

1. Sigma-Umgebung
2. σ -Umgebung Ergebnisse Regeln
3. σ -Umgebung mit der Normalverteilung
4. $z\sigma$ -Umgebung
5. $z = \Phi^{-1}(\frac{1+\alpha}{2})$
6. Sigma-Regeln Graphen für 68,3%, 90%, 95%
7. σ -Regeln, Prognoseintervall
8. Wie wirkt sich eine Vergrößerung von n ?
9. $z\sigma$ -Umgebung für Mittelwerte
10. $P(X > k \cdot n)$ für größer werdendes n
11. $1/\sqrt{n}$ - Gesetz
12. Aufgaben Bernoullikette Tabellarische Lösung

Für den Anfang geeignet

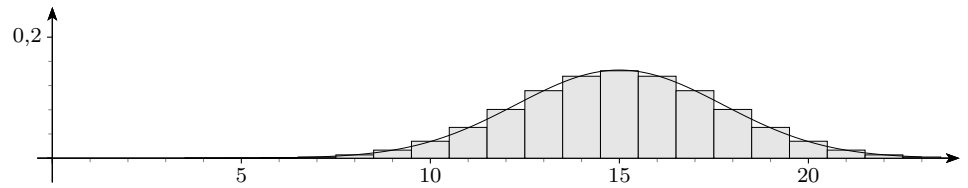
↑ Sigma-Umgebung

Vergleichen wir die beiden Binomialverteilungen:

$$n = 30$$

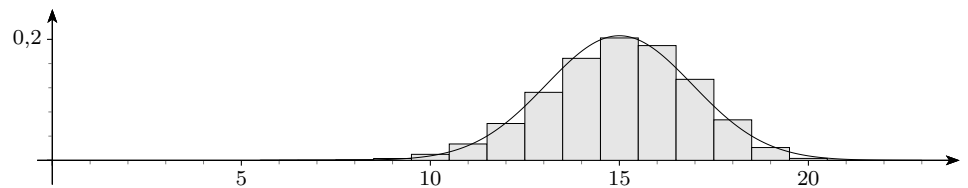
$$p = 0,5$$

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



$$n = 20$$

$$p = 0,75$$



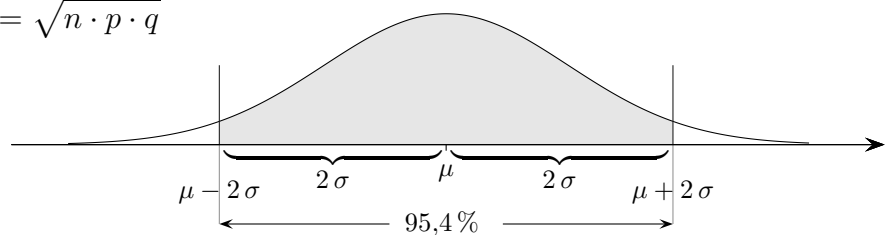
Der Erwartungswert beider Binomialverteilungen beträgt 15. Jedoch sind die Wahrscheinlichkeiten, mit denen die Zufallsvariable X um den Erwartungswert streut, verschieden.

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 (den 2σ -Strebereich). 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, falls die Laplace-Bedingung $\sigma > 3$ erfüllt ist: Die Wahrscheinlichkeit (genähert), dass die Anzahl der Treffer in der

<i>1σ-Umgebung liegt, beträgt 68,3% $\approx \frac{2}{3}$ (1σ-Regel),</i>	<code>normalcdf(-1, 1)</code>
<i>2σ-Umgebung liegt, beträgt 95,4% (2σ-Regel),</i>	<code>normalcdf(-2, 2)</code>
<i>3σ-Umgebung liegt, beträgt 99,7% (3σ-Regel).</i>	<code>normalcdf(-3, 3)</code>

2. $\mu = 6,44$

$\sigma = 2,54$

$\mu - 2\sigma = 1,36$

$\mu + 2\sigma = 11,51$

2σ -Umgebung = [2, 11]

