11. Approximation der Binomialverteilung durch die Normalverteilung

Die Berechnung der Binomialverteilung ist wegen der Binomialkoeffizienten nicht unproblematisch. Man kann sie deshalb in gewissen Fällen durch Näherungsverteilungen z. B. die Normalverteilung approximieren. Dies ist eine Folge des Zentralen Grenzwertsatzes.

Die Idee wird am folgenden Beispiel erläutert.

Eine Laplacemünze wird zwanzigmal geworfen. Die Wahrscheinlichkeit in n = 20 Würfen genau x-mal Kopf zu werfen ist:

$$P_{20}(x) = \binom{20}{x} \cdot \left(\frac{1}{2}\right)^{20}$$

Die Tabelle der Binomialverteilung liefert die folgenden Wahrscheinlichkeiten:

x_i	0	1	2	3	4	5	6	7	8	9	10
$P_{20}(x_i)$	0.0	0.0	0.0	0.1	0.5	1.5	3.7	7.4	12.0	16.0	17.6

Die Wahrscheinlichkeit in 20 Würfen mindestens 8 und höchstens 13 mal Kopf zu werfen kann als Inhalt des grün markierten Treppenflächenstücks interpretiert werden. In der Skizze auf der nächsten Seite ist zusätzlich der Graph der Normalverteilung

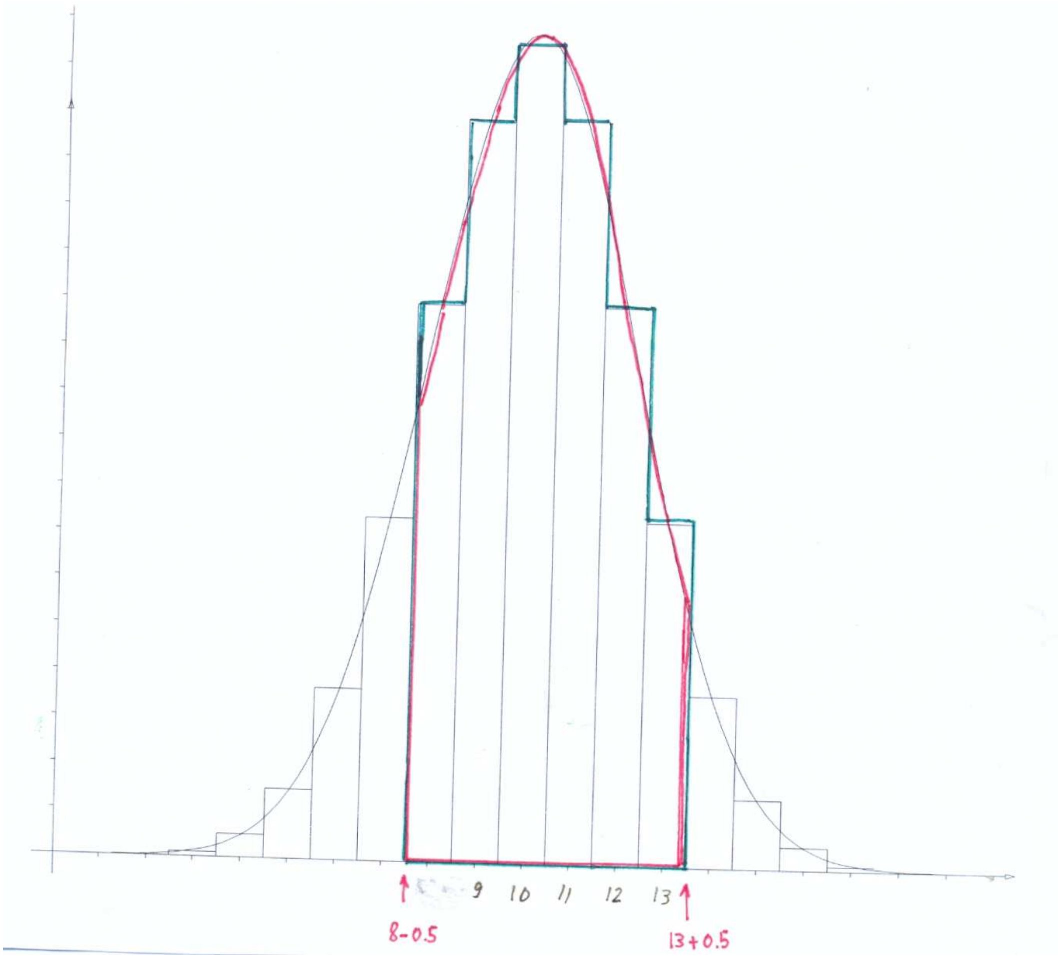
$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ gezeichnet.

