Inhalt von Abschnitt 4

Eine Variable oder ein Merkmal X heißt stetig, falls zu zwei Werten $a < b$ auch jeder Zwischenwert im Intervall $[a, b]$ möglich ist.
Falls die Werte von X als Ergebnisse eines Zufallsvorgangs resultieren, wird X zu einer stetigen Zufallsvariable.

Stetige Zufallsvariablen und Dichten

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

$$P(a \leq X \leq b) = \int_a^b f(x) \mathrm{d}x$$

gilt. Die Funktion $f(x)$ heißt (Wahrscheinlichkeits-)Dichte von X.

Wahrscheinlichkeiten stetiger Zufallsvariablen

Für stetige Zufallsvariablen X gilt

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

und

$$P(X = x) = 0 \quad \text{für jedes } x \in \mathbb{R}.$$

Die Dichte $f(x)$ besitzt einige Analogien zur Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariable. Deshalb wird auch die gleiche Bezeichnung verwendet.

Ein wesentlicher Unterschied ist jedoch, dass die Werte $f(x)$ einer stetigen Dichte keine Wahrscheinlichkeiten sind. Somit können Dichten auch Werte $f(x) > 1$ annehmen.

Verteilungsfunktion einer stetigen Zufallsvariable

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

Eigenschaften der Verteilungsfunktion stetiger ZVn

1. $F(x)$ ist stetig und monoton wachsend mit Werten im Intervall $[0,1]$.
2. Für die Grenzen gilt

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0,$$
$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

3. Für Werte von x, an denen $f(x)$ stetig ist, gilt

$$F'(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x),$$

d.h. die Dichte ist die Ableitung der Verteilungsfunktion.

4. Für Intervalle erhält man

$$P(a \leq X \leq b) = F(b) - F(a),$$
$$P(X \geq a) = 1 - F(a).$$

Stetige Gleichverteilung

Definition 4.1 (Stetige Gleichverteilung)

Eine stetige Zufallsvariable heißt gleichverteilt auf dem Intervall $[a, b]$, wenn sie eine *Dichte*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

besitzt. Eine auf $[0, 1]$ gleichverteilte Zufallsvariable nennt man auch standardgleichverteilt.

Stetige Gleichverteilung (Forts.)

Die Verteilungsfunktion ergibt sich dann für $a \leq x \leq b$ zu

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{a}^{x} \frac{1}{b-a}\mathrm{d}t = \frac{x-a}{b-a}.$$

$$\Rightarrow F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b. \end{cases}$$

Exponentialverteilung

Genese: Stetige Wartezeitverteilung bei "Prozessen ohne Gedächtnis", Grenzwert der geometrischen Verteilung.

Definition 4.2 (Exponentialverteilung)

Eine positive ZV X ist *exponential*verteilt mit Parameter λ $(X \sim Ex(\lambda))$, wenn sie die Dichte

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0 \end{cases}$$

hat. Die Verteilungsfunktion ist

$$F(x) = 1 - e^{-\lambda x},\, x \geq 0.$$

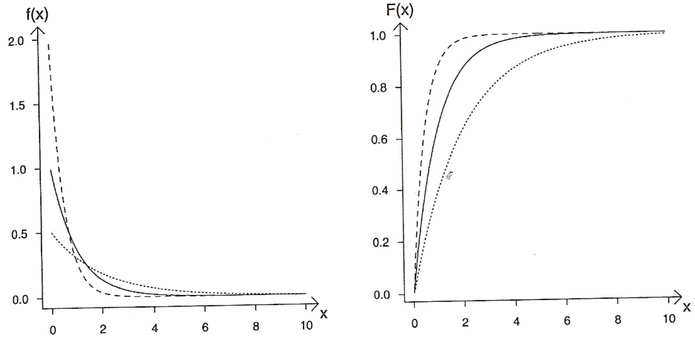


ABBILDUNG 6.8: Dichte und Verteilungsfunktion der Exponentialverteilung für $\lambda = 0.5$ (····), $\lambda = 1.0$ (——) und $\lambda = 2.0$ (– – –)

Figure: aus Fahrmeir et al.