3. $\mu = 2701,58$

$\sigma = 36,23$

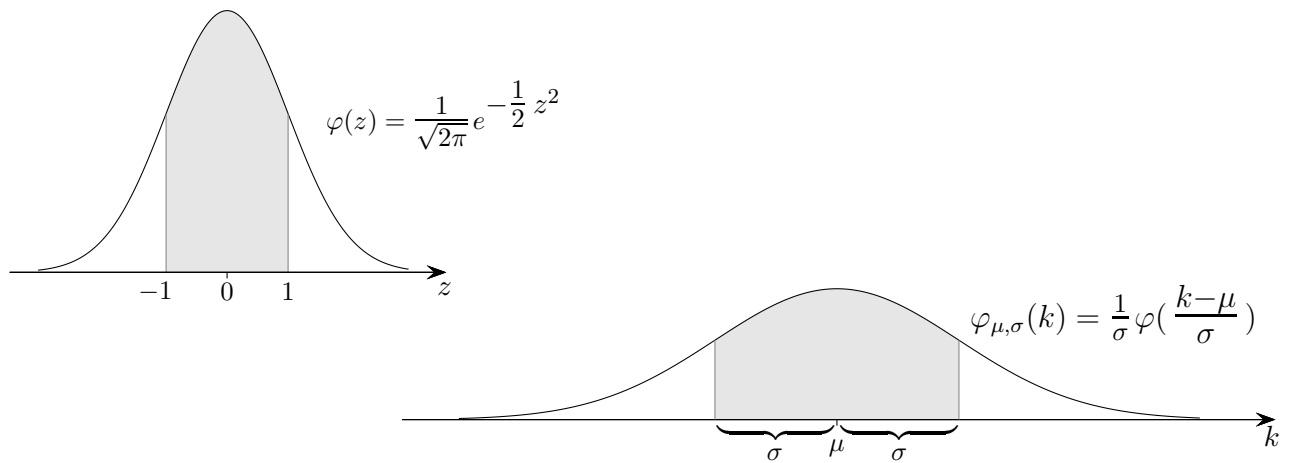
$\mu - 2\sigma = 2629,11$

$\mu + 2\sigma = 2774,05$

2σ -Umgebung = [2630, 2774]

Das Ereignis ist nicht ungewöhnlich.

↑ σ -Umgebung mit der Normalverteilung



Die (Dichte-)Funktion $\varphi_{\mu,\sigma}$ der Normalverteilung ergibt sich aus der Gaußschen Glockenkurve φ durch Strecken und Stauchen um σ und verschieben um μ (siehe Normalverteilung).

Einem z -Wert entspricht $k = \mu + z\sigma$.

Hierbei geht 0 in μ über und das Intervall $[-1, 1]$ in die σ -Umgebung $[\mu - \sigma, \mu + \sigma]$ von μ .

$$\begin{aligned} k &= \mu + z\sigma \\ \Leftrightarrow k - \mu &= z\sigma \\ \Leftrightarrow z &= \frac{k - \mu}{\sigma} \end{aligned}$$

Zur Berechnung von Wahrscheinlichkeiten kann daher stets die Gaußsche Glockenkurve herangezogen werden.

$$\begin{aligned} P(\mu - \sigma \leq k \leq \mu + \sigma) &= 68,3\% \approx \frac{2}{3} & P(-1 \leq z \leq 1) &= \text{normalcdf}(-1, 1, \overbrace{0, 1}^{\text{kann entfallen}}) \\ P(\mu - 2\sigma \leq k \leq \mu + 2\sigma) &= 95,4\% & P(-2 \leq z \leq 2) &= \text{normalcdf}(-2, 2) \\ P(\mu - 3\sigma \leq k \leq \mu + 3\sigma) &= 99,7\% & P(-3 \leq z \leq 3) &= \text{normalcdf}(-3, 3) \end{aligned}$$

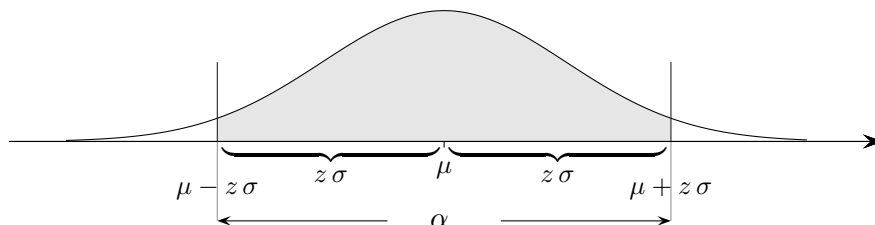
Für n -stufige Bernoulli-Versuche gilt, falls die Laplace-Bedingung $\sigma > 3$ erfüllt ist:
Die Wahrscheinlichkeit (genähert), dass die Anzahl der Treffer in der