Die Werte der Parameter μ bzw. σ ergeben sich aus den entsprechenden Werten für die Binomialverteilung zu $\mu = np = 20 \cdot \frac{1}{2} = 10$ und $\sigma^2 = npq = 5$ bzw. $\sigma = \sqrt{5}$. Offenbar kann die gesuchte Wahrscheinlichkeit gut durch den Inhalt unter der Kurve der Normalverteilung

$$\varphi(x) = \frac{1}{\sqrt{10\pi}} e^{-\frac{(x-10)^2}{10}}$$

im Intervall [8 - 0.5 , 13 + 0.5] (rot eingefärbt) angenähert werden kann. Näherungsweise gilt also:

$$P_{20}(8 \le X \le 13) \approx \int_{7.5}^{13.5} \varphi(x)dx$$



Die Abweichungen der obern bzw. untern Grenze vom Erwartungswert in σ- Einheiten ergeben sich zu

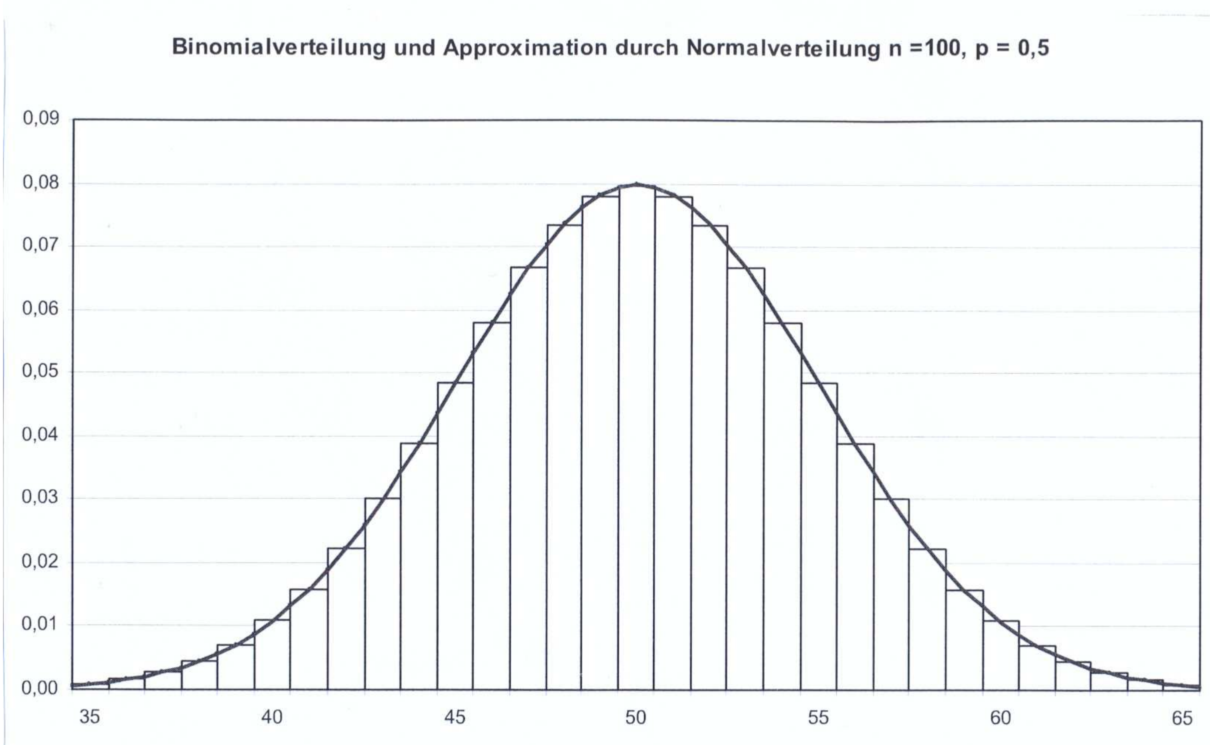
$$\frac{3.5}{\sqrt{5}} \approx 1.565 \text{ bzw. } \frac{-2.5}{\sqrt{5}} \approx -1.118$$

$$p(8 \le X \le 13) \approx \Phi(1.565) - \Phi(-1.118) = \Phi(1.565) - (1 - \Phi(1.118))$$

≈ 0.9412 - (1 - 0.8682) = **0.8094**

Aus der Tabelle für die Binomialverteilung erhält man 0.942 - 0.132 = **0.810**.

