3.3. Verteilungsfunktion und Erwartungswert der hypergeometrischen Verteilung

Musterbeispiel
Eine Urne enthält N = 5 Kugeln, davon r = 3 rote. Es werden n = 4 Kugeln ohne Zurücklegen gezogen. Gesucht ist die Wahrscheinlichkeit, dass genau k = 2 rote dabei sind.

Um auf die Zahl der möglichen Kombinationen zu kommen, fasst man jeweils alle Ereignisse zusammen, die die gleiche Auswahl aufweisen und sich nur in der Anordnung unterscheiden. Z.B. kann die Auswahl 1234 in 4! verschiedenen Anordnungen auftreten. Ebenso tritt aber auch die Auswahl 1235, die Auswahl 1345 und jede andere Auswahl genau 4! mal auf. Die insgesamt $5\cdot 4\cdot 3\cdot 2$ Ereignisse lassen sich also in Ergebnisgruppen von jeweils 4! Ereignissen aufteilen, die in der Auswahl übereinstimmen und sich nur in der Anordnung unterscheiden. Damit erhält man $\frac{5\cdot 4\cdot 3\cdot 2}{4!} = 5$ verschiedene Ergebnisse bzw. Kombinationen für die Ziehung von n = 4 Kugeln aus einer Urne mit N = 5 verschiedenen Kugeln. Entsprechend gilt: Bei Ziehung von n = 3 bzw. 2 bzw. 1 Kugeln aus einer Urne mit N = 5 verschiedenen Kugeln treten $\frac{5\cdot 4\cdot 3}{3!} = 20$ bzw. $\frac{5\cdot 4}{2!} = 10$ bzw. $\frac{5}{1} = 5$ Kombinationen auf.

Im Musterbeispiel gibt es also $\binom{N}{n} = \binom{5}{4} = 5$ **mögliche** Kombinationen. Für die Berechnung der Laplace-Wahrscheinlichkeit benötigt man noch die Zahl der **günstigen** Kombinationen. Eine Kombination ist günstig, wenn k = 2 der r = 3 roten Kugeln ($\binom{r}{k} = \binom{3}{2} = 3$ Teilkombinationen) **und** n − k = 2 der N − r = 2 schwarzen Kugeln ($\binom{N-r}{n-k} = \binom{2}{2} = 1$ Teilkombinationen) gezogen werden. Jede günstige Teilkombination der roten Kugeln kann mit jeder günstigen Teilkombination der schwarzen Kugeln kombiniert werden, so dass man insgesamt $\binom{r}{k}\cdot\binom{N-r}{n-k} = \binom{3}{2}\cdot\binom{2}{2} = 3\cdot 1 = 3$ günstige Kombinationen erhält.

Die gesuchte Wahrscheinlichkeit im Musterbeispiel ist also

$$
\begin{aligned}
\text{P(genau 3 rote und 2 schwarze Kugeln)} &= \frac{\text{Zahl der günstigen Ergebnisse}}{\text{Zahl der möglichen Ergebnisse}} \\
&= \frac{\text{2 von 3 roten Kugeln kombiniert mit 2 von 2 schwarzen Kugeln}}{\text{4 von 5 Kugeln insgesamt}} \\
&= \frac{\binom{r}{k}\cdot\binom{N-r}{n-k}}{\binom{N}{n}} \\
&= \frac{\binom{3}{2}\cdot\binom{2}{2}}{\binom{5}{4}} \\
&= \frac{3\cdot 1}{5} \\
&= 0{,}6.
\end{aligned}
$$

Satz über die hypergeometrische Verteilung
Eine Urne enthält **N** Kugeln, davon **r** rote. Es werden **n** Kugeln **ohne** Zurücklegen gezogen.

Die Wahrscheinlichkeit, genau **k** rote Kugeln zu ziehen, ist dann $H_{n, r, N}(k) = \dfrac{\binom{r}{k} \cdot \binom{N-r}{n-k}}{\binom{N}{n}}$.

