Universität Basel

Mathematik 2
Dr. Thomas Zehrt

Zufallsvariablen 2:
Stetige Standardverteilungen

Benötigtes Vorwissen: Der Stoff der Vorlesung „Statistik“ wird als bekannt vorausgesetzt, insbesondere das **Kapitel 8 „Zufallsvariablen“** und das **Kapitel 9 „Diskrete Standardverteilungen“**

Inhaltsverzeichnis

1 Die stetige Gleichverteilung

Definition 1.1 *Eine stetige Zufallsvariable X mit der Dichte*

$$f_{Re}(t) = \begin{cases} \dfrac{1}{b-a} & t \in [a,b] \\ 0 & \textit{sonst} \end{cases}$$

heisst (stetig) <u>gleichverteilt</u> oder auch <u>rechtecksverteilt</u> auf dem Intervall $[a,b]$.

$$f_{Re}(t) = \begin{cases} \dfrac{1}{b-a} & t \in [a,b] \\ 0 & \text{sonst} \end{cases}$$

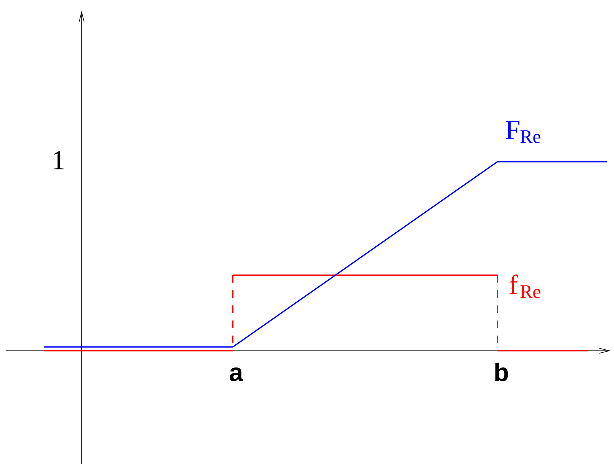
$$E(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$$

Wir wollen die Verteilungsfunktion F_{Re} der stetigen Gleichverteilung auf dem Intervall $[a,b]$ bestimmen und skizzieren. Es gilt:

$$F_{Re}(x) = \int_{-\infty}^{x} f_{Re}(t)\, dt = \begin{cases} 0 & x < a \\ \dfrac{x}{b-a} + c & x \in [a,b] \\ 1 & x > b \end{cases}$$

Die Integrationskonstante c muss nun so gewählt werden, dass $F_{Re}(a) = 0$ (oder $F_{Re}(b) = 1$) gilt. Insbesondere kann man im allgemeinen **nicht** $c = 0$ setzen.

$$F_{Re}(a) = \frac{a}{b-a} + c = 0 \leftrightarrow c = \frac{-a}{b-a}$$



2 Die Exponentialverteilung

Definition 2.1 *Eine stetige Zufallsvariable X mit der Dichte*

$$f_{Ex}(t;\lambda) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & sonst \end{cases}$$

heisst exponentialverteilt mit dem Parameter λ. Kurz: $X \sim Ex(\lambda)$

$$\boxed{\begin{aligned} f_{Ex}(t;\lambda) &= \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & \text{sonst} \end{cases} \\ E(X) &= \frac{1}{\lambda} \\ Var(X) &= \frac{1}{\lambda^2} \end{aligned}}$$

Wir wollen wieder die Verteilungsfunktion F_{Ex} der Exponentialverteilung bestimmen. Es gilt:

$$\begin{aligned} F_{Ex}(x) = \int_{-\infty}^{x} f_{Ex}(t)\, dt &= \begin{cases} 0 & x < 0 \\ \int_0^x \lambda \cdot e^{-\lambda t}\, dt & x \geq 0 \end{cases} \\ &= \begin{cases} 0 & x < 0 \\ \lambda \cdot \frac{1}{-\lambda} \left[e^{-\lambda t}\right]_0^x & x \geq 0 \end{cases} \\ &= \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \end{aligned}$$

Aufgabe 2.1 *Ermitteln Sie für eine mit* $\lambda = 0.4$ *exponentialverteilte Zufallsvariable* X *die folgenden Wahrscheinlichkeiten (bzw. bedingten Wahrscheinlichkeiten):*
$P(X < 3)$, $P(2 < X \leq 3)$, $P(X \geq 3|X > 2)$ *und* $P(X \leq 3|X > 2)$.

Lösungen: 0.6988, 0.1481, 0.6703, 0.3297

$$P(X \leq 3|X > 2) \;=\; \frac{P(X \leq 3 \cap X > 2)}{P(X > 2)} \;=\; \frac{P(2 < X \leq 3)}{P(X > 2)} \;=\; \frac{F_{Ex}(3) - F_{Ex}(2)}{1 - F_{Ex}(2)} \;=\; 1 - e^{-0.4}$$

Aufgabe 2.2 X *sei eine exponentialverteilte Zufallsvariable. Beweisen Sie die folgende Gleichung*

$$P(X \leq w + x|X > w) \;=\; P(X \leq x).$$

3 Die Normalverteilung

Die Normalverteilung kommt in der Praxis relativ häufig vor. Beispiele für normalverteilte Zufallsvariablen sind:

- der Intelligenzquotient,
- die täglichen Renditen einer Firma,
- das Schlachtgewicht von Mastochsen ...