Definition 4.3 (Unabhängigkeit von stetigen Zufallsvariablen)

Zwei stetige Zufallsvariablen X und Y sind unabhängig, wenn für alle $x \in \mathbb{R}$ und $y \in \mathbb{R}$

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

gilt. Dabei ist F_X bzw. F_Y die Verteilungsfunktion von X bzw. Y.

Allgemeiner sind die stetigen Zufallsvariablen $X_1, \ldots, X_n$ unabhängig, wenn für $x_1, \ldots, x_n \in \mathbb{R}$

$$P(X \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n)$$

gilt.

Es lässt sich – wiederum in Analogie zum diskreten Fall – zeigen, dass aus dieser Definition auch die Unabhängigkeit von allgemeineren durch die Zufallsvariablen definierten Ereignissen folgt: Falls $X_1, \ldots, X_n$ unabhängig sind, so gilt für beliebige zulässige Ereignisse $A_1, \ldots, A_n$, insbesondere für Intervalle $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n).$$

Erwartungswert

Definition 4.4 (Erwartungswert)

Der *Erwartungswert* $\mathbb{E}(X)$ einer stetigen Zufallsvariable X mit Dichte $f(x)$ ist

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) \mathrm{d}x.$$

Eigenschaften von Erwartungswerten

1. Transformationen:
   Sei $g(x)$ eine reelle Funktion. Dann gilt für $Y = g(X)$

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x.$$

2. Lineare Transformationen:
   Für $Y = aX + b$ ist

$$\mathbb{E}(Y) = \mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b.$$

Eigenschaften von Erwartungswerten (Forts.)

3. Symmetrische Verteilungen:
   Ist die Dichte $f(x)$ symmetrisch um den Punkt c, d.h. ist $f(c-x) = f(x+c)$ für alle x, so gilt

$$\mathbb{E}(X) = c.$$

4. Additivität:
   Für zwei Zufallsvariablen X und Y ist

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

5. Allgemeiner gilt mit beliebigen Konstanten $a_1, \ldots, a_n$

$$\mathbb{E}(a_1 X_1 + \ldots + a_n X_n) = a_1 \, \mathbb{E}(X_1) + \ldots + a_n \, \mathbb{E}(X_n).$$

Beispiel 4.5 (Stetige Gleichverteilung)

Sei X auf $[a, b]$ gleichverteilt. Dann ist

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) \mathrm{d}x = \int_a^b x \frac{1}{b-a} \mathrm{d}x$$
$$= \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{(b-a)(b+a)}{2(b-a)},$$

also
$$\mathbb{E}(X) = \frac{a+b}{2}.$$

Beispiel 4.6 (Exponentialverteilung)

Für eine exponentialverteilte Zufallsvariable $X \sim Ex(\lambda)$ ist

$$
\begin{aligned}
\mathbb{E}(X) &= \lambda \int_0^\infty x e^{-\lambda x} \mathrm{d}x \\
&= \left[x \cdot (-1) e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} \mathrm{d}x \\
&= 0 + \frac{1}{\lambda}, \\
\mathbb{E}(X) &= \frac{1}{\lambda}.
\end{aligned}
$$

Dabei wurde partielle Integration benutzt.

Modus, Median, Quantile

Definition 4.7 (Modus)

Jeder x-Wert, für den $f(x)$ ein Maximum besitzt, ist Modus, kurz x_{mod}. Falls das Maximum eindeutig ist und $f(x)$ keine weiteren lokalen Maxima besitzt, heißt $f(x)$ unimodal.

Definition 4.8 (Median und Quantile)

Für $0 < p < 1$ ist das p-Quantil x_p die Zahl auf der x-Achse, für die

$$F(x_p) = p$$

gilt. Der Median x_{med} ist das 50%-Quantil, es gilt also

$$F(x_{\text{med}}) = \frac{1}{2}.$$

Für streng monotone Verteilungsfunktionen $F(x)$ sind p-Quantil und Median eindeutig bestimmt.

