2.4 Stetige Zufallsvariable

Beispiel. Abfüllung von 500–Gramm–Packungen einer bestimmten Ware auf einer automatischen Abfüllanlage.

Die Zufallsvariable X beschreibe die Füllmenge einer zufällig ausgewählten Packung in g.

X kann jeden Wert (in der Nähe von 500 g) annehmen.

Bei beliebig genauer Messung wird i.a. jeder konkrete Wert nur mit Wkt. Null auftreten:

$$P(X = 500) \; = \; 0\,,$$

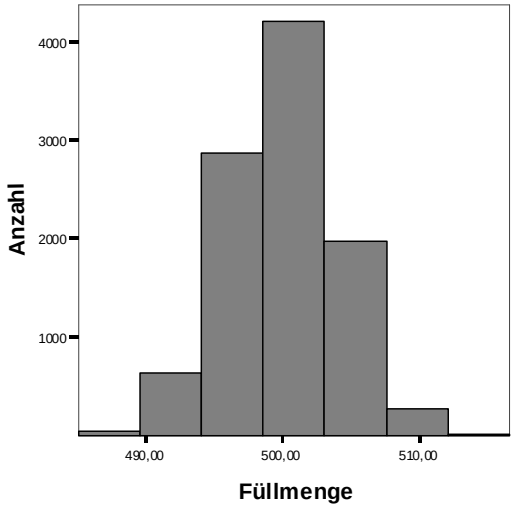
aber $P(495 < X < 505)$ wird i.a. positiv sein, und diese Wahrscheinlichkeit beschreibt den Anteil der Packungen mit einer Füllmenge zwischen 495 g und 505 g unter allen abgefüllten Packungen.

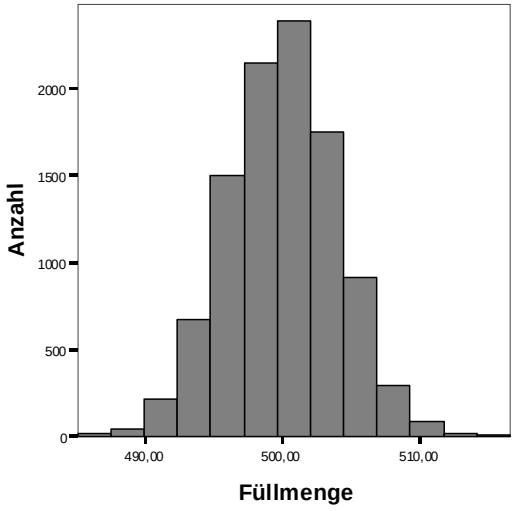
Nun: Betrachtung von 'immer feineren' Histogrammen für eine Stichprobe mit sehr vielen kontrollierten Packungen
siehe nächste Seite

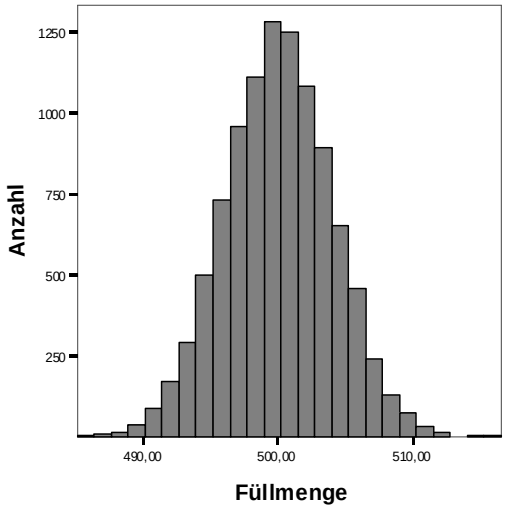
Bei beliebiger Verfeinerung (und immer größeren Stichprobenumfängen) wird eine Funktion erkennbar.

