11 Diskrete Zufallsvariablen

11.1 Wahrscheinlichkeitsfunktion und Verteilungsfunktion

In Kapitel 2 wurde zwischen diskreten und stetigen Zufallsvariablen unterschieden. Eine Zufallsvariable X wurde als *diskret* bezeichnet, wenn sie nur endlich viele, höchstens aber abzählbar unendlich viele Ausprägungen annehmen kann.[1] Beispiele für diskrete Zufallsvariablen sind etwa die Merkmale „Augenzahl beim Wurf mit einem Würfel“ (sechs Ausprägungen) oder „Anteil der SPD-Zweitstimmen in % bei den Bundestagswahlen im Zeitraum 1990 - 2005“ (fünf Ausprägungen). Zählvariablen sind stets diskret. Als *stetig* gelten Zufallsvariablen, bei denen die Menge der Ausprägungen Intervalle sind. Die Anzahl der Ausprägungen ist hier nicht mehr abzählbar.

Unterscheidung von Daten- und Modellebene

Im Folgenden geht es um die Wahrscheinlichkeitsverteilung diskreter Zufallsvariablen – zunächst allgemein, bevor dann spezielle diskrete Verteilungsmodelle vorgestellt werden, die häufiger verwendet werden. Zwischen den in Kapitel 4 im Rahmen der beschreibenden Statistik behandelten empirischen Verteilungen (Häufigkeitsverteilungen) und den theoretischen Verteilungen von Zufallsvariablen (Wahrscheinlichkeitsverteilungen) gibt es auffällige Analogien. Dennoch ist eine klare Unterscheidung beider Konzepte wichtig. **Empirische Verteilungen** basieren auf Daten, während **theoretische Verteilungen** Modelle sind, mit denen man die Realität näherungsweise abzubilden versucht. Auch die Kenngrößen empirischer und theoretischer Verteilungen zur Charakterisierung von Lage und Streuung sollten auseinander gehalten und unterschiedlich benannt werden.

Option 1: Beschreibung diskreter Zufallsvariablen anhand der Wahrscheinlichkeitsfunktion

Betrachtet sei eine diskrete Zufallsvariable X, die k Werte $x_1, \ldots, x_k$ annehmen kann. Letztere definieren die **Trägermenge** der Zufallsvariablen X. Das Verhalten von X ist vollständig definiert, wenn für jede Realisation x_i die Eintrittswahrscheinlichkeit $p_i = P(X = x_i)$ bekannt ist; $i = 1, \ldots, k$. Die Funktion f, die jeder Ausprägung x_i eine Eintrittswahrscheinlichkeit p_i zuordnet, heißt **Wahrscheinlichkeitsfunktion** von X. Damit die Wahrscheinlichkeitsfunktion nicht nur auf der Trägermenge $\{x_1, \ldots, x_k\}$, sondern für alle reellen Zahlen x erklärt ist, setzt man sie Null für alle x mit $x \neq x_i$:

$$f(x) = \begin{cases} p_i & \text{für } x = x_i;\ i = 1, 2, \ldots, k \\ 0 & \text{für alle sonstigen } x. \end{cases} \tag{11.1}$$

Die Funktion $f(x)$ lässt sich anhand eines Stab- oder Säulendiagramms mit k Stäben bzw. Säulen der Länge $p_1, p_2, \ldots, p_k$ darstellen. Besonders einfach ist der Fall, dass alle Ausprägungen x_i die gleiche Eintrittswahrscheinlichkeit $p = \frac{1}{k}$ besitzen, also $p_i \equiv p$ gilt (lies: p-i identisch p). Man spricht dann von einer **diskreten Gleichverteilung** oder genauer von einer diskreten Gleichverteilung mit Parameter p. Eine Gleichverteilung

[1] Zum Begriff „abzählbar unendlich“ vgl. erneut die Fußnote in Abschnitt 2.2.

trifft z. B. für das Merkmal „Augenzahl X beim Würfeln mit einem Würfel“ zu, wenn man mehrfach würfelt und einen „fairen“ Würfel voraussetzt, also einen Würfel, bei dem alle Augenzahlen mit gleicher Wahrscheinlichkeit auftreten. Die Variable X hat hier sechs Ausprägungen $x_1 = 1, x_2 = 2, \ldots, x_6 = 6$, die alle die Eintrittswahrscheinlichkeit $p = \frac{1}{6}$ aufweisen. Die Wahrscheinlichkeitsfunktion $f(x)$ der zugehörigen Gleichverteilung ist im oberen Teil von Abbildung 11.1 wiedergegeben.

Die diskrete Gleichverteilung

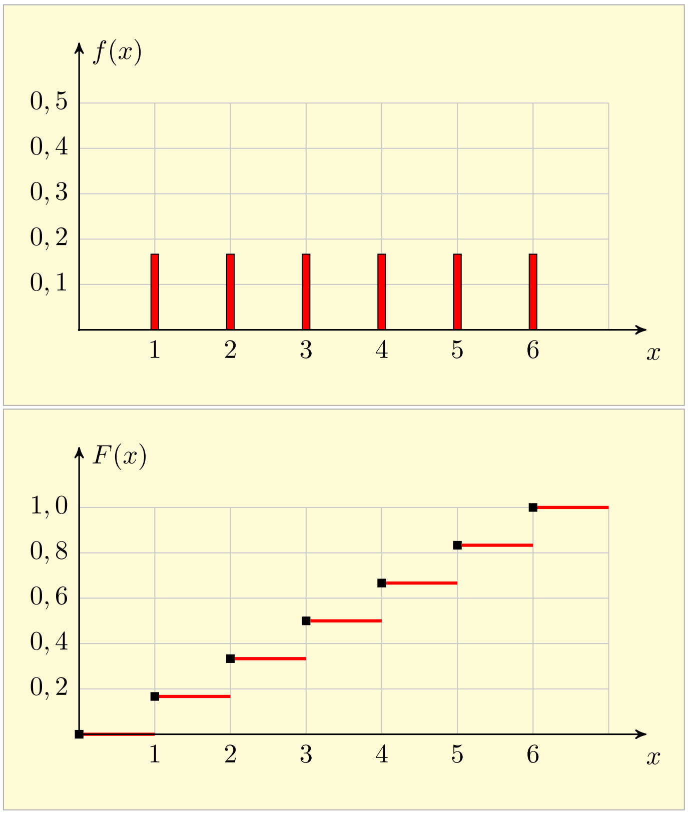


Abb. 11.1: *Wahrscheinlichkeits- und Verteilungsfunktion der diskreten Gleichverteilung mit $p = \frac{1}{6}$ (Würfeln mit einem Würfel)*

Eine Wahrscheinlichkeitsfunktion kann nur nicht-negative Werte annehmen. Ferner muss die Summe der Eintrittswahrscheinlichkeiten $p_1, p_2, \ldots, p_k$ in (11.1) stets 1 sein. Hier besteht eine Analogie zu den in Kapitel 4 behandelten relativen Häufigkeitsverteilungen , denn auch relative Häufigkeiten sind nicht-negativ und summieren sich zu 1 auf.

Zur Beschreibung des Verhaltens der durch (11.1) definierten diskreten Zufallsvariablen X kann man anstelle der Wahrscheinlichkeitsfunktion auch die **Verteilungsfunktion**[2]

$$F(x) = P(X \leq x) \tag{11.2}$$

von X heranziehen, die man zwecks Unterscheidung von der empirischen Verteilungsfunktion für einen Datensatz präziser **theoretische Verteilungsfunktion** nennt.

[2] Wenn man die Wahrscheinlichkeitsverteilungen zweier Zufallsvariablen unterscheiden will, kann man durch einen tiefgestellten Index deutlich machen, welche Verteilung gerade gemeint ist. Für eine Variable X würde man also z. B. zwecks Präzisierung $f_X(x)$ und $F_X(x)$ anstelle von $f(x)$ und $F(x)$ schreiben.

Option 2: Beschreibung diskreter Zufallsvariablen anhand der Verteilungsfunktion

Offenbar hat $F(x)$ für $x < x_1$ den Wert Null und springt in $x = x_1$ auf den Wert $F(x_1) = p_1$. Der Funktionswert bleibt auf dem Niveau p_1 bis zur Stelle $x = x_2$, an der ein erneuter Sprung nach oben erfolgt, nun auf $F(x_2) = p_1 + p_2$, usw. Die Werte der Funktion $F(x)$ ergeben sich also dadurch, dass an den Stellen $x = x_i$ jeweils ein positiver Beitrag p_i hinzukommt, d.h. $F(x)$ ist eine monoton wachsende Treppenfunktion mit Sprungstellen in $x = x_i$. Bei der letzten Sprungstelle, also in $x = x_k$, erreicht $F(x)$ den Wert 1. Anstelle von (11.2) kann man demnach hier auch schreiben:

$$F(x) = \begin{cases} 0 & \text{für } x < x_1 \\ p_1 & \text{für } x_1 \leq x < x_2 \\ \vdots & \vdots \\ p_1 + p_2 + \ldots + p_{k-1} & \text{für } x_{k-1} \leq x < x_k \\ 1 & \text{für } x \geq x_k. \end{cases} \tag{11.3}$$

