### 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 $\Omega_X$, d.h. jeder Wert innerhalb eines Intervalls $[a, b]$ ist ein mögliches Ergebnis.

- Vorstellung: Auswertung eines *stetigen* Merkmals $X$ an zufällig ausgewählter Person aus einer Grundgesamtheit. (in Erweiterung der Situation aus 1.50)

- 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

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

Die diskrete Gleichverteilung auf diesem Gitter ergibt die Wahrscheinlichkeitsfunktion

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

$\Rightarrow$ Wählt man $n$ gerade und lässt nun $n \to \infty$, macht man also das Gitter beliebig fein, so geht die Wahrscheinlichkeit für jeden einzelnen Wert 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

$$P(X \leq 0.5) = \frac{n}{2} \cdot \frac{1}{n} = 0.5$$

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. 1.55.

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

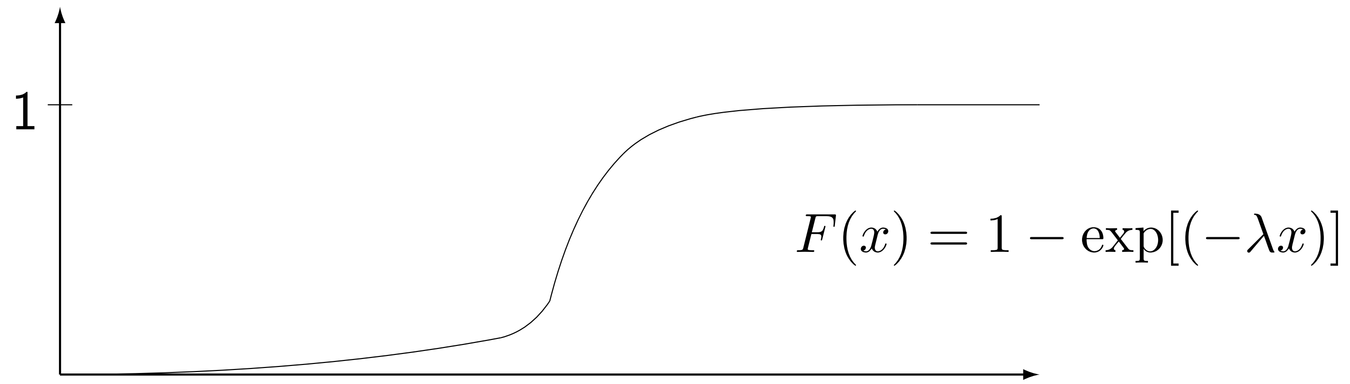
$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) \\ &= P(a < X < b) = F(b) - F(a) \end{aligned}$$

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

**Bem. 1.56.**

- Stetige Zufallsvariablen sind in der Wahrscheinlichkeitsrechnung und in der induktiven Statistik 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 1.57.**

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. 1.58.**

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 1.59.**

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 $a < b$

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

## Bem. 1.60.

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.

