3.3. Spezielle Verteilungen für mehrstufige Experimente

Baumdiagramme für mehrstufige Experimente werden bei vielen Wiederholungen schnell unübersichtlich. Für das wiederholte Ziehen mit und ohne Zurücklegen aus einer Urne, die nur zwei Sorten von Kugeln enthält, lassen sich aber Verteilungsformeln angeben, die das Baumdiagramm ersetzen.
Beispiel:
Wie groß ist die Wahrscheinlichkeit, beim dreimaligen (zehnmaligen) Würfeln einmal (zweimal) die sechs zu würfeln?

3.3.1. Kombinatorik

Kombinatorische Formeln dienen dazu, die Zahl der Ergebnisse (= Kombinationen der Ergebnisse der einzelnen Stufen) bei mehrstufigen Experimenten zu berechnen.

Beispiel: Aufgaben zur hypergeometrischen Verteilung: Nr. 1 und 2

Allgemeines Zählprinzip bei mehrstufigen Experimenten:
Besteht ein Zufallsexperiment aus k voneinander unabhängigen Stufen mit jeweils n_1, n_2, ..., n_k Ergebnissen, so hat das Gesamtexperiment $n_1 \cdot n_2 \cdot ... \cdot n_k$ kombinierte Ergebnisse.
Spezialfälle:
1. Aus einer Urne mit n Kugeln wird k mal **mit** Zurücklegen gezogen.
   Dann gibt es $n \cdot ... \cdot n = n^k$ verschiedenen Ergebnisse.
2. Aus einer Urne mit n Kugeln wird k mal **ohne** Zurücklegen gezogen.
   Dann gibt es $n \cdot (n-1) \cdot ... \cdot (n-k+1)$ verschiedenen Ergebnisse.
3. Aus einer Urne mit n Kugeln werden **alle n** Kugeln **ohne** Zurücklegen gezogen.
   Dann gibt es $n \cdot (n-1) \cdot ... \cdot 2 \cdot 1 = n!$ (lies n **Fakultät**) verschiedenen Ergebnisse.

Übungen: Aufgaben zur hypergeometrischen Verteilung: Nr. 3 - 6

Beispiel: Aufgaben zur hypergeometrischen Verteilung: Nr. 7 a) und b)

Satz über die Zahl der Kombinationen/Teilmengen
Aus einer Urne mit n Kugeln wird k mal **ohne** Zurücklegen gezogen. Jeweils alle k! Ergebnisse mit gleicher Auswahl aber unterschiedlicher Reihenfolge der k Kugeln werden als **Kombination** zusammengefasst
Dann gibt es $\frac{n \cdot (n-1) \cdot ... \cdot (n-k+1)}{k!} = \frac{n!}{(n-k)! \cdot k!} = \binom{n}{k}$ mögliche **Kombinationen**.
Die Zahlen $\binom{n}{k}$ (lies „k aus n“) heißen **Binomialkoeffizienten**.
Alternative Formulierung:
Es gibt $\binom{n}{k}$ Möglichkeiten, aus n verschiedenen Objekten k Objekte auszuwählen.

Übungen: Aufgaben zur hypergeometrischen Verteilung: Nr. 7 c) - g)

3.3.2. Das Ziehen ohne Zurücklegen und die hypergeometrische Verteilung

Beispiel: Aufgaben zur hypergeometrischen Verteilung: Nr. 8

Satz über da Ziehen ohne Zurücklegen (Hypergeometrische Verteilung)
Eine Urne enthält **N** Kugeln, davon **r** rote und **N − r** schwarze. Es werden **n** Kugeln **ohne** Zurücklegen gezogen.
Die Wahrscheinlichkeit, genau **k** rote Kugeln zu ziehen, ist dann

$$H_{n,\, r/N}(k) = \binom{n}{k} \cdot \underbrace{\frac{r}{N} \cdot \frac{r-1}{N-1} \cdot ... \cdot \frac{r-k}{N-k}}_{\text{k mal rot}} \cdot \underbrace{\frac{N-r}{N-k-1} \cdot \frac{N-r-1}{N-k-2} \cdot ... \cdot \frac{N-r-n+k}{N-n}}_{\text{n − k mal schwarz}}$$

Übungen: Aufgaben zur hypergeometrischen Verteilung: Nr. 9 - 11

3.3.3. Das Ziehen mit Zurücklegen und die Binomialverteilung

Beispiel: Aufgaben zur Binomialverteilung: Nr. 1

Satz über das Ziehen mit Zurücklegen (Binomialverteilung)
Eine Urne enthält **N** Kugeln, davon **r** rote und **N − r** schwarze. Es werden **n** Kugeln **mit** Zurücklegen gezogen.
Die Wahrscheinlichkeit, genau **k** rote Kugeln zu ziehen, ist dann

$$B_{n,\, r/N}(k) = \binom{n}{k} \cdot \left(\frac{r}{N}\right)^k \cdot \left(\frac{N-r}{N}\right)^{n-k}$$

$$= \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

mit $p = \frac{r}{N}$ = Wahrscheinlichkeit, bei **einer** Ziehung eine rote Kugel zu ziehen

