Einführung

Definition: Ein **Zufallsexperiment** ist ein Experiment, das die folgenden Bedingungen erfüllt:
1. Das Experiment lässt sich auf gleiche Weise beliebig oft durchführen.
2. Das Experiment lässt mehrere, sich gegenseitig ausschließende Ergebnisse zu.
3. Das Ergebnis des Experiments lässt sich nicht vorhersagen.

Definition:
1. Die möglichen, sich gegenseitig ausschließenden Ergebnisse eines Zufallsexperiments heißen die **Elementarereignisse** $\omega_1, \omega_2, \omega_3, \dots$.
2. Die Menge der Elementarereignisse heißt die **Ergebnismenge** Ω, d.h. $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$.

Definition:
1. Jede Teilmenge A der Ergebnismenge Ω eines Zufallsexperiments heißt ein **Ereignis**.
2. Die Menge aller Ereignisse einer Ergebnismenge Ω heißt der **Ereignisraum**.

Beispiel: Das Werfen eines idealen (homogenen) Würfels ist ein Zufallsexperiment.
Die Elementarereignisse sind 1, 2, 3, 4, 5, 6.
Die Ergebnismenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$
Die Ereignisse sind alle Teilmengen von Ω, also

$\{\ \} = \emptyset$, das ist $\binom{6}{0} = 1$ Ereignis

$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$, das sind $\binom{6}{1} = 6$ Ereignisse

$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}$ ... , $\{5,4\}, \{5,6\}$, $\binom{6}{2} = 15$ Ereignisse

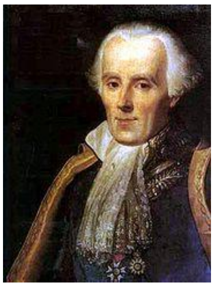
$\vdots$

$\{1, 2, 3, 4, 5, 6\}$, das ist $\binom{6}{6} = 1$ Ereignis.

Insgesamt gibt es $\binom{6}{0} + \binom{6}{1} + \binom{6}{2} + \dots + \binom{6}{6} = (1+1)^6 = 2^6 = 64$ Ereignisse.

Satz: Eine Ergebnismenge von n Elementen besitzt 2^n Ereignisse.

Definition: Ein Zufallsexperiment mit der Ergebnismenge $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ heißt **Laplace-Experiment**, wenn alle n Elementarereignisse gleichberechtigt sind.
Dann beträgt die Wahrscheinlichkeit für jedes Elementarereignis $P(\omega_i) = \frac{1}{n}$.



Pierre-Simon Marquis de Laplace, französischer Mathematiker (1749 – 1827)

Beispiel 1: Das Werfen eines idealen (homogenen) Würfels ist ein Laplace-Experiment mit $P(i) = \frac{1}{6}$ für $i = 1, \dots, 6$.

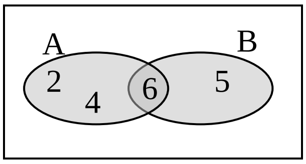
Beispiel 2: Das Werfen von zwei unterscheidbaren idealen Würfeln ist ein Laplace-Experiment. Es hat 36 Elementarereignisse, jeweils mit der Wahrscheinlichkeit $\frac{1}{36}$.

Beispiel 3: In einer Urne befinden sich 4 rote und 8 schwarze Kugeln. Man zieht eine Kugel. Dann gilt $P(r) = \frac{4}{12} = \frac{1}{3}$ und $P(s) = \frac{8}{12} = \frac{2}{3}$.

Beispiel 4: Werfen eines idealen Würfels. Die Ereignisse A und B sind definiert durch:
$A = \{2, 4, 6\}$, d.h. die Augenzahl ist gerade, mit $P(A) = \frac{3}{6} = \frac{1}{2}$.

$B = \{5, 6\}$, d.h. die Augenzahl ist größer als 4 mit $P(B) = \frac{2}{6} = \frac{1}{3}$.

Dann ist die Vereinigung gleich $A \cup B = \{2,4,5,6\}$, d.h. die Augenzahl ist gerade **oder** größer als 4 mit $P(A \cup B) = \frac{4}{6} = \frac{2}{3}$,
und der Durchschnitt gleich $A \cap B = \{6\}$, d.h. die Augenzahl ist gerade **und** größer als 4 mit $P(A \cap B) = \frac{1}{6}$.



Additionssatz für Ereignisse: $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$.

Man muss $P(A \cap B)$ einmal subtrahieren, da der Durchschnitt sowohl in A als auch in B vorkommt.

Speziell für zwei sich ausschließende Ereignisse A und B: $\boxed{A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)}$.

<u>Die Axiome von Kolmogorow</u>

Andrei Nikolajewitsch Kolmogorow (Андрей Николаевич Колмогоров), 1903-1987, russischer Mathematiker.

Gegeben ist ein Zufallsexperiment mit der Ergebnismenge Ω. Dann wird jedem Ereignis A eine reelle Zahl $P(A)$, genannt Wahrscheinlichkeit von A, zugeordnet, so dass folgende drei Axiome gelten:

1. Für alle Ereignisse A gilt $0 \le P(A) \le 1$.
2. Für das „sichere Ereignis“ gilt $P(\Omega) = 1$.
3. Für zwei sich ausschließende Ereignisse A_1 und A_2 gilt $P(A_1 \cup A_2) = P(A_1) + P(A_2)$, d.h. für zwei Ereignisse mit $A_1 \cap A_2 = \emptyset$ gilt, dass die Wahrscheinlichkeit für $A_1 \cup A_2$ gleich der Summe von $P(A_1)$ und $P(A_2)$ ist.

Folgerungen:
1. $P(\emptyset) = 0$.
   Nach 3. ist $P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset)$, also $P(\emptyset) = 0$.
2. Für das Gegenereignis (komplementäres Ereignis) $\overline{A}$ von A gilt $P(\overline{A}) = 1 - P(A)$.
   Nach 2. und 3. ist $1 = P(\Omega) = P(\overline{A} \cup A) = P(\overline{A}) + P(A)$.
3. Es gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
   Nach 3. ist $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$.
   Mit $P(A \setminus B) = P(A) - P(A \cap B)$ und $P(B \setminus A) = P(B) - P(A \cap B)$ folgt
   $P(A \cup B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

Die Binomialverteilung

Def.: Ein Zufallsexperiment heißt **Bernoulli-Experiment**, wenn es nur zwei Ausgänge (z.B. 0 oder 1) hat. Jakob Bernoulli (1655 Basel – 1705 Basel)

Beispiel: Ziehe ein Los. Es ist entweder eine Niete (0) oder ein Treffer (1).
Die Wahrscheinlichkeit für einen Treffer bezeichnen wir mit p, die Wahrscheinlichkeit für eine Niete mit q. Dabei ist $q = 1 - p$.

Das Zufallsexperiment werde nun unter gleichen Voraussetzungen n mal durchgeführt. X sei die **Zufallsvariable**, welche die Anzahl der Treffer bezeichnet.

n = 1: $P(X=0)=q$

$P(X=1)=p$

n = 2: $P(X=0)=P\big((0,0)\big)=q\cdot q=q^2$.

$P(X=1)=P\big((1,0)\vee(0,1)\big)=p\cdot q+q\cdot p=2pq$

$P(X=2)=P\big((1,1)\big)=p\cdot p=p^2$

n = 3: $P(X=0)=P\big((0,0,0)\big)=q\cdot q\cdot q=q^3$.

$P(X=1)=P\big((1,0,0)\vee(0,1,0)\vee(0,0,1)\big)=p\cdot q\cdot q+q\cdot p\cdot q+q\cdot q\cdot p=3pq^2$

$P(X=2)=P\big((1,1,0)\vee(1,0,1)\vee(0,1,1)\big)=p\cdot p\cdot q+p\cdot q\cdot p+q\cdot p\cdot p=3p^2q$

$P(X=3)=P\big((1,1,1)\big)=p\cdot p\cdot p=p^3$

n = 4: $P(X=0)=q^4=\mathbf{1}\cdot p^0q^4=\binom{\mathbf{4}}{\mathbf{0}}\cdot p^0q^4$, $P(X=1)=4pq^3=\mathbf{4}\cdot p^1q^3=\binom{\mathbf{4}}{\mathbf{1}}\cdot p^1q^3$,

$P(X=2)=6p^2q^2=\mathbf{6}\cdot p^2q^2=\binom{\mathbf{4}}{\mathbf{2}}\cdot p^2q^2$, $P(X=3)=4p^3q=\mathbf{4}\cdot p^3q^1=\binom{\mathbf{4}}{\mathbf{3}}\cdot p^3q^1$,

$P(X=4)=p^4=\mathbf{1}\cdot p^4q^0=\binom{\mathbf{4}}{\mathbf{4}}\cdot p^4q^0$.

Die Koeffizienten $\binom{n}{k}$ für $k=0,1,2,\dots,n$ heißen **Binomialkoeffizienten**. $\binom{n}{k}$ gibt an, wie viele Möglichkeiten es gibt, bei n gleichartigen Versuchen k Treffer zu erzielen.

Beispiel: $\binom{4}{2}=6$. Die sechs Möglichkeiten sind (1,1,0,0), (1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1), (0,0,1,1).

Allgemein gilt $\boxed{P(X=k)=\binom{n}{k}\cdot p^k\cdot q^{n-k}\quad \text{für}\quad k=0,1,2,\dots,n}$.

Wir suchen eine Formel für $\binom{n}{k}$: Wie viele Möglichkeiten gibt es, bei n gleichartigen Versuchen k Treffer zu erzielen?

Modell: Wir denken uns n Plätze (sie stehen für n Versuche), von denen wir k (für die Treffer) auswählen.
Für den 1. Platz gibt es n Möglichkeiten.
Für den 2. Platz gibt es dann nur noch $n-1$ Möglichkeiten.
Also gibt es für die ersten beiden Plätze $n\cdot(n-1)$ Möglichkeiten.
Für den 3. Platz gibt es dann nur noch $n-2$ Möglichkeiten.
Also gibt es für die ersten drei Plätze $n\cdot(n-1)\cdot(n-2)$ Möglichkeiten.
Für den 4. Platz gibt es dann nur noch $n-3$ Möglichkeiten.
Also gibt es für die ersten vier Plätze $n\cdot(n-1)\cdot(n-2)\cdot(n-3)$ Möglichkeiten.
Für den k-ten Platz gibt es dann nur noch $n-(k-1)$ Möglichkeiten.
Also gibt es für die k Plätze $n\cdot(n-1)\cdot(n-2)\cdot(n-3)\cdot\ldots\cdot\big(n-(k-1)\big)$ Möglichkeiten.

Dieses Produkt lässt sich vereinfachen zu

$$n\cdot(n-1)\cdot(n-2)\cdot(n-3)\cdot\ldots\cdot\big(n-(k-1)\big)=$$

$$\frac{n\cdot(n-1)\cdot(n-2)\cdot(n-3)\cdot\ldots\cdot\big(n-(k-1)\big)\cdot(n-k)\cdot(n-k-1)\cdot(n-k-2)\cdot\ \ldots\ \cdot3\cdot2\cdot1}{(n-k)\cdot(n-k-1)\cdot(n-k-2)\cdot\ \ldots\ \cdot3\cdot2\cdot1}=\frac{n!}{(n-k)!}.$$

Unter n!, gelesen „n Fakultät“, versteht man das Produkt $1\cdot2\cdot3\cdot\ldots\cdot(n-2)\cdot(n-1)\cdot n$.

Beispiele: $1!=1$, $2!=2$, $3!=6$, $4!=24$. Man definiert $\boxed{0!=1}$, was wir später noch brauchen.

Die Zahl $\frac{n!}{(n-k)!}$ ist aber noch nicht $\binom{n}{k}$.

Beispiel: $n = 4$ und $k = 3$. $\frac{n!}{(n-k)!} = \frac{4!}{1!} = 4 \cdot 3 \cdot 2 = 24$, während $\binom{4}{3} = 4$ ist.

N bedeute Niete, T bedeute Treffer. Die 24 Fälle sind hier aufgelistet.

(T_1, T_2, T_3, N), (T_1, T_3, T_2, N), (T_2, T_1, T_3, N), (T_2, T_3, T_1, N), (T_3, T_1, T_2, N), (T_3, T_2, T_1, N),
(T_1, T_2, N, T_3), (T_1, T_3, N, T_2), (T_2, T_1, N, T_3), (T_2, T_3, N, T_1), (T_3, T_1, N, T_2), (T_3, T_2, N, T_1),
(T_1, N, T_2, T_3), (T_1, N, T_3, T_2), (T_2, N, T_1, T_3), (T_2, N, T_3, T_1), (T_3, N, T_1, T_2), (T_3, N, T_2, T_1),
(N, T_1, T_2, T_3), (N, T_1, T_3, T_2), (N, T_2, T_1, T_3), (N, T_2, T_3, T_1), (N, T_3, T_1, T_2), (N, T_3, T_2, T_1).

Bei uns fallen je $k! = 3! = 6$ Möglichkeiten zusammen, die in einer Zeile stehen, da die Reihenfolge der drei Treffer keine Rolle spielt. Somit bleiben die vier Möglichkeiten (T, T, T, N), (T, T, N, T), (T, N, T, T) und (N, T, T, T) übrig.

Die gesuchte Formel lautet somit $\boxed{\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}}$.

In Dreiecksform schreibt man für diese Binomialkoeffizientenoeffizienten:

$$
\begin{array}{lccccccccc}
n = 0 & & & & & \binom{0}{0} & & & & \\
n = 1 & & & & \binom{1}{0} & & \binom{1}{1} & & & \\
n = 2 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
n = 3 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
n = 4 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}
$$

usw.

Mit dem Taschenrechner erhält man $\binom{n}{k}$ durch die Tastenfolge n $\boxed{\text{nCr}}$ k.

Beispiele:

$\binom{4}{0} = \frac{4!}{0! \cdot (4-0)!} = \frac{4!}{4!} = 1$, $\quad \binom{n}{0} = \frac{n!}{0! \cdot (n-0)!} = \frac{n!}{n!} = 1$

$\binom{4}{1} = \frac{4!}{1! \cdot (4-1)!} = \frac{4!}{3!} = 4$, $\quad \binom{n}{1} = \frac{n!}{1! \cdot (n-1)!} = \frac{n!}{(n-1)!} = n$

$\binom{4}{2} = \frac{4!}{2! \cdot (4-2)!} = \frac{4!}{2! \cdot 2!} = \frac{4 \cdot 3}{2 \cdot 1} = 6$, $\quad \binom{n}{2} = \frac{n!}{2! \cdot (n-2)!} = \frac{n \cdot (n-1)}{2}$

$\binom{4}{3} = \frac{4!}{3! \cdot (4-3)!} = \frac{4!}{3! \cdot 1!} = \frac{4 \cdot 3 \cdot 2}{3 \cdot 2 \cdot 1} = 4$, $\quad \binom{n}{3} = \frac{n!}{3! \cdot (n-3)!} = \frac{n \cdot (n-1) \cdot (n-2)}{3!} = \frac{n \cdot (n-1) \cdot (n-2)}{6}$

$\binom{4}{4} = \frac{4!}{4! \cdot (4-4)!} = \frac{4!}{4! \cdot 0!} = 1$, $\quad \binom{n}{4} = \frac{n!}{4! \cdot (n-4)!} = \frac{n \cdot (n-1) \cdot (n-2) \cdot (n-3)}{4!} = \frac{n \cdot (n-1) \cdot (n-2) \cdot (n-3)}{24}$

Man erkennt, dass die Koeffizienten gerade die Zahlen des Pascalschen Dreiecks sind:

$$
\begin{array}{lccccccccccccccc}
n = 0 & & & & & & & & 1 & & & & & & & \\
n = 1 & & & & & & & 1 & & 1 & & & & & & \\
n = 2 & & & & & & 1 & & 2 & & 1 & & & & & \\
n = 3 & & & & & 1 & & 3 & & 3 & & 1 & & & & \\
n = 4 & & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
n = 5 & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
n = 6 & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
n = 7 & 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

usw.

Blaise Pascal 1632-1662 franz. Mathematiker, Physiker und Religionsphilosoph

① An den Rändern steht je eine 1.

② Jeder Koeffizient ist die Summe der beiden darüberstehenden Koeffizienten $\boxed{\binom{n}{k-1}+\binom{n}{k}=\binom{n+1}{k}}$

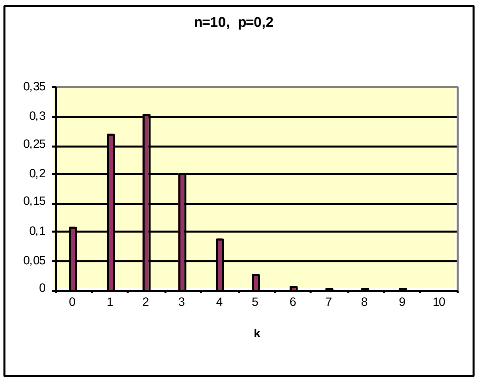
③ Das Pascalsche Dreieck ist symmetrisch $\boxed{\binom{n}{k}=\frac{n!}{k!\cdot(n-k)!}=\binom{n}{n-k}}$

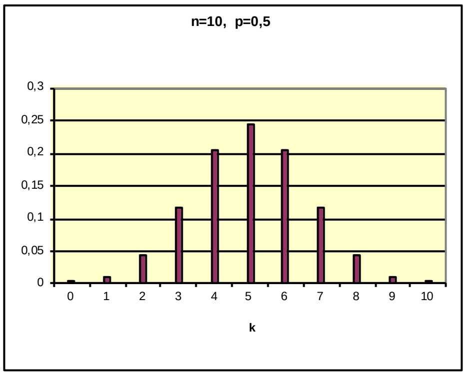
Beispiele: $\binom{n}{0}=\binom{n}{n}$, $\binom{n}{1}=\binom{n}{n-1}$, $\binom{n}{2}=\binom{n}{n-2}$, $\binom{100}{98}=\binom{100}{2}=\frac{100\cdot 99}{1\cdot 2}=4950$,

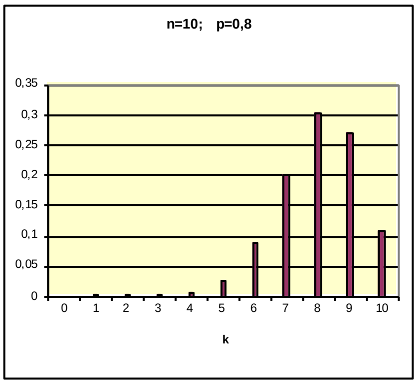
$\binom{50}{47}=\binom{50}{3}=\frac{50\cdot 49\cdot 48}{1\cdot 2\cdot 3}=19600$, $\binom{49}{43}=\binom{49}{6}=\frac{49\cdot 48\cdot 47\cdot 46\cdot 45\cdot 44}{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}=13983816$, die Anzahl unterschiedlich ausgefüllter Lotto-Scheine.

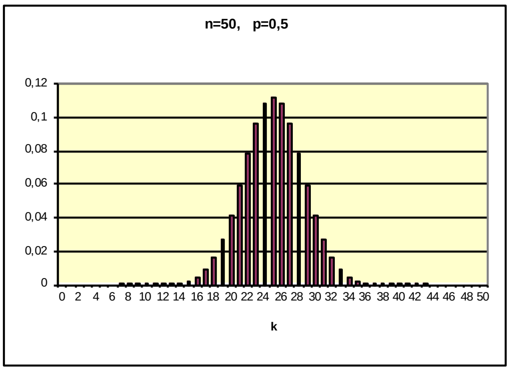
Der binomische Lehrsatz lautet: $\boxed{(a+b)^n=\binom{n}{0}a^n b^0+\binom{n}{1}a^{n-1}b^1+\binom{n}{2}a^{n-2}b^2+\ldots+\binom{n}{n}a^0 b^n}$.

Somit gilt: Wenn ein Bernoulli-Experiment n mal durchgeführt wird, dann beträgt die Wahrscheinlichkeit für genau k Treffer: $\boxed{P(X=k)=B_{n;p}(k)=\binom{n}{k}p^k q^{n-k} \quad \text{mit } q=1-p}$. Man sagt dann: „Die Zufallsvariable X ist binomialverteilt mit den Parametern n und p" oder kurz: $\boxed{X \text{ ist } B_{n;p}-\text{verteilt.}}$









Satz: Bei gegebenem n ist die Summe aller Wahrscheinlichkeiten $B_{n;p}(k)$ gleich 1, denn nach dem binomischen Lehrsatz gilt:

$$\sum_{k=0}^{n} B_{n;p}(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots + \binom{n}{n} p^n q^0 = (q+p)^n = 1^n = 1 .$$

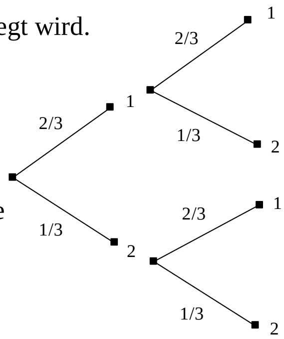
Das arithmetische Mittel wird bestimmt nach der Formel $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{1}{n}\sum_{j=1}^{m} h_j \cdot x_j = \sum_{j=1}^{m} f_j \cdot x_j = \sum_{j=1}^{m} x_j \cdot f_j$ mit den relativen Häufigkeiten f_j der Werte x_j. Überträgt man diese Formel auf die Binomialverteilung, so erhält man $\sum_{k=0}^{n} k \cdot P(X = k)$. Diese Summe nennt man den Erwartungswert $E(X) = \mu = \sum_{k=0}^{n} k \cdot P(X = k)$.

Definition: Eine **Zufallsvariable X** eines Zufallsexperiments ist eine Funktion, die jedem Ergebnis dieses Experiments eine Zahl x_i zuordnet. Z.B. X ist die Anzahl der Treffer.

Ihr Erwartungswert ist $\boxed{\mathbf{E(X) = \mu = \sum_i x_i \cdot P(X = x_i)}}$.

Beispiel: In einer Urne befinden sich drei Kugeln, welche sich nur in der Aufschrift 1, 1 und 2 unterscheiden.

a. Man zieht nacheinander zwei Kugeln, wobei die zuerst gezogene Kugel zurückgelegt wird.
Das Baumdiagramm veranschaulicht die möglichen Ergebnisse.
Es sei X die Zahl beim 1. Zug und Y die Zahl beim 2. Zug.



Es gelten die beiden **Pfadregeln:**

1. Der Multiplikationssatz: Die Wahrscheinlichkeit eines Ereignisses am Ende eines Pfades ist das Produkt der Wahrscheinlichkeiten entlang dieses Pfades.

Beispiele: $P(X=1 \wedge Y=1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$, $P(X=1 \wedge Y=2) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$,

$P(X=2 \wedge Y=1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$, $P(X=2 \wedge Y=2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

2. Der Additionssatz: Die Wahrscheinlichkeit für ein Ereignis, das aus mehreren Pfaden besteht, ist die Summe der Wahrscheinlichkeiten der einzelnen Pfade.

Beispiel: $P(Y=1) = P\big((X=1 \wedge Y=1) \vee (X=2 \wedge Y=1)\big) = P(X=1 \wedge Y=1) + P(X=2 \wedge Y=1) = \frac{4}{9} + \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$.

Die Vierfeldertafel:

$X \setminus Y$	$Y=1$	$Y=2$	$\sum$
$X=1$	$P(X=1 \wedge Y=1) = \frac{4}{9}$	$P(X=1 \wedge Y=2) = \frac{2}{9}$	$P(X=1) = \frac{2}{3}$
$X=2$	$P(X=2 \wedge Y=1) = \frac{2}{9}$	$P(X=2 \wedge Y=2) = \frac{1}{9}$	$P(X=2) = \frac{1}{3}$
$\sum$	$P(Y=1) = \frac{2}{3}$	$P(Y=2) = \frac{1}{3}$	1

Die Vierfeldertafel ist ein Spezialfall der Kontingenztafel (Kreuztabelle).
In diesem Beispiel sind die beiden Zufallsvariablen X und Y unabhängig voneinander.
Wie bei der Unabhängigkeit von Variablen (siehe bivariate Verteilungen) erhält man die vier Felder als Produkt der betreffenden Zeilensumme und der betreffenden Spaltensumme.

Z.B. $P(X=1 \wedge Y=1) = P(X=1) \cdot P(Y=1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$.

Und nun zum Erwartungswert:

Es ist $E(X) = 1 \cdot P(X=1) + 2 \cdot P(X=2) = 1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{3} = \frac{4}{3}$

Mit Hilfe der Vierfeldertafel folgt $E(Y) = 1 \cdot P(Y=1) + 2 \cdot P(Y=2) = 1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{3} = \frac{4}{3}$, wie $E(X)$.

Die Zufallsvariable $\mathbf{X} + \mathbf{Y}$ kann die Werte 2, 3 oder 4 annehmen.

$P(X+Y=2) = P(X=1 \wedge Y=1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$

$P(X+Y=3) = P\big((X=1 \wedge Y=2) \vee (X=2 \wedge Y=1)\big) = \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$

$P(X+Y=4) = P(X=2 \wedge Y=2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

Und somit $E(X+Y) = 2 \cdot \frac{4}{9} + 3 \cdot \frac{4}{9} + 4 \cdot \frac{1}{9} = \frac{8}{3}$, und das ist gleich $E(X) + E(Y)$.

Die Zufallsvariable $\mathbf{X} \cdot \mathbf{Y}$ kann die Werte 1, 2 oder 4 annehmen.

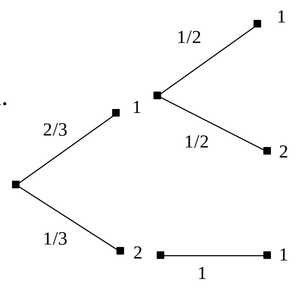
$P(X \cdot Y=1) = P(X=1 \wedge Y=1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$

$P(X \cdot Y=2) = P\big((X=1 \wedge Y=2) \vee (X=2 \wedge Y=1)\big) = \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$

$P(X \cdot Y=4) = P(X=2 \wedge Y=2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

Und somit $E(X \cdot Y) = 1 \cdot \frac{4}{9} + 2 \cdot \frac{4}{9} + 4 \cdot \frac{1}{9} = \frac{16}{9}$, und das ist gleich $E(X) \cdot E(Y)$.

b. Man zieht nacheinander zwei Kugeln, ohne die zuerst gezogene Kugel zurückzulegen.
Das Baumdiagramm veranschaulicht die möglichen Ergebnisse.
Es sei wieder X die Zahl beim 1. Zug und Y die Zahl beim 2. Zug.



