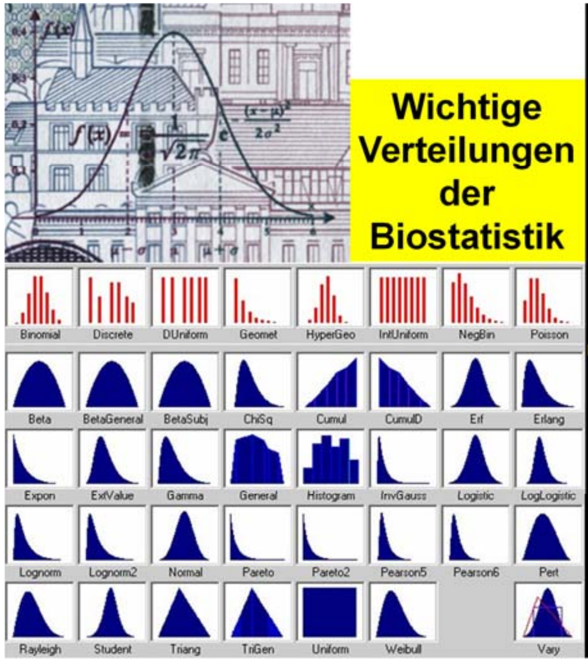
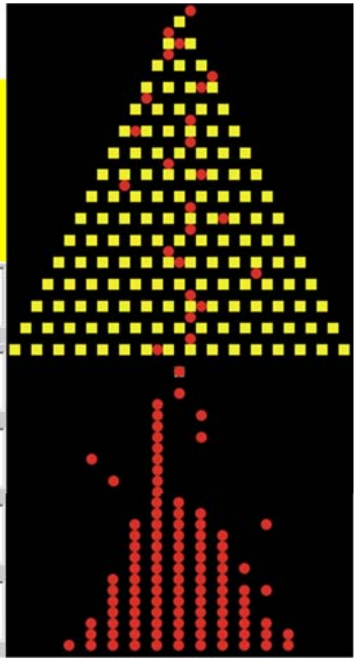
Wichtige Verteilungen der Biostatistik





KAD 2015.10.01

Bernoulli-Experiment

das Zufallsexperiment hat zwei mögliche Ausfälle:
E (Erfolg), $\bar{E}$ (Misserfolg)

das Experiment wird ***n* mal wiederholt** (n Stufen)

die Wahrscheinlichkeit **in einem Einzelversuch**
für E ist: $p(E) = p$, für $p(\bar{E}) = 1\text{-}p = q$

- Die Wahrscheinlichkeit für das Eintreten des betrachteten Ereignisses E ist in jedem Einzelversuch **gleich**.
- In jedem Einzelversuch ist das Ergebnis **unabhängig** von den Ergebnissen aller anderen Versuche.

Wir suchen die Wahrscheinlichkeit, dass wir bei einem n-mal wiederholten Experiment *(n-stufiges Bernoulli Experiment)* genau k-mal Erfolg haben.

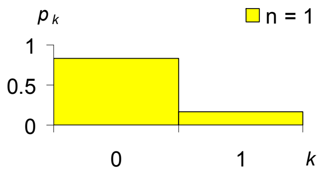
http://ccl.northwestern.edu/netlogo/models/GaltonBox

Würfelversuch als Bernoulli-Experiment

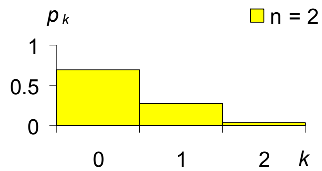
E (Erfolg) entspricht der Augenzahl 6 (AZ6), $p(E) = \frac{1}{6} = 0.166$

$\bar{E}$ (Misserfolg) entspricht den Augenzahlen nicht 6, $p(\bar{E}) = \frac{5}{6} = 0.833$

für $n = 1$
AZ6 0-mal: $p_0 = 1\text{-}0.166 = p(\bar{E})$
AZ6 1-mal: $p_1 = 0.166 = p(E)$
($\Sigma p_k = 0.833+0.166=1$)



für $n = 2$
AZ6 0-mal: $p_0=0.833\cdot 0.833$
AZ6 1-mal: $p_1=0.833\cdot 0.166+0.166\cdot 0.833$
AZ6 2-mal: $p_2=0.166\cdot 0.166$
($\Sigma p_k = 1$)



für $n = 3$

AZ6 0-mal: $p_0 = 1\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \frac{125}{216}$ $\bar{E}\,\bar{E}\,\bar{E}$

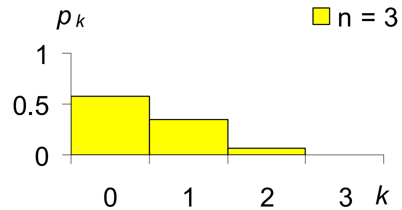
AZ6 1-mal: $p_1 = 3\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{75}{216}$ $E\,\bar{E}\,\bar{E}$ $\bar{E}\,E\,\bar{E}$ $\bar{E}\,\bar{E}\,E$