1σ -Umgebung liegt, beträgt $68,3\% \approx \frac{2}{3}$	(1σ -Regel),	$\text{normalcdf}(-1, 1)$
2σ -Umgebung liegt, beträgt $95,4\%$	(2σ -Regel),	$\text{normalcdf}(-2, 2)$
3σ -Umgebung liegt, beträgt $99,7\%$	(3σ -Regel).	$\text{normalcdf}(-3, 3)$

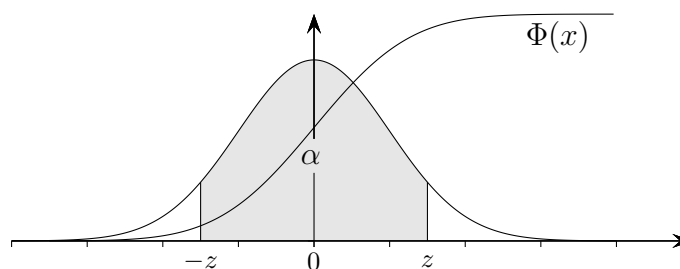
↑ $z\sigma$ -Umgebung

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.



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

α	z
0,90	1,64
0,95	1,96
0,954	2
0,99	2,58
0,997	3

Beispiel:

$$n = 500$$

$$p = 0,5$$

$$\alpha = 98\%$$

$$\mu = 250$$

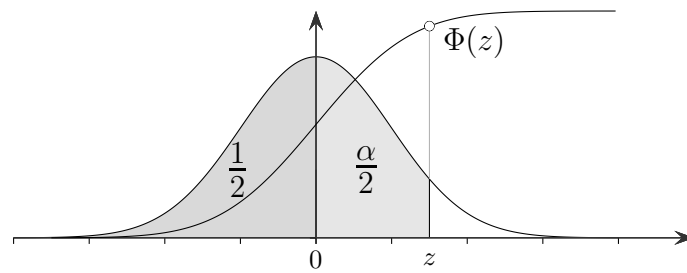
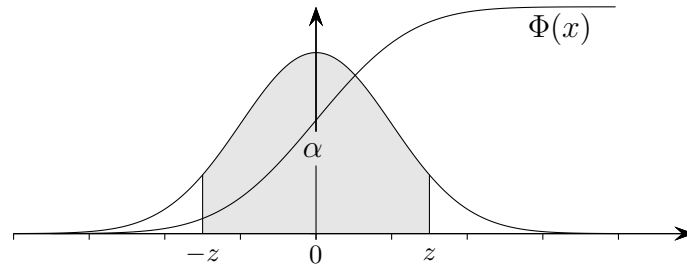
$$\sigma = 11,18$$

$$z = 2,33 \quad -z = \text{invNorm}(0,01)$$

$$[224, 276]$$

Dieser Bereich kann mit dem GTR auch direkt ermittelt werden:
 $223,9 = \text{invNorm}(0,01, 250, 11,18)$ und $276,0 = \text{invNorm}(0,99, 250, 11,18)$,
 jedoch ist die obige Überlegung zur Berechnung von Konfidenzintervallen nützlich.