Charakterisierung der diskreten Gleichverteilung

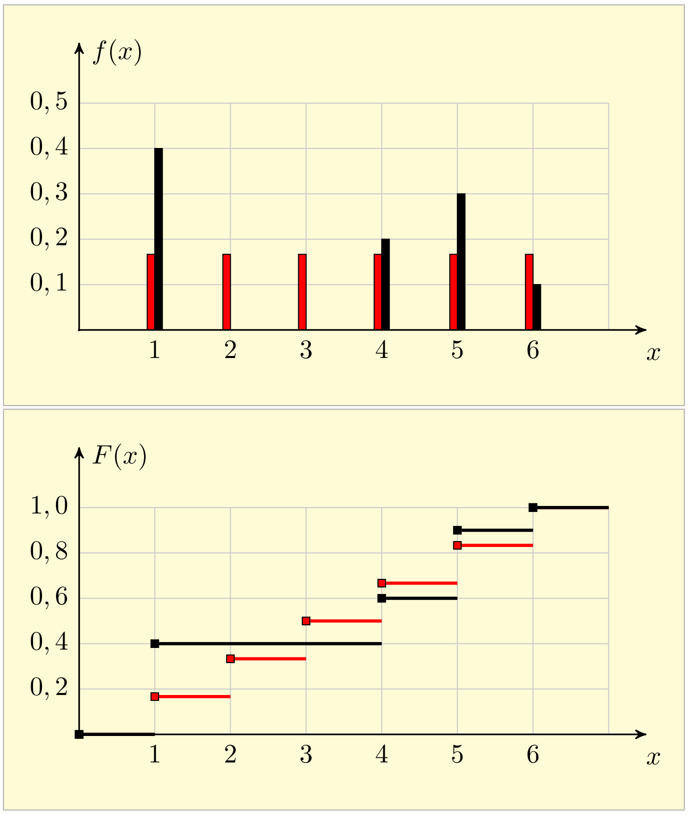
Wahrscheinlichkeits- und Verteilungsfunktion einer diskreten Gleichverteilung mit k Ausprägungen $x_1, x_1, \ldots, x_k$ gehen aus (11.1) und (11.3) als Spezialfall hervor, wenn dort für alle Eintrittswahrscheinlichkeiten p_i der Wert $p = \frac{1}{k}$ eingesetzt wird. Der untere Teil von Abbildung 11.1 zeigt die Verteilungsfunktion $F(x)$ des mit $p = \frac{1}{6}$ diskret gleichverteilten Merkmals „Augenzahl X beim Würfeln mit einem Würfel". Die Funktion weist an den Stellen $x_1 = 1, x_2 = 2, \ldots, x_6 = 6$ jeweils Sprünge der festen Höhe $\frac{1}{6}$ auf.

Es gibt eine weitere Parallele zwischen den in Kapitel 4 behandelten relativen Häufigkeitsverteilungen und den Verteilungen diskreter Zufallsvariablen. Durch Aufsummieren relativer Häufigkeiten kommt man zu empirischen Verteilungsfunktionen, die – wie in Kapitel 4 gezeigt – ebenfalls monoton wachsende Treppenfunktionen sind, welche bis zum ersten Sprung den Wert 0 aufweisen und nach der letzten Sprungstelle den Wert 1 erreichen.

Java-Applet „Würfeln mit einem Würfel (mit Modell)"

Zusammenhänge zwischen Verteilungen diskreter Merkmale (theoretische Verteilungen) und relativen Häufigkeitsverteilungen (empirische Verteilungen) lassen sich anhand des statistischen Experiments „Würfeln mit einem Würfel" gut sichtbar machen, wenn man das Experiment n-mal durchführt mit hinreichend groß gewähltem n. Abbildung 11.2 zeigt im oberen Teil die per Simulation gewonnenen relativen Häufigkeiten in Form schwarzer Säulen für die sechs möglichen Ausprägungen bei nur 10-facher Durchführung des statistischen Experiments ($n = 10$). Im unteren Teil ist, ebenfalls in Schwarz, die hieraus resultierende empirische Verteilungsfunktion wiedergegeben. Zu Vergleichszwecken ist auch das schon in Abbildung 11.1 dargestellte Modell der diskreten Gleichverteilung mit dem Parameter $p = \frac{1}{6}$ eingezeichnet. Neben der Abbildung sind in einer Tabelle – dort in der mittleren Spalte – die beobachteten relativen Häufigkeiten f_i für die einzelnen Augenzahlen und in der letzten Spalte die hier wieder mit F_i abgekürzten Werte der empirischen Verteilungsfunktion an den Stellen $x = x_i$ aufgeführt ($i = 1, 2, \ldots, 6$). Die Tabelle zeigt, dass bei den 10 Würfen viermal die Augenzahl 1, zweimal die Augenzahl 4, dreimal die Augenzahl 5 und einmal die Augenzahl 6 erschien.

Abbildung 11.3 zeigt erneut die relativen Häufigkeiten und die daraus abgeleitete empirische Verteilungsfunktion, nun aber für den Fall $n = 100$. Auch hier ist zusätzlich



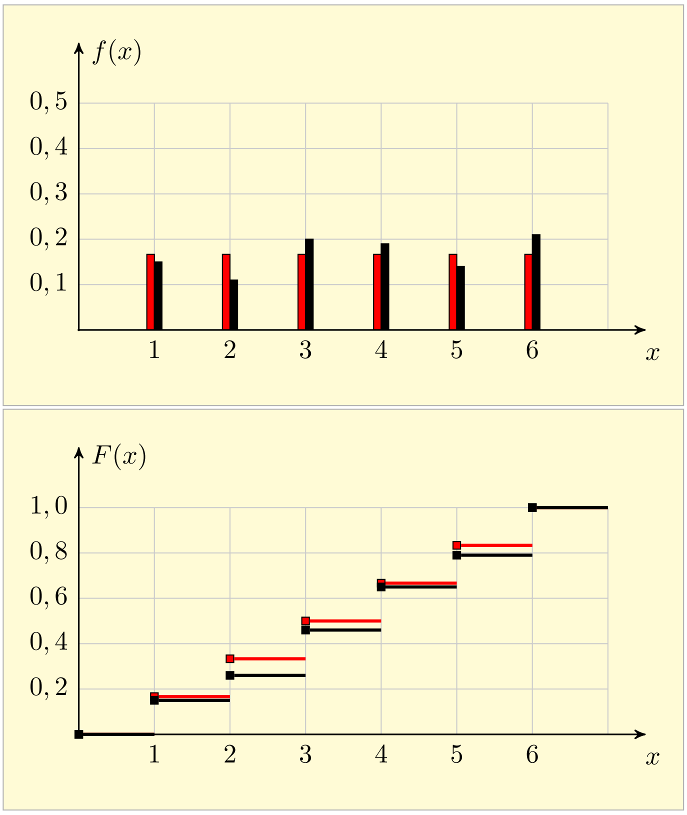
x	$f(x)$	$F(x)$
1	0,4	0,4
2	0	0,4
3	0	0,4
4	0,2	0,6
5	0,3	0,9
6	0,1	1,0

Abb. 11.2: *Relative Häufigkeiten für die sechs Augenzahlen bei 10-fachem Würfeln mit einem Würfel und Vergleich mit der diskreten Gleichverteilung mit* $p = \frac{1}{6}$

das Modell der diskreten Gleichverteilung mit $p = \frac{1}{6}$ dargestellt. Ferner sind erneut die relativen Häufigkeiten f_i und die kumulierten Häufigkeiten F_i tabellarisch ausgewiesen. Man erkennt beim Vergleich von Abbildung 11.3 mit Abbildung 11.2, dass das theoretische Verteilungsmodell die Simulationsergebnisse bei größerem n tendenziell besser beschreibt – die im Experiment beobachteten relativen Häufigkeiten f_i nähern sich den Werten $f(x_i) = \frac{1}{6}$ der Wahrscheinlichkeitsfunktion mit Vergrößerung von n an.[3]

Der erste Wert $f_1 = 0,4$ der vorletzten Spalte in der Tabelle neben Abbildung 11.2 besagt z. B., dass in 40 % der Fälle, also bei 4 der $n = 10$ Würfe, die Augenzahl $x_1 = 1$ beobachtet wurde. Der entsprechende Wert $f_1 = 0,15$ in der Tabelle neben Abbildung 11.3, der sich auf $n = 100$ bezieht und 15 % der Würfe mit Augenzahl $x_1 = 1$ beinhaltet, liegt schon viel näher am theoretischen Wert $f(x_1) = \frac{1}{6} \approx 0,17$. Eine analoge Feststellung gilt etwa für den Vergleich der Werte $F_5 = 0,90$ (Wert der Verteilungsfunktion im Falle $n = 10$) und $F_5 = 0,79$ (Wert im Falle $n = 100$) in der letzten Spalte der beiden Tabellen. Diese Werte geben die kumulierten relativen Häufigkeiten für das Auftreten einer Augenzahl bis einschließlich 5 an und stellen Schätzwerte für den Wert $F(x_5) = \frac{5}{6} \approx 0,83$ der Verteilungsfunktion an der Stelle $x_5 = 5$ dar. Der Wert $F_5 = 0,79$ in der Tabelle neben Abbildung 11.3 liegt näher am theoretischen Wert $F(x_5)$.

[3]Nur um die Notation nicht zu sehr zu komplizieren, wird in diesem Manuskript sowohl für die empirische als auch für die theoretische Verteilungsfunktion dieselbe Bezeichnung $F(x)$ verwendet.



x	$f(x)$	$F(x)$
1	0,15	0,15
2	0,11	0,26
3	0,20	0,46
4	0,19	0,65
5	0,14	0,79
6	0,21	1,00

Abb. 11.3: *Relative Häufigkeiten für die sechs Augenzahlen bei 100-fachem Würfeln mit einem Würfel und Vergleich mit der diskreten Gleichverteilung mit* $p = \frac{1}{6}$

JACOB I. BERNOULLI

Neben der diskreten Gleichverteilung ist noch ein weiterer einfacher Spezialfall einer diskreten Verteilung zu erwähnen, nämlich die nach dem Schweizer Mathematiker Jacob I. BERNOULLI (1655 - 1705) benannte **Bernoulli-Verteilung**, für die man auch die Bezeichnung **Zweipunkt-Verteilung** findet. Diese Verteilung liegt vor, wenn eine Zufallsvariable X nur zwei Ausprägungen aufweist, etwa x_1 und x_2 oder A und $\overline{A}$. Die Variable X spricht man auch als **binäre Zufallsvariable** an. Bezeichnet $p_1 = p$ die Eintrittswahrscheinlichkeit für den Fall $x = x_1$ und p_2 die für den Fall $x = x_2$, so ist offenbar $p_2 = 1-p$. Die Wahrscheinlichkeitsfunktion (11.1) hat dann die spezielle Gestalt

$$f(x) = \begin{cases} p & \text{für } x = x_1; \\ 1-p & \text{für } x = x_2; \\ 0 & \text{für alle sonstigen } x. \end{cases} \tag{11.4}$$

Charakterisierung der Bernoulli-Verteilung

Durch die Wahrscheinlichkeitsfunktion (11.4) oder die aus ihr ableitbare Verteilungsfunktion

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x < x_1; \\ p & \text{für } x_1 \leq x < x_2; \\ 1 & \text{für } x \geq x_2 \end{cases} \tag{11.5}$$

ist eine Bernoulli-Verteilung vollständig definiert. Ihre Gestalt hängt natürlich vom Parameter p ab. Eine mit dem Parameter p bernoulli-verteilte Zufallsvariable X bezeichnet man auch als $Be(p)$-verteilt und verwendet hierfür die Notation $X \sim Be(p)$ (lies: X

ist *bernoulli-verteilt* mit dem Parameter p). Das Merkmal „Ergebnis eines Münzwurfexperiments“ (einmaliger Münzwurf mit den möglichen Realisationen „Zahl“ und „Kopf“) ist z. B. bernoulli-verteilt mit $p = 0,5$, wenn man eine „faire“ Münze voraussetzt. Ein statistisches Experiment, dessen Ausgang durch ein bernoulli-verteiltes Merkmal beschrieben wird, heißt **Bernoulli-Experiment**.