$q = 1 - p$ = Wahrscheinlichkeit, bei **einer** Ziehung eine schwarze Kugel zu ziehen

Eingabe im GTR:
z.B. $B_{200\;0,3}(5 \leq X \leq 15)$ = sum(seq(200 nCr X*0,3^X*0,7^(200−X),X,5,15)) mit
sum = LIST/MATH/5, seq = LIST/OPS/5 und nCr = MATH/PRB/3

Übungen: Aufgaben zur Binomialverteilung: Nr. 2 - 10

3.3.4. Näherung der hypergeometrischen Verteilung durch die Binomialverteilung

Beispiel:
Berechne die Wahrscheinlichkeit für die Ziehung zweier roter Kugeln **mit und ohne Zurücklegen** aus einer Urne mit
a) N = 5 Kugeln, davon r = 3 rote
b) N = 50 Kugeln, davon r = 30 rote

Lösung

a) Bei der ersten Ziehung ist $p = \frac{r}{N} = \frac{3}{5} = 0{,}6$. Bei der zweiten Ziehung ist p entweder nur noch $\frac{2}{4} = 0{,}5$ (falls zuvor eine rote Kugel gezogen wurde) oder $\frac{3}{4} = 0{,}75$ (falls zuvor eine schwarze Kugel gezogen wurde).

b) Bei der ersten Ziehung ist $p = \frac{r}{N} = \frac{30}{50} = 0{,}6$. Bei der zweiten Ziehung ist p entweder $\frac{29}{49} = 0{,}59$ (falls zuvor eine rote Kugel gezogen wurde) oder $\frac{30}{49} = 0{,}61$ (falls zuvor eine schwarze Kugel gezogen wurde).

Man kann also unabhängig von dem Ergebnis der ersten Ziehung auch für die zweite Ziehung $p \approx 0{,}6$ als Näherungswert verwenden. Wenn für beide Ziehungen die gleichen Wahrscheinlichkeiten gelten, entspricht dies aber dem Ziehen mit Zurücklegen und führt auf die Binomialverteilung.

Näherung der hypergeometrischen Verteilung durch die Binomialverteilung
Bei vielen Wiederholungen ($N \geq 20$) lässt sich auch das Ziehen **ohne** Zurücklegen (d.h. die hypergeometrische Verteilung) durch die Binomialverteilung annähern. Befinden sich nämlich sehr viele Kugeln in der Urne, so werden die Mengenverhältnisse bzw. Wahrscheinlichkeiten durch die Entnahme weniger Kugeln kaum geändert. Man kann also bei der 2., 3., usw. Kugel näherungsweise die gleichen Wahrscheinlichkeiten verwenden wie bei der ersten Kugel

3.3.5. Erwartungswert und Varianz

Beispiel: Aufgaben zur Binomialverteilung Nr. 11

Wahrscheinlichkeitsfunktion und Verteilungsfunktion einer Zufallvariablen X
Gegeben sei eine **Zufallsvariable X** ,die jedem Ergebnis ω_1, ω_2, ... eine reelle Zahl $X(\omega_1)$, $X(\omega_2)$, .. zuordnet.

Die **Wahrscheinlichkeitsfunktion f** gibt für jedes $k \in \mathbb{R}$ die zugehörige Wahrscheinlichkeit $f(k) = P(X = k)$ an.

Die **Verteilungsfunktion F** gibt für jedes $k \in \mathbb{R}$ die zugehörige Wahrscheinlichkeit $F(k) = P(X \leq k)$ an.

Übungen: Aufgaben zur Binomialverteilung Nr. 12 und 13

Beispiel: Aufgaben zur Binomialverteilung Nr. 14

Erwartungswert, Varianz und Standardabweichung
Der **Erwartungswert** $\mu = E(X)$ einer Zufallsvariablen X ist der mit den entsprechenden Wahrscheinlichkeiten $P(\omega_i)$ gewichtete **Mittelwert** aller möglichen Werte $X(\omega_i)$: $E(X) = X(\omega_1)\cdot P(\omega_1) + X(\omega_2)\cdot P(\omega_2) + ...$
Die **Varianz** $V(X) = E((X - \mu)^2)$ ist ein Maß für die **mittlere quadratische Abweichung vom Mittelwert.**
Die **Standardabweichung** $\sigma = \sqrt{V(X)}$ ist ein Maß für die **mittlere Abweichung vom Mittelwert**

