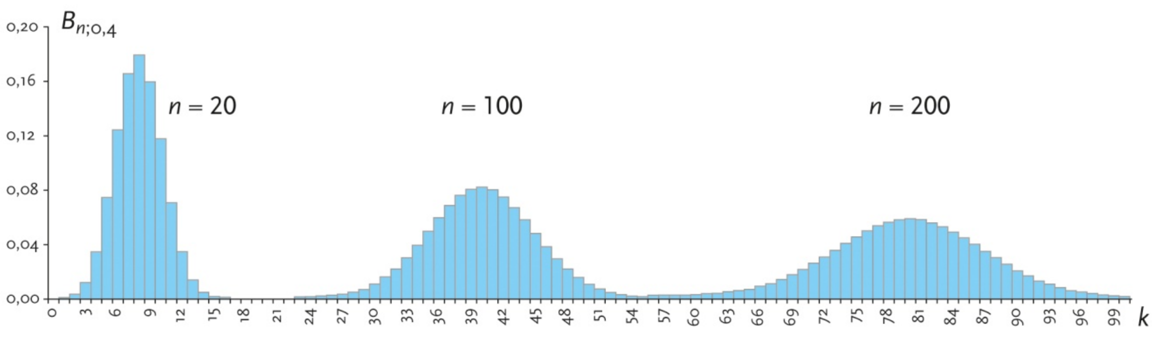
STETIGE VERTEILUNGEN

1. Die Näherungsformel von Moivre-Laplace



Betrachtet man die Binomialverteilungen $B_{n;p}$ für wachsendes n bei konstantem p, so werden die Histogramme einer binomialverteilten Zufallsvariablen breiter und symmetrischer um den Erwartungswert $\mu = n \cdot p$.

Die Wahrscheinlichkeit eines einzelnen Ergebnisses wird immer kleiner, da die Flächensumme der Rechtecke immer die Gesamtwahrscheinlichkeit 1 ergibt.

Die Histogramme erhalten zunehmend Glockenform, wobei sich die (Symmetrie-)Achse an der Stelle $\mu = n \cdot p$ immer mehr nach rechts verschiebt.

Um das Verhalten von $B_{n;p}$ für große Werte von n besser untersuchen zu können, verschiebt man die Schaubilder so, dass der Erwartungswert $\mu = n \cdot p$ auf der 2. Koordinatenachse liegt und gleicht somit die Verschiebung der (Symmetrie-) Achse aus. Jeder Wert X=k wird um μ Einheiten nach links verschoben.

Gleichzeitig streckt man die Rechteckshöhen, die $B_{n;p}(k)$ – Werte, mit dem Faktor $\sigma = \sqrt{n \cdot p \cdot q}$ und die ursprünglichen Rechtecksbreiten mit 1LE mit dem Faktor $\frac{1}{\sigma}$. Damit gleicht man das Flacherwerden der Glockenform aus und hat gleichzeitig die Konstanz der Flächenmaßzahlen der Rechtecke (der Einzelwahrscheinlichkeiten) gewahrt.

Damit gilt:

Man erhält eine neu Zufallsvariable, ein **standartisierte Zufallsvaiable** $Z = \frac{1}{\sigma} \cdot (k - \mu) = \frac{k - \mu}{\sigma}$.

Für $k > \mu$ nimmt die standartisierte Zufallsvariable positive, für $k < \mu$ negative Werte an.
Eine solche Verteilung heißt **standartisierte Binomialverteilung**: $\sigma \cdot P(Z)$

De Moivre hat erkannt, dass die Histogramme bestimmter standartisierter Binomialverteilungen trotz unterschiedlicher Parameter n und p in guter Näherung einen fast identischen Verlauf zeigen. Diese Histogramme haben einen glockenförmigen Verlauf.

Laplace hat diese Überlegungen weitergeführt und erkannt, dass die Histogramme standartisierter Binomialverteilungen um so besser von glockenförmigen Graphen umrandet werden, je größer die Standardabweichung σ ist .(Faustregel: Wenn die Laplace-Bedingung $\sigma > 3$ erfüllt ist)