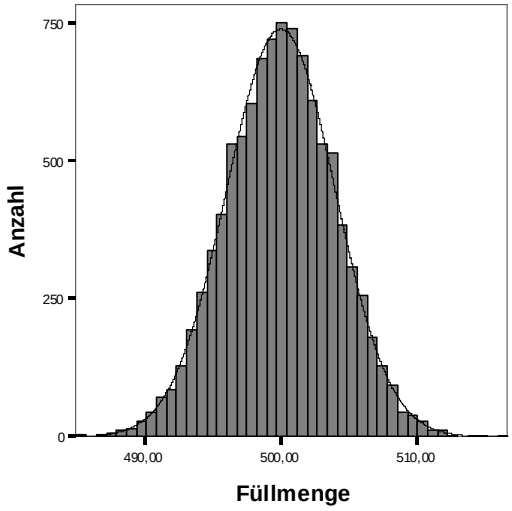
Die Verteilungsdichte f_X modelliert das Histogramm der relativen Häufigkeiten der Füllmengen aller abgefüllten Pakete, beziehungsweise das 'Abfüllverhalten' der Maschine, wenn diese Werte in immer feinere Intervalle einsortiert werden.

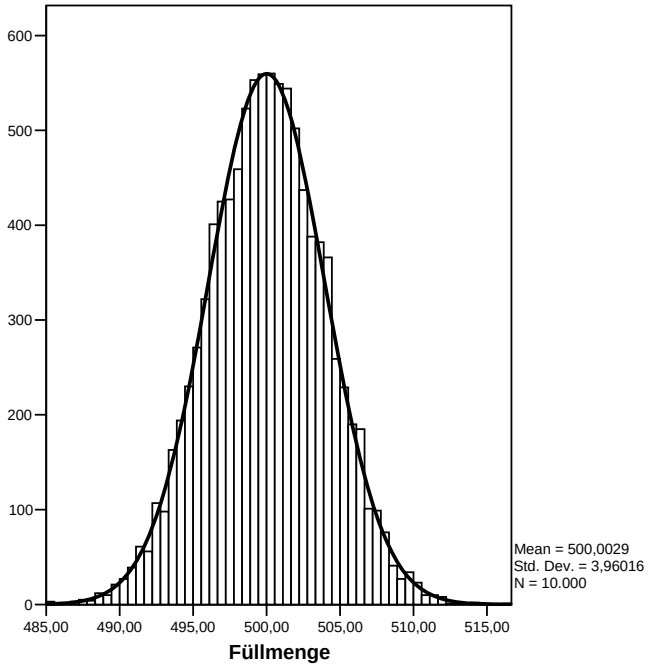
In jedem Teilintervall entspricht die Fläche unter f_X der Wkt., dass die ZV Werte in diesem Intervall annimmt.


Anzahl
4000
3000
2000
1000
490,00
500,00
510,00
Füllmenge


Anzahl
2000
1500
1000
500
0
490,00
500,00
510,00
Füllmenge


Anzahl
1250
1000
750
500
250
490,00
500,00
510,00
Füllmenge


Anzahl
750
500
250
0
490,00
500,00
510,00
Füllmenge

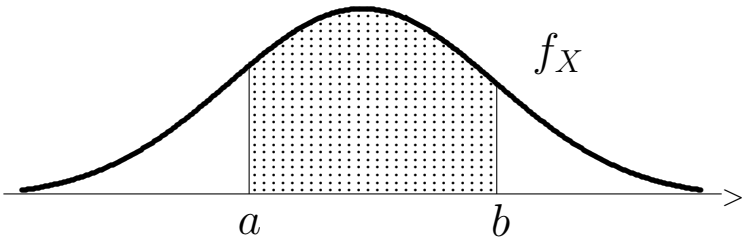

600
500
400
300
200
100
0
485,00
490,00
495,00
500,00
505,00
510,00
515,00
Füllmenge
Mean = 500,0029
Std. Dev. = 3,96016
N = 10.000

Definition: Eine Zufallsvariable, deren Werte alle reellen Zahlen in einem bestimmten (endlichen oder unendlichen) Intervall annehmen können, und für die eine Funktion $f_X \geq 0$ existiert, so dass

$$P(a < X < b) = \int_a^b f_X(x)dx$$

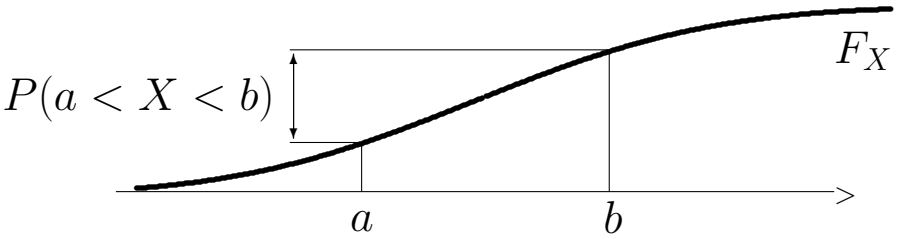
gilt, heißt stetige Zufallsvariable.

