4. Spezielle diskrete Verteilungen

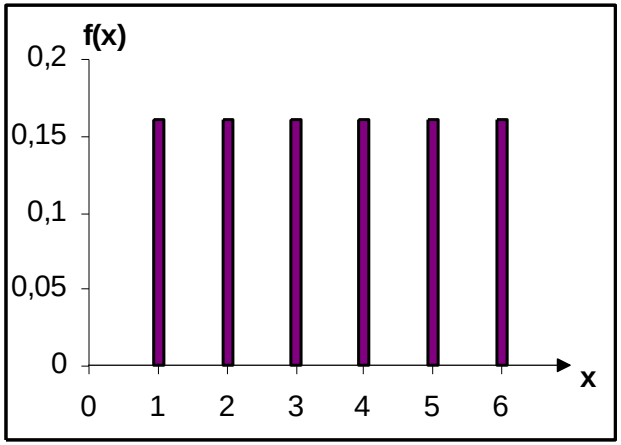
Bisher wurden Wahrscheinlichkeitsverteilungen in allgemeiner Form dargestellt. In der Praxis treten häufig ganz bestimmte Wahrscheinlichkeitsverteilungen auf.

4.1 Diskrete Gleichverteilung

Falls bei einer diskreten Zufallsvariablen X die Wahrscheinlichkeit für jeden Wert $x_1, x_2, \ldots, x_m$ gleich ist, gilt für ihre Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} 1/m & \text{für } x = x_1, \ldots, x_m \\ 0 & \text{sonst} \end{cases}$$

Diskrete Gleichverteilung (m = 6)



Diskret gleichverteilte Zufallsvariable X

Erwartungswert	Varianz
$E(X) = \frac{1}{m}\sum_{j=1}^{m} x_j$	$Var(X) = \frac{1}{m}\sum_{j=1}^{m}\left[x_j - E(X)\right]^2$
Bei natürlichen Zahlen 1, 2, … , m: $E(X) = \frac{m+1}{2}$	Bei natürlichen Zahlen 1, 2, … , m: $Var(X) = \frac{m^2 - 1}{12}$

Beweis für die Formel zum Erwartungswert bei natürlichen Zahlen 1, 2, … , m

Verwendung der Summenformel für eine arithmetische Reihe:

$$\sum_{j=1}^{m} j = \frac{m(m+1)}{2}$$

Daraus folgt:

$$E(X) = \frac{1}{m}\sum_{j=1}^{m} x_j = \frac{1}{m}\sum_{x=1}^{m} x = \frac{1}{m}\frac{m(m+1)}{2} = \frac{m+1}{2}$$

Beweis für die Formel zur Varianz bei natürlichen Zahlen 1, 2, … , m

Formel für Summe der quadrierten natürlichen Zahlen von 1 bis m:

$$\sum_{j=1}^{m} j^2 = \frac{m(m+1)(2m+1)}{6}$$

Daraus folgt:

$$E(X^2) = \frac{1}{m}\sum_{j=1}^{m} x_j^{\,2} = \frac{1}{m}\sum_{x=1}^{m} x^2 = \frac{1}{m}\frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2$$

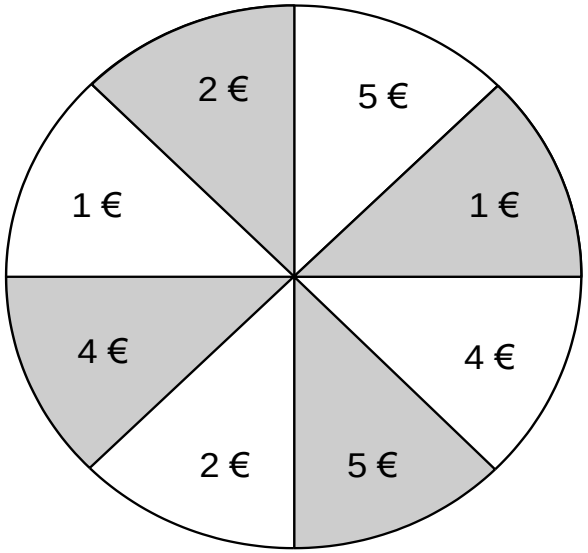
$$Var(X) = \frac{2(m+1)(2m+1)}{12} - \frac{m^2 + 2m + 1}{4}$$

