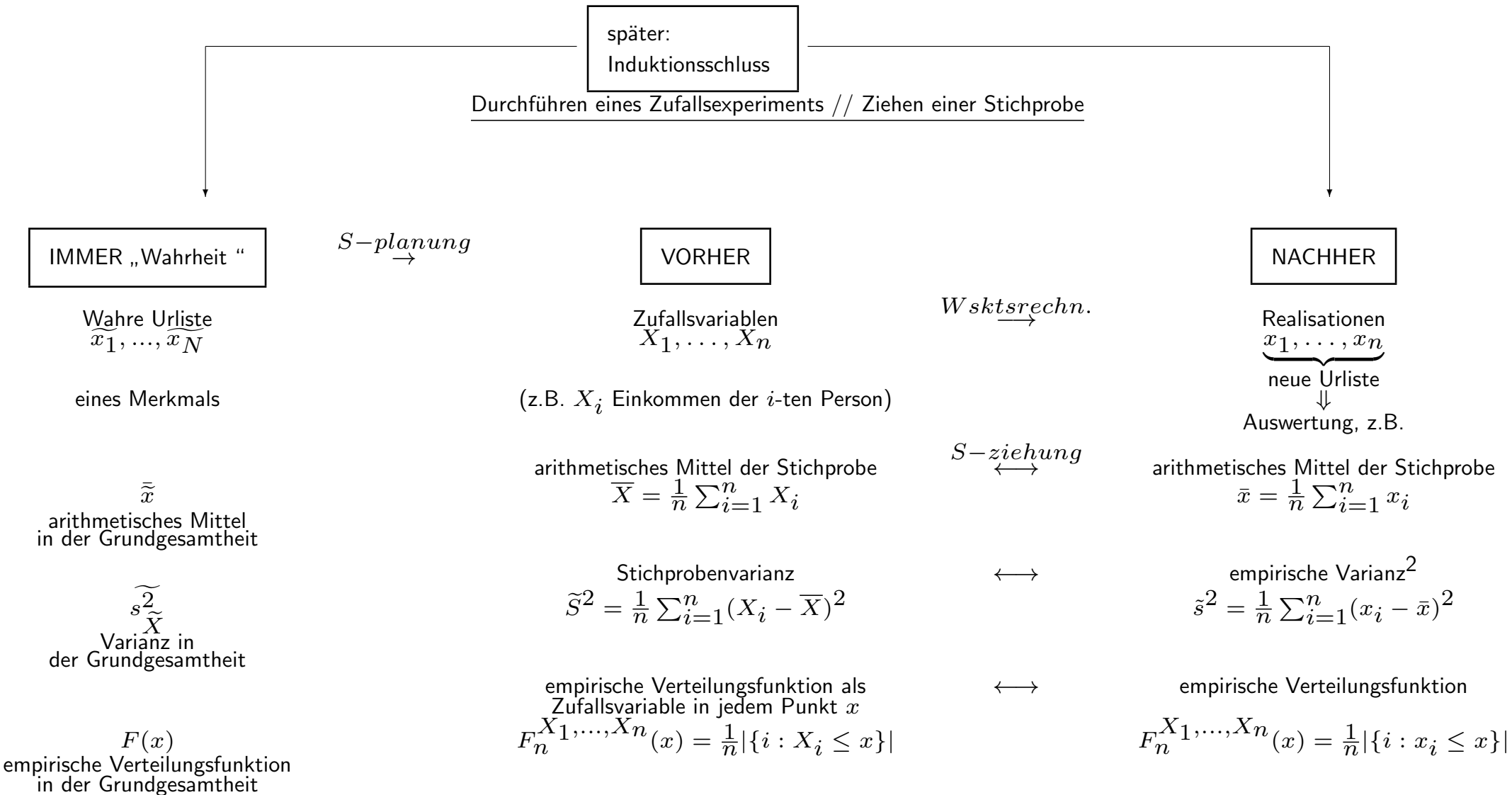
Dies bildet insbesondere die Situation ab, in der X_1, \dots, X_n eine Stichprobe eines Merkmals \tilde{X} bei reiner Zufallsauswahl sind.