Varianz und Standardabweichung

Definition 4.9 (Varianz und Standardabweichung einer stetigen Zufallsvariable)

Die Varianz $\mathbb{V}(X)$ einer stetigen Zufallsvariable X mit Dichte $f(x)$ ist

$$\sigma^2 = \mathbb{V}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x$$

mit $\mu = \mathbb{E}(X)$. Die Standardabweichung ist

$$\sigma = +\sqrt{\mathbb{V}(X)}.$$

Varianz und Standardabweichung (Forts.)

- Eigenschaften von Varianzen

1. Es gilt $\mathbb{V}(X) = \mathbb{E}((X - \mu)^2)$
2. Verschiebungsregel

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mu^2$$

bzw. $$\mathbb{V}(X) = \mathbb{E}((X - c)^2) - (\mu - c)^2.$$

3. Lineare Transformationen
   Für $Y = aX + b$ ist

$$\mathbb{V}(Y) = \mathbb{V}(aX + b) = a^2\,\mathbb{V}(X),$$
$$\sigma_Y = |a|\sigma_X.$$

Varianz und Standardabweichung (Forts.)

4. Varianz der Summe von unabhängigen Zufallsvariablen
   Falls X und Y bzw. $X_1, \ldots, X_n$ unabhängig sind gilt

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

bzw. mit beliebigen Konstanten $a_1, \ldots, a_n$

$$\mathbb{V}(a_1 X_1 + \ldots + a_n X_n) = a_1^2\,\mathbb{V}(X_1) + \ldots + a_n^2\,\mathbb{V}(X_n).$$

Beispiel 4.10 (Stetige Gleichverteilung)

Es ist zunächst

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = \int_a^b x^2 \frac{1}{b-a}\mathrm{d}x = \frac{b^3 - a^3}{3(b-a)}$$

Aus dem Verschiebungssatz erhalten wir

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Die Varianz wächst also quadratisch und die Standardabweichung $\sigma = (b-a)/\sqrt{12}$ linear mit der Länge des Intervalls.

Beispiel 4.11 (Exponentialverteilung)

Es ergibt sich für $X \sim Ex(\lambda)$

$$\mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Die Varianz ist also umgekehrt proportional zu λ^2: Je größer λ ist, umso steiler geht die Dichte $f(x)$ gegen null, umso näher ist also die Verteilung bei null konzentriert und umso kleiner ist somit die Varianz.

Definition 4.12 (Standardisierung von Zufallsvariablen)

Eine Zufallsvariable X mit $\mu = \mathbb{E}(X)$ und $\sigma^2 = \mathbb{V}(X)$ kann man ähnlich wie ein Merkmal bzw. Daten standardisieren. Man geht über zu

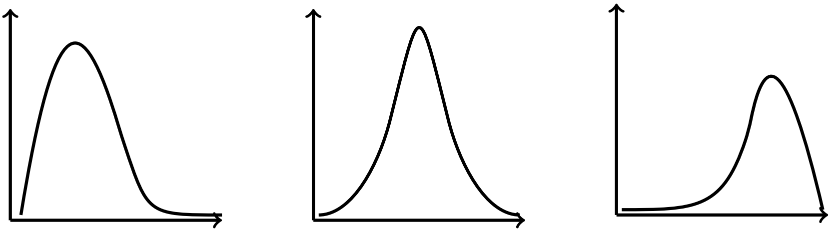
$$Z = \frac{X - \mu}{\sigma}.$$

$$\Rightarrow \quad \mathbb{E}(Z) = 0 \quad \text{und} \quad \mathbb{V}(Z) = 1$$

Symmetrie und Schiefe stetiger Verteilungen

Wie bei empirischen Verteilungen ist auch bei Verteilungen von Zufallsvariablen nach Lage und Streuung die Schiefe die nächstwichtige, die Form der Verteilung charakterisierende Eigenschaft.

Die Abbildung zeigt für stetige Zufallsvariablen die Dichten einer linkssteilen (oder rechtsschiefen), einer symmetrischen und einer rechtssteilen (oder linksschiefen) Verteilung.