Java-Applet “Münzwurf“

Die nachstehende Grafik vermittelt eine Aussage, die schon in Abbildung 10.2 visualisiert wurde. Sie zeigt den Entwicklungspfad der relativen Häufigkeiten $f_j = f_j(\text{Zahl})$ des Auftretens der Ausprägung „Zahl“ nach j Münzwurfexperimenten, wobei $j = 1, 2, \ldots, n$ mit $n = 500$. Es wird also ein Bernoulli-Experiment wiederholt durchgeführt – man spricht in diesem Zusammenhang unter der hier erfüllten Voraussetzung der Unabhängigkeit der Einzelexperimente auch von einer **Bernoulli-Kette** – und die Zwischenstände $f_j(\text{Zahl})$ bis zum Endstand fortlaufend visualisiert. Abbildung 11.4 zeigt nicht nur einen, sondern zwei Entwicklungspfade, also zwei Bernoulli-Ketten. Der Endstand $f_{500}(\text{Zahl})$ der beobachteten relativen Häufigkeit liegt in beiden Fällen sehr dicht am Wert $p = 0,5$ der Eintrittswahrscheinlichkeit für „Zahl“, gegen den die Bernoulli-Ketten für $n \to \infty$ stochastisch konvergieren. Wenn man die Ausprägungen x_1 und x_2 zu 1 und 0 umcodiert (vgl. (11.18)), wird eine Bernoulli-Verteilung auch **Null-Eins-Verteilung** genannt.

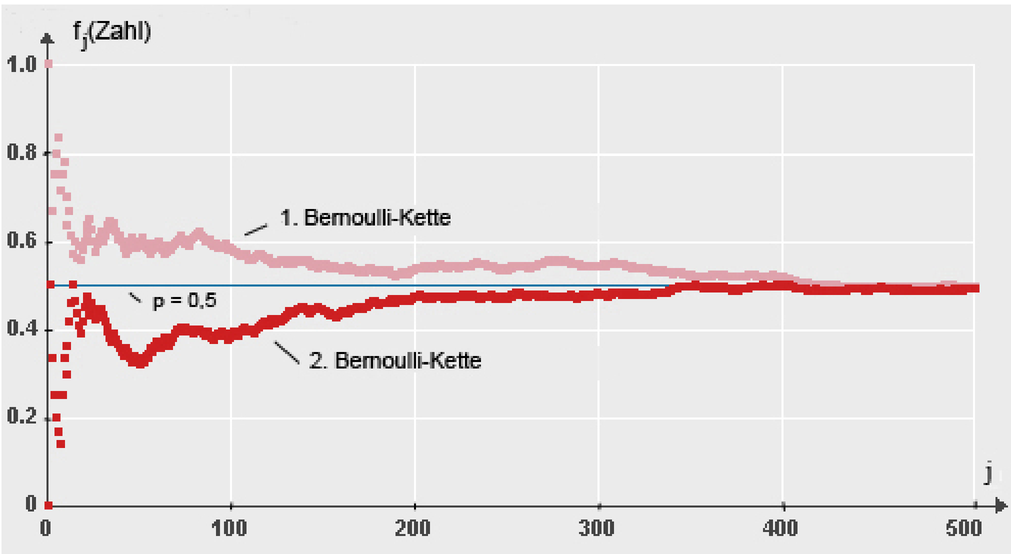


Abb. 11.4: *Relative Häufigkeit für „Zahl“ bei 500-fachem Münzwurf und Vergleich mit dem Parameter $p = 0,5$ der Bernoulli-Verteilung*

11.2 Kenngrößen diskreter Verteilungen

In Kapitel 5 wurden empirische Verteilungen durch wenige Kenngrößen charakterisiert. Zu nennen sind hier insbesondere die Lageparameter Mittelwert und Median , mit denen der Schwerpunkt einer Verteilung beschrieben wurde, sowie die Streuungsparameter Spannweite, Standardabweichung und Varianz, mit denen die Variabilität eines Datensatzes ausgedrückt werden kann. Auch theoretische Verteilungen werden durch Lage- und Streuungsmaße charakterisiert. Die Analogien zwischen empirischen und theoretischen Verteilungen sind bei den diskreten Zufallsvariablen besonders augenfällig.

Das arithmetische Mittel $\overline{x}$ einer empirischen Verteilung eines Datensatzes $x_1, x_2, \ldots, x_n$, der sich auf ein diskretes Merkmal X mit k Ausprägungen $a_1, a_2, \ldots a_k$ bezieht, lässt sich gemäß (5.4) als Summe der mit den relativen Häufigkeiten gewichteten Merkmalsausprägungen darstellen, also durch $a_1 f_1 + a_2 f_2 + \ldots + a_k f_k$. In ähnlicher Weise lässt sich auch der Schwerpunkt der Verteilung der diskreten Zufallsvariablen (11.1) charakterisieren. Man bildet hier die Summe $x_1 p_1 + x_2 p_2 + \ldots + x_k p_k$ der mit den Eintrittswahrscheinlichkeiten $p_1, p_2, \ldots, p_k$ gewichteten Realisationen. Diese Summe wird als **Erwartungswert** bezeichnet und mit $E(X)$ oder kürzer mit μ bezeichnet. Der Erwartungswert $E(X)$ (lies: *Erwartungswert von* X) einer nach (11.1) definierten diskreten Zufallsvariablen ist also gegeben durch

Erwartungswert und Varianz einer diskreten Zufallsvariablen

$$\mu := E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k = \sum_{i=1}^{k} x_i p_i. \tag{11.6}$$

Aufgabe 11.1

Die Merkmalsausprägungen $a_1, a_2, \ldots a_k$ und die relativen Häufigkeiten $f_1, f_2, \ldots f_k$ aus Kapitel 5 werden also hier, bei der Charakterisierung theoretischer Verteilungsmodelle, durch die Realisationen $x_1, x_2, \ldots x_k$ einer diskreten Zufallsvariablen und deren Eintrittswahrscheinlichkeiten $p_1, p_2, \ldots p_k$ ersetzt. Die gleichen Ersetzungen kann man auch in den Formeln (5.8) und (5.7) für die empirische Standardabweichung bzw. die empirische Varianz vornehmen. Man erhält so für die mit $V(X)$ oder σ^2 (lies: *Varianz von* X resp. *sigma-Quadrat*) abgekürzte **Varianz** der diskreten Zufallsvariablen (11.1) mit $\mu = E(X)$ die Darstellung

$$\sigma^2 := V(X) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \ldots + (x_k - \mu)^2 p_k = \sum_{i=1}^{k} (x_i - \mu)^2 p_i. \tag{11.7}$$

Die Darstellung (11.6) geht offenbar in (11.7) über, wenn man in (11.6) anstelle von X den Term $(X - \mu)^2$ einsetzt. Die Varianz $\sigma^2 = V(X)$ einer Zufallsvariablen ist also nichts anderes als der Erwartungswert der quadrierten Differenz zwischen X und $\mu = E(X)$:

$$\sigma^2 = E\left[(X - \mu)^2\right]. \tag{11.8}$$

Für die Varianz ist manchmal die Darstellung

$$\sigma^2 = E(X^2) - \mu^2 \tag{11.9}$$

nützlich, die sich aus (11.8) ergibt, wenn man dort den Term in der eckigen Klammer ausmultipliziert und dann den Erwartungswert gliedweise anwendet – s. hierzu auch die noch folgenden Formeln (11.11) und (11.13). Der Varianzdarstellung (11.9) entspricht auf der empirischen Ebene die Zerlegungsformel (5.7).

Die mit σ (lies: *sigma*) bezeichnete **Standardabweichung** von X ist definiert durch

$$\sigma = \sqrt{V(X)}. \tag{11.10}$$

Zwischen den Kenngrößen empirischer und theoretischer Verteilungen wird in der Lehrbuchliteratur oft nicht klar unterschieden. Der Mittelwert bezieht sich auf eine empirische, der Erwartungswert immer auf eine theoretische Verteilung. Wenn von der Varianz die Rede ist, kann man durch die Verwendung der präziseren Bezeichnungen „empirische Varianz" bzw. „theoretische Varianz" deutlich machen, ob die Varianz eines Datensatzes (empirische Ebene) oder die einer Zufallsvariablen (Modellebene) gemeint ist. Eine analoge Aussage gilt für die Standardabweichung.

Lineartransformationen bei Zufallsvariablen

In der Praxis unterzieht man eine Zufallsvariable X mit Erwartungswert oft einer Lineartransformation $Y = aX + b$. Die Addition von b entspricht einer Verschiebung des Nullpunkts, während die Multiplikation von X mit einem von Null verschiedenen Wert a eine Streckung oder Stauchung der zur Messung verwendeten Skala beinhaltet (im Fall $a < 0$ kommt noch ein Vorzeichenwechsel hinzu). Lineartransformationen sind z. B. relevant, wenn man eine andere Skala bei der Messung verwendet (etwa Temperaturmessung in Kelvin statt in Celsius) oder wenn man X in eine Zufallsvariable Y mit Erwartungswert $E(Y) = 0$ und Varianz $V(Y) = 1$ überführen will (**Standardisierung**).