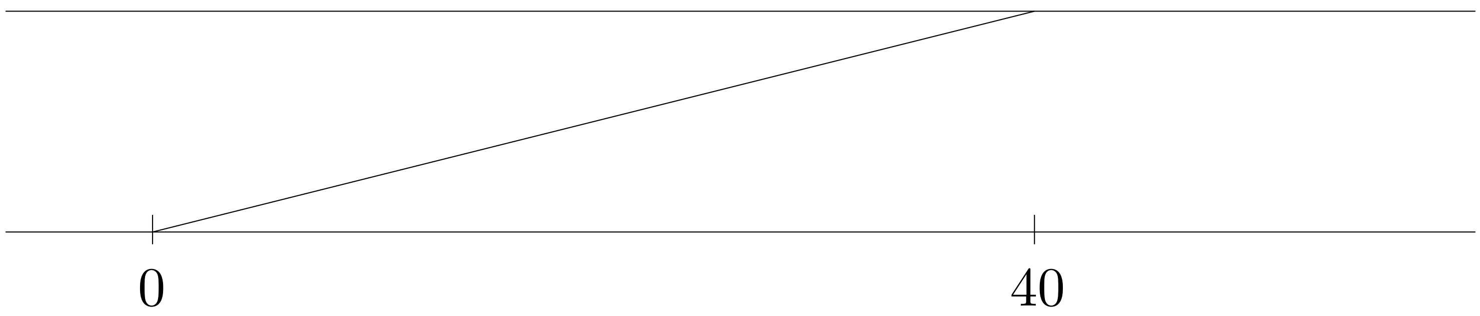
**Bsp. 1.61.**

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

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

und das Wahrscheinlichkeitsmaß $P$ über

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

## Bsp. 1.62.

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 $P(X \in [\frac{1}{4}, \frac{3}{4}])$ !

alternativ:

$$
\begin{aligned}
P\left(X \in \left[\frac{1}{4}; \frac{3}{4}\right]\right) &= F\left(\frac{3}{4}\right) - F\left(\frac{1}{4}\right) \\
&= \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 \\
&= \frac{9}{16} - \frac{1}{16} = \frac{1}{2}
\end{aligned}
$$

### 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.
- Brüderl, Josef, Peter Preisendörfer and Rolf Ziegler (1996): *Der Erfolg neugegründeter Betriebe: Eine empirische Studie zu den Chancen und Risiken von Unterneh-*

*mensgründungen [The Success of Newly Founded Firms: An Empirical Study]* Berlin: Duncker & Humblot. 309 pages. 2. edition 1998. 3. enlarged edition 2007.

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 1.63.

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.

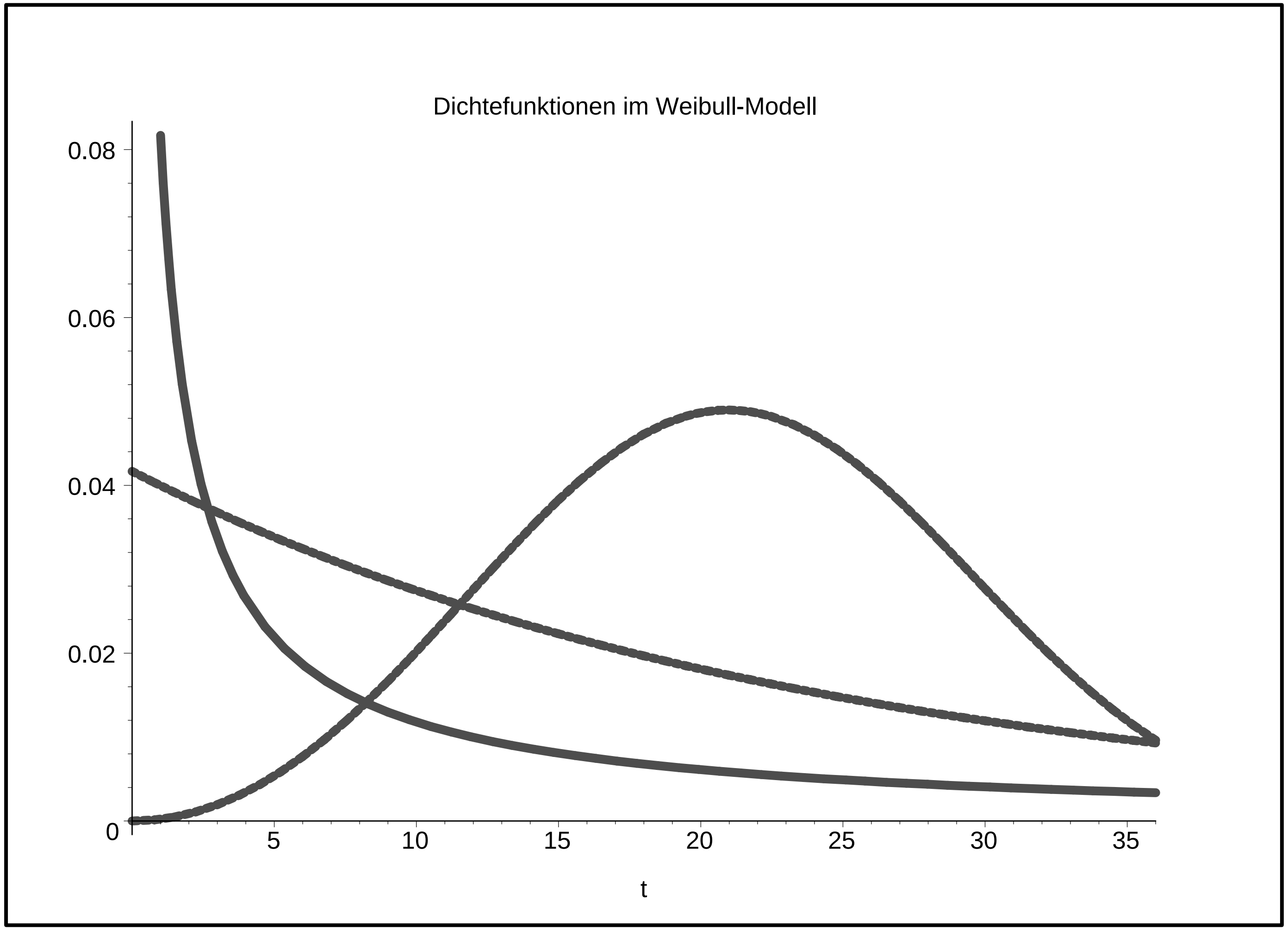
ii) Es gelten folgende Zusammenhänge (mit $\exp(z) := e^z$)