Ein weiteres Beispiel zur Approximation der Binomialverteilung durch die Normalverteilung für n = 100 und p = 0.5 dagestellt



Allgemein gilt:

Ist $n > \dfrac{9}{pq}$, dann kann die Binomialverteilung durch die Normalverteilung approximiert werden und es gilt:

$$p(a \le X \le b) = \sum_{x=a}^{b} P_n(x) \approx \Phi\left(\frac{b - np + 0.5}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - 0.5}{\sqrt{npq}}\right) \qquad n > \frac{9}{pq}$$

Bem.:
Obschon im Beispiel n = 20 und p = 0.5 die Faustregel verletzt ist, ist die Näherung sehr gut. Bei kleinen Erfolgswahrscheinlichkeiten verwendet man die sogenannte Poissonverteilung als Näherungsverteilung.

Die Näherung kann auch für die Berechnung eines einzelnen Werts der Binomialverteilung verwendet werden.

Aufgabe:
Wie gross ist die Wahrscheinlichkeit in 120 Würfen mit einem Laplacewürfel genau 21 Vierer zu werfen?

a) Wert der Binomialverteilung: $p(X=21)=\binom{120}{21}\cdot(\tfrac{1}{6})^{21}\cdot(\tfrac{5}{6})^{99}\approx 0.0927$

b) Aprroximation durch die Normalverteilung:

$\mu = np = 120\cdot\tfrac{1}{6} = 20$ $\sigma^2 = npq = 120\cdot\tfrac{1}{6}\cdot\tfrac{5}{6} = \tfrac{100}{6}$

$$p(X=21)\approx\Phi\left(\frac{21.5-20}{4.0825}\right)-\Phi\left(\frac{20.5-20}{4.0825}\right)=\Phi(0.367)-\Phi(0.122)=0.6432-0.5486=0.0946$$

Aufgabe:
40% der Stimmberechtigten haben bei der letzten Wahl die Partei Fiat Justitia gewählt. Wie gross ist die Wahrscheinlichkeit, dass von 600 zufällig ausgewählten Stimmberechtigten
a) höchstens 260 b) mehr als 260 c) weniger als 220 d) zwischen 220 und 260 diese Partei wählen.

n = 600 $\mu = np = 600\cdot 0.6 = 240$ und $\sigma^2 = npq = 240\cdot 0.6 = 144$ bzw. $\sigma = 12$

a) p_1 = $p(X\le 260)\approx\Phi\left(\frac{260.5-240}{12}\right)\approx\Phi(1.708)\approx 0.9555$

b) und c) $p_2 = 1-p_1$ = 0.0455

d) $p_3 = 1-2p_2$ = 0.9110

Aufgabe:
Wir nehmen an, es sei bekannt, dass bei einer bestimmten Vorlage 1050 von 491050 Stimmberechtigten „ja“ stimmen, während die übrigen 490000 sich mit einer Wahrscheinlichkeit von 50% für „ja“ entscheiden. Wie gross ist die Wahrscheinlichkeit, dass die Vorlage angenommen wird?

Lösung:
Sei X die Anzahl der zufälligen „ja“-Stimmenden. Bei einem absoluten Mehr von 245‘526 werden für eine Annahme der Vorlage noch mindestens 245‘526 – 1050 = 244‘476 zufällig „ja“-Stimmende benötigt, d.h. die Zufallsvariable X muss mindestens den Wert 244‘476 erreichen.

$\mu = np = 490000\cdot\tfrac{1}{2} = 245000$ und $\sigma^2 = npq = 245000\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = 122500 = 350^2$ bzw. $\sigma = 350$

$$p(X\ge 244476)=p(X\le 490000-244476)=\Phi\left(\frac{245524-245000}{350}\right)\approx\Phi(1.497)=0.9328$$