Unterzieht man eine Zufallsvariable X mit Erwartungswert $\mu = E(X)$ einer Lineartransformation $Y = aX + b$, so ergeben sich Erwartungwert und Varianz von Y unter Rückgriff auf die Definitionen (11.6) und (11.7) nach

$$E(Y) = E(aX + b) = a \cdot E(X) + b \tag{11.11}$$

$$V(Y) = V(aX + b) = a^2 \cdot V(X). \tag{11.12}$$

Für den Erwartungswert und die Varianz der Summe zweier unabhängiger Zufallsvariablen X und Y gelten ferner die hier nicht bewiesenen Darstellungen [4]

$$E(X + Y) = E(X) + E(Y) \tag{11.13}$$

$$V(X + Y) = V(X) + V(Y). \tag{11.14}$$

Die Darstellungen (11.13) und (11.14) gelten entsprechend auch für die Summen von n unabhängigen Zufallsvariablen ($n \geq 2$).

Kenngrößen der Null-Eins-Verteilung

Erwartungswert und Varianz der Null-Eins-Verteilung ergeben sich unmittelbar aus

[4] Der Begriff der „Unabhängigkeit" von zwei oder mehreren Zufallsvariablen wird in Abschnitt 13.1 formalisiert. Im Gegensatz zu (11.13) gilt die Darstellung (11.14) nicht mehr bei Abhängigkeit von X und Y; in diesem Falle ist sie durch (13.14) zu ersetzen.

den Formeln (11.6) und (11.7) für den Erwartungswert bzw. die Varianz diskreter Zufallsvariablen, wenn man dort $k = 2$ sowie $x_1 = 1$, $p_1 = p$, $x_2 = 0$ und $p_2 = 1 - p$ einsetzt und bei der Varianzberechnung auf die Verschiebungsregel zurückgreift:

$$\mu = 1 \cdot p + 0 \cdot (1 - p) = p. \tag{11.15}$$

$$\sigma^2 = E(X^2) - \mu^2 = p - p^2 = p(1 - p). \tag{11.16}$$

Beispiel 11.1 **Erwartungswert und Varianz der Augenzahl beim Würfeln**

In Abbildung 11.1 wurde die Wahrscheinlichkeitsverteilung der Zufallsvariablen „Augenzahl X beim Würfeln mit einem Würfel" (Gleichverteilung mit Parameter $p = \frac{1}{6}$) anhand ihrer Wahrscheinlichkeitsfunktion $f(x)$ und ihrer Verteilungsfunktion $F(x)$ veranschaulicht. Da X die Ausprägungen $x_i = i$ mit den festen Eintrittswahrscheinlichkeiten $p_i = p = \frac{1}{6}$ besitzt $(i = 1, 2, \ldots, 6)$ erhält man für den Erwartungswert $\mu = E(X)$ und die Varianz $\sigma^2 = V(X)$ aus (11.6) und (11.7)

$$\mu = \sum_{i=1}^{6} x_i \cdot p_i = \frac{1}{6} \sum_{i=1}^{6} i = \frac{21}{6} = 3,5$$

$$\sigma^2 = \sum_{i=1}^{6} (x_i - \mu)^2 p_i = \frac{1}{6} \sum_{i=1}^{6} (i - 3,5)^2 = \frac{17,5}{6} \approx 2,92.$$

Wie bei empirischen Verteilungen kann man auch bei theoretischen Verteilungen **Quantile** zur näheren Charakterisierung heranziehen. Das **p-Quantil** einer Verteilung ist durch

Quantile als weitere Kenngrößen

$$F(x_p) = p \quad (0 < p < 1) \tag{11.17}$$

definiert, also durch den Wert x_p der Verteilungsfunktion $F(x)$, an dem $F(x)$ den Wert p annimmt. Der **Median** $\widetilde{x} = x_{0,5}$ sowie das **untere Quartil** $x_{0,25}$ und das **obere Quartil** $x_{0,75}$ einer theoretischen Verteilung sind wieder spezielle Quantile, die sich bei Wahl von $p = 0,5$ resp. von $p = 0,25$ und $p = 0,75$ ergeben.

Bei diskreten Verteilungen sind die Quantile durch (11.17) noch nicht eindeutig festgelegt. Bei der im unteren Teil von Abbildung 11.1 wiedergegebenen Verteilungsfunktion einer speziellen diskreten Gleichverteilung gilt z. B. $F(x) = 0,5$ für jeden Wert x aus dem Intervall $3 \leq x < 4$. Man benötigt daher hier – analog zur eindeutigen Festlegung von empirischen Quantilen – noch eine Zusatzbedingung. Man kann z. B. den linken Randpunkt des Intervalls wählen, d. h. das p-Quantil x_p so festlegen, dass $F(x_p) \geq p$ gilt und gleichzeitig $F(x) < p$ für $x < x_p$. Für die diskrete Gleichverteilung in Abbildung 11.1 erhält man so für den Median $\widetilde{x} = x_{0,5}$ den Wert $\widetilde{x} = 3$.

11.3 Die Binomialverteilung

Es fällt nicht schwer, in verschiedenen Lebensbereichen Beispiele für Merkmale X zu finden, die nur zwei mögliche Ausprägungen haben, also den Charakter von Binärvariablen haben. Das Ergebnis eines Münzwurfexperiments wurde schon genannt. Praxisrelevantere Beispiele sind etwa die Geschlechterverteilung bei Geburten, die Verteilung eines Gendefekts in einer Population (nicht betroffene / betroffene Individuen), der beim Mikrozensus erfragte Erwerbsstatus einer Person (erwerbstätig / nicht erwerbstätig) oder der Qualitätsstatus von Produkten bei Serienfertigungen (spezifikationskonform / nicht-spezifikationskonform). Aber auch Merkmale mit mehr als zwei Ausprägungen können stets auf Binärvariablen zurückgeführt werden, wenn man sich nur dafür interessiert, ob eine bestimmte Realisation eintritt oder nicht. Das Würfeln mit einem Würfel lässt sich z. B. als Bernoulli-Experiment interpretieren, wenn man sich darauf beschränkt, nur zwischen den Ereignissen „Augenzahl ist 6 / nicht 6“ oder „Augenzahl ist größer als 2 / nicht größer als 2“ zu unterscheiden.

Hat man ein Bernoulli-Experiment mit den möglichen Ausgängen $x_1 = A$ und $x_2 = \overline{A}$ und den Eintrittswahrscheinlichkeiten $P(A) = p$ bzw. $P(\overline{A}) = 1-p$ mehrfach und unabhängig voneinander durchgeführt, so interessiert man sich oft dafür, wie häufig eine der beiden Realisationen auftritt, etwa A. Beim Münzwurfexperiment könnte dies z. B. die Anzahl der Ausgänge mit „Zahl“ sein. Ist n die Anzahl der unabhängig durchgeführten Bernoulli-Experimente und bezeichnet X die Anzahl der Ausgänge A, so ist die Zählvariable X offenbar eine diskrete Zufallsvariable mit den Ausprägungen $i \quad (i = 0, 1, \ldots, n)$. Wenn man den Ausgang jedes der n Bernoulli-Experimente anhand einer Indikatorvariablen

$$X_i = \begin{cases} 1 & \text{bei Eintritt von } x_1 = A \\ 0 & \text{bei Eintritt von } x_2 = \overline{A} \end{cases} \tag{11.18}$$

beschreibt (null-eins-verteilte Zufallsvariable), so lässt sich X als Summe

$$X = \sum_{i=1}^{n} X_i \tag{11.19}$$

der n voneinander unabhängigen Indikatorvariablen schreiben. Die Verteilung der Zählvariablen X heißt **Binomialverteilung**. Die Binomialverteilung ist für die statistische Praxis von großer Bedeutung. Die Null-Eins-Verteilung ist offenbar ein Spezialfall der Binomialverteilung ($n = 1$).

Kenngrößen der Binomialverteilung

Aus (11.19) kann man leicht den Erwartungswert $E(X)$ und die Varianz $V(X)$ der binomialverteilten Variablen X ableiten. Die in (11.19) eingehenden n Indikatorvariablen X_i sind voneinander unabhängig und folgen alle einer Null-Eins-Verteilung, besitzen demnach wegen (11.15) und (11.16) alle den Erwartungswert $E(X_i) = p$ und die Varianz $V(X_i) = p(1-p)$. Mit den Formeln (11.13) und (11.14), die sich auch für die Summe von n unabhängigen Zufallsvariablen formulieren lassen ($n \geq 2$), folgen hieraus für die Kenngrößen $\mu = E(X)$ und $\sigma^2 = V(X)$ einer Binomialverteilung die Darstellungen

$$\mu = n \cdot p \tag{11.20}$$

$$\sigma^2 = n \cdot p(1-p). \tag{11.21}$$

Da eine diskrete Zufallvariable aber noch nicht durch Erwartungswert und Varianz alleine, sondern erst durch die Wahrscheinlichkeitsfunktion (11.1) oder – alternativ – durch die Verteilungsfunktion (11.2) vollständig beschrieben ist, sei noch die Wahrscheinlichkeitsfunktion der Binomialverteilung abgeleitet. Hierzu werde erst einmal die Wahrscheinlichkeit dafür betrachtet, dass bei dem Bernoulli-Experiment am Anfang genau x-mal der Ausgang A und danach $(n-x)$-mal der Ausgang $\overline{A}$ beobachtet wird, die Bernoulli-Kette also die spezielle Gestalt $A, A, \ldots, A, \overline{A}, \ldots \overline{A}$ hat mit zwei homogenen Teilketten der Längen x bzw. $n-x$. Die Wahrscheinlichkeit für den Eintritt dieser speziellen Ergebnisfolge, die für die Zählvariable X zum Wert x führt, ist wegen der Unabhängigkeit der einzelnen Bernoulli-Experimente $p^x(1-p)^{n-x}$. Nun gibt es aber nicht nur eine Ergebnisfolge, sondern nach Tabelle 10.1 insgesamt $\binom{n}{x}$ mögliche Ausprägungen einer Bernoulli-Kette der Länge n, bei der insgesamt x-mal der Ausgang A auftritt. Die Reihenfolge des Auftretens der Ausgänge A innerhalb einer Ergebnisfolge hat ja keinen Effekt auf den Wert der Zählvariablen X. Die Wahrscheinlichkeit $P(X = x)$ dafür, dass die Anzahl der Ausgänge A innerhalb der Bernoulli-Kette die Ausprägung einen bestimmten Wert x annimmt, ist damit gegeben durch das $\binom{n}{x}$-fache von $p^x(1-p)^{n-x}$. Es gilt also für die **Wahrscheinlichkeitsfunktion** $f(x) = P(X = x)$ der Binomialverteilung