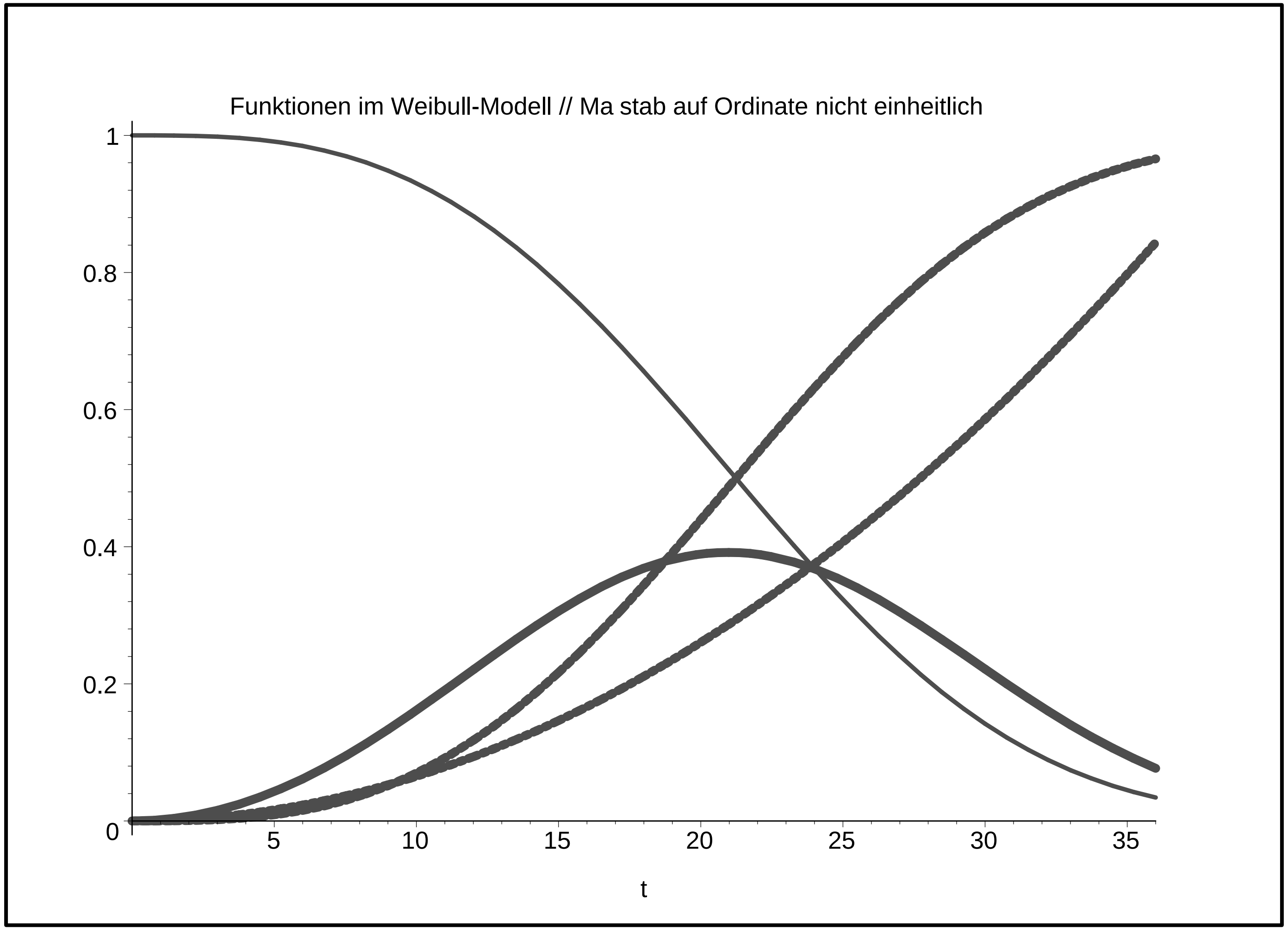
$$\begin{aligned}
S(x) &= 1 - F(x) \\
S(x) &= \exp\left(-\int_0^x \lambda(u)du\right) \\
F(x) &= 1 - \exp\left(-\int_0^x \lambda(u)du\right) \\
f(x) &= \lambda(x) \cdot S(x)
\end{aligned}$$