Lageregel

Symmetrische unimodale Verteilung:	$x_{\text{mod}} = x_{\text{med}} = \mathbb{E}(X)$
Linkssteile Verteilung:	$x_{\text{mod}} < x_{\text{med}} < \mathbb{E}(X)$
Rechtssteile Verteilung:	$x_{\text{mod}} > x_{\text{med}} > \mathbb{E}(X)$

Normalverteilung

Definition 4.13 (Normalverteilung)

Eine Zufallsvariable X heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, kurz $X \sim N(\mu, \sigma^2)$, wenn sie die Dichte

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

besitzt. Es gilt

$$\mathbb{E}(X) = \mu, \mathbb{V}(X) = \sigma^2.$$

Speziell für $\mu = 0$, $\sigma^2 = 1$ erhält man die Standardnormalverteilung $N(0, 1)$ mit der Dichte

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

- Die Dichte ist symmetrisch zu μ, d.h. es gilt

$$f(\mu - x) = f(\mu + x), \quad x \in \mathbb{R}.$$

- Die Dichte der Normalverteilung hat Glockenform mit dem Maximum an der Stelle μ und den Wendepunkten bei $\mu \pm \sigma$.

- Die Verteilungsfunktion ist definitionsgemäß durch

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

gegeben. Dieses Integral lässt sich nicht analytisch berechnen und in geschlossener Form schreiben. Dies gilt ebenso für die Verteilungsfunktion $\Phi(x)$ der *Standardnormalverteilung*

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t.$$

Deshalb muss $\Phi(x)$ durch numerische Verfahren am Computer berechnet werden und ist für bequemeres Rechnen tabelliert.

Symmetrie-Beziehung:

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

Standardisierung

Ist X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable, so ist die standardisierte Zufallsvariable

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

standardnormalverteilt, d.h. $Z \sim N(0, 1)$.

Damit lässt sich die Verteilungsfunktion F einer $N(\mu, \sigma^2)$-verteilten Zufallsvariable durch die Verteilungsfunktion Φ der Standardnormalverteilung ausdrücken.

$$
\begin{aligned}
F(x) = P(X \leq x) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

Rückführung auf Standardnormalverteilung

$$
F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(z) \quad \text{mit } z = \frac{x-\mu}{\sigma}
$$

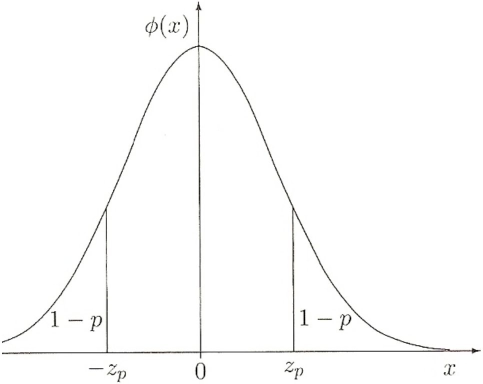
Quantile

Die Quantile z_p der Standardnormalverteilung sind durch die Gleichung

$$\Phi(z_p) = p, \quad 0 < p < 1,$$

bestimmt. Das p-Quantil z_p teilt damit die Fläche unter der Dichte $\phi(z)$ in eine Fläche mit Inhalt p links von z_p und eine Fläche mit Inhalt $1-p$ rechts davon auf. Wegen der Symmetrie gilt

$$z_p = -z_{1-p}.$$



Carl Friedrich Gauss (Germany, 1777–1855)

Wichtige Quantile der Standardnormalverteilung sind

p	50%	75%	90%	95%	97.5%	99%
z_p	0.0 (Median)	0.67	1.28	1.64	1.96	2.33

Quantile

$$z_p = \frac{x_p - \mu}{\sigma} \quad \text{bzw.} \quad x_p = \mu + \sigma z_p$$

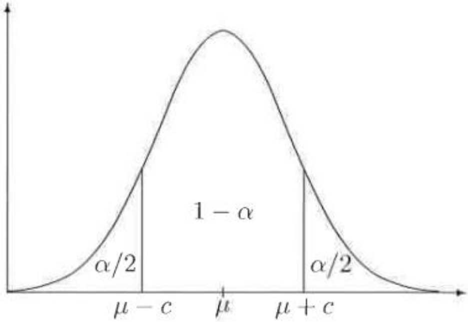


ABBILDUNG 6.16: Zentrales Schwankungsintervall

Figure: aus Fahrmeir et al.