Die Vierfeldertafel:
Mit Hilfe des Multiplikationssatzes erhält man aus dem Baumdiagramm:

$X \setminus Y$	$Y=1$	$Y=2$	$\sum$
$X=1$	$P(X=1 \wedge Y=1) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$	$P(X=1 \wedge Y=2) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$	$P(X=1) = \frac{2}{3}$
$X=2$	$P(X=2 \wedge Y=1) = \frac{1}{3} \cdot 1 = \frac{1}{3}$	$P(X=2 \wedge Y=2) = \frac{1}{3} \cdot 0 = 0$	$P(X=2) = \frac{1}{3}$
$\sum$	$P(Y=1) = \frac{2}{3}$	$P(Y=2) = \frac{1}{3}$	1

Die Variablen X und Y sind sicher abhängig voneinander. Dies zeigt sich auch daran, dass nicht jedes Feld das Produkt aus der betreffenden Zeilensumme und der betreffenden Spaltensumme ist.

Der Begriff der **bedingten Wahrscheinlichkeit $\mathbf{P(A\,|\,B)}$**.

$P(A\,|\,B)$ bezeichnet die Wahrscheinlichkeit für A unter der Bedingung B.

$P(Y=1\,|\,X=1) = \frac{1}{2}$ bezeichnet die Wahrscheinlichkeit für $Y=1$ unter der Bedingung $X=1$.

$P(Y=1\,|\,X=2) = 1$ bezeichnet die Wahrscheinlichkeit für $Y=1$ unter der Bedingung $X=2$.

$P(Y=2\,|\,X=1) = \frac{1}{2}$ bezeichnet die Wahrscheinlichkeit für $Y=2$ unter der Bedingung $X=1$.

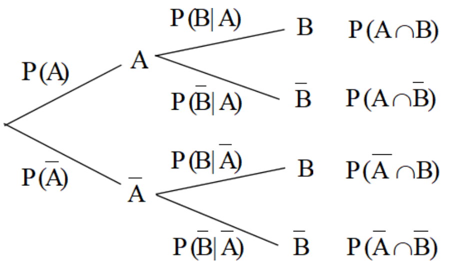
$P(Y=2\,|\,X=2) = 0$ bezeichnet die Wahrscheinlichkeit für $Y=2$ unter der Bedingung $X=2$.

$P(Y)$ hängt also vom vorhergehenden Wert von X ab.

Es ist $E(X)=1\cdot\frac{2}{3}+2\cdot\frac{1}{3}=\frac{4}{3}$ und

$E(Y)=1\cdot P(Y=1)+2\cdot P(Y=2)=1\cdot\frac{2}{3}+2\cdot\frac{1}{3}=\frac{4}{3}$, wie $E(X)$.

Allgemeine Form von Baumdiagramm und Vierfeldertafel:



	B	$\overline{B}$	$\sum$
A	$P(A\cap B)$	$P(A\cap\overline{B})$	$P(A)$
$\overline{A}$	$P(\overline{A}\cap B)$	$P(\overline{A}\cap\overline{B})$	$P(\overline{A})$
$\sum$	$P(B)$	$P(\overline{B})$	1

Allgemein gilt $\boxed{\mathbf{P(A)\cdot P(B\,|\,A)=P(A\cap B)}}$ bzw $\boxed{\mathbf{P(B\,|\,A)=\frac{P(A\cap B)}{P(A)}}}$.

Die Zufallsvariable $\mathbf{X+Y}$ kann die Werte 2 oder 3 annehmen.

$$P(X+Y=2)=P(X=1\wedge Y=1)=\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{3}$$

$$P(X+Y=3)=P\big((X=1\wedge Y=2)\vee(X=2\wedge Y=1)\big)=\frac{2}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 1=\frac{2}{3}$$

Und somit $E(X+Y)=2\cdot\frac{1}{3}+3\cdot\frac{2}{3}=\frac{8}{3}$, und das ist gleich $E(X)+E(Y)$.

Die Zufallsvariable $\mathbf{X\cdot Y}$ kann die Werte 1 oder 2 annehmen.

$$P(X\cdot Y=1)=P(X=1\wedge Y=1)=\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{3}$$

$$P(X\cdot Y=2)=P\big((X=1\wedge Y=2)\vee(X=2\wedge Y=1)\big)=\frac{2}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 1=\frac{2}{3}$$

Und somit $E(X\cdot Y)=1\cdot\frac{1}{3}+2\cdot\frac{2}{3}=\frac{5}{3}$, während $E(X)\cdot E(Y)=\frac{4}{3}\cdot\frac{4}{3}=\frac{16}{9}$ einen anderen Wert ergibt.

Satz: Für zwei **beliebige** Zufallsvariablen X und Y gilt: $\boxed{\mathbf{E(X+Y)=E(X)+E(Y)}}$.

Beweis: Die Zufallsvariable X nehme die Werte $x_1,x_2,\dots,x_n$ an, die Zufallsvariable Y nehme die Werte $y_1,y_2,\dots,y_m$ an. Dann gilt

$$E(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)\cdot P(X=x_i\wedge Y=y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}\Big[x_i\cdot P(X=x_i\wedge Y=y_j)+y_j\cdot P(X=x_i\wedge Y=y_j)\Big]=$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m}x_i\cdot P(X=x_i\wedge Y=y_j)\ +\sum_{i=1}^{n}\sum_{j=1}^{m}y_j\cdot P(X=x_i\wedge Y=y_j)=$$

$$\sum_{i=1}^{n}x_i\cdot\sum_{j=1}^{m}P(X=x_i\wedge Y=y_j)\ +\sum_{j=1}^{m}y_j\cdot\sum_{i=1}^{n}P(X=x_i\wedge Y=y_j)=$$

$$\sum_{i=1}^{n}x_i\cdot P(X=x_i)+\sum_{j=1}^{m}y_j\cdot P(Y=y_j)\ =\ E(X)+E(Y).\ \text{q.e.d.}$$

Die Zufallsvariable X sei $B_{n;p}$ -verteilt, d.h. X gibt die Anzahl der Treffer bei n Versuchen an, wobei p die Wahrscheinlichkeit für einen Treffer bedeutet. Dann ist $X = X_1 + X_2 + \ldots + X_n$, wobei für die Zufallsvariable X_i gilt:

$X_i = \begin{cases} 1 & \text{falls Versuch Nr. i ein Treffer ist} \\ 0 & \text{falls Versuch Nr. i eine Niete ist} \end{cases}$. Dann ist

$E(X_i) = \mu_i = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = 1 \cdot p = p$, so dass $E(X) = \mu = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = n \cdot p$.

Dieses Ergebnis war zu erwarten, denn bei n Versuchen mit der jeweiligen Trefferwahrscheinlichkeit p erwartet man insgesamt $n \cdot p$ Treffer. Also gilt: $\boxed{\textbf{X sei } \mathbf{B_{n;p}}\textbf{-verteilt} \quad \Rightarrow \quad \mathbf{E(X) = \mu = n \cdot p}}$.

Bemerkung: Betrachten wir noch einmal die beiden Beispiele a. (mit Zurücklegen) und b. (ohne Zurücklegen). Im Beispiel a. sind die beiden Zufallsvariablen X und Y unabhängig voneinander. Das Ergebnis des zweiten Zuges hängt nicht vom Ergebnis des ersten Zuges ab. $P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$

Im Beispiel b. sind die beiden Zufallsvariablen X und Y abhängig voneinander. Das Ergebnis des zweiten Zuges hängt vom Ergebnis des ersten Zuges ab. $P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j \mid X = x_i)$

Daraus folgt für die **bedingte Wahrscheinlichkeit** $\boxed{P(Y = y_j \mid X = x_i) = \dfrac{P(X = x_i \wedge Y = y_j)}{P(X = x_i)}}$

Definition: Zwei **Zufallsvariablen** X, Y heißen (stochastisch) **unabhängig**, falls

$\boxed{P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j)}$ für alle möglichen i, j.

Diese Definition in symmetrisch in X und Y. Somit folgt für alle möglichen i, j

$P(Y = y_j \mid X = x_i) = \dfrac{P(X = x_i \wedge Y = y_j)}{P(X = x_i)} = \dfrac{P(X = x_i) \cdot P(Y = y_j)}{P(X = x_i)} = P(Y = y_j)$ ist unabhängig von $X = x_i$,

$P(X = x_i \mid Y = y_j) = \dfrac{P(X = x_i \wedge Y = y_j)}{P(Y = y_j)} = \dfrac{P(X = x_i) \cdot P(Y = y_j)}{P(Y = x_j)} = P(X = x_i)$ ist unabhängig von $Y = y_j$.

Beispiel 1: Für die Zufallsvariablen X und Y gelte $P(X = x_1) = 0,4$, $P(Y = y_1) = 0,2$, $P(Y = y_2) = 0,5$. Man bestimme $P(X = x_i \wedge Y = y_j)$ so, dass X und Y unabhängig sind.

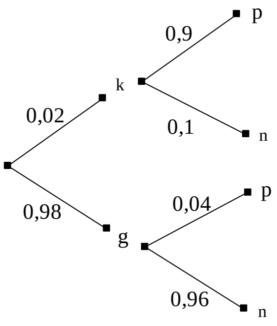
	$Y = y_1$	$Y = y_2$	$Y = y_3$	$\sum$
$X = x_1$	$P(X = x_1 \wedge Y = y_1)$ = 0,08	$P(X = x_1 \wedge Y = y_2)$ = 0,20	$P(X = x_1 \wedge Y = y_3)$ = 0,12	$P(X = x_1) = 0,4$
$X = x_2$	$P(X = x_2 \wedge Y = y_1)$ = 0,12	$P(X = x_2 \wedge Y = y_2)$ = 0,30	$P(X = x_2 \wedge Y = y_3)$ = 0,18	$P(X = x_2) = 0,6$
$\sum$	$P(Y = y_1) = 0,2$	$P(Y = y_2) = 0,5$	$P(Y = y_3) = 0,3$	1

Für die Unabhängigkeit von X und Y muss $P(X = x_i \wedge Y = y_j) = P(X = x_1) \cdot P(Y = y_j)$ sein. Dann ist z.B.

$P(X = x_1 \mid Y = y_1) = \dfrac{P(X = x_1 \wedge Y = y_1)}{P(Y = y_1)} = \dfrac{P(X = x_1) \cdot P(Y = y_1)}{P(Y = y_1)} = P(X = x_1) = 0,4$ unabhängig von Y.

Beispiel 2: In einer Bevölkerung haben durchschnittlich 2% aller Personen eine bestimmte Krankheit. Ein medizinischer Test erkennt in 90% aller Fälle diese Krankheit; bei den Gesunden zeigt der Test in 4% aller Fälle irrtümlich diese Krankheit an.
Es bedeute: $k \mathrel{\widehat{=}} \text{krank}$, $g \mathrel{\widehat{=}} \text{gesund}$, $p \mathrel{\widehat{=}} \text{Test positiv}$, $n \mathrel{\widehat{=}} \text{Test negativ}$
Zeichnen Sie ein Baumdiagramm und eine Vierfeldertafel.

Baumdiagramm



Vierfeldertafel

	p	n	
k	$0{,}02\cdot 0{,}9 = 0{,}018$	$0{,}02\cdot 0{,}1 = 0{,}002$	0,02
g	$0{,}98\cdot 0{,}04 = 0{,}0392$	$0{,}98\cdot 0{,}96 = 0{,}9408$	0,98
	0,0572	0,9428	1

$P(k\mid p) = \dfrac{P(k\wedge p)}{P(p)} = \dfrac{0{,}018}{0{,}0572} \approx 0{,}3147$, d.h. von den positiv getesteten Personen sind nur 31,47% krank.

$P(g\mid p) = \dfrac{P(g\wedge p)}{P(p)} = \dfrac{0{,}0392}{0{,}0572} \approx 0{,}6853$, d.h. von den positiv getesteten Personen sind 68,53% gesund.

Diese Zahl ist erstaunlich groß. Das liegt daran, dass von der großen Zahl der gesunden Personen 4% als krank getestet werden. Man müsste den Test dahingehend verbessern, diese 4% zu verringern.

$P(k\mid n) = \dfrac{P(k\wedge n)}{P(n)} = \dfrac{0{,}002}{0{,}9428} \approx 0{,}0021$, d.h. von den negativ getesteten Personen sind nur 0,21% krank

$P(g\mid n) = \dfrac{P(g\wedge n)}{P(n)} = \dfrac{0{,}9408}{0{,}9428} \approx 0{,}9979$, d.h. von den negativ getesteten Personen sind 99,79% gesund.

Satz: Für zwei **unabhängige** Zufallsvariablen X und Y gilt: $\boxed{\mathbf{E(X\cdot Y) = E(X)\cdot E(Y)}}$.

Beweis: $E(X\cdot Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i\cdot y_j)\cdot P(X=x_i \wedge Y=y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[x_i \cdot P(X=x_i)\cdot y_j\cdot P(Y=y_j)\right] =$

$$\underbrace{\sum_{i=1}^{n} x_i\cdot P(X=x_i)}_{=E(X)} \cdot \underbrace{\sum_{j=1}^{m} y_j \cdot P(Y=y_j)}_{=E(Y)} = E(X)\cdot E(Y) \quad \text{q.e.d.}$$

Und nun zur **Varianz**: $\boxed{s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(a_i-\bar{x}\right)^2 = \frac{1}{n}\sum_{j=1}^{m} h_j\cdot\left(x_j-\bar{x}\right)^2 = \sum_{j=1}^{m} f_j\cdot\left(x_j-\bar{x}\right)^2 = \sum_{j=1}^{m} f_j\cdot x_j^2 - \bar{x}^2 = \overline{x^2}-\bar{x}^2}$.

Wir übertragen diese Formeln auf die Wahrscheinlichkeitsrechnung:

Varianz $V(X) = Var(X) = \sum_i \left(x_i - E(X)\right)^2\cdot P(X=x_i) = \sum_i\left(x_i^2 - 2\cdot x_i\cdot E(X) + E(X)^2\right)\cdot P(X=x_i) =$

$$\underbrace{\sum_i x_i^2\cdot P(X=x_i)}_{=E\left(X^2\right)} - 2E(X)\cdot\underbrace{\sum_i x_i\cdot P(X=x_i)}_{=E(X)} + E(X)^2\cdot\underbrace{\sum_i P(X=x_i)}_{=1} = E\left(X^2\right) - E(X)^2 .$$

so erhält man

$$\boxed{\begin{array}{l} Var(X) = \sigma^2 = \sum_i\left(x_i - E(X)\right)^2\cdot P(X=x_i) = E(X^2) - E(X)^2 \quad \text{(Verschiebungssatz der Varianz)} \\ \text{oder auch} \quad Var(X) = \sigma^2 = \sum_i\left(x_i - E(X)\right)^2\cdot P(X=x_i) = E\left(\left(X - E(X)\right)^2\right) \end{array}}.$$

Satz: Für zwei **unabhängige** Zufallsvariablen X und Y gilt $\boxed{\mathbf{Var(X+Y)=Var(x)+Var(Y)}}$.

Beweis: $\mathrm{Var}(X+Y)=E\left((X+Y)^2\right)-E(X+Y)^2=E\left(X^2+2XY+Y^2\right)-\left(E(X)+E(Y)\right)^2=$

$E(X^2)+2\cdot E(X\cdot Y)+E(Y)^2-E(X)^2-2\cdot E(X)\cdot E(Y)-E(Y)^2=E(X^2)-E(X)^2+E(Y)^2-E(Y)^2=$
$=\mathrm{Var}(X)+\mathrm{Var}(Y)$, da wegen der Unabhängigkeit $E(X\cdot Y)=E(X)\cdot E(Y)$ gilt. q.e.d.

Die Zufallsvariable X sei nun wieder $B_{n;p}$-verteilt, d.h. X gibt die Anzahl der Treffer bei n Versuchen an, wobei p die Wahrscheinlichkeit für einen Treffer bedeutet. Dann ist $X=X_1+X_2+\ldots+X_n$, wobei die Zufallsvariablen X_i unabhängig sind und nur die Werte 0 oder 1 annehmen. Dann folgt wegen $E(X_i)=\mu_i=p$

$V(X_i)=(1-\mu_i)^2\cdot p+(0-\mu_i)^2\cdot q=(1-p)^2\cdot p+(0-p)^2\cdot q=q^2\cdot p+p^2\cdot q=p\cdot q\cdot(q+p)=p\cdot q$, oder

$V(X_i)=E(X_i{}^2)-E(X_i)^2=0^2\cdot q+1^2\cdot p-p^2=p-p^2=p\cdot(1-p)=p\cdot q$. Somit folgt

$V(X)=V(X_1+X_2+\ldots+X_n)=V(X_1)+V(X_2)+V(X_n)=n\cdot p\cdot q$.

Und damit gilt $\boxed{\mathbf{X\ sei\ B_{n;p}\text{-}verteilt} \;\Rightarrow\; \mathbf{Var(X)=\sigma^2=n\cdot p\cdot q}}$.

Beispiel 1: Bei einer statistischen Erhebung wird festgestellt, dass an einem Stichtag in einer Bevölkerung jede 20. Person krank oder verletzt ist. Wie groß darf am Stichtag eine Gesamtheit von Personen höchstens sein, wenn die Wahrscheinlichkeit, dass alle Personen dieser Gesamtheit gesund sind, mindestens 80% betragen soll?

k	$P(X=k)$
7	0,089561677
8	0,113719761
9	0,127685345
10	0,128357373
11	0,116688521
12	0,096728643
13	0,073623420
$\sum$	0,74636474

Lösung: Die Zufallsvariable X bezeichne die Anzahl der kranken Personen und $p=1/20=0,05$ die Wahrscheinlichkeit, krank zu sein. Der Ansatz lautet $P(X=0)\geq 0,8$, d.h. $B_{n;0,05}(0)\geq 0,8$, wobei n gesucht ist. Es folgt $\binom{n}{0}0,05^0\cdot 0,95^{\,n}\geq 0,8$, d.h. $0,95^n\geq 0,8$. Logarithmieren liefert $n\cdot\ln 0,95\geq\ln 0,8$ und nach Division durch $\ln 0,95$, was negativ ist, folgt $n\leq\frac{\ln 0,8}{\ln 0,95}\approx 4,35$. Man darf also höchstens 4 Personen befragen. Wenn mehr als 4 befragt werden, dann sinkt die Wahrscheinlichkeit, dass ausschließlich Gesunde ausgewählt werden auf unter 80%.

Beispiel 2: In einer Fabrik werden die hergestellten Teile von einem Kontrolleur überprüft, der jeweils mit einer Wahrscheinlichkeit von 95% richtig beurteilt, ob das Teil normgerecht ist. Es sei X die Anzahl der falschen Entscheidungen des Kontrolleurs bei 200 untersuchten Teilen.

a. Berechnen Sie den Erwartungswert μ_X und die Standardabweichung σ_X.

b. Mit welcher Wahrscheinlichkeit liegt die Anzahl der falsch beurteilten Teile im Intervall $[\mu_X-\sigma_X/\mu_X+\sigma_X]$?

Lösung: Zur vorgegebenen Zufallsvariablen X gehört die Wahrscheinlichkeit $p=0,05$, dass ein Teil falsch beurteilt wird.

a. $\mu_X=n\cdot p=200\cdot 0,05=10$. Im Mittel wird der Kontrolleur bei 200 untersuchten Teilen 10 falsch beurteilen. $\sigma_X=\sqrt{npq}=\sqrt{200\cdot 0,05\cdot 0,95}\approx 3,08$.

b. Das gefragte Intervall ist also $[6,9/13,1]$, d.h. wir müssen $P(7\leq X\leq 13)$ bestimmen.

Es ist $\sum_{k=7}^{13}P(X=k)=\sum_{k=7}^{13}B_{200;0,05}(k)=\sum_{k=7}^{13}\binom{200}{k}\cdot 0,05^k\cdot 0,95^{200-k}\approx 0,746$.

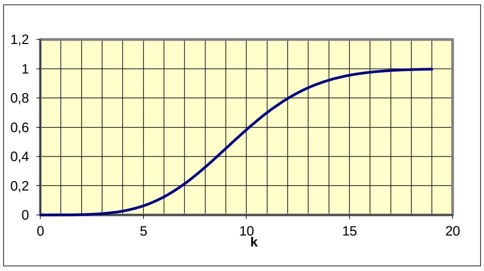
Es gibt aber auch Tafeln, in denen die Summenverteilung angegeben ist: $P(X\leq 13)=0,87010777$

und $P(X\leq 6)=0,12374303$,

also $P(7\leq X\leq 13)=0,87010777-0,12374303=0,74636474$.

Oder man verwendet die Verteilungsfunktion

$P(X\leq k)=\sum_{i=0}^{k}B_{200;0,05}(i)$ im Schaubild.



Zusätze zur Varianz und Kovarianz

Für die **Varianz** hatten wir die Formel $s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2$ für eine Zufallsvariable X

übertragen zu $\boxed{\mathrm{Var}(X) = \sigma^2 = E\left(\left(X - E(x)\right)^2\right) = \sum_{i=1}^{n}\left(x_i - E(X)\right)^2 \cdot P(X = x_i) = E(X^2) - E(X)^2}$.

Für zwei Zufallsvariablen X, Y ist die **Kovarianz** definiert durch

$$\boxed{\mathrm{Cov}(X,Y) = \sigma_{XY}^2 = E\left[\left(X - E(x)\right)\cdot\left(Y - E(Y)\right)\right] = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i - E(X)\right)\cdot\left(y_j - E(Y)\right)\cdot P(X = x_i \wedge Y = y_j) = E(X\cdot Y) - E(X)\cdot E(Y)}$$

Beispiel:

	$x_1 = 1$	$x_2 = 2$	$x_3 = 3$	
$y_1 = 4$	0,1	0,2	0,1	0,4
$y_2 = 5$	0	0,4	0,2	0,6
	0,1	0,6	0,3	1

Mit $E(X) = 1\cdot 0{,}1 + 2\cdot 0{,}6 + 3\cdot 0{,}3 = 2{,}2$ und $E(Y) = 4\cdot 0{,}4 + 5\cdot 0{,}6 = 4{,}6$ folgt

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{3}\sum_{j=1}^{2}\left(x_i - E(X)\right)\cdot\left(y_j - E(Y)\right)\cdot P(X = x_i \wedge Y = y_j) =$$
$$= (1-2{,}2)\cdot(4-4{,}6)\cdot 0{,}1 + (1-2{,}2)\cdot(5-4{,}6)\cdot 0 +$$
$$+ (2-2{,}2)\cdot(4-4{,}6)\cdot 0{,}2 + (2-2{,}2)\cdot(5-4{,}6)\cdot 0{,}4 +$$
$$+ (3-2{,}2)\cdot(4-4{,}6)\cdot 0{,}1 + (3-2{,}2)\cdot(5-4{,}6)\cdot 0{,}2 =$$
$$= 0{,}072 - 0{,}008 + 0{,}016 = 0{,}08 .$$

Andererseits ist $E(X\cdot Y) = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i \cdot y_j \cdot P(X = x_i \wedge Y = y_j) =$

$$= \left(1\cdot 4\cdot 0{,}1 + 1\cdot 5\cdot 0\right) + \left(2\cdot 4\cdot 0{,}2 + 2\cdot 5\cdot 0{,}4\right) + \left(3\cdot 4\cdot 0{,}1 + 3\cdot 5\cdot 0{,}2\right) = 0{,}4 + 5{,}6 + 7{,}2 = 10{,}2$$

Insgesamt folgt $\mathrm{Cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y) = 10{,}2 - 2{,}2\cdot 4{,}6 = 0{,}08$ wie oben.

Bemerkung: Bei der linearen Regression gibt es nur Paare mit gleichen Indizes, also (x_1, y_1), (x_2, y_2), …, bzw. $P(X = x_i \wedge Y = y_j) = 0$ für $i \neq j$, so dass sich die Doppelsumme zur Einfachsumme vereinfacht:

$$\sum_i\sum_j\left(x_i - E(X)\right)\cdot\left(y_j - E(Y)\right)\cdot P(X = x_i \wedge Y = y_j) = \sum_i\left(x_i - E(X)\right)\cdot\left(y_i - E(Y)\right)\cdot P(X = x_i \wedge Y = y_i) .$$

Bei n Punkten wird dies $s_{XY} = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)\cdot\left(y_i - \overline{y}\right)\cdot\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\cdot\left(y_i - \overline{y}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x}\cdot\overline{y} = \overline{x\cdot y} - \overline{x}\cdot\overline{y}$.

Satz: Für zwei Zufallsvariablen X, Y gilt:

1. $E(X+Y) = E(X) + E(Y)$, $E(a\cdot X + b\cdot Y + c) = a\cdot E(X) + b\cdot E(Y) + c$ für $a,b,c \in \mathbb{R}$.
2. $E(X\cdot Y) = E(X)\cdot E(Y) + \mathrm{Cov}(X,Y)$, bzw. $\mathrm{Cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y)$
3. $V(X \pm Y) = V(X) + V(Y) \mp 2\cdot \mathrm{Cov}(X,Y)$, $V(a\cdot X + b) = a^2 \cdot V(x)$
4. **Wenn X, Y unabhängig sind, dann gilt $\mathbf{E(X\cdot Y) = E(X)\cdot E(Y)}$, $\mathbf{V(X\pm Y) = V(X) + V(Y)}$ und $\mathbf{Cov(X,Y) = 0}$. Nicht umkehrbar!** (Siehe Aufgaben)
5. $\mathrm{Cov}(X,X) = V(X)$, $\mathrm{Cov}(Y,X) = \mathrm{Cov}(X,Y)$, $\mathrm{Cov}(a\cdot X + b, c\cdot Y + d) = a\cdot c\cdot \mathrm{Cov}(X,Y)$, $\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$
6. $|\mathrm{Cov}(X,Y)| \le \mathrm{Var}(X)\cdot \mathrm{Var}(Y)$ bzw. $-\mathrm{Var}(X)\cdot\mathrm{Var}(Y) \le \mathrm{Cov}(X,Y) \le \mathrm{Var}(X)\cdot\mathrm{Var}(Y)$.