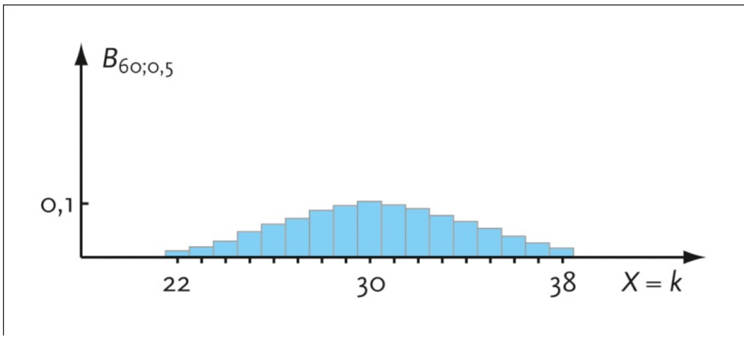
Das Schaubild der Funktion $\boldsymbol{\varphi : x \to \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}}$; $\mathbf{x \in IR}$ liefert die „Grenzkurve“ ,die Glockenkurve (als Grenzlage der Histogramme für $n \to \infty$)

Diese Funktion φ heißt **Gauß-Funktion**, ihr Schaubild heißt **Gauß’sche Glockenkurve.**

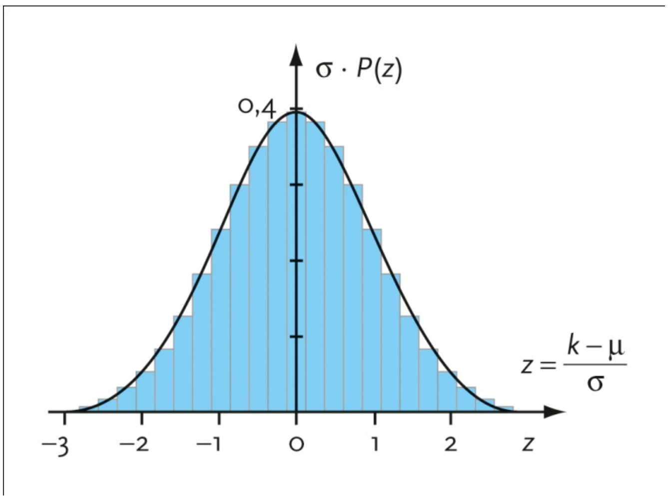
Diese Glockenkurve ist symmetrisch zur y-Achse und hat die x-Achse als Asymptote.

Moivre hat diese Glockenkurve für p=0,5 untersucht, Laplace zeigte, dass sich auch im Fall $p \neq 0{,}5$ für große Werte von n dieselbe Grenzkurve ergibt.

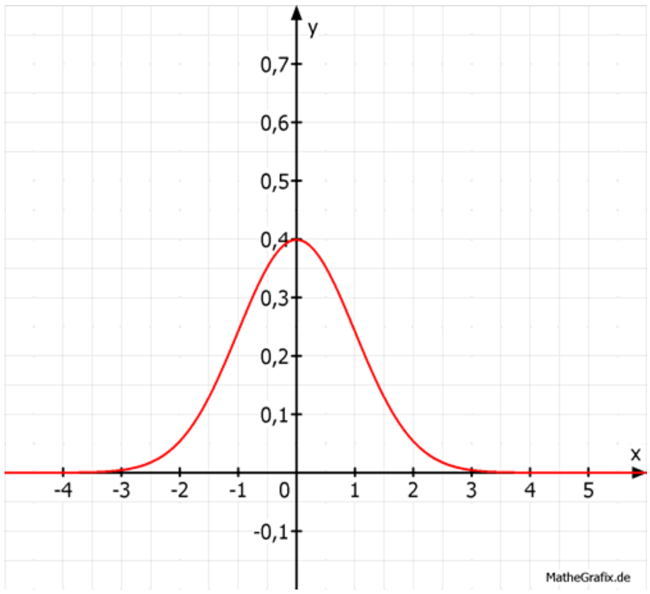
Beispiel: Binomialverteilung mit n=60, p=0,5, $\mu = n \cdot p = 30$, $\sigma = \sqrt{n \cdot p \cdot q} \approx 2{,}74$



Standartisierte Binomialverteilung:



Gauss-Kurve:



Der Flächeninhalt zwischen der Gauß-Kurve und der x-Achse entspricht somit dem der Summe der Inhalte aller Rechtecksflächen des Histogramms einer binomialverteilten Zufallsvariablen X ebenso wie die der dazugehörigen standardisierten Zufallsvariablen Z und hat der Wert 1: $\int_{-\infty}^{\infty} \varphi(x)dx = 1$