Gemeinsamkeit: Es sind Grössen, die aus der Summe vieler verschiedener (und kleiner) zufälliger Einflüsse entstehen.

Einflüsse auf den Intelligenzquotienten:

Gene, Eltern, Geschwister, Schule, Lehrer, Ernährung, wie lange man als Säugling einen Schnuller hatte usw.

Definition 3.1 *Eine stetige Zufallsvariable X heisst* <u>*normalverteilt*</u> *oder $N(\mu, \sigma^2)$-verteilt, mit den Parametern*

- μ *($-\infty < \mu < \infty$) und*
- σ^2 *($0 < \sigma^2$),*

falls ihre Dichte durch die Funktion

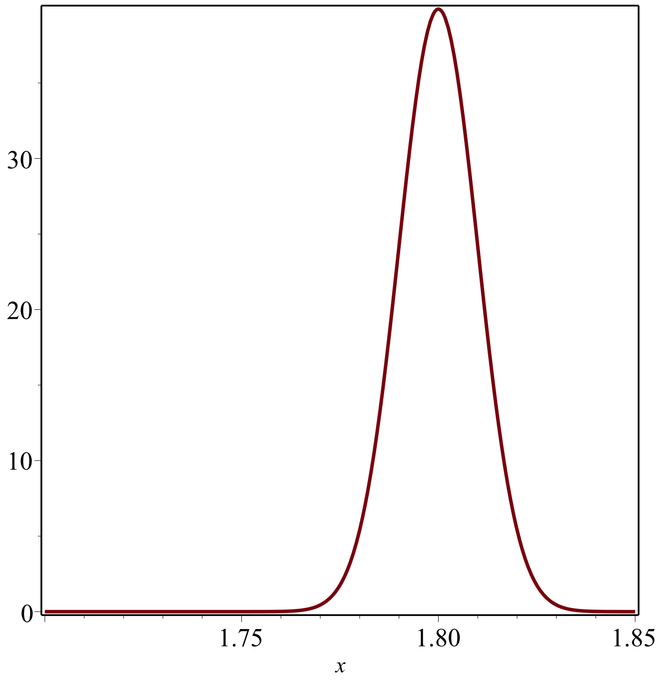
$$f(t) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} \;=:\; \phi(t;\mu,\sigma^2)$$

gegeben ist.

Der Graph von f wird auch als <u>*Gaußsche Glockenkurve*</u> *bezeichnet.*

Beispiel 3.1

$$\phi(t;1.8,0.01^2) \quad = \quad \frac{1}{0.01\sqrt{2\pi}}\, e^{-\frac{(t-1.8)^2}{2\cdot 0.01^2}}$$



Satz 1 (Eigenschaften von $\phi(t;\mu,\sigma^2)$) *Die Zufallsvariable X sei $N(\mu,\sigma^2)$-verteilt mit der Dichtefunktion $\phi(t;\mu,\sigma^2)$. Dann gilt:*

1. *$\phi(t;\mu,\sigma^2)$ ist symmetrisch bezüglich der Achse $x=\mu$, d.h. es gilt $\phi(\mu+t;\mu,\sigma^2) = \phi(\mu-t;\mu,\sigma^2)$ für alle reellen Zahlen t.*
2. *$\phi(t;\mu,\sigma^2)$ hat ein lokales Maximum an der Stelle $t=\mu$ und sonst keine Extremalstellen.*
3. *Die einzigen beiden Wendestellen von $\phi(t;\mu,\sigma^2)$ sind $\mu+\sigma$ und $\mu-\sigma$.*
4. $\lim\limits_{t\to\pm\infty} \phi(t;\mu,\sigma^2) = 0$
5. $E(X) \;=\; \mu$ *und* $Var(X) \;=\; \sigma^2$