Vollständige Beschreibung der Binomialverteilung

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x = 0, 1, \ldots, n \\ 0 & \text{für alle sonstigen } x. \end{cases} \tag{11.22}$$

Flash-Animation „Galton-Brett und Binomialverteilung“

Die **Verteilungsfunktion** $F(x) = P(X \leq x)$ ist auf der Trägermenge $\{0, 1, \ldots, n\}$ durch

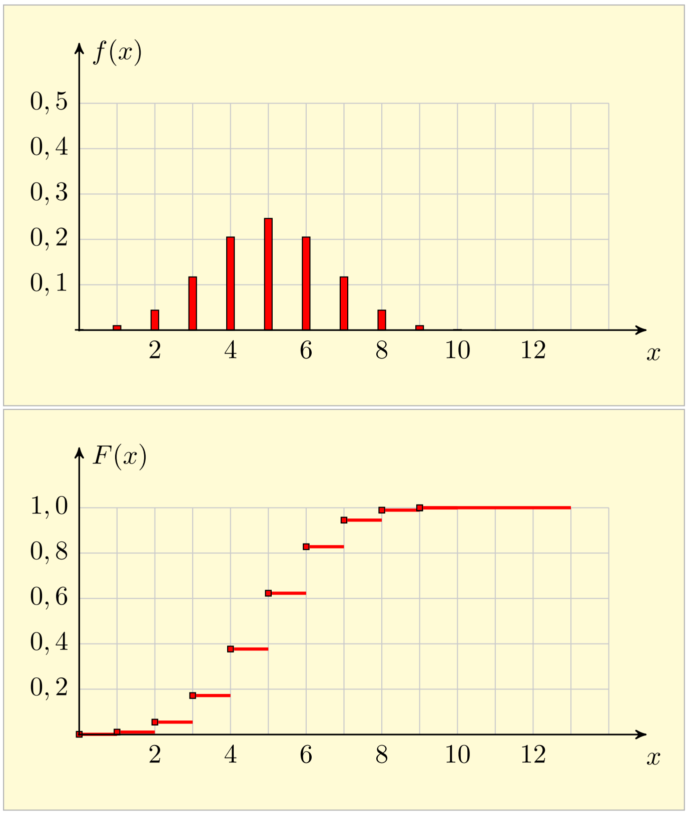
$$F(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \qquad x = 0, 1, \ldots, n. \tag{11.23}$$

definiert. Zwischen zwei benachbarten Elementen der Trägermenge bleibt $F(x)$ auf dem Niveau des kleineren Elements (Treppenfunktion), um dann an der Stelle $x = n$ den Endwert 1 zu erreichen. Eine mit Parametern n und p binomialverteilte Zufallsvariable X bezeichnet man in der Literatur auch als $B(n, p)$-verteilt und schreibt dafür $X \sim B(n, p)$ (lies: X ist *binomialverteilt* mit den Parametern n und p). Die Aussagen $X \sim B(1, p)$ und $X \sim Be(p)$ sind identisch, weil die Bernoulli-Verteilung eine Binomialverteilung mit $n = 1$ ist.

Abbildung 11.5 zeigt Wahrscheinlichkeits- und Verteilungsfunktion einer $B(10; 0,5)$-verteilten Zufallsvariablen. Neben Werten beider Funktionen ist auch der Erwartungswert $\mu = E(X)$ wiedergegeben, der sich als $\mu = 10 \cdot 0,5 = 5$ errechnet. Der Tabelle neben der Grafik entnimmt man z. B., dass die Verteilungsfunktion $F(x)$ an der Stelle $x = 3$ den Wert $F(3) = 0,1719$ annimmt. Dieser Wert ist wegen $F(3) = P(X \leq 3)$ die Summe der Werte $f(0) = 0,0010$, $f(1) = 0,0098$, $f(2) = 0,0439$ und $f(3) = 0,1172$ der Wahrscheinlichkeitsfunktion bis $x = 3$. Durch Aufsummieren von Werten der Wahrscheinlichkeitsfunktionen ergeben sich also die Werte der Verteilungsfunktion. Umgekehrt kann man aus $F(x)$ durch Differenzenbildung Werte der Wahrscheinlichkeitsfuntion $f(x)$ gewinnen. Der Wert $f(3) = P(X = 3) = 0,1172$ ergibt sich in Abbildung 11.5 etwa als

Java-Applet „Die Binomialverteilung“

Differenz von $F(3) = P(X \leq 3) = 0,1719$ und $F(2) = P(X \leq 2) = 0,0547$. Es genügt also eine der beiden Funktionen $f(x)$ und $F(x)$ zu tabellieren.



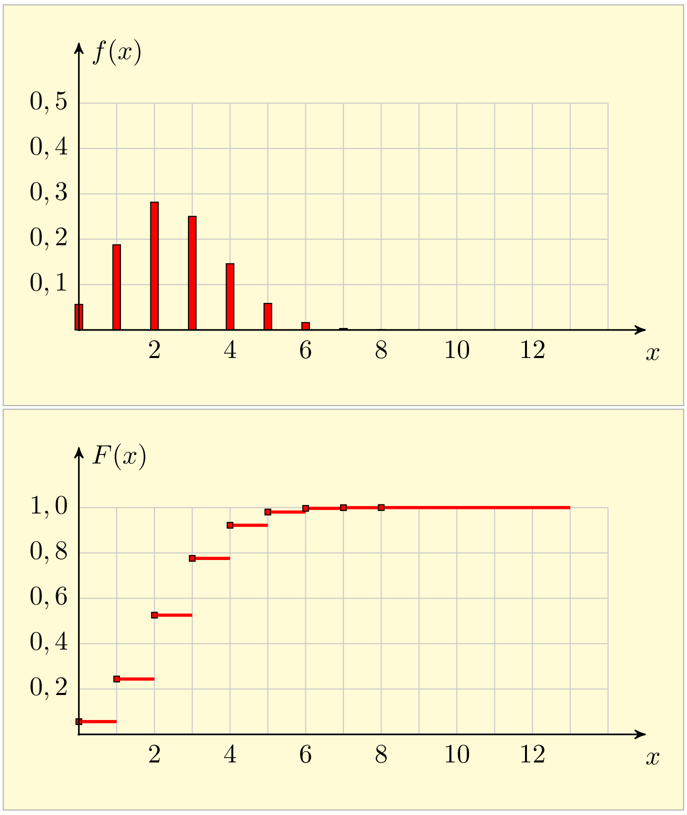
i	f_i	F_i
0	0,0010	0,0010
1	0,0098	0,0107
2	0,0439	0,0547
3	0,1172	0,1719
4	0,2051	0,3770
5	0,2461	0,6230
6	0,2051	0,8281
7	0,1172	0,9453
8	0,0439	0,9893
9	0,0098	0,9990
10	0,0010	1,0000
11	0,0000	1,0000
12	0,0000	1,0000

Abb. 11.5: *Binomialverteilung mit $n = 10$ und $p = 0,50$*

Die Wahrscheinlichkeitsfunktion (11.22) der Binomialverteilung ist für $p = 0,5$ symmetrisch bezüglich des Erwartungswerts. Für $p \neq 0,5$ gilt dies nicht mehr, wie Abbildung 11.6 beispielhaft illustriert. Die Wahrscheinlichkeitsfunktion weist hier links vom Erwartungswert $\mu = 2,5$ einen steileren Verlauf auf.

In Tabelle 19.1 des Anhangs sind Verteilungsfunktionen von Binomialverteilungen für $n = 1, 2, \ldots, 20$ und $p = 0,05, 0,10, \ldots, 0,50$ tabelliert. Werte der Wahrscheinlichkeits- und der Verteilungsfunktion einer Binomialverteilung und auch anderer Verteilungen lassen sich auch mit jedem gängigen Statistiksoftwarepaket, z. B. SPSS, und auch mit EXCEL oder der kostenfreien Statistiksoftware *R* berechnen – bezüglich *R* vgl. z. B. die umfassende Einführung von Gross (2010). [5]

[5] Bei Verwendung von SPSS findet man die Wahrscheinlichkeitsfunktion und die Verteilungsfunktion der Binomialverteilung sowie anderer Verteilungen im Menü „Transformieren / Variable berechnen". Wahrscheinlichkeitsfunktion und Verteilungsfunktion der Binomialverteilung sind dort mit PDF.BINOM(..) resp. mit CDF.BINOM(..) abgekürzt. Dabei steht „pdf" für „probability density function" und „cdf" für „cumulative density function". Auch in EXCEL findet man die gängigen Wahrscheinlichkeitkeitsverteilungen. Die Verteilungsfunktion der Binomialverteilung ist hier z. B. über „Einfügen / Funktion" unter BINOMVERT zugänglich.



x	$f(x)$	$F(x)$
0	0,0563	0,0563
1	0,1877	0,2440
2	0,2816	0,5256
3	0,2503	0,7759
4	0,1460	0,9219
5	0,0584	0,9803
6	0,0162	0,9965
7	0,0031	0,9996
8	0,0004	1,0000
9	0,0000	1,0000
10	0,0000	1,0000
11	0,0000	1,0000
12	0,0000	1,0000