Zentrale Schwankungsintervalle, $k\sigma$-Bereiche

Für $X \sim N(\mu, \sigma^2)$ gilt

$$P(\mu - z_{1-\alpha/2}\sigma \leq X \leq \mu + z_{1-\alpha/2}\sigma) = 1 - \alpha.$$

Für $z_{1-\alpha/2} = k$ erhält man die $k\sigma$-Bereiche

$$\begin{aligned}
k = 1: &\qquad P(\mu - \ \sigma \leq X \leq \mu + \ \sigma) = 0.6827\\
k = 2: &\qquad P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545\\
k = 3: &\qquad P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973
\end{aligned}$$

Lineare Transformation

Für $X \sim N(\mu, \sigma^2)$ ist die linear transformierte Variable $Y = aX + b$, $a \neq 0$, wieder normalverteilt mit

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Addition

Sind $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$ normalverteilt und unabhängig, so gilt

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Sind $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, unabhängig, so ist jede Linearkombination $Y = a_1X_1 + \ldots + a_nX_n$ wieder normalverteilt mit

$$Y \sim N(a_1\mu_1 + \ldots + a_n\mu_n, a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2).$$

Logarithmische Normalverteilung

Verteilungen von nichtnegativen Zufallsvariablen, etwa Lebensdauern, Wartezeiten oder Einkommen, sind häufig linkssteil. Häufig logarithmiert man X zu $Y = \ln(X)$ und hofft, dass Y zumindest annähernd normalverteilt ist.

Definition 4.14 (Logarithmische Normalverteilung)

Eine nichtnegative Zufallsvariable X heißt logarithmisch normalverteilt mit Parametern μ und σ^2, kurz $X \sim LN(\mu, \sigma^2)$, wenn $Y = \ln(X)$ $N(\mu, \sigma^2)$-verteilt ist. Es gilt

$$\mathbb{E}(X) = e^{\mu+\sigma^2/2}, \quad \mathbb{V}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

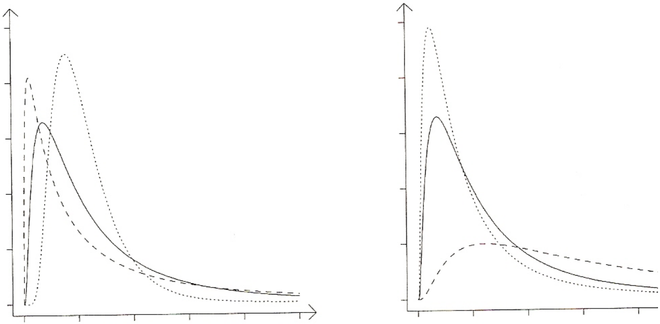


ABBILDUNG 6.17: Dichten der logarithmischen Normalverteilung für $\mu = 0$ und $\sigma^2 = 0.25$ ($\cdots$), $\sigma^2 = 1$ (——) und $\sigma^2 = 2.25$ (- - -) links bzw. $\sigma^2 = 1$ und $\mu = -0.4$ ($\cdots$), $\mu = 0$ (——) sowie $\mu = 1.2$ (- - -) rechts

Figure: aus Fahrmeir et al.

Gamma-Verteilung

Die Gamma-Verteilung kann hergeleitet werden als eine Wartezeit-Verteilung in einem Poisson-Prozess.