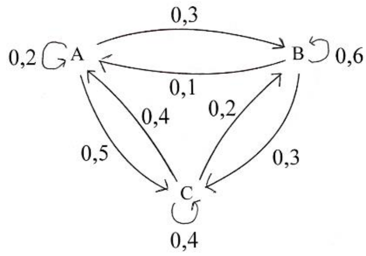
   Der Bruch $\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}$, genannt **Korrelationskoeffizient**, liegt zwischen -1 und 1 und ist ein Maß für den Zusammenhang zwischen X und Y.

Markov-Ketten

(Andrei Andrejewitsch Markov,(Андрей Андреевич Марков) russischer Mathematiker, 1856 – 1922)

Gegeben sei ein System, das mehrere Zustände annehmen kann. Die Wahrscheinlichkeiten eines Wechsels von einem Zustand in einen anderen möglichen Zustand dürfen nur vom momentanen Zustand und nicht von den früheren Zuständen abhängen. Dies ist die sog. Markov-Bedingung.

Beispiel: In einer Stadt gebe es drei Fahrradstationen A, B und C. An jeder kann ein Fahrrad ausgeliehen werden, und es ist an einer der drei Stationen wieder zurückzugegeben. Die jeweiligen Übergangswahrscheinlichkeiten können dem Diagramm (Prozessdiagramm) entnommen werden. Dabei gilt für jede Station, dass die Summen der abgehenden Wahrscheinlichkeiten immer 1 ist.
Die Übergangswahrscheinlichkeiten lassen sich auch in einer sog. Übergangsmatrix darstellen.



$$P = \begin{pmatrix} P(AA) & P(BA) & P(CA) \\ P(AB) & P(BB) & P(CB) \\ P(AC) & P(BC) & P(CC) \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix}.$$ Man beachte die Anordnung.

Dabei bedeutet z.B. $P(AB) = 0,3$, dass ein in A ausgeliehenes Fahrrad mit der Wahrscheinlichkeit 0,3 in B wieder zurückgegeben wird. Diese Wahrscheinlichkeit hängt natürlich nicht davon ab, an welchen Stationen das Fahrrad zuvor ausgeliehen war, so dass die Markov-Bedingung erfüllt ist.
Man kann $P(AB)$ auch interpretieren als bedingte Wahrscheinlichkeit $P(B \mid A)$, dass das Fahrrad in B zurückgegeben wird unter der Bedingung, dass es zuvor in A stand.
In der Übergangsmatrix beträgt jede Spaltensumme gleich 1, da ein ausgeliehenes Fahrrad an einer der drei Stationen zurückgegeben werden muss.

Nun wird ein bestimmtes Fahrrad betrachtet. Es befinde sich mit den Wahrscheinlichkeiten p_A, p_B, p_C in einer der Stationen A, B, C, wobei $p_A + p_B + p_C = 1$ gilt. Mit welchen Wahrscheinlichkeiten p'_A, p'_B, p'_C befindet es sich nach der Rückgabe in den Stationen A, B oder C?

Es gilt $p'_A = P(AA)\cdot p_A + P(BA)\cdot p_B + P(CA)\cdot p_C = 0,2\cdot p_A + 0,1\cdot p_B + 0,4\cdot p_C$
und $p'_B = P(AB)\cdot p_A + P(BB)\cdot p_B + P(CB)\cdot p_C = 0,3\cdot p_A + 0,6\cdot p_B + 0,2\cdot p_C$
und $p'_C = P(AC)\cdot p_A + P(BC)\cdot p_B + P(CC)\cdot p_C = 0,5\cdot p_A + 0,3\cdot p_B + 0,4\cdot p_C$.

Diese Rechnungen lassen sich übersichtlich als Matrizenmultiplikation darstellen:

$$\begin{pmatrix} p'_A \\ p'_B \\ p'_C \end{pmatrix} = P \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix} = \begin{pmatrix} P(AA) & P(BA) & P(CA) \\ P(AB) & P(BB) & P(CB) \\ P(AC) & P(BC) & P(CC) \end{pmatrix} \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix} \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}.$$

In der Literatur wird die Übergangsmatrix P auch in transponierter Form geschrieben. Dann muss die Multiplikationsreihenfolge geändert werden:

$$(p'_A \;\; p'_B \;\; p'_C) = (p_A \;\; p_B \;\; p_C) \bullet P = (p_A \;\; p_B \;\; p_C) \bullet \begin{pmatrix} P(AA) & P(AB) & P(AC) \\ P(BA) & P(BB) & P(BC) \\ P(CA) & P(CB) & P(CC) \end{pmatrix} = (p_A \;\; p_B \;\; p_C) \bullet \begin{pmatrix} 0,2 & 0,3 & 0,5 \\ 0,1 & 0,6 & 0,3 \\ 0,4 & 0,2 & 0,4 \end{pmatrix}.$$

Definition: Eine Markov-Kette heißt **irreduzibel**, wenn es möglich ist, von jedem Zustand in endlich vielen Schritten in jeden Zustand zu wechseln.
Die oben gegebene Markov-Kette ist irreduzibel.

Definition: Eine Markov-Kette heißt **aperiodisch**, wenn für jeden Zustand gilt: Die Anzahlen aller möglichen Schritte, um zum gleichen Zustand zurückzukehren, haben den größten gemeinsamen Teiler 1, d.h. ggT=1.

Die oben gegebene Markov-Kette ist aperiodisch. Um z.B. von A aus wieder zu A zurückzukehren, benötigt man 1 oder 2 oder 3, .. Schritte, so dass der ggT dieser Anzahlen gleich 1 ist.

Ein Fahrrad befinde sich nun in Station A. Mit welchen Wahrscheinlichkeiten ist es nach der Rückgabe in A, B oder C anzutreffen?

$\begin{pmatrix} p'_A \\ p'_B \\ p'_C \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,2 \\ 0,3 \\ 0,5 \end{pmatrix}$. Das sind gerade die gegebenen Wahrscheinlichkeiten, siehe Schaubild.

Mit welchen Wahrscheinlichkeiten ist dieses Fahrrad nach der zweiten Rückgabe in A, B oder C anzutreffen?

$\begin{pmatrix} p''_A \\ p''_B \\ p''_C \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix} \bullet \begin{pmatrix} 0,2 \\ 0,3 \\ 0,5 \end{pmatrix} = \begin{pmatrix} 0,27 \\ 0,34 \\ 0,39 \end{pmatrix}$ oder

$\begin{pmatrix} p''_A \\ p''_B \\ p''_C \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix}^2 \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,27 & 0,20 & 0,26 \\ 0,34 & 0,45 & 0,32 \\ 0,39 & 0,35 & 0,42 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,27 \\ 0,34 \\ 0,39 \end{pmatrix}$.

Für die 3. Rückgabe gilt $\begin{pmatrix} p'''_A \\ p'''_B \\ p'''_C \end{pmatrix} = \begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix}^3 \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,244 & 0,225 & 0,252 \\ 0,363 & 0,400 & 0,354 \\ 0,393 & 0,375 & 0,394 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,244 \\ 0,363 \\ 0,393 \end{pmatrix}$

Die Frage ist nun: Wo ist das Fahrrad nach beliebig vielen Rückgaben anzutreffen?

Dazu bestimmt man zunächst die **stationäre Verteilung** $\begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}$, für die $P \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix} = \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}$ gilt.

Aus $\begin{pmatrix} 0,2 & 0,1 & 0,4 \\ 0,3 & 0,6 & 0,2 \\ 0,5 & 0,3 & 0,4 \end{pmatrix} \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix} = \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}$ folgt das LGS $\begin{array}{rcrcrcl|l|l} -0,8p_A & + & 0,1p_B & + & 0,4p_C & = & 0 & \cdot 0,3 & \cdot 1 \\ 0,3p_A & - & 0,4p_B & + & 0,2p_C & = & 0 & \cdot 0,8 & \cdot 1 \\ 0,5p_A & + & 0,3p_B & - & 0,6p_C & = & 0 & & \cdot 1 \end{array}$

also $\begin{array}{rcrcrcl} -0,8p_A & + & 0,1p_B & + & 0,4p_C & = & 0 \\ & - & 0,29p_B & + & 0,28p_C & = & 0 \\ & & & & 0 & = & 0 \end{array}$. Dazu kommt noch die Gleichung $p_A + p_B + p_C = 1$.

Das Ergebnis ist $p_A = \frac{6}{25} = 0,24$, $p_B = \frac{28}{75} = 0,37\overline{3}$, $p_C = \frac{29}{75} = 0,38\overline{6}$.

Ohne Beweis: $\overline{P} = \lim\limits_{n\to\infty} P^n = \begin{pmatrix} \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{28}{75} & \frac{28}{75} & \frac{28}{75} \\ \frac{29}{75} & \frac{29}{75} & \frac{29}{75} \end{pmatrix}$ enthält gerade dreimal den stationären Spaltenvektor $\begin{pmatrix} \overline{p}_A \\ \overline{p}_B \\ \overline{p}_C \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \\ \frac{28}{75} \\ \frac{29}{75} \end{pmatrix}$.

Somit befindet sich unser ursprünglich in A stehendes Fahrrad nach beliebig vielen Rückgaben mit $\overline{p}_A = \frac{6}{25} = 0,24$, $\overline{p}_B = \frac{28}{75} = 0,37\overline{3}$, $\overline{p}_C = \frac{29}{75} = 0,38\overline{6}$ in Station A, B oder C.

Die Matrix $\overline{P} = \lim\limits_{n\to\infty} P^n$ heißt auch **Gleichgewichtsmatrix**.

Dasselbe Ergebnis folgt für ein Fahrrad mit beliebigem Ausgangspunkt, denn $\begin{pmatrix} \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{28}{75} & \frac{28}{75} & \frac{28}{75} \\ \frac{29}{75} & \frac{29}{75} & \frac{29}{75} \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \\ \frac{28}{75} \\ \frac{29}{75} \end{pmatrix}$,

$\begin{pmatrix} \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{28}{75} & \frac{28}{75} & \frac{28}{75} \\ \frac{29}{75} & \frac{29}{75} & \frac{29}{75} \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \\ \frac{28}{75} \\ \frac{29}{75} \end{pmatrix}$ und ebenso $\begin{pmatrix} \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{28}{75} & \frac{28}{75} & \frac{28}{75} \\ \frac{29}{75} & \frac{29}{75} & \frac{29}{75} \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \\ \frac{28}{75} \\ \frac{29}{75} \end{pmatrix}$.

Es gilt sogar $\begin{pmatrix} \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{28}{75} & \frac{28}{75} & \frac{28}{75} \\ \frac{29}{75} & \frac{29}{75} & \frac{29}{75} \end{pmatrix} \bullet \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \cdot (p_A + p_B + p_C) \\ \frac{28}{75} \cdot (p_A + p_B + p_C) \\ \frac{29}{75} \cdot (p_A + p_B + p_C) \end{pmatrix} = \begin{pmatrix} \frac{6}{25} \\ \frac{28}{75} \\ \frac{29}{75} \end{pmatrix}$, da $p_A + p_B + p_C = 1$ gilt, d.h. nach langer Zeit befinden sich $\frac{6}{25} = 24\%$ aller Fahrräder in Station A, $\frac{28}{75} \approx 37{,}3\%$ aller Fahrräder in Station B und $\frac{29}{75} \approx 38{,}7\%$ aller Fahrräder in Station C.

Satz: Zu jeder irreduziblen und aperiodischen Markov-Kette gibt es mindestens eine stationäre Verteilung, also eine Gleichgewichtsmatrix. (Ohne Beweis)

Die Normalverteilung

In unseren Beispielen haben wir gesehen, dass je nach n und p unterschiedliche Verteilungsfunktionen entstehen. Die Frage ist, ob es nicht näherungsweise eine universellere Verteilung gibt, die unabhängig von n und p anwendbar ist.

Def.: Eine Zufallsvariable mit dem Erwartungswert 0 und der Varianz 1 heißt **standardisierte Zufallsvariable.**

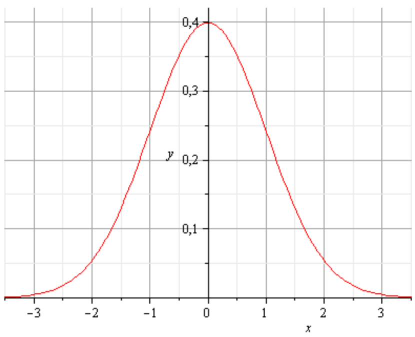
Satz: Es sei X eine Zufallsvariable mit dem Erwartungswert μ und der Standardabweichung σ. Dann ist $\frac{X-\mu}{\sigma}$ eine **standardisierte Zufallsvariable**.

Beweis: $E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma}\left(E(X)-\mu\right) = \frac{1}{\sigma}(\mu-\mu) = 0$ und

$\operatorname{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\operatorname{Var}(X-\mu) = \frac{1}{\sigma^2}\operatorname{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1$. q.e.d.

Johann Carl Friedrich Gauß, 1777-1855, (latinisiert **Carolus Fridericus Gauss**) leitete (veröffentlicht 1809) eine passende Formel für die **Dichtefunktion** (stetige Wahrscheinlichkeitsverteilung, $x \in \mathbb{R}$) $\varphi(x)$ der **Standardnormalverteilung** her:

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



0	± 0,5	± 1	± 1,5	± 2	± 2,5	± 3	± 3,5
0,39894228	0,35206533	0,24197072	0,1295176	0,05399097	0,0175283	0,00443185	0,00087268

$\varphi(x)$ ist wirklich eine Dichtefunktion, denn es ist $\int\limits_{-\infty}^{\infty}\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = 1$.

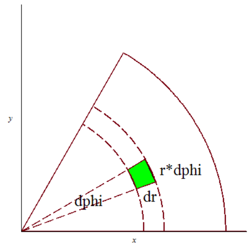
Beweis: Für den Beweis benötigt man einen Trick:

$$I^2 = \left(\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx\right)^2 = \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx\cdot\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \int\limits_{x=-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx\cdot\int\limits_{y=-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy = \int\limits_{x=-\infty}^{\infty}\int\limits_{y=-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy\,.$$

Veranschaulichung: $\left(f(x_1)\cdot\Delta x + f(x_2)\cdot\Delta x + f(x_3)\cdot\Delta x\right)\cdot\left(f(y_1)\cdot\Delta y + f(y_2)\cdot\Delta y + f(y_3)\cdot\Delta y\right) =$
$= f(x_1)\cdot f(y_1)\cdot\Delta x\cdot\Delta y + f(x_1)\cdot f(y_2)\cdot\Delta x\cdot\Delta y + f(x_1)\cdot f(y_3)\cdot\Delta x\cdot\Delta y + \ldots + f(x_3)\cdot f(y_3)\cdot\Delta x\cdot\Delta y$.

Dies ist eine Summe aus $3\cdot 3 = 9$ Summanden über alle Kombinationen $f(x_i)\cdot f(y_j)$.

Bei diesem zweidimensionalen Integral wird über die ganze zweidimensionale Ebene integriert. $dx\,dy$ ist der Flächeninhalt der kleinen Rechtecke, in welche die Ebene zerlegt wird. Nun führt man Polarkoordinaten r und φ ein: $x = r\cdot\cos\varphi$, $y = r\cdot\sin\varphi$ und $x^2+y^2 = r^2$. Mit ihrer Hilfe zerlegt man die Ebene in kleine Flächen vom Inhalt $r\cdot d\varphi\cdot dr = r\,dr\,d\varphi$.



Dann wird $I^2 = \int\limits_{r=0}^{\infty}\int\limits_{\varphi=0}^{2\pi} e^{-\frac{1}{2}r^2}\,r\,dr\,d\varphi = \int\limits_{r=0}^{\infty}\left[e^{-\frac{1}{2}r^2}\cdot r\cdot\varphi\right]_{\varphi=0}^{2\pi} dr = 2\pi\int\limits_{r=0}^{\infty} e^{-\frac{1}{2}r^2}\cdot r\cdot dr =$

$= 2\pi\cdot\left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty} = 2\pi\cdot(0+1) = 2\pi$. Und somit $I = \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}$, also $\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = 1$.

Es gilt $E(X) = \mu = \int\limits_{-\infty}^{\infty} x\cdot\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x\cdot e^{-\frac{1}{2}x^2}\,dx = 0$, da das Schaubild von $x\cdot\varphi(x)$ punktsymmetrisch zum Ursprung ist.

Oder $E(X) = \mu = \int\limits_{-\infty}^{\infty} x\cdot\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x\cdot e^{-\frac{1}{2}x^2}\,dx = \frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{1}{2}x^2}\right]_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}}(0-0) = 0$.

Und wegen $\mu = 0$ folgt für die Varianz

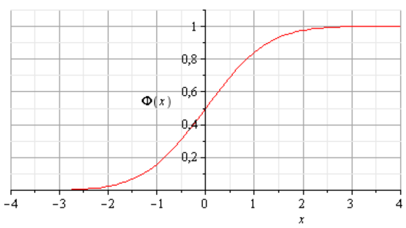
$$Var(x) = \sigma^2 = \int\limits_{-\infty}^{\infty}(x-\mu)^2\cdot\varphi(x)\,dx = \int\limits_{-\infty}^{\infty} x^2\cdot\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x^2\cdot e^{-\frac{1}{2}x^2}\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} -x\,e^{-\frac{1}{2}x^2}\cdot(-x)\,dx\,.$$

In dieser Form kann man mit Hilfe von $\int f(x)\cdot g(x)\,dx = F(x)\cdot g(x) - \int F(x)\cdot g'(x)\,dx$ partiell integrieren:

$$Var(x) = \sigma^2 = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x\,e^{-\frac{1}{2}x^2}\cdot x\,dx = \frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{1}{2}x^2}\cdot x\right]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} -e^{-\frac{1}{2}x^2}\cdot 1\,dx = (0-0) + \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = 1\,,$$

wie oben gezeigt.

Die **Verteilungsfunktion** der Standardnormalverteilung ist $\boxed{\boldsymbol{\Phi(x) = \int\limits_{-\infty}^{x}\varphi(t)\,dt = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt}}$.



Man erkennt: $\Phi(0) = 0{,}5$ und $\Phi(\infty) = 1$.

Das Schaubild von $\Phi(x)$ hat den Wendepunkt $(0/0{,}5)$, denn für die Ableitung gilt $\Phi'(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ und $\Phi''(x) = -x\cdot\varphi(x)$.

Wie kann man $\Phi(x)$ numerisch bestimmen? Wir wissen bisher nur, dass $\Phi(-\infty) = 0$, $\Phi(0) = 0{,}5$ und $\Phi(\infty) = 1$.

Nach Taylor kann eine Funktion f um die Stelle a entwickelt werden: $f(x) = \sum\limits_{n=0}^{\infty} a_n\cdot(x-a)^n = \sum\limits_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}\cdot(x-a)^n$.

Diese unendliche Reihe konvergiert für $|x-a| < r = \lim\limits_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right|$. Speziell folgt für $f(x) = e^x$ und $a = 0$ und alle $x \in \mathbb{R}$ die Reihe $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$. Für $f(x) = e^{-\frac{1}{2}x^2}$ ergibt sich

$$e^{-\frac{1}{2}x^2} = \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{1}{2}x^2\right)^n = 1 - \frac{1}{2}x^2 + \frac{1}{8}x^4 - \frac{1}{48}x^6 + \frac{1}{384}x^8 - \frac{1}{3840}x^{10} + \frac{1}{46080}x^{12} - \frac{1}{645120}x^{14} + \frac{1}{10321920}x^{16} \mp \ldots$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt = 0{,}5 + \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{1}{2}t^2} dt = 0{,}5 + \frac{1}{\sqrt{2\pi}} \left(x - \frac{1}{6}x^3 + \frac{1}{40}x^5 - \frac{1}{336}x^7 + \frac{1}{3456}x^9 - \frac{1}{42240}x^{11} \pm \ldots \right).$$

Beispiel: $\Phi(1)$:

Reihe bis x^1: $\Phi(1) \approx 0{,}8989$ Reihe bis x^3: $\Phi(1) \approx 0{,}83245$

Reihe bis x^5: $\Phi(1) \approx 0{,}842425$ Reihe bis x^7: $\Phi(1) \approx 0{,}841238$

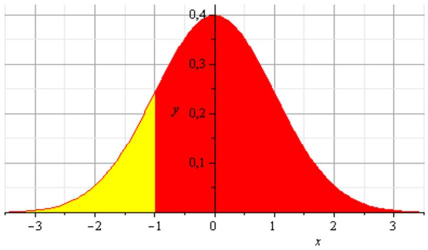
Reihe bis x^9: $\Phi(1) \approx 0{,}8413536$ Reihe bis x^{11}: $\Phi(1) \approx 0{,}84134412$

$\Phi(1) \approx 0{,}841344746$ auf neun Nachkommastellen genau.

Die Standardnormalverteilungstabelle: Z.B. $\Phi(1{,}23) = 0{,}8907$

x	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Bemerkungen zur Tabelle:



1. Für $x \geq 3{,}9$ ist $\Phi(x) = 1{,}0000$ auf 4 Nachkommastellen gerundet.
2. Es gilt $\boxed{\Phi(-x) = 1 - \Phi(x)}$.
   Beispiel: $\Phi(-1) = 1 - \Phi(1) = 1 - 0{,}8413 = 0{,}1587$.
   $\Phi(-1)$ ist durch die linke Fläche dargestellt.
   $\Phi(1)$ ist wegen der Achsensymmetrie gleich der rechten Fläche.
   Man erkennt, dass $\Phi(-1) + \Phi(1) = 1$ gilt.
3. Für $x \leq -3{,}9$ ist $\Phi(x) = 0{,}0000$ auf 4 Nachkommastellen gerundet.

Wie sieht die Verteilungsfunktion $f(x)$ aus, wenn die Varianz σ^2 und der Erwartungswert μ beträgt?

Bei $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ gilt $E(X) = \mu = \int_{-\infty}^{\infty} x \cdot \varphi(x)\,dx = 0$ und $Var(x) = \sigma^2 = \int_{-\infty}^{\infty} x^2 \cdot \varphi(x)\,dx = 1$; siehe oben.

Wenn nun die Varianz z.B. 2 ist, dann sieht das Schaubild der zugehörigen Verteilungsfunktion breiter aus. Es bietet sich $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ an. Denn z.B. verläuft auch das Schaubild der Parabel $y = \left(\frac{x}{2}\right)^2 = \frac{x^2}{4}$ viel breiter als dasjenige von $y = x^2$.

Und wenn nun die Varianz z.B. $\frac{1}{2}$ beträgt, dann sieht das Schaubild der zugehörigen Verteilungsfunktion schmaler aus. Auch hier bietet sich der gleiche Ansatz $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ an. Denn z.B. verläuft das Schaubild der Parabel $y = \left(\frac{x}{1/2}\right)^2 = 4x^2$ auch schmaler als dasjenige von $y = x^2$.

Eine einfache Rechnung ergibt aber $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = \sigma$ statt 1. Also nehmen wir $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$.

Denn dann wird $\int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = 1$, wie es sein soll.

Außerdem ist $E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = 0$ und $Var(X) = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = \sigma^2$.

Nun muss das Schaubild von $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ noch um μ nach rechts verschoben werden: Ersetzt man x durch $x - \mu$, dann folgt die Dichtefunktion zu $\boxed{\mathbf{f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}}$.

Analog, wenn z.B. die Parabel $y = a \cdot x^2$ um 3 nach rechts verschoben wird, dann erhält man $y = a \cdot (x-3)^2$.

Mit der Substitution $u = \frac{x-\mu}{\sigma}$ folgt $\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du = 1$.

Analog folgt $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \mu$ und $Var(x) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx = \sigma^2$.

Definition: Eine stetige Zufallsvariable X heißt **normalverteilt** mit $E(X) = \mu$ und $Var(X) = \sigma^2$, wenn für ihre Dichtefunktion gilt

$$\boxed{\mathbf{f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{für } x\in\mathbb{R}}} \quad \text{mit } \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}.$$

(10 DM-Schein seit 1991 mit Gauß und seiner Glockenkurve)

Mit der Substitution $u=\frac{x-\mu}{\sigma}$ folgt $du=\frac{1}{\sigma}dx$ oder $dx=\sigma\, du$ folgt für normalverteilte Zufallsvariablen X

$$P(X\le k)=\int_{-\infty}^{k}\frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=\frac{1}{\sigma\cdot\sqrt{2\pi}}\int_{-\infty}^{\frac{k-\mu}{\sigma}}e^{-\frac{1}{2}u^2}\sigma\, du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{k-\mu}{\sigma}}e^{-\frac{1}{2}u^2}du=\Phi\left(\frac{k-\mu}{\sigma}\right).$$

Allgemein gilt $\boxed{\mathbf{P(X\le x)=\Phi\left(\frac{x-\mu}{\sigma}\right)}}$.

Zentraler Grenzwertsatz: Bei n **unabhängigen** Wiederholungen X_i einer beliebigen Zufallsvariablen X mit $E(X_i)=E(X)=\mu$ und $Var(X_i)=Var(X)=\sigma^2$ konvergiert die Wahrscheinlichkeitsverteilung der Summe $S_n=\sum_{i=1}^{n}X_i$ und auch die des Mittelwerts $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ mit zunehmendem n gegen die Normalverteilung mit den Parametern $E(S_n)=n\cdot\mu$ und $Var(S_n)=n\cdot\sigma^2$ bzw. $E(\overline{X}_n)=E\left(\frac{1}{n}S_n\right)=\frac{1}{n}\cdot E(S_n)=\frac{1}{n}\cdot n\cdot\mu=\mu$ und $Var(\overline{X}_n)=Var\left(\frac{1}{n}S_n\right)=\frac{1}{n^2}\cdot Var(S_n)=\frac{1}{n^2}\cdot n\cdot\sigma^2=\frac{\sigma^2}{n}$. Dabei ist es gleichgültig, welche Verteilung die X_i haben, sie müssen nur identisch verteilt sein. (Ohne Beweis)

Beispiel 1: Bei einmaligem Würfeln sind die 6 Ergebnisse identisch verteilt, ja sogar gleichverteilt mit der jeweiligen Wahrscheinlichkeit $1/6$. Der Erwartungswert beträgt $\mu=\frac{1}{6}\cdot(1+2+...6)=3{,}5$ und die Varianz

$$\sigma^2=\frac{1}{6}\left((1-3{,}5)^2+(2-3{,}5)^2+...+(6-3{,}5)^2\right)=\frac{35}{12}.$$

Bei n-maligem Würfeln liegt die Augensumme S zwischen n und $6n$. Der Erwartungswert beträgt $E(S)=n\cdot 3{,}5$ und die Varianz $Var(S)=n\cdot\frac{35}{12}$.