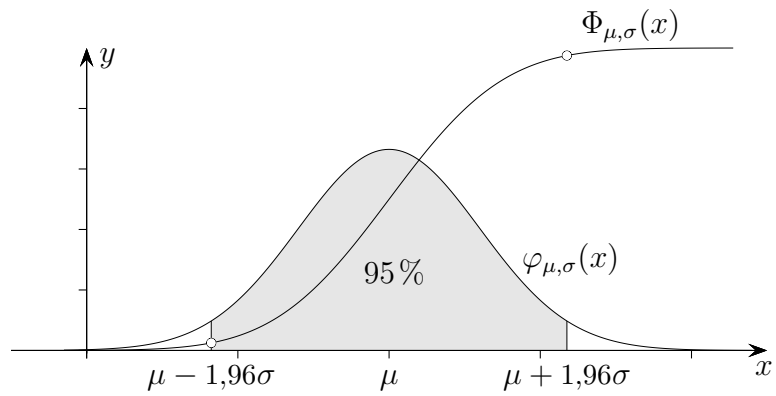
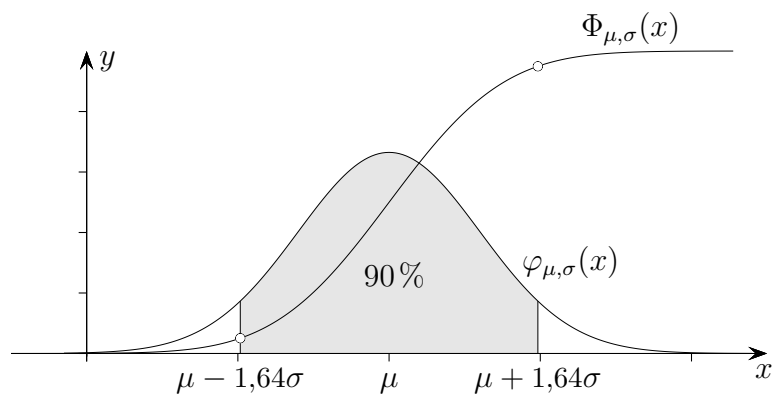
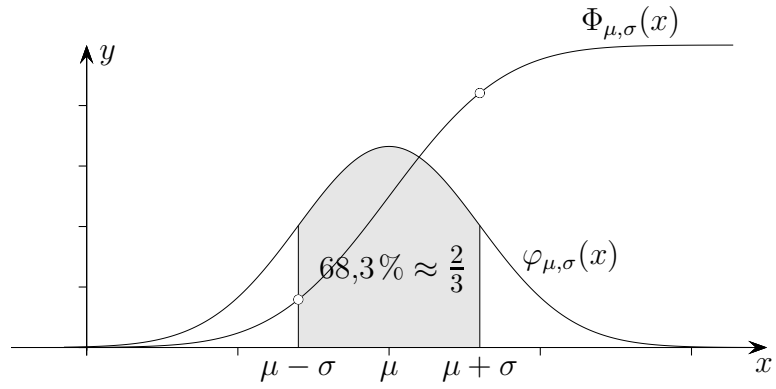
$$\uparrow z = \Phi^{-1}\left(\frac{1+\alpha}{2}\right)$$



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

\uparrow

↑ Sigma-Regeln



↑ σ -Regeln, Prognoseintervall

Für n -stufige Bernoulli-Versuche, $\mu = n \cdot p$, $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$, gilt, falls die Laplace-Bedingung $\sigma > 3$ erfüllt ist:

(Ist X eine normalverteilte Zufallsgröße, so gilt:)

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68,3\% \approx \frac{2}{3}$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95,4\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99,7\%$$

$$P(\mu - 1,64\sigma \leq X \leq \mu + 1,64\sigma) \approx 90,0\%$$

$$P(\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma) \approx 95,0\%$$

$$P(\mu - 2,58\sigma \leq X \leq \mu + 2,58\sigma) \approx 99,0\%$$

Für eine Punktschätzung wird der Erwartungswert bestimmt: $\mu = n \cdot p$

Ein Prognoseintervall $[\mu - z\sigma; \mu + z\sigma]$ wird mit den σ -Regeln ermittelt.

$$n = 100, \quad p = 0,2$$

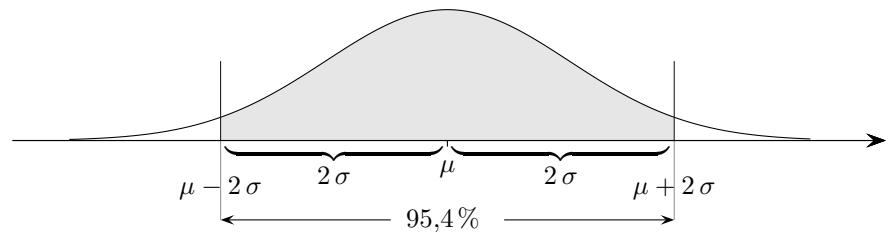
$$\mu = 20, \quad \sigma = 4$$

$$[\mu - 1,96\sigma; \mu + 1,96\sigma] = [12,16; 27,84]$$

Das 95%-Prognoseintervall ist $[12; 28]$.