Übungen: Aufgaben zur Binomialverteilung Nr. 15 und 16

Erwartungswert und Standardabweichung der Binomialverteilung
Ist eine Zufallsvariable binomialverteilt mit $P(X = k) = B_{n\,p}(k)$, so gilt:
Erwartungswert μ $= n\cdot p$
= Wert mit der größten Wahrscheinlichkeit
$\Rightarrow$ **Maximum** der Wahrscheinlichkeitsfunktion bei $k = \mu$
Standardabweichung σ $= \sqrt{n \cdot p \cdot q}$
= mittlere **Streuung** um den Erwartungswert
$\Rightarrow$ **Wendepunkte** der Wahrscheinlichkeitsfunktion bei $k = \mu \pm \sigma$

Beweis:
Der Beweis vereinfacht sich stark durch die Verwendung der **Linearität** des Erwartungswertes: Für zwei Zufallsvariablen X und Y sowie reelle Zahlen a und b gilt nämlich
$E(aX + bY) = (aX(\omega_1) + bY(\omega_1))\cdot P(\omega_1) + \ldots = a\cdot(X(\omega_1)\cdot P(\omega_1) + \ldots) + b\cdot(Y(\omega_1)\cdot P(\omega_1) + \ldots) = a\cdot E(X) + b\cdot E(Y)$.

Sei $X_i = 1$, wenn die i-te Kugel rot ist und $X_i = 0$, wenn sie schwarz ist. Der Erwartungswert für diese Zufallsvariable ist dann $E(X_i) = X_i(r)\cdot P(r) + X_i(s)\cdot P(s) = 1\cdot p + 0\cdot(1 - p) = p$.
Außerdem gilt $X = X_1 + X_2 + \ldots X_n$ und daher entsprechend $\mathbf{E(X)} = E(X_1) + \ldots + E(X_n) = \mathbf{n\cdot p}$.

Weiter gilt

$$
\begin{aligned}
E(X^2) &= E((X_1 + \ldots + X_n)^2) \\
&= \underbrace{X_1(r)\cdot X_1(r)\cdot P(r) + \ldots \quad + X_n(r)\cdot X_n(r)\cdot P(r,r)}_{\text{n Summanden p}} + \left.\begin{array}{l} \underbrace{X_1(r)\cdot X_2(r)\cdot P(r,r) + \ldots + X_1(r)\cdot X_n(r)\, P(r,r)}_{\text{n – 1 Summanden } p^2} \\ + \underbrace{X_2(r)\cdot X_1(r)\cdot P(r,r) + \ldots + X_2(r)\cdot X_n(r)\, P(r,r)}_{\text{n – 1 Summanden } p^2} \\ \quad\vdots \\ + \underbrace{X_n(r)\cdot X_1(r)\cdot P(r,r) + \ldots + X_n(r)\cdot X_{n-1}(r)\, P(r,r)}_{\text{n – 1 Summanden } p^2} \end{array}\right\} \text{n Summen} \\
&= n\cdot p + n(n - 1)\cdot p^2
\end{aligned}
$$

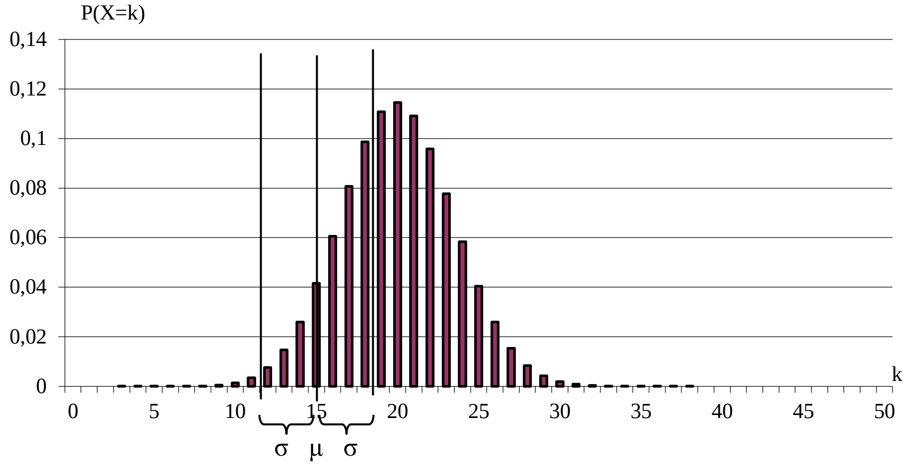
und damit

$\mathbf{V(X)} = E((X - \mu)^2) = E((X - np)^2) = E(X^2 - 2npX + (np)^2) = E(X^2) - 2np(E(X) + (np)^2 = n\cdot p + n(n - 1)\cdot p^2 - 2(np)^2 + (np)^2 = np - np^2 = np(1 - p) = \mathbf{n\cdot p\cdot q}$.