Definition 4.15 (Gamma-Verteilung)

Eine Zufallsvariable X heißt *Gamma-verteilt* $(X \sim \Gamma(\nu, \alpha))$, wenn die Dichte bestimmt ist durch

$$f(x) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0. \end{cases}$$

Dabei bezeichnet $\Gamma(\nu)$ die Gamma-Funktion

$$\Gamma(\nu) = \int_0^\infty t^{\nu-1} e^{-t} dt, \quad \nu > 0.$$

$\Gamma(\nu)$ ist als Verallgemeinerung der Fakultät zu verstehen. Für ganzzahlige ν gilt die Beziehung $\Gamma(\nu) = (\nu - 1)!$, für beliebiges ν die Beziehung $\Gamma(\nu + 1) = \nu \cdot \Gamma(\nu)$.
Die Erscheinungsform der Dichte ist durch den Formparameter ν bestimmt, α heißt Skalenparameter. Für $0 < \nu < 1$ fällt die Dichte monoton, für $\nu > 1$ gilt $f(0) = 0$, die Dichte besitzt ein Maximum bei $(\nu - 1)/\alpha$ und fällt wieder gegen null. Im Spezialfall $\nu = 1$ ist die Gammaverteilung identisch mit der Exponentialverteilung $E(\alpha)$.

Kennwerte

$$\mathbb{E}(X) = \frac{\nu}{\alpha}, \mathbb{V}(X) = \frac{\nu}{\alpha^2}.$$

Gelegentlich ist es günstiger, die Verteilung anders zu parametrisieren, indem neben dem Formparameter ν der Erwartungswert benutzt wird.

Erwartungswert-Parametrisierung

Man erhält mit

$$\mu = \nu/\alpha \text{ bzw. } \alpha = \nu/\mu$$

die Gamma-Verteilung $(\Gamma(\nu, \mu))$ in *Erwartungswert-Parametrisierung* mit

$$\mathbb{E}(X) = \mu \text{ und } \mathbb{V}(x) = \mu^2/\nu.$$

Chi-Quadrat-Verteilung

Viele Teststatistiken besitzen, vor allem für sogenannte Anpassungstests, unter geeigneten Voraussetzungen zumindest approximativ eine Chi-Quadrat-Verteilung. Diese lässt sich als Verteilung der Summe von unabhängigen und quadrierten standardnormalverteilten Zufallsvariablen herleiten.

Definition 4.16 (χ^2-Verteilung)

Seien $X_1, \ldots, X_n$ unabhängige und identisch $N(0,1)$-verteilte Zufallsvariablen. Dann heißt die Verteilung der Zufallsvariablen

$$Z = X_1^2 + \ldots + X_n^2$$

(zentrale) Chi-Quadrat-Verteilung mit n Freiheitsgraden, kurz $\chi^2(n)$-Verteilung, und Z heißt (zentral) $\chi^2(n)$-verteilt, kurz $Z \sim \chi^2(n)$. Es gilt

$$\mathbb{E}(Z) = n, \mathbb{V}(Z) = 2n.$$

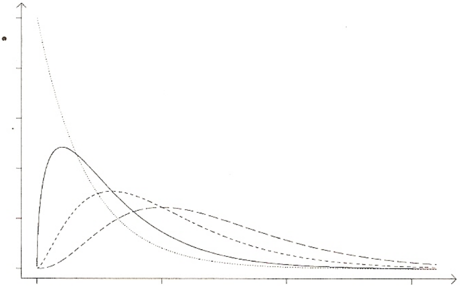


ABBILDUNG 6.18: Dichten von χ^2-Verteilungen für $n = 2$ $(\cdots)$, $n = 3$ (—) $n = 5$ (- - -) und $n = 7$ (– –) Freiheitsgrade

Student-Verteilung

Diese Verteilung findet besonders bei Parametertests und bei Konfidenzintervallen für Parameter Verwendung. Sie wird häufig auch als Students t-Verteilung oder kurz t-Verteilung bezeichnet.