AZ6 2-mal: $p_2 = 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = \frac{15}{216}$ $\bar{E}\,E\,E$ $E\,\bar{E}\,E$ $E\,E\,\bar{E}$

AZ6 3-mal: $p_3 = 1\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = \frac{1}{216}$ $E\,E\,E$

($\Sigma p_k = 1$)

$$\boxed{p_k = \binom{n}{k} p^k (1-p)^{n-k}}$$



n Experimente, k Erfolge, $\binom{n}{k}$: Anzahl der möglichen Reihenfolgen für $n = 7$

$$p_k = \binom{7}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{7-k}$$

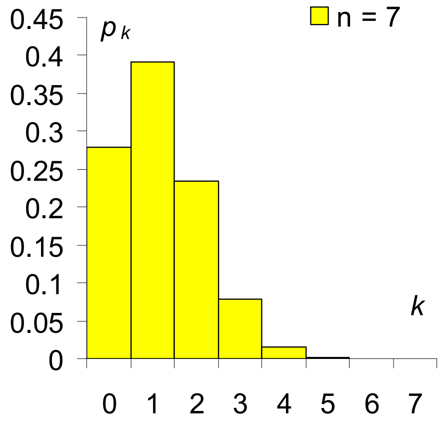
$$p_0 = \binom{7}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^7 = 0.279$$

$$p_1 = \binom{7}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^6 = 0.391$$

$$p_2 = \binom{7}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^5 = 0.234$$

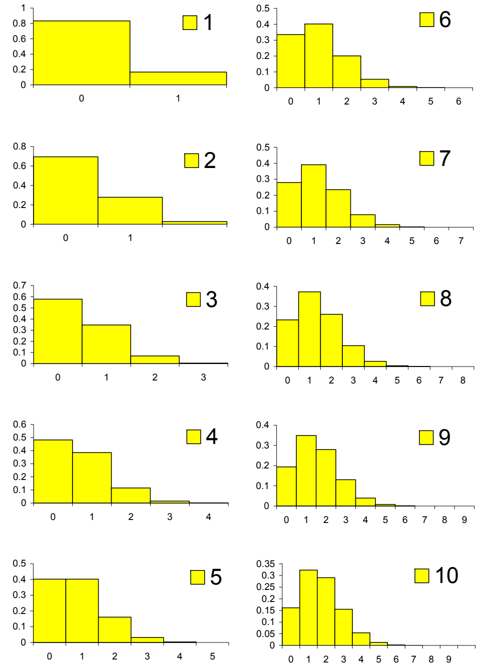
$$p_3 = \binom{7}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^5 = 0.078$$

...



$$p(E) = \frac{1}{6}$$

Binomialverteilung



Bernoulli-Experiment ➡ Binomialverteilung

p_k ist die Wahrscheinlichkeit, dass wir bei einem n-stufigen Bernoulli-Experiment genau k-mal Erfolg haben

die Wahrscheinlichkeit dass die ersten k Versuche Erfolg haben und die anderen misserfolglich sind:

$$p(\underbrace{E,E,\ldots E}_{k},\underbrace{\bar{E},\bar{E},\ldots\bar{E}}_{n-k}) = \underbrace{p(E)\cdot p(E)\cdot\ldots p(E)}_{k}\cdot\underbrace{p(\bar{E})\cdot p(\bar{E})\cdot\ldots p(\bar{E})}_{n-k} = p^k(1-p)^{n-k}$$

Jede andere Reihenfolge der k Ergebnisse und n-k Gegenergebnisse haben die selbe $p^k(1-p)^{n-k}$ Wahrscheinlichkeit.
Wieviele mögliche Reihenfolgen gibt es?

Kombinatorik:
Sollen k Objekte in beliebiger Reihenfolge aus n Objekten ausgewählt werden, ergeben sich:

$$\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{1\cdot 2\cdot\ldots\cdot k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Möglichkeiten, wobei $n! = 1\cdot 2\cdot\ldots\cdot(n-1)\cdot n$.
(Kombination aus n Elementen zur Klasse k ohne Wiederholung)

zB: $n = 3$, $k = 2$ $\quad \binom{3}{2} = \frac{3!}{2!\cdot 1!} = \frac{1\cdot 2\cdot 3}{1\cdot 2\;\cdot\;1} = 3$

E E Ē   E Ē E   Ē E E

Fakultät von n	1*2*3*...*n
0! =	1
1! =	1
2! =	2
3! =	6
4! =	24
5! =	120
6! =	720
7! =	5,040
8! =	40,320
9! =	362,880
10! =	3,628,800
11! =	39,916,800
12! =	479,001,600
13! =	6,227,020,800
14! =	87,178,291,200