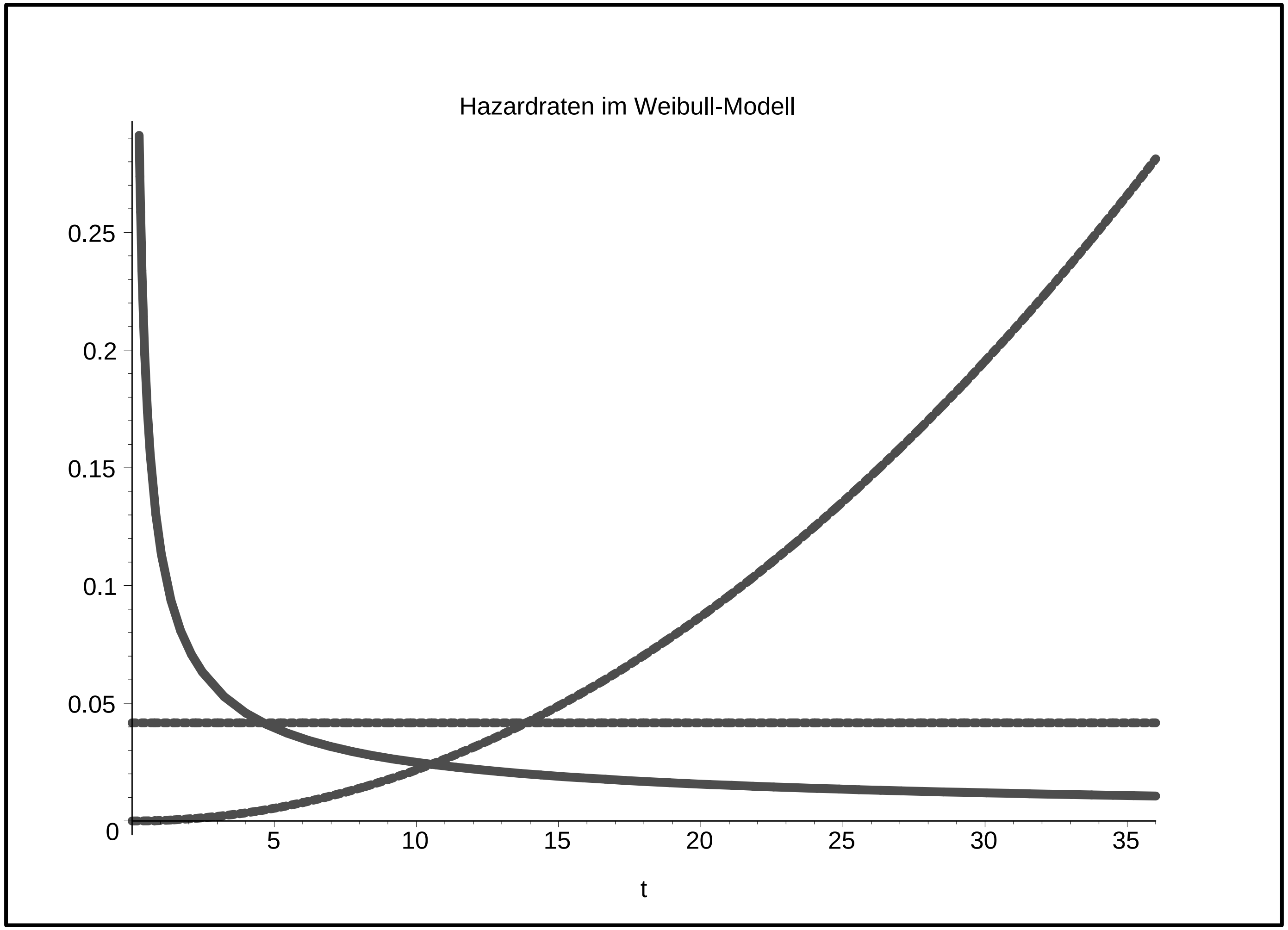
Zur Interpretation der Hazardrate: (Formel am geschicktesten oft von innen nach außen zu lesen)