Definition 4.17 (t-Verteilung, Student-Verteilung)

Seien $X \sim N(0,1)$, $Z \sim \chi^2(n)$ sowie X und Z unabhängig. Dann heißt die Verteilung der Zufallsvariable

$$T = \frac{X}{\sqrt{Z/n}}$$

(zentrale) t-Verteilung mit n Freiheitsgraden, kurz $t(n)$-Verteilung. Die Zufallsvariable T heißt (zentral) $t(n)$-verteilt, kurz $T \sim t(n)$. Es gilt

$$\mathbb{E}(T) = 0 \quad (n \geq 2), \quad \mathbb{V}(T) = \frac{n}{n-2} \quad (n \geq 3).$$

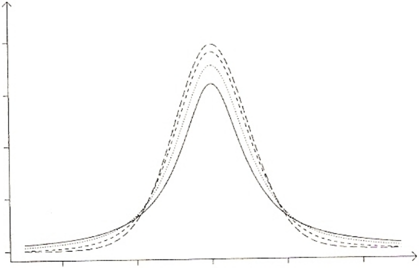
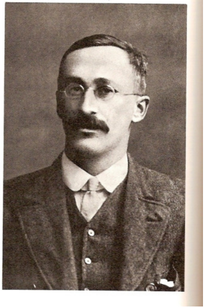


ABBILDUNG 6.19: Dichten von t-Verteilungen für $n = 1$ (—) $n = 2$ (··
(- - -) und $n = 20$ (−−) Freiheitsgrade

W.S. Gosset ("Student")



Für $n \to \infty$ konvergiert die Dichtekurve gegen die Dichte ϕ der Standardnormalverteilung. Ab $n > 30$ ist die Approximation bereits sehr gut. Deshalb sind die Quantile nur bis $n = 30$ vertafelt.

Quantile der Fisher-Verteilung werden vor allem bei Testverfahren der Regressions und Varianzanalyse benötigt.

Definition 4.18 (Fisher-Verteilung)

Seien $X \sim \chi^2(m)$ und $Y \sim \chi^2(n)$-verteilt und voneinander unabhängig. Dann heißt die Verteilung der Zufallsvariable

$$Z = \frac{X/m}{Y/n}$$

(zentral) Fisher- oder F-verteilt mit den Freiheitsgraden m und n, kurz $Z \sim F(m, n)$. Es gilt

$$\mathbb{E}(Z) = \frac{n}{n-2} \quad \text{für } n \geq 3,$$

$$\mathbb{V}(Z) = \frac{2n^2(n+m-2)}{m(n-4)(n-2)^2} \quad \text{für } n \geq 5.$$

Weibull-Verteilung

Geeignet zur Modellierung von Zeiten und Überlebenszeiten etc.

Definition 4.19 (Weibull-Verteilung)

Eine positive ZV X ist *Weibull*-verteilt mit Parametern $\gamma > 0$ und $\lambda > 0$ ($X \sim W(\gamma, \lambda)$), wenn sie die Dichte

$$f(x) = \begin{cases} \lambda\gamma x^{\gamma-1} e^{-\lambda x^{\gamma}} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0 \end{cases}$$

hat. Die Verteilungsfunktion ist

$$F(x) = 1 - e^{-\lambda x^{\gamma}},\ x \geq 0.$$

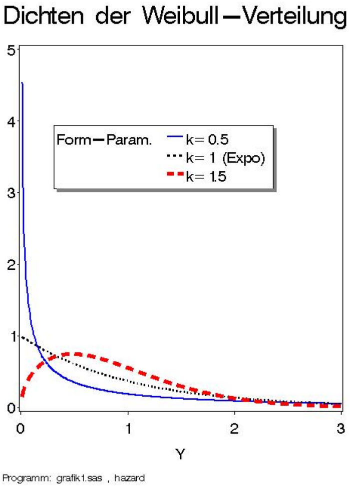
Weibull-Verteilung (Forts.)

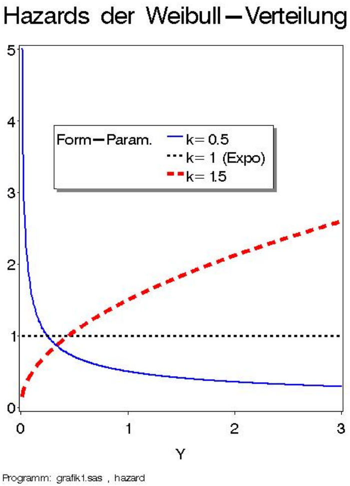
Der Form-Parameter γ hat entscheidenden Einfluss auf Gestalt der Dichte f; er bestimmt die zeitliche Entwicklung der "Ereignisgeschwindigkeit" ($\gamma = 1$: Exponentialverteilung).