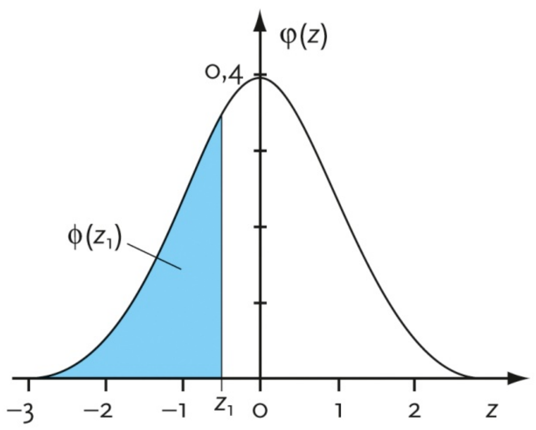
Die Summenwahrscheinlichkeit $P(X \le k) = B_{n;p}(X \le k)$ kann dann näherungsweise durch den Inhalt der Teilfläche, die von der Gauss-Kurve und der x-Achse (bzw. z-Achse) im Intervall $(-\infty; z]$ eingeschlossen wird, berechnet werden:

Für die integrale Näherung der Summenwahrscheinlichkeit $\mathbf{P(X \le k) = B_{n;p}(X \le k)}$ gilt:

$$\mathbf{B_{n;p}(X \le k) = \int_{-\infty}^{z} \varphi(x)dx \quad mit \quad z = \frac{k-\mu}{\sigma}}$$

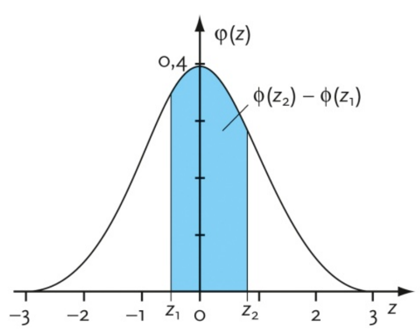
Diese Gleichung bezeichnet man als integrale Näherungsformel von de Moivre-Laplace.

Die Integrale $\int_{-\infty}^{z} \varphi(x)dx = \Phi(z)$ sind Näherungswerte der Summenwahrscheinlichkeiten $B_{n;p}(X \le k)$.



Folgerungen:

(1) $P(X>k) = 1 - P(X \le k) = 1 - B_{n;p}(X \le k) = 1 - \Phi(z)$

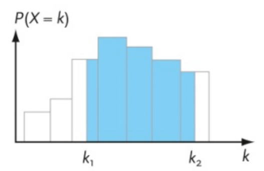


(2) Die Intervallwahrscheinlichkeit $P(k_1 \le X \le k_2)$ ist näherungsweise gleich der Differenz

$$\Phi(z_2) - \Phi(z_1) = \int_{-\infty}^{z_2} \varphi(z)dz - \int_{-\infty}^{z_1} \varphi(z)dz$$

mit $z_1 = \frac{k_1 - \mu}{\sigma}$ und $z_2 = \frac{k_2 - \mu}{\sigma}$

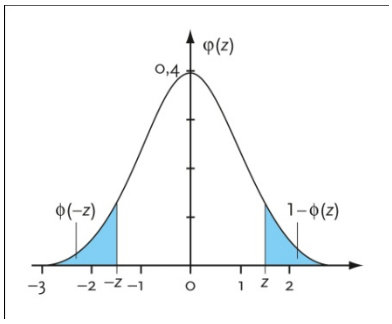
Korrekturglieder:
Da k_1 und k_2 die Mitten der Rechtecksbreiten sind,
rückt man vor der Standardisierung k_1 um 0,5 LE nach links und k_2 um



0,5 LE nach rechts.
Damit hat man einen besseren Näherungswert der Intervallwahrscheinlichkeit.
Für die standardisierten Variablen gilt dann:

$z_1 = \frac{k_1 - 0{,}5 - \mu}{\sigma}$ und $z_2 = \frac{k_2 + 0{,}5 - \mu}{\sigma}$

und für die integrale Näherung der Intervallwahrscheinlichkeit



$$B_{n;p}(k_1 < X < k_2) = \Phi\left(\frac{k_2 + 0{,}5 - \mu}{\sigma}\right) - \Phi\left(\frac{k_1 - 0{,}5 - \mu}{\sigma}\right)$$

Diese Korrektur ist insbesondere dann wichtig, wenn n und σ nicht hinreichend groß sind.

(3) Ist z<0, ist $\Phi(z) < 0{,}5$

(4) Es gilt: $\Phi(-z) < 1 - \Phi(z)$

2. Die Gauß'sche Glockenfunktion

Die Gauß'sche Glockenfunktion φ mit $\varphi(x) = \frac{1}{2\pi} e^{-\frac{1}{2}x^2}$, mit der man die Konturen von Binomialverteilungen beschreiben kann, heißt **Standard-Glockenfunktion**.