Die Anzahl der Möglichkeiten, dabei die Augensumme S zu erhalten, fand ich im Internet unter dem Nickname „Schadowking“ zu $\sum_{k=0}^{\frac{S-n}{6}}(-1)^k\cdot\binom{n}{k}\cdot\binom{S-1-6k}{n-1}$.

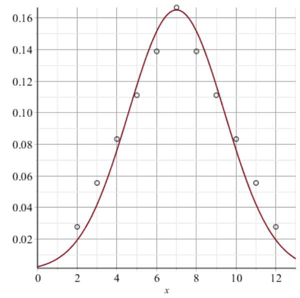
1. Würfle zweimal und betrachte die Augensumme S_2. Dann gilt

S_2	2	3	4	5	6	7	8	9	10	11	12
$P(S_2)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Dabei ist $E(S_2) = 2 \cdot 3{,}5 = 7$ und $Var(S_2) = 2 \cdot \frac{35}{12} = \frac{35}{6}$.

Für die zugehörige Normalverteilung gilt dann

$$f(x) = \frac{1}{\sqrt{35/6} \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-7}{\sqrt{35/6}}\right)^2} .$$



Vergleich der Einzelwerte mit der Normalverteilung.

Da $n = 2$ klein ist, ist keine bessere Übereinstimmung zu erwarten.

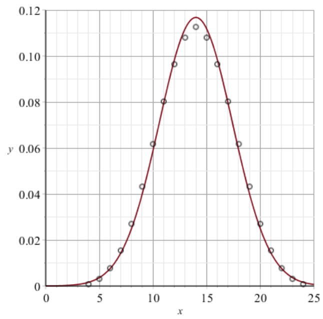
2. Bei viermaligem Würfeln gilt $E(S_4) = 4 \cdot 3{,}5 = 14$ und $Var(S_4) = 4 \cdot \frac{35}{12} = \frac{35}{3}$. Die Augensumme S_4 kann die 21 Werte 4, 5, … 24 annehmen. Hier der Anfang der Wertetabelle. Dies genügt, denn sie ist symmetrisch zum elften Wert $S_4 = 14$.

S_4	4	5	6	7	8	9	10	11	12	13	14	15	16
$P(S_4) \cdot 6^4$	1	4	10	20	35	56	80	104	125	140	146	140	125

Für die zugehörige Normalverteilung gilt dann

$$f(x) = \frac{1}{\sqrt{35/3} \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-14}{\sqrt{35/3}}\right)^2} .$$

Für $n = 4$ ist die Übereinstimmung schon besser.



3. Bei achtmaligem Würfeln ist die Übereinstimmung der exakten Wahrscheinlichkeit mit der Näherung

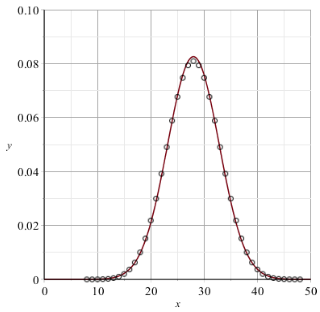
$$f(x) = \frac{1}{\sqrt{70/3} \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-28}{\sqrt{70/3}}\right)^2}$$

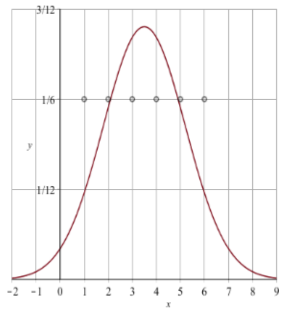
hervorragend. (Linkes Bild)

4. Total unbrauchbar ist die Näherung bei nur einmaligem Würfeln. Da gilt

$$f(x) = \frac{1}{\sqrt{35/12} \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-3{,}5}{\sqrt{35/12}}\right)^2}$$, während

die exakten Wahrscheinlichkeiten konstant $\frac{1}{6}$ sind. (Rechtes Bild)

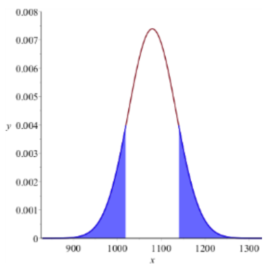




Beispiel 2: Aus langjähriger Erfahrung benötigt ein Dozent für die Korrektur einer Klausur im Mittel $\mu = 30$ Minuten bei einer Standardabweichung von $\sigma = 9$ Minuten. Nun stehen $n = 36$ Klausuren zur Korrektur an. Die erwartete Korrekturzeit beträgt also $36 \cdot \mu = 36 \cdot 30$ Minuten = 1080 Minuten =18 Stunden bei einer Standardabweichung von $\sqrt{n \cdot \sigma^2} = \sqrt{36 \cdot 81} = 54$ Minuten. Da die einzelnen Korrekturzeiten sowohl unabhängig voneinander als auch identisch verteilt sind (Verteilung ist unbekannt), ist die gesamte Korrekturzeit X nach dem zentralen Grenzwertsatz näherungsweise normalverteilt.

Der Dozent möchte nun wissen, mit welcher Wahrscheinlichkeit er höchstens 17 Stunden = 1020 Minuten Korrekturzeit benötigt: $P(X \le 1020) = \Phi\left(\frac{1020-1080}{54}\right) = \Phi\left(-\frac{10}{9}\right) = 1 - \Phi\left(\frac{10}{9}\right) \approx 1 - 0{,}8665 = 0{,}1335$.

$$P(X \le 1020) = \Phi\left(\frac{1020-1080}{54}\right) = \Phi\left(-\frac{10}{9}\right) = 1-\Phi\left(\frac{10}{9}\right) \approx 1-0,8665 = 0,1335 .$$



Und zu seinem Trost, die Wahrscheinlichkeit, dass er mehr als 19 Stunden = 1140 Stunden korrigieren muss, beträgt auch nur

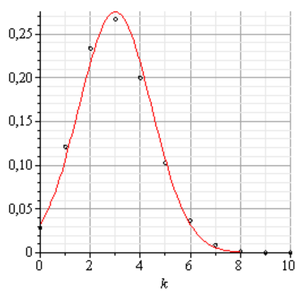
$$P(X > 1140) = 1-P(X \le 1140) = 1-\Phi\left(\frac{1140-1080}{54}\right) = 1-\Phi\left(\frac{10}{9}\right) \approx 1-0,8665 = 0,1335 .$$

Folgerung: Für nicht zu kleine Werte von n (Faustformel $\sigma^2 = n \cdot p \cdot q > 9$) gilt die **Näherungsformel von Moivre-Laplace**:
(Abraham de Moivre 1667-1754, ein Freund von Newton): Spezialfall $p = 1/2$ um 1733)
(Pierre-Simon Marquis de Laplace 1749-1827: allgemein um 1810)

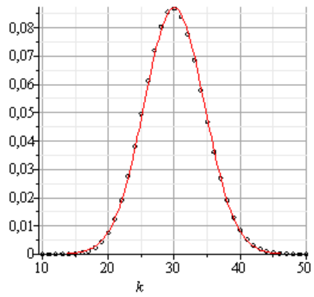
$$\boxed{\mathbf{B_{n;p}(k) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}} \quad \text{mit} \quad \mathbf{\mu = n \cdot p,} \quad \mathbf{\sigma = \sqrt{n \cdot p \cdot q}} \quad \text{und} \quad \mathbf{q = 1-p}}$$

Beispiele: Gezeichnet sind jeweils die Punkte von $B_{n;p}(k)$ und das Schaubild von $f(k) = \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$.

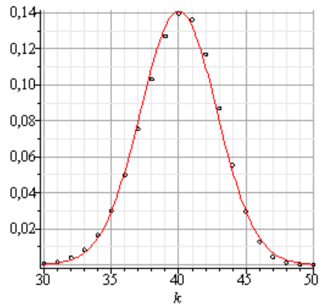
$n = 10, \quad p = 0,3$



$n = 100, \quad p = 0,3$



$n = 50, \quad p = 0,8$



Zusätze zur Normalverteilung:

1. Die Zufallsvariable X sei normalverteilt mit dem Erwartungswert μ und der Varianz σ^2. Dann gilt

$$P(\mu-\sigma \le X \le \mu+\sigma) = \Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right) = \Phi(1)-\Phi(-1) = \Phi(1)-\left(1-\Phi(1)\right) = 2\cdot\Phi(1)-1 \approx 0,6827$$

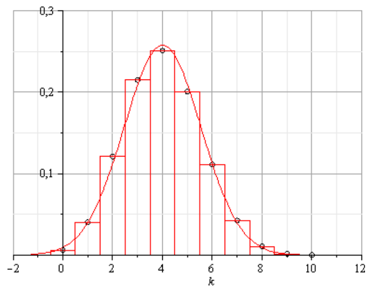
d.h. **68% aller beobachteten Werte liegen im Intervall** $[\mu-\sigma / \mu+\sigma]$. Die sog. σ**-Regel**.

Analog folgt $P(\mu-2\sigma \le X \le \mu+2\sigma) = 2\cdot\Phi(2)-1 \approx 0,9545$,
d.h. **95% aller beobachteten Werte liegen im Intervall** $[\mu-2\sigma / \mu+2\sigma]$, die sog. 2σ**-Regel**.

Ebenso: $P(\mu-3\sigma \le X \le \mu+3\sigma) = 2\cdot\Phi(3)-1 \approx 0,9981 \approx 99,8\%$.

2. Die Zufallsvariable X sei $B_{n;p}$-verteilt. Uns interessiert $P(X \le x) = \sum_{k=0}^{x} B_{n;p}(k) = \sum_{k=0}^{x} \binom{n}{k} \cdot p^k \cdot q^{n-k}$.

Beispiel: $n = 10, \quad p = 0,4$. Dann ist $\mu = 4$ und $\sigma = \sqrt{npq} = \sqrt{2,4}$.
Im Schaubild sind die 11 Werte von $B_{10;0,4}(k)$ als Punkte, das zugehörige Histogramm aus 11 gleich breiten Rechtecken und die Näherungskurve $y = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ gezeichnet.



Es sei x eine ganze Zahl. Dann kann man

$P(X \le x) = \sum_{k=0}^{x} B_{n;p}(k) = \sum_{k=0}^{x} \binom{n}{k} \cdot p^k \cdot q^{n-k}$ interpretieren als:

1. Summe der Wahrscheinlichkeitswerte $B_{n;p}(k)$, oder
2. Summe der gezeichneten Rechteckflächeninhalte bis x, oder
3. näherungsweise als Inhalt der Fläche unter der Näherungskurve von $-\infty$ bis zum Wert $x + \frac{1}{2}$, denn die Rechtecke um den Wert k gehen von $k - \frac{1}{2}$ bis $k + \frac{1}{2}$.

Z.B. sei $x = 6$. Dann gilt

$P(X \le 6) = \sum_{k=0}^{6} B_{10;0,4}(k) = \sum_{k=0}^{6} \binom{10}{k} \cdot 0,4^k \cdot 0,6^{10-k} \approx 0,9452$.

Mit Hilfe der Näherungsformel folgt $P(X \le x) \simeq \int_{-\infty}^{x+\frac{1}{2}} \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt = \int_{-\infty}^{x+\frac{1}{2}} \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\cdot\left(\frac{t-\mu}{\sigma}\right)^2} dt$.

Dieses Integral löst man durch Substitution: $u = \frac{t-\mu}{\sigma}$, bzw. $t = \mu + \sigma u$. Dann folgt $dt = \sigma du$

und folglich $P(X \le x) \approx \int_{-\infty}^{\frac{x+0,5-\mu}{\sigma}} \frac{1}{\sigma} \varphi(u)\, \sigma du = \int_{-\infty}^{\frac{x+0,5-\mu}{\sigma}} \varphi(u)\, du = \Phi\left(\frac{x+0,5-\mu}{\sigma}\right)$.

Für $x = 6$ folgt $P(X \le 6) \approx \Phi\left(\frac{6+0,5-4}{\sqrt{2,4}}\right) \approx \Phi(1,61) \approx 0,9467$.

Ohne die 0,5 wäre $P(X \le 6) \approx \Phi\left(\frac{6-4}{\sqrt{2,4}}\right) \approx \Phi(1,29) \approx 0,9016$, ein wesentlich schlechterer Wert. Das liegt daran, dass das n mit $n = 10$ doch sehr klein ist.

Satz: Die Zufallsvariable X sei $B_{n;p}$-verteilt. Dann gilt näherungsweise:

$$\boxed{\mathbf{P(X \le x) \approx \Phi\left(\frac{x+0,5-\mu}{\sigma}\right) \quad mit \quad \mu = n \cdot p \quad und \quad \sigma = \sqrt{n \cdot p \cdot q}}}$$ bzw.

$$\boxed{\mathbf{P(x_1 \le X \le x_2) \approx \Phi\left(\frac{x_2+0,5-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-0,5-\mu}{\sigma}\right) \quad mit \quad \mu = n \cdot p \quad und \quad \sigma = \sqrt{n \cdot p \cdot q}}}.$$

3. Die Zufallsvariable X sei $B_{n;p}$-verteilt. Wie lässt sich mit Hilfe der Vertafelung von $\Phi(x)$ der Wert von $B_{n;p}(k)$ bestimmen?

   Mit Hilfe der Rechtecke im Histogramm erkennt man, dass $B_{n;p}(k) \approx \Phi\left(\frac{k+0,5-\mu}{\sigma}\right) - \Phi\left(\frac{k-0,5-\mu}{\sigma}\right)$.

   Beispiel 1: $B_{10;0,4}(3) = \binom{10}{3} \cdot 0,4^3 \cdot 0,6^7 \approx 0,215$. Und mit der Näherung folgt

   $B_{10;0,4}(3) \approx \Phi\left(\frac{3+0,5-4}{\sqrt{2,4}}\right) - \Phi\left(\frac{3-0,5-4}{\sqrt{2,4}}\right) \approx \Phi(-0,3227) - \Phi(-0,9682) \approx 0,3734 - 0,1665 = 0,207$.

   Beispiel 2: $B_{100;0,4}(30) = \binom{100}{30} \cdot 0,4^{30} \cdot 0,6^{70} \approx 0,0100$. Und mit der Näherung folgt

   $B_{100;0,4}(30) \approx \Phi\left(\frac{30+0,5-40}{\sqrt{24}}\right) - \Phi\left(\frac{30-0,5-40}{\sqrt{24}}\right) \approx \Phi(-1,939) - \Phi(-2,143) \approx 0,0262 - 0,0160 = 0,0102$.

Man erkennt, je größer das n, desto besser ist die Übereinstimmung.

Bemerkung: In der Formel $P(X \le x) \approx \Phi\left(\frac{x+0,5-\mu}{\sigma}\right)$ darf man **nur bei großem n die 0,5 weglassen.**
Bei kleinem n sieht man den Unterschied am besten:

Im Schaubild $n=4$ und $p=0,4$. Gezeichnet sind die 5 Punkte für $P(X \le x) = \sum_{k=0}^{x} B_{4;0,4}(k) = \sum_{k=0}^{x} \binom{4}{k} \cdot 0,4^k \cdot 0,6^{4-k}$ für $k=0, 1, 2, 3, 4$, die untere Kurve $y = \Phi\left(\frac{x-1,6}{\sqrt{0,96}}\right)$ und die obere Kurve $y = \Phi\left(\frac{x+0,5-1,6}{\sqrt{0,96}}\right)$.

Man erkennt die Güte der oberen Kurve.

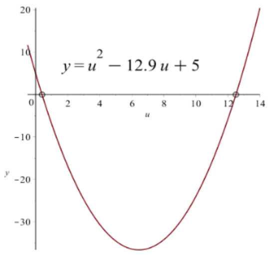
Aufgabe 1:
Schätzen Sie ab, wie oft man eine ideale Münze werfen muss, so dass die relative Häufigkeit für Bild mit einer Wahrscheinlichkeit von mindestens 99% in dem zu $p=0,5$ symmetrischen Intervall $[0,4/0,6]$ liegt.

Lösung:
Es sei X die Anzahl der Bilder bei n Würfen. Bei einer idealen Münze ist die Wahrscheinlichkeit für Bild $p=0,5$, der Erwartungswert ist $\mu = p = 0,5$ und die Varianz $\sigma^2 = p \cdot q = 0,5 \cdot 0,5 = 0,25$. Da die Ergebnisse der einzelnen Würfe unabhängig voneinander sind, addieren sich die Erwartungswerte p und auch die Varianzen $p \cdot q$ zu $\mu_X = n \cdot p = n \cdot 0,5$ und $\sigma_X^2 = n \cdot p \cdot q = 0,25n$, also $\sigma_X = 0,5\sqrt{n}$. Mit $\overline{X} = \frac{X}{n}$, der mittleren Bild-Quote, folgt $P(0,4 \le \overline{X} \le 0,6) = P(0,4 \cdot n \le X \le 0,6 \cdot n) = \Phi\left(\frac{0,6 \cdot n + 0,5 - 0,5 \cdot n}{0,5\sqrt{n}}\right) - \Phi\left(\frac{0,4 \cdot n - 0,5 - 0,5 \cdot n}{0,5\sqrt{n}}\right) =$

$$= \Phi\left(\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) - \Phi\left(-\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) = \Phi\left(\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) - 1 + \Phi\left(\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) = 2 \cdot \Phi\left(\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) - 1 \ge 0,99.$$

Daraus folgt $\Phi\left(\frac{0,2 \cdot n + 1}{\sqrt{n}}\right) \ge 0,995$. Nach obiger Tabelle folgt $\frac{0,2 \cdot n + 1}{\sqrt{n}} \ge 2,58$.

Dies ergibt $0,2n - 2,58\sqrt{n} + 1 \ge 0$ bzw. $u^2 - 12,9u + 5 \ge 0$, eine quadratische Ungleichung für $u = \sqrt{n}$.
Es folgt $\sqrt{n} = u \le 0,4$ oder $\sqrt{n} = u \ge 12,5$, folglich $n \ge 12,5^2 = 156,25$. Die Münze muss also mindestens 157 Mal geworfen werden.



Variante: Wir gehen von einer perfekten Normalverteilung aus und lassen die 0,5-Korrektur weg. Dann folgt

$$P(0,4 \le \overline{X} \le 0,6) = P(0,4 \cdot n \le X \le 0,6 \cdot n) = \Phi\left(\frac{0,6 \cdot n - 0,5 \cdot n}{0,5 \cdot \sqrt{n}}\right) - \Phi\left(\frac{0,4 \cdot n - 0,5 \cdot n}{0,5 \cdot \sqrt{n}}\right) =$$

$= \Phi\left(0,2 \cdot \sqrt{n}\right) - \Phi\left(-0,2 \cdot \sqrt{n}\right) = \Phi\left(0,2 \cdot \sqrt{n}\right) - 1 + \Phi\left(0,2 \cdot \sqrt{n}\right) = 2 \cdot \Phi\left(0,2 \cdot \sqrt{n}\right) - 1 \ge 0,99$. Daraus folgt $\Phi\left(0,2 \cdot \sqrt{n}\right) \ge 0,995$, also $0,2 \cdot \sqrt{n} \ge 2,58$, so dass $\sqrt{n} \ge 12,9$, d.h. $n \ge 166,41$.

2. Lösungsweg: Wir verwenden eine andere Zufallsvariable X mit $X = \begin{cases} 1 & \text{bei Bild} \\ 0 & \text{bei Zahl} \end{cases}$.

Mit $\mu_{\overline{X}} = \mu_X = p = 0,5$ und $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{\sqrt{p \cdot q}}{\sqrt{n}} = \frac{\sqrt{0,5 \cdot 0,5}}{\sqrt{n}} = \frac{0,5}{\sqrt{n}}$ folgt

$$P(0,4 \le \overline{X} \le 0,6) = \Phi\left(\frac{0,6-0,5}{0,5/\sqrt{n}}\right) - \Phi\left(\frac{0,4-0,5}{0,5/\sqrt{n}}\right) = \Phi\left(0,2 \cdot \sqrt{n}\right) - \Phi\left(-0,2 \cdot \sqrt{n}\right) = 2 \cdot \Phi\left(0,2 \cdot \sqrt{n}\right) - 1$$ wie oben.

Verallgemeinerung: Wir gehen von einer Normalverteilung der Zufallsvariablen X aus, lassen also die 0,5-Korrektur weg.

Gegeben sei
1. die Trefferwahrscheinlichkeit p,
2. die Länge ℓ des zum (nicht notwendig bekannten) Erwartungswert symmetrischen Vertrauensintervalls,
3. die Mindestwahrscheinlichkeit **α** für dieses Vertrauensintervall, das sog. **Konfidenzniveau**.

Gesucht ist der dazu erforderliche **Stichprobenumfang n.**

Die Zufallsvariable X sei wie beim 2. Lösungsweg definiert. Dann ist $\mu_{\overline{X}} = \mu_X = p$ und $\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n} = \frac{p \cdot q}{n}$.

$$P\left(\mu_{\overline{X}} - \frac{\ell}{2} \le \overline{X} \le \mu_{\overline{X}} + \frac{\ell}{2}\right) = \Phi\left(\frac{\mu_{\overline{X}} + \frac{\ell}{2} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right) - \Phi\left(\frac{\mu_{\overline{X}} - \frac{\ell}{2} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right) = \Phi\left(\frac{\ell}{2\sigma_{\overline{X}}}\right) - \Phi\left(-\frac{\ell}{2\sigma_{\overline{X}}}\right) = 2 \cdot \Phi\left(\frac{\ell}{2\sigma_{\overline{X}}}\right) - 1 =$$

$2 \cdot \Phi\left(\frac{\ell}{2} \cdot \sqrt{\frac{n}{p \cdot q}}\right) - 1 \ge \alpha$, so dass $\Phi\left(\frac{\ell}{2} \cdot \sqrt{\frac{n}{p \cdot q}}\right) \ge \frac{1+\alpha}{2}$. Aus der Tabelle der Normalverteilung ergibt sich dann der Wert z, so dass $\frac{\ell}{2} \cdot \sqrt{\frac{n}{p \cdot q}} \ge z$ und folglich $\boxed{n \ge \frac{4z^2 \cdot p \cdot q}{\ell^2}}$, wobei $\boxed{\Phi(z) = \frac{1+\alpha}{2}}$.

Diese Formel ist **unabhängig vom Erwartungswert** μ_X.

Für obiges Beispiel ergibt dies $n \ge \frac{4z^2 \cdot p \cdot q}{\ell^2} = \frac{4 \cdot 2{,}58^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}2^2} = 166{,}41$ wie oben.

Zusatz: Falls der Wert von p nicht bekannt ist, dann wird in der obigen Formel sicherheitshalber den größtmöglichen Wert von $p \cdot q = p \cdot (1-p) = p - p^2$ eigesetzt. Und dieser Wert ist 0,25 für $p = 0{,}5$. Also $\boxed{n \ge \frac{z^2}{\ell^2}}$.

Aufgabe 2:
Eine Zufallsvariable X sei $B_{300;0,5}$-verteilt. Bestimmen Sie ein möglichst kleines Intervall $[\mu_X - \varepsilon / \mu_X + \varepsilon]$ so, dass X mit einer Wahrscheinlichkeit von 90% Werte aus diesem Intervall annimmt.

Lösung:
Es ist $\mu_X = n \cdot p = 150$ und $\sigma_X^2 = n \cdot p \cdot q$, also $\sigma_X = \sqrt{n \cdot p \cdot q} = \sqrt{300 \cdot 0{,}5 \cdot 0{,}5} = 5\sqrt{3}$. Dann folgt

$$P(|X - \mu_X| \le \varepsilon) = P(150 - \varepsilon \le X \le 150 + \varepsilon) = \Phi\left(\frac{150 + \varepsilon + 0{,}5 - 150}{5\sqrt{3}}\right) - \Phi\left(\frac{150 - \varepsilon - 0{,}5 - 150}{5\sqrt{3}}\right) =$$

$= \Phi\left(\frac{\varepsilon + 0{,}5}{5\sqrt{3}}\right) - \Phi\left(-\frac{\varepsilon + 0{,}5}{5\sqrt{3}}\right) = 2 \cdot \Phi\left(\frac{\varepsilon + 0{,}5}{5\sqrt{3}}\right) - 1 = 0{,}9$. Also $\Phi\left(\frac{\varepsilon + 0{,}5}{5\sqrt{3}}\right) = 0{,}95$. Die Tabelle liefert

$\frac{\varepsilon + 0{,}5}{5\sqrt{3}} = 1{,}645$, d.h. $\varepsilon = 13{,}7$. Das Intervall lautet also $[136{,}3/163{,}7]$, d.h. $[137/163]$.

Zusatz: Bisher hatten wir angenommen, dass die Zufallsvariable annähernd normalverteilt ist. Aber auch für beliebige Verteilungen lässt sich die Wahrscheinlichkeit für die Abweichung vom Mittelwert abschätzen:

Tshebyscheff'sche Ungleichung: $\boxed{P(|X - \mu| \ge c) \le \frac{\sigma^2}{c^2}}$ für $c > 0$.

Pafnuti Lwowitsch Tschebyscheff (Пафнутий Львович Чебышёв), 1821 – 1894, russischer Mathematiker.

Beweis: $\sigma^2 = \sum_{i=1}^{m} (x_i - \mu)^2 \cdot P(X = x_i) \ge \sum_{|x_i - \mu| \ge c} (x_i - \mu)^2 \cdot P(X = x_i) \ge \sum_{|x_i - \mu| \ge c} c^2 \cdot P(X = x_i) =$

$= c^2 \cdot \sum_{|x_i - \mu| \ge c} P(X = x_i) = c^2 \cdot P(|X - \mu| \ge c)$. Und daraus folgt sie Behauptung.

Folgerung: $\boxed{P(|X-\mu|<c) \geq 1-\frac{\sigma^2}{c^2}}$ für $c>0$.

Beispiele: Es sei X eine beliebige Zufallsvariable. Dann gilt

$P(\mu-\sigma<X<\mu+\sigma)=P(-\sigma<X-\mu<\sigma)=P(|X-\mu|<\sigma)\geq 1-\frac{\sigma^2}{\sigma^2}=0$. Dies ist eine leere Aussage!