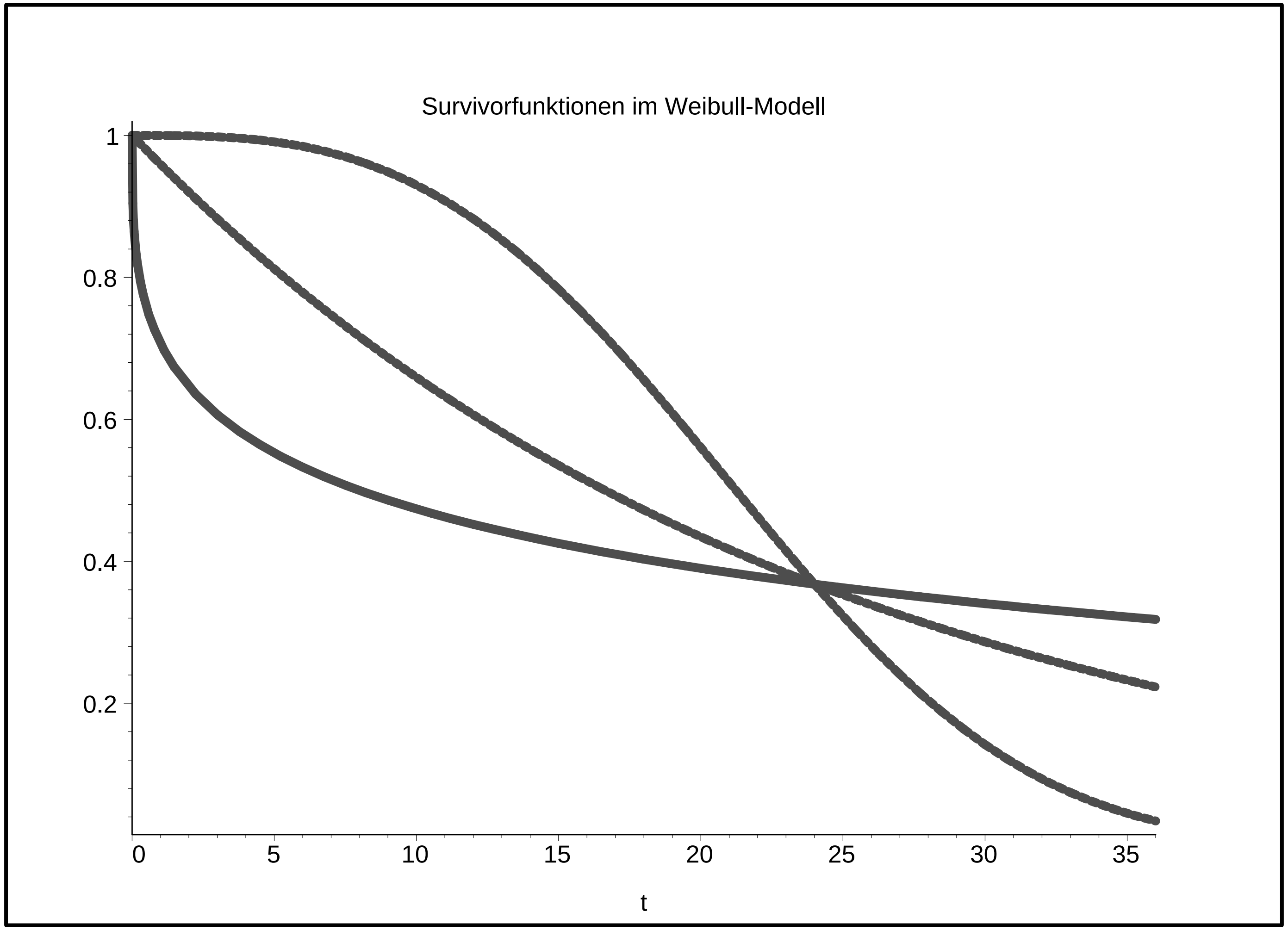
- Beachte: $\lambda(x)$ kann Werte zwischen 0 und unendlich annehmen, ist also insbesondere keine Wahrscheinlichkeit.