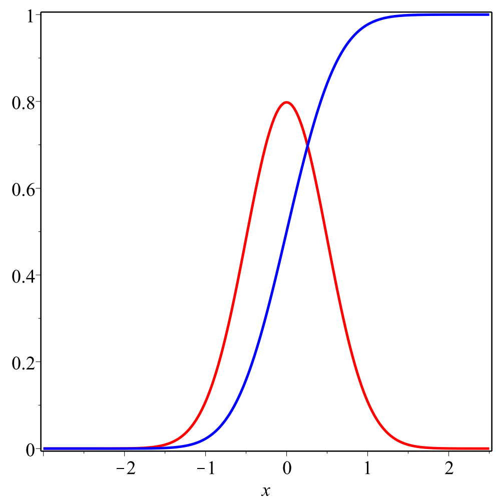
Die Verteilungsfunktion ist dann

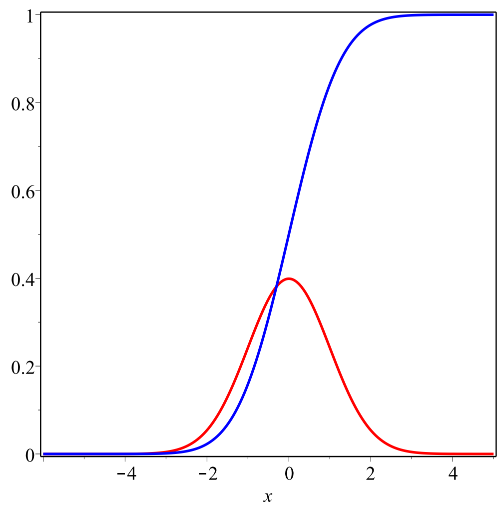
$$F(x) \;=\; \int\limits_{-\infty}^{x} f(t)\,dt \;=\; \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt \;=:\; \Phi(x;\mu,\sigma^2).$$

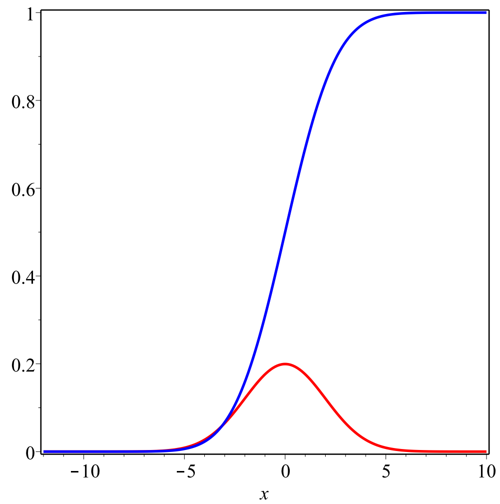
Bemerkung:

Die Dichtefunktion f ist nicht elementar integrierbar, d.h. f hat keine Stammfunktion, die aus den elementaren Funktionen (Polynome, rationale Funktionen, $e^x, \log(x)$, $\sin(x), \cos(x), \ldots$) zusammengesetzt ist.

Beispiel 3.2 *Dichtefunktion (rot) und Verteilungsfunktion (blau) für* $\mu = 0$ *und* $\sigma = 0.5, 1$ *und* 2







Die Standardnomalverteilung

Von spezieller Bedeutung ist die Standardnormalverteilung eine Normalverteilung mit den Parametern $\mu = 0$ und $\sigma^2 = 1$. Die Dichte ist

$$f(t) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}} \;=:\; \phi(t)$$

und die zugehörige Verteilungsfunktion schreibt sich als:

$$F(x) \;=\; \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \;=:\; \Phi(x).$$

Satz 2 (Einfache Eigenschaften von Φ)

- $\Phi(0) = 1/2$
- $\Phi(-x) = 1 - \Phi(x)$

Beweisidee:

Beide Eigenschaften beruhen darauf, dass die Dichtefunktion ϕ symmetrisch zur $y-$Achse ist, d.h. es gilt stets $\phi(-t) = \phi(t)$. Deshalb befindet sich genau die Hälfte der Gesamtfläche unter dem Graphen von ϕ links von 0 und es gilt $\Phi(0) = 1/2$.

Ist $x > 0$ so erkennt man auch, dass die Flächen unter dem Graphen links von $-x$ und rechts von x gleich gross sind. Daraus folgt die zweite Eigenschaft. $\square$

Satz 3 *Ist die Zufallsvariable X $N(\mu, \sigma^2)$-verteilt, so ist die Zufallsvariable* $\dfrac{X-\mu}{\sigma}$ *standardnormalverteilt:*

$$\boxed{\Phi(x;\mu,\sigma^2) \;=\; \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

Meist werden wir die folgende Formel benötigen, die diesen Zusammenhang auf konkrete Wahrscheinlichkeitsbestimmungen anwendet:

$$\boxed{\int\limits_a^b \phi(x;\mu,\sigma^2)\,dx \;=\; \int\limits_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \phi(x)\,dx}$$

bzw.

$$\boxed{\Phi\left(b;\mu,\sigma^2\right) - \Phi\left(a;\mu,\sigma^2\right) \;=\; \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

Beweis: Wir wählen die Substitution $u = u(x) = \frac{x-\mu}{\sigma}$ und damit $du = \frac{1}{\sigma}\,dx$ und erhalten damit schrittweise:

$$\begin{aligned}
\int\limits_a^b \phi(x;\mu,\sigma^2)\,dx &= \frac{1}{\sigma\sqrt{2\pi}} \int\limits_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}u^2}\,du \\
&= \underbrace{\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}x^2}\,dx}_{=\Phi\left(\frac{b-\mu}{\sigma}\right)} - \underbrace{\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-\frac{1}{2}x^2}\,dx}_{=\Phi\left(\frac{a-\mu}{\sigma}\right)},
\end{aligned}$$

Es genügt also, die Funktionswerte von $\Phi(x)$ bestimmen zu können, um die Funktionswerte jeder der Funktionen $\Phi(x;\mu,\sigma^2)$ zu berechnen!!