Münzenversuch als Bernoulli-Experiment

zwei Münzenversuche: E (Erfolg) entspricht Zahl 

p=1/2, 1- p =1/2, n=2

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

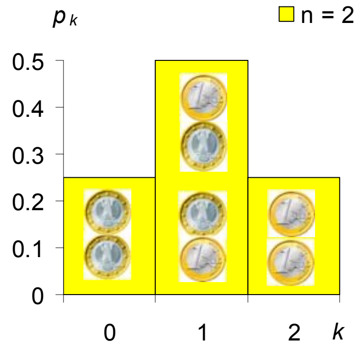
für p=1/2 vereinfacht sich: $p_k = \binom{n}{k}\left(\frac{1}{2}\right)^n$

$$p_0 = 1\left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$p_1 = 2\left(\frac{1}{2}\right)^2 = \frac{2}{4} = \frac{1}{2}$$

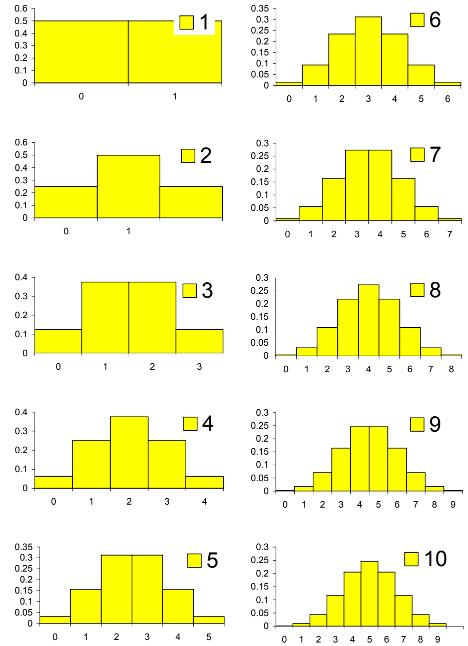
$$p_2 = 1\left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$p_0+p_1+p_2=1$



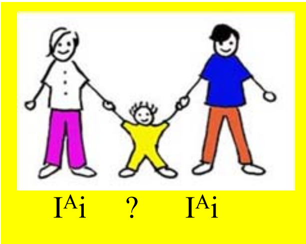
$$p(E) = \frac{1}{2}$$

Binomialverteilung



Blutgruppenversuch als Bernoulli-Experiment

Genotyp	$I^A I^A$	$I^A i$	$i I^A$	ii
Wahrscheinlichkeit	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Fenotyp	A			0
Wahrscheinlichkeit	$\frac{3}{4}$			$\frac{1}{4}$



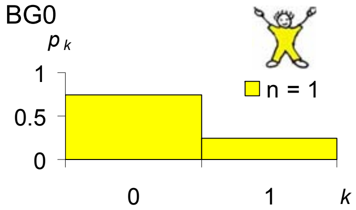
Nennen wir als „Erfolg" (E), wenn das Kind eine Blutgruppe 0 hat und „Misserfolg" ($\bar{E}$), wenn seine Blutgruppe nicht 0 ist.

$p(E)$=1/4=0.25 $p(\bar{E})$=3/4=0.75 $p(E)+p(\bar{E})$=1

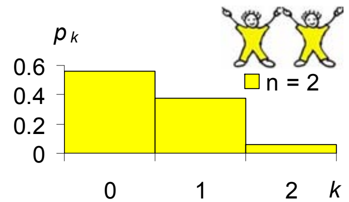
Die Familie hat ***n*** Kinder. Wie hoch ist die Wahrscheinlichkeit, dass ***k*** Kinder eine Blutgruppe 0 haben?

n: Anzahl der Kinder, k: Anzahl der Kinder mit BG0

für n = 1
(BG0 0-mal:) p_0=0,75=$p(\bar{E})$
(BG0 1-mal:) p_1=0,25=$p(E)$
(Σp_k = 0,75+0,25=1)



für n = 2
(BG0 0-mal:) p_0=0,75·0,75 =0,5625
(BG0 1-mal:) p_1=0,75·0,25+0,25·0,75=0,375
(BG0 2-mal:) p_2=0,25·0,25 = 0,0625
(Σp_k = 0,5625+0,375+0,0625=1)



allgemein: Sei: $p(E)=p$ $p(\bar{E})=1-p=q$

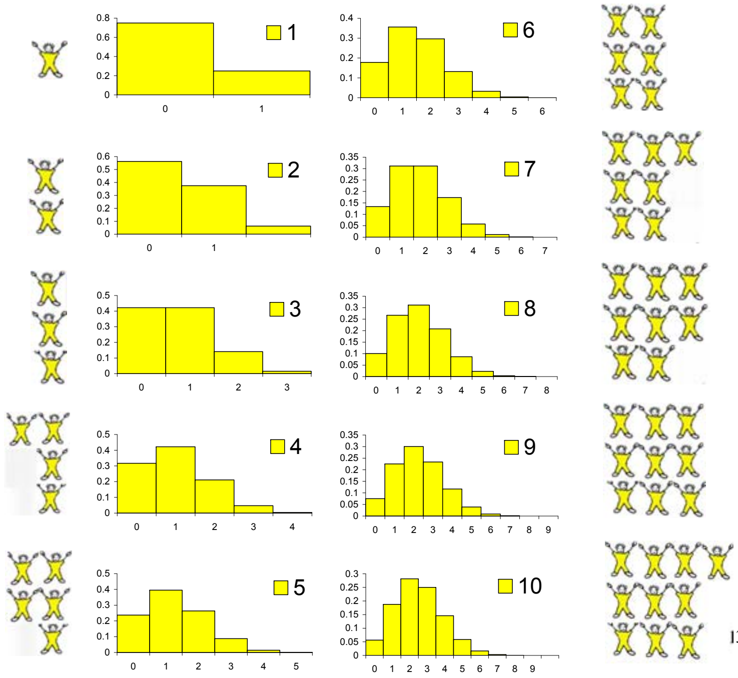
$$\boxed{p_k = \binom{n}{k} p^k (1-p)^{n-k}}$$