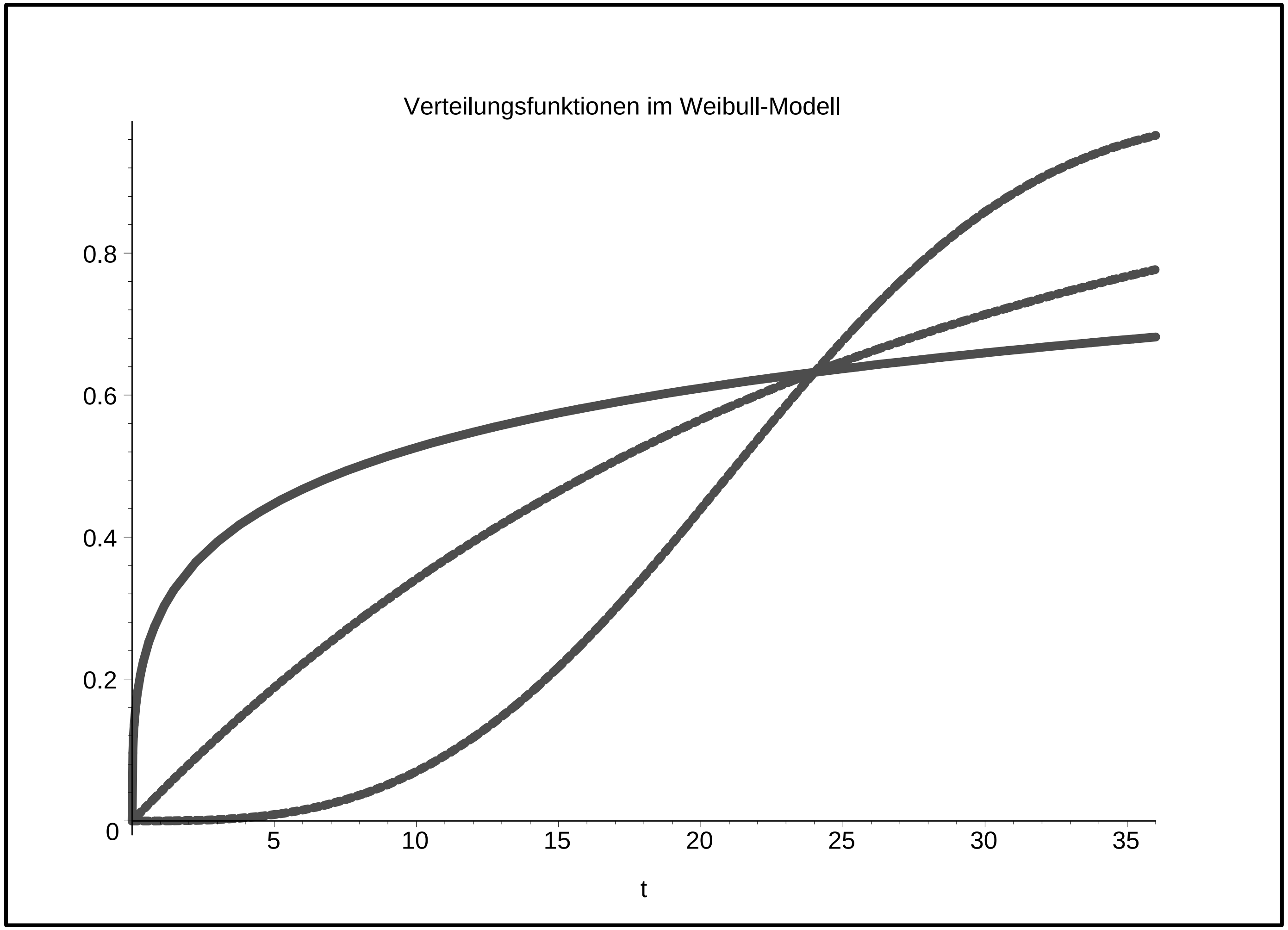
- Sehr anschauliches Instrument zur Beschreibung von Lebensdauerverteilungen.