Uebungsaufgabe
Bei einer Versicherung sind 2000 Agenten beschäftigt, die unabhängig voneinander 75% ihrer Arbeitszeit im Aussendienst verbringen
a)
Mit welcher Wahrscheinlichkeit sind höchstens 483 Agenten im Büro?
b)
Wie viele Schreibtische müssen angeschafft werden, damit jeder im Büro Anwesende in 95% der Fälle einen Schreibtisch vorfindet?
Lösung:
a)
$\mu = np = 2000 \cdot \frac{1}{4} = 500 \qquad \sigma^2 = npq = 2000 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{1500}{4} = 375$

$$p(X \le 483) \approx \Phi\left(\frac{483.5 - 500}{\sqrt{375}}\right) = \Phi(-0.852) = 0.1971$$

b)
$\Phi(u) = 0.95 \qquad$ u = 1.645 $\qquad$ x = 500 + 1.645·≈ 532

Aufgabe (*):
Wir nehmen an, dass Hoteldirektoren davon ausgehen, dass ein angemeldeter Gast mit der Wahrscheinlichkeit 0.8 auch wirklich eintrifft.
Im Hotel A mit 84 Betten werden 100 Anmeldungen entgegen genommen.
Wie viele Anmeldungen dürfen im Hotel B mit 168 Betten entgegen genommen werden, wenn die Wahrscheinlichkeit einen Gast abweisen zu müssen in beiden Hotels übereinstimmen soll?
Hotel A:
n = 100 $\quad \mu = np = 100 \cdot 0.8 = 80$ und $\sigma^2 = npq = n \cdot 0.8 \cdot 0.2 = 0.16 \cdot n$ bzw. $\sigma = 4$
X: Anzahl der in einer bestimmten Nacht eintreffenden Gäste. Die Wahrscheinlichkeit, dass Gäste abgewiesen werden müssen beträgt:

$$p(X \ge 85) = 1 - p(X \le 84) = 1 - \Phi\left(\frac{84.5 - 80}{4}\right) \approx 1 - \Phi(1.125) = 1 - 0.8697 = 0.1303 \quad (1)$$

Hotel B:
n: Anzahl Anmeldungen. Für den von n abhängigen Erwartungswert der ankommenden Gäste gilt dann:
$\mu = np = 0.8n$ und $\sigma^2 = npq = n \cdot 0.8 \cdot 0.2 = 0.16 \cdot n$ bzw. $\sigma = 0.4\sqrt{n}$

Y: Anzahl der in einer bestimmten Nacht im Hotel eintreffenden Gäste.
Die Wahrscheinlichkeit, dass Gäste abgewiesen werden müssen beträgt:

$$p(Y \ge 169) = 1 - p(Y \le 168) = 1 - \Phi\left(\frac{168.5 - 0.8n}{0.4 \cdot \sqrt{n}}\right)$$

Da die beiden Wahrscheinlichkeiten übereinstimmen müssen gilt wegen (1):

$$\Phi\left(\frac{84.5 - 80}{4}\right) = \Phi(1.125) = \Phi\left(\frac{168.5 - 0.8n}{0.4\sqrt{n}}\right)$$

Damit muss gelten:

$$\frac{168.5 - 0.8n}{0.4\sqrt{n}} = 1.125$$

Dies führt mit der Substitution $z = \sqrt{n}$ auf eine quadratische Gleichung und schliesslich auf die Lösung n = 202.6. Hotel B kann damit höchstens 203 Anmeldungen entgegen nehmen.