Anwendung Lottospiel:
Die Urne enthält N = 49 Kugeln, von denen r = 6 Stück vom Spieler vorher als „richtig“ gewählt wurden. Es werden n = 6 Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit, daß unter den n = 6 gezogenen Kugeln k = 4 „richtige“ sind, ist dann

$$H_{6, 6, 49}(4) = \frac{\binom{6}{4} \cdot \binom{49-6}{6-4}}{\binom{49}{6}} = \frac{\binom{6}{4} \cdot \binom{43}{2}}{\binom{49}{6}} = 0{,}000968 \approx 0{,}1\%$$

Anwendung Kartenspiele:
Beim Skatspiel enthält das Spiel N = 32 Karten, darunter r = 4 der begehrten Buben. Jeder Spieler erhält („zieht“) n = 10 Karten. Die Wahrscheinlichkeit, dass unter den n = 10 ausgeteilten Karten k = 3 Buben sind, ist dann

$$H_{10, 4, 32}(4) = \frac{\binom{4}{3} \cdot \binom{32-4}{10-3}}{\binom{32}{10}} = \frac{\binom{4}{3} \cdot \binom{28}{7}}{\binom{32}{10}} = 0{,}0734 \approx 7{,}3\%$$

Wahrscheinlichkeitsfunktion und Verteilungsfunktion
Um einen Überblick über die Wahrscheinlichkeiten P(X = k) aller möglichen Ergebnisse X = k eines Experiments zu bekommen, trägt man
- die Wahrscheinlichkeitsfunktion $f_X(k) = P(X = k)$ und
- die Verteilungsfunktion $F_X(k) = P(X \leq k) = P(X = 1) + ... + P(X = k) = f_X(1) + ... + f_X(k)$

in ein Koordinatensystem über der „Trefferzahl“ k auf.

Ist die Zufallsvariable **hypergeometrisch** verteilt, so erhält man

- die **Wahrscheinlichkeitsfunktion** $f_{n, r, N}(k) = H_{n, r, N}(k) = \dfrac{\binom{r}{k} \cdot \binom{N-r}{n-k}}{\binom{N}{n}} = P(X = k)$ und
- die **Verteilungsfunktion** $F_{n, r, N}(k) = f_{n, r, N}(1) + ... + f_{n, r, N}(k) = P(X \leq k)$

Dabei sind
- die Zahl n der gezogenen Kugeln
- die Zahl r der insgesamt vorhandenen richtigen Kugeln
- die Zahl N der insgesamt vorhandenen Kugeln

feste **Parameter** und
- die Zahl k der gezogenen richtigen Kugeln

die **Variable**.

Übung: Aufgaben zu Verteilungsfunktionen Nr. 1

Die Trefferzahl mit der höchsten Wahrscheinlichkeit (bzw. relativen Häufigkeit) ist jeweils $k = 5 \cdot \frac{r}{10} = n \cdot \frac{r}{N} = n \cdot p$, wobei $p = \frac{r}{N}$ die Wahrscheinlichkeit ist, bei **einer** Ziehung eine richtige Kugel zu ziehen. Da die Verteilungskurven symmetrisch zu diesem Maximum sind, liegt dort auch der **Erwartungswert**:

Erwartungswert und Varianz der Hypergeometrischen Verteilung
Ist eine Zufallsvariable $H_{n,\,r,\,N}$ (k)-verteilt, so ist

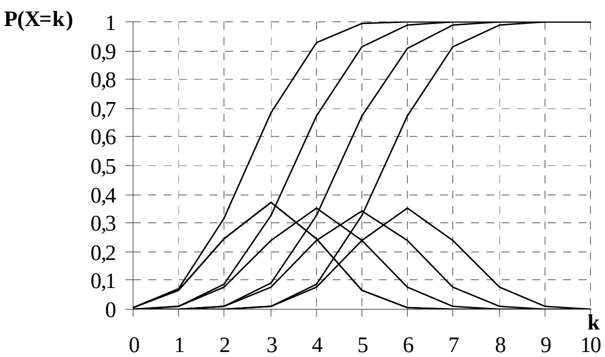
- der **Erwartungswert** $E(X) = n \cdot p$

- die **Varianz** $V(X) = npq\,\frac{N-n}{N-1}$,

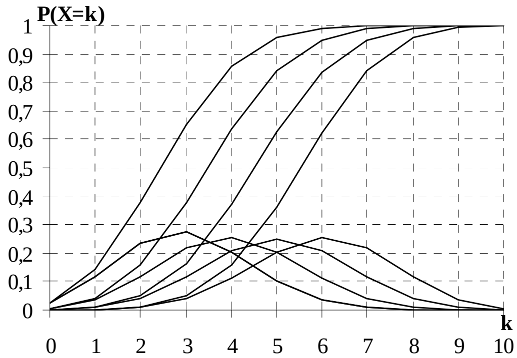
wobei $p = \frac{r}{N}$ und $q = 1 - p$.