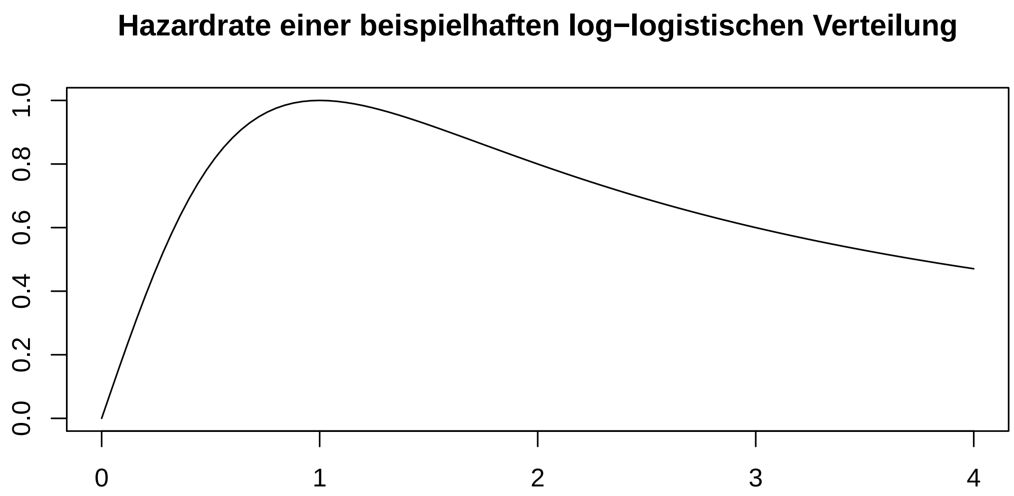

Dichtefunktionen im Weibull-Modell
0.08
0.06
0.04
0.02
0
5
10
15
20
25
30
35
t


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
0.2
0.4
0.6
0.8
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 1.64.**

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. 1.65.

- 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.