

Summe, Produkt, Quotient von Zufallsvariablen

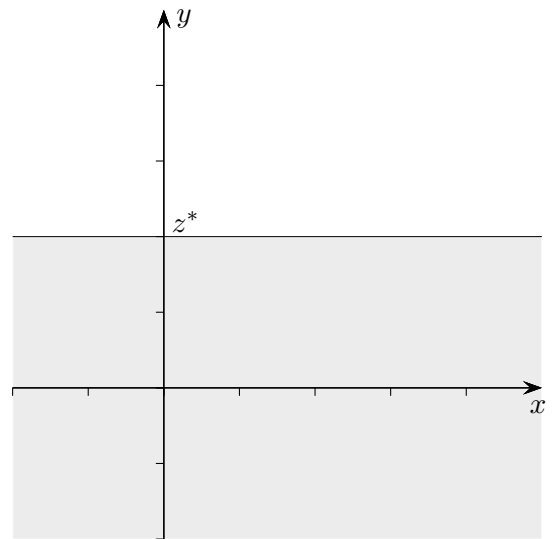
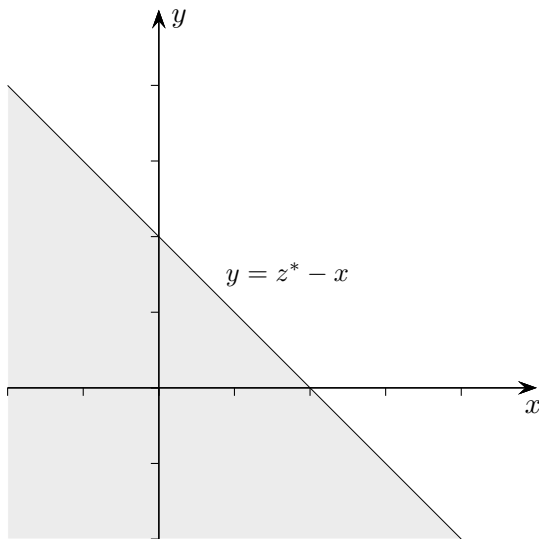
1. Summe von Zufallsvariablen
2. Produkt von Zufallsvariablen
3. Quotient von Zufallsvariablen
4. Differenz von Zufallsvariablen
5. t -Verteilung
6. Verteilung von $g(X)$
7. Verteilung der Zufallsvariablen $\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$
8. t -Test

↑ Summe von Zufallsvariablen

Die Zufallsvariablen X und Y haben die gemeinsame Dichte $f(x, y)$.
 Falls sie unabhängig sind, gilt $f(x, y) = f_X(x) \cdot f_Y(y)$.

Untersuchen wir die Verteilungsfunktion der Summe $X + Y$:

$$P(X + Y \leq z^*) = \int \int_{x+y \leq z^*} f(x, y) \, dx \, dy$$



Mit der Transformation $(x, z - x) \longleftarrow (x, z)$ Probe: $x + (z - x) = z$
 kann die Integration vereinfacht werden:

$$P(X + Y \leq z^*) = \int_{-\infty}^{z^*} \underbrace{\int_{-\infty}^{\infty} f(x, z - x) \, dx}_{\text{Dichte von } X + Y} \, dz$$