Mit einer Wahrscheinlichkeit von mindestens 95% wird man mindestens 12 und höchstens 28 Treffer erzielen. Im Sachkontext wurde nach außen gerundet, $P(12 \leq X \leq 28) = 96,7\%$.

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

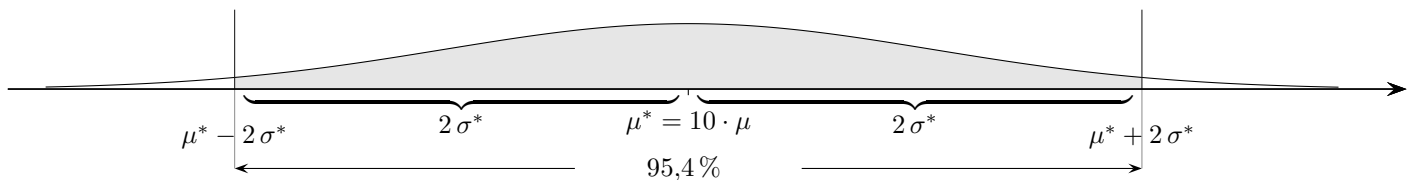


Nehmen wir an, n werde verzehnfacht.

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

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

mit

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

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

übertragen werden.

Hierbei ist $E(X_i) = \mu$, $V(X_i) = \sigma^2$.

Die X_i sind voneinander unabhängig.

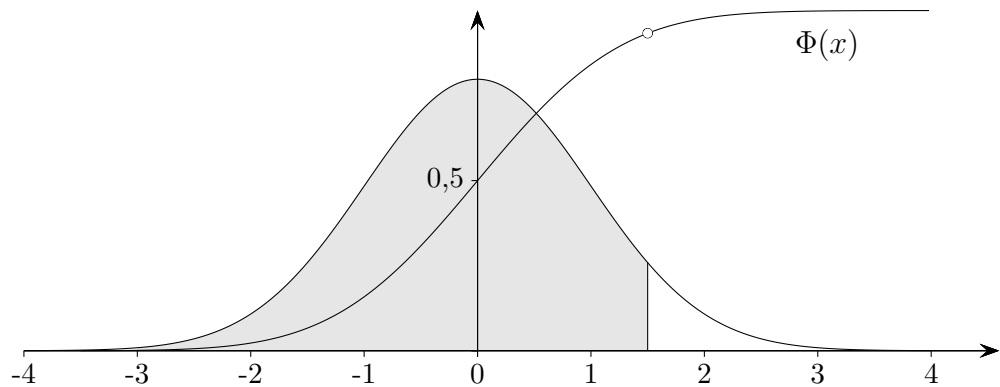
X sei binomialverteilt mit n und p , $k \in [0, 1]$.
Untersuche

$P(X > k \cdot n)$ für größer werdendes n .

X sei binomialverteilt mit n und p , $k \in [0, 1]$.
 Untersuche

$P(X > k \cdot n)$ für größer werdendes n .