Jede Funktion von X_1, \dots, X_n ist wieder eine Zufallsvariable, z.B. das arithmetische Mittel oder die Stichprobenvarianz

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Vor dem Ziehen der Stichprobe: Wahrscheinlichkeitsaussagen möglich \implies Wahrscheinlichkeitsrechnung anwenden

- Gerade bei diesen Zufallsgrößen ist die Abhängigkeit von n oft wichtig, man schreibt dann \bar{X}_n, \tilde{S}_n^2
- Sind X_1, \dots, X_n jeweils $\{0, 1\}$ -Variablen, so ist \bar{X}_n gerade die empirische *relative Häufigkeit* von Einsen in der Stichprobe vom Umfang n . Notation: H_n

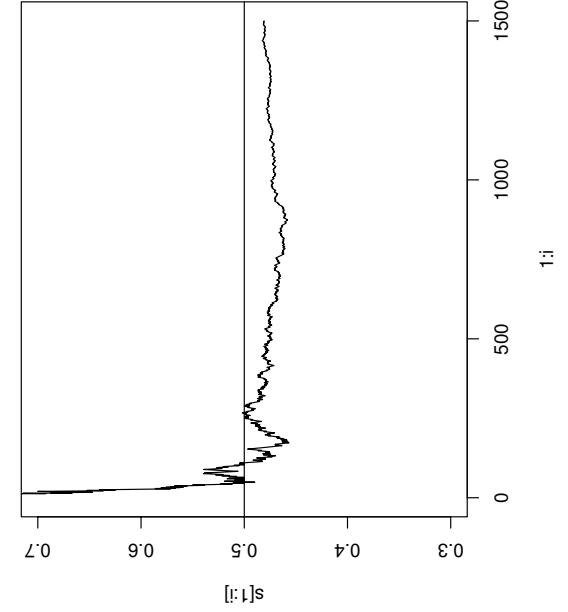
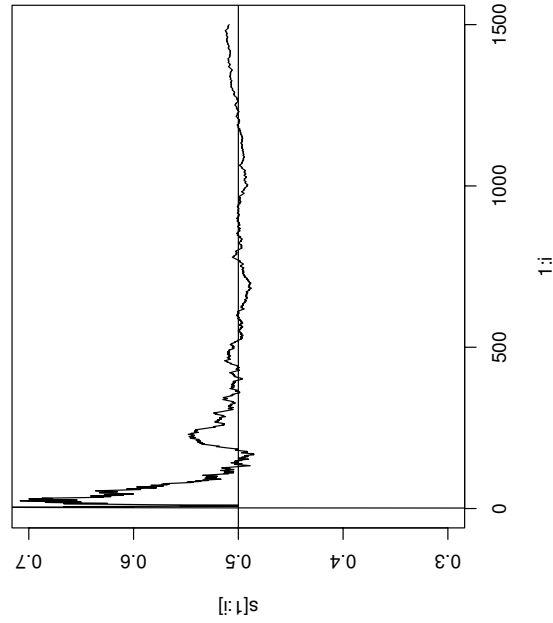
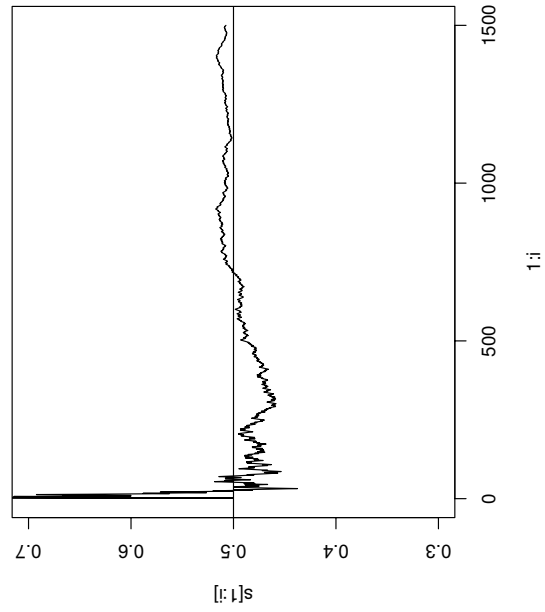