$P(\mu-2\sigma<X<\mu+2\sigma)=P(-2\sigma<X-\mu<2\sigma)=P(|X-\mu|<2\sigma)\geq 1-\frac{\sigma^2}{(2\sigma)^2}=0{,}75$, d.h. eine beliebige Zufallsvariable X nimmt mit einer Wahrscheinlichkeit von mindestens 75% Werte im Intervall $\mu-2\sigma\leq X\leq\mu+2\sigma$ an.

$P(\mu-3\sigma<X<\mu+3\sigma)=P(-3\sigma<X-\mu<3\sigma)=P(|X-\mu|<3\sigma)\geq 1-\frac{\sigma^2}{(3\sigma)^2}=\frac{8}{9}=88{,}9\%$, d.h. eine beliebige Zufallsvariable X nimmt mit einer Wahrscheinlichkeit von mindestens 88,9% Werte im Intervall $\mu-3\sigma<X<\mu+3\sigma$ an.

Die Näherungsformel von Poisson (Siméon Denis Poisson, 1837)

Die Poisson-Verteilung lässt sich aus der Binomialverteilung herleiten, und zwar für großes n und kleines p. Wir setzen $n\cdot p=\lambda$, d.h. $p=\frac{\lambda}{n}$. Dann gilt

$$\lim_{n\to\infty} B_{n;p}(k)=\lim_{n\to\infty}\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}=\lim_{n\to\infty}\frac{n!}{(n-k)!\cdot k!}\cdot\left(\frac{\lambda}{n}\right)^k\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$=\lim_{n\to\infty}\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$=\lim_{n\to\infty}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n}\right)\cdot\frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k}.$$

Wegen $\lim_{n\to\infty}\frac{n}{n}=\lim_{n\to\infty}\frac{n-1}{n}=\ldots=\lim_{n\to\infty}\frac{n-k+1}{n}=1$ und $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}$ und $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1^{-k}=1$ folgt

Def.: Eine Zufallsvariable X heißt **Poisson-verteilt**, wenn gilt $\boxed{P(X=k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda} \text{ für } k=0,1,2,\ldots}$

Satz: Für eine Poisson-verteilte Zufallsvariable gilt $\boxed{E(X)=Var(X)=\lambda}$, d.h. $\boxed{\mu=\sigma^2=\lambda}$.

Beweis:

Es gilt $\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\cdot e^{-\lambda}=e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1$.

$$E(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\lambda\cdot e^{-\lambda}\cdot\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda\cdot e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda\cdot e^{-\lambda}\cdot e^{\lambda}=\lambda.$$

λ entspricht also der durchschnittlichen Anzahl des Auftretens eines Ereignisses pro Zeit, Strecke, Fläche,…

$$Var(X)=\sum_{k=0}^{\infty}(k-\lambda)^2\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=0}^{\infty}(k^2-2k\lambda+\lambda^2)\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=e^{-\lambda}\cdot\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}-2\lambda\cdot\underbrace{\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}}_{=E(X)=\lambda}+\lambda^2\cdot e^{-\lambda}\cdot\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}}$$

Dann wird $Var(X)=e^{-\lambda}\cdot\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}-2\lambda^2+\lambda^2=e^{-\lambda}\cdot\sum_{k=0}^{\infty}\left(k\cdot(k-1)+k\right)\frac{\lambda^k}{k!}-\lambda^2$

$$=e^{-\lambda}\cdot\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!}+\underbrace{\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}}_{=E(X)=\lambda}-\lambda^2=e^{-\lambda}\cdot\lambda^2\cdot\underbrace{\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}}_{=e^{\lambda}}+\lambda-\lambda^2=\lambda. \quad \text{q.e.d.}$$

Beispiel 1: Ein idealer Würfel wird 6-mal geworfen. Wie groß ist die Wahrscheinlichkeit, genau einmal die 6 zu erhalten?

Es sei X die Anzahl der erhaltenen Sechsen.

Mit der Binomialverteilung folgt: $P(X=1)=\binom{6}{1}\cdot\left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^5=0{,}4019$.

Mit der Poisson-Näherung folgt wegen $\lambda=E(X)=1$ die Wahrscheinlichkeit $P(X=1)=\frac{1^1}{1!}\cdot e^{-1}=0{,}3679$.

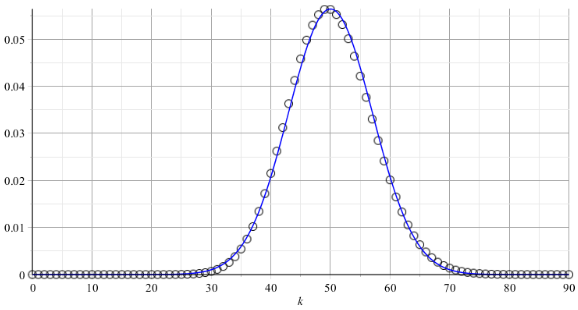
Die Poisson-Näherung ist nicht so gut, da $n=6$ zu klein ist.

Falls nun der Würfel $n=6000$ mal geworfen wird, dann gilt für die Wahrscheinlichkeit, dass die Sechs 1000 mal erscheint:

$$P(X=1000)=\binom{6000}{1000}\cdot\left(\frac{1}{6}\right)^{1000}\cdot\left(\frac{5}{6}\right)^{5000}\approx 0{,}0138.$$

Nach Poisson folgt mit $\lambda=\mu=1000$ und $k=1000$: $P(X=k=1000)=\frac{\lambda^k}{k!}\cdot e^{-k}=\frac{1000^{1000}}{1000!}\cdot e^{-1000}\approx 0{,}0126$.

Beispiel 2: Im Mittel passieren 50 Radfahrer pro Stunde eine bestimmte Stelle. Die Anzahl X der Radfahrer, die diese Stelle passieren, sei Poisson-verteilt.



Die Kreise geben die Poisson-Verteilung wieder:

$$P(X=k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\frac{50^k}{k!}\cdot e^{-50} \quad \text{mit } \lambda=\mu=50.$$

Die durchgezogene Linie ist die Gauß-Verteilung

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

mit $\mu=\sigma^2=50$.

Zusatz: Bestimmen Sie $P(45\le X\le 55)$ bei $\lambda=\mu=\sigma^2=50$.

Nach Poisson folgt $P(45\le X\le 55)=\sum_{k=45}^{55}P(X=k)=\sum_{k=45}^{55}\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=45}^{55}\frac{50^k}{k!}\cdot e^{-50}=0{,}5634$ nach meinem PC.

Nach dem zentralen Grenzwertsatz können wir wegen $\mu=\sigma^2=50$ rechnen:

$$P(45\le X\le 55)=\Phi\left(\frac{55+0{,}5-50}{\sqrt{50}}\right)-\Phi\left(\frac{45-0{,}5-50}{\sqrt{50}}\right)=\Phi(0{,}78)-\Phi(-0{,}78)=2\cdot\Phi(0{,}78)-1=2\cdot 0{,}7823-1=0{,}5646$$

mit Hilfe unserer Tafel.

Eine genauere Rechnung mit dem PC liefert $\Phi\left(\frac{55+0{,}5-50}{\sqrt{50}}\right)-\Phi\left(\frac{45-0{,}5-50}{\sqrt{50}}\right)=0{,}5633$, traumhaft!

Ohne diese 0,5 erhielte man $\Phi\left(\frac{55-50}{\sqrt{50}}\right)-\Phi\left(\frac{45-50}{\sqrt{50}}\right)=0{,}5205$, enttäuschend!

Die zweidimensionale (bivariate) Normalverteilung

Gegeben sind zwei normalverteilte Zufallsvariablen X und Y mit den Dichtefunktionen

$$f_X(x)=\frac{1}{\sigma_X\cdot\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \quad\text{und}\quad f_Y(y)=\frac{1}{\sigma_Y\cdot\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}.$$

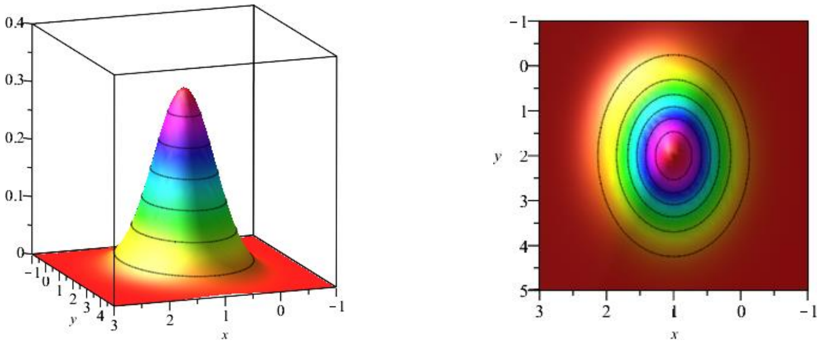
① Wenn die beiden Zufallsvariablen unabhängig sind, dann gilt wegen $P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$

für die gemeinsame Dichte $f(x,y) = f_X(x) \cdot f_Y(y) = \dfrac{1}{\sigma_X \cdot \sigma_Y \cdot 2\pi} e^{-\frac{1}{2}\cdot\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$.

$f(x,y)$ ist wirklich eine Dichte, denn es gilt

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)\cdot f_y(y)\,dx\,dy = \int_{-\infty}^{\infty} f_Y(y) \underbrace{\int_{-\infty}^{\infty} f_x(x)\cdot dx}_{=1}\,dy = \int_{-\infty}^{\infty} f_Y(y)\,dy = 1.$$

Beispiel 1a: $\mu_X = 1$, $\sigma_X = 0{,}5$, $\mu_Y = 2$, $\sigma_X = 1$. Die Höhenlinien sind Ellipsen.

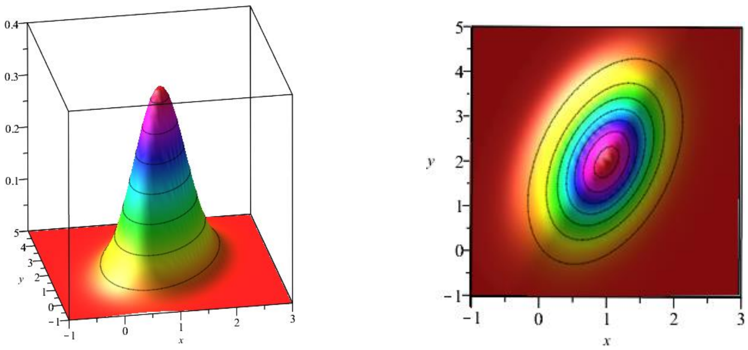


② Wenn die beiden Zufallsvariablen abhängig sind, dann gilt (ohne Herleitung)

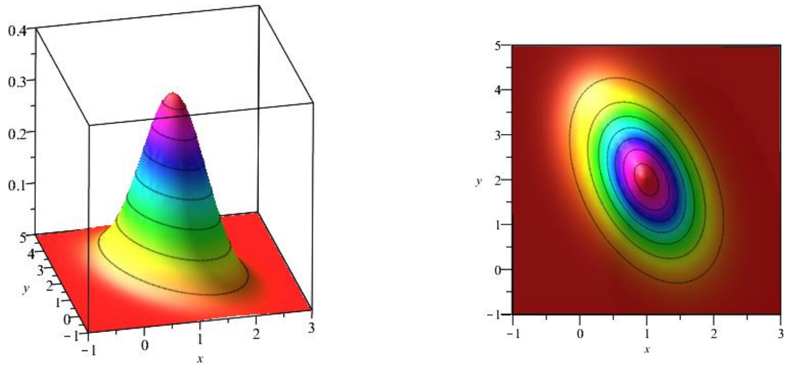
$f(x,y) = \dfrac{1}{\sigma_X \cdot \sigma_Y \cdot 2\pi \cdot \sqrt{1-\rho^2}} e^{-\frac{1}{2\cdot(1-\rho^2)}\cdot\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\cdot\frac{x-\mu_X}{\sigma_X}\cdot\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$ mit dem Korrelationskoeffizienten

$\rho = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(x)\cdot\mathrm{Var}(Y)}} = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \dfrac{E(X\cdot Y) - E(X)\cdot E(Y)}{\sqrt{E(X^2) - E(X)^2}\cdot\sqrt{E(Y^2)-E(Y)^2}}$, der zwischen -1 und 1 liegt.

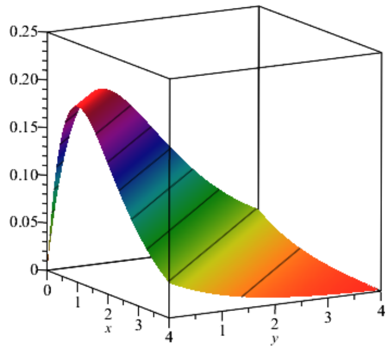
Beispiel 1b: $\mu_X = 1$, $\sigma_X = 0{,}5$, $\mu_Y = 2$, $\sigma_X = 1$, $\rho = 0{,}4$. ($\rho > 0$ bedeutet: mit x nimmt auch y zu)



Beispiel 1c: $\mu_X = 1$, $\sigma_X = 0{,}5$, $\mu_Y = 2$, $\sigma_X = 1$, $\rho = -0{,}4$. ($\rho < 0$ bedeutet: mit wachsendem x nimmt y ab)



Beispiel 2: Die beiden Zufallsvariablen X und Y haben die gemeinsame Dichtefunktion $f(x,y)=\begin{cases}\frac{1}{2}\cdot(x+y)\cdot e^{-x-y} & \text{für } x,y\ge 0\\ 0 & \text{sonst}\end{cases}$.

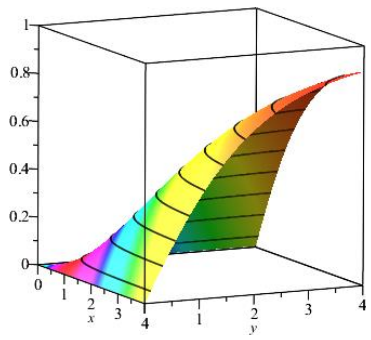


Es liegt wirklich eine Dichtefunktion vor, denn

$$\int_0^\infty\int_0^\infty f(x,y)\,dx\,dy=\frac{1}{2}\int_0^\infty\int_0^\infty (x+y)\cdot e^{-x-y}dx\,dy=-\frac{1}{2}\int_0^\infty\left[(x+y+1)\cdot e^{-x-y}\right]_0^\infty dy=$$

$$=\frac{1}{2}\int_0^\infty (y+1)\cdot e^{-y}dy=-\frac{1}{2}\left[(y+2)\cdot e^{-y}\right]_0^\infty=1\,.$$

Für die Verteilungsfunktion $F(x,y)$ gilt



$$F(x,y)=P(X\le x\wedge Y\le y)=\frac{1}{2}\int_0^y\int_0^x (x+y)\cdot e^{-x-y}dx\,dy=$$

$$=\frac{1}{2}\int_0^y\left((y+1)\cdot e^{-y}-(x+2+y)\cdot e^{-x-y}\right)dy=$$

$$=1-\frac{1}{2}\left((x+2)\cdot e^{-x}+(y+2)\cdot e^{-y}-(x+2+y)\cdot e^{-x-y}\right).$$

Z.B. $P(X\le 0\wedge Y\le 0)=F(0,0)=0$, $P(X\le 4\wedge Y\le 0)=F(4,0)=0$,
$P(X\le 0\wedge Y\le 4)=F(0,4)=0$, $P(X\le 4\wedge Y\le 4)=F(4,4)=1-6e^{-4}+5e^{-8}\approx 0{,}8918$.

Wie erhält man die Dichtefunktionen $f_X(x)$ und $f_Y(y)$ für x und y allein? Dazu erinnern wir uns an die Kontingenztafel mit Wahrscheinlichkeiten, z.B.

	y_1	y_2	y_3	
x_1	0,2	0,1	0,1	$P(X=x_1)=0{,}4$
x_2	0,3	0,2	0,1	$P(X=x_2)=0{,}6$
	$P(Y=y_1)=0{,}5$	$P(Y=y_2)=0{,}3$	$P(Y=y_3)=0{,}2$	1

Die Verteilung von X erhält man durch Addition über alle möglichen y-Werte:

$P(x_1)=P(x_1\wedge y_1)+P(x_1\wedge y_2)+P(x_1\wedge y_3)=0{,}2+0{,}1+0{,}1=0{,}4$

$P(x_2)=P(x_2\wedge y_1)+P(x_2\wedge y_2)+P(x_2\wedge y_3)=0{,}3+0{,}2+0{,}1=0{,}6$.

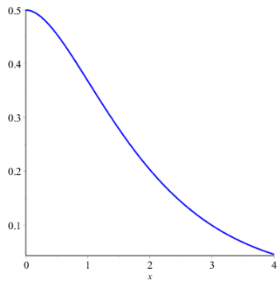
Die Verteilung von Y erhält man durch Addition über alle möglichen x-Werte:

$P(y_1)=P(x_1\wedge y_1)+P(x_2\wedge y_1)=0{,}2+0{,}3=0{,}5$

$P(y_2)=P(x_1\wedge y_2)+P(x_2\wedge y_2)=0{,}1+0{,}2=0{,}3$

$P(y_3)=P(x_1\wedge y_3)+P(x_2\wedge y_3)=0{,}1+0{,}1=0{,}2$.

Beim Übergang zu stetigen Funktionen ist die Addition einfach nur durch ein Integral zu ersetzen. Die beiden Dichtefunktionen sind



$$f_X(x)=\int_0^\infty f(x,y)\,dy=\frac{1}{2}\int_0^\infty (x+y)\cdot e^{-x-y}dy=-\frac{1}{2}\left[(x+y+1)\cdot e^{-x-y}\right]_{y=0}^\infty=\frac{1}{2}(x+1)\cdot e^{-x},$$

$$f_Y(y)=\int_0^\infty f(x,y)\,dx=\frac{1}{2}\int_0^\infty (x+y)\cdot e^{-x-y}dx=-\frac{1}{2}\left[(x+y+1)\cdot e^{-x-y}\right]_{x=0}^\infty=\frac{1}{2}(y+1)\cdot e^{-y}.$$

Sind die beiden Zufallsvariablen X und Y (stochastisch) unabhängig?

Für $x, y \ge 0$ ist $f(x,y) = \frac{1}{2}\cdot(x+y)\cdot e^{-x-y}$, während $f_X(x)\cdot f_Y(y) = \frac{1}{4}(x+1)(y+1)\cdot e^{-x-y}$ ein anderes Ergebnis liefert. Also sind X und Y abhängig.

Jetzt lassen sich auch die beiden Erwartungswerte $E(X)$ und $E(Y)$ bestimmen.

$E(X) = \int_0^\infty x\cdot f_X(x)\,dx = \frac{1}{2}\int_0^\infty x(x+1)e^{-x}\,dx = -\frac{1}{2}\Big[(x^2+3x+3)\cdot e^{-x}\Big]_0^\infty = \frac{3}{2}$. Analog folgt $E(Y) = \frac{3}{2}$.

Mit $E(X^2) = \int_0^\infty x^2\cdot f_X(x)\,dx = \frac{1}{2}\int_0^\infty x^2(x+1)e^{-x}\,dx = -\frac{1}{2}\Big[(x^3+4x^2+8x+8)e^{-x}\Big]_0^\infty = 4$ folgt

$\mathrm{Var}(X) = E(X^2) - E(X)^2 = 4 - \left(\frac{3}{2}\right)^2 = \frac{7}{4}$. Ebenso folgt $\mathrm{Var}(Y) = \frac{7}{4}$.

Außerdem ist

$$E(X\cdot Y) = \int_0^\infty\int_0^\infty x\cdot y\cdot f(x,y)\,dx\,dy) = \int_0^\infty\int_0^\infty x\cdot y\cdot(x+y)\cdot e^{-x-y}\,dx\,dy = -\frac{1}{2}\int_0^\infty\Big[(x^2+xy+2x+y+2)\cdot y\cdot e^{-x-y}\Big]_{x=0}^\infty dy =$$

$$= \frac{1}{2}\int_0^\infty (y+2)\cdot y\cdot e^{-y}\,dy = -\frac{1}{2}\Big[(y^2+4y+4)\cdot e^{-y}\Big]_0^\infty = 2\,.$$

Insgesamt folgt $\mathrm{Cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y) = 2 - \frac{3}{2}\cdot\frac{3}{2} = -\frac{1}{4}$, bzw. der Korrelationskoeffizient

$$r = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\cdot\sqrt{\mathrm{Var}(y)}} = \frac{-1/4}{\sqrt{3/2}\cdot\sqrt{3/2}} = -\frac{1}{6}\,.$$

Das negative Vorzeichen zeigt, dass mit wachsendem x das y überwiegend abnimmt.

Wegen $\mathrm{Cov}(X,Y) \ne 0$ sind X und Y abhängig, denn es gilt der Satz:
Wenn X, Y unabhängig, dann $\mathrm{Cov}(X,Y) = 0$.
Und die Kontraposition: Wenn $\mathrm{Cov}(X,Y) \ne 0$, dann X, Y nicht unabhängig.

Die geometrische Verteilung (Warten auf den ersten Erfolg)

Beim **Bernoulli-Experiment** gibt es genau zwei Ausgänge; Treffer – Niete oder auch Erfolg – Misserfolg.
Die Wahrscheinlichkeit für einen Treffer sei p, die für eine Niete ist dann $q = 1 - p$.
Die **Binomialverteilung** gibt Auskunft über die Wahrscheinlichkeit, bei n Versuchen k Treffer zu erzielen:
$P(X = k) = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$, da es bei k Treffern $n - k$ Nieten geben muss.

Die geometrische Verteilung gibt Auskunft über die Wahrscheinlichkeit, erst beim k-ten Versuch einen Treffer zu erzielen, wobei $k \in \mathbb{N} = \{1,2,3,\ldots\}$.

Wenn X die Nummer des Versuchs angibt, bei dem der erste Treffer auftritt, dann gilt
$P(X=1) = p$, $P(X=2) = p\cdot(1-p) = p\cdot q$, $P(X=3) = p\cdot(1-p)^2 = p\cdot q^2$,...

Allgemein: $\boxed{P(X=k) = p\cdot(1-p)^{k-1} = p\cdot q^{k-1},\quad k\in\mathbb{N}}$

Beispiel: Beim idealen Würfel wird eine sechs mit der Wahrscheinlichkeit $p = \frac{1}{6}$ gewürfelt. Dann gilt

$P(X=1) = \frac{1}{6} = 0{,}1\overline{6}$, $P(X=2) = \frac{1}{6}\cdot\frac{5}{6} = \frac{5}{36} = 0{,}13\overline{8}$, $P(X=3) = \frac{1}{6}\cdot\left(\frac{5}{6}\right)^2 = \frac{25}{216} = 0{,}115\overline{740}$, ...

$P(X=k)$ ist eine streng monoton fallende Folge.

Ist die Summe aller Wahrscheinlichkeiten wirklich 1?

$\sum_{k=1}^{\infty} P(X=k) = \sum_{k=1}^{\infty} p \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} q^{k-1} = p \cdot (1+q+q^2+q^3+...) = p \cdot s$. Dabei ist $s = 1+q+q^2+q^3+...$ eine unendliche geometrische Reihe. Man bildet die Differenz $s - q \cdot s = (1+q+q^2+q^3+...) - (q+q^2+q^3+...) = 1$ und erhält $s \cdot (1-q) = 1$, also $s = \frac{1}{1-q} = \frac{1}{p}$ für $-1 < q < 1$, so dass dann $\sum_{k=1}^{\infty} P(X=k) = p \cdot s = p \cdot \frac{1}{p} = 1$ gilt.

Behauptung: $\boxed{E(X) = \frac{1}{p}}$

$$E(X) = \sum_{k=1}^{\infty} k \cdot P(X=k) = \sum_{k=1}^{\infty} k \cdot p \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} \frac{d}{dq} q^k = p \cdot \frac{d}{dq} \sum_{k=1}^{\infty} q^k = p \cdot \frac{d}{dq}\left(q+q^2+q^3+...\right) = .$$

$$p \cdot \frac{d}{dq}\left(1+q+q^2+q^3+...-1\right) = p \cdot \frac{d}{dq}\left(\frac{1}{1-q} - 1\right) = p \cdot \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Beispiel: Im Mittel erhält man beim 6. Wurf eine sechs, da $E(X) = \frac{1}{1/6} = 6$.

Bemerkung; Hätte man auf den ersten Blick nicht eine kleinere Zahl, etwa 3,5 erwartet?
Man erhält dann eine kleinere Zahl, wenn eine einmal erhaltene Zahl nicht nochmals auftreten darf (ohne Zurücklegen):

$E = 1 \cdot$ P(sechs beim 1. Zug) + $2 \cdot$ P(sechs erst beim 2. Zug) $+ ... +$ $6 \cdot$ P(sechs erst beim 6. Zug) =

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{5} + 3 \cdot \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{1}{4} + \quad + 6 \cdot \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = (1+2+3+4+5+6) \cdot \frac{1}{6} = \frac{21}{6} = 3{,}5 .$$

Behauptung: $\boxed{V(X) = \frac{1-p}{p^2} = \frac{q}{p^2}}$

Es ist $V(X) = E(X^2) - E(X)^2$.

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \cdot P(X=k) = \sum_{k=1}^{\infty} k^2 \cdot p \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} k^2 \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot (k+1) \cdot q^{k-1} - p \cdot \sum_{k=1}^{\infty} k \cdot q^{k-1} =$$

$= p \cdot \sum_{k=1}^{\infty} k \cdot (k+1) \cdot q^{k-1} - \frac{1}{p}$, da der zweite Summand gerade $E(X)$ ist.

$$= p \cdot \sum_{k=1}^{\infty} \frac{d^2}{dq^2} q^{k+1} - \frac{1}{p} = p \cdot \frac{d^2}{dq^2} \sum_{k=1}^{\infty} q^{k+1} - \frac{1}{p} = p \cdot \frac{d^2}{dq^2}\left(q^2+q^3+q^4+...\right) - \frac{1}{p} = p \cdot \frac{d^2}{dq^2}\left(\frac{1}{1-q} - 1 - q\right) - \frac{1}{p} =$$

$$= p \cdot \frac{2}{(1-q)^3} - \frac{1}{p} = \frac{2p}{p^3} - \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}.$$

Insgesamt folgt $V(X) = E(X^2) - E(X)^2 = \frac{2}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$.

Beispiel: Für $p = \frac{1}{6}$ folgt $V(X) = \frac{1-p}{p^2} = \frac{1-1/6}{1/36} = 30$.