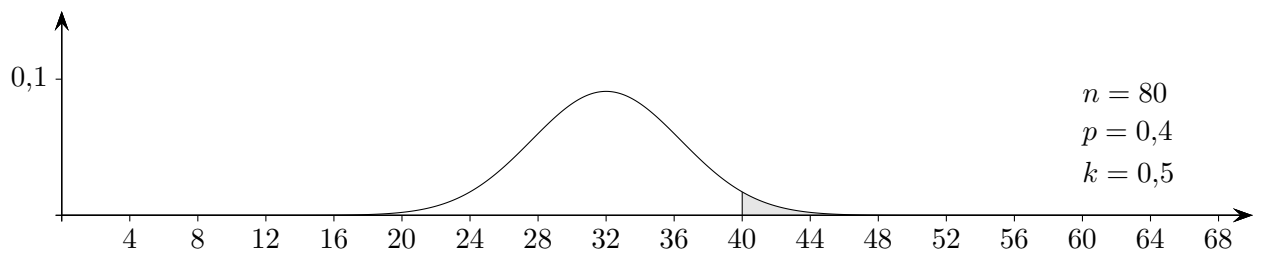
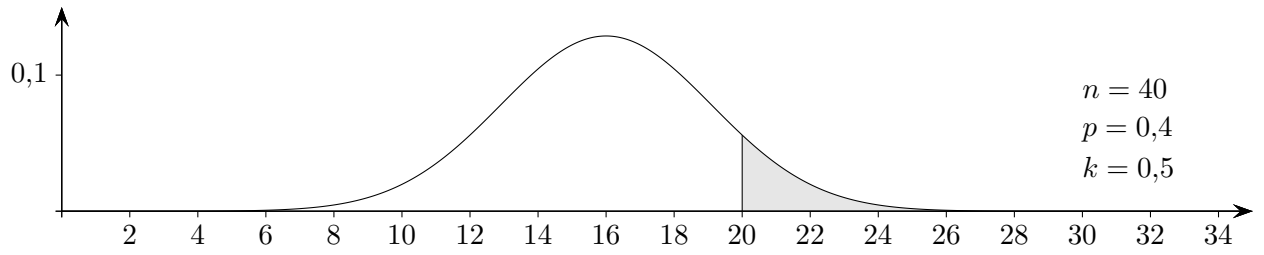
$$\begin{aligned}
 P(X > k \cdot n) &= 1 - P(X \leq k \cdot n) \\
 &= 1 - \Phi\left(\frac{k \cdot n - n \cdot p}{\sqrt{n \cdot p \cdot q}}\right) \\
 &= 1 - \Phi\left(\frac{n \cdot (k - p)}{\sqrt{n \cdot p \cdot q}}\right) \\
 &= 1 - \Phi\left(\frac{\sqrt{n} \cdot (k - p)}{\sqrt{p \cdot q}}\right)
 \end{aligned}$$



$k = p$: $P(X > k \cdot n) = \frac{1}{2}$ $k \cdot n$ stimmt mit dem Erwartungswert $n \cdot p$ überein.

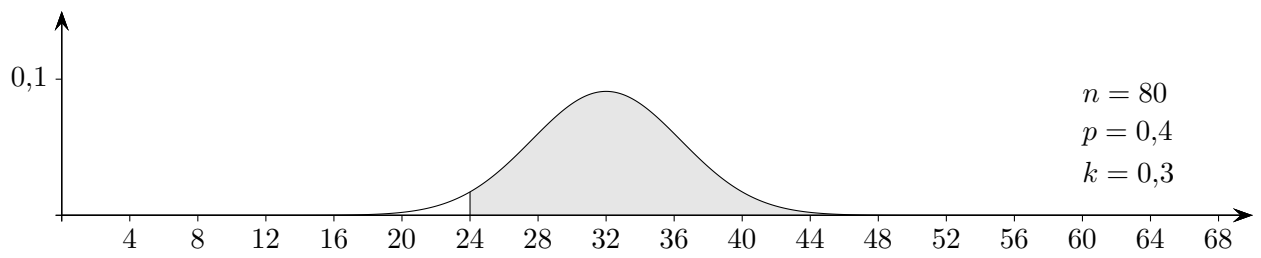
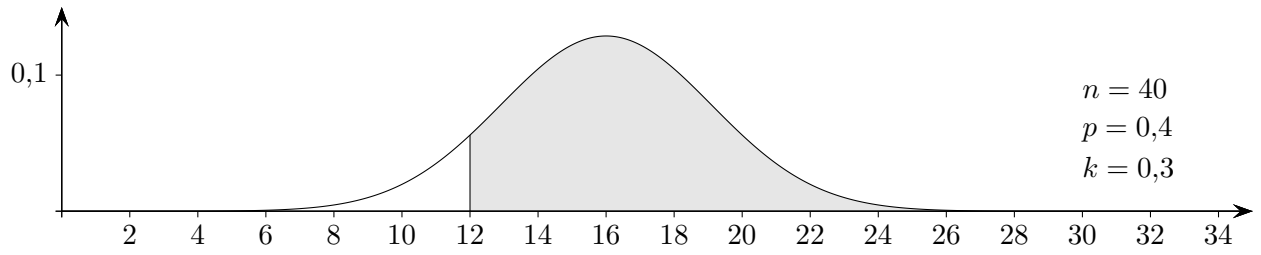
$k > p$: $P(X > k \cdot n) \rightarrow 0$ $k \cdot n$ liegt rechts vom Erwartungswert $n \cdot p$.