Funktionen $\varphi_{\mu;\sigma}$ mit $\varphi_{\mu;\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi\mathrm{F}}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ heißen **Gauß'sche Glockenfunktionen** mit den festen Parametern μ und σ .

Sie besitzen eine Maximalstelle bei $x = \mu$ mit dem Maximalwert $y_{max} = \frac{1}{\sigma\sqrt{2\pi}}$

und zwei Wendestellen bei $x = \mu \pm \sigma$ mit dem Funktionswert $y_W = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}}$

Und es gilt: $\int_{-\infty}^{\infty} \varphi_{\mu;\sigma}(x)dx = 1$

Für $\mu = 0$ und $\sigma = 1$ erhält man die Standard-Glockenfunktion.

Anmerkungen:

(1) Die Gauß'schen Glockenfunktionen lassen sich nur numerisch integrieren.

(2) Verschiedene Schaubilder:

3. Stetige Zufallsvariable und Wahrscheinlichkeitsdichte

Diskrete Zufallsvariable sind Zufallsvariable, deren Werte abzählbar sind und durchnummeriert werden können. Je zwei benachbarte Ergebnisse sind eindeutig unterscheidbar.
Ihre Wahrscheinlichkeiten kann man in Tabellen darstellen.
Beispiel: 100mal würfeln, die Zufallsvariable X gibt die Anzahl der 6en an.

Stetige Zufallsvariable sind Zufallsvariable, deren Werte beliebig dicht liegende reelle Zahlen sind, wobei zwischen zwei beliebig dicht liegenden Werten einer stetigen Zufallsvariablen immer noch unendlich viele Werte liegen.

Beispiel für eine stetige Zufallsvariable:
In einer Zentrifuge befindet sich ein kleines Holzkügelchen, das durch mehrere Öffnungen die Zentrifuge verlassen kann. Die Winkelgeschwindigkeit der Zentrifuge wird innerhalb von 2 Minuten auf einen maximalen Wert hochgefahren. Die Zufallsvariable X gibt an, wie viel Zeit vergeht, bis das Kügelchen innerhalb dieser 2 Minuten die Zentrifuge verlassen hat (wobei die Kugel auf jeden Fall innerhalb von 2 Min die Zentrifuge verlässt.)
Es gibt also unendlich viele Werte für die Zufallsvariable im Intervall (0:2],
alle Zahlen x mit $0<x\leq 2$ sind möglich. Die Zufallsvariable ist stetig.

Eine Funktion f, aus der man Wahrscheinlichkeiten durch Integrieren erhält, nennt man **Wahrscheinlichkeitsdichte.**

Eine Funktion f heißt **Wahrscheinlichkeitsdichte** über einem Intervall I=[a;b] (oder I=(a;b)…), wenn gilt:

(1) $f(x) \geq 0$ für alle $x \in \mathbf{I}$

(2) $\int_a^b f(x)dx = 1$

Anmerkungen:

1. Durch (1) ist gewährleistet, dass die Wahrscheinlichkeiten von Teilintervallen nicht negativ sind.
2. Die Wahrscheinlichkeit des gesamten Intervalls beträgt 1=100%
3. Man nennt f auch **Dichtefunktion.**
4. Eine Zufallsvariable X mit reellen Werten im Intervall I heißt **stetig verteilt**, wenn gilt:

$$P(k_1 \leq X \leq k_2) = \int_{k_1}^{k_2} f(x)dx \qquad \text{für alle } k_1, k_2 \in \mathrm{IR}$$

5. Die Funktionswerte f(x) sind keine Wahrscheinlichkeiten.
   Denn die Wahrscheinlichkeit, dass die Zufallsvariable genau den Wert k annimmt, berechnet sich **durch** $P(X = k) = \int_k^k f(x)dx = 0$,d.h. die Einzelwahrscheinlichkeiten sind exakt null.

4. Normalverteilte Zufallsvariablen

Eine stetige Zufallsvariable, die durch ihren Erwartungswert μ und ihre Standardabweichung σ festgelegt ist und nach der Standardisierung $z = \frac{k-\mu}{\sigma}$ als Dichtefunktion die Gauß'sche Glockenfunktion $\varphi_{\mu;\sigma}$ besitzt,

heißt **normalverteilte Zufallsvariable mit den Parametern μ und σ.**
Man sagt: **X ist $N(\mu;\sigma)$ – verteilt** .