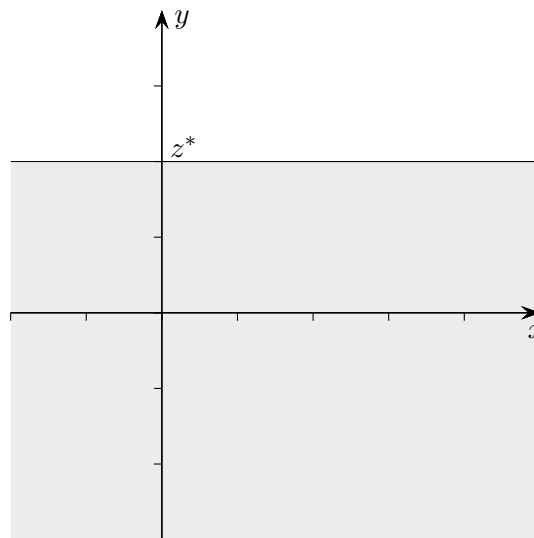
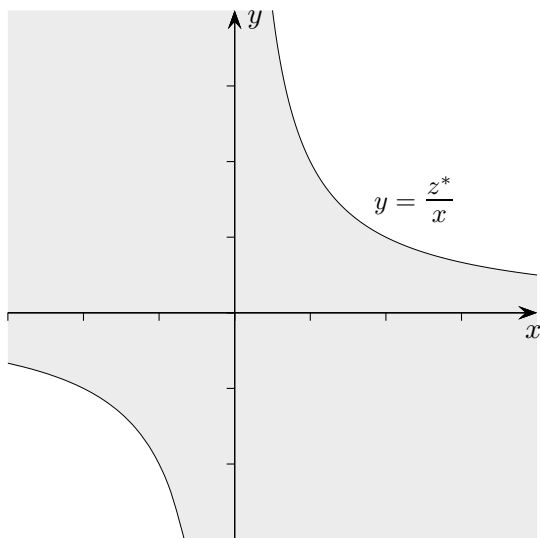
Allgemein gilt für eine Transformation (siehe Verschiedenes, krummlinige Koordinaten)

$$(u, v) \longrightarrow (x(u, v), y(u, v))$$

$$dA^* = \left| \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right| \, du \, dv \quad \text{und in diesem Fall} \quad \left| \begin{vmatrix} 1 & 0 \\ \dots & 1 \end{vmatrix} \right| = 1.$$

↑ Produkt von Zufallsvariablen

$$P(X \cdot Y \leq z^*) = \int \int_{x \cdot y \leq z^*} f(x, y) \, dx \, dy$$



Die Transformation lautet:

$$(x, \frac{z}{x}) \longleftarrow (x, z)$$

Probe: $x \cdot \frac{z}{x} = z$

$$P(X \cdot Y \leq z^*) = \int_{-\infty}^{z^*} \underbrace{\int_{-\infty}^{\infty} f(x, \frac{z}{x}) \left| \frac{1}{x} \right| \, dx}_{\text{Dichte von } X \cdot Y} \, dz$$

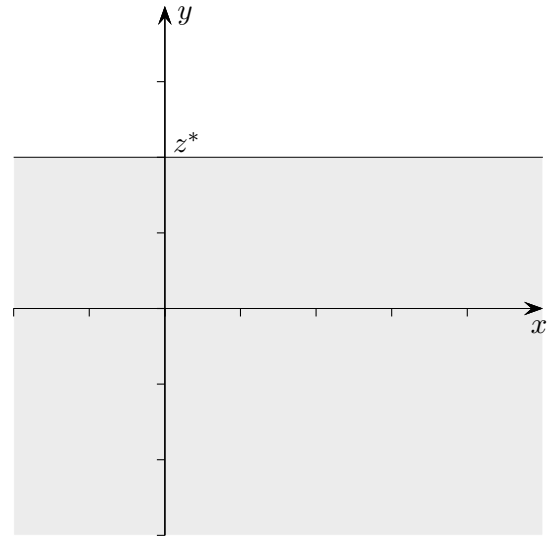
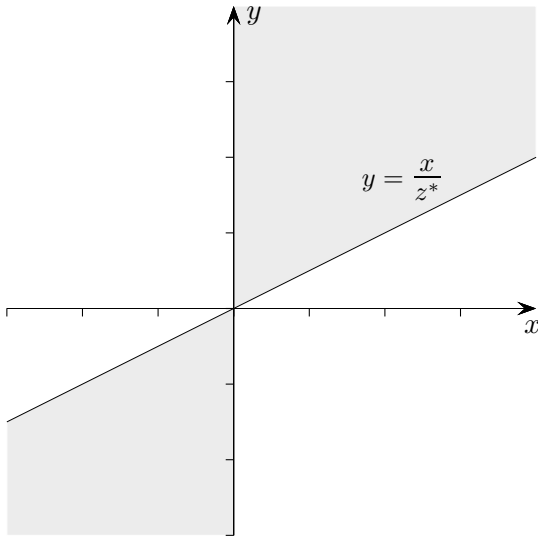
$$(u, v) \longrightarrow (x(u, v), y(u, v))$$

$$dA^* = \left| \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right| \, du \, dv$$

in diesem Fall $\left| \begin{vmatrix} 1 & 0 \\ \dots & \frac{1}{x} \end{vmatrix} \right| = \left| \frac{1}{x} \right|$