$\square$

Früher gab es umfangreiche Tafelwerke, in denen die Funktionswerte von Φ tabelliert waren.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.866	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982

Aufgabe 3.1 *Die Zufallsvariable sei normalverteilt mit $E(X) = 1$ und $Var(X) = 4$. Drücken Sie die folgenden Wahrscheinlichkeiten durch Funktionswerte der Funktion Φ aus: $P(|X| > 4)$ und $P(|X - 1| > 6)$.*

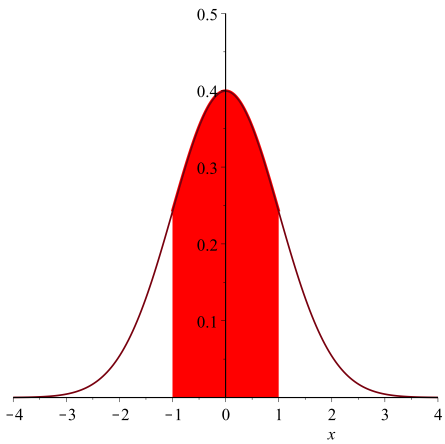
Lösungen:

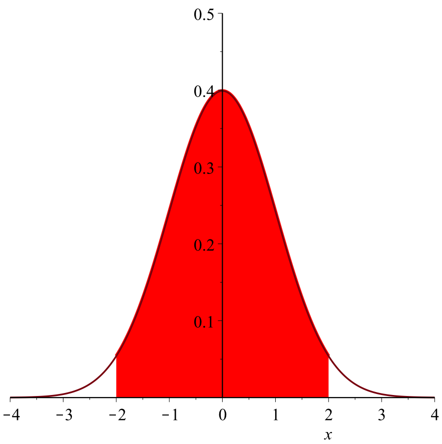
$$\begin{aligned} P(|X| > 4) &= P(X < -4 \text{ oder } X > 4) = \Phi(-4; 1, 4) + (1 - \Phi(4; 1, 4)) = \ldots \\ &= 2 - \Phi(5/2) - \Phi(3/2) \end{aligned}$$

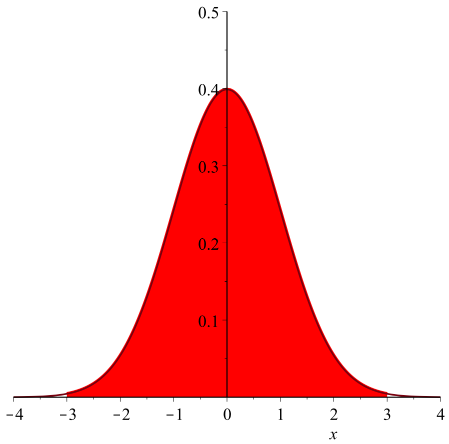
$$\begin{aligned} P(|X - 1| > 6) &= P(X < -5 \text{ oder } X > 7) = \Phi(-5; 1, 4) + (1 - \Phi(7; 1, 4)) = \ldots \\ &= 2 - 2\Phi(3) \end{aligned}$$

Satz 4 ($k\sigma$-Bereiche der Normalverteilungen) *Die folgenden Wahrscheinlichkeiten sind unabhängig von der konkreten Wahl der Parameter μ und σ:*

$$
\begin{aligned}
P(\mu - \sigma \le X \le \mu + \sigma) &= 0.683\\
P(\mu - 2\sigma \le X \le \mu + 2\sigma) &= 0.955\\
P(\mu - 3\sigma \le X \le \mu + 3\sigma) &= 0.997
\end{aligned}
$$







Beweis:

Die Zufallsvariable X sei $N(\mu, \sigma^2)$-verteilt. Dann gilt:

$$
\begin{aligned}
& P(\mu - \sigma \le X \le \mu + \sigma)\\
= \; & \Phi(\mu + \sigma; \mu, \sigma) - \Phi(\mu - \sigma; \mu, \sigma)\\
= \; & \Phi\left((\mu + \sigma - \mu)/\sigma\right) - \Phi\left((\mu - \sigma - \mu)/\sigma\right)\\
= \; & \Phi(1) - \Phi(-1)\\
= \; & 2\Phi(1) - 1\\
\approx \; & 0.683.
\end{aligned}
$$

Die anderen beiden Regeln können analog bewiesen werden.

$\square$

Wir wollen noch darauf hinweisen, dass Summen von normalverteilten Zufallsvariablen selbst wieder normalverteilte Zufallsvariablen sind.

Satz 5

Seien $X_1, X_2, \ldots, X_n$ unabhängige und normalverteilte Zufallsvariablen mit $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1, 2, \ldots, n$. Dann ist die Zufallsvariable $X_1 + X_2 + \ldots + X_n$ perfekt (nicht nur asymptotisch) $N(\mu_1 + \mu_2 + \ldots + \mu_n, \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$-verteilt.

4 Die Testverteilungen

Aus der Normalverteilung lassen sich drei wesentliche Verteilungen, die sogenannten

Prüf- oder Testverteilungen,

gewinnen. Diese Verteilungen spielen eine grosse Rolle beim Prüfen von Hypothesen über Nomalverteilungen.