$$= \frac{2(2m^2 + m + 2m + 1) - 3(m^2 + 2m + 1)}{12}$$

$$= \frac{4m^2 + 2m + 4m + 2 - 3m^2 - 6m - 3}{12} = \frac{m^2 - 1}{12}$$

Beispiel:

Auf einem Straßenfest ist ein Glücksrad aufgebaut, bei dem jede Besucher*in mit einem Einsatz von 3 € einen bestimmten Gewinn (= Auszahlung minus Einsatz) erzielt. Das Glücksrad ist in acht Felder eingeteilt, die mit der jeweiligen Auszahlung beschriftet sind:

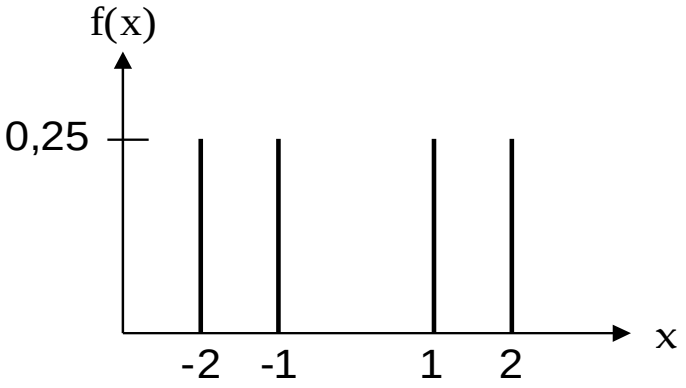


2 von 8 gleich große Sektoren führen zu einer Auszahlung von 1 €. Der entsprechende Gewinn mit einer Wahrscheinlichkeit von 2/8 = 1/4 beträgt:

Gewinn = Auszahlung - Einsatz = 1 - 3 = -2 [€]

Entsprechendes gilt für die übrigen Auszahlungen. Man erhält folgende Wahrscheinlichkeitsfunktion für die Zufallsvariable X, die den Gewinn darstellt:

$$f(x) = \begin{cases} 0{,}25 & \text{für} \quad x = -2 \\ 0{,}25 & \text{für} \quad x = -1 \\ 0{,}25 & \text{für} \quad x = 1 \\ 0{,}25 & \text{für} \quad x = 2 \\ 0 & \text{sonst} \end{cases}$$



Der Erwartungswert des Gewinns beträgt:

$$E(X) = \frac{1}{4}\sum_{j=1}^{4} x_j = \frac{1}{4}(-2 - 1 + 1 + 2) = \frac{1}{4}0 = 0\,[€]$$

Damit handelt es sich um ein „faires” Spiel. Die Varianz des Gewinns beträgt:

$$Var(X) = \frac{1}{4}\sum_{j=1}^{4}(x_j - 0)^2 = \frac{1}{4}\left[(-2-0)^2 + (-1-0)^2 + (1-0)^2 + (2-0)^2\right]$$

$$= \frac{1}{4}(4 + 1 + 1 + 4) = \frac{1}{4}10 = 2{,}5\,[€^2]$$

Die Standardabweichung des Gewinns beträgt:

$$\sigma = \sqrt{2{,}5} = 1{,}58\,[€]$$

4.2 Bernoulli-Verteilung

Ein Bernoulli-Prozess ist eine endliche oder abzählbar unendliche Folge von stochastisch unabhängigen Zufallsvorgängen bei denen nur ein Ereignis A oder dessen Komplementärereignis $\bar{A}$ eintreten können.