↑ Quotient von Zufallsvariablen

$$P\left(\frac{X}{Y} \leq z^*\right) = \int \int_{\frac{x}{y} \leq z^*} f(x, y) \, dx \, dy$$



Eine Transformation lautet:

$$\left(x, \frac{x}{z}\right) \longleftarrow (x, z)$$

Probe: $x : \frac{x}{z} = z$

$$P\left(\frac{X}{Y} \leq z^*\right) = \int_{-\infty}^{z^*} \underbrace{\int_{-\infty}^{\infty} f\left(x, \frac{x}{z}\right) \left|\frac{x}{z^2}\right| \, dx}_{\text{Dichte von } \frac{X}{Y}} \, dz$$

$$(u, v) \longrightarrow (x(u, v), y(u, v))$$

$$dA^* = \left| \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right| \, du \, dv \quad \text{in diesem Fall} \quad \left| \begin{vmatrix} 1 & 0 \\ \dots & -\frac{x}{z^2} \end{vmatrix} \right| = \left| -\frac{x}{z^2} \right|$$

und eine weitere:

$$(z \cdot x, x) \longleftarrow (x, z)$$

$$P\left(\frac{X}{Y} \leq z^*\right) = \int_{-\infty}^{z^*} \underbrace{\int_{-\infty}^{\infty} f(z \cdot x, x) |x| \, dx}_{\text{Dichte von } \frac{X}{Y}} \, dz$$

$$\left| \begin{vmatrix} z & x \\ 1 & 0 \end{vmatrix} \right| = |x|$$

↑

↑ Differenz von Zufallsvariablen

Für die Verteilungsfunktion der Zufallsvariablen $X - Y$ gilt:

$$P(X - Y \leq z^*) = \int \int_{x - y \leq z^*} f(x, y) \, dx \, dy$$

Mit der Transformation $(x, x - z) \longleftarrow (x, z)$
kann die Integration vereinfacht werden:

Probe: $x - (x - z) = z$

$$P(X - Y \leq z^*) = \int_{-\infty}^{z^*} \underbrace{\int_{-\infty}^{\infty} f(x, x - z) \, dx}_{\text{Dichte von } X - Y} \, dz$$

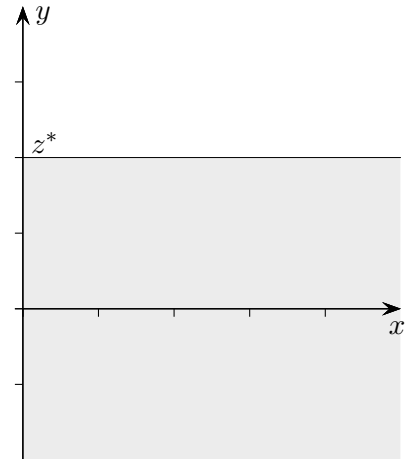
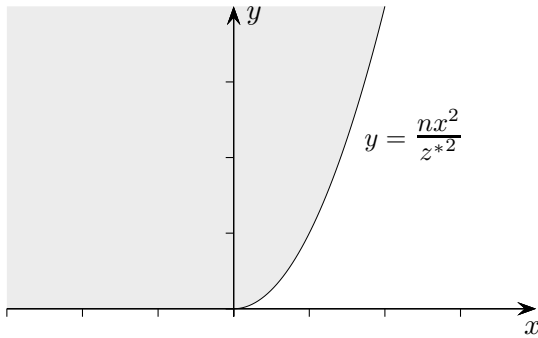
$$(u, v) \longrightarrow (x(u, v), y(u, v))$$

$$dA^* = \left| \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right| \, du \, dv \qquad \left| \begin{vmatrix} 1 & 0 \\ \dots & -1 \end{vmatrix} \right| = 1.$$

↑ t -Verteilung

X sei $N(0, 1)$ -verteilt und Y χ_n^2 -verteilt.