Wahrscheinlichkeiten für Bereiche lassen sich durch Integration der Dichtefunktion bestimmen: $\Phi_{\mu;\sigma}(z) = \int_{-\infty}^{z} \varphi_{\mu;\sigma}(x)dx$

Speziell:

Die Normalverteilung für $\mu = 0$ und $\sigma = 1$ (mit der Standard-Glockenfunktion φ mit $\varphi(x) = \frac{1}{2\pi} e^{-\frac{1}{2}x^2}$) heißt **Standardnormalverteilung N(0;1).**
Die Wahrscheinlichkeit $P(X \le k)$ einer standardnormalverteilten (stetigen) Zufallsvariablen X berechnet sich durch: $P(X \le k) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx = \Phi(z)$ mit $z = \frac{k-\mu}{\sigma}$

Anmerkungen:

1. Die Normalverteilung ist eine spezielle stetige Verteilung.
2. Einsatz des GTR:

   Mit normalcdf(a,b, μ,σ)erhält man den Wert für $\int_{a}^{b} \varphi_{\mu;\sigma}(x)dx$

   Mit normalcdf($-\infty, x, \mu, \sigma$) ist die zugehörige Integralfunktion $\Phi_{\mu;\sigma}(x) = \int_{-\infty}^{x} \varphi_{\mu;\sigma}(t)dt$ gegeben, man kann Funktionswerte an jeder Stelle x=x_0 bestimmen.
   (Bei manchen TR muss man statt $-\infty$ einen Näherungswert (z.B. -100) einsetzen.)
   Mit invNorm(Y, μ,σ) kann man die Stelle x bestimmen, an der die Integralfunktion den Wert Y annimmt: $Y = \Phi_{\mu;\sigma}(x) = \int_{-\infty}^{x} \varphi_{\mu;\sigma}(t)dt$

   (Bsp.: invNorm(0.75,7.8,1.5) ergibt ungefähr den Wert 8.81, d.h. an der Stelle x=8,81 hat die Integralfunktion $\Phi_{7,8;1,5}$ den Wert 0,75.

3. Bei der Normalverteilung ist μ der Erwartungswert und σ die Standardabweichung.
   Begr. LS S.369

4. Es gilt $P(a \leq X \leq b) = P(a < X < b)$

5. Für jede normalverteilte Zufallsvariable gelten die Sigma-Regeln:

   $$P(\mu-\sigma \le X \le \mu+\sigma) = \Phi(1)-\Phi(-1) \approx 68,3\%$$

   (Begründung:

   $$P(\mu-\sigma \le X \le \mu+\sigma) = \int_{\mu-\sigma}^{\mu+\sigma} \varphi_{\mu;\sigma}(x)dx = \Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right) = \Phi(1)-\Phi(-1) \approx 68,3\%)$$

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

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

   bzw.

   $$P(\mu-1,64\sigma \le X \le \mu+1,64\sigma) = 90\%$$

   $$P(\mu-1,96\sigma \le X \le \mu+1,96\sigma) = 95\%$$

   $$P(\mu-2,58\sigma \le X \le \mu+2,58\sigma) = 99\%$$

6. Binomialverteilung und Normalverteilung

Die Gauß'schen Glockenfunktionen sind einerseits Wahrscheinlichkeitsdichten stetiger Zufallsvariablen.

Andererseits beschreiben sie die Kontur von Binomialverteilungen unter bestimmten Bedingungen:

Für binomialverteilte Zufallsvariable X mit $\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ mit $\sigma > 3$ gelten die Näherungen:

(1) $P(X=k) = B_{n;p}(k) \approx \varphi_{\mu;\sigma}(k)$ (lokale Näherungsformel)

(2) $P(a \le X \le b) = \int_{a-0,5}^{b+0,5} \varphi_{\mu;\sigma}(x)dx$ (integrale Näherungsformel mit Stetigkeitskorrektur)

Diese Annäherung der Binomialverteilung durch die Normalverteilung begründet die Sigma-Regel für binomialverteilte Zufallsvariable:

$$P(\mu-\sigma \le X \le \mu+\sigma) \approx \int_{\mu-\sigma-0,5}^{\mu+\sigma+0,5} \varphi_{\mu;\sigma}(x)dx \approx \int_{\mu-\sigma}^{\mu+\sigma} \varphi_{\mu;\sigma}(x)dx \approx 0,68$$

Weitere spezielle Wahrscheinlichkeitsverteilungen:

Die Poisson-Verteilung

Geometrische Verteilung

Hypergeometrische Verteilung