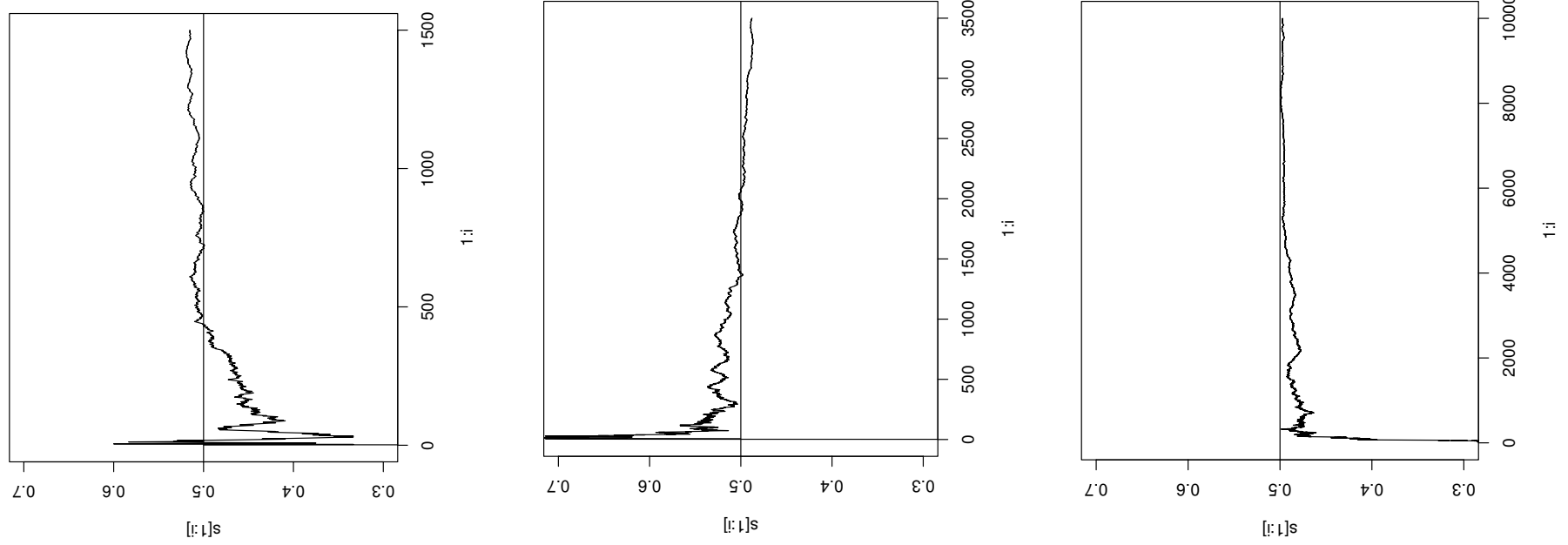
²Gehört nicht zur Grundgesamtheit; hier „ $\tilde{}$ “ für empirische Version

1.7.2 Das schwache Gesetz der großen Zahlen

Betrachte für wachsenden Stichprobenumfang n :

- X_1, \dots, X_n i.i.d.
- $X_i \in \{0, 1\}$ binäre Variablen mit $\pi = P(X_i = 1)$
- $H_n =$ die relative Häufigkeit der Einsen in den ersten n Versuchen.





