

## 10 Verteilungsfunktionen (Teil 2)

### 10.1 Nomenklatur für Verteilungsparameter

- Ein Ortsparameter (location parameter)  $\gamma$  verändert nur die Lage der Verteilung auf der x-Achse.

- Beispiel: Konstante Verteilung mit  $x(i) = \begin{cases} 1 & \text{für } i = \gamma \\ 0 & \text{sonst} \end{cases}$

- Ein Skalierungsparameter (scale parameter)  $\beta$  verändert den Erwartungswert und die Varianz aber nicht den Variationskoeffizienten der ZV.

- Generell: Sei  $X(\beta)$  nach einer Verteilung mit Skalierungsparameter  $\beta$  verteilt. Dann ist  $Y=c \cdot X(\beta)$  nach der gleichen Verteilung mit Skalierungsparameter  $c \cdot \beta$  verteilt, d.h  $X(c \cdot \beta)$ .

- Beispiel:  $\beta=1/\lambda$  bei der Exponentialverteilung

$$\begin{aligned} P(c \cdot X(\beta) \leq t) &= P\left(X(\beta) \leq \frac{t}{c}\right) = 1 - e^{-\frac{1}{\beta} \cdot \left(\frac{t}{c}\right)} = 1 - e^{-\frac{1}{c \cdot \beta} \cdot t} \\ &= P(X(c \cdot \beta) \leq t) \end{aligned}$$

- Ein Formparameter (shape parameter)  $\alpha$  verändert auch die Form der Verteilung.

- Beispiel:  $\alpha=k$  bei der Erlang-k-VerteilungVDF

## 10.2 Kontinuierliche Verteilungen (LK 6.2.2, 6.2.4 und 8.3)

### 10.2.1 Dreiecksverteilung: ZV $X \sim \text{triang}(a, b, c)$ (LK 8.3.15)

- VDF:  $f(x) = \begin{cases} \frac{2 \cdot (x - a)}{(b - a) \cdot (c - a)} & \text{if } a \leq x \leq c \\ \frac{2 \cdot (b - x)}{(b - a) \cdot (b - c)} & \text{if } c \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

- VF:  $F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{(x - a)^2}{(b - a) \cdot (c - a)} & \text{if } a \leq x \leq c \\ 1 - \frac{(b - x)^2}{(b - a) \cdot (b - c)} & \text{if } c \leq x \leq b \\ 1 & \text{if } b < x \end{cases}$

- Range  $[a, b]$
- $E[X] = (a + b + c) / 3$
- $\text{VAR}[X] = (a^2 + b^2 + c^2 - ab - ac - bc) / 18$
- Mode:  $c$
- Erzeugung von  $X \sim \text{triang}(0, 1, c)$  mit  $0 < c < 1$  durch Inversion:

$$U \sim U(0,1).$$

falls  $U < c$

$$\text{dann } X = \sqrt{cU}$$

$$\text{sonst } X = 1 - \sqrt{(1 - c)(1 - U)}.$$

- Transformationseigenschaft:

$$X \sim \text{triang}\left(0, 1, \frac{c - a}{b - a}\right), \quad a < c < b \Rightarrow a + (b - a) X \sim \text{triang}(a, b, c)$$

- Anwendung von Orts-, Skalierungs- und Formparameter
  - Ortsparameter  $\gamma = a$ : Linke Dreiecksgrenze
  - Skalierungsparameter  $\beta = b/a$ : Dreiecksbreite  $b = a \cdot \beta$  (nur für  $a > 0$  anwendbar)
  - Formparameter  $\alpha = c/(b - a)$ : Relative Position der Dreiecksspitze

### 10.2.2 Gewichtete Gleichverteilungen

- m Intervalle  $[a_i, b_i)$  mit Wahrscheinlichkeit  $p_i$ ,  $0 \leq i < m$
- Range  $[a_0, b_{m-1}]$

$$\bullet \quad VDF : f(x) = \begin{cases} 0 & \text{if } x \leq a_0 \\ \frac{p_i}{b_i - a_i} & \text{if } a_i < x \leq b_i \\ 0 & \text{if } b_{m-1} < x \end{cases}$$

$$\bullet \quad VF : F(x) = \begin{cases} 0 & \text{if } x \leq a_0 \\ \sum_{0 \leq j < i} p_j + \frac{x - a_i}{b_i - a_i} \cdot p_i & \text{if } a_i < x \leq b_i \\ 1 & \text{if } b_{m-1} < x \end{cases}$$

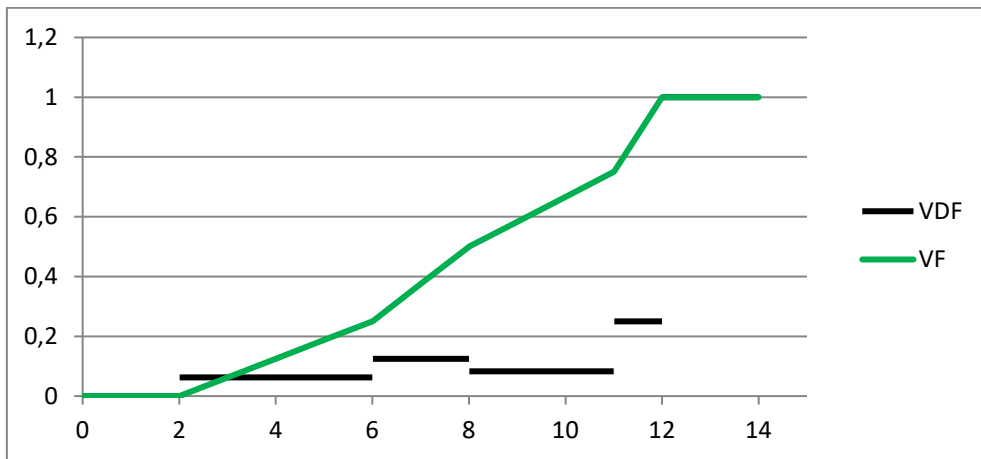


Abbildung 10-1: Verteilungsdichte- und Verteilungsfunktion einer gewichteten Gleichverteilung.

- Erzeugung durch Inversion:

$$U \sim U(0,1)$$

$$i = \max_j \left( j : \left( \sum_{0 \leq k < j} p_k \right) \leq U \right)$$

$$X = a_i + \frac{U - \sum_{0 \leq k < i} p_k}{p_i} \cdot (b_i - a_i)$$

Keine Inversion des ganzen Streckenzuges nötig, sondern nur jeweils einer Teilstrecke

- Bei der Methode muss gelten  $b_i = a_{i+1}$ , sie kann aber leicht erweitert werden, so dass diese Einschränkung nicht nötig ist.
- Diese Methode kann sehr rechenintensiv sein, wenn die Anzahl der Intervalle  $n$  groß ist. In diesen Fällen sollten intelligente Datenstrukturen verwendet werden z.B. Indexe.

### 10.2.3 Spezialfall der vorherigen Verteilung: empirische Verteilung basierend auf einzelnen Stichproben (LK 6.2.4, 8.3.16)

- Gegeben:  $n+1$  reellwertige Stichproben  $X_0, X_1, \dots, X_n$ , so dass keine zwei identisch sind
- Aufsteigende Sortierung:  $X_{(0)} \leq X_{(1)} \leq \dots \leq X_{(n)}$   
Beispiel:  $n=4$ ,  $X_{(0)}=2$ ,  $X_{(1)}=6$ ,  $X_{(2)}=8$ ,  $X_{(3)}=11$ ,  $X_{(4)}=12$
- Diese induzieren  $n$  Intervalle mit Grenzen  $[X_{(i)}, X_{(i+1)})$   
für  $0 \leq i < n$  und jedes dieser Intervalle ist gleichwahrscheinlich mit  $p_i = \frac{1}{n}$
- Range  $[X_0, X_n]$
- Problem: Momente der empirischen kontinuierlichen Verteilung und der Stichprobe stimmen i.A. nicht überein.
- Siehe Beispiel in 10.2.2

#### 10.2.4 Gamma Verteilung: ZV $X \sim \text{gamma}(\alpha, \beta)$ (LK 8.3.4)

- Die Gamma-Verteilung ist eine Erweiterung der Erlang-k-Verteilung auf  $k \in \mathbb{R}^+$ :  $k\text{-Erlang}(\lambda) \sim \text{gamma}(k, 1/\lambda)$  für  $k \in \mathbb{N}$

- VDF: 
$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\beta^{-\alpha} \cdot x^{\alpha-1} \cdot e^{-x/\beta}}{\Gamma(\alpha)} & \text{if } 0 < x \end{cases}$$