Abb. 11.6: *Binomialverteilung mit* $n = 10$ *und* $p = 0,25$

Beispiel 11.2 **Berechnung von Wahrscheinlichkeiten via Binomialverteilung**

Wenn man eine Münze n-mal wirft, so ist die Anzahl X der Ereignisse „Zahl“ eine $B(n, p)$-verteilte Zufallsvariable. Der Erwartungswert ist hier durch $\mu = np = \frac{n}{2}$ und die Varianz durch $V(X) = np(1-p) = \frac{n}{4}$ gegeben. Bei Verwendung einer „fairen“ Münze, also einer Münze mit gleichen Eintrittswahrscheinlichkeiten für „Zahl (Z)“' und „Kopf (K)“, gilt $p = 0,5$. Die Wahrscheinlichkeit $P(X \leq 2)$ dafür, bei 3 Würfen *höchstens* 2-mal den Ausgang „Zahl“ zu erhalten, ist dann durch den Wert $F(2)$ der Verteilungsfunktion der Binomialverteilung mit $n = 3$ und $p = 0,5$ gegeben, nach Tabelle 19.1 also durch $F(2) = 0,875$. Die Wahrscheinlichkeit $f(2) = P(X = 2)$ dafür, bei den drei Würfen *genau* zweimal „Zahl“ zu erzielen, errechnet sich als Differenz der Funktionswerte $F(2) = P(X \leq 2) = 0,875$ und $F(1) = P(X \leq 1) = 0,500$, also als $0,375$. Der letztgenannte Wert wurde auch schon in Beispiel 10.2 elementar unter Verwendung des Laplace-Ansatzes (10.5) über die Kombinatorik abgeleitet.

Bei größeren Werten n werden aber kombinatorische Überlegungen recht aufwändig, insbesondere, wenn ein Wert $p \neq 0,5$ ins Spiel kommt. Zieht man etwa aus einer Lostrommel, in der ein Anteil p Gewinne und ein Anteil von $1-p$ Nieten sind, nacheinander n Lose und legt nach jeder Einzelziehung das Los in die Trommel zurück, so ist die Wahrscheinlichkeit nach 20 Ziehungen genau 4 Gewinne gezogen zu haben, offenbar errechenbar als Differenz $F(4) - F(3)$ von zwei Werten der Verteilungsfunktion einer $B(20; 0,05)$-

Aufgabe 11.2

verteilten Zufallsvariablen. Man erhält im Falle $p = 0,05$ mit Tabelle 19.1 den Wert $0,9974 - 0,9841 = 0,0133$. Für die Wahrscheinlichkeit dafür, im Falle $n = 20$ und $p = 0,05$ mindestens 4 Gewinne zu ziehen, ermittelt man den Wert $1 - F(3) = 0,0159$.

Exkurs 11.1 Fiasko beim Zentralabitur NRW 2008

In Nordrhein-Westfalen wurde 2008 zum zweiten Male ein Zentralabitur organisiert. Dabei gab es hinterher erhebliche Kritik, u. a. an einer Aufgabe zur Wahrscheinlichkeitsrechnung, die für Schüler der Mathematik-Leistungskurse konzipiert war. Die Kritik ging durch alle Medien und führte am Ende zu einem Rücktritt des Pressesprechers der zuständigen Schulministerin in Düsseldorf und für die betroffenen Schüler zum Angebot der Wiederholung der gesamten Mathematikprüfung.

Die umstrittene Aufgabe bezog sich auf Trefferquoten von D. Nowitzki, Mannschaftsführer der deutschen Basketball-Mannschaft bei der Olympiade 2008 in Peking. Der erste Teil der Aufgabe lautete wie folgt:[6]

Der deutsche Basketball-Profi Dirk Nowitzki spielte in der amerikanischen Profiliga beim Club Dallas Mavericks. In der Saison 2006/07 erzielte er bei Freiwürfen eine Trefferquote von 90,4 Prozent. Berechnen Sie die Wahrscheinlichkeit dafür, dass er

(1) genau 8 *Treffer bei* 10 *Versuchen erzielt,*
(2) höchstens 8 *Treffer bei* 10 *Versuchen erzielt,*
(3) höchstens viermal nacheinander bei Freiwürfen erfolgreich ist.

Basketballer Nowitzki (Quelle: dpa)

An der Aufgabe lässt sich sehr gut verdeutlichen, wie wichtig es ist, klar zwischen empirischen Befunden (Datenebene) und Modellansätzen zur approximativen Beschreibung solcher Befunde (Modellebene) zu unterscheiden. Die Analogie zum Münzwurfexperiment liegt auf der Hand; auch bei einem Freiwurf gibt es zwei mögliche Ausgänge (Korb wird getroffen / verfehlt). Sei zunächst erneut die Situation beim Münzwurfexperiment betrachtet. Man kann hier davon ausgehen, dass die Wahrscheinlichkeit p für den Eintritt des interessierenden Ereignisses, etwa „Zahl“, sich nicht ändert, wenn man eine Münze n-mal wirft. Bei einer „fairen“ Münze ist $p = 0,5$. Ob eine Münze wirklich fair ist, d. h. gleiche Eintrittswahrscheinlichkeiten für „Zahl“ und „Kopf“ aufweist, kann man anhand einer Bernoulli-Kette mit größerem n testen. Abbildung 11.4 zeigt zwei Bernoulli-Ketten, die die Vermutung einer fairen Münze zumindest visuell stützen. In der Regel wird man aufgrund von Symmetrieüberlegungen ohne nähere Prüfung von der Annahme $p = 0,5$ ausgehen. Wenn man bei einem Münzwurf Anhaltspunkte dafür hat, dass die Wahrscheinlichkeit p für „Zahl“ nicht der Bedingung $p = 0,5$ genügt, kann man p schätzen (vgl. hierzu Kapitel 13) und den Schätzwert $\widehat{p}$ heranziehen. In beiden Fällen – faire oder nicht-faire Münze – ist die Eintrittswahrscheinlichkeit für „Zahl“ von Wurf zu Wurf eine feste Größe und für die Anzahl X der „Treffer“ (Beobachtung von „Zahl“) gilt $X \sim B(n, p)$ bzw. $X \sim B(n, \widehat{p})$.

Wenn die Trefferquote des Basketballers Nowitzki in der Saison 2006/07 als repräsentativ für seine Leistung angesehen werden darf, müsste man dies in der Aufgabe durch die explizite Annahme einer festen Trefferquote von $0,904$ zum Ausdruck bringen. Eine konstante

[6] Die vollständige Fassung der Aufgabe wurde im Juni 2008 von *Spiegel online* ins Netz gestellt.

Trefferquote ist aber keineswegs selbstverständlich; gerade im Sport sind größere Formschwankungen an der Tagesordnung. Ohne die Annahme einer festen Trefferquote p resp. $\widehat{p}$ ist die Information über die Trefferquote in der letzten Saison lediglich ein – möglicherweise einmaliger – empirischer Befund, der nicht ausreicht für die Berechnung der verlangten Trefferwahrscheinlichkeiten. Das Modell der Binomialverteilung setzt ja voraus, dass die n in (11.19) eingehenden Indikatorvariablen X_i (mit dem Wert 1 bei Eintritt des Ereignisses „Treffer" und 0 beim Ereignis „kein Treffer") alle derselben Bernoulli-Verteilung (11.4) folgen.

Würde man allerdings die Aufgabe um die Annahme einer konstanten Trefferquote $0,904$ ergänzen, wäre die Wahrscheinlichkeit für die Erzielung von genau 8 Treffern bzw. von höchstens 8 Treffern bei 10 Wurfversuchen durch den Wert $f(8)$ der Wahrscheinlichkeitsfunktion $f(x)$ resp. den Wert $F(8)$ der Verteilungsfunktion $F(x)$ einer $B(10; 0,904)$-verteilten Zufallsvariablen gegeben. Man würde dann z. B. für $f(8)$ nach (11.22)

$$f(8) = \binom{10}{8} \cdot 0,904^8 \cdot 0,096^2 = 45 \cdot 0,4460129 \cdot 0,009216 \approx 0,185.$$

errechnen. Die Wahrscheinlichkeit für die Erzielung von vier Treffern in Folge lässt sich allerdings auch bei Annahme einer festen Trefferquote noch nicht beantworten, weil in Aufgabenteil (c) die Gesamtzahl n der Würfe nicht angegeben ist, von der das Ergebnis abhängt. Die Aufgabe (c) ist also eigentlich nicht lösbar. Unterstellt man, dass auch hier $n = 10$ gemeint war – die amtliche Musterlösung des Ministeriums ging allerdings ohne nachvollziehbaren Grund von $n = 5$ aus – und codiert man „Treffer" mit „1" und das Komplementärereignis „kein Treffer" mit „0", hätte man aus den insgesamt $2^{10} = 1024$ möglichen Ergebnisfolgen diejenigen heraus zu suchen, bei denen nie mehr als vier Einsen in Folge erscheinen. Die Ergebnisfolge $(1, 0, 1, 1, 1, 1, 0, 0, 1, 1)$ wäre ein Beispiel für eine Ergebnisfolge, die dem Erfordernis „höchstens vier Treffer in Folge" genügt.