$$P\left(\frac{X}{\sqrt{\frac{Y}{n}}} \leq z^*\right) = \int \int_{\frac{x}{\sqrt{\frac{y}{n}}} \leq z^*} f(x, y) \, dx \, dy$$



Transformation: $(z\sqrt{\frac{x}{n}}, x) \longleftarrow (x, z)$

$$P\left(\frac{X}{\sqrt{\frac{Y}{n}}} \leq z^*\right) = \int_{-\infty}^{z^*} \underbrace{\int_0^{\infty} f(z\sqrt{\frac{x}{n}}, x) \cdot \sqrt{\frac{x}{n}} \, dx}_{\text{Dichte von } \frac{X}{\sqrt{\frac{Y}{n}}}} \, dz$$

$$(u, v) \longrightarrow (x(u, v), y(u, v))$$

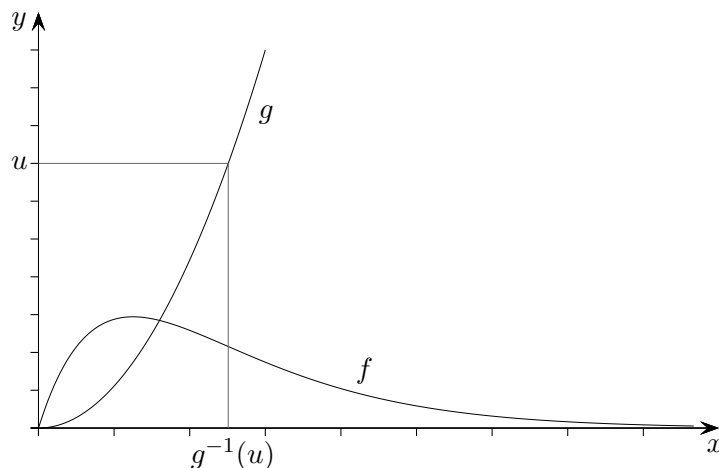
$$dA^* = \left| \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right| \, du \, dv \quad \text{in diesem Fall} \quad \left| \begin{vmatrix} \cdots & \sqrt{\frac{x}{n}} \\ 1 & 0 \end{vmatrix} \right| = \sqrt{\frac{x}{n}}$$

Die Dichte hat die Form: $g_n(x) = \dots \frac{1}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}$ (Konstanten werden mit der Γ -Funktion formuliert).

↑ Verteilung von $g(X)$

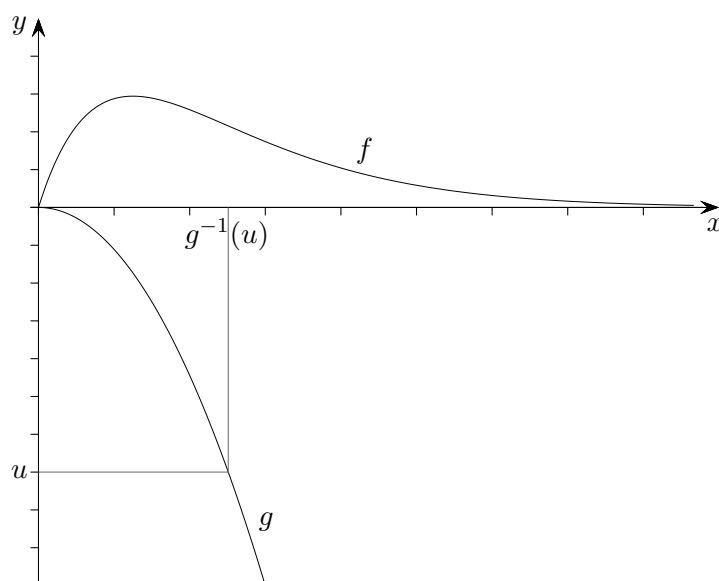
Sei die Dichte von X f , die Verteilungsfunktion F , g umkehrbar.
Wie lautet die Dichte von $g(X)$?

$$\begin{aligned} P(g(X) \leq u) &= P(X \leq g^{-1}(u)) \\ &= F(g^{-1}(u)) \\ \frac{d}{du} F(g^{-1}(u)) &= f(g^{-1}(u)) \cdot (g^{-1}(u))' \end{aligned}$$



Sei nun g monoton fallend.