$k < p$: $P(X > k \cdot n) \rightarrow 1$ $k \cdot n$ liegt links vom Erwartungswert $n \cdot p$.



$k > p$: $P(X > k \cdot n) \rightarrow 0$ $k \cdot n$ liegt rechts vom Erwartungswert $n \cdot p$.

Die Wahrscheinlichkeit für mindestens $k \cdot 100\%$ Treffer von n strebt für wachsendes n gegen null.

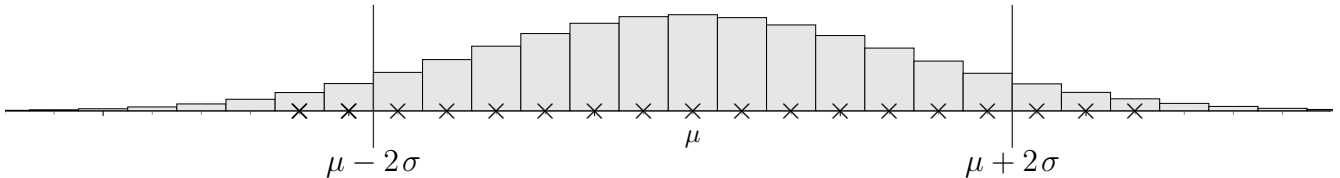


$k < p$: $P(X > k \cdot n) \rightarrow 1$ $k \cdot n$ liegt links vom Erwartungswert $n \cdot p$.

Die Wahrscheinlichkeit für mindestens $k \cdot 100\%$ Treffer von n strebt für wachsendes n gegen 1.

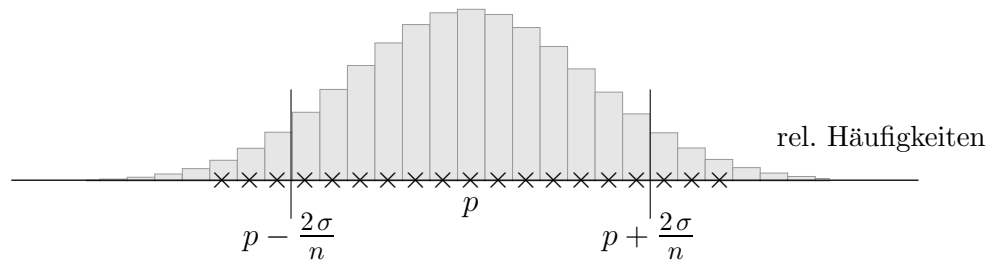
↑ $1/\sqrt{n}$ - Gesetz

Wir gehen von einer Binomialverteilung einer Stichprobe (n , p sind gegeben) und der 2σ -Umgebung des Erwartungswerts zur Sicherheitswahrscheinlichkeit $\alpha = 95,4\%$ aus.



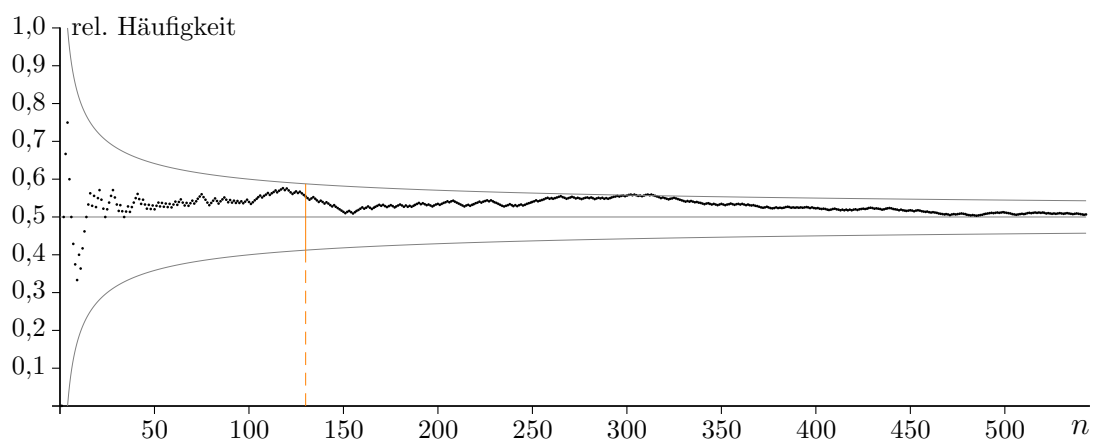
Die Stichprobenergebnisse werden mit der Sicherheitswahrscheinlichkeit α in der 2σ -Umgebung liegen. Diese heißt auch Schwankungs- oder Prognoseintervall für die Stichprobenergebnisse.