n Experimente, k Erfolge

$\binom{n}{k}$: Anzahl der möglichen Reihenfolgen

Binomialverteilung

$$p(E) = \frac{1}{4}$$



Bernoulli-Experiment, Binomialverteilung

zB: 60% der Patienten haben Grippe.
Heute kommen 4 Patienten.
p=0,6 n=4

$$p_k = \binom{4}{k}(0.6)^k(0.4)^{4-k}$$

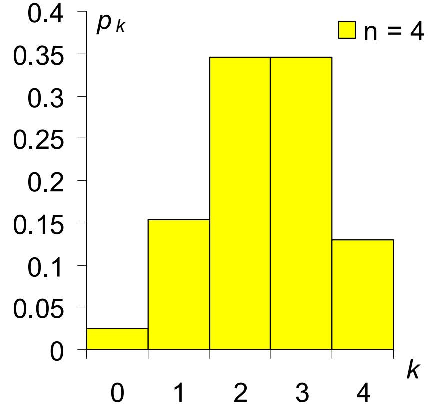
$$p_0 = \binom{4}{0}(0.6)^0(0.4)^4 = 0.0256$$

$$p_1 = \binom{4}{1}(0.6)^1(0.4)^3 = 0.1536$$

$$p_2 = \binom{4}{2}(0.6)^2(0.4)^2 = 0.3456$$

$$p_3 = \binom{4}{3}(0.6)^3(0.4)^1 = 0.3456$$

$$p_4 = \binom{4}{4}(0.6)^4(0.4)^0 = 0.1296$$

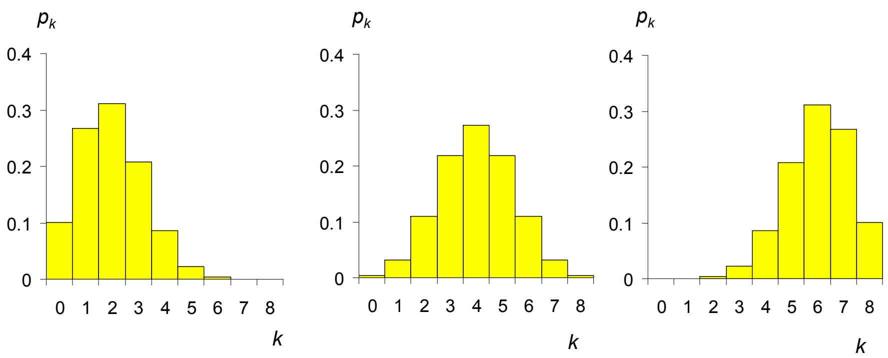


Binomialverteilung

p = 0.25 < 0.5 rechtsschief

p = 0.5 symmetrisch

p = 0.75 > 0.5 linksschief



Die Funktion wird generell für wachsendes n immer „symmetrischer”.

Erwartungswert und Streuung der Binomialverteilung

Erwartungswert: $\boxed{\mu = np}$

Tritt ein bestimmtes Ergebnis mit Wahrscheinlichkeit p ein, so haben wir bei n-maliger Wiederholung etwa np solche Ereignisse zu „erwarten”.

$$\text{Erwartungswert}(x_i) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\text{Erwartungswert}(X) = \sum_i E(x_i) = np$$

(identische Einzelprozesse)

theoretische Streuung: $\boxed{\sigma = \sqrt{np(1-p)}}$

$$\text{Varianz}(x_i) = E(x_i^2) - (E(x_i))^2 = 1^2 p + 0^2(1-p) - p^2 = p - p^2 = p(1-p)$$

$$\text{Varianz}(X) = \sum_i \text{Varianz}(x_i) = np(1-p)$$

zB: 12 Würfelexperimente:
Wie oft tritt „6” auf?
n = 12, p = 1/6, μ = 2

$$\sigma = \frac{\sqrt{60}}{6} \approx 1.29$$

Poisson: Verteilung der seltenen Ereignisse

$$p_k = \binom{n}{k} p^k q^{n-k} \xrightarrow[\left.\begin{matrix} n \to \infty \\ p \to 0 \end{matrix}\right\} np=\lambda]{} \boxed{p_k = \frac{\lambda^k}{k!} e^{-\lambda}}$$