Beobachtungen:

Theorem 1.76. [Theorem von Bernoulli]

Seien X_1, \dots, X_n , i.i.d. mit $X_i \in \{0, 1\}$ und $P(X_i = 1) = \pi$. Dann gilt für

$$H_n = \frac{1}{n} \sum_{i=1}^n X_i$$

(relative Häufigkeit der „Einsen“) und beliebig kleines $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|H_n - \pi| \leq \epsilon) = 1$$

Anschauliche Interpretation: Die relative Häufigkeit eines Ereignisses nähert sich praktisch sicher mit wachsender Versuchszahl an die Wahrscheinlichkeit des Ereignisses an.

Zwei wichtige Konsequenzen:

- 1) Häufigkeitsinterpretation von Wahrscheinlichkeiten:
- 2) Induktion: Man kann dieses Ergebnis nutzen, um Information über eine unbekannte Wahrscheinlichkeit ($\pi \hat{=}$ Anteil in einer Grundgesamtheit) zu erhalten.

Sei z.B. π der (unbekannte) Anteil der SPD Wähler, so ist die relative Häufigkeit in der Stichprobe eine „gute Schätzung für π “. Je größer die Stichprobe ist, umso größer ist die Wahrscheinlichkeit, dass die relative Häufigkeit sehr nahe beim wahren Anteil π ist.

Das Ergebnis lässt sich verallgemeinern auf Mittelwerte beliebiger Zufallsvariablen:

Schwaches Gesetz der großen Zahl: Gegeben seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit (existierendem) Erwartungswert μ und (existierender) Varianz σ^2 . Dann gilt für

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$$

und beliebiges $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1$$

Schreibweise:

$$\bar{X}_n \xrightarrow{P} \mu$$

(„Stochastische Konvergenz“, „ \bar{X}_n konvergiert in Wahrscheinlichkeit gegen μ “.)

Konsequenz für die Interpretation des Erwartungswerts:

1.7.3 Der Hauptsatz der Statistik

Satz 1.77. [Hauptsatz der Statistik]

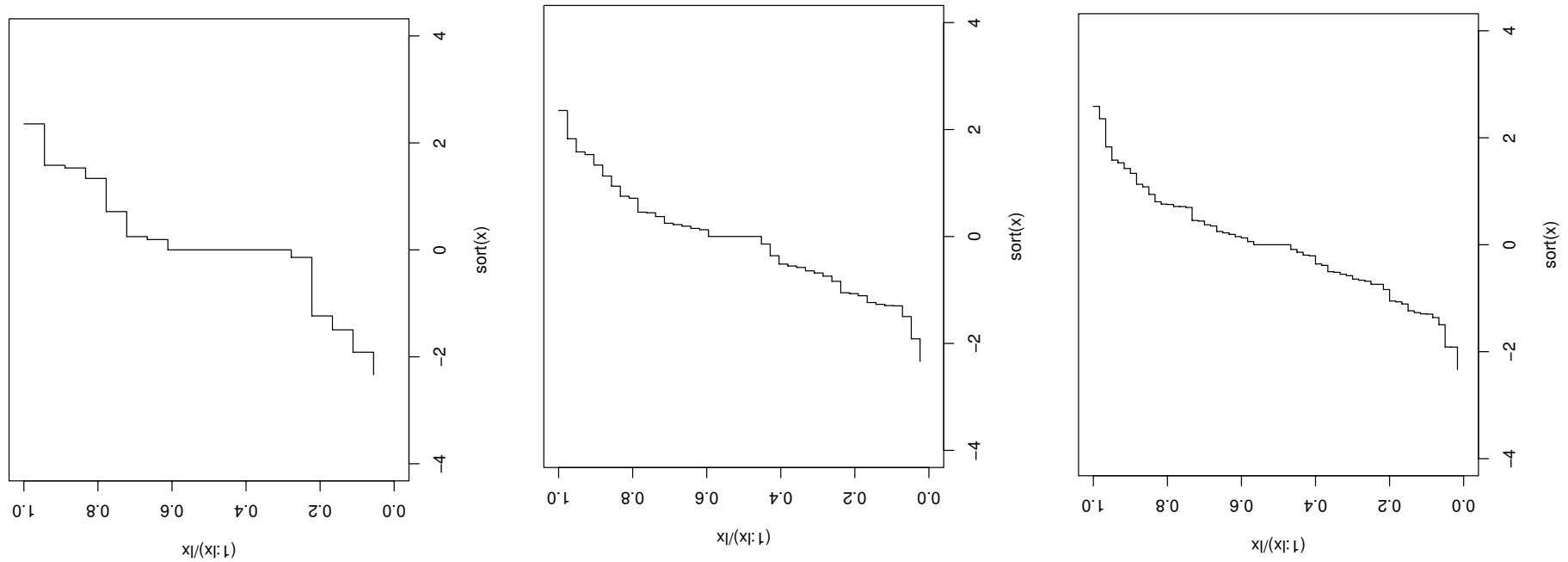
Seien X_1, \dots, X_n i.i.d. mit Verteilungsfunktion F und sei $F_n(x)$ die empirische Verteilungsfunktion der ersten n Beobachtungen. Mit