Beispiel 1: Wie oft muss man einen Würfel werfen, bis jede der sechs Zahlen mindestens einmal auftritt?

Man würfelt zum ersten Mal. Es erscheint eine Zahl mit der Wahrscheinlichkeit $p_1 = \frac{6}{6}$.

Die Wahrscheinlichkeit, beim zweiten Wurf eine andere Zahl zu erhalten, beträgt $p_2 = \frac{5}{6}$.

Die Wahrscheinlichkeit, beim dritten Wurf eine andere Zahl zu erhalten, beträgt $p_3 = \frac{4}{6}$, ...

Die Wahrscheinlichkeit, beim sechsen Wurf die noch nicht vorkommende Zahl zu erhalten, beträgt $p_6 = \frac{1}{6}$.

Somit muss man im Mittel $\frac{1}{p_1}+\frac{1}{p_2}+...+\frac{1}{p_6}=\frac{6}{6}+\frac{6}{5}+\frac{6}{4}+\frac{6}{3}+\frac{6}{2}+\frac{6}{1}=\frac{147}{10}=14,7$ mal würfeln, bis jede der sechs Zahlen mindestens einmal auftritt.

Beispiel 2: Die Erfolgswahrscheinlichkeit bei einem Experiment sei p. Wie oft (n) muss dieses Experiment mindestens durchgeführt werden, damit dabei mit einer Wahrscheinlichkeit von mindestens α zumindest ein Erfolg auftritt?

Es ist $(1-p)^n$ die Wahrscheinlichkeit, dass bei n aufeinander folgenden Versuchen nur Misserfolge erzielt werden. Folglich ist $1-(1-p)^n$ die Wahrscheinlichkeit, bei n aufeinander folgenden Versuchen mindestens einen Erfolg zu haben. Aus $1-(1-p)^n \geq \alpha$ folgt $(1-p)^n \leq 1-\alpha$, so dass $n\cdot\ln(1-p)\leq\ln(1-\alpha)$, also $n \geq \frac{\ln(1-\alpha)}{\ln(1-p)}$.

Zahlenbeispiel: Der Ball geht bei einem Schuss mit $p=0,3$ ins Tor. Mit der Wahrscheinlichkeit von $\alpha=0,9$ soll ein Treffer erzielt werden. Wegen $n\geq\frac{\ln(1-0,9)}{\ln(1-0,3)}=\frac{\ln(0,1)}{\ln(0,7)}\approx 6,5$ muss die Mannschaft zu mindestens 7 Torschüssen kommen.

Die negative Binomialverteilung (Warten auf den r-ten Erfolg)

Ein Bernoulli-Experiment mit der Trefferwahrscheinlichkeit p soll mehrfach durchgeführt werden. Bei der geometrischen Verteilung haben wir untersucht, mit welcher Wahrscheinlichkeit erst beim k-ten Versuch ein Treffer erzielt wird.

Wenn also X die **Nummer des Versuchs mit dem ersten Treffer** bezeichnet, dann gilt nach der geometrischen Verteilung $P(X=k)=p\cdot(1-p)^{k-1}=p\cdot q^{k-1}$ für $k=1,2,3,...$. Außerdem ist $E(X)=\frac{1}{p}$ und $V(X)=\frac{1-p}{p^2}=\frac{q}{p^2}$.

Falls aber X die **Anzahl der Nieten vor dem ersten Treffer** angibt, dann gilt $P(X=k)=p\cdot(1-p)^k=p\cdot q^k$ für $k=0,1,2,...$ und $E(X)=\frac{1}{p}\cdot q=\frac{q}{p}$, da in diesem Fall jede Wahrscheinlichkeit mit q multipliziert wird.

Die Varianz $V(X)=\frac{1-p}{p^2}=\frac{q}{p^2}$ bleibt aber unverändert. Herleitung wie bei der geometrischen Verteilung.

Im Folgenden sei X die Anzahl der Nieten vor dem der r-ten Treffer, $r=1, 2, 3, ...$. Außerdem sei $P(X=k)$ für $k=0,1,2,...$ die Wahrscheinlichkeit, vor dem r-ten Treffer k Nieten zu haben.

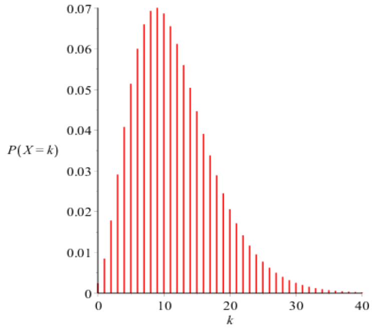
Wie groß ist $P(X=k)$?
Bei $X=k$ liegen vor dem r-ten Treffer genau k Nieten und $r-1$ Treffer, also insgesamt $k+r-1$ Versuche.
Die Wahrscheinlichkeit für eine beliebige Reihenfolge dieser $r-1$ Treffer und k Nieten beträgt $p^{r-1}\cdot(1-p)^k$.
Und die Wahrscheinlichkeit, dass daraufhin ein Treffer erscheint, beträgt somit $p^{r-1}\cdot(1-p)^k\cdot p=p^r\cdot(1-p)^k$.
Zur Bestimmung der Anzahl der verschiedenen Reihenfolgen müssen die k Nieten und $r-1$ Treffer auf die $k+r-1$ Plätze verteilt werden, und zwar ohne Berücksichtigung der Reihenfolge. Diese Anzahl beträgt $\binom{k+r-1}{k}$ oder auch $\binom{k+r-1}{r-1}$. Folglich ist $P(X=k)=\binom{k+r-1}{k}\cdot p^r\cdot(1-p)^k$.

Zum Erwartungswert: Man kann das Warten auf den r-ten Treffer als Folge von r-fachem Warten auf den nächsten Treffer interpretieren. Somit werden die r Erwartungswerte $\frac{q}{p}$ für jedes einzelne Warten addiert, was

$\boxed{E(X)=r\cdot\frac{q}{p}}$ ergibt. Analog folgt $\boxed{V(X)=r\cdot\frac{q}{p^2}}$.

Definition: Eine Zufallsvariable X heißt **negativ binomialverteilt (Nb(r,p)-verteilt)** oder auch **Pascal-verteilt** mit den Parametern r und p, wenn

$$\boxed{P(X=k)=\binom{k+r-1}{k}\cdot p^r\cdot(1-p)^k}$$. Dabei ist $r\in\mathbb{N}$, $0<p<1$ und $k\in\mathbb{N}_0$.



Beispiel: Ein Bernoulli-Experiment mit $p=0,3$ wird mehrfach ausgeführt. Es sei X die Anzahl der Nieten vor dem r = 5-ten Treffer. Dann gilt

$$P(X=k)=\binom{k+r-1}{k}\cdot p^r\cdot(1-p)^k=$$
$$=\binom{k+4}{k}\cdot 0,3^5\cdot 0,7^k \quad \text{für } k=0,1,2,\ldots$$

$P(X=0)=\binom{4}{0}\cdot 0,3^5\cdot 0,7^0=0,3^5=0,00243$, $P(X=1)=\binom{5}{1}\cdot 0,3^5\cdot 0,7^1=0,008505$,

$P(X=2)=\binom{6}{2}\cdot 0,3^5\cdot 0,7^2=0,0178605$, $P(X=9)=\binom{13}{9}\cdot 0,3^5\cdot 0,7^9=0,07011237447$ als Maximum.

Außerdem ist $E(X)=r\cdot\frac{q}{p}=5\cdot\frac{0,7}{0,3}=\frac{35}{3}\approx 11,7$, d.h. im Mittel treten 11,7 Nieten vor dem fünften Treffer auf.

Die Varianz ergibt sich zu $V(X)=5\cdot\frac{0,7}{0,3^2}=\frac{350}{9}\approx 38,9$.

Begründung für den Namen „negativ":

Es ist $\binom{k+r-1}{k}=\frac{(k+r-1)!}{k!\cdot(r-1)!}=\frac{r\cdot(r+1)\cdot\ldots\cdot(k+r-2)\cdot(k+r-1)}{k!}$. Der Zähler besteht aus genau k Faktoren.

Somit kann man dies umformen in $(-1)^k\cdot\frac{(-r)\cdot(-r-1)\cdot\ldots\cdot(-r-(k-2))\cdot(-r-(k-1))}{k!}=(-1)^k\cdot\binom{-r}{k}$.

Somit gilt auch

$$\boxed{P(X=k)=\binom{k+r-1}{k}\cdot p^r\cdot(1-p)^k=(-1)^k\cdot\binom{-r}{k}p^r\cdot(1-p)^k=(-1)^k\cdot\binom{-r}{k}p^r\cdot q^k=\binom{-r}{k}p^r\cdot(-q)^k}.$$

Zur Erinnerung: $\binom{8}{3}=\frac{8!}{3!\cdot 5!}=\frac{8\cdot 7\cdot 6}{3\cdot 2\cdot 1}$, d.h. im Zähler und Nenner jeweils drei absteigende Faktoren.

Weitere Beispiele: $\binom{-2}{3}=\frac{-2\cdot(-3)\cdot(-4)}{3\cdot 2\cdot 1}=-4$, $\binom{-0,3}{4}=\frac{-0,3\cdot(-1,3)\cdot(-2,3)\cdot(-3,3)}{4\cdot 3\cdot 2\cdot 1}=0,1233375$.

Zur Erinnerung: Für $|x|<1$ folgt mit Hilfe der Taylor-Reihe um $x=0$ die binomische Reihe

$$(1+x)^\alpha=\binom{\alpha}{0}\cdot x^0+\binom{\alpha}{1}\cdot x^1+\binom{\alpha}{2}\cdot x^2+\ldots=\sum_{k=0}^{\infty}\binom{\alpha}{k}\cdot x^k \quad\text{mit}\quad \binom{\alpha}{k}=\frac{\alpha\cdot(\alpha-1)\cdot(\alpha-2)\cdot\ldots\cdot(\alpha-k+1)}{k!}.$$

Beispiele zur binomischen Reihe:

① $(1+x)^2=\sum_{k=0}^{\infty}\binom{2}{k}\cdot x^k=\binom{2}{0}\cdot x^0+\binom{2}{1}\cdot x^1+\binom{2}{2}\cdot x^2+\binom{2}{3}\cdot x^3+\binom{2}{4}\cdot x^4+\ldots=1+2x+x^2$, denn

$\binom{2}{3}=\frac{2\cdot 1\cdot 0}{3!}=0$, $\binom{2}{4}=\frac{2\cdot 1\cdot 0\cdot(-1)}{4!}=0$, … .

② $\sqrt{1+x} = (1+x)^{1/2} = \sum_{k=0}^{\infty} \binom{1/2}{k} \cdot x^k = \binom{1/2}{0} \cdot x^0 + \binom{1/2}{1} \cdot x^1 + \binom{1/2}{2} \cdot x^2 + \binom{1/2}{3} \cdot x^3 + \binom{1/2}{4} \cdot x^4 + \ldots =$

$$= 1 + \frac{\frac{1}{2}}{1!} x + \frac{\frac{1}{2} \cdot \left(-\frac{1}{2}\right)}{2!} x^2 + \frac{\frac{1}{2} \cdot \left(-\frac{1}{2}\right) \cdot \left(-\frac{3}{2}\right)}{3!} x^3 + \ldots = 1 + \frac{1}{2} x - \frac{1}{8} x^2 + \frac{1}{16} x^3 - \frac{5}{128} x^4 \pm \ldots \ .$$

③ $\frac{1}{\sqrt{2-3x}} = \frac{1}{\sqrt{2} \cdot \sqrt{1-\frac{3}{2}x}} = \frac{1}{\sqrt{2}} \cdot \left(1-\frac{3}{2}x\right)^{-1/2} =$

$$= \frac{1}{\sqrt{2}} \cdot \left(\binom{-1/2}{0} \cdot \left(-\tfrac{3}{2}x\right)^0 + \binom{-1/2}{1} \cdot \left(-\tfrac{3}{2}x\right)^1 + \binom{-1/2}{2} \cdot \left(-\tfrac{3}{2}x\right)^2 + \binom{-1/2}{3} \cdot \left(-\tfrac{3}{2}x\right)^3 + \ldots \right) =$$

$$= \frac{1}{\sqrt{2}} \cdot \left(1 + \left(-\tfrac{1}{2}\right) \cdot \left(-\tfrac{3}{2}x\right) + \frac{-\frac{1}{2} \cdot \left(-\frac{3}{2}\right)}{2!} \cdot \left(-\tfrac{3}{2}x\right)^2 + \frac{-\frac{1}{2} \cdot \left(-\frac{3}{2}\right) \cdot \left(-\frac{5}{2}\right)}{3!} \cdot \left(-\tfrac{3}{2}x\right)^3 + \frac{-\frac{1}{2} \cdot \left(-\frac{3}{2}\right) \cdot \left(-\frac{5}{2}\right) \cdot \left(-\frac{7}{2}\right)}{4!} \cdot \left(-\tfrac{3}{2}x\right)^4 + \ldots \right) =$$

$$= \frac{1}{\sqrt{2}} \cdot \left(1 + \tfrac{3}{4}x + \tfrac{27}{32}x^2 + \tfrac{135}{128}x^3 + \tfrac{2835}{2048}x^4 + \ldots\right) \text{ für } |x| < \tfrac{2}{3}.$$

Die hypergeometrische Verteilung

Gegeben ist eine Menge von N Elementen. M von ihnen, $M \le N$, haben eine gewünschte Eigenschaft. Nun zieht man n Kugeln ohne Zurücklegen. Mit welcher Wahrscheinlichkeit $P(X = k)$ befinden sich unter diesen n Kugeln genau k mit der gewünschten Eigenschaft?

Beispiel: Beim Zahlenlotto „6 aus 49“ werden aus $N = 49$ Zahlen genau $M = 6$ Gewinnzahlen gezogen. Auf dem Lottoschein werden $n = 6$ Zahlen angekreuzt. X sei die Anzahl der richtig angekreuzten Zahlen. Dann gilt

$$P(X=0) = \frac{\binom{43}{6}}{\binom{49}{6}} = \frac{6\,096\,454}{13983816} \approx 0{,}435965\,, \quad P(X=1) = \frac{\binom{6}{1} \cdot \binom{43}{5}}{\binom{49}{6}} = \frac{5\,775\,588}{13983816} \approx 0{,}413019\,,$$

$$P(X=2) = \frac{\binom{6}{2} \cdot \binom{43}{4}}{\binom{49}{6}} = \frac{1851150}{13983816} \approx 0{,}132378\,, \quad P(X=3) = \frac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}} = \frac{246820}{13983816} \approx 0{,}017650\,,$$

$$P(X=4) = \frac{\binom{6}{4} \cdot \binom{43}{2}}{\binom{49}{6}} = \frac{13545}{13983816} \approx 0{,}000969\,, \quad P(X=5) = \frac{\binom{6}{5} \cdot \binom{43}{1}}{\binom{49}{6}} = \frac{258}{13983816} \approx 0{,}000018\,,$$

$$P(X=6) = \frac{\binom{6}{6} \cdot \binom{43}{0}}{\binom{49}{6}} = \frac{1}{13983816} \approx 0{,}000\,000\,071511\,, \quad \text{analog } P(X=0) = \frac{\binom{43}{6}}{\binom{49}{6}} = \frac{\binom{6}{0} \cdot \binom{43}{6}}{\binom{49}{6}}.$$

Allgemein gilt $\boxed{P(X=k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}}$.

Außerdem gilt

$$E(X)=\sum_{k=0}^{n}k\cdot\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}=\sum_{k=1}^{n}k\cdot\frac{\frac{M}{k}\cdot\binom{M-1}{k-1}\cdot\binom{N-M}{n-k}}{\frac{N}{n}\cdot\binom{N-1}{n-1}}=n\cdot\frac{M}{N}\cdot\sum_{k=1}^{n}\frac{\binom{M-1}{k-1}\cdot\binom{N-1-(M-1)}{n-1-(k-1)}}{\binom{N-1}{n-1}}=$$

$$n\cdot\frac{M}{N}\cdot\sum_{i=0}^{n-1}\frac{\binom{M-1}{i}\cdot\binom{N-1-(M-1)}{n-1-i}}{\binom{N-1}{n-1}}.$$

Diese Summe hat den Wert 1, da alle Wahrscheinlichkeiten addiert werden, die zur hypergeometrischen Verteilung mit $N-1$ Elementen gehören, von denen $M-1$ die gewünschte Eigenschaft haben. Und man zieht $n-1$ Kugeln. Somit gilt $E(X)=n\cdot\frac{M}{N}$.

Beim Lotto ($n=6$ Kreuze, $M=6$ Gewinnzahlen, $N=49$ Zahlen zur Auswahl) heißt dies, dass im Mittel $E(X)=6\cdot\frac{6}{49}=\frac{36}{49}\approx 0,7347$ Zahlen richtig angekreuzt sind.

Faltung von Wahrscheinlichkeitsverteilungen

Definition: Es seien X und Y zwei unabhängige Zufallsvariablen, die auf demselben Wahrscheinlichkeitsraum definiert sind. Dann heißt die Verteilung von $X+Y$ eine Faltung der beiden Verteilungen von X und Y.

Beispiel: Es werde mit zwei idealen Würfeln gewürfelt. X_1 und X_2 seien die Zufallsvariablen für die Augenzahlen von Würfel 1 und 2. Dann ist $S=X_1+X_2$ die Zufallsvariable für die Summe der beiden Augenzahlen. Dann gilt wegen der Unabhängigkeit von X_1 und X_2 :

$$P(S=2)=P(X_1=1\wedge X_2=1)=P(X_1=1)\cdot P(X_2=1)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}.$$

$$P(S=3)=P\big((X_1=1\wedge X_2=2)\vee(X_1=2\wedge X_2=1)\big)=P(X_1=1\wedge X_2=2)+P(X_1=2\wedge X_2=1)=$$

$$=P(X_1=1)\cdot P(X_2=2)+P(X_1=2)\cdot P(X_2=1)=\frac{1}{36}+\frac{1}{36}=\frac{2}{36}.$$

$$P(S=4)=\sum_{i=1}^{3}P(X_1=i)\cdot P(X_2=4-i)=\frac{3}{36}.$$ Für dieses Beispiel gilt $P(S=s)=\frac{6-|7-s|}{36}$.

Allgemein gilt im **diskreten** Fall, wenn also X und Y nur ganzzahlige Werte annehmen können,

$$\boxed{P(X+Y=k)=\sum_{i\in\mathbb{Z}}P(X=i\wedge Y=k-i)=\sum_{i\in\mathbb{Z}}P(X=i)\cdot P(Y=k-i)}.$$ **Diskrete Faltungsformel**

Wegen der Unabhängigkeit von X und Y gilt $P(X=i\wedge Y=j)=P(X=i)\cdot P(Y=j)$.

Im **stetigen** Fall ergibt sich $\boxed{P(X+Y=k)=\int_{-\infty}^{\infty}f_X(t)\cdot f_Y(k-t)\,dt}$. **Stetige Faltungsformel**

Satz: Gegeben seien zwei binomialverteilte Zufallsvariablen X_1 und X_2 mit den Verteilungen $B_{n_1,p}$ und $B_{n_2,p}$ zur gleichen Wahrscheinlichkeit p. Die Summe $Z=X_1+X_2$ ist dann $B_{n_1+n_2,p}-\text{verteilt}$.

Beweis: $P(Z=k)=\sum_{i=0}^{k}P(X_1=i)\cdot P(X_2=k-i)=\sum_{i=0}^{k}\binom{n_1}{i}\cdot p^i\cdot q^{n_1-i}\cdot\binom{n_2}{k-i}\cdot p^{k-i}\cdot q^{n_2-k+i}=$

$$=p^k\cdot q^{n_1+n_2-k}\cdot\sum_{i=0}^{k}\binom{n_1}{i}\cdot\binom{n_2}{k-i}=\binom{n_1+n_2}{k}\cdot p^k\cdot q^{n_1+n_2-k}\cdot\sum_{i=0}^{k}\frac{\binom{n_1}{i}\cdot\binom{n_2}{k-i}}{\binom{n_1+n_2}{k}}=\binom{n_1+n_2}{k}\cdot p^k\cdot q^{n_1+n_2-k}.$$

Die obige Summe hat den Wert 1, da hier die Wahrscheinlichkeiten einer hypergeometrischen Verteilung addiert werden.

Man sagt, die Binomialverteilungen zu einem festen p bilden eine **Faltungsfamilie**.

Dieser Satz ist einleuchtend, denn wenn ein Bernoulli-Experiment mit der Trefferwahrscheinlichkeit p zuerst n_1-mal und dann n_2-mal durchgeführt wird, dann ist auch eine $(n_1+n_2)-$fache Durchführung binomial-verteilt.

Satz: Gegeben seien zwei unabhängige Poisson-verteilte Zufallsvariablen X_1 und X_2 mit den Parametern λ_1 und λ_2. Die Summe $Z = X_1 + X_2$ ist dann ebenfalls Poisson-verteilt mit dem Parameter $\lambda_1+\lambda_2$. Die Poisson-Verteilungen bilden also eine Faltungsfamilie.

Beweis: Es sei $P(X_1=k)=\frac{\lambda_1{}^k}{k!}\cdot e^{-\lambda_1}$ und $P(X_2=k)=\frac{\lambda_2{}^k}{k!}\cdot e^{-\lambda_2}$. Dann gilt

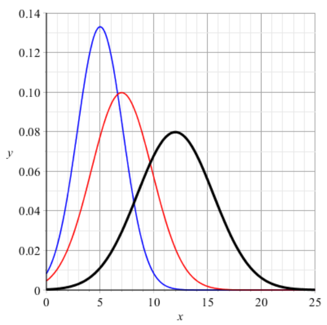
$$P(X_1+X_2=k)=\sum_{i=0}^{k}P(X_1=i)\cdot P(X_2=k-i)=\sum_{i=0}^{k}\frac{\lambda_1{}^i}{i!}\cdot e^{-\lambda_1}\cdot\frac{\lambda_2{}^{k-i}}{(k-i)!}\cdot e^{-\lambda_2}=e^{-\lambda_1-\lambda_2}\cdot\frac{1}{k!}\cdot\sum_{i=0}^{k}\frac{k!}{i!\cdot(k-i)!}\cdot\lambda_1{}^i\cdot\lambda_2{}^{k-i}=$$

$$=e^{-\lambda_1-\lambda_2}\cdot\frac{1}{k!}\cdot\sum_{i=0}^{k}\binom{k}{i}\cdot\lambda_1{}^i\cdot\lambda_2{}^{k-i}=e^{-\lambda_1-\lambda_2}\cdot\frac{1}{k!}\cdot(\lambda_1+\lambda_2)^k=\frac{(\lambda_1+\lambda_2)^k}{k!}\cdot e^{-\lambda_1-\lambda_2}.$$

Satz: Gegeben seien zwei unabhängige normalverteilte Zufallsvariablen X_1 und X_2 mit den Parametern $\mu_1,\ \sigma_1$ und $\mu_2,\ \sigma_2$. Die Summe $Z = X_1 + X_2$ ist dann ebenfalls normalverteilt mit den Parametern $\mu=\mu_1+\mu_2$ und $\sigma^2=\sigma_1^2+\sigma_2^2$.
Ohne Beweis.

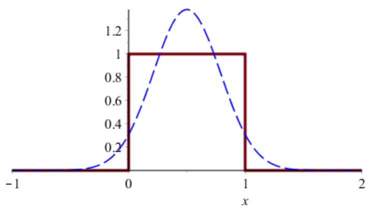
Die weiter unten behandelten Exponentialverteilungen bilden keine Familie.

Beispiel: Die Dauer der Telefonate eines Ehepaares seien unabhängig voneinander und normalverteilt mit den Parametern $\mu_1=5$, $\sigma_1=3$ und $\mu_2=7$, $\sigma_2=4$, jeweils in Minuten. Dann ist die Dauer zweier unabhängig voneinander geführten Gespräche normalverteilt mit $\mu=12$, $\sigma=\sqrt{3^2+4^2}=5$, in Minuten. Die schwarze Kurve zeigt die Summe (Faltung).



Beispiel: Ein Zufallsgenerator erzeuge Zufallszahlen x von 0 bis 1 einschließlich. Die dazugehörige Dichtefunktion ist $f(x)=\begin{cases}1 \text{ für } 0\le x\le 1\\ 0 \quad \text{sonst}\end{cases}$.

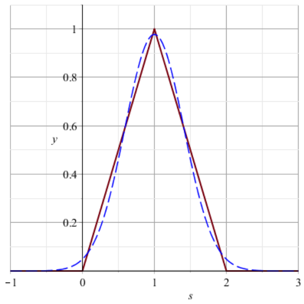
Siehe Schaubild.



Außerdem ist $\mu=\int_0^1 x\cdot f(x)\,dx=\frac{1}{2}$ und $\sigma^2=\int_0^1 (x-\mu)^2\cdot f(x)\,dx=\frac{1}{12}$.

Das Schaubild der zugehörigen Normalverteilung $n(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ ist gestrichelt eingezeichnet.

Nun betrachten wir die Summe von **zwei** unabhängig erzeugten Zufallszahlen von 0 bis 1 einschließlich. Für die Dichtefunktion $f_2(s)$ gilt dann

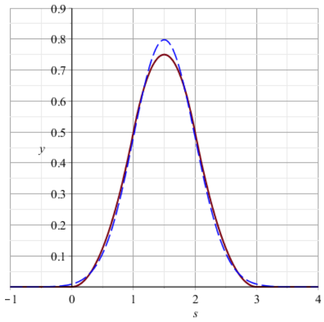


$$f_2(s)=\int_0^2 f(x)\cdot f(s-x)\,dx=\begin{cases}\int_0^2 f(x)\cdot 0\,dx=0 & \text{für } s<0\\ \int_0^s 1\cdot 1\,dx=s & \text{für } 0\le s\le 1\\ \int_{s-1}^1 1\cdot 1\,dx=2-s & \text{für } 1<s\le 2\\ \int_0^2 0\cdot f(s-x)\,dx=0 & \text{für } s>2\end{cases}.$$

Außerdem ist $\mu=\int_0^2 x\cdot f_2(x)\,dx=1$ und $\sigma^2=\int_0^2 (x-\mu)^2\cdot f_2(x)\,dx=\frac{1}{6}$.

Das Schaubild der zugehörigen Normalverteilung ist gestrichelt eingezeichnet.

