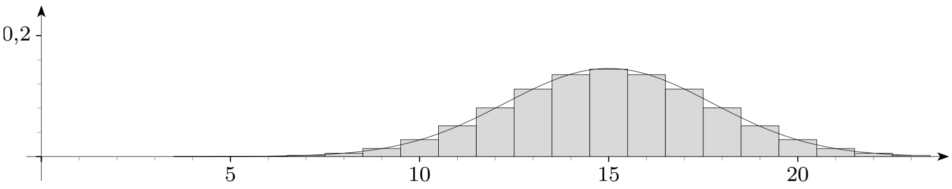
Sigma-Umgebung

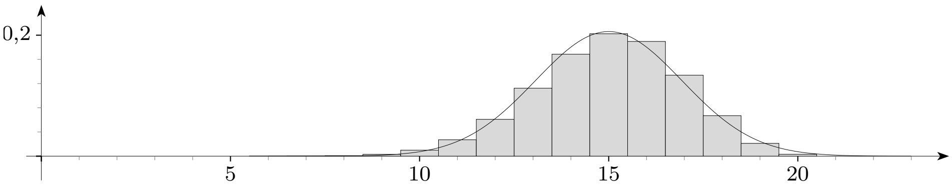
Vergleichen wir die beiden Binomialverteilungen:

$n = 30$
$p = 0{,}5$

(z.B. 30-maliges Werfen einer Münze, X Anzahl von "Zahl")



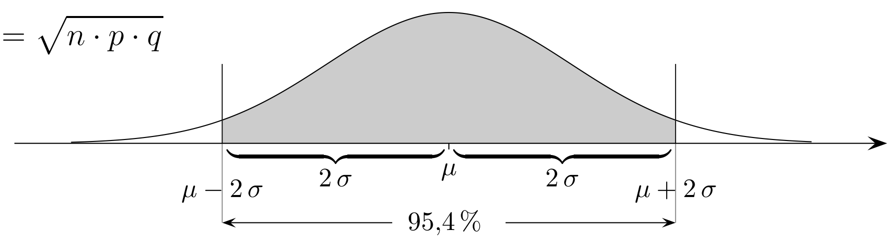
$n = 20$
$p = 0{,}75$



Der Erwartungswert beider Binomialverteilungen beträgt 15. Jedoch sind die Wahrscheinlichkeiten, mit denen die Zufallsvariable X um den Erwartungswert streut, verschieden.
Wie weit darf die Zufallsvariable vom Erwartungswert höchstens abweichen, damit kein Zweifel an der Unverfälschtheit der Münze (allgemeiner: an der zugrundegelegten Wahrscheinlichkeit) berechtigt ist? Um solche und ähnliche Fragen beantworten zu können, benötigen wir zunächst ein Maß für die Streuung um den Erwartungswert.
Wenn wir die Binomialverteilung durch eine symmetrische Kurve annähern, so charakterisiert die Lage der beiden Wendepunkte die Streuung um den Erwartungswert.

Der Abstand vom Erwartungswert zur x-Koordinate eines Wendepunkts heißt daher Standardabweichung und wird mit σ (lies: sigma) bezeichnet.

Mit Mitteln der Analysis kann $\sigma = \sqrt{n \cdot p \cdot q}$ bestimmt werden.



1. Bestimmen Sie für $p = 0{,}5$ und $n = 20$ (50) die 2σ-Umgebung des Erwartungswerts. Mit welcher Wahrscheinlichkeit liegen die Werte von X in dieser Umgebung?

2. Jede Woche werden beim Lotto 6 aus 49 ca. 90 Millionen Tipps abgegeben. Wie viele Tipps mit 6 Richtigen werden dabei sein? (Wahrscheinlichkeit für 6 Richtige beträgt 1/13 983 816)

3. Ein ungewöhnlicher Knabenüberschuss beschäftigte 1982 in Bonn die Mediziner. Es wurden 2762 Jungen und 2494 Mädchen geboren. Liegt hier eine signifikante (bedeutsame) Abweichung vom bisherigen Anteil der Jungengeburten von 51,4 % vor?

σ-Umgebung   Ergebnisse

1.

n	μ	σ	2σ-Umgebung	3σ-Umgebung
20	10	2,24	6, ... , 14	4, ... , 16
50	25	3,54	18, ... , 32	15, ... , 35

In der 2σ-Umgebung sind alle Stichprobenergebnisse - und diese sind stets ganzzahlig - aus $[\mu - 2\sigma \mid \mu + 2\sigma]$, d.h. es wird hier "nach innen" gerundet.

n	Wahrscheinlichkeit		
20		0,959	0,997
50		0,967	0,997

Trotz der Unterschiede im Stichprobenumfang n sind die Wahrscheinlichkeiten für die σ-Umgebungen ungefähr gleich.
Auch wenn wir die zugrundegelegten Trefferwahrscheinlichkeiten variierten, ergäbe sich kein anderes Bild.

Für n-stufige Bernoulli-Versuche gilt unter der sogenannten Laplace-Bedingung $\sigma > 3$:

Die Wahrscheinlichkeit, dass die Anzahl der Treffer in der 2σ-Umgebung des Erwartungswerts liegt, beträgt ca. 95,4 % (2σ-Regel), für die 3σ-Umgebung sind es ca. 99,7 % (3σ-Regel).