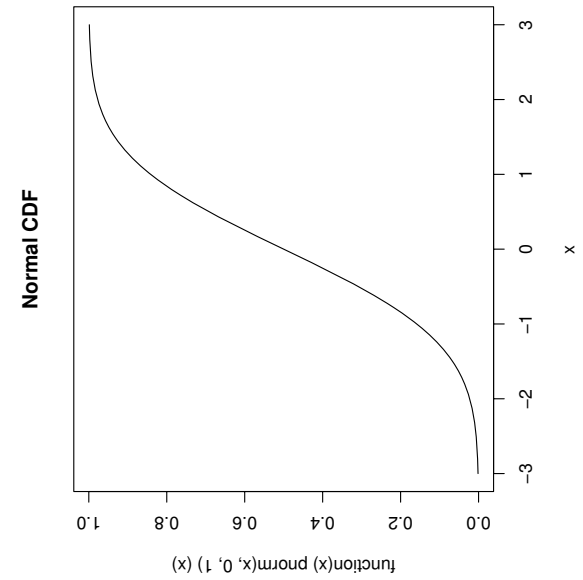
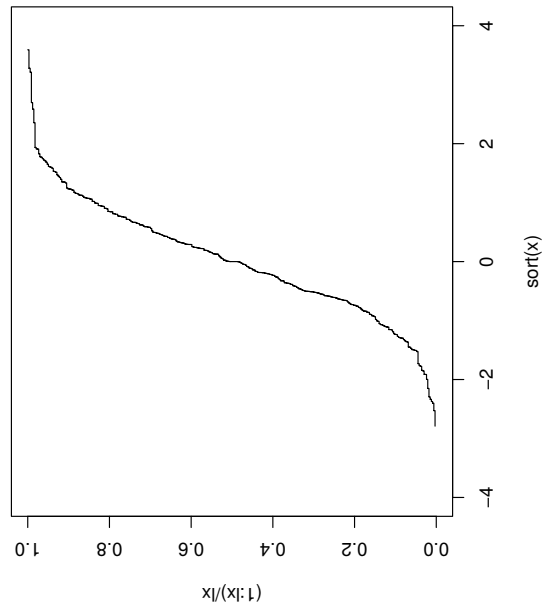
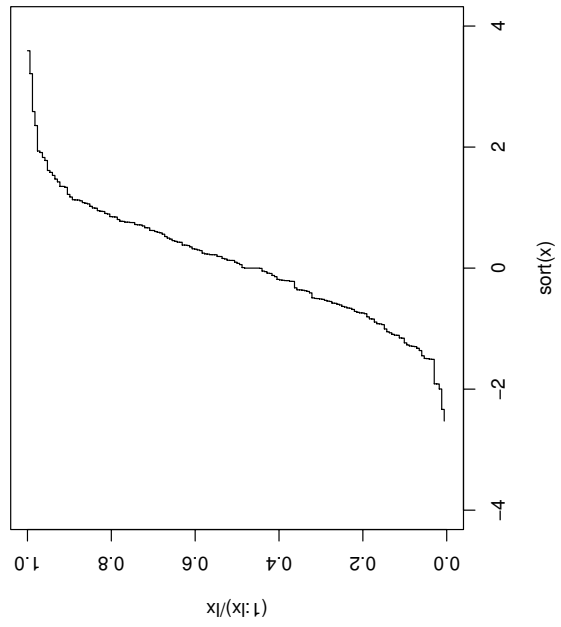
$$D_n := \sup_x |F_n(x) - F(x)|,$$