Es ist eine gute Näherung für große n und kleine p Werte

Schätzung der seltener Ereignisse

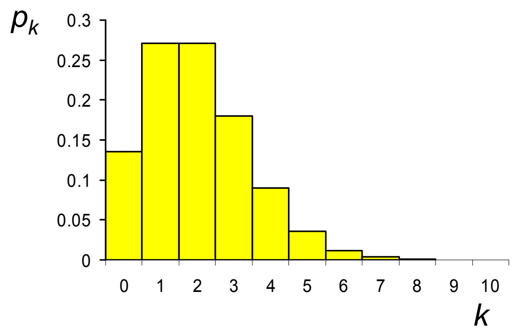
Beispiel:

Die Wahrscheinlichkeit, dass man eine Krankheit bekommt, ist p=0.001. Wie groß ist die Wahrscheinlichkeit, dass es in einer Stadt mit n = 2000 Einwohner k = 0,1,2… Kranken gibt?

Poisson Verteilung: Beispiel

Die Wahrscheinlichkeit, dass man eine Krankheit bekommt, ist p=0.001. Wie groß ist die Wahrscheinlichkeit, dass es in einer Stadt mit n = 2000 Einwohner k = 0,1,2… Kranken gibt?

$\lambda = np = 2$ $\qquad p_k = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{2^k}{k!} e^{-2}$



$$p_0 = \frac{2^0}{0!} e^{-2} = e^{-2} = 0{,}135$$

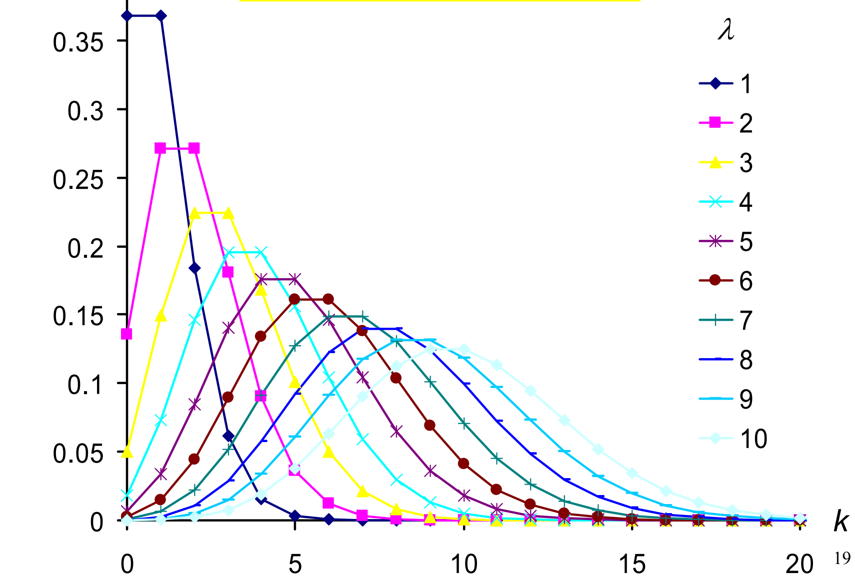
$$p_1 = \frac{2^1}{1!} e^{-2} = 2e^{-2} = 0{,}271$$

$$p_2 = \frac{2^2}{2!} e^{-2} = 2e^{-2} = 0{,}271$$

$$p_3 = \frac{2^3}{3!} e^{-2} = \frac{8}{6} e^{-2} = 0{,}181$$

…

Poisson Verteilung



Erwartungswert und Streuung der Poisson-Verteilung

Erwartungswert: $\boxed{\mu = \lambda}$

Varianz: Varianz(x)=λ

Theoretische Streuung: $\boxed{\sigma = \sqrt{\lambda}}$

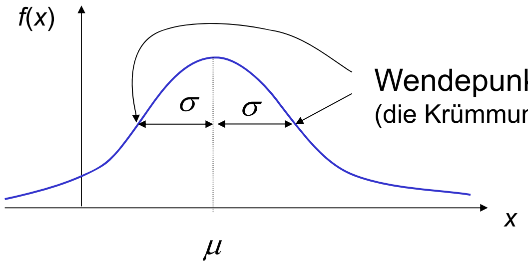
Eine Verteilung, wo der Erwartungswert und die Streuung voneinander nicht unabhängig sind!