Kennwerte

$$\mathbb{E}(X) = \lambda^{-\frac{1}{\gamma}} \Gamma(\frac{\gamma+1}{\gamma}),$$
$$\mathbb{V}(X) = \lambda^{-\frac{2}{\gamma}} (\Gamma(\frac{\gamma+2}{\gamma}) - \Gamma(\frac{\gamma+1}{\gamma})).$$

Weibull-Verteilung (Forts.)





Transformation einer Zufallsvariable

Man hat öfter die Situation, dass man die Verteilung einer ZV X kennt und an der Verteilung der transformierten ZV $Z = g(X)$ interessiert ist, wobei $g(x)$ eine streng monotone Funktion ist.

Satz 4.20

Sei X eine reelle ZV mit Dichte f_X und sei $g(x)$ eine streng monotone, stetig differenzierbare Funktion. Sei die ZV Z definiert als $Z = g(X)$. Dann gilt für die Verteilungsfunktion F_Z

$$F_Z(z) = \begin{cases} F_X(g^{-1}(z)), & \textit{falls } g(x) \textit{ monoton wachsend,} \\ 1 - F_X(g^{-1}(z)), & \textit{falls } g(x) \textit{ monoton fallend.} \end{cases}$$

Für die Dichte von Z gilt

$$f_Z(z) = \left|\frac{\mathrm{d}g^{-1}(z)}{\mathrm{d}z}\right| \cdot f_X(g^{-1}(z)).$$

Definition 4.21 (Hazardfunktion)

Sei T positive ZV mit Dichte $f(t)$ und Verteilungsfunktion $F(t)$. Dann heißt

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

Hazardfunktion von T.

- $\lambda(t) \cdot \Delta t \approx$ Wahrsch., dass Ereignis im Intervall $(t, t + \Delta t]$ eintritt unter der Bedingung, dass es vor dem Zeitpunkt t noch nicht eingetreten ist.

$$\lambda(t) = \lim \frac{P(t < T \leq t + dt \mid T > t)}{dt}.$$

Hazard

$$\lambda(t) = -\frac{d}{dt}\ln(1 - F(t)),$$

$$\Longrightarrow$$

$$F(t) = 1 - \exp(-\int_0^t \lambda(s)ds)$$

$\int \lambda(s)ds$ heißt *kumulativer Hazard.*

Weibullverteilung

Hazardfunktion: $\lambda(t) = \lambda\gamma t^{\gamma-1}$

$\gamma = 1$	$\lambda(t) = \lambda = const.$	kein Altern, Gedächtnislosigkeit: Exponentialvtlg.; Hazardrate hängt nicht von Zeit ab
$\gamma > 1$	$\lambda(t)$ monoton wachsend	"positives Altern": Hazardrate nimmt mit steigendem Alter zu
$\gamma < 1$	$\lambda(t)$ monoton fallend	"negatives Altern": $h(t)$ fällt

Hazardraten in männlicher US-Bevölkerung

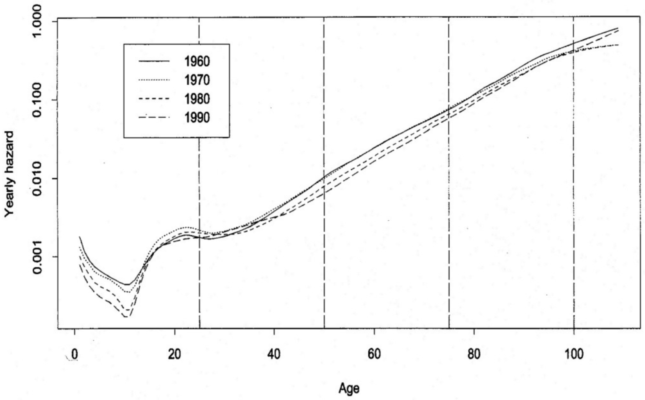


Figure: aus Therneau (1999)