gilt für jedes $c > 0$

$$\lim_{n \rightarrow \infty} P(D_n > c) = 0.$$

Interpretation:





1.7.4 Der zentrale Grenzwertsatz

- Gibt es für große Stichprobenumfänge Regelmäßigkeiten im Verteilungstyp?
- Gibt es eine Standardverteilung, mit der man oft bei großen empirischen Untersuchungen rechnen kann? Damit kann man dann insbesondere Fehlermengen einheitlich behandeln.

Satz 1.78. [Zentraler Grenzwertsatz]

Seien X_1, \dots, X_n i.i.d. mit $E(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2 > 0$ sowie

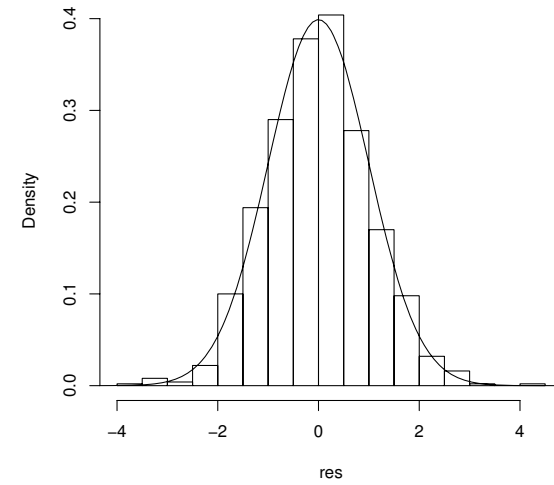
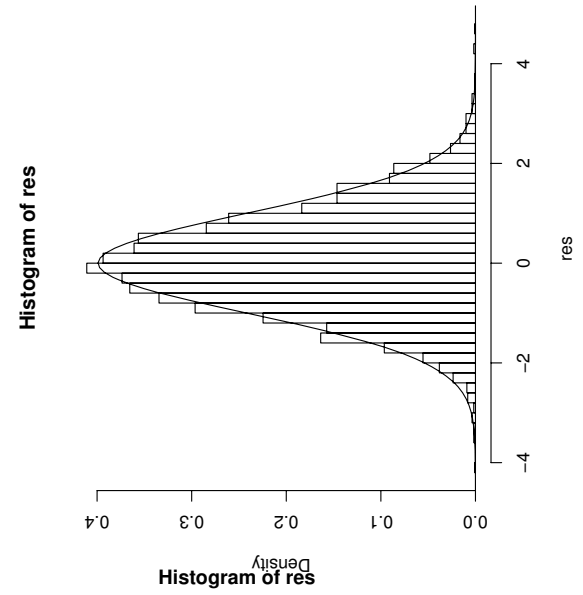
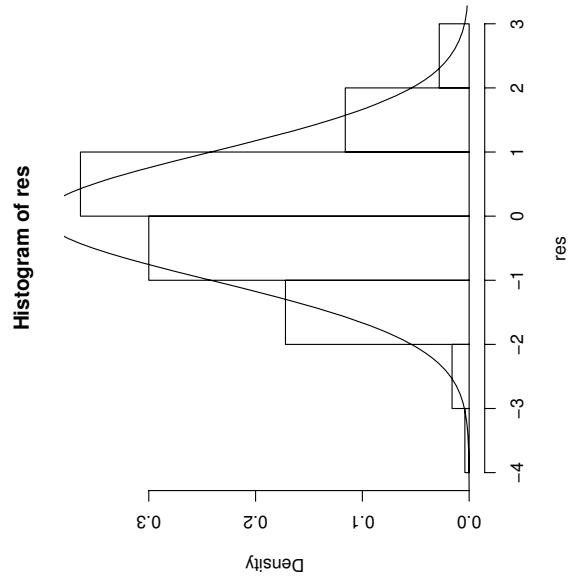
$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right).$$