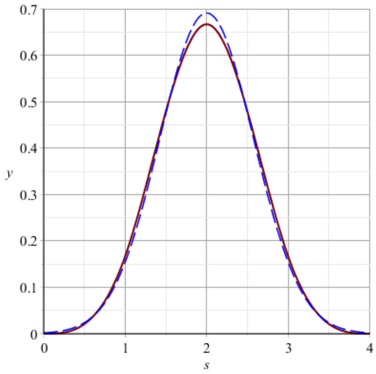
Für die Dichtefunktion der Summe von **drei** Zufallszahlen gilt



$$f_3(s)=\int_0^3\int_0^3 f(x)\cdot f(y)\cdot f(s-x-y)\,dy\,dx=\begin{cases}0 & \text{für } s<0\\ \frac{1}{2}s^2 & \text{für } 0\le s<1\\ -s^2+3s-\frac{3}{2} & \text{für } 1\le s<2\\ \frac{1}{2}s^2-3s+\frac{9}{2} & \text{für } 2\le s\le 3\\ 0 & \text{für } s>3\end{cases}.$$

Außerdem ist $\mu=\int_0^3 x\cdot f_3(x)\,dx=\frac{3}{2}$ und $\sigma^2=\int_0^3 (x-\mu)^2\cdot f_3(x)\,dx=\frac{1}{4}$.

Das Schaubild der zugehörigen Normalverteilung ist gestrichelt eingezeichnet.
Für die Dichtefunktion der Summe von **vier** Zufallszahlen gilt



$$f_4(s)=\int_0^4\int_0^4\int_0^4 f(x)\cdot f(y)\cdot f(z)\cdot f(s-x-y-z)\,dz\,dy\,dx=$$

$$=\begin{cases}0 & \text{für } s<0\\ \frac{1}{6}s^3 & \text{für } 0\le s<1\\ -\frac{1}{2}s^3+2s^2-2s+\frac{2}{3} & \text{für } 1\le s<2\\ \frac{1}{2}s^3-4s^2+10s-\frac{22}{3} & \text{für } 2\le s<3\\ -\frac{1}{6}s^3+2s^2-8s+\frac{32}{3} & \text{für } 3\le s\le 4\\ 0 & s>4\end{cases}.$$

Außerdem ist $\mu=\int_0^4 x\cdot f_4(x)\,dx=2$ und $\sigma^2=\int_0^4 (x-\mu)^2\cdot f_4(x)\,dx=\frac{1}{3}$.

Das Schaubild der zugehörigen Normalverteilung ist gestrichelt eingezeichnet.

Die χ^2 - Verteilung, der χ^2 - Unabhängigkeitstest

Im Kapitel bivariate Verteilung hatten wir mit Hilfe des korrigierten C-Koeffizienten $C^*=\sqrt{\frac{M}{M-1}\cdot\frac{\chi^2}{\chi^2+n}}$ nach Karl Pearson die Unabhängigkeit zweier Eigenschaften untersucht. Es gilt $0\le C^*\le 1$, wobei $C^*=0$ völlige Unabhängigkeit und $C^*=1$ völlige Abhängigkeit bedeutet. Dieser Unabhängigkeitstest lässt sich verfeinern.

Da bei der χ^2-Verteilung die Werte h^e_{ij} nicht unter 5 sein sollten, wurden alle Werte h_{ij} der Tabelle aus dem Kapitel bivariate Verteilung mit 10 multipliziert.

	Prüfergebnis						
Maschine	***1***		***2***		***3***		$\sum$
1	350	325	100	100	50	75	500
	625	25/13	0	0	625	25/3	
2	250	260	80	80	70	60	400
	100	5/13	0	0	100	5/3	
3	50	65	20	20	30	15	100
	225	45/13	0	0	225	15	
$\sum$	650		200		150		**n = 1000**

mit

h_{ij}	h^e_{ij}
$(h_{ij}-h^e_{ij})^2$	$\frac{(h_{ij}-h^e_{ij})^2}{h^e_{ij}}$

Der χ^2-Koeffizient wird $\chi^2=\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\left(h_{ij}-h^e_{ij}\right)^2}{h^e_{ij}}=\left(\frac{25}{13}+0+\frac{25}{3}\right)+\left(\frac{5}{13}+0+\frac{5}{3}\right)+\left(\frac{45}{13}+0+15\right)=\frac{400}{13}\approx 30{,}8$,

das Zehnfache des damaligen Ergebnisses. Aber $C^*=\sqrt{\frac{M}{M-1}\cdot\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{3}{2}\cdot\frac{400/13}{400/13+1000}}=\sqrt{\frac{3}{67}}\approx 0{,}212$

bleibt unverändert. Wir haben also keine Unabhängigkeit.

<u>Nullhypothese: Wir gehen von der Unabhängigkeit der beiden Eigenschaften aus</u> und fragen, ob unser C^* bzw. χ^2 die Abhängigkeit nur vortäuscht, also die beiden Eigenschaften in Wirklichkeit unabhängig sind.

Wenn man nun die drei Maschinen mehrfach 1000 Teile herstellen lässt, dann ist jede absolute Häufigkeit h_{ij} binomialverteilt mit dem Erwartungswert h^e_{ij}. Bei ausreichend großen Stückzahlen kann man die Binomialverteilung durch die Normalverteilung ausreichend genau annähern. Somit sind $h_{ij}-h^e_{ij}$ und auch

$\frac{h_{ij}-h^e_{ij}}{\sqrt{h^e_{ij}}}$ normalverteilt mit dem Erwartungswert 0. Nach Quadrieren gilt $\left(\frac{h_{ij}-h^e_{ij}}{\sqrt{h^e_{ij}}}\right)^2=\frac{\left(h_{ij}-h^e_{ij}\right)^2}{h^e_{ij}}\geq 0$.

Die Summe $\chi^2=\sum_i\sum_j\frac{\left(h_{ij}-h^e_{ij}\right)^2}{h^e_{ij}}$ von Normalverteilungen ist eine sogenannte $\boldsymbol{\chi^2}$ **- Verteilung**. Ihre Werte können nicht negativ werden.

Die Gestalt der Dichtefunktion hängt von der Zahl der Summanden ab; genauer gesagt, von der Zahl der unabhängigen Summanden.

In unserer Tabelle enthält die Zeile für Maschine 1 die Werte $h_{11}=350$, $h_{12}=100$ und $h_{13}=50$. Wegen der Summe $h_{11}+h_{12}+h_{13}=500$ sind nur zwei dieser drei Werte frei wählbar, der dritte lässt sich dann berechnen. Man spricht von zwei **Freiheitsgraden**

Wegen $h_{21}+h_{22}+h_{23}=400$ haben wir in der Zeile für die Maschine 2 auch genau zwei Freiheitsgrade.
Weil die Spaltensummen gegeben sind, lassen sich alle Werte in der Zeile der Maschine 3 berechnen. Es kommen keine Freiheitsgrade mehr dazu.
Unsere Tabelle hat vier Freiheitsgrade.

Allgemein gilt: Wenn in einer Kreuztabelle r die Anzahl der Reihen (rows) und c die Anzahl der Spalten (columns) ist, dann gilt für die Anzahl n der Freiheitsgrade $\boxed{n=(r-1)\cdot(c-1)}$.
Auch wird **df = degrees of freedom** für die Anzahl n der Freiheitsgrade verwendet.

Man kann nun die Dichtefunkton von χ^2 bei n Freiheitsgraden herleiten: $\boxed{f_n(x) = \begin{cases} 0 & \text{für} \quad x \le 0 \\ \dfrac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} & \text{für} \quad x > 0 \end{cases}}$

Dabei ist $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt$ die Eulersche Gammafunktion. Speziell gilt

① $\Gamma(1) = \int_0^\infty e^{-t}\, dt = \left[-e^{-t}\right]_0^\infty = 1$

② $\Gamma(x+1) = x \cdot \Gamma(x)$, denn durch partielle Integration, $\left(\int f \cdot g = F \cdot g - \int F \cdot g'\right)$, folgt $\Gamma(x+1) = \int_0^\infty e^{-t} t^x\, dt =$

$\left[-e^{-t} \cdot t^x\right]_{t=0}^{t=\infty} - \int_0^\infty -e^{-t}\, x \cdot t^{x-1}\, dt = 0 + x \cdot \int_0^\infty e^{-t}\, t^{x-1}\, dt = x \cdot \Gamma(x)$.

Folgerung: $\boxed{\Gamma(n+1) = n!}$ für $n = 0, 1, 2, 3, \ldots$, d.h., dass $\Gamma(x+1)$ die Fakultäten interpoliert.

③ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, denn mit der Substitution $t = \frac{1}{2}u^2$ folgt $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t} t^{-1/2}\, dt = \int_0^\infty e^{-\frac{1}{2}u^2} u^{-1}\, 2^{\frac{1}{2}} u\, du = \sqrt{2} \int_0^\infty e^{-\frac{1}{2}u^2} du$.

Wegen $\int_{-\infty}^\infty e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$, (siehe Normalverteilung), also $\int_0^\infty e^{-\frac{1}{2}x^2} dx = \frac{1}{2}\sqrt{2\pi}$, folgt $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

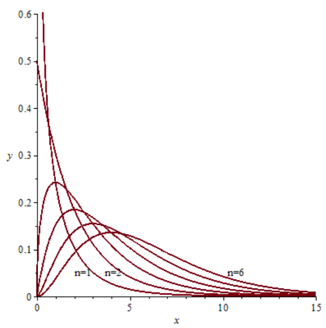
① Wie verhält sich $f_n(x)$ für $x \to 0$, wobei $x > 0$ gilt?

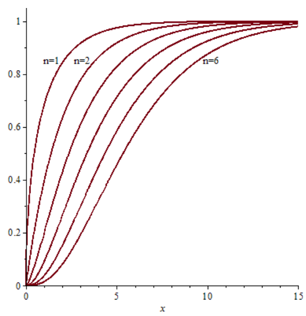
Für $n = 1$ und $x > 0$ gilt $f_1(x) = \dfrac{x^{-\frac{1}{2}} \cdot e^{-\frac{x}{2}}}{2^{\frac{1}{2}} \cdot \Gamma\left(\frac{1}{2}\right)} = \dfrac{e^{-\frac{x}{2}}}{\sqrt{2 \cdot \pi \cdot x}} \to \infty$ für $x \to 0$. Dies zeigt auch das Schaubild.

Für $n = 2$ und $x > 0$ gilt $f_2(x) = \dfrac{e^{-\frac{x}{2}}}{2 \cdot \Gamma(1)} = \dfrac{e^{-\frac{x}{2}}}{2} \to \dfrac{1}{2}$ für $x \to 0$ wie das Schaubild zeigt.

Für $n = 3, 4, 5, \ldots$ gilt $f_n(x) \to 0$ für $x \to 0$. Dies zeigt auch das Schaubild.

② Wie verhält sich $f_n(x)$ für $x \to \infty$? Wegen des Faktors $e^{-\frac{x}{2}}$ gilt $f_n(x) \to 0$ für $x \to \infty$.



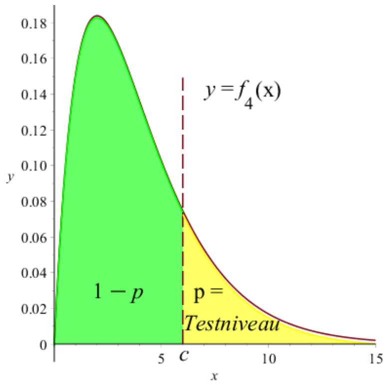


Im linken Schaubild sind die Dichtefunktionen $f_n(x)$, im rechten die Verteilungsfunktionen

$P\left(\chi^2 \le x\right) = F_n(x) = \int_0^x f_n(t)\, dt$ für $n = 1, 2, \ldots, 6$ dargestellt.

Für jede Dichtefunktion gilt $F_n(\infty) = \int_0^\infty f_n(x)\,dx = 1$. Dies erkennt man auch am rechten Schaubild.

Außerdem ist der Erwartungswert $E\left(\chi^2\right) = n$ und die Varianz $V\left(\chi^2\right) = 2n$; ohne Herleitung.
Alles immer bei angenommener Unabhängigkeit der beiden Eigenschaften.



Vorgehen beim Test:

1. Man bestimmt die Größe $\chi^2 = \sum_i \sum_j \frac{\left(h_{ij} - h_{ij}^e\right)^2}{h_{ij}^e}$.
2. Man bestimmt die Anzahl $n = (r-1)\cdot(c-1)$ der Freiheitsgrade und gibt das Testniveau p in % vor. Das heißt, dass die Nullhypothese der Unabhängigkeit mit einer Wahrscheinlichkeit von höchstens p% irrtümlich abgelehnt wird.
   Mit diesen beiden Werten entnimmt man der Tabelle den kritischen Wert c, der die beiden Flächen trennt.
3. Bei **angenommener Unabhängigkeit** der beiden Eigenschaften fällt χ^2 mit der Wahrscheinlichkeit $1-p$ in den linken Bereich und **mit der Wahrscheinlichkeit p in den rechten Bereich**. **Im rechten Bereich wird auf Abhängigkeit geschlossen.**

Wie bestimmt man die Werte dieser Tabelle?

Zu gegebenen Werten von n und p löst man die Gleichung $P\left(\chi^2 \geq c\right) = \int_c^\infty f_n(x)\,dx = p$ nach c auf.

c		**Testniveau p**								
		0,1%	**1%**	**2%**	**2,5%**	**5%**	**10%**	**20%**	**25%**	**50%**
Freiheitsgrad n	**1**	10,828	6,635	5,412	5,024	3,841	2,706	1,642	1,323	0,455
	2	13,816	9,210	7,824	7,378	5,991	4,605	3,219	2,773	1,386
	3	16,266	11,345	9,837	9,348	7,815	6,251	4,642	4,108	2,366
	4	18,467	13,277	11,668	11,143	9,488	7,779	5,989	5,385	3,357
	5	20,515	15,086	13,388	12,833	11,070	9,236	7,289	6,626	4,351
	6	22,458	16,812	15,033	14,449	12,592	10,645	8,558	7,841	5,348
	7	24,322	18,475	16,622	16,013	14,067	12,017	9,803	9,037	6,346
	8	26,124	20,090	18,168	17,535	15,507	13,362	11,030	10,219	7,344
	9	27,877	21,666	19,679	19,023	16,919	14,684	12,242	11,389	8,343
	10	29,588	23,209	21,161	20,483	18,307	15,987	13,442	12,549	9,342

In unserem Beispiel ist $n = (3-1)\cdot(3-1) = 4$ und $\chi^2 \approx 30{,}8$.

Unter der Annahme der Unabhängigkeit beider Eigenschaften wäre in nur $p = 0{,}1\%$ aller Tests $\chi^2 \geq 18{,}467$.
Mit $\chi^2 \approx 30{,}8$ liegen wir sogar weit rechts von $c = 18{,}467$, so dass die beiden Eigenschaften Maschine und Prüfergebnis so gut wie sicher abhängig sind.

In nur $p = 1\%$ aller Tests wäre bei Unabhängigkeit der beiden Eigenschaften $\chi^2 \geq 13{,}277$.
In nur $p = 2\%$ aller Tests wäre bei Unabhängigkeit der beiden Eigenschaften $\chi^2 \geq 11{,}668$.
….
In $p = 100\%$ aller Tests wäre bei Unabhängigkeit der beiden Eigenschaften $\chi^2 \geq 0$.

Mein Computer liefert $\int_{30{,}8}^{\infty} f_4(x)\,dx \approx 0{,}0000034$. Das bedeutet, dass unter der Annahme der Unabhängigkeit der beiden Eigenschaften in nur $0{,}00034\%$ aller Fälle (Tests mit $n = 1000$ produzierten Teilen) $\chi^2 \geq 30{,}8$ wäre.

Scheinbares Problem: Wenn z.B. die Häufigkeiten h_{ij} verdoppelt werden, dann verdoppelt sich auch der Wert von $\chi^2 = \sum_i \sum_j \frac{\left(h_{ij} - h_{ij}^e\right)^2}{h_{ij}^e}$. Somit kommt man mit zunehmenden Werten von h_{ij} immer zu größeren Werten von χ^2, die dann zur Ablehnung der Unabhängigkeit führen.

Lösung: Bei Unabhängigkeit müssen die Werte von $\frac{\left(h_{ij} - h_{ij}^e\right)^2}{h_{ij}^e}$ so gut wie Null sein, so dass χ^2 nicht so groß werden kann.

Der McNemar-Test bei der Vierfeldertafel

Der Test ist nach dem amerikanischen Psychologen Quinn Michael McNemar (1900-1986) benannt.
Dieser Test ist ein χ^2 – Test, bei dem zwei verbundene (an den gleichen Probanden erhoben) Stichproben für ein dichotomes (zwei verschiedene Ausprägungen) Merkmal erhoben werden.

Beispiel: Eine Firma möchte den Umsatz ihres Vanilleeises steigern. Dazu spendiert diese Firma den Studenten eines Statistik-Kurses jeweils eine große Portion dieses Eises. Vorher und nachher wurden die Studenten gefragt, ob sie dieses Eis mögen. Das Ergebnis ist in einer Vierfeldertafel dargestellt.

		Nachher		
		Nein	Ja	
Vorher	Nein	a	b	$a+b$
	Ja	c	d	$c+d$
		$a+c$	$b+d$	$a+b+c+d$

Die Frage ist nun, ob die Eisspende die Begeisterung für dieses Eis verändert hat.
Dabei interessieren die Werte a und d nicht, da sie keine Änderung angeben. Die Zahlen b und c beschreiben die Änderung.

Nullhypothese H_0: Es wechseln genauso viele Meinungen von ja nach nein wie umgekehrt und die Spende der Firma ist umsonst. Dann erwartet man für die beiden Werte b und c jeweils ihren Mittelwert $\frac{b+c}{2}$.
Somit ergibt sich

$$\chi^2 = \frac{\left(b - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} + \frac{\left(c - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} = \frac{\left(\frac{b-c}{2}\right)^2}{\frac{b+c}{2}} + \frac{\left(\frac{-b+c}{2}\right)^2}{\frac{b+c}{2}} = \frac{2\cdot\left(\frac{b-c}{2}\right)^2}{\frac{b+c}{2}} = 2\cdot\frac{(b-c)^2}{4}\cdot\frac{2}{b+c} = \frac{(b-c)^2}{b+c} = \frac{|b-c|^2}{b+c}.$$

Der Term $\frac{|b-c|^2}{b+c}$ ist χ^2 – verteilt mit $n = (\text{Zeilenzahl} - 1)\cdot(\text{Spaltenzahl} - 1) = 1\cdot 1 = 1$ Freiheitsgrad.
Hier der Ausschnitt der obigen Tabelle für $n = 1$.

	Testniveau p								
	0,1%	**1%**	**2%**	**2,5%**	**5%**	**10%**	**20%**	**25%**	**50%**
g	10,828	6,635	5,412	5,024	3,841	2,706	1,642	1,323	0,455

Die Werte b und c sind ganzzahlig, während die χ^2 – Verteilung stetig ist. Somit muss wie bei der Normalverteilung eine Korrektur angebracht werden, ähnlich wie die $0{,}5$ bei $P(X \le x) \approx \Phi\left(\frac{x + 0{,}5 - \mu}{\sigma}\right)$.

Yates-Korrektur: $\boxed{\chi^2 = \frac{\left(|b-c| - 0{,}5\right)^2}{b+c}}$. Diese Formel verwenden wir!

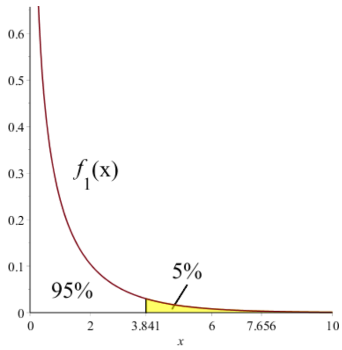
Edwards-Korrektur: $\chi^2 = \frac{(|b-c|-1)^2}{b+c}$.

Mediziner verwenden gerne $\chi^2 = \frac{(|b-c|-1)^2}{b+c+1}$ für $8 \le b+c \le 30$ und $\chi^2 = \frac{(b-c)^2}{b+c+1}$ für $b+c > 30$.
Die Korrekturen fallen nur bei kleinen Zahlen ins Gewicht.

Zahlenbeispiel:

		Nachher: Nein	Nachher: Ja	
Vorher	Nein	8	29	37
	Ja	11	34	45
		19	63	82

Yates ergibt: $\chi^2 = \frac{(|b-c|-0,5)^2}{b+c} = \frac{(|29-11|-0,5)^2}{40} = \frac{17,5^2}{40} = 7,65625$.



Im Beispiel sei nun das Signifikanzniveau auf 5% festgesetzt. Das bedeutet, dass die Nullhypothese mit einer Wahrscheinlichkeit von höchstens 5% irrtümlich abgelehnt wird.
Laut Tabelle ist dann $g = 3,841$.

Da nun $7,65625 > 3,841$, wird unsere Nullhypothese mit einer Irrtumswahrscheinlichkeit von höchstens 5% abgelehnt. Folglich gibt es durch die Eisspende der Firma bei den Studenten eine signifikante Änderung in der Liebe zum Vanilleeis. Und wegen $29 > 11$ hat die Liebe zu Vanilleeis zu- und nicht abgenommen.

Genauere Betrachtung: Die Dichtefunkton von χ^2 bei nur einem Freiheitsgrad lautet $f_1(x) = \frac{e^{-\frac{x}{2}}}{\sqrt{2 \cdot \pi \cdot x}}$.
Dann gilt für die Wahrscheinlichkeit für eine irrtümliche Ablehnung der Nullhypothese nach meinem PC sogar nur $P(\chi^2 \ge 7,65625) = \int_{7,65625}^{\infty} f_1(x)\,dx = 0,005657597815 \approx 0,57\%$. Dies ist vereinbar mit obiger Tabelle, nach der c zwischen 0,1% und 1% liegen muss.

Exakte Betrachtung: Die Werte von b und c sind jeweils binomialverteilt mit $n = b+c = 40$ und $p = 0,5$. Ihr Mittelwert ist $40/2 = 20$. X sei die Anzahl der Wechsler. Dann gilt für zur Zahl 20 symmetrische Intervalle

$$P(15 \le X \le 25) = \sum_{k=15}^{25} \binom{40}{k} \cdot 0,5^k \cdot 0,5^{40-k} = 0,5^{40} \cdot \sum_{k=15}^{25} \binom{40}{k} = 0,9193 < 95\%,$$

d.h. dass mit der Wahrscheinlichkeit $1 - P(15 \le X \le 25) = 0,0807 > 5\%$ die Nullhypothese abgelehnt würde. Und dieser Wert ist zu groß.

$$P(14 \le X \le 26) = \sum_{k=14}^{26} \binom{40}{k} \cdot 0,5^k \cdot 0,5^{40-k} = 0,5^{40} \cdot \sum_{k=14}^{26} \binom{40}{k} = 0,9615 > 95\%,$$

d.h. dass mit der Wahrscheinlichkeit von nur $1 - P(14 \le X \le 26) = 0,0385 = 3,85\%$ die Nullhypothese abgelehnt würde. Und dieser Wert liegt ja unter 5%.

Da unsere Werte $b = 29$ und $c = 11$ außerhalb des Intervalls $[14;26]$ liegen, wird die Nullhypothese abgelehnt.

Die Exponentialverteilung

Beim radioaktiven Zerfall eines Präparats hat man festgestellt, dass in einem festen Zeitabschnitt immer der gleiche Anteil λ der noch nicht zerfallenen Atome zerfällt. λ heißt auch **Ausfallrate** oder **Zerfallskonstante**.

$f(t)$ bezeichne die Anzahl der zur Zeit t, $t \ge 0$, noch nicht zerfallenen Atome. Im darauffolgenden kleinen Zeitintervall Δt zerfällt dann der Anteil $\lambda \cdot \Delta t$ von $f(t)$, also zerfallen $\lambda \cdot \Delta t \cdot f(t)$ Atome. Daraus ergibt sich die Beziehung $\frac{f(t+\Delta t)-f(t)}{\Delta t} = \frac{[f(t)-\lambda\cdot\Delta t\cdot f(t)]-f(t)}{\Delta t} = \frac{-\lambda\cdot\Delta t\cdot f(t)}{\Delta t} = -\lambda \cdot f(t)$.
Führt man nun den Grenzübergang $\Delta t \to 0$ aus, so folgt die Differenzialgleichung $f'(t) = -\lambda \cdot f(t)$.

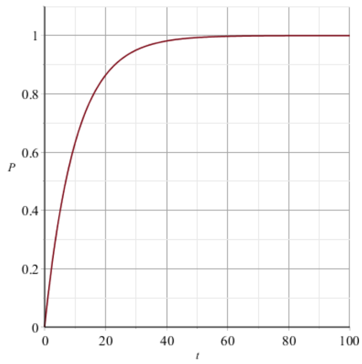
Zur Lösung wird diese Gleichung umgeschrieben in $\frac{df}{dt} = -\lambda \cdot f(t)$ bzw. $\frac{1}{f(t)} df = -\lambda\, dt$. Die Integration ergibt $\ln(|f(t)|) = -\lambda \cdot t + k$ mit einer Konstanten k. Folglich ist $|f(t)| = e^{-\lambda\cdot t+k} = e^k \cdot e^{-\lambda\cdot t}$, d.h. $f(t) = \pm e^k \cdot e^{-\lambda\cdot t}$.
Da $f(t)$ eine positive Zahl sein muss, gilt $f(t) = c\cdot e^{-\lambda\cdot t}$ mit der positiven Konstanten $c = f(0)$.

Es sei nun **X die Lebensdauer eines Atoms**. Dann gibt $P(X \le t)$ die Wahrscheinlichkeit an, dass ein Atom im Zeitraum von 0 bis t zerfällt, also höchstens die Zeit t lebt,

$$P(X \le t) = \frac{\text{Anzahl der in der Zeit t zerfallenen Atome}}{\text{Anzahl der zur Zeit 0 nicht zerfallenen Atome}} = \frac{f(0)-f(t)}{f(0)} = \frac{f(0)-f(0)\cdot e^{-\lambda\cdot t}}{f(0)} = 1-e^{-\lambda\cdot t}.$$

Im Schaubild ist $F(x) = P(X \le x) = 1-e^{-0{,}1x}$ für $\lambda = 0{,}1$ gezeichnet.