◇

11.4 Die hypergeometrische Verteilung

Die Binomialverteilung beschreibt das Zufallsverhalten der Zählvariablen X aus (11.19) bei einem n-fach durchgeführten Bernoulli-Experiment, wobei die einzelnen Experimente voneinander unabhängig sind. Die Zählvariable weist aus, wie häufig einer der beiden möglichen Ausgänge $x_1 = A$ und $x_2 = \overline{A}$ und $P(A) = p$ bzw. $P(\overline{A}) = 1 - p$ innerhalb der Bernoulli-Kette auftrat. Als Beispiele wurden Münzwurf- oder auch Würfelexperimente angeführt, wenn man bei letzteren nur zwischen zwei Ausgängen differenziert (etwa „gerade / ungerade Augenzahl"). Die Grundsituation lässt sich anhand des Urnenmodells beschreiben. Eine Urne (Behälter) enthalte eine Menge roter und schwarzer Kugeln. Der Urne wird n-mal eine Kugel entnommen und man zählt die Anzahl X der roten Kugeln. Nach jeder Ziehung wird die entnommene Kugel in die Urne zurückgelegt. Der Quotient „Anzahl roter Kugeln / Anzahl aller Kugeln", der die Wahrscheinlichkeit für die Entnahme einer roten Kugel bestimmt, bleibt hier von Ziehung zu Ziehung konstant. Die Binomialverteilung lässt sich also anschaulich durch das **Urnenmodell mit Zurücklegen** veranschaulichen. Dieses Modell ist z. B. beim wiederholten Münzwurf passend, weil die Ausgangslage sich nicht von Wurf zu Wurf verändert. Es ist so, als ob man einer Urne, die zwei Zettel mit der Aufschrift „Zahl" bzw. „Kopf" enthält, jeweils einen Zettel entnimmt und den gezogenen Zettel vor der nächsten Ziehung zurücklegt.

Varianten des Urnenmodells

In der Realität gibt es Situationen, bei denen das beschriebene Modell des Ziehens mit Zurücklegen nicht oder nur näherungsweise passt – man denke nur an die Ziehung der Lottozahlen oder an Befragungen von Personen auf der Basis zufälliger Stichproben. Auch in der Wareneingangsprüfung bei einem Unternehmen wird man bei Entnahme einer Stichprobe von n Elementen aus einem Warenlos ein entdecktes nichtspezifikationskonformes Element vor der Entnahme eines weiteren Elements nicht zurücklegen. In solchen Fällen wird das **Urnenmodell ohne Zurücklegen** verwendet.

Flash-Animation „Hypergeometrische Verteilung“

Wenn man einer Urne mit N Kugeln, von denen M rot und die restlichen $N - M$ schwarz sind, nacheinander n Kugeln ohne Zurücklegen entnimmt, so repräsentiert die Ziehung jeder Kugel zwar weiterhin ein Bernoulli-Experiment, die Einzelexperimente sind aber nicht mehr voneinander unabhängig. Die Eintrittswahrscheinlichkeit für das interessierende Ereignis „Kugel ist rot“ wird jetzt nicht nur von M, sondern auch vom Umfang N der Grundgesamtheit beeinflusst. Die Verteilung der durch (11.19) definierten Zählvariablen X ist bei Annahme einer Stichprobenentnahme ohne Zurücklegen nicht mehr durch eine Biomialverteilung gegeben, sondern durch die **hypergeometrische Verteilung**. Letztere ist durch drei Parameter beschrieben, nämlich durch N, M und n, und man schreibt hierfür $X \sim H(n; M; N)$ (lies: X *ist hypergeometrisch verteilt mit den Parametern* n, M *und* N). Erwartungswert $\mu = E(X)$ und Varianz $\sigma^2 = V(X)$ der hypergeometrischen Verteilung seien hier der Vollständigkeit ohne Beweis angegeben.[7]

Kenngrößen der hypergeometrischen Verteilung

$$\mu = n \cdot \frac{M}{N} \tag{11.24}$$

$$\sigma^2 = n \cdot \frac{M}{N}(1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}. \tag{11.25}$$

Erwartungswert und Varianz einer $H(n, M, N)$-verteilten Zufallsvariablen X stimmen nach (11.20) und (11.21) offenbar mit dem Erwartungswert bzw. der Varianz einer $B(n; p)$-verteilten Variablen mit $p = \frac{M}{N}$ überein – mit dem einzigen Unterschied, dass bei der Varianz der Binomialverteilung der in (11.25) auftretende Bruchterm $\frac{N-n}{N-1}$ fehlt. Da dieser Term für $n > 1$ kleiner als 1 ist (für $n = 1$ ist er 1), hat die hypergeometrische Verteilung im Vergleich zur Binomialverteilung eine kleinere Varianz, wobei die Unterschiede mit wachsendem N vernachlässigbar werden. Dass die hypergeometrische Verteilung eine kleinere Varianz aufweist, ist einleuchtend, denn beim Ziehen ohne Zurücklegen wird die in einem gezogenen Stichprobenelement steckende Information (Kugel ist rot oder schwarz) nicht immer wieder verschenkt, d. h. es gibt weniger Unsicherheit über den verbleibenden Inhalt der Urne im Vergleich zum Ziehen mit Zurücklegen. Im Extremfall der sukzessiven Ziehung aller in der Urne befindlichen Elemente ($n = N$) ohne Zurücklegen liegt vollständige Information über den Urneninhalt vor. Die Zählvariable X ist dann keine Zufallsvariable mehr, sondern eine deterministische Größe mit dem Wert M. Man erkennt den nicht-stochastischen Charakter von X im Falle $n = N$ auch aus der Varianzformel (11.25), denn es gilt dann $\frac{N-n}{N-1} = 0$ und somit $V(X) = 0$.

Die Angabe der **Trägermenge** einer $H(n; M; N)$-verteilten Zufallsvariablen, also der Menge der möglichen Ausprägungen der Zählvariablen X, ist nicht trivial. Sie ist durch

[7] Eine Herleitung von Erwartungswert, Varianz und auch der Wahrscheinlichkeitsfunktion $f(x)$ der hypergeometrischen Verteilung findet man z. B. bei Mosler / Schmid (2011, Abschnitt 2.3.4).

$T = \{x_{min}, \ldots, x_{max}\}$ gegeben mit $x_{min} = max(0; n - N + M)$ als dem kleinsten und $x_{max} = min(n; M)$ als dem größten Element der Trägermenge (s. hierzu den Exkurs am Ende dieses Abschnitts). Die **Wahrscheinlichkeitsfunktion** $f(x) = P(X = x)$ der hypergeometrischen Verteilung ist ebenfalls nicht ganz so einfach ableitbar wie die der Binomialverteilung. Es gilt die Darstellung

Vollständige Beschreibung der hypergeometrischen Verteilung

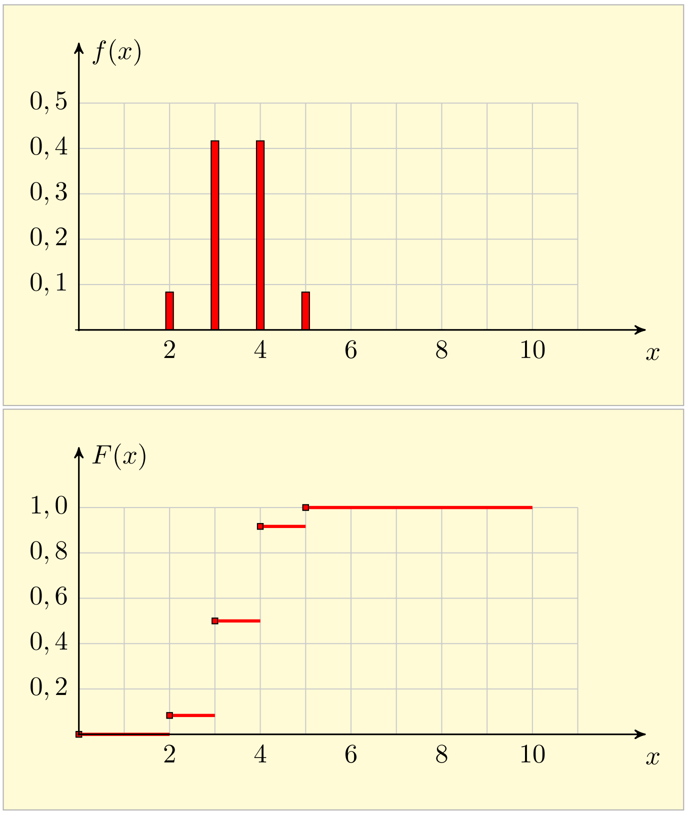
$$f(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{für } x \in T \\ 0 & \text{für alle sonstigen } x, \end{cases} \tag{11.26}$$

Java-Applet „Die hypergeometrische Verteilung“

deren Herleitung im Exkurs am Ende dieses Abschnitts skizziert wird. Für die **Verteilungsfunktion** $F(x) = P(X \leq x)$ gilt dann auf der Trägermenge

$$F(x) = \sum_{k=0}^{x} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad x \in T. \tag{11.27}$$

Da die Wahrscheinlichkeitsfunktion für $x \notin T$ stets 0 ist, bleibt $F(x)$ zwischen zwei benachbarten Elementen der Trägermenge auf dem Niveau des kleineren Werts, um dann in $x_{max} = min(n; M)$ den Endwert 1 anzunehmen (Treppenfunktion).



x	$f(x)$	$F(x)$
0	0,0000	0,0000
1	0,0000	0,0000
2	0,0833	0,0833
3	0,4167	0,5000
4	0,4167	0,9167
5	0,0833	1,0000
6	0,0000	1,0000
7	0,0000	1,0000
8	0,0000	1,0000
9	0,0000	1,0000
10	0,0000	1,0000

Abb. 11.7: *Hypergeometrische Verteilung mit $n = 5$, $M = 7$ und $N = 10$*

Abbildung 11.7 veranschaulicht die Wahrscheinlichkeits- und die Verteilungsfunktion einer $H(5; 7; 10)$-verteilten Zufallsvariablen. Der Erwartungswert $\mu = E(X)$ errechnet sich hier als $\mu = 5 \cdot \frac{7}{10} = 3,5$. Neben den Graphen sind einige Werte der beiden

Funktionen tabelliert. Die Trägermenge T der dargestellten hypergeometrischen Verteilung ist durch $T = \{x_{min}, \ldots, x_{max}\}$ gegeben mit $x_{min} = max(0; 5 - 10 + 7) = 2$ und $x_{max} = min(5; 7) = 5$. Der Abbildung 11.7, insbesondere der letzten Spalte der Tabelle, entnimmt man z. B., dass die Verteilungsfunktion $F(x)$ an der Stelle $x_{max} = 5$ auf den Endwert 1 springt.