Zur Beschreibung der Dichtefunktionen benötigt man zunächst die folgenden beiden Funktionen. Dabei sind diese Funktionen hier nur dazu nötig, die Normalisierungskonstanten (die gewährleisten, dass der Flächeninhalt unter den Graphen der Dichtefunktionen genau 1 ist) einheitlich ausdrücken zu können.

- Die Gammafunktion ($x > -1$)

$$\Gamma(x) \;=\; \int_0^\infty t^{x-1}\, e^{-t}\, dt$$

  Es gilt z.B.

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty t^0\, e^{-t}\, dt \;=\; \lim_{c\to\infty} [-e^{-t}]_0^c \;=\; 1 \\
\Gamma(1/2) &= \sqrt{\pi} \\
\Gamma(n+1) &= n \cdot \Gamma(n) \qquad \text{für alle } n \in \mathbb{N} \quad (\text{oder sogar } n \in \mathbb{R}_+)
\end{aligned}$$

  Diese Eigenschaft kann durch partielle Integration bewiesen werden und zeigt, dass die Gammafunktion die Fakultät interpoliert, die nur für natürliche Zahlen definiert ist. Dass die Gammafunktion dabei so definiert ist, dass nicht $\Gamma(n)$ sondern $\Gamma(n+1)$ gleich $n!$ ist, hat historische Gründe.

- Die Betafunktion ($p, q > 0$)

$$B(p,q) \;=\; \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

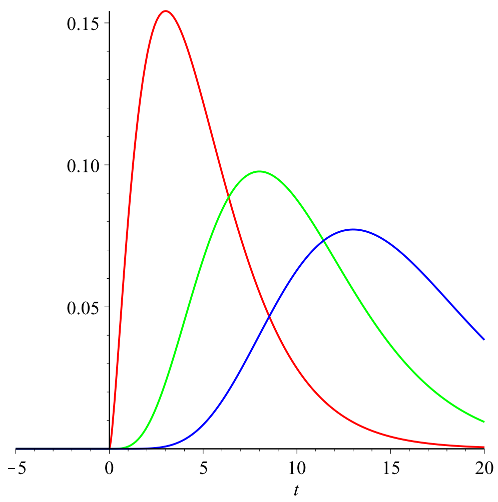
χ^2-Verteilungen

Definition 4.1 *$Z_1, Z_2, \ldots, Z_n$ seien n unabhängige und identisch $N(0,1)$-verteilte Zufallsvariablen. Dann ist die Summe ihrer Quadrate $Z = \sum_{i=1}^n Z_i^2$ $\underline{\chi^2\text{-verteilt}}$ mit n Freiheitsgraden. Kurz: $Z \sim \chi_n^2$*

Die Dichtefunktion einer χ_n^2-verteilten Zufallsvariablen mit n Freiheitsgraden ist

$$f(t) = \begin{cases} 0 & t \leq 0 \\ \dfrac{1}{2^{\frac{n}{2}}\,\Gamma(\frac{n}{2})}\, t^{\frac{n}{2}-1}\, e^{-\frac{t}{2}} & t > 0 \end{cases}$$

Die Dichtefunktionen für χ_5^2 (rot), χ_{10}^2 (grün) und χ_{15}^2 (blau):



Satz 6 (Additionssatz) *Die Summe zweier unabhängiger χ_n^2-verteilter bzw. χ_m^2-verteilter Zufallsvariablen ist χ_{n+m}^2-verteilt.*

Wesentliches Beispiel: Für unabhängige Zufallsvariablen $X_i \sim N(\mu, \sigma^2)$ $(i = 1, 2, \ldots, n)$ gilt für die Stichprobenvarianz

$$S_X^2 \;=\; \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

tatsächlich

$$\frac{n-1}{\sigma^2} \cdot S_X^2 \;\sim\; \chi_{n-1}^2.$$

t-Verteilungen

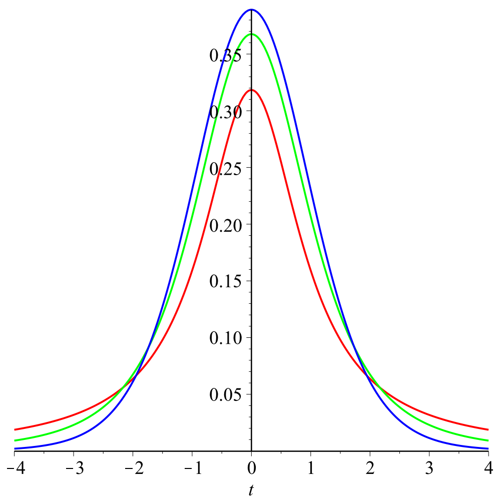
Definition 4.2 *Sind X und Y unabhängige Zufallsvariablen mit $X \sim N(0,1)$ und $Y \sim \chi_n^2$ so besitzt der Quotient*

$$\frac{X}{\sqrt{Y/n}} \;\sim\; t_n$$

eine <u>t_n-Verteilung</u> (Student-Verteilung) mit n Freiheitsgraden.
Die Dichtefunktion einer t_n-verteilten Zufallsvariablen mit n Freiheitsgraden ist

$$f(t) \;=\; \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi\, n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Die Dichtefunktionen für t_1 (rot), t_3 (grün) und t_{10} (blau):