- Range:  $[0, \infty)$

- $E[X] = \alpha \cdot \beta$ ,  $c_{var} = 1/\sqrt{\alpha}$

- VF: keine geschlossene Form außer für  $\alpha \in \mathbb{N}$  (Erlang-k Verteilung)

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-x/\beta} \cdot \sum_{0 \leq i < \alpha} \frac{(x/\beta)^i}{i!} & \text{if } 0 < x \end{cases}$$

- Mode  $\begin{cases} 0 & \text{if } \alpha < 1 \\ \beta \cdot (\alpha - 1) & \text{if } 1 \leq \alpha \end{cases}$

- $\lim_{x \rightarrow 0} (f(x)) = \begin{cases} \infty & \text{if } \alpha < 1 \\ \frac{1}{\beta} & \text{if } \alpha = 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$

- Besonderheiten

- Da  $\beta$  Skalierungsparameter ist, gilt:

$$X \sim \text{gamma}(\alpha, \beta), c > 0 \Rightarrow c \cdot X \sim \text{gamma}(\alpha, c \cdot \beta)$$

- Außerdem:  $\sum_{0 \leq i < m} \text{gamma}(\alpha_i, \beta) \sim \text{gamma}\left(\left(\sum_{0 \leq i < m} \alpha_i\right), \beta\right)$

- $X_i \sim \text{gamma}(\alpha_i, \beta), i=1,2 \Rightarrow X_1/(X_1+X_2) \sim \text{beta}(\alpha_1, \alpha_2)$

- $\chi^2(k) \sim \text{gamma}(k/2, 2)$

- Beispiele

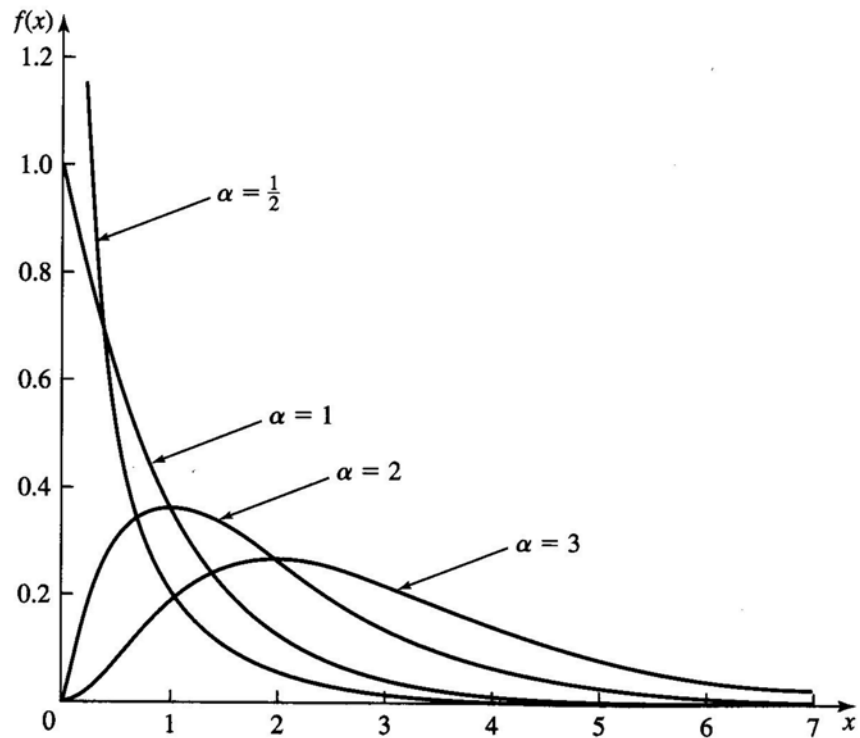


Abbildung 3.2: Verteilungsdichtefunktion von Gamma-Verteilungen (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 302)

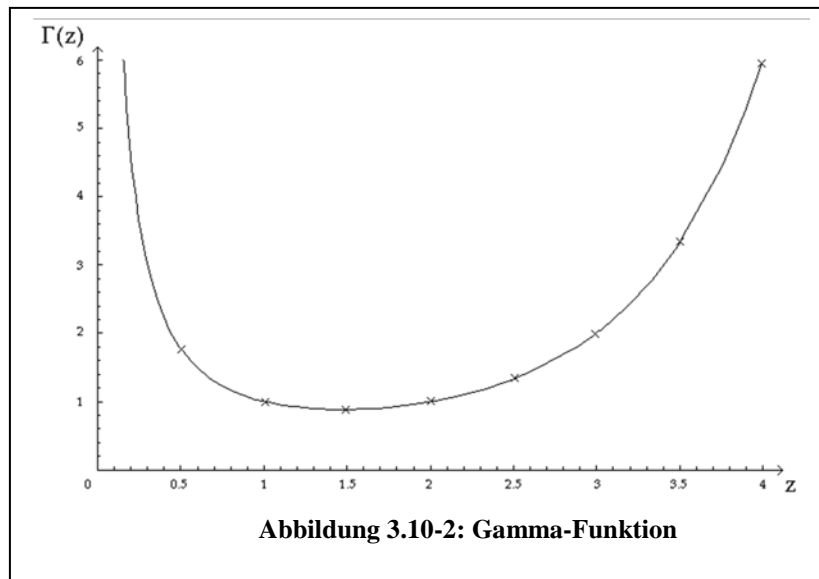
- Zum Hintergrund der Gamma-Funktion:

$$\Gamma(z) = \begin{cases} \int_0^{\infty} t^{z-1} \cdot e^{-t} dt & \text{if } 0 \leq x \\ 0 & \text{if } x < 0 \end{cases}$$

$$\Gamma(z+1) = z \cdot \Gamma(z) \text{ für } z > 0,$$

$$\Gamma(k) = (k-1)! \text{ für } k \in \mathbb{N}_0$$

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$



- Erzeugung von  $\text{gamma}(\alpha, \beta)$ -verteilten Zufallsvariablen
  1. Rückführung von  $X \sim \text{gamma}(\alpha, \beta)$  auf  $X = \beta \cdot Y$  mit  $Y \sim \text{gamma}(\alpha, 1)$
  2. Erzeugung von  $X \sim \text{gamma}(\alpha, 1)$  durch Accept-Reject Methode:
    1. Fall:  $0 < \alpha < 1$

$$t(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^{\alpha-1}}{\Gamma(\alpha)} & 0 < x \leq 1 \\ \frac{e^{-x}}{\Gamma(\alpha)} & 1 < x \end{cases} \Rightarrow r(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\alpha x^{\alpha-1}}{b} & 0 < x \leq 1 \\ \frac{\alpha e^{-x}}{b} & 1 < x \end{cases},$$

$$\text{mit } b = \frac{e + \alpha}{e}, \text{ da } c = \int_0^{\infty} t(x) dx = \frac{b}{\alpha \cdot \Gamma(\alpha)}$$

für die Erzeugung von  $Y$  benötigt man die Inverse der VF  $R(x)$

$$R(x) = \begin{cases} \frac{x^{\alpha}}{b} & 0 \leq x \leq 1 \\ 1 - \frac{\alpha e^{-x}}{b} & 1 < x \end{cases} \Rightarrow R^{-1}(u) = \begin{cases} (bu)^{1/\alpha} & u < \frac{1}{b} \\ -\ln \frac{b(1-u)}{\alpha} & \text{sonst} \end{cases}$$

wegen  $\frac{f(Y)}{t(Y)} = \begin{cases} e^{-Y} & 0 \leq Y \leq 1 \\ Y^{\alpha-1} & 1 < Y \end{cases}$  ergibt sich folgender Algorithmus:

1.  $U_1 \sim U(0,1)$ ,  $P = bU_1$ ,

2. falls  $P \leq 1$

$$Y = P^{1/\alpha}, \text{ (ergibt } 0 < Y \leq 1)$$

$$U_2 \sim U(0,1); \text{ falls } U_2 \leq e^{-Y} \text{ Accept, ansonsten Reject}$$

falls  $P > 1$

$$Y = -\ln[(b - P)/\alpha], \text{ (ergibt } 1 < Y)$$

$$U_2 \sim U(0,1); \text{ falls } U_2 \leq Y^{\alpha-1} \text{ Accept, ansonsten Reject}$$

2. Fall:  $\alpha > 1$

$$t(x) = c r(x) \text{ mit } c = \frac{4 \alpha^\alpha e^{-\alpha}}{\lambda \Gamma(\alpha)}, \lambda = \sqrt{2\alpha - 1}$$

$$r(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\lambda \mu x^{\lambda-1}}{(\mu + x^\lambda)^2} & \text{if } 0 < x \end{cases} \text{ mit } \mu = \alpha^\lambda$$

$$\Rightarrow R(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{x^\lambda}{\mu + x^\lambda} & \text{if } 0 < x \end{cases} \Rightarrow R^{-1}(u) = \left( \frac{\mu u}{1-u} \right)^{1/\lambda}$$



durch geeignete Vorverarbeitung erhält man mit

$$a = \frac{1}{\sqrt{2\alpha - 1}}, \quad b = \alpha - \ln 4, \quad q = \alpha + \frac{1}{a}, \quad \theta = 4.5, \quad d = 1 + \ln \theta$$

folgenden Algorithmus:

- $U_1, U_2 \sim U(0,1)$ .
- $V = a \ln \frac{U_1}{1-U_1}, Y = \alpha \cdot e^V, Z = U_1^2 \cdot U_2, W = b + q \cdot V - Y$ .
- falls  $(W + d - \theta \cdot Z \geq 0)$  dann Accept  
     ansonsten{falls  $(W \geq \ln Z)$  dann Accept ansonsten Reject}

Der Vortest  $(W + d - \theta \cdot Z \geq 0)$  könnte auch ausgelassen werden. Er dient aber der Effizienz, da dadurch in vielen Fällen die Berechnung des Logarithmus im Haupttest vermieden werden kann.