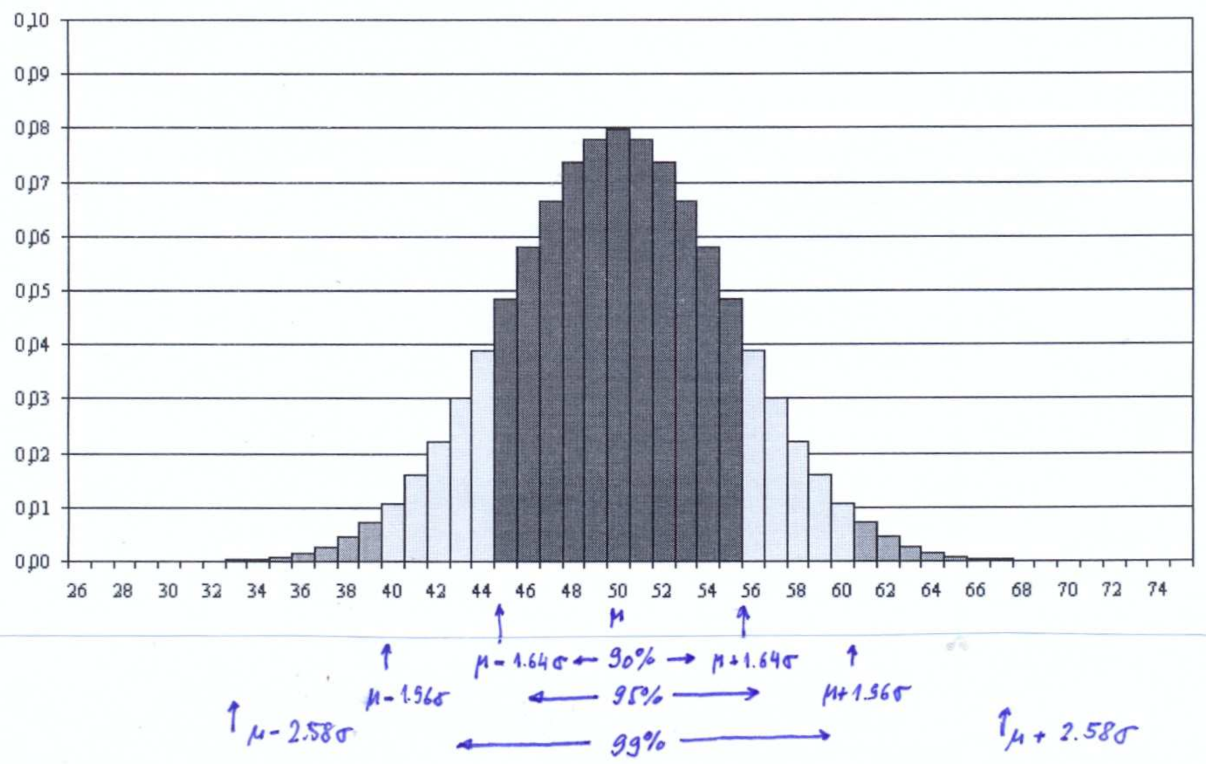
Entsprechend den Regeln für die Normalverteilung gelten für die Binomialverteilung die folgenden Regeln (sofern $n > \frac{9}{pq}$)

Intervall	Wahrscheinlichkeit	
$[\mu - 1.64\sigma, \mu + 1.64\sigma]$	ca. 90%	$\Phi^*(1.64) \approx 0.9$
$[\mu - 1.96\sigma, \mu + 1.96\sigma]$	ca. 95%	$\Phi^*(1.96) \approx 0.95$
$[\mu - 2.58\sigma, \mu + 2.58\sigma]$	ca. 99%	$\Phi^*(2.58) \approx 0.99$

Beispiel:
Wirft man eine Laplacemünze 100 mal dann liegt das Ereignis „Anzahl Kopf“ mit den angegebenen Wahrscheinlichkeiten in dem angegebenen Intervall.
Wegen n = 100, p = 0.5

$\mu = np = 100 \cdot \frac{1}{2} = 50$ $\qquad \sigma^2 = npq = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$ $\qquad \sigma = 5$

$[\mu - 1.64\sigma, \mu + 1.64\sigma] = [42,58]$	ca. 90%	$\Phi^*(1.64) \approx 0.9$
$[\mu - 1.96\sigma, \mu + 1.96\sigma] = [40,60]$	ca 95%	$\Phi^*(1.96) \approx 0.95$
$[\mu - 2.58\sigma, \mu + 2.58\sigma] = [37,63]$	ca. 99%	$\Phi^*(2.58) \approx 0.99$



Beispiel:
Wird ein Laplacewürfel 1200 mal geworfen, dann werden mit einer Wahrscheinlichkeit von etwa 90% mindestens 179 und höchstens 221 Würfe mit der Augenzahl 1 vorkommen.
Begründung:
$\mu = np = 1200 \cdot \frac{1}{6} = 200 \qquad \sigma^2 = npq = 1200 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{1000}{6}$
Intervall: $[\mu - 1.64\sigma, \mu + 1.64\sigma]$ = $[178, 221]$

Uebungsaufgabe:
Wie oft muss ein Laplacewürfel geworfen werden, wenn die relative Häufigkeit der geworfenen Sechser mit 97% Sicherheit im Intervall $[\frac{1}{6} - 0.01, \frac{1}{6} + 0.01]$ liegen soll?

$\mu = \frac{1}{6} \cdot n \qquad \sigma = \frac{1}{6} \cdot \frac{5}{6} \cdot n = \frac{5}{36} \cdot n \qquad \Phi^*(u) = 0.97 \qquad \Phi(u) = 0.985$ u = 2.17

$$\frac{2.17 \cdot \sigma}{n} = \frac{2.17 \cdot \frac{1}{6}\sqrt{5n}}{n} = \frac{2.17 \cdot \sqrt{5}}{6 \cdot \sqrt{n}} = \frac{1}{100} \qquad n = \left(\frac{217 \cdot \sqrt{5}}{6}\right)^2 \approx 6540.1$$