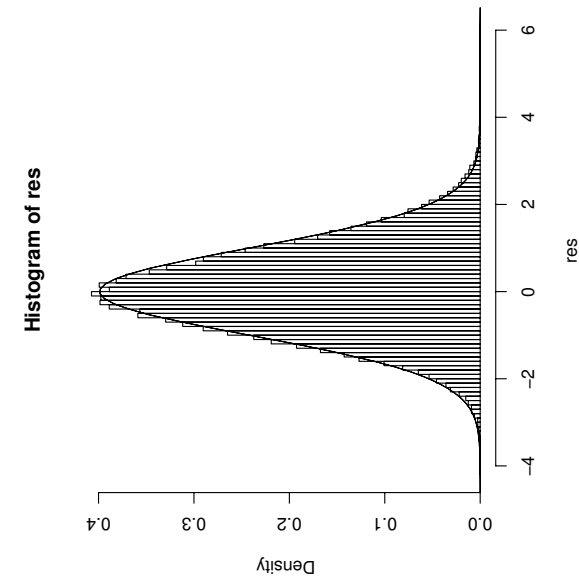
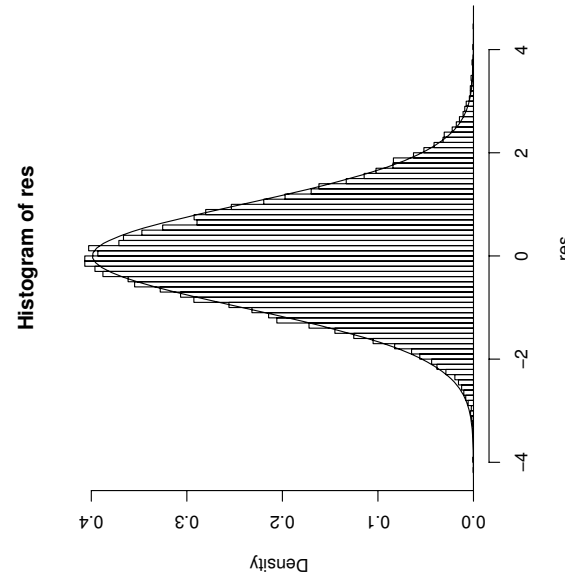
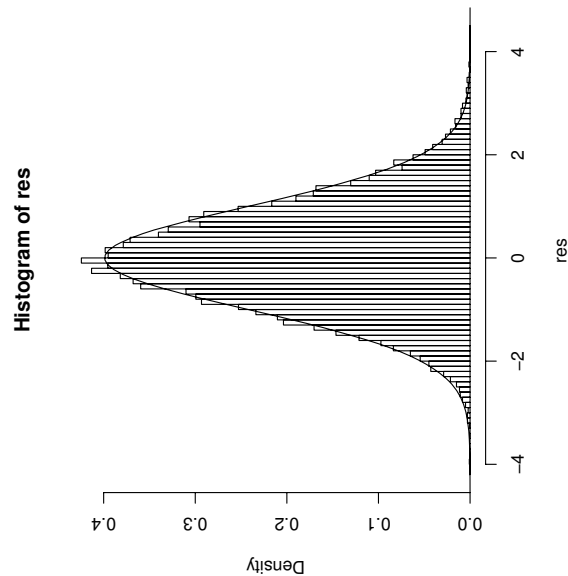
Dann gilt: Z_n ist *asymptotisch standardnormalverteilt*, in Zeichen: $Z_n \stackrel{a}{\sim} \mathcal{N}(0; 1)$, d.h. es gilt für jedes z

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z).$$

Für die Eingangsfragen gilt also:

- Ja, wenn man die Variablen geeignet mittelt und standardisiert, dann kann man bei großem n näherungsweise mit der Normalverteilung rechnen. Dabei ist für festes n die Approximation umso besser, je „symmetrischer“ die ursprüngliche Verteilung ist.