Beispiel 4.1 *Im Falle $n = 1$ entsteht die so genannte* **Cauchy-Verteilung** *mit der Dichtefunktion*

$$f(t) \;=\; \frac{1}{\pi}\,\frac{1}{1+t^2}.$$

Wir könne hier noch relativ leicht bestätigen (falls wir die Stammfunktion kennen), dass es sich hier wirklich um eine Dichtefunktion handelt. Es gilt zunächst

$$\frac{1}{\pi}\int_a^b \frac{1}{1+t^2}\,dt \;=\; \frac{1}{\pi}(\arctan(b) - \arctan(a))$$

und somit

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+t^2}\,dt \;=\; \frac{1}{\pi}\lim_{z\to\infty}(\arctan(z) - \arctan(-z)) \;=\; \frac{1}{\pi}\,(\pi/2 - (-\pi/2)) \;=\; 1$$

Satz 7 *Seien $X_1, X_2, \ldots, X_n$ unabhängige und identisch $N(\mu, \sigma^2)$-verteilte Zufallsvariablen. Dann sind $\overline{X}$ und S_X^2 unabhängig. Der folgende Quotient ist t_{n-1}-verteilt:*

$$\frac{(\overline{X}-\mu)\sqrt{n}}{S_X} = \frac{(\overline{X}-\mu)\sqrt{n}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}} \sim t_{n-1}$$

Die linke Seite kann auch wie folgt (vergl. Definition oben) geschrieben werden:

$$\frac{(\overline{X}-\mu)\sqrt{n}}{S_X} = \frac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}{\sqrt{\dfrac{\dfrac{n-1}{\sigma^2} S_X^2 \sim \chi^2_{n-1}}{n-1}}}.$$

F-Verteilungen

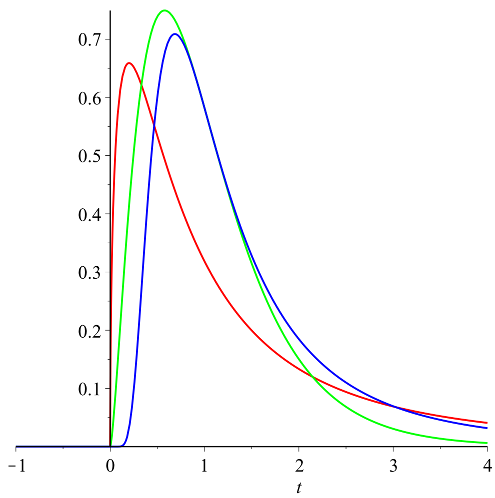
Definition 4.3 *Sind X und Y unabhängige Zufallsvariablen mit $X \sim \chi^2_m$ und $Y \sim \chi^2_n$ so besitzt der Quotient*

$$\frac{X/m}{Y/n} \sim F_{m,n}$$

eine Fisher'sche F-Verteilung mit (m,n) Freiheitsgraden.
Die Dichtefunktion einer F-verteilten Zufallsvariablen mit (m,n) Freiheitsgraden ist

$$f(t) = \begin{cases} 0 & t \leq 0 \\ \dfrac{m^{\frac{m}{2}}\, n^{\frac{n}{2}}}{B(\frac{m}{2},\frac{n}{2})} \, \dfrac{t^{\frac{m}{2}-1}}{(m\,t+n)^{\frac{m+n}{2}}} & t > 0 \end{cases}$$

Die Dichtefunktionen für $F_{3,3}$ (rot), $F_{5,50}$ (grün) und $F_{50,5}$ (blau):



Für $X \sim N(\mu_X, \sigma^2)$ und $Y \sim N(\mu_Y, \sigma^2)$ betrachten wir zunächst die beiden Stichprobenvarianzen

$$\begin{aligned} S_X^2 &= \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \overline{X})^2 \quad \text{Stichprobenumfang } m \\ S_Y^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \quad \text{Stichprobenumfang } n \end{aligned}$$

Dann gilt für das Verhältnis der beiden Stichprobenvarianzen:

$$\frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

Anmerkung: Ist eine Zufallsvariable X $F_{m,n}$-verteilt, so ist $1/X$ natürlich $F_{n,m}$-verteilt. Das ist offensichtlich, wenn man die Definition als Quotient betrachtet. Der folgende Satz ist in der Testtheorie manchmal recht nützlich.