Die Funktion f_X heißt Dichtefunktion der ZV X.



Verteilungsfunktion F_X der Zufallsvariablen X

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)dy\,.$$



Im Beispiel ist also $F(495)$ die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Paket eine Füllmenge von höchstens 495 g aufweist und modelliert den Anteil aller solcher Pakete an der Gesamtproduktion dieser Maschine.
Außerdem gilt $P(495 < X < 505) = F_X(505) - F_X(495)$.

Häufig interessieren Wktn. der Form (X stetige ZV):

$P(a \leq X \leq b)$ ... zweiseitiger Toleranzbereich

$P(X \leq b)$ ... einseitiger - linksseitiger - Toleranzbereich

$P(a \leq X)$ ... einseitiger - rechtsseitiger - Toleranzbereich

Es gilt:

$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a) \\ P(X \leq b) &= F_X(b) \\ P(a < X) &= 1 - F_X(a) \end{aligned}$$

Diese Formeln gelten auch für diskrete ZV, dort ist es aber häufig einfacher, mit den Einzelwahrscheinlichkeiten zu rechnen.

<u>Hinweis:</u> Da für jeden Wert x gilt: $P(X = x) = 0$, braucht man bei stetigen ZV nicht zwischen

$P(a < X < b)$, $P(a < X \leq b)$, $P(a \leq X < b)$ und $P(a \leq X \leq b)$ (u.s.w.) zu unterscheiden!

Erwartungswert einer stetigen Zufallsvariablen X:

$$E(X) = \int_{-\infty}^{\infty} x \; f_X(x) \; dx$$

Rechenregeln:

X Zufallsvariable, dann auch $a \cdot X + b, \; a, b \in R$

$$E(aX + b) = a\, E(X) + b$$

Beispiel: Stetige gleichmäßige Verteilung über dem Intervall [a,b] aus R:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & , x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

"Rechteckverteilung"; jedes Teilintervall von [a,b] gleicher Länge hat die gleiche Wahrscheinlichkeit

$$F_X(x) = \begin{cases} 0 & , x < a \\ \dfrac{x-a}{b-a} & , x \in [a, b] \\ 1 & , x > b \end{cases}$$

$$E(X) = \int_a^b x \cdot \frac{1}{b-a} \; dx \; = \; \frac{a+b}{2}$$

<u>Streuung</u> (Varianz)

$$
\begin{aligned}
var(X) \;=\; D^2X \;=\; \sigma_X^2 \;&=\; E(X - E(X))^2 \\
&=\; \int_{-\infty}^{\infty} (x - E(X))^2 \; f_X(x) \; dx \\
&\qquad\qquad \text{! Zahl} \\
&=\; E(X^2) - (E(X))^2 \\
&\qquad \uparrow \int_{-\infty}^{\infty} x^2 f_X(x) dx
\end{aligned}
$$

<u>Beispiel:</u> Stetige gleichmäßige Verteilung über [a,b]

$$var(X) = \frac{(b-a)^2}{12}$$

Rechenregel:

$$var(a \cdot X \;+\; b) \;=\; a^2 \; var(X) \qquad (a, b \in R)$$

Die Normalverteilung

1. wichtigste stetige Verteilung
2. viele (physikalische und biologische) Größen sind (näherungsweise!) normalverteilt
   Gauß $\longrightarrow$ Beschreibung von Messfehlern
3. Bedeutung folgt aus sogenannten Grenzwertsätzen