Beispiel: gamma(2,1), d.h.  $f(x) = x e^{-x}$  für  $x > 0$ ; damit ergibt sich:

$\lambda = \sqrt{3}, \mu = 2\sqrt{3}, c = \frac{16e^{-2}}{\sqrt{3}}$  für Fall 2. Wegen gamma(2,1) ~ 2-Erlang(1) kann

diese Verteilung auch anders erzeugt werden.

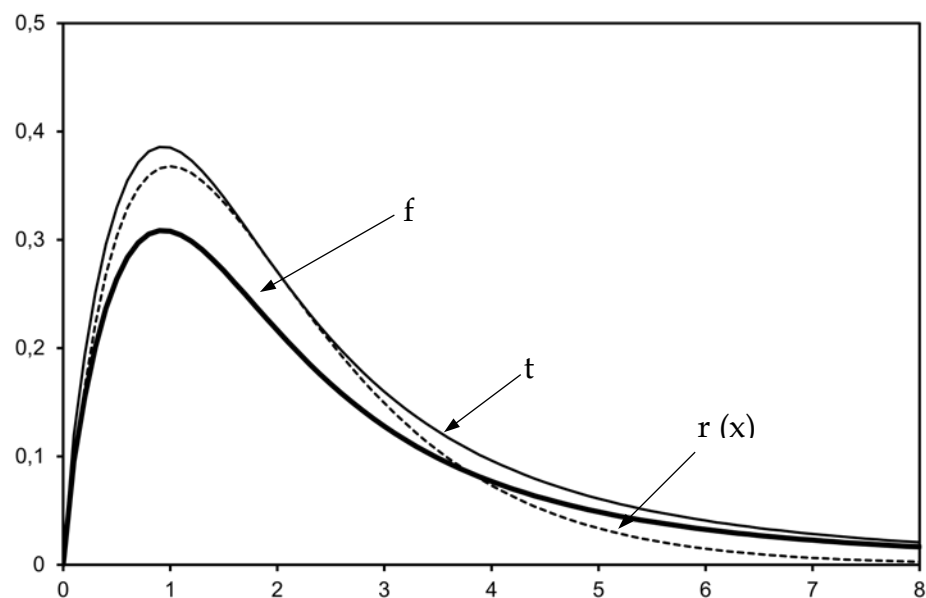
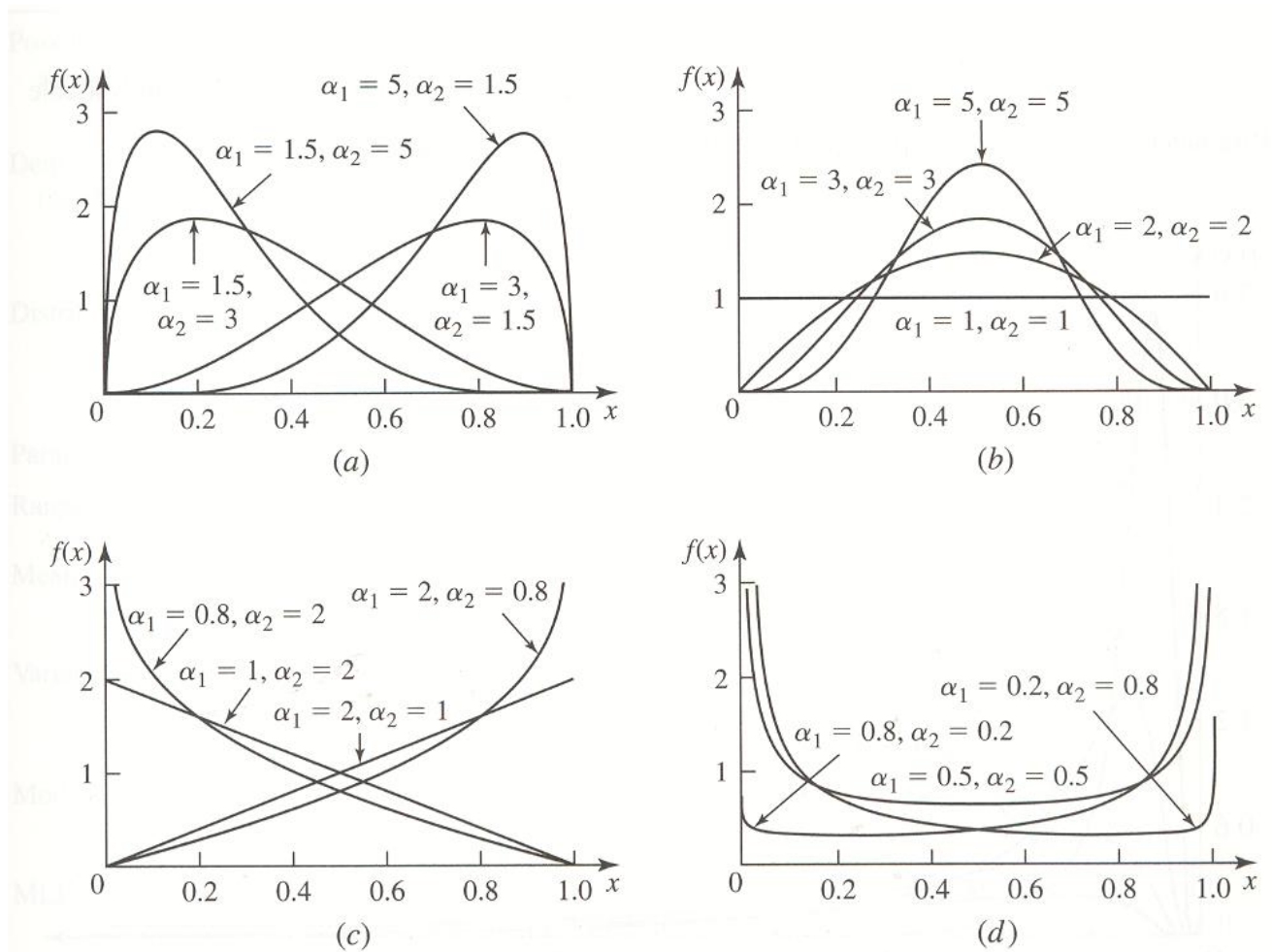


Abbildung 3.3: VDF  $f(x)$ , Majorante  $t(x)$  und dazugehörige VDF  $r(x)$  zur Erzeugung von Gamma (2, 1) durch die Accept/Reject-Methode

### 10.2.5 Beta Verteilung: ZV $X \sim \text{beta}(\alpha_1, \alpha_2)$ (LK 8.3.8)

- VDF:  $f(x) = \begin{cases} \frac{x^{\alpha_1-1} \cdot (1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$
- Range:  $[0, 1]$
- Beta-Funktion:  $B(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1-1} \cdot (1-x)^{\alpha_2-1} dx = \frac{\Gamma(\alpha_1) \cdot \Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$
- Formparameter:  $\alpha_1, \alpha_2$
- VF: keine geschlossene Form
- $E[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$
- $\text{VAR}[X] = \frac{\alpha_1 \cdot \alpha_2}{(\alpha_1 + \alpha_2)^2 \cdot (\alpha_1 + \alpha_2 + 1)}$
- Mode:  $\begin{cases} \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2} & \text{für } 1 < \alpha_1, \alpha_2 \\ 0 \text{ and } 1 & \text{für } \alpha_1, \alpha_2 < 1 \\ 0 & \text{für } (\alpha_1 < 1 \wedge 1 \leq \alpha_2) \text{ oder } (\alpha_1 = 1 \wedge 1 < \alpha_2) \\ 1 & \text{für } (\alpha_2 < 1 \wedge 1 \leq \alpha_1) \text{ oder } (\alpha_2 = 1 \wedge 1 < \alpha_1) \\ \text{Gleichverteilung} & \text{für } \alpha_1 = \alpha_2 = 1 \end{cases}$
- Besonderheit:  $Y_i \sim \text{gamma}(\alpha_i, \beta) \Rightarrow \frac{Y_1}{Y_1 + Y_2} \sim \text{beta}(\alpha_1, \alpha_2)$
- Erzeugung durch Komposition zweier  $\text{gamma}(\alpha_i, \beta)$  Verteilungen:  $X = \frac{Y_1}{Y_1 + Y_2}$

- Beispiele



**FIGURE 6.11**

beta( $\alpha_1, \alpha_2$ ) density functions.