2. $\mu = 6{,}44$
   $\sigma = 2{,}54$
   $\mu - 2\sigma = 1{,}36$
   $\mu + 2\sigma = 11{,}51$

   2σ-Umgebung $= [2,\ 11]$

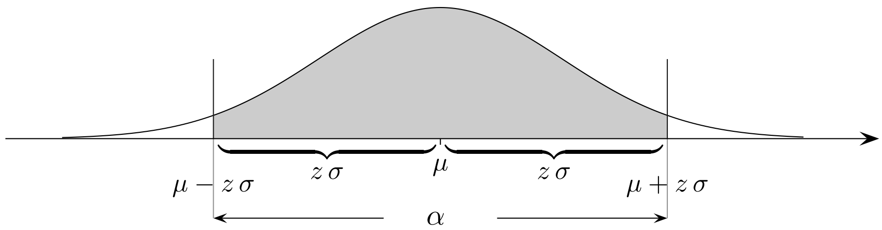
3. $\mu = 2701{,}58$
   $\sigma = 36{,}23$
   $\mu - 2\sigma = 2629{,}11$
   $\mu + 2\sigma = 2774{,}05$

   2σ-Umgebung $= [2630,\ 2774]$

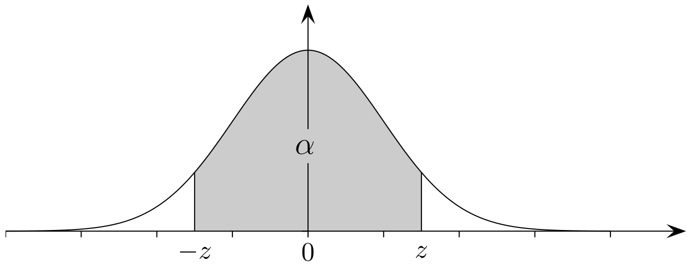
   Das Ereignis ist nicht ungewöhnlich.

$z\sigma$-Umgebung

Die Sicherheitswahrscheinlichkeit α (z. B. 98 %) sei vorgegeben.
Gesucht ist das Vielfache z von σ, so dass die Wahrscheinlichkeit des Bereichs $[\mu - z\sigma \mid \mu + z\sigma]$ α beträgt. X weicht dann mit Wahrscheinlichkeit α höchstens um $z\sigma$ vom Erwartungswert μ ab.



z wird mit der Normalverteilung ermittelt.



$$z = \Phi^{-1}\left(\frac{1}{2} + \frac{\alpha}{2}\right) = \Phi^{-1}\left(\frac{1+\alpha}{2}\right)$$

α	z
0,90	1,64
0,95	1,96
0,954	2
0,99	2,58
0,9973	3

Beispiel:

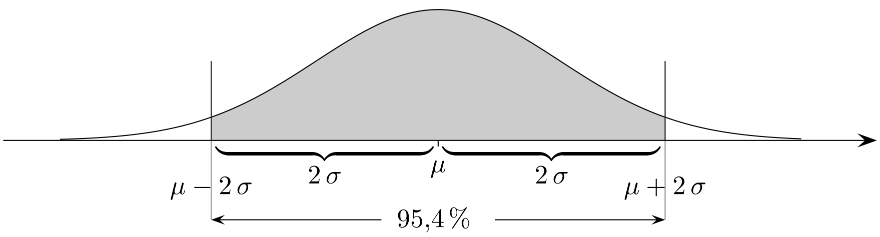
$n = 500$
$p = 0{,}5$
$\alpha = 98\,\%$

$\mu = 250$
$\sigma = 11{,}18$
$z = 2{,}33$ $\quad -z = \text{invNorm}(0{,}01)$

[224, 276]

Dieser Bereich kann mit dem GTR auch direkt ermittelt werden:
223,9 = invNorm(0.01, 250, 11.18) und 276,0 = invNorm(0.99, 250, 11.18),
jedoch ist die obige Überlegung zur Berechnung von Konfidenzintervallen nützlich.

Wie wirkt sich eine Vergrößerung von n auf die Größe der 2σ-Umgebung aus?

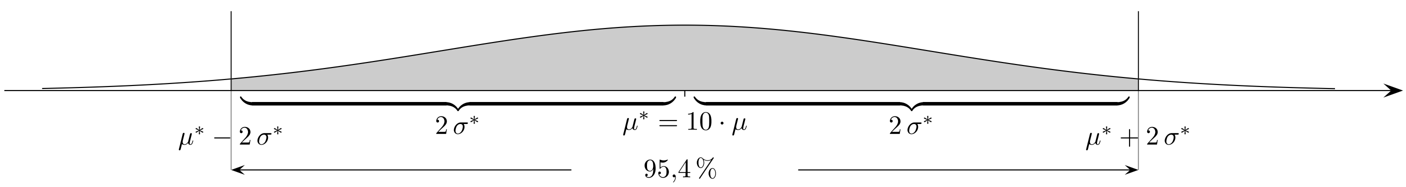


Nehmen wir an, n werde verzehnfacht.

Die Standardabweichung beträgt dann

$$\begin{aligned}\sigma^* &= \sqrt{10 \cdot n \cdot p \cdot q} \\ &= \sqrt{10} \cdot \sigma\end{aligned}$$

Die Größe der $2\sigma^*$-Umgebung unterscheidet sich also von der 2σ-Umgebung um den Faktor $\sqrt{10}$.
Für die Grafik wurde der Faktor 2 gewählt.



Beginnen wir mit dem Vergleich bei $n = 1$, so vergrößert sich die 2σ-Umgebung mit zunehmendem n um den Faktor $\sqrt{n}$ ($\sqrt{n}$-Gesetz).
Für $z\sigma$-Umgebungen ändert sich hieran nichts.

$z\sigma$-Umgebungen für Mittelwerte

Das bisher Gesagte kann auf die Zufallsvariable

$$\overline{X} = \frac{1}{n}\sum X_i$$

mit

$$\mu_{\overline{X}} = \mu$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

übertragen werden.
Hierbei ist $E(X_i) = \mu,\ V(X_i) = \sigma^2$.
Die X_i sind voneinander unabhängig.