   Summen sehr vieler kleiner (unabhängiger) Summanden $\longrightarrow$ NV

Dichtefunktion der NV: Gaußsche Glockenkurve

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

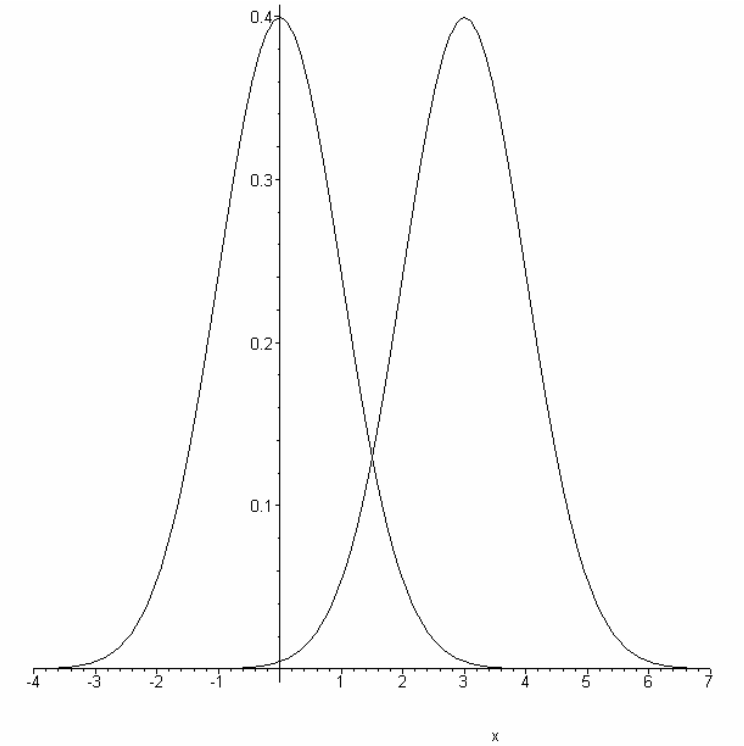
NV mit dem Parameter μ und σ

$$X \sim N(\mu, \sigma^2)$$

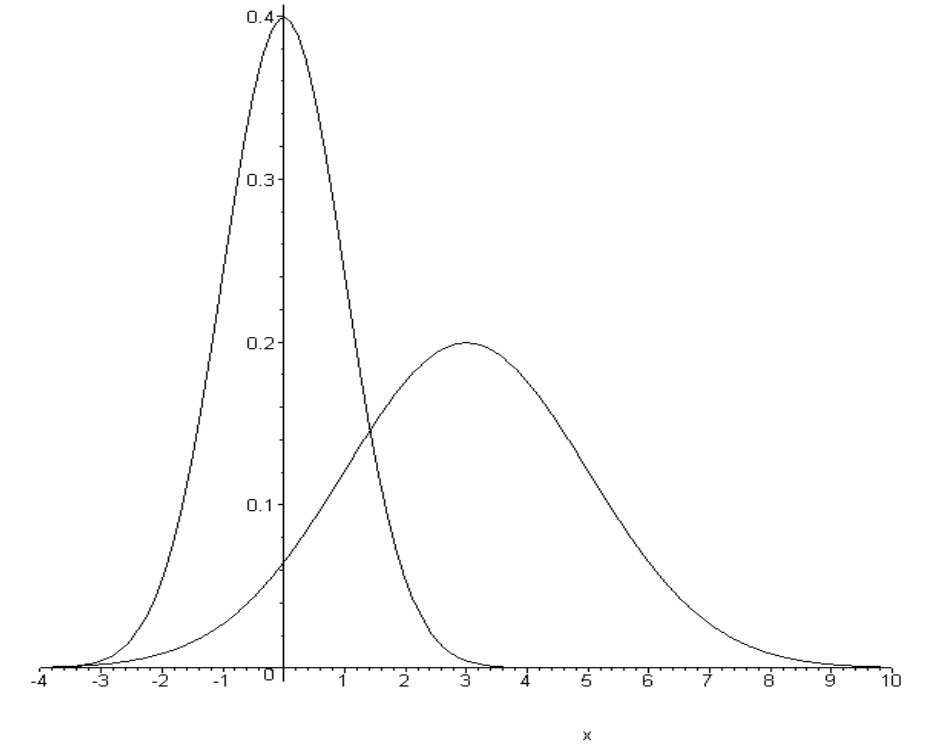
μ: Zentrum

σ: Form

$E(X) = \mu$, $var(X) = \sigma^2$



linke Kurve: Dichtefunktion einer N(0,1)-verteilten ZV
rechte Kurve: Dichtefunktion einer N(3,1)-verteilten ZV



linke Kurve: Dichtefunktion einer N(0,1)-verteilten ZV
rechte Kurve: Dichtefunktion einer N(3,4)-verteilten ZV