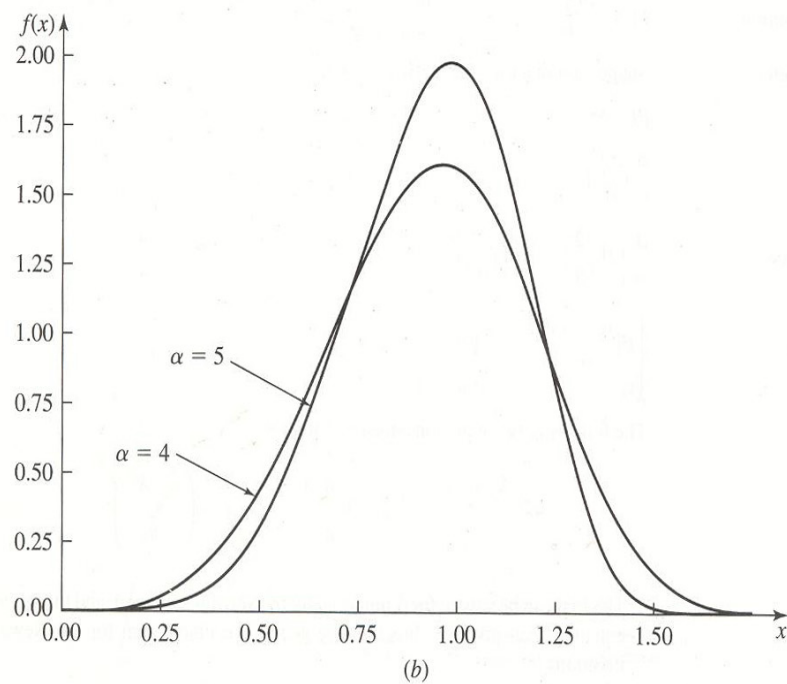
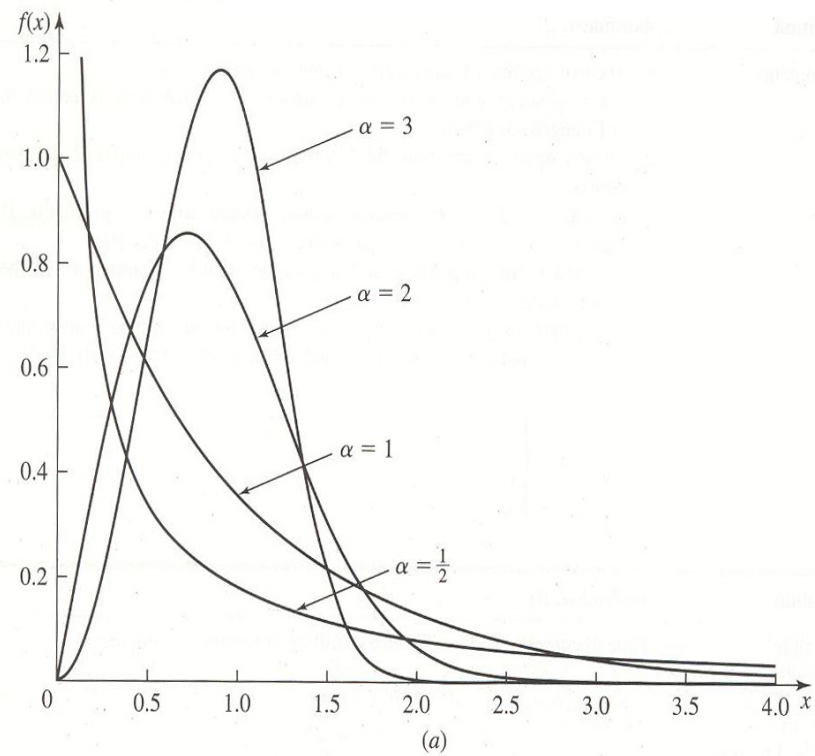
(continued)

Abbildung 3.4: Verteilungsdichtefunktion von Beta-Verteilungen (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 309)

### 10.2.6 Weibull Verteilung: $ZV \sim \text{Weibull}(\alpha, \beta)$ (LK 8.3.5)

- Nach Waloddi Weibull benannt
- VF:  $F(t) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha}$ 
  - Für  $\alpha = 1$  und  $\beta = \frac{1}{\lambda}$  erhalten wir die Exponentialverteilung
- VDF:  $f(t) = \alpha \cdot \beta^{-\alpha} \cdot t^{\alpha-1} \cdot e^{-\left(\frac{t}{\beta}\right)^\alpha}$
- Formparameter  $\alpha$  und Skalierungsparameter  $\beta$
- Range  $[0, \infty)$
- $E[X] = \frac{\beta}{\alpha} \cdot \Gamma\left(\frac{1}{\alpha}\right)$
- $VAR[X] = \frac{\beta^2}{\alpha} \cdot \left\{ 2 \cdot \Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \cdot \left[ \Gamma\left(\frac{1}{\alpha}\right) \right]^2 \right\}$
- $\text{Mode} \begin{cases} \beta \cdot \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} & \alpha \geq 1 \\ 0 & \alpha < 1 \end{cases}$
- $X \sim \text{Exp}(1) \Rightarrow Y = \beta \cdot {}^\alpha\sqrt{X} \sim \text{Weibull}(\alpha, \beta)$ 
  - Beweis:  $F_Y(t) = P(Y \leq t) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha} = P\left(X \leq \left(\frac{t}{\beta}\right)^\alpha\right) = P(\beta \cdot {}^\alpha\sqrt{X} \leq t)$
- Erzeugung von Weibull-ZV durch Inversion:  $U \sim U(0,1), Y = \beta \cdot {}^\alpha\sqrt{-\ln U}$
- Vergleich mit Gamma-Verteilung
  - Dieselben Grenzwerte für  $t \rightarrow 0$  bei gegebenem  $\alpha \Rightarrow$  ähnliches Aussehen
  - Kann auch beliebige Erwartungswerte und Variationskoeffizienten modellieren
  - Variationskoeffizient auch nur von  $\alpha$  abhängig (muss so sein, denn  $\beta$  ist nur Skalierungsparameter)
  - Anderes Verhalten hinsichtlich Modellierung von Raten, was im Folgenden ausgeführt wird.

- Beispiele



**FIGURE 6.8**  
Weibull( $\alpha, 1$ ) density functions.

**Abbildung 3: Verteilungsdichtefunktion von Weibull-Verteilungen (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 304)**

### 10.2.7 (Relative Ausfall-)Raten

- Zeitabhängige (relative Ausfall-)Rate=Anzahl von Ausfällen innerhalb einer Zeiteinheit bezogen auf eine Restmenge von Elementen nach Zeit  $t$
- Beispiel (Quelle: <http://de.wikipedia.org/wiki/Ausfallrate>)
  - 10000 Glühbirnen werden gemessen. Am 19. Tag blieben noch 9600 Birnen übrig und an diesem Tag fielen 5 Glühbirnen aus.
  - Zeitabhängige Ausfallrate =  $\frac{5/9600}{24 \text{ h}} = \frac{21,7}{10^6 \text{ h}} = 2,17 \cdot 10^{-5} \cdot \frac{1}{\text{h}}$
  - Diese Rate ist nur typisch für Tag 19. Man kann nicht einfach den Kehrwert bilden (46080 h) und davon ausgehen, dass eine Lampe im Durchschnitt solange brennt, weil die Ausfallrate mit zunehmender Betriebszeit steigt.