Die ausgezeichnete kontinuierliche Verteilung: Normalverteilung

Verteilungsdichtefunktion: $$\boxed{f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Parameter der Normalverteilung: Erwartungswert: μ

Streuung: σ

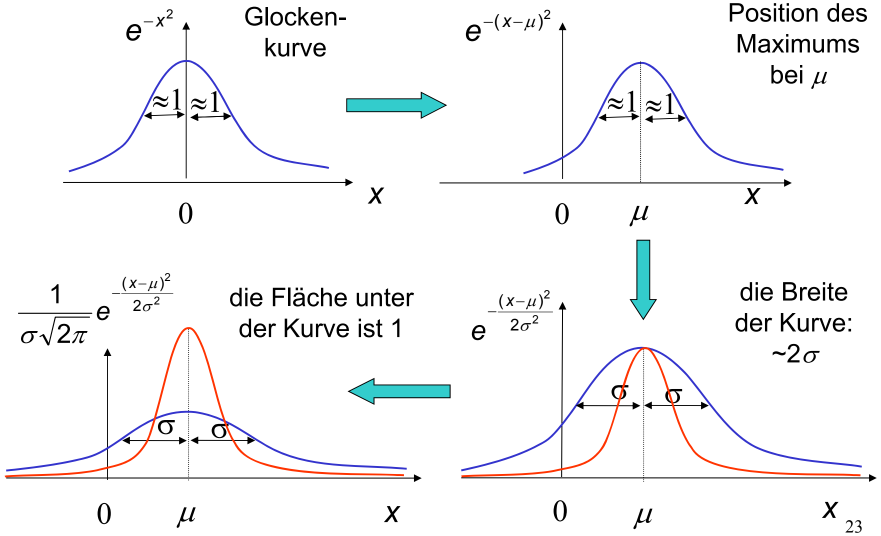


Wendepunkte
(die Krümmung der Kurve ändert sich)

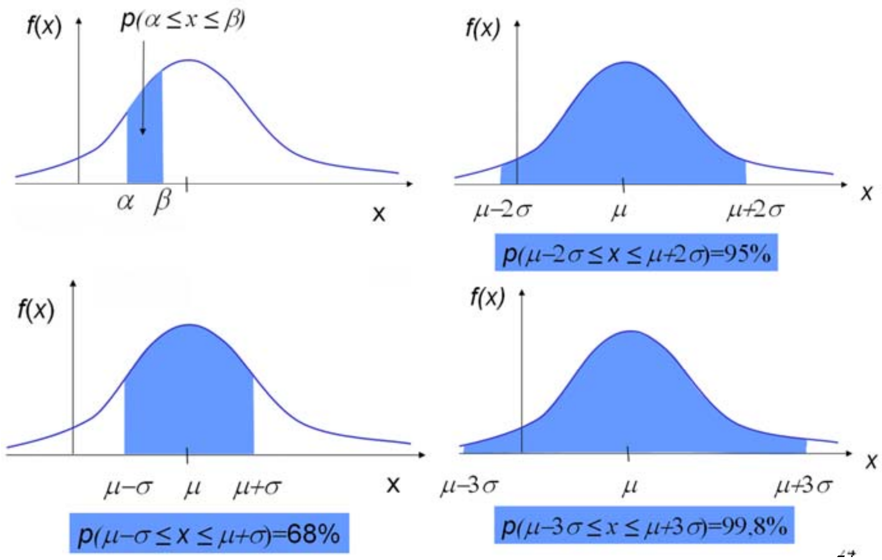
Normalverteilung (Gauss-Verteilung)

für die dargestellte Funktion: $\mu = 3$, $\sigma = 1$

Position des Maximums und die Breite der Kurve



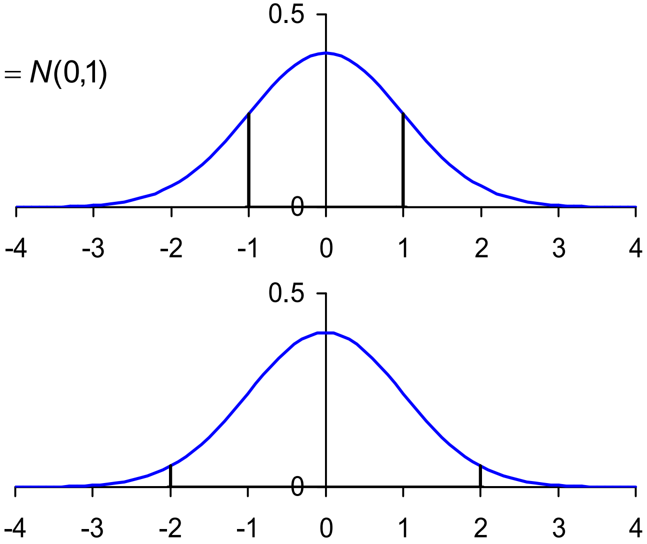
Normalverteilung



Standard - Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = N(0,1)$$

$\boxed{\mu = 0 \quad \sigma = 1}$



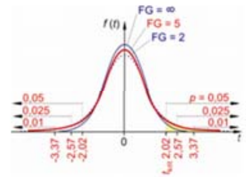
1. STATISTISCHE TABELLEN

t-VERTEILUNG

Freiheitsgrad (FG)	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
	0,4	**0,25**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
	p (Irrtumswahrscheinlichkeit, zweiseitiger Test)						
	0,8	**0,5**	**0,2**	**0,1**	**0,05**	**0,02**	**0,01**
1	0,325	1,000	3,078	6,314	12,70	31,82	63,65
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763
29	0,256	0,683	1,311	1,699	2,045	2,462	2,756
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704
60	0,255	0,679	1,296	1,671	2,000	2,390	2,66
120	0,254	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,250	0,674	1,282	1,645	1,960	2,326	2,576

Pr.Buch Anhang S.27.2

t-Verteilungsfamilie



„Glockenkurven"

Je größer ist der Freiheitsgrad, desto schmaler ist die Kurve.