Satz 8 *Sei $F_{m,n;p}$ das p-Quantil der $F_{m,n}$-Verteilung. Dann gilt*

$$F_{n,m;1-p} = \frac{1}{F_{m,n;p}}.$$

Beweis: Sei X eine $F_{m,n}$-verteilte Zufallsvariable und $1/X$ (natürlich) $F_{n,m}$-verteilt. Wir starten mit der üblichen Definition eines p-Quantils und formen diese um.

$$\begin{aligned} & P(X \leq F_{m,n;p}) = p \\ \leftrightarrow \quad & P\left(\frac{1}{X} \geq \frac{1}{F_{m,n;p}}\right) = p \\ \leftrightarrow \quad & P\left(\frac{1}{X} \leq \frac{1}{F_{m,n;p}}\right) = 1 - p \end{aligned}$$

Die Wahrscheinlichkeit, dass die $F_{n,m}$-verteilte Zufallsvariable einen Wert von maximal $\frac{1}{F_{m,n;p}}$ annimmt, ist also $1 - p$. Daraus folgt die Behauptung. □

5 *Zufallszahlen*

Ein konkretes in der Praxis (Simulation und Monte-Carlo-Algorithmen) auftretendes Problem ist die Erzeugung von Zufallszahlen, die einer vorgegebenen Verteilung folgen. Im Allgemeinen wird man dabei zunächst versuchen, stetig gleichverteilte Zufallszahlen auf dem Intervall $[0, 1]$ zu erzeugen. Dann wird man im nächsten Schritt versuchen, die erhaltene Sequenz so zu bearbeiten, dass die neue Sequenz von Zufallszahlen der gewünschten Verteilung folgt.

Erzeugung gleichverteilter Zufallszahlen Um ein Folge echter (gleichverteilter) Zufallszahlen am Computer zu erzeugen, benötigt man einen zufälligen physikalischen Prozess. Insbesondere kann ein (deterministisches) Computerprogramm keine echten Zufallszahlen erzeugen, daher benötigt man spezielle Hardware. Die Firma Intel nutzt beispielsweise das thermische Rauschen eines Widerstandes, um daraus Zufallszahlen zu generieren.

Oftmals genügen aber auch so genannte Pseudozufallszahlen, die von speziellen Computerprogrammen, so genannten Pseudozufallszahlgeneratoren, deterministisch erzeugt werden. Einfache Pseudozufallszahlgeneratoren sind beispiesweise so genannte Schieberegister oder Kongruenzgeneratoren. Diese Programme erzeugen zufällig erscheinende Zahlenfolgen. Kennt man allerdings den Algorithmus, kann man jede Zahl der Folge vorhersagen.

Richtig verteilte Zufallszahlen Nehmen wir an, dass wir eine Folge auf dem Intervall $[0, 1]$ gleichverteilter Zufallszahlen erzeugt haben. Wie kann man daraus Zufallszahlen konstruieren, die irgend einer anderen Verteilung folgen?

Satz 9 *Sei F die Verteilungsfunktion einer gewünschten (stetigen) Verteilung mit der zugehörigen Umkehrfunktion F^{-1}. Weiterhin sei U eine auf dem Intervall $[0, 1]$ stetig gleichverteilte Zufallsvariable.*

Dann ist die Zufallsvariable $X := F^{-1}(U)$ eine Zufallsvariable mit Verteilungsfunktion F.

Beweis:

Zunächst gilt sicher $F^{-1}(U) \leq x \iff U \leq F(x)$ und somit folgt:

$$P(X \leq x) \;=\; P(F^{-1}(U) \leq x) \;=\; P(U \leq F(x)) \;=\; F(x).$$

Dabei ist die letzte Umformung korrekt, da U stetig gleichverteilt auf dem Intervall $[0, 1]$ ist, also ganz allgemein $P(U \leq y) = y$ gilt. $\square$

Zufallszahlen in R R enthält Befehle, die beliebig verteilte (Pseudo)zufallszahlen erzeugen können.

```
> x <- runif(1000, min=0, max=1)
```

erzeugt 1000 auf $[0, 1]$ gleichverteilte Zufallszahlen (Vektorkomponeten von x)

```
> x <- rnorm(1000, mean=0, sd=1)
```