Es gilt: Ist $X \sim N(\mu, \sigma^2)$, dann ist

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

standard-normalverteilt: $Z \sim N(0, 1)$
mit Dichtefunktion:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

und Verteilungsfunktion:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

(Gaußsche Fehlerfunktion, Erfc)

Begriff: Standardisieren

Hinweis: Die Verteilungsfunktion einer Normalverteilung (z.B. $\Phi(x)$) ist keine mit einer Formel darstellbare elementare Funktion. Man verwendet deshalb Tafeln für $\Phi(x)$!

Bestimmung von Wahrscheinlichkeiten aus Tafeln

!!! Tafeln existieren nur für die Werte der Verteilungsfunktion $\Phi(x)$ der N(0,1)-Verteilung für positive Werte x

Ablesebeispiele:

$$\begin{aligned}\Phi(1) &= 0,841345\\ \Phi(1,1) &= 0,864334\\ \Phi(1,11) &= 0,866500\end{aligned}$$

Es gilt

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

also z.B.

$$\Phi(-1) = 1 - \Phi(1) = 1 - 0,841345 = 0,158655$$

$X \sim N(\mu, \sigma^2)$, dann

$$\begin{aligned}P(X < a) &= \Phi\left(\frac{a-\mu}{\sigma}\right)\\ P(a < X < b) &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\\ P(X > b) &= 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)\end{aligned}$$

Achtung: Division durch σ, nicht durch σ^2

Beispiel:

$X \sim N(8,\ 9)$, Gesucht $P(X < 10)$

$$P(X < 10) = \Phi\left(\frac{10-8}{3}\right) = \Phi\left(\frac{2}{3}\right) = 0,747$$

Quantile der (standardisierten) Normalverteilung

$$\Phi(z_q) \;= q$$

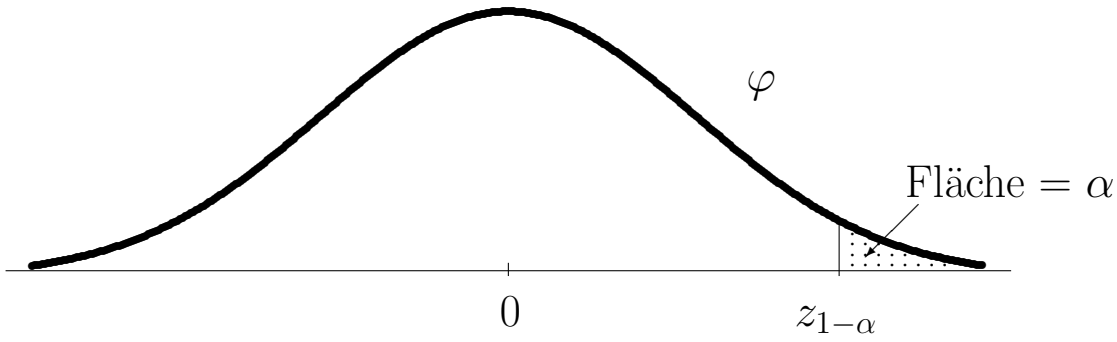
$z_q \ldots$ Quantil der Ordnung q

$$z_{0,95} \;=\; 1,645$$

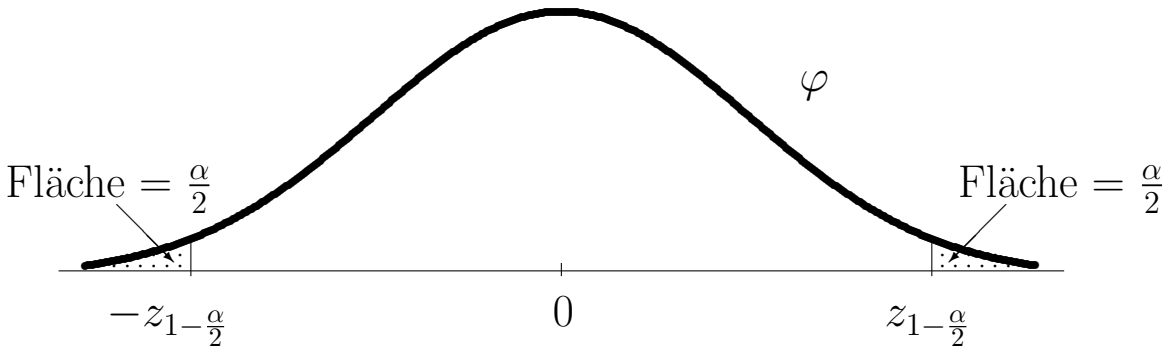
Andere Bezeichnung möglich!