Beispiel:

$B_{50\ 0,4}(k)$ mit $\mu = 20$ und $\sigma = 3{,}46$

$\Rightarrow$ **Maximum** bei $k = 20$ mit $B_{50\ 0,4}(20) = 11{,}45$ %
Wendepunkte bei $k \approx 17$ mit $B_{50\ 0,4}(17) = 8{,}07$ %
und $k \approx 23$ mit $B_{50\ 0,4}(23) = 7{,}78$ %:



Übungen: Aufgaben zur Binomialverteilung Nr. 17 und 18

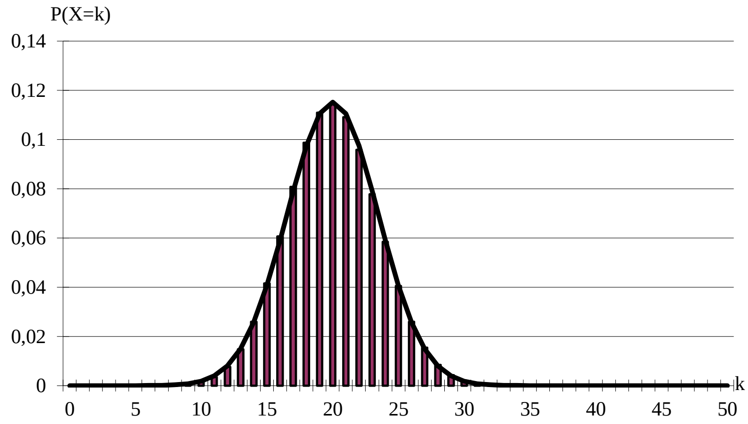
3.3.6. Näherung der Binomialverteilung durch die Normalverteilung

Beispiel: Aufgaben zur Normalverteilung und Hypothesentests Nr. 1 und 2

Lokaler Grenzwertsatz von de Moivre und Laplace
Für große n lässt sich die Binomialverteilung durch eine geeignet verschobene und gestreckte **Normalverteilung**
$\Phi(k) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}k^2}$ annähern: $\lim\limits_{n\to\infty} B_{n\,p}(k) = \frac{1}{\sigma}\phi\left(\frac{k-\mu}{\sigma}\right)$.
Faustregel: Die Näherung ist ausreichend genau, wenn $n \cdot p \cdot q > 9$ ist.
EInstellungen im GTR: $X_{MIN} = \mu - \sigma^2$, $X_{MAX} = \mu + \sigma^2$, $Y_{MIN} = 0$, $Y_{MAX} = P(X = \mu)$

Beispiel:

$B_{50\ 0,4}(k)$ mit $\mu = 20$ und $\sigma = 3{,}46 \Rightarrow B_{50\ 0,4}(k) \approx \frac{1}{3,4641}\phi\left(\frac{k-20}{3,4641}\right)$



Übungen: Aufgaben zur Normalverteilung und Hypothesentests Nr. 3 - 5

3.3.7. Hypothesentests

Beispiel: Aufgaben zur Normalverteilung und Hypothesentests Nr. 6

Hypothesentest

1. Möchte man Aussagen über Wahrscheinlichkeiten bei z.B. Kunden (Kaufentscheidung), Wählern (Wahlentscheidung), Produkten (Ausschussanteil) oder Medikamenten (Wirkungsanteil) gewinnen, so muss man sich auf die Untersuchung einer **Stichprobe** oder **Statistik** von begrenztem **Umfang n** beschränken.
2. Man geht von der **Hypothese** aus, dass das zu untersuchende Ereignis mit der Wahrscheinlichkeit p_1 eintritt. Häufig wird eine **Gegenhypothese** oder **Alternative** mit der Wahrscheinlichkeit p_2 ebenfalls untersucht.
3. Die Aufgabe des Statistikers besteht darin, eine vernünftige **Entscheidungsgrenze k** festzulegen: Tritt das Ereignis in weniger oder mehr als k von n Fällen ein, so wird die Hypothese angenommen, ansonsten wird sie abgelehnt bzw. die Alternative angenommen.
4. Dabei können zwei Arten von **Fehlentscheidungen** auftreten:
   Fehler 1. Art: Die Hypothese wird **abgelehnt**, obwohl sie **erfüllt** ist.
   Risiko 1. Art (Signifikanzniveau, Irrtumswahrscheinlichkeit) α = Wahrscheinlichkeit für Fehler 1. Art
   Fehler 2. Art: Die Hypothese wird **angenommen**, obwohl sie **nicht erfüllt** ist.
   Risiko 2. Art β = Wahrscheinlichkeit für Fehler 2. Art

Übungen: Aufgaben zur Normalverteilung und Hypothesentests Nr. 7 - 11