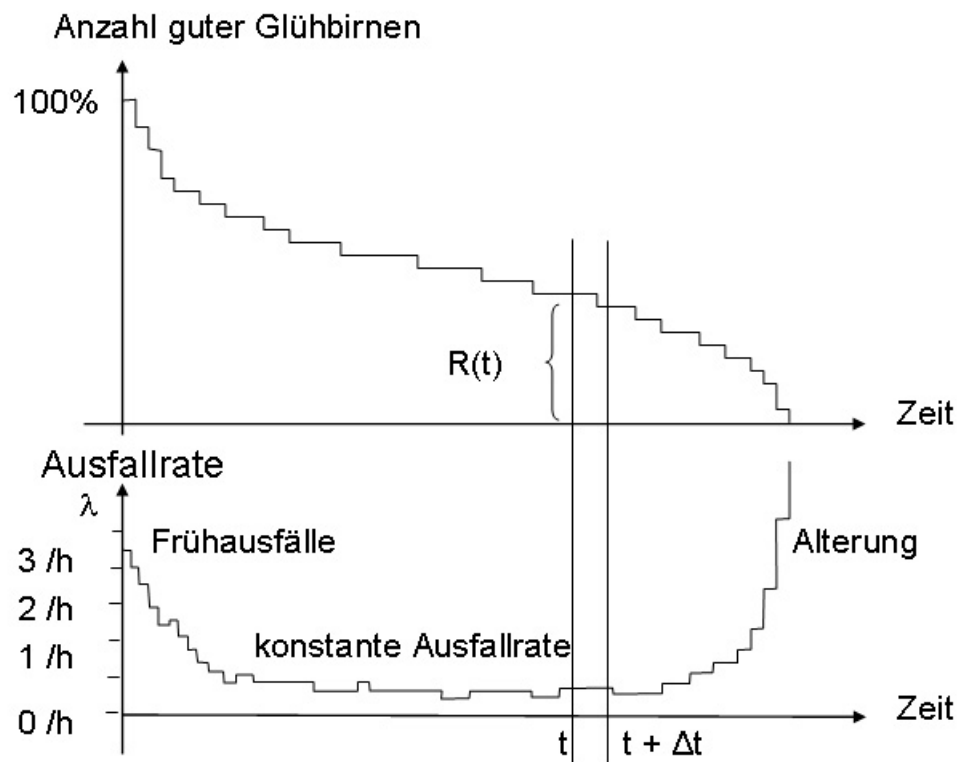


Abbildung 4: Anzahl der Glühbirnen und Ausfallrate in Abhängigkeit der Zeit. (Wikipedia)

- Gegeben: Verteilungs- und Verteilungsdichtefunktion  $F(t)$  und  $f(t)$  der Zeit bis zu einem Ausfall
- Die Überlebensfunktion beschreibt die Wahrscheinlichkeit, dass nach  $t$  Zeit noch kein Ausfall eingetreten ist:  $R(t) = F^c(t) = 1 - F(t)$

- Mathematische Definition einer (relativen Ausfall-)Rate
  - Hilfskonzept: Wahrscheinlichkeit nach t Zeit ein Ereignis innerhalb eines kleinen Intervalls der Länge  $\Delta t$  zu haben:  $\int_t^{t+\Delta t} f(\tau) d\tau$

- Die zeitabhängige (relative Ausfall-)Rate  $h(t)$  ist der Grenzwert

$$h(t) = \lim_{\Delta t \rightarrow 0} (h(t, \Delta t)) = \frac{\left( \frac{\int_t^{t+\Delta t} f(\tau) d\tau}{\Delta t} \right)}{R(t)} = \frac{f(t)}{1-F(t)},$$

falls  $f(t)$  stetig und beschränkt.

- Beispiele
  - Die Exponentialverteilung als Überlebensfunktion führt zu einer konstanten (relativen Ausfall-)Rate:  $h(t) = \frac{f(t)}{1-F(t)} = \frac{\lambda \cdot e^{-\lambda \cdot t}}{e^{-\lambda \cdot t}} = \lambda$

- Die Weibull-Verteilung als Überlebensfunktion führt zu einer mit der

$$\text{Zeit } t \text{ wachsende Rate: } h(t) = \frac{f(t)}{1-F(t)} = \frac{\alpha \cdot \beta^{-\alpha} \cdot t^{\alpha-1} \cdot e^{-\left(\frac{t}{\beta}\right)^\alpha}}{e^{-\left(\frac{t}{\beta}\right)^\alpha}} = \alpha \cdot \beta^{-\alpha} \cdot t^{\alpha-1},$$

welche zur Modellierung von Fehlerraten besser geeignet ist als eine konstante Fehlerrate.

- Allgemein

$$\blacksquare f(t) = g(t) \cdot e^{G(t)} \text{ mit } G(t) = \int_{-\infty}^t g(\tau) d\tau \Rightarrow F(t) = 1 - e^{G(t)}$$

$$\blacksquare h(t) = \frac{f(t)}{1-F(t)} = \frac{g(t) \cdot e^{G(t)}}{e^{G(t)}} = g(t)$$

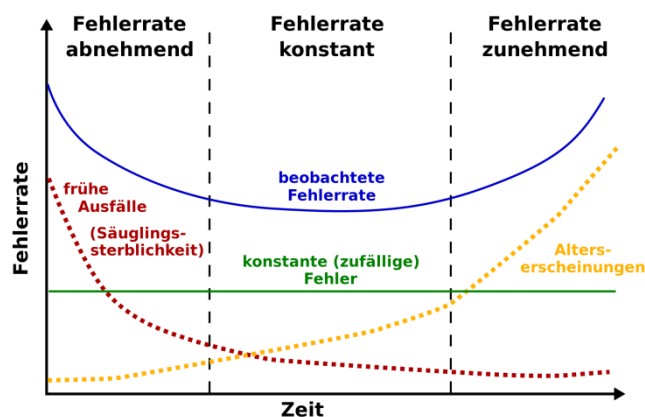


Abbildung 5: Realistische zeitabhängige Fehlerratenkurve. (Quelle: Wikipedia)

### 10.2.8 Pareto-Verteilung: ZV $X \sim \text{Par}(k, x_{\min})$

- Quelle: Wikipedia: <http://de.wikipedia.org/wiki/Pareto-Verteilung>
- Benannt nach Vilfredo Pareto, hat aber nichts mit Pareto-Optimum zu tun
- VF:  $F(x) = 1 - \left(\frac{x_{\min}}{x}\right)^k, x \geq x_{\min}$
- Range:  $[x_{\min}, \infty)$
- VDF:  $f(x) = \frac{k}{x_{\min}} \cdot \left(\frac{x_{\min}}{x}\right)^{k+1}, x \geq x_{\min}$

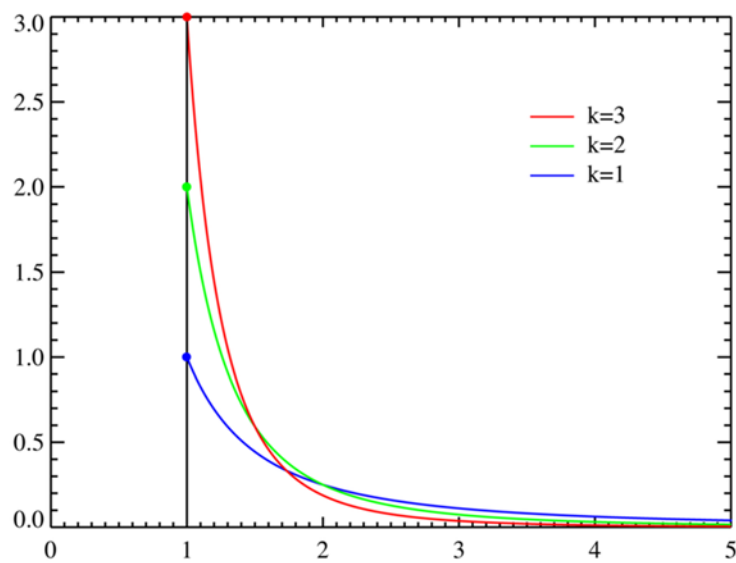


Abbildung 10-6: VDF der Pareto-Verteilung mit  $x_{\min}=1$ .

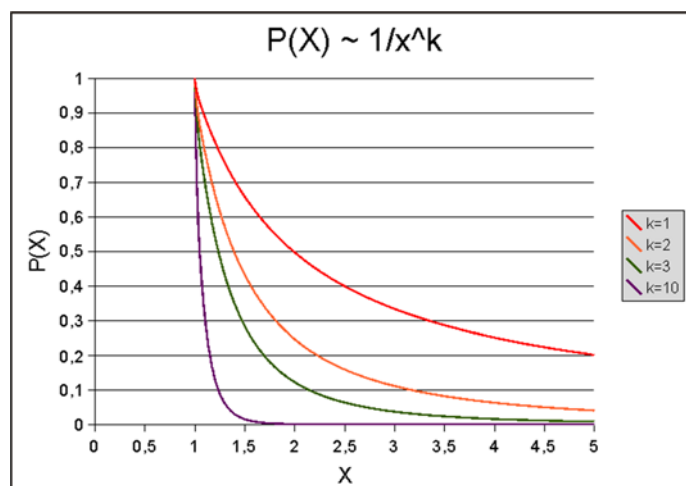


Abbildung 10-7: Komplementäre VF der Pareto-Verteilung mit  $x_{\min}=1$ .