Zusammenhang mit sogenannten kritischen Werten:

$z_{1-\alpha}\,, \quad \Phi(z_{1-\alpha}) \;= 1-\alpha$ ("einseitige Fragestellung"):



$z_{1-\alpha/2}\,, \quad \Phi(z_{1-\alpha/2}) \;= 1-\alpha/2$ ("zweiseitige Fragestellung"):

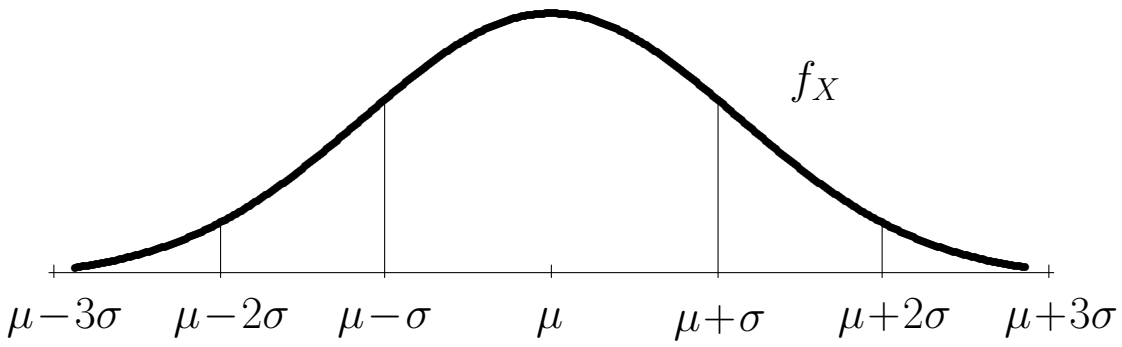


"$k \cdot \sigma$ Grenzen" für die Normalverteilung:

Für $X \sim N(\mu, \sigma^2)$ gilt

$P(|X - \mu| \leq k\sigma) = 2\Phi(k) - 1$ und für $k = 1, 2, 3$ ergibt sich

$1 \cdot \sigma$ Grenze: 0,6826 ,
$2 \cdot \sigma$ Grenze: 0,9546 ,
$3 \cdot \sigma$ Grenze: 0,9974 .



Hinweis: Im Zusammenhang mit der Behandlung von Aufgaben der schließenden Statistik benötigt man verschiedene weitere stetige Verteilungen und deren Tafeln.

Des Weiteren hat man die gemeinsame Verteilung von Zufallsvariablen und den Begriff der Unabhängigkeit von Zufallsvariablen zu definieren. Aus Zeitgründen soll der Begriff der Unabhängigkeit nur "naiv" eingeführt werden.

<u>Beispiel</u>

- Die Milchleistungen von zwei zufällig aus einer Herde ausgewählten Kühen können durch zwei unabhängige normalverteilte Zufallsvariablen modelliert werden.
- Die Milchleistung und der Fettgehalt der Milch einer zufällig aus einer Herde ausgewählten Kuh sollten durch zwei abhängige normalverteilte Zufallsvariablen modelliert werden.

Es gilt

- Sind X und Y unabhängig und normalverteilt, $X \sim N(\mu_X; \sigma_X^2)$ und $Y \sim N(\mu_Y; \sigma_Y^2)$, so ist auch $X + Y$ normalverteilt:

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

  allgemeiner gilt mit $a\,, b \in \mathbb{R}$:

  $$aX + bY \sim N(a\mu_X + b\mu_Y\,;\ a^2\sigma_X^2 + b^2\sigma_Y^2)$$

  insbesondere also z.B.:

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