Beispiel: Für $\lambda = 0{,}1 s^{-1}$ gilt:



① $P(X \le 0) = 0$, da am Anfang $(X = 0)$ noch kein Atom zerfallen ist.

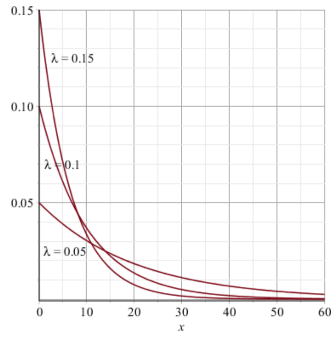
② $P(X \le 1) = 0{,}095$. Die Wahrscheinlichkeit, dass ein Atom höchstens eine Sekunde überlebt, beträgt 0,095. Oder: Innerhalb der ersten Sekunde zerfallen 9,5% der vorhandenen Atome.

③ $P(X \le 20) = 0{,}865$. Die Wahrscheinlichkeit, dass ein Atom höchstens zwanzig Sekunden überlebt, beträgt 0,865. Oder: Innerhalb der ersten zwanzig Sekunden zerfallen 86,5% der vorhandenen Atome.

④ $P(1 < X \le 2) = P(X \le 2) - P(X \le 1) = 0{,}086$. Die Wahrscheinlichkeit, dass ein Atom in der zweiten Sekunde zerfällt, beträgt 8,6%.

Definition: Eine Zufallsgröße X heißt **exponentialverteilt mit dem Parameter $\lambda > 0$**, wenn für ihre Verteilungsfunktion $F(x)$ gilt: $F(x) = \begin{cases} 1-e^{-\lambda\cdot x} & \text{für } x \ge 0 \\ 0 & \text{für } x < 0 \end{cases}$.

Für die Dichtefunktion gilt dann $f(x) = F'(x) = \begin{cases} \lambda\cdot e^{-\lambda\cdot x} & \text{für } x \ge 0 \\ 0 & \text{für } x < 0 \end{cases}$.



Die Schaubilder von $f(x)$ sind für $\lambda = 0{,}05$, $\lambda = 0{,}1$ und $\lambda = 0{,}15$ gezeichnet.

Die Funktionen $f(x)$ und $F(x)$ für $x \ge 0$ sind stetige Funktionen, ebenso wie $\varphi(x)$ und $\Phi(x)$ für $x \in \mathbb{R}$ bei der Normalverteilung.

Es gilt $\int_0^\infty f(x)\,dx = \int_0^\infty \lambda\cdot e^{-\lambda\cdot x}\,dx = -\int_0^{-\infty} e^u\,du = -\left[e^u\right]_0^{-\infty} = -(0-1) = 1$ mit der Substitution $u = -\lambda\cdot x$. Somit handelt es sich wirklich um eine Wahrscheinlichkeitsverteilung.

Für den Erwartungswert gilt $E(X)=\int_0^\infty x\cdot f(x)\,dx=\int_0^\infty x\cdot\lambda\cdot e^{-\lambda\cdot x}dx=\int_0^{-\infty}\frac{u}{\lambda}\cdot e^u du=\frac{1}{\lambda}\int_0^{-\infty}u\cdot e^u du=\frac{1}{\lambda}$. Dieses Integral erhält man durch partielle Integration. $\boxed{E(X)=\frac{1}{\lambda}}$

Wenn z.B. $\lambda=0{,}1s^{-1}$ beträgt, d.h. wenn 10% der vorhandenen Atome pro Sekunde zerfallen, dann beträgt die mittlere Lebensdauer eines Atoms $\frac{1}{\lambda}=\frac{1}{0{,}1s^{-1}}=10\,s$.

Umgekehrt, wenn wir Geräte mit einer mittleren Lebensdauer von 5 Jahren betrachten, dann ist $E(X)=\frac{1}{\lambda}=5$ Jahre, also $\lambda=\frac{1}{5\text{ Jahre}}=0{,}2/\text{Jahr}=20\%\ /\text{Jahr}$.

X bezeichnet die Lebensdauer. $P(X\le x)=1-e^{-\lambda\cdot x}$
$E(X)=\frac{1}{\lambda}$ ist die mittlere Lebensdauer.

Für die Varianz gilt $V(X)=\int_0^\infty\left(x-\frac{1}{\lambda}\right)^2\cdot f(x)\,dx=\int_0^\infty\left(x-\frac{1}{\lambda}\right)^2\cdot\lambda\cdot e^{-\lambda\cdot x}\,dx=\frac{1}{\lambda^2}$. $\boxed{V(X)=\frac{1}{\lambda^2}}$

Es sei t_0 ein beliebiger Zeitpunkt und $t>t_0$ ebenfalls beliebig. $P(X>t)$ ist die Wahrscheinlichkeit, dass das Atom den Zeitpunkt t unzerfallen überlebt. Hängt diese Wahrscheinlichkeit davon ab, dass es den früheren Zeitpunkt t_0 unzerfallen überlebt hat?

Beispielsweise bei Elektrogeräten trifft das nicht zu. Wenn sie ein gewisses Alter t_0 überschritten haben, dann steigt die Wahrscheinlichkeit, in nächster Zeit zu versagen. Nicht so beim Zerfall von Atomen. Ihre Wahrscheinlichkeit, in der nächsten Sekunde zu zerfallen, hängt nach dem folgenden Satz nicht von ihrem Alter ab.

Satz: Die Exponentialverteilung besitzt kein Gedächtnis, sie ist gedächtnislos.

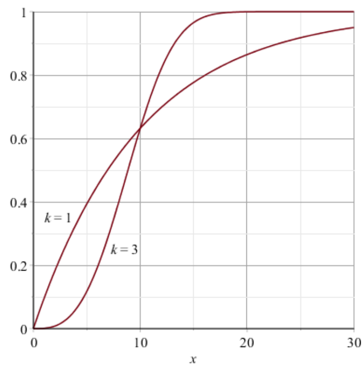
Beweis: Es sei $\Delta t>0$ ein beliebiges Zeitintervall. Dann ist zu zeigen, dass für die bedingte Wahrscheinlichkeit $P(X>t_0+\Delta t\mid X>t_0)=P(X>\Delta t)$ gilt, also unabhängig von t_0 ist.

$$P(X>t_0+\Delta t\mid X>t_0)=\frac{P(t_0+\Delta t\ \wedge\ X>t_0)}{P(X>t_0)}=\frac{P(X>t_0+\Delta t)}{P(X>t_0)}=\frac{1-P(X\le t_0+\Delta t)}{1-P(X\le t_0)}=\frac{1-(1-e^{-\lambda\cdot(t_0+\Delta t)})}{1-(1-e^{-\lambda\cdot t_0})}=$$

$$\frac{e^{-\lambda\cdot(t_0+\Delta t)}}{e^{-\lambda\cdot t_0}}=e^{-\lambda\cdot\Delta t}=1-(1-e^{-\lambda\cdot\Delta t})=1-P(X\le\Delta t)=P(X>\Delta t).$$

Zusatz: Wenn man berücksichtigen möchte, dass ein Gerät am Anfang nicht so häufig versagt, kann die **Weibull-Verteilung** verwendet werden: $F(x)=P(X\le x)=1-e^{-(\lambda\cdot x)^k}$ mit $x\ge 0$ und $\lambda,\ k>0$. k heißt Formparameter.

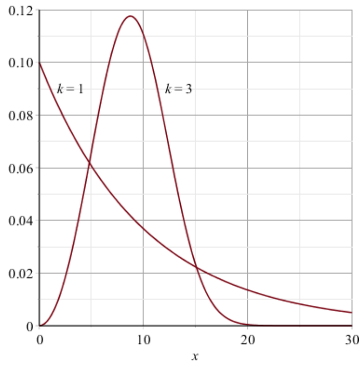
Für $k=1$ ergibt sich die Exponentialverteilung.
Für die beiden Schaubilder von $F(x)$ ist $\lambda=0{,}1$ gewählt. k ist 1 bzw. 3. Man sieht z.B., dass bei Weibull (k = 3) die Wahrscheinlichkeit für eine Lebensdauer von nur 5 Jahren $F(X=5)=0{,}118$ beträgt, während die Exponentialverteilung dafür den Wert $F(X=5)=0{,}393$ liefert, so dass das Gerät innerhalb der ersten 5 Jahre mit einer größeren Wahrscheinlichkeit versagt.



Durch logarithmische Differenziation erhält man die Dichtefunktion

$f(x) = F'(x) = \lambda \cdot k \cdot (\lambda \cdot x)^{k-1} \cdot e^{-(\lambda \cdot x)^k}$

Für die beiden Schaubilder von $f(x)$ ist $\lambda = 0{,}1$ gewählt.
k ist 1 (d.h. exponentialverteilt) bzw. 3.



Der Erwartungswert ist $E(X) = \frac{1}{\lambda} \cdot \Gamma(1+1/k)$ und die Varianz

$V(X) = \frac{1}{\lambda^2} \cdot \left(\Gamma(1+2/k) - \Gamma^2(1+1/k)\right)$. Dabei ist $\Gamma(x) = \int\limits_0^\infty t^{x-1} \cdot e^{-t}\, dt$

die Eulersche Gammafunktion. Für sie gilt $\Gamma(1) = 1$ und $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$.

Die Maximum-Likelihood-Methode (ML-Methode)

Dabei handelt es sich um die **Methode der größten Plausibilität.** Sie dient der Schätzung von Parametern. Das allgemeine Verfahren stammt vom englischen Statistiker Ronald Aylmer Fisher (1890 − 1962). Gauß hatte diese Methode in Spezialfällen auch schon verwendet.

Beispiel 1: Gegeben ist eine Urne mit sehr vielen gleichartigen Kugeln, die entweder weiß oder schwarz sind. Man möchte erfahren, wieviel Prozent der Kugeln weiß sind. p bezeichne die unbekannte Wahrscheinlichkeit für weiß.

Um p zu schätzen, führen wir n voneinander unabhängige Messungen durch. Bei der Messung Nummer i, $i = 1, \ldots, n$, werden n_i Kugeln gezogen, von denen x_i Kugeln weiß und $n_i - x_i$ Kugeln schwarz sind. Die Wahrscheinlichkeit dafür beträgt $P(X_i = x_i) = \binom{n_i}{x_i} \cdot p^{x_i} \cdot (1-p)^{n_i - x_i}$, ohne Berücksichtigung der Reihenfolge.

Die L-Funktion besteht aus dem **Produkt der einzelnen Wahrscheinlichkeiten** $P(X_i = x_i)$. Und wir bestimmen nun p so, dass diese L-Funktion ein Maximum annimmt.

$L(p) = \prod\limits_{i=1}^{n} \binom{n_i}{x_i} \cdot p^{x_i} \cdot (1-p)^{n_i - x_i} = \left(\prod\limits_{i=1}^{n} \binom{n_i}{x_i}\right) \cdot p^x \cdot (1-p)^{N-x}$. Dabei ist $N = \sum\limits_{i=1}^{n} n_i$ die Anzahl aller gezogenen Kugeln und $x = \sum\limits_{i=1}^{n} x_i$ die Anzahl aller gezogenen weißen Kugeln.

Um das Maximum von $L(p)$ zu finden, leiten wir nach p ab.

$$L'(p) = \prod_{i=1}^{n} \binom{n_i}{x_i} \cdot \left[x \cdot p^{x-1} \cdot (1-p)^{N-x} - p^x \cdot (N-x) \cdot (1-p)^{N-x-1}\right] = \prod_{i=1}^{n} \binom{n_i}{x_i} \cdot \left[p^{x-1} \cdot (1-p)^{N-x-1} \cdot \left(x \cdot (1-p) - p \cdot (N-x)\right)\right] =$$

$= \prod\limits_{i=1}^{n} \binom{n_i}{x_i} \cdot \left[p^{x-1} \cdot (1-p)^{N-x-1} \cdot (x - p \cdot N)\right] = 0$. **Am plausibelsten erscheint** $x - p \cdot N = 0$, also $p = \frac{x}{N} = \frac{\sum\limits_{i=1}^{n} x_i}{\sum\limits_{i=1}^{n} n_i}$.

Und das ist vernünftig! Nicht sinnvoll wäre der Mittelwert $\frac{1}{n} \cdot \sum\limits_{i=1}^{n} \frac{x_i}{n_i}$ aller Anteile $\frac{x_i}{n_i}$, da diese dann gleichwertig behandelt würden. Aber $\frac{x_i}{n_i}$ liegt umso näher bei p liegen, je größer n_i ist.

Beispiel 2: Es sei bekannt, dass die Zufallsvariable X Poisson-verteilt ist, d.h. es gilt $P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ mit dem Parameter $\lambda > 0$. λ ist aber unbekannt und soll mit Hilfe mehrerer Messungen geschätzt werden.

Zum Beispiel bestimmen wir mehrfach ($n \in \mathbb{N}$) die Anzahl X_i der Radler, die innerhalb einer Stunde an einer bestimmten Stelle vorbeifahren. Die einzelnen Zufallsvariablen X_i, $i = 1...n$ sind unabhängig voneinander und jeweils Poisson-verteilt; man sagt auch, sie sind **iid** – verteilt: *independent and identically distributed* .

Nun führen wir n voneinander unabhängige Messungen mit den Zufallsvariablen X_i durch und erhalten eine Stichprobe von n Messwerten x_i, $i = 1...n$. Wir bilden damit die L-Funktion als Produkt der einzelnen Wahrscheinlichkeiten: $L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$. Der Parameter λ ist nun so zu bestimmen, dass $L(\lambda)$ maximal wird.

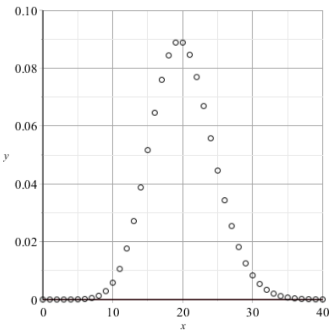
Die L-Funktion lautet

$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda} = \frac{\lambda^{x} \cdot e^{-n\cdot\lambda}}{\prod_{i=1}^{n} x_i!}$ mit der Abkürzung $x = \sum_{i=1}^{n} x_i$. Durch Ableiten nach λ folgt

$L'(\lambda) = \frac{x \cdot \lambda^{x-1} \cdot e^{-n\cdot\lambda} - n \cdot \lambda^{x} \cdot e^{-n\cdot\lambda}}{\prod_{i=1}^{n} x_i!} = \frac{\lambda^{x-1} \cdot e^{-n\cdot\lambda} \cdot (x - n\cdot\lambda)}{\prod_{i=1}^{n} x_i!} = 0$, also $\lambda = \frac{x}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

Dieses Ergebnis ist vernünftig, denn einerseits ist bei der Poisson-Verteilung gerade dieses λ der Erwartungswert, andererseits wurden bei den n Beobachtungen insgesamt $\sum_{i=1}^{n} x_i$ Radler gezählt, d.h. im Mittel wurden pro Beobachtung gerade $\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ gezählt.

Zusatz 1: Da die natürliche Logarithmusfunktion streng monoton steigend ist, kann statt des Maximums von $L(\lambda)$ das Maximum von $\ln\left(L(\lambda)\right)$ bestimmt werden. Beide Maxima treten deshalb an der gleichen Stelle auf.



$\ln\left(L(\lambda)\right) = \ln\left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}\right) = \sum_{i=1}^{n} \ln\left(\frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}\right) = \sum_{i=1}^{n} \left(x_i \cdot \ln(\lambda) - \lambda - \ln(x_i!)\right)$,

also $L'(\lambda) = \sum_{i=1}^{n} \left(x_i \cdot \frac{1}{\lambda} - 1\right) = \frac{1}{\lambda} \cdot \left(\sum_{i=1}^{n} x_i\right) - n = 0$. Und daraus folgt ebenfalls

$\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

Z.B. wurden 10 Messungen 25, 11, 18, 30, 16, 21, 15, 31, 20, 13 durchgeführt: Folglich $\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{10} \cdot 200 = 20$, also $P(X = k) = \frac{20^k}{k!} \cdot e^{-20}$. Siehe Schaubild.

Zusatz 2:

a. $\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ ist eine erwartungstreue Schätzfunktion für λ, denn

$E\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right) = \frac{1}{n} \cdot E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X) = \frac{1}{n} \cdot n \cdot E(X) = E(X)$. $E(x_i)$ ist dabei die Anzahl der pro Stunde vorbeifahrenden Radler bei Messung Nummer i.

b. $\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ ist eine konsistente Schätzfunktion für λ, d.h. $\lim_{x\to\infty} \mathrm{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right) = 0$, so dass die Mittelwerte mit wachsendem n immer weniger streuen. Wegen der Unabhängigkeit der einzelnen Messungen gilt

$\mathrm{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} \cdot \mathrm{Var}\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \mathrm{Var}(x_i) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n} \to 0$ für $n \to \infty$.

Beispiel 3: Es sei bekannt, dass die Zufallsvariable X normalverteilt ist mit $f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Jetzt sollen die beiden Parameter μ und σ geschätzt werden. Wir führen n voneinander unabhängige Messungen mit den Zufallsvariablen X_i durch und erhalten n Messwerte x_i, $i=1...n$. Die L-Funktion lautet

$$L(\mu,\sigma)=\prod_{i=1}^{n}\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}=\frac{1}{\sqrt{2\pi}^n}\sigma^{-n}\cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

. Und diese Funktion soll partiell nach μ und nach σ abgeleitet werden. Oh je! Wir greifen auf den natürlichen Logarithmus zurück:

$$\ln\left(L(\mu,\sigma)\right)=\sum_{i=1}^{n}\ln\left(\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\right)=\sum_{i=1}^{n}\left(-\ln(\sigma)-\ln\left(\sqrt{2\pi}\right)-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right).$$

Wir leiten nun partiell nach μ und nach σ ab.

$$\frac{\partial}{\partial\mu}L(\mu,\sigma)=\sum_{i=1}^{n}\left(-\frac{1}{2}\cdot\frac{2\cdot(x_i-\mu)\cdot(-1)}{\sigma^2}\right)=\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2}=\frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\mu)=0.$$ Also

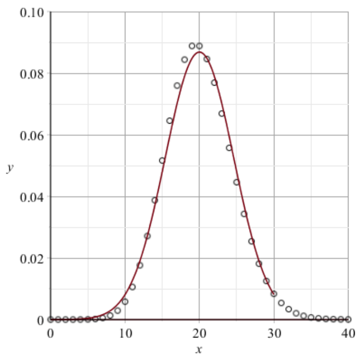
$\sum_{i=1}^{n}(x_i-\mu)=\left(\sum_{i=1}^{n}x_i\right)-n\cdot\mu=0$, so dass wir μ schätzen durch $\frac{1}{n}\cdot\sum_{i=1}^{n}x_i$, was sehr vernünftig ist.

$$\frac{\partial}{\partial\sigma}L(\mu,\sigma)=\sum_{i=1}^{n}\left(-\frac{1}{\sigma}+\frac{(x_i-\mu)^2}{\sigma^3}\right)=\frac{1}{\sigma^3}\cdot\sum_{i=1}^{n}\left(-\sigma^2+(x_i-\mu)^2\right)=\frac{1}{\sigma^3}\cdot\left(-n\cdot\sigma^2+\sum_{i=1}^{n}(x_i-\mu)^2\right)=0.$$

Und daraus ergibt sich ein Schätzwert von σ^2 zu $\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2$, was hocherfreulich ist.

Mit den Zahlenwerten 25, 11, 18, 30, 16, 21, 15, 31, 20, 13 von Beispiel 2 folgt $\mu=20$ und $\sigma=\sqrt{21{,}1}$.

Im Schaubild sind die Punkte aus Beispiel 2 und die Glockenkurve nach Gauß aus diesem Beispiel gezeichnet. Eine gute Übereinstimmung!



Die Momenten-Methode (MoM-Methode)

Sie dient, ebenso wie die Maximum-Likelihood-Methode, der Schätzung von Parametern. Sie stammt von Karl Pearson (1857-1936).

Definition: Das **k-te Moment** m_k , $k\in\mathbb{N}$, einer Zufallsvariablen X ist der Erwartungswert der k-ten Potenz von X, also $m_k=E(X^k)$.

Speziell: Das erste Moment ist $m_1=E(X)=\mu$ der Erwartungswert.

Speziell: Das zweite Moment ist $m_2=E(X^2)$.

Das zweite Moment kommt in der Varianz vor: $\sigma^2=V(X)=E(X^2)-E(X)^2=m_2-m_1^2$.

Definition: Das **k-te empirische Moment** $\hat{m}_k$, $k \in \mathbb{N}$, der Stichprobe $x_1, x_2, ..., x_n$ ist $\hat{m}_k = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k$. Das „Dach“ bedeutet Schätzwert.

Speziell: Das erste empirische Moment $\hat{m}_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_1 = \overline{x}$ ist das arithmetische Mittel der Stichprobe.

Speziell: Das zweite empirische Moment ist $\hat{m}_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 = \overline{x^2}$.

Wir hatten früher hergeleitet, dass $\frac{1}{n}\sum_{i=1}^{n}(x - x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \overline{x^2} - \overline{x}^2$ gilt.

Idee der Momenten-Methode: Die k-ten Momente werden durch die k-ten empirischen Momente ersetzt.

Speziell: Der Erwartungswert $m_1 = E(X) = \mu$ wird geschätzt durch $\hat{\mu} = \hat{m}_1 = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Speziell: Die Varianz $V(X) = \sigma^2$ wird geschätzt durch $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x - x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \overline{x^2} - \overline{x}^2$.

Im Rahmen der induktiven Statistik werden wir erfahren, dass streng genommen $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x - x_i)^2$ gilt.

Ein Unterschied ist aber nur für kleine Werte von n relevant.

Beispiel 1: Gegeben ist eine Urne mit sehr vielen gleichartigen Kugeln, die entweder weiß oder schwarz sind. Man möchte erfahren, wieviel Prozent der Kugeln weiß sind. Es sei p die Wahrscheinlichkeit für eine weiße Kugel. Die Zufallsvariable X sei definiert durch $X = \begin{cases} 1 & \text{die gezogene Kugel ist weiß} \\ 0 & \text{die gezogene Kugel ist schwarz} \end{cases}$.

Der Erwartungswert von X ist $\mu = E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$.

Die Wahrscheinlichkeit wird geschätzt durch $\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} n_i}$, so dass wir $p \approx \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} n_i}$ erhalten, genau wie nach der Maximum-Likelihood-Methode.

Beispiel 2: Es sei bekannt, dass die Zufallsvariable X Poisson-verteilt ist, d.h. es gilt $P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ mit dem Parameter $\lambda > 0$. λ ist aber unbekannt und soll mit Hilfe mehrerer Messungen geschätzt werden.
Zum Beispiel bestimmen wir mehrfach ($n \in \mathbb{N}$) die Anzahl X_i der Radler, die innerhalb einer Stunde an einer bestimmten Stelle vorbeifahren. Die einzelnen Zufallsvariablen X_i , $i = 1...n$ sind unabhängig voneinander und jeweils Poisson-verteilt.
Wir wissen, dass der Erwartungswert bei der Poisson-Verteilung gleich λ ist. Diesen Erwartungswert schätzen wir durch das arithmetische Mittel $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, so dass wir wie oben $\lambda \approx \frac{1}{n}\sum_{i=1}^{n} x_i$ erhalten.

Beispiel 3: Es sei bekannt, dass die Zufallsvariable X normalverteilt ist mit $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Jetzt sollen die beiden Parameter μ und σ geschätzt werden. Wir führen n voneinander unabhängige Messungen mit den Zufallsvariablen X_i durch und erhalten n Messwerte x_i , $i = 1...n$.

Wir schätzen den Erwartungswert durch das arithmetische Mittel: $\hat{\mu} = \hat{m}_1 = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ und die Varianz σ^2

durch $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x - x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \overline{x^2} - \bar{x}^2$, d.h. zweites empirisches Moment minus Quadrat des ersten empirischen Moments.

Exakter wäre $\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2$, wie wir später zeigen werden.

Beispiel 4: Ein Zufallsgenerator liefert uns Zufallszahlen aus einem Intervall $[a, b]$, dessen Grenzen a und b nicht bekannt sind. Durch eine Stichprobe $x_1, x_2, ..., x_n$ wollen wir diese beiden Grenzen abschätzen.

Die Verteilungsfunkton lautet $f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \le x \le b \\ 0 & \text{für } x < a \text{ oder } x > b \end{cases}$.

Für den Erwartungswert erhalten wir $\mu = E(X) = \int_a^b x \cdot f(x)\,dx = \frac{1}{b-a}\cdot\int_a^b x\,dx = \frac{1}{b-a}\cdot\frac{1}{2}\left(b^2 - a^2\right) = \frac{1}{2}(b+a)$.

Also liegt das arithmetische Mittel aller Zufallszahlen etwa in der Intervallmitte $\frac{1}{2}(b+a)$.

Für die Varianz gilt

$$\sigma^2 = \int_a^b (x-\mu)^2 \cdot f(x)\,dx = \frac{1}{b-a}\int_a^b (x-\mu)^2\,dx = \frac{1}{b-a}\cdot\frac{1}{3}\cdot\left((b-\mu)^2 - (a-\mu)^2\right) = \frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right) =$$

$$= \frac{1}{b-a}\cdot\frac{1}{3}\cdot\left(\left(\frac{b-a}{2}\right)^3 + \left(\frac{b-a}{2}\right)^3\right) = \frac{1}{b-a}\cdot\frac{1}{3}\cdot 2\cdot\frac{(b-a)^3}{8} = \frac{1}{12}\cdot(b-a)^2.$$ Wegen $a < b$ folgt $b - a = \sqrt{12\sigma^2}$.

Aus den beiden Gleichungen $a + b = 2\mu$ und $b - a = \sqrt{12\sigma^2}$ folgt $a = \mu - \sqrt{3\sigma^2}$ und $b = \mu + \sqrt{3\sigma^2}$.

Zahlenbeispiel: Der Generator erzeugt die fünf Zahlen 2, 4, 6, 8, 10. Wir schätzen $\mu \approx \frac{1}{5}\sum_{i=1}^{n} x_i = 6$ und

$\sigma^2 \simeq \frac{1}{(5-1)}\sum_{i=1}^{5}(x_1 - 6)^2 = \frac{25}{3}$. Also $a = \mu - \sqrt{3\sigma^2} \approx 6 - \sqrt{25} = 1$ und $b = \mu + \sqrt{3\sigma^2} \approx 6 + \sqrt{25} = 11$.