erzeugt 1000 standardnormalverteilte Zufallszahlen

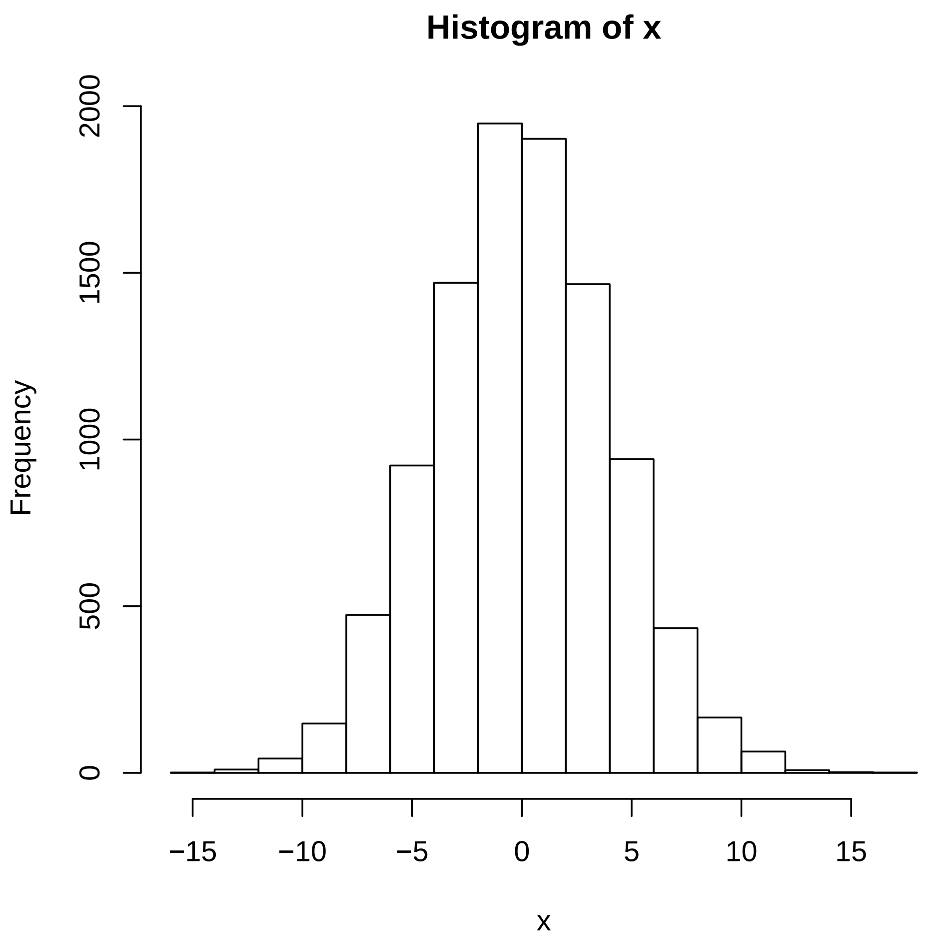
> x < − rexp(1000, rate=1)

erzeugt 1000 exponentialverteilte Zufallszahlen (mit Parameter 1)

Beispiel 5.1 *Geben wir in* R *die folgenden beiden Kommandos ein, sollte eine ähnliche Skizze erzeugt werden.*

> x < − rnorm(10000, mean=0, sd=4)

> hist(x)



6 Übungsaufgaben

1. Eine Reifenfirma untersucht die Lebensdauer X eines neu entwickelten Reifens. Dabei zeigt sich, dass die ermittelte Lebensdauer der Reifen gut durch eine Normalverteilung mit den Parametern $\mu = 36'000$ km und $\sigma = 4'000$ km angenähert werden kann.

   (a) Wie gross ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Reifen höchstens 48'000 km hält?

   (b) Wie gross ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Reifen mehr als 28'000 km hält?

   (c) Wie gross ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Reifen zwischen 30'000 und 40'000 km hält?

   (d) Wie gross ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Reifen, der schon über 30'000 km gehalten hat, höchstens 40'000 km hält?
   Hinweis: Bedingte Wahrscheinlichkeit.

   (e) Welche Lebensdauer wird von 95 Prozent der Reifen nicht überschritten?
   Hinweis: Bestimme l, so dass $P(X \leq l) = 0.95$.

2. Sei Z eine Zufallsvariable mit $E(Z) = 0$ und $Var(Z) = 1$. Sie wollen ein Intervall $[-c, c]$ finden, so dass Z mit einer Wahrscheinlichkeit von mindestens 0.97 einen Wert in diesem Intervall annimmt, d.h. es soll $P(Z \in [-c, c]) \geq 0.97$ gelten.

   (a) Welches c liefert die Ungleichung von Tschebyschev?

   (b) Bestimmen Sie (das optimale) c, falls $Z \sim N(0; 1)$.

3. Eine Glühlampe sei ununterbrochen in Betrieb, bis sie ausfällt. Die Zufallsgrösse X, die die (zufällige) Lebensdauer der Glühlampe angibt, sei exponentialverteilt.

   Weiterhin sei bekannt, dass derartige Glühlampen eine mittlere Lebenszeit von 6000 Stunden haben.

   (a) Wie muss λ gewählt werden, damit $E(X)$ gleich der mittleren Lebenszeit ist?

   (b) Wie gross ist die Wahrscheinlichkeit, dass die Glühlampe nach 6000 Stunden bereits ausgefallen ist?

   (c) Eine Glühlampe sei bereits 3000 Stunden in Betrieb. Wie gross ist die Wahrscheinlichkeit, dass sie noch mindestens 3000 Stunden weiterbrennt?

4. Bestimmen Sie mittels der Tabellen in der Formelsammlung die folgenden Quantile.

   (a) Z standardnormalverteilt: $z_{0.95}$, $z_{0.99}$ und $z_{0.995}$

   (b) χ^2-Verteilung: $\chi^2_{10;0.95}$, $\chi^2_{20;0.99}$ und $\chi^2_{100;0.95}$

   (c) t-Verteilung: $t_{10;0.9}$, $t_{30;0.99}$ und $t_{120;0.99}$

   (d) F-Verteilung: $F_{10,10;0.95}$ und $F_{14,10;0.95}$

   Skizzieren Sie (grob) die zugehörigen Dichtefunktionen. Wo kann man für das jeweilige p-Quantil den Wert p „sehen“ ?