Die relativen Häufigkeiten $\frac{k}{n}$ der Stichprobenergebnisse liegen dann im Intervall $[p - \frac{2\sigma}{n} | p + \frac{2\sigma}{n}]$, dem Prognoseintervall für die relativen Häufigkeiten.



Zu gegebenem p sind also Aussagen über die zu erwartenden Stichprobenergebnisse bzw. über die relativen Häufigkeiten möglich.

Während die Prognoseintervalle für die Stichprobenergebnisse mit wachsendem n immer größer werden, ziehen sich die Prognoseintervalle für die relativen Häufigkeiten immer mehr zusammen.



$1/\sqrt{n}$ -Gesetz (hier für $p = 0,5$, dann ist $2\sigma/n = 1/\sqrt{n}$)
 Für n gilt: Die rel. Häufigkeit liegt mit ca. 95% Sicherheit im Intervall $[0,5 - 1/\sqrt{n}, 0,5 + 1/\sqrt{n}]$.
 Wird n vervierfacht, halbiert sich die Länge des Intervalls.

Wie oft muss man eine Münze werfen, damit die relative Häufigkeit für „Wappen“ mit einer Sicherheit von 95% im Intervall $[0,48; 0,52]$ liegt?

$$1/\sqrt{n} = 0,02 \implies n = 2500 \text{ (mindestens)}$$

1. Bernoullikette, $n = 100$, $p = 0,3$, X Anzahl der Treffer, $\alpha = 5\%$
Suche das
 - a) größte ganzzahlige k mit $P(X \leq k) \leq \alpha$,
 - b) kleinste ganzzahlige k mit $P(X \geq k) \leq \alpha$.

2. Bernoullikette, $n = 320$, $p = 0,05$, X Anzahl der Treffer
Bestimmen Sie das kleinste um den Erwartungstert von X symmetrische Intervall,
in dem X mit einer Wahrscheinlichkeit von mindestens 90% liegt.

↑

1. Bernoullikette, $n = 100$, $p = 0,3$, X Anzahl der Treffer, $\alpha = 5\%$
 Suche das

a) größte ganzzahlige k mit $P(X \leq k) \leq \alpha$,

Tabellarische Lösung: $\backslash Y1 = \text{binomcdf}(n, p, X)$

X	\Y1
22	0,04787
23	0,07553

$[0; 22]$ ist das größte Intervall mit den geforderten Eigenschaften.

b) kleinste ganzzahlige k mit $P(X \geq k) \leq \alpha$.

Tabellarische Lösung: $\backslash Y1 = 1 - \text{binomcdf}(n, p, X-1)$

X	\Y1
38	0,05305
39	0,03398

$[39; 100]$ ist das größte Intervall mit den geforderten Eigenschaften.

2. Bernoullikette, $n = 320$, $p = 0,05$, X Anzahl der Treffer

Bestimmen Sie das kleinste um den Erwartungstert von X symmetrische Intervall,
 in dem X mit einer Wahrscheinlichkeit von mindestens 90% liegt.

Gesucht ist das kleinste k mit $P(16 - k \leq X \leq 16 + k) \geq 90\%$.

Tabellarische Lösung: $\backslash Y1 = \text{binomcdf}(n, p, 16+X) - \text{binomcdf}(n, p, 16-X-1)$

X	\Y1
5	0,84397
6	0,90688

$[10; 22]$ ist das kleinste Intervall mit den geforderten Eigenschaften.