- Mittelwert:  $E[X] = \begin{cases} x_{\min} \cdot \frac{k}{k-1} & k > 1 \\ \infty & k \leq 1 \end{cases}$
- Varianz:  $Var[X] = \begin{cases} x_{\min}^2 \cdot \left( \frac{k}{k-2} - \frac{k^2}{(k-1)^2} \right) & k > 2 \\ \infty & k \leq 2 \end{cases}$
- Variationskoeffizient:  $c_{var}[X] = \frac{1}{\sqrt{k \cdot (k-2)}}$
- Erzeugung durch Inversion:  $X = \frac{x_{\min}}{\sqrt[k]{U}}$
- Für  $x_{\min}=1$  gilt:  $\ln(X) = \ln\left(\frac{1}{\sqrt[k]{U}}\right) = -\frac{1}{k} \ln(U)$

Somit gilt:  $X \sim \text{Par}(k, x_{\min}=1) \Rightarrow \ln(X) \sim \text{expo}(\lambda)$  mit  $\lambda=k$

- Modellierung von Dateigrößen im Internet
- Die Pareto-Verteilung beschreibt das statistische Phänomen, wenn eine kleine Anzahl von hohen Werten einer Wertemenge mehr zu deren Gesamtwert beiträgt, als die hohe Anzahl der kleinen Werte dieser Menge.
- Bsp.: Pareto untersuchte die Verteilung des Volksvermögens in Italien und fand, dass ca. 80% des Vermögens bei ca. 20% der Familien konzentriert war. Banken sollten sich also vornehmlich um diese 20% der Menschen kümmern und ein Großteil ihrer Auftragslage wäre gesichert.
- Daraus leitet sich das **Pareto-Prinzip** ab, auch „80-zu-20-Regel“, „80-20-Verteilung“ oder „Pareto-Effekt“ genannt. Es besagt, dass sich viele Aufgaben mit einem Mitteleinsatz von ca. 20% zu 80% erledigen lassen. Beispielsweise bedient ein Versandhandel ca. 80% der Kunden mit einem Arbeitsaufwand von ca. 20%. Die weitaus meiste Arbeit verursachen die wenigen Kunden, die sich beschweren und Lieferungen beanstanden.

- Weiteres Beispiel: Verteilung von Städtegrößen in Deutschland

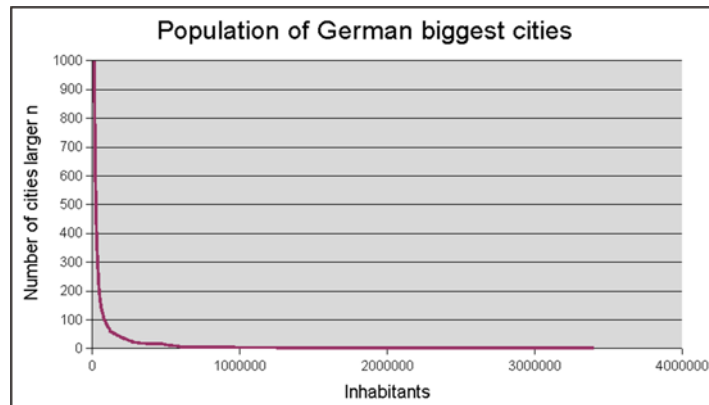


Abbildung 3-7: Anzahl von Städten in Deutschland, die eine bestimmte Bevölkerungszahl überschreiten (lin-lin-scale).

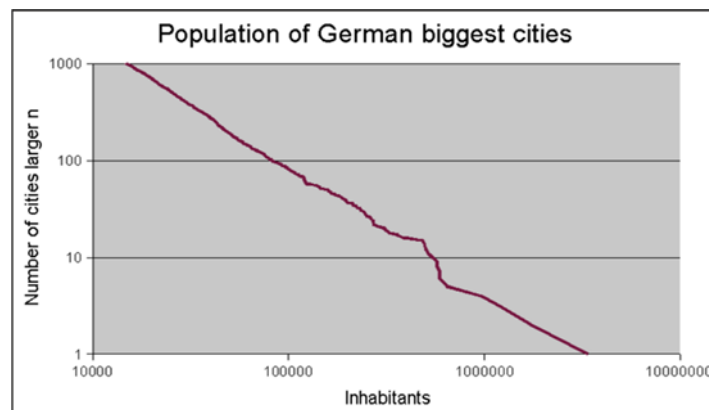
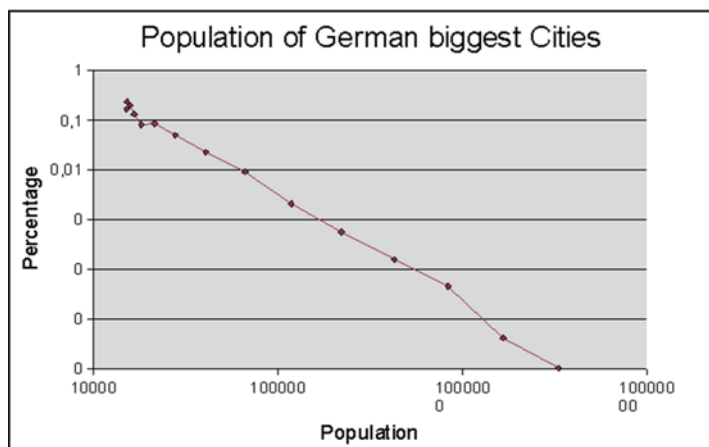


Abbildung 10-8: Anzahl von Städten in Deutschland, die eine bestimmte Bevölkerungszahl überschreiten (log-log-scale).



### 10.2.9 Lognormal-Verteilung: ZV $X \sim \text{LN}(\mu, \sigma^2)$ (LK 8.3.7)

- Relation zu  $N(\mu, \sigma^2)$ :  $Y \sim N(\mu, \sigma^2) \Rightarrow e^Y \sim \text{LN}(\mu, \sigma^2)$ .
- Erzeugung durch Komposition:  $Y \sim N(\mu, \sigma^2) \Rightarrow X = e^Y$ .
- Range:  $[0, \infty)$
- Achtung:  $\mu$  und  $\sigma^2$  sind nicht Mittelwert und Varianz von  $\text{LN}(\mu, \sigma^2)$ !
- $E[X] = e^{\mu + \sigma^2/2}$
- $\text{VAR}[X] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$
- Wahl der geeigneten Parameter für die erzeugende Normalverteilung:

$$\mu = E[Y] = \ln\left(\frac{E[X]^2}{\sqrt{E[X]^2 + \text{VAR}[X]}}\right) \quad \sigma^2 = \text{VAR}[Y] = \ln\left(\frac{\text{VAR}[X] + E[X]^2}{E[X]^2}\right)$$

- Beispiel

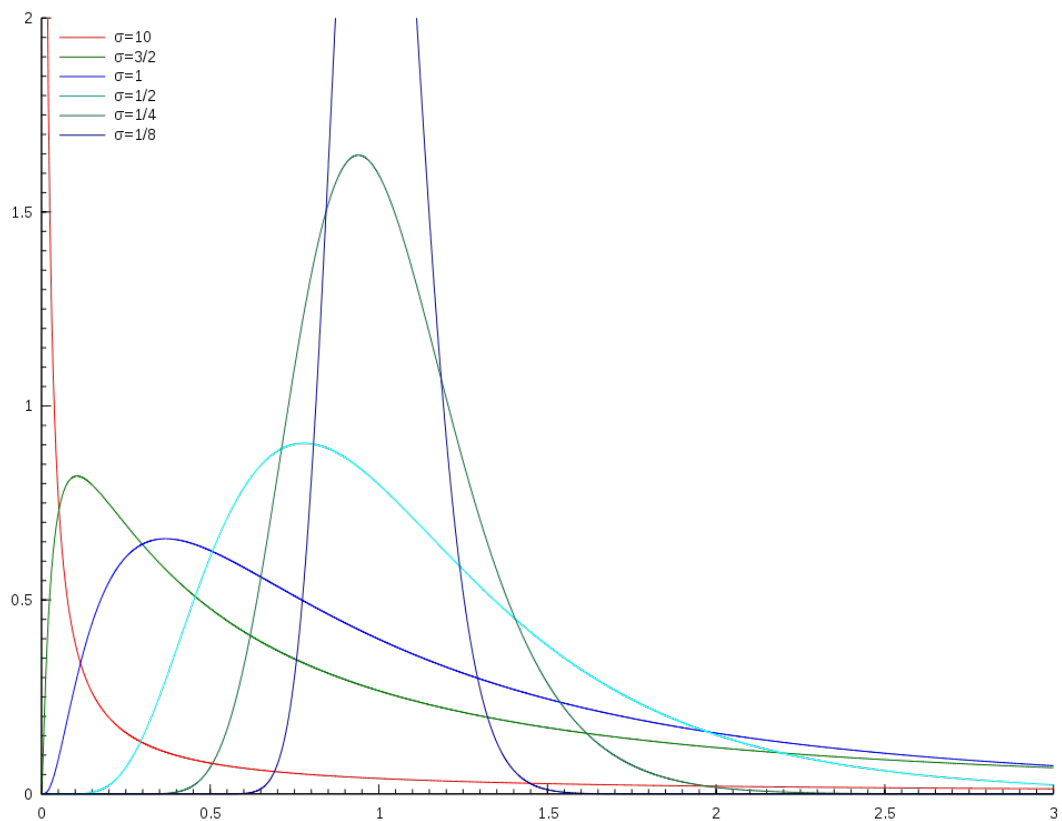


Abbildung 8: Dichtefunktionen von Lognormalverteilungen mit  $\mu = 0$ .