Approximation der hypergeometrischen Verteilung

Die Tabellierung der Wahrscheinlichkeits- oder der Verteilungsfunktion ist für die hypergeometrische Verteilung viel aufwändiger als bei der Binomialverteilung, weil die Tabellen hier von drei Parametern abhängen. Im vorliegenden Manuskript wird aus diesem Grunde auf eine Tabellierung verzichtet. In der Praxis wendet man anstelle der hypergeometrischen Verteilung meist die einfacher handhabbare Binomialverteilung an, wenn der Umfang N der Grundgesamtheit im Vergleich zum Umfang der Stichprobe n groß ist (Faustregel: $\frac{n}{N} < 0,05$). In diesem Falle kann man für eine $H(n; M; N)$-verteilte Zufallsvariable X in guter Näherung annehmen, dass sie $B(p; n)$-verteilt ist mit $p = \frac{M}{N}$. Die Tragfähigkeit der Approximation liegt darin begründet, dass die Unterschiede zwischen den Situationen „Ziehen ohne / mit Zurücklegen“ mit Verkleinerung des Auswahlsatzes $\frac{n}{N}$ immer weniger ins Gewicht fallen.

Bernoulli-Verteilung als Spezialfall

Die Binomialverteilung und die hypergeometrische Verteilung charakterisieren also beide das Zufallsverhalten der Indikatorvariablen (11.19), allerdings unter verschiedenen Bedingungen. Die Indikatorvariable (11.19) zählt, wie oft bei n-facher Durchführung eines Bernoulli-Experiments (n-faches Ziehen einer Kugel aus einer Urne mit roten und schwarzen Kugeln) mit den möglichen Ausgängen $x_1 = A$ (Kugel ist rot) und $x_2 = \overline{A}$ eines der beiden Ereignisse, etwa A, beoachtet wird. Beim Ziehen *mit* Zurücklegen ist die Zählvariable binomialverteilt, beim Ziehen *ohne Zurücklegen* folgt sie einer hypergeometrischen Verteilung. Beide Verteilungen gehen im Sonderfall $n = 1$ offenbar in die Bernoulli-Verteilung über. Beim Ziehen einer einzigen Kugel aus einer Urne mit roten und schwarzen Kugeln und der Wahrscheinlichkeit p für das Ereignis A entfällt nämlich eine Unterscheidung von Ziehen mit oder ohne Zurücklegen und die Wahrscheinlichkeitsfunktion (11.4) beschreibt den Ausgang des einmaligen Bernoulli-Experiments.

Beispiel 11.3 Berechnung von Wahrscheinlichkeiten beim Lottospiel

Lotto wird in Europa nicht einheitlich gespielt. In Deutschland gibt es z. B. das Lottospiel „6 aus 49“, in der Schweiz „6 aus 45“ und in Italien “6 aus 90“. Die Wahrscheinlichkeiten für die Ereignisse „6 Richtige“, „0 Richtige“, „mindestens 4 Richtige“ o. ä. beim deutschen Lotto lassen sich anhand der hypergeometrischen Verteilung mit den Parametern $n = 6$, $M = 6$ und $N = 49$ berechnen. Dabei beinhaltet n hier die Anzahl der Kreuze auf dem Lottoschein (beim Urnenmodell die Anzahl der gezogenen Kugeln), M die maximale Anzahl der Treffer (beim Urnenmodell die Anzahl der „roten“ Kugeln in der Urne) und N die Anzahl der die Lottozahlen präsentierenden Kugeln in der Trommel (bzw. in der Urne). Der Erwartungswert für die Anzahl X der Richtigen beim Lottospiel „6 aus 49“ ist nach (11.24) durch $\mu = \frac{36}{49} \approx 0,735$ gegeben.

Für die Berechnung von Wahrscheinlichkeiten der Art „x Richtige“ oder „mindestens x Richtige“ ist der Rückgriff auf eine Tabelle mit Werten der Wahrscheinlichkeitsfunktion

$f(x) = P(X = x)$ oder der Verteilungsfunktion $F(x) = P(X \leq x)$ der genannten hypergeometrischen Verteilung am einfachsten. Wenn man nicht über eine solche Tabelle

Abb. 11.8: *„Lottofee" (ARD-Ziehung der Lottozahlen; Quelle: Hessischer Rundfunk)*

verfügt, kann man die gesuchten Wahrscheinlichkeiten direkt aus (11.26) bzw. aus (11.27) bestimmen. Für das Ereignis „0 Richtige" erhält man z. B. nach (11.26) mit Einsetzen von $n = 6$, $M = 6$ und $N = 49$ bei Beachtung von $\binom{6}{0} = 1$ die Darstellung

$$f(0) = \frac{\binom{6}{0}\binom{49-6}{6-0}}{\binom{49}{6}} = \frac{\binom{43}{6}}{\binom{49}{6}}.$$

Der Nennerterm, für den man mit (10.10) den schon in Beispiel 10.3 angegebenen Wert

$$\binom{49}{6} = \frac{49!}{43! \cdot 6!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13983816$$

ermittelt, repräsentiert die Anzahl der möglichen Ausgänge einer Lottoziehung. Für den Zählerterm, der die Anzahl der möglichen Ausgänge mit 0 Richtigen wiedergibt, folgt

$$\binom{43}{6} = \frac{43!}{37! \cdot 6!} = \frac{43 \cdot 42 \cdot 41 \cdot 40 \cdot 39 \cdot 38}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 6096454.$$

Die Wahrscheinlichkeit $f(0) = P(X = 0)$ für das Ereignis „0 Richtige" ist somit

$$f(0) = \frac{\binom{43}{6}}{\binom{49}{6}} = \frac{6096454}{13983816} \approx 0,436,$$

also ca. 43,6 % – ein Wert, der manchen Lottofreund irritieren dürfte. Die Wahrscheinlichkeit $f(6) = P(X = 6)$ für „6 Richtige" ließe sich analog bestimmen. Da allerdings von den 13983816 möglichen Ausgängen einer Lottoziehung nur ein einziger Ausgang „6 Richtige" beinhaltet, kann man $f(6) = P(X = 6)$ einfacher über

$$f(6) = \frac{1}{13983816} \approx 0,0000000715 = 7,15 \cdot 10^{-8}$$

errechnen. An Lottospieler, die 6 Richtige reklamieren können (Gewinnklasse 2), werden 8% der Lottoeinnahmen verteilt. Man sollte wissen, dass 50% der Lottoeinnahmen als Steuern abgeführt oder für festgelegte Zwecke verwendet und mithin gar nicht an Spieler verteilt werden.

Aufgabe 11.3

Die ohnehin schon verschwindend kleine Wahrscheinlichkeit für einen Volltreffer verringert sich beim deutschen Lotto noch um den Faktor 10 auf $\frac{1}{139838160} \approx 7,15 \cdot 10^{-9}$, wenn man das Spiel „6 aus 49 mit Superzahl" spielt. Die „Superzahl" ist eine Zusatzzahl, die aus der Menge $\{0, 1, ..., 8, 9\}$ gezogen wird („1 aus 10"). Um an den legendären Jackpot zu kommen (Gewinnklasse 1 mit einer Ausschüttungsquote von 10%), muss man „6 Richtige aus 49" haben *und* die korrekte Zusatzzahl „1 aus 10" vorweisen können. Diese Gewinnklasse ist, nicht überraschend, häufig gar nicht besetzt – die vorgesehenen Gewinne werden dann auf die nächste Ziehung übertragen. Noch geringer als ein Erreichen der Gewinnklasse 1 beim deutschen Lotto ist die Wahrscheinlichkeit eines Volltreffers bei der italienischen Lottovariante „6 aus 90". Sie entspricht dem Wert $f(6) = \frac{1}{622614630} \approx 1,61 \cdot 10^{-9}$ der Wahrscheinlichkeitsfunktion einer hypergeometrischen Verteilung mit $n = 6$, $M = 6$ und $N = 90$.

Exkurs 11.2 **Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung**

Um die Trägermenge einer $H(n; M; N)$-verteilten Zufallsvariablen zu bestimmen, sind nur die kleinst- und die größtmögliche Ausprägung der durch (11.19) erklärten Zählvariablen X im Urnenmodell ohne Zurücklegen zu ermitteln. Die Variable X, die sich hier als Anzahl der gezogenen roten Kugeln nach n Ziehungen interpretieren lässt, kann im Falle $n \leq M$ den Wert n offenbar nicht überschreiten – es können nicht mehr rote Kugeln gezählt als gezogen werden. Im Falle $n > M$ ist hingegen M die Obergrenze – es können nicht mehr rote Kugeln gezogen werden als in der Urne vorhanden sind. Das größte Element x_{max} der Trägermenge hat also den Wert $x_{max} = min(n; M)$. Ferner gilt, dass $n - (N - M)$ die Anzahl der roten Kugeln nach n Ziehungen darstellt, diese aber auch nicht kleiner als 0 sein kann, d. h. $x_{min} = max(0; n - N + M)$ definiert den kleinstmöglichen Wert.

Bei der Herleitung der Wahrscheinlichkeitsfunktion (11.26) kann man auf Tabelle (10.1) zurückgreifen, die Basisformeln der Kombinatorik ausweist. Der Nenner von (11.26) repräsentiert die Anzahl der Möglichkeiten, aus einer Urne mit N Kugeln insgesamt n Kugeln ohne Zurücklegen zu entnehmen. Nach Tabelle (10.1) ist diese Anzahl durch $\binom{N}{n}$ gegeben, weil es auf die Reihenfolge der Ergebnisse der Ziehungen hier nicht ankommt. Der Produktterm im Zähler von (11.26) ergibt sich aus folgender Überlegung: In der Urne befinden sich vor Beginn der Ziehung M rote und $N - M$ schwarze Kugeln. Es gibt $\binom{M}{x}$ Möglichkeiten, x rote Kugeln aus M roten Kugeln auszuwählen. Damit nach n Ziehungen ohne Zurücklegen die Anzahl der gezogenen roten Kugeln genau x ist, müssen aus dem Anfangsvorrat von $N - M$ schwarzen Kugeln $n - x$ schwarze Kugeln gezogen werden. Es gibt $\binom{N-M}{n-x}$ Möglichkeiten der Auswahl dieser $n - x$ Kugeln.

◇