$\boxed{t_\infty \equiv N(0,1)}$

Verteilung der Summe von Zufallsvariablen

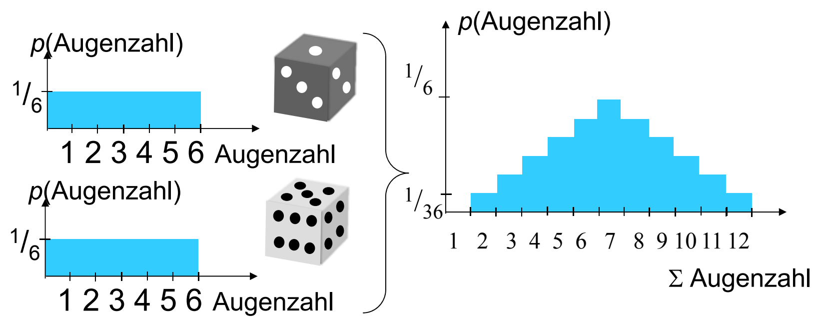
Beispiel

Wir werfen mit ein/zwei Würfeln.

Welche Verteilung hat die Summe der Augenzahlen?

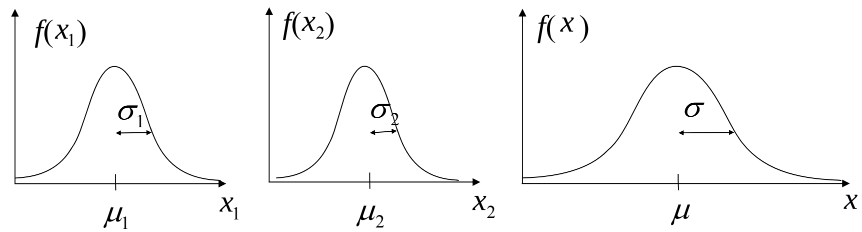
mit einem Würfel – Gleichverteilung

mit zwei Würfeln? – keine Gleichverteilung!



Verteilung der Summe von Zufallsvariablen

x_1 und x_2 sind **unabhängige** Zufallsvariablen. Beide folgen dergleichen Verteilung mit Erwartungswerten μ_1 bzw. μ_2 und Streuungen σ_1 bzw. σ_2.
Welcher Verteilung folgt die Summe: $x = x_1 + x_2$?



x hat eine Verteilung mit Erwartungswert $\mu = \mu_1 + \mu_2$
und Varianz $\sigma^2 = \sigma_1^2 + \sigma_2^2$

die additive Größe ist die Varianz, nicht die Streuung!

Stabile und nicht stabile Verteilungen

Der Verteilungstyp ist ***stabil,*** wenn die Summe von zwei unabhängigen indentisch verteilten Zufallsvariablen auch denselben Verteilungstyp besitzt.

stabile Verteilungen: Normal-, Binomial-, Poisson-

nicht stabile Verteilungen: Gleich-

Beispiel:

x_1 und x_2 sind unabhängige Zufallsgrößen.
Beide folgen einer Normalverteilung mit Erwartungswerten μ_1 bzw. μ_2 und Streuungen σ_1 bzw. σ_2.
Welcher Verteilung folgt der Durchschnitt $\bar{x} = \frac{x_1 + x_2}{2}$?

x hat eine Normalverteilung mit Erwartungswert $\mu = \frac{\mu_1 + \mu_2}{2}$

und Varianz $\sigma^2 = \frac{\sigma_1^{\;2} + \sigma_2^{\;2}}{2^2}$ $\left(\sigma = \frac{\sqrt{(\sigma_1^2 + \sigma_2^2)}}{2} \right)$

Zentraler Grenzwertsatz

Es seien x_1, $x_{2,\ldots}x_n$ unabhängige Zufallsgrößen, die alle derselben Verteilung haben. Die **Verteilung der Summe** $S_n = \sum_{i=1}^{n} x_i$ nähert sich einer **Normalverteilung,** wenn $n \to \infty$.

Die Summe der Verteilungsfunktionen konvergiert gegen eine Normalverteilung auch wenn die einzelnen Zufallsgrößen keine Normalverteilung haben.

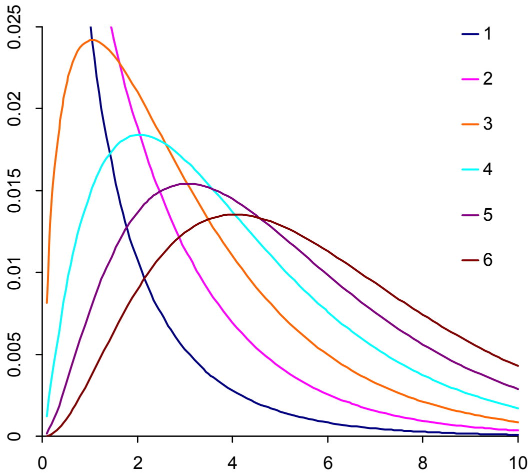
Biologische Bedeutung:
Wenn ein Parameter (zB. Körpergröße, Blutzuckerkonzentration) durch viele ($n \to \infty$) anderen Faktoren (Zufallsgrößen) beeinflusst wird, folgt dieser Parameter einer Normalverteilung.