## 10.3 Diskrete Verteilungen (LK 6.2.3 und 8.4)

### 10.3.1 Hypergeometrische Verteilung $h(k_0|n;n_0;k)$

- Quelle: Feller: "An Introduction to Probability Theory and Its Applications", volume I, 3rd edition, p. 233.
- „Ziehen ohne Zurücklegen“
- Urne mit  $n_0$  roten und  $n_1$  schwarzen Kugeln,  $n=n_0+n_1$
- Anzahl gezogener Kugeln:  $k$
- Anzahl der roten Kugeln in der Stichprobe:  $k_0$

- Verteilung  $P(X = k_0) = \frac{\binom{n_0}{k_0} \cdot \binom{n_1}{k-k_0}}{\binom{n}{k}} = \frac{\binom{n_0}{k_0} \cdot \binom{n-n_0}{k-k_0}}{\binom{n}{k}} = \frac{\binom{k}{k_0} \cdot \binom{n-k}{n_0-k_0}}{\binom{n}{n_0}}$

- $E[X] = \sum_{0 \leq k_0 \leq k} k_0 \cdot \frac{\binom{n_0}{k_0} \cdot \binom{n_1}{k-k_0}}{\binom{n}{k}} = k \cdot \frac{n_0}{n}$

- $VAR[X] = \frac{k \cdot n_0 \cdot (n-n_0)}{n^2} \cdot \left(1 - \frac{k-1}{n-1}\right)$

- Vergleich mit „Ziehen mit Zurücklegen“
  - Anzahl roter Kugeln  $k_0$  entspricht einer Binomialverteilung mit  $\text{Binom}(k, n_0/n)$
  - $E[X] = k \cdot n_0/n$
  - $VAR[X] = \frac{k \cdot n_0 \cdot (n-n_0)}{n^2}$

### 10.3.2 Erzeugung der negativ-binomialen Verteilung mit realem $s$ durch die Gamma-Verteilung (LK 8.4.6)

- NegBin-Verteilung:  $p(k) = \frac{\Gamma(s+k)}{\Gamma(s) \cdot k!} \cdot p^s \cdot (1-p)^k$
- VDF der Gamma-Verteilung:  $f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\beta^{-\alpha} \cdot x^{\alpha-1} \cdot e^{-x/\beta}}{\Gamma(\alpha)} & \text{if } 0 < x \end{cases}$
- Gamma-Funktion:  $\Gamma(z) = \begin{cases} \int_0^\infty t^{z-1} \cdot e^{-t} dt & \text{if } 0 \leq x \\ 0 & \text{if } x < 0 \end{cases}$
- Die Negbin-Verteilung  $\text{Negbin}(s, p)$  kann als Kombination einer  $\text{Poisson}(y)$ -Verteilungen dargestellt werden, wobei  $y$  nach  $\text{Gamma}(s, (1-p)/p)$  verteilt ist.
- $\text{NegBin}(k|s, p) = \int_0^\infty \text{Poisson}(k|y) \cdot \text{Gamma}\left(y|s, \frac{1-p}{p}\right) dy =$

$$\begin{aligned} & \int_0^\infty \frac{y^k}{k!} \cdot e^{-y} \cdot \frac{\left(\frac{1-p}{p}\right)^{-s} \cdot y^{s-1} \cdot e^{-y \cdot \frac{p}{1-p}}}{\Gamma(s)} dy = \\ & \frac{\left(\frac{1-p}{p}\right)^{-s}}{k! \cdot \Gamma(s)} \cdot \int_0^\infty y^{k+s-1} \cdot e^{-\frac{y}{1-p}} dy = \\ & \frac{\left(\frac{1-p}{p}\right)^{-s}}{k! \cdot \Gamma(s)} \cdot \Gamma(s+k) \cdot (1-p)^{s+k} = \\ & \frac{\Gamma(s+k)}{k! \cdot \Gamma(s)} \cdot (1-p)^k \cdot p^s \end{aligned}$$

- Daraus ergibt sich eine weitere Erzeugungsregel, die auch für reelle  $s$  Bestand hat:
  1. Erzeuge  $y \sim \text{Gamma}(s, (1-p)/p)$  (Accept-Reject Methode)
  2. Erzeuge  $\text{Negbin}(s, p) \sim \text{Poisson}(y)$  (Konstruktionsmethode)
 In der Praxis werden effizientere Algorithmen eingesetzt.

### 10.3.3 Zipf-Verteilung

- Quelle: Wikipedia: [http://de.wikipedia.org/wiki/Zipfsches\\_Gesetz](http://de.wikipedia.org/wiki/Zipfsches_Gesetz)
- Entstehung: die Zipf-Verteilung beschrieb ursprünglich die Wahrscheinlichkeit der häufigsten Worte in einem Sprachkorporus und wurde nach dem Harvard-Linguisten George Kingsley Zipf benannt. Der Sprachkorporus umfasste N unterschiedliche Worte. Diese wurden nach ihrer Häufigkeit sortiert und die Position k eines Wortes in dieser Ordnung wurde mit Rang ( $1 \leq k \leq N$ ) bezeichnet. Die Häufigkeit des Wortes vom Rang k ist Zipf-verteilt.
- Diskretes Gegenstück zur kontinuierlichen Pareto-Verteilung
- Unterschiede: Zipf ist ganzzahlig, endlich, und zeigt das Potenzgesetz für die Verteilung, nicht für die Verteilungsfunktion.
- Wertebereich:  $1 \leq k \leq N$
- $P(X = k | N, s) = \frac{1/k^s}{H_{N,s}}$  (Verteilung)
- $H_{N,s} = \sum_{1 \leq n \leq N} 1/n^s$  ist die harmonische Zahl der Ordnung N von s und wird zur Normierung benutzt.
- $P(X \leq k | N, s) = \frac{H_{k,s}}{H_{N,s}}$  (Verteilungsfunktion)
- Momente:  $E[X^k] = \frac{H_{N,s-k}}{H_{N,s}}$
- Es kann  $N=\infty$  sein.  $H_{\infty,s}$  wird auch die Riemannsche Zeta-Funktion genannt und die entsprechende Verteilung heißt dann auch manchmal Zeta-Verteilung.
- Die Zipf-verteilte ZVs müssen nicht notwendigerweise die Auftrittswahrscheinlichkeit von  $X=k$  beschreiben, sie können auch die Auftrittswahrscheinlichkeiten von gegebenen Elementen  $x_k$  beschreiben.
- Weitere Anwendung: Modellierung von Flussgrößen im Internet
- „power law“ (Potenzgesetz) Verteilung erkennt man im loglog-plot; im Gegensatz zur Pareto-Verteilung wird hier aber die Verteilung und nicht die Verteilungsfunktion betrachtet.