$$\begin{aligned} P(g(X) \leq u) &= 1 - F(g^{-1}(u)) \\ \frac{d}{du} (1 - F(g^{-1}(u))) &= -f(g^{-1}(u)) \cdot (g^{-1}(u))' \\ &= f(g^{-1}(u)) \cdot |(g^{-1}(u))'| \end{aligned}$$



Für eine $N(0, 1)$ -verteilte Zufallsvariable X und $g(x) = \sigma x + \mu$ erhalten wir die Dichte

$$\psi(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

Mit $g(x) = \sqrt{\frac{x}{n}}$ und der Quotientenbildung für Zufallsvariablen ist eine weitere Herleitung der t -Verteilung möglich.

↑

© Roofs

↑ Verteilung der Zufallsvariablen $\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$

X sei normalverteilt und $n = 2$.

$$\begin{aligned} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 &= \frac{1}{\sigma^2} [(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2] \\ &= \frac{1}{\sigma^2} \left[\left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 \right] \\ &= \frac{1}{\sigma^2} \left[\left(\frac{X_1 - X_2}{2}\right)^2 + \left(\frac{X_1 - X_2}{2}\right)^2 \right] \\ &= \frac{1}{\sigma^2} 2 \left(\frac{X_1 - X_2}{2}\right)^2 \\ &= \left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2 \end{aligned}$$

$X_1 - X_2$ ist $N(0, \sqrt{2}\sigma)$ verteilt (Varianzen addieren sich).

Daher sind $\frac{X_1 - X_2}{\sqrt{2}\sigma}$ $N(0, 1)$ - und $\left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2$ χ_1 -verteilt.

Allgemein gilt:

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \text{ ist } \chi_{n-1}\text{-verteilt.}$$

Der Übergang von μ zu \bar{X} verringert also den Freiheitsgrad um 1.

Beweisskizze:

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n ((X_i - \bar{X}) + (\bar{X} - \mu))^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2(\bar{X} - \mu) \cdot \underbrace{\sum_{i=1}^n (X_i - \bar{X})}_{=0} + n(\bar{X} - \mu)^2 \\ \underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2}_{\chi_n^2\text{-verteilt}} &= \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \underbrace{\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right)^2}_{\chi_1^2\text{-verteilt}} \end{aligned}$$

Aus einer Unabhängigkeitsüberlegung folgt, dass die fehlende Verteilung eindeutig bestimmt ist. Daher muss es die χ_{n-1}^2 -Verteilung sein.

↑ t -Test

Um für eine normalverteilte Zufallsvariable X ein Schwankungsintervall für den Mittelwert μ zu ermitteln, wird die $N(0, 1)$ -Verteilung von $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ herangezogen.

Die Standardabweichung σ muss bekannt sein.

Falls dies nicht der Fall ist, kann σ mit der Stichprobenvarianz

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

geschätzt werden.

s für σ eingesetzt, ergibt die Zufallsvariable:

$$T = \frac{\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{(\bar{X} - \mu) \sqrt{n}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}}$$

Für einen kleinen Stichprobenumfang n ist T nicht normalverteilt, sondern unterliegt der sogenannten t -Verteilung. Dies konnte Gosset 1908 nachweisen.

Hierzu wird der Bruch mit $\frac{1}{\sigma}$ erweitert und etwas umgeformt:

$$T = \frac{\frac{(\bar{X} - \mu) \sqrt{n}}{\sigma}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}}} = \frac{\frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2}}$$

Nun ist zu erkennen, dass T vom Typ $\frac{Y}{\frac{Z}{\sqrt{n-1}}}$ ist, wobei Y $N(0, 1)$ - und Z χ_{n-1}^2 -verteilt ist.

T ist daher t -verteilt mit $n - 1$ Freiheitsgraden.

Es liege das Stichprobenergebnis X_1, X_2, \dots, X_n vor.

Dann gilt (Sicherheitswahrscheinlichkeit α):

$$-c \leq \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \leq c$$

$$\Leftrightarrow \bar{X} - c \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + c \cdot \frac{s}{\sqrt{n}}$$

c wird mit $P(T \leq c) = \frac{1+\alpha}{2}$ bestimmt (analog zur Normalverteilung), $n - 1$ Freiheitsgrade.