Die Lage des Erwartungswertes lässt sich noch deutlicher erkennen, wenn man von größeren N und n ausgeht:

a) N = 20, n= 10, r = 6, 8, 10 und 12



b) N = 200, n = 100, r = 60, 80, 100 und 120:



Alternative Herleitung der Hypergeometrischen Verteilung

Bei der obigen Herleitung der Formel $H_{n,\,r,\,N}(k) = \frac{\binom{r}{k}\cdot\binom{N-r}{n-k}}{\binom{N}{n}}$ wurden sowohl bei der Zahl $\binom{N}{n}$ der **möglichen Ergebnisse** im Nenner als auch bei der Zahl $\binom{r}{k}\cdot\binom{N-r}{n-k}$ der **günstigen Ergebnisse** im Zähler nur die **Kombinationen** gezählt. Ereignisse, die sich nur in der Anordnung unterscheiden und von der Auswahl her gleich sind, werden durch den Übergang zu Binomialkoeffizienten (bzw. Teilen durch die entsprechende Gruppengröße = Zahl der möglichen Anordnungen einer Kombination) zu Ergebnissen zusammengefaßt.
Die gleiche Formel lässt sich aber auch gewinnen, wenn man auf die Zusammenfassung der Ereignisse mit gleicher Kombination verzichtet und sowohl im Nenner als auch im Zähler direkt alle **Ereignisse** zählt. Die Zahl der **möglichen Ereignisse** im Musterbeispiel ist dann $5\cdot4\cdot3\cdot2 = 5_4 = 120$. Die Zahl der **günstigen Ereignisse** erhält man folgendermaßen: Es gibt $3\cdot2 = 3_2 = 6$ Möglichkeiten, die ersten beiden k = 2 Kugeln in beliebiger Anordnung aus den vorhandenen r = 3 roten Kugeln zu ziehen und $2\cdot1 = 2_2 = 2$ Möglichkeiten, die folgenden n − k = 2 Kugeln in beliebiger Anordnung aus den vorhandenen N − r = 2 schwarzen Kugeln zu ziehen. Insgesamt ergeben sich also $3_2\cdot2_2 = 6\cdot2 = 12$ Möglichkeiten, die ersten beiden Kugeln rot und die folgenden beiden Kugeln schwarz zu ziehen. Ebenso viele Möglichkeiten hat man, die 1. + 3. Kugel, die 1. + 4. Kugel, die 2. + 3. Kugel, die 2. + 4. Kugel und die 3. + 4. Kugel rot und jeweils die beiden anderen schwarz zu ziehen.
Insgesamt ergeben sich also $\binom{4}{2}\cdot3_2\cdot2_2 = 6\cdot12 = 72$ Möglichkeiten, zwei rote und zwei schwarze Kugeln in beliebiger Anordnung zu ziehen. Die gesuchte Wahrscheinlichkeit im Musterbeispiel ist also

$$P(\text{genau 3 rote und 2 schwarze Kugeln}) = \frac{\text{Zahl der günstigen Ereignisse}}{\text{Zahl der möglichen Ereignisse}}$$

$$= \frac{\text{2 von 3 roten K. kombiniert mit 2 von 2 schwarzen K. in bel. Reihenfolge}}{\text{4 von 5 Kugeln in beliebiger Reihenfolge}}$$

$$= \frac{\binom{n}{k}\cdot r_k \cdot (N-r)_{n-k}}{N_n}$$

$$= \frac{\binom{3}{2} \cdot 3_2 \cdot 2_2}{5_4}$$

$$= \frac{72}{120}$$

$$= 0{,}6.$$

Alternative Herleitung der hypergeometrische Verteilung

Eine Urne enthält **N** Kugeln, davon **r** rote. Es werden **n** Kugeln **ohne** Zurücklegen gezogen. Die Wahrscheinlichkeit, genau **k** rote Kugeln zu ziehen, ist dann

$$H_{n,\,r,\,N}(k) = \frac{\binom{n}{k} \cdot r_k \cdot (N-r)_{n-k}}{N_n}$$

$$= \frac{\frac{n!}{(n-k)!\,k!} \cdot r \cdot .. \cdot (r-k+1) \cdot (N-r) \cdot ... \cdot (N-r-n+k+1)}{N \cdot ... \cdot (N-n+1)}$$

$$= \frac{\frac{r \cdot ... \cdot (r-k+1)}{k!} \cdot \frac{(N-r) \cdot ... \cdot (N-r-n+k+1)}{(n-k)!}}{\frac{N \cdot ... \cdot (N-n+1)}{n!}}$$

$$= \frac{\binom{r}{k} \cdot \binom{N-r}{n-k}}{\binom{N}{n}}$$