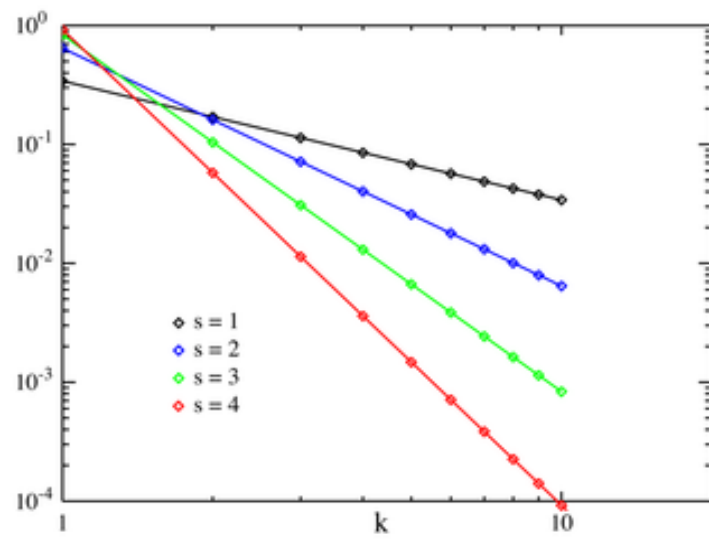


Abbildung 10-9: Zipf-Verteilung

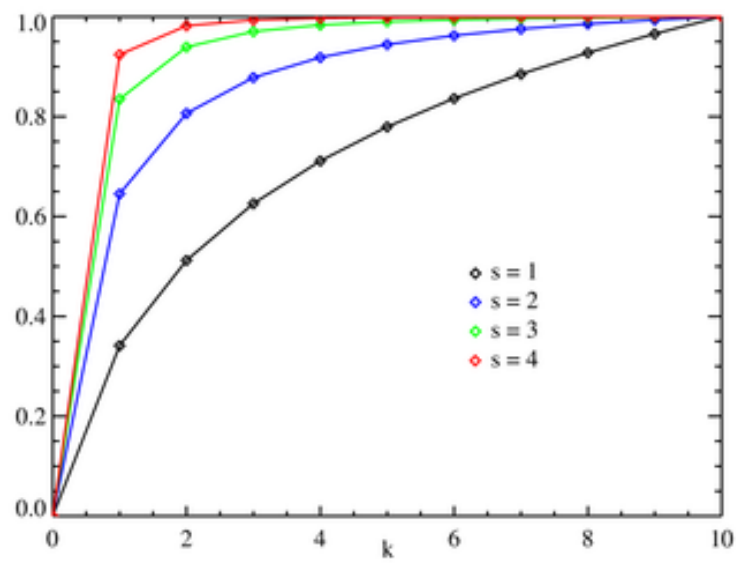


Abbildung 10-10: Zipf-Verteilungsfunktion

## 10.4 Kombinierte Verteilungen

Manchmal gibt es bei der Modellierung von Systemen Randbedingungen zu beachten.

- Untere Verteilungsgrenze
  - Problem: Diskrete Zwischenankunfts- und Bedienzeiten sind niemals Null
  - Lösung: Kombination aus deterministischer Verteilung und einer anderen:  $X=Y+Z$   
 Ziel:  $E[X]$ ,  $VAR[X]$ ,  $X \geq c$   
 konstante Verteilung:  $Y=c \Rightarrow E[Y]=c$ ,  $VAR[Y]=0$   
 Anforderungen an  $Z=X-Y$ 
    - $E[Z]=E[X]-c$
    - $VAR[Z]=VAR[X]-0$
    - Beispiel:  $E[X]=10$ ,  $VAR[X]=6$ ,  $X \geq 5$ , unendlicher Wertebereich
      - $Y=5$ ,  $E[Z]=5$ ,  $VAR[Z]=6$
      - $Z \sim \text{NegBin}(25, 5/6)$  wegen  $p=E/VAR$ ,  $s=E \cdot p / (1-p)$
- Untere und obere Verteilungsgrenze
  - Problem: Technische Größen haben Minimal- und Maximalwerte
  - Lösungen
    - Nimm Verteilungen mit endlichem Wertebereich, z.B. Binomialverteilung, Gleichverteilung, Beta-Verteilung
    - Beispiel:  $E[X]=10$ ,  $VAR[X]=4$ ,  $X \geq 5$ , endlicher Wertebereich
      - $Y=5$ ,  $E[Z]=5$ ,  $VAR[Z]=4$
      - $Z \sim \text{Binom}(25, 1/5)$  wegen  $p=1-VAR/E$ ,  $n=E/p$
    - Schneide Verteilungen mit unendlichem Wertebereich ab und normalisiere VDF zur Berechnung der Momente ( $\cong$  verwerfen aller  $X > c$ )



- Zusammengesetzte Verteilungen (composite distribution)
  - Diskrete ZV  $N$  (beschreibt Gruppengröße)
  - Diskrete oder kontinuierliche iid ZVn  $X_i$
  - Zusammengesetzte ZV  $Y = \sum_{0 < i \leq N} X_i$
  - Beispiel
    - Gruppenbearbeitungsprozess mit einer Bearbeitungszeit eines einzelnen Elements von  $X \sim \text{Expo}(\lambda)$
    - Falls Gruppengröße  $N$  fest: Modellierung der Gruppenbedienzeit mit  $Y \sim k\text{-Erlang}(k, \lambda)$
    - Falls Gruppengröße  $N$  variabel: Modellierung der Gruppenbedienzeit  $Y$  nur mit zusammengesetzter Verteilung
  - Analytische Eigenschaften (allgemein)
    - $E[Y] = E[N] \cdot E[X]$  (ist intuitiv) (2.27e)
    - $\text{VAR}[Y] = E[N] \cdot \text{VAR}[X] + \text{VAR}[N] \cdot E[X]^2$  (2.27f)
    - Beweis siehe Tran-Gia, Übungsaufgabe 1.9
    - Falls  $N$  konstant, dann ist  $\text{VAR}[0] = 0!$

## 10.5 Übersicht behandelter Verteilungen

### 10.5.1 Kontinuierliche Verteilungen

1. Gleichverteilung
2. Dreiecksverteilung
3. Empirische Verteilung basierend auf gewichteten Gleichverteilungen
  - a. Empirische Verteilung basierend auf einzelnen Stichproben
4. Exponentielle Verteilung ( $\lambda$ ): Zeit zwischen Anrufen
5. Erlang-k-Verteilung ( $k, \lambda$ ): Summe von  $k$  exponentiellen Phasen
6. Hyperexponentielle Verteilung: höhere Varianz
7. Gamma-Verteilung ( $\alpha, \beta$ ): Erweiterung der Erlang-k-Verteilung auf  $k \in \mathbb{R}^+$  sowie Generalisierung von exponentieller, Erlang-k, und Hyperexponentieller Verteilung
8. Beta-Verteilung ( $\alpha_1, \alpha_2$ ): Verteilung mit unterschiedlichen VDF-Formen auf beschränktem Intervall
9. Weibull-Verteilung( $k, \lambda$ ): entspricht exponentieller Verteilung mit zeitabhängiger Rate; Modellierung von zeitabhängigen Fehlerraten
10. Pareto-Verteilung( $x_{\min}, k$ ): modelliert oft empirische Verteilungen recht gut (Städtegrößen, ...); Potenzgesetz
11. Normalverteilung ( $\mu, \sigma^2$ ): Summe sehr vieler iid Zufallsvariablen
12. Lognormal-Verteilung ( $\mu, \sigma^2$ ): Anwendung im Mobilfunk

### 10.5.2 Diskrete Verteilungen

1. Diskrete Gleichverteilung
2. Beliebige diskrete Verteilung
3. Bernoulli-Verteilung  $\text{Bernoulli}(p)$ : Münzwurf
4. Binomial-Verteilung  $\text{Binom}(n,p)$ : Wiederholter Münzwurf; Ziehen von Kugeln mit Zurücklegen; Ankünfte mit geometrisch verteilten Zwischenankunftszeiten  $\text{Geom}(p)$  in einem Intervall ganzzahliger Länge  $n$
5. Geometrische Verteilung  $\text{Geom}(0)(p)$ : Anzahl von Misserfolgen bis beim wiederholten Münzwurf zum ersten Mal gewonnen wurde; ganzzahliges Gegenstück zur exponentiellen Verteilung
6. Um 1 verschobene geometrische Verteilung  $\text{Geom}(1)(p)$ : Anzahl aller Versuche bis beim wiederholten Münzwurf zum ersten Mal gewonnen wurde
7. Negativ-binomiale Verteilung  $\text{NegBin}(s,p)$ : Wiederholter Münzwurf: Anzahl von Fehlversuchen bis  $s$  Mal gewonnen wurde;
8. Erweiterung der negativ-binomialen Verteilung  $\text{NegBin}(s,p)$  auf reelle Werte von  $s$
9. Poisson-Verteilung  $\text{Poisson}(y)$ : Anzahl von Ankünften mit exponentiell verteilten Zwischenankunftszeiten in einem Intervall fester Länge
10. Hypergeometrische Verteilung  $h(k_0 | n; n_0; k)$ : Ziehen von Kugeln ohne Zurücklegen
11. Zipf-Verteilung: modelliert oft empirische Verteilungen recht gut (Flussgrößen im Internet, ...); ganzzahliges Gegenstück zur kontinuierlichen Pareto-Verteilung