Wichtige Verteilungen der analytischen Statsitik

Khi-Quadrat (χ^2) Verteilung

Wenn x_1, $x_{2,\ldots}x_n$ unabhängige standardnormalverteilte Zufallsgrößen sind, dann hat die Zufallsgröße
$\chi_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2$
eine sogenannte
χ^2 –Verteilung mit
n Freiheitsgraden

$$\boxed{\begin{gathered}\mu = n \\ \sigma^2 = 2n\end{gathered}}$$



der Modalwert der Verteilungen mit Freiheitsgrad $n > 2$: (n-2)

Wichtige Verteilungen der analytischen Statistik

t-Verteilung (Student-Verteilung)

x, x_1, x_2 ,... x_n sind unabhängige, standardnormalverteilte Zufallsgrößen. Dann die Größe

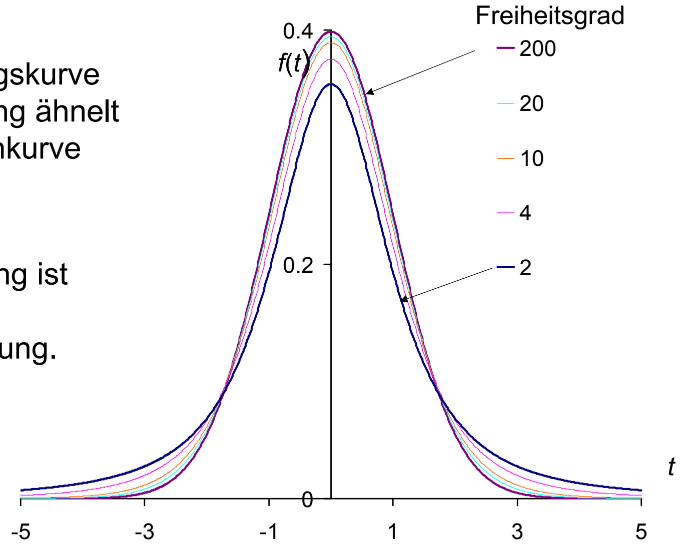
$$t_n = \frac{x}{\sqrt{\frac{(x_1^2 + x_2^2 + \ldots + x_n^2)}{n}}}$$

folgt einer t-Verteilung mit Freiheitsgrad n.
Die t-Verteilung ist symmetrisch, μ_n=0 ($n \geq 2$),
die Streuung ist: $\sigma = \sqrt{\frac{n}{n-2}}$ ($n \geq 3$)

Die Kurven der *t*-Verteilungen

Die Verteilungskurve der *t*-Verteilung ähnelt einer Glockenkurve (wenn $n \to \infty$).

Die *t*-Verteilung ist breiter als die Normalverteilung.



Überblickstabelle

	Lagemaße	Variabilitätsmaße
diskrete Verteilung	$\mu = \sum_i x_i p(x_i)$	$\sigma^2 = \sum_i (x_i - \mu)^2 p(x_i)$
empirische Werte	$\bar{x} = \frac{\sum_i x_i}{n}$ $\bar{x} = \sum_i x_i h(x_i)$	$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$ $s^2 = \frac{\sum_i x_i^2}{n} - \left(\frac{\sum_i x_i}{n} \right)^2$
kont. Verteilung	$\mu = \int_{-\infty}^{+\infty} x f(x) dx$	$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$

Überblickstabelle

Diskrete Verteilungen

	diskrete Gleich-	Binomial-	Poisson-
p_k	$\frac{1}{n}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$\frac{\lambda^k}{k!} e^{-\lambda}$
μ	$\frac{n+1}{2}$	np	λ
σ^2	$\frac{n^2-1}{12}$	$np(1-p)$	λ

Überblickstabelle

Kontinuierliche Verteilungen

$x_1^2 + x_2^2 + \ldots + x_n^2$ $\frac{x}{\sqrt{\frac{(x_1^2 + x_2^2 + \ldots + x_n^2)}{n}}}$

	kontinuier. Gleich-	Normal-	Standard-normal-	Khi-Quadrat-	*t*-
$f(x)$	$\frac{1}{b-a}$, $a \le x \le b$	$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$	$\frac{1}{\left(\frac{n}{2}-1\right)! 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$	$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ $f(t) = \frac{K}{\left(1 + \frac{t^2}{f}\right)^{\frac{f+1}{2}}}$
μ	$\frac{a+b}{2}$	μ	0	n	0 n ≥ 2
σ^2	$\frac{(b-a)^2}{12}$	σ^2	1	$2n$	$\frac{n}{n-2}$ n ≥ 3