Anwendung des zentralen Grenzwertsatz auf \bar{X} :

Gemäß dem Gesetz der großen Zahlen weiß man: $\bar{X}_n \longrightarrow \mu$

Für die Praxis ist es aber zudem wichtig, die konkreten Abweichungen bei großem aber endlichem n zu quantifizieren, etwa zur Beantwortung folgender Fragen:

- Gegeben eine Fehlermarge ε und Stichprobenumfang n : Wie groß ist die Wahrscheinlichkeit, dass \bar{X} höchstens um ε von μ abweicht?

- Gegeben eine Fehlermarge ε und eine „Sicherheitswahrscheinlichkeit“ γ : Wie groß muss man n mindestens wählen, damit mit mindestens Wahrscheinlichkeit γ das Stichprobenmittel höchstens um ε von μ abweicht (*Stichprobenplanung*)?

Aus dem zentralen Grenzwertsatz folgt:

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) &= \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \cdot \sigma} \\ &= \frac{n\bar{X}_n - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\approx} \mathcal{N}(0, 1)\end{aligned}$$

oder auch

$$\bar{X}_n \stackrel{a}{\approx} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Wichtige Anwendung: Approximation der Binomialverteilung

Sei $X \sim B(n, \pi)$. Kann man die Verteilung von X approximieren?

Damit lässt sich der zentrale Grenzwertsatz anwenden:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{Y_i - \pi}{\sqrt{\pi(1 - \pi)}} \right) &= \frac{1}{\sqrt{n}} \frac{\sum Y_i - n \cdot \pi}{\sqrt{\pi(1 - \pi)}} \\ &= \frac{\sum Y_i - n \cdot \pi}{\sqrt{n \cdot \pi(1 - \pi)}} \stackrel{a}{\approx} \mathcal{N}(0, 1) \end{aligned}$$

und damit

$$\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \stackrel{a}{\approx} \mathcal{N}(0, 1)$$

so dass

$$P(X \leq x) \approx \Phi \left(\frac{x - n \cdot \pi}{\sqrt{n \cdot \pi(1 - \pi)}} \right)$$

falls n groß genug.

Trotzdem werden oft Faustregeln angegeben, ab wann diese Approximation gut ist, z.B.

$$n \cdot \pi \geq 5 \quad \text{und} \quad n \cdot (1 - \pi) \geq 5$$

$$n \cdot \pi(1 - \pi) \geq 9$$

Stetigkeitskorrektur: Durch die Approximation der *diskreten* Binomialverteilung durch die *stetige* Normalverteilung geht der diskrete Charakter verloren. Man erhält als Approximation $P(X = x) \approx 0$ für jedes $x \in \mathbb{N}$, was gerade für mittleres n unerwünscht ist.

Benutze deshalb

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

bei ganzzahligem $x \in \mathbb{N}$.

Man erhält als bessere Approximation

$$P(X \leq x) \approx \Phi \left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right)$$

$$P(X = x) \approx \Phi \left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right) - \Phi \left(\frac{x - 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right)$$

Fiktives Beispiel: Ein Politiker ist von einer gewissen umstrittenen Maßnahme überzeugt und überlegt, ob es taktisch geschickt ist, zur Unterstützung der Argumentation eine Mitgliederbefragung zu dem Thema durchzuführen. Er wählt dazu 200 Mitglieder zufällig aus und beschließt, eine Mitgliederbefragung zu „riskieren“, falls er in der Stichprobe mindestens 52% Zustimmung erhält.

Wie groß ist die Wahrscheinlichkeit, in der Stichprobe mindestens 52% Zustimmung zu erhalten, obwohl der wahre Anteil nur 48% beträgt?