Die Wahrscheinlichkeiten der beiden Ereignisse werden bei einem Bernoulli-Prozess als konstant vorausgesetzt:

$P(A) = p$ und $P(\bar{A}) = 1 - p$

Beispiele für Bernoulli-Prozesse:

- Münzwurf: Die möglichen Ausgänge sind Kopf (A) und Zahl ($\bar{A}$)
- Ausspielen eines Würfels: Die geworfene Augenzahl kann gleich 1 (A) oder größer als 1 ($\bar{A}$) sein
- Produktionsprozess: Ein entnommenes Teil weist einen Defekt (A) oder keinen Defekt ($\bar{A}$) auf
- Dauer von Telefongesprächen: Ein Gespräch kann höchstens 5 Minuten (A) oder länger ($\bar{A}$) dauern

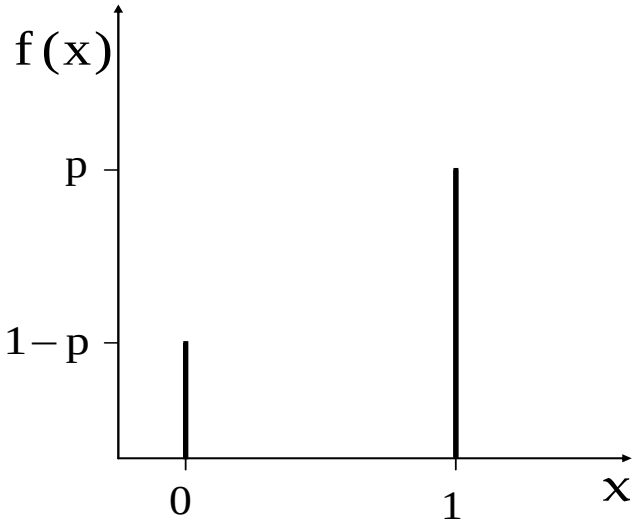
Bei der Bernoulli-Verteilung wird der Bernoulli-Prozess einmal durchgeführt.

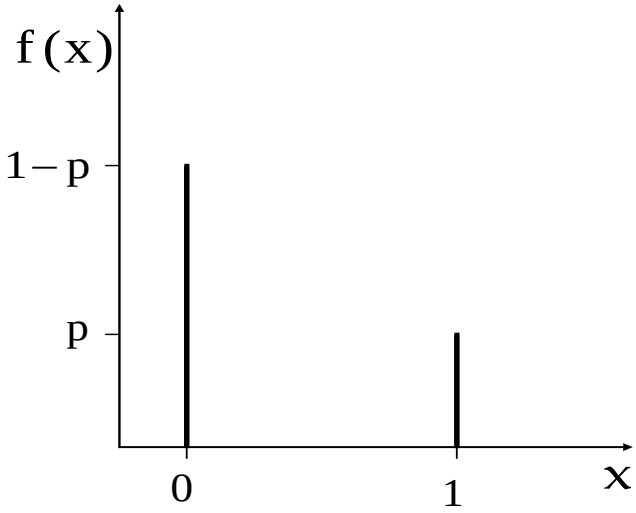
Die Zufallsvariable X, die angibt, wie oft das Ereignis A eingetreten ist, kann dann nur die Werte 0 und 1 annehmen. Für die Wahrscheinlichkeiten gilt:

$P(X = 1) = P(A) = p$ und $P(X = 0) = P(\bar{A}) = 1 - p$

Die Wahrscheinlichkeitsfunktion einer Bernoulli-verteilten Zufallsvariablen X lautet:

$$f(x) = \begin{cases} 1 - p & \text{für } x = 0 \\ p & \text{für } x = 1 \\ 0 & \text{sonst} \end{cases}$$





Bernoulli-verteilte Zufallsvariable X

Erwartungswert	Varianz
$E(X) = p$	$Var(X) = p(1 - p)$

Beweis für die Formeln zu Erwartungswert und Varianz:

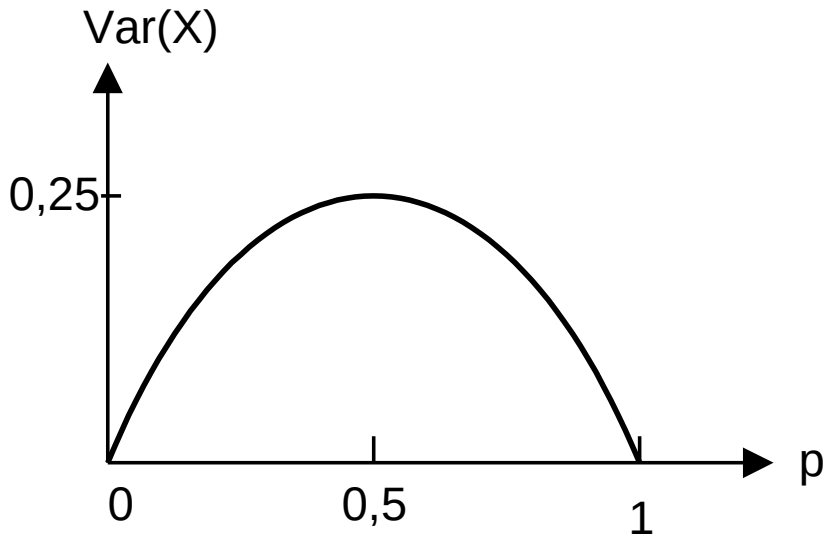
$$E(X) = \sum_{j=1}^{2} x_j f(x_j) = 0(1-p) + 1p = p$$

$$Var(X) = \sum_{j=1}^{2} \left[x_j - E(X)\right]^2 f(x_j) = \sum_{j=1}^{2} (x_j - p)^2 f(x_j)$$

$$= (0-p)^2 (1-p) + (1-p)^2 p = p^2 (1-p) + (1-p)^2 p$$

$$= p(1-p)\left[p + (1-p)\right] = p(1-p)$$

Varianz einer Bernoulli-verteilten Zufallsvariablen X



Die Varianz nimmt zunächst mit steigendem p zu, bis sie bei p = 0,5 ihr Maximum erreicht. Danach sinkt sie wieder mit zunehmendem p bis auf 0 ab.

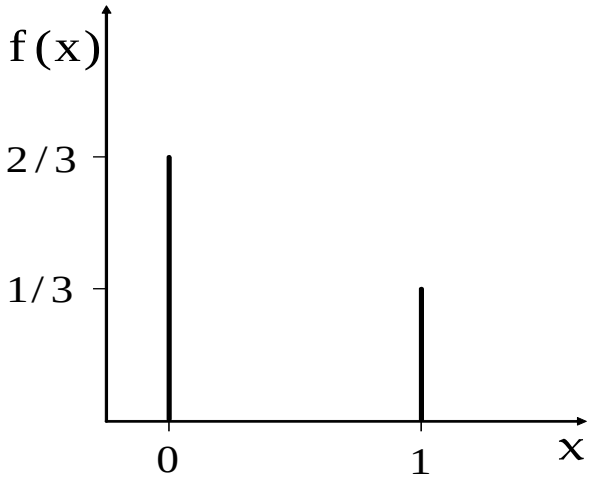
Beispiel:

Das Ereignis A sei als „Augenzahl kleiner oder gleich 2“ beim Ausspielen eines Würfels definiert. Die Wahrscheinlichkeit für A beträgt dann 1/3:

p = P(A) = 2/6 = 1/3

Die Wahrscheinlichkeitsverteilung der Zufallsvariablen X „Anzahl von A“ ist dann wie folgt gegeben:

$$f(x) = \begin{cases} 2/3 & \text{für } x = 0 \\ 1/3 & \text{für } x = 1 \\ 0 & \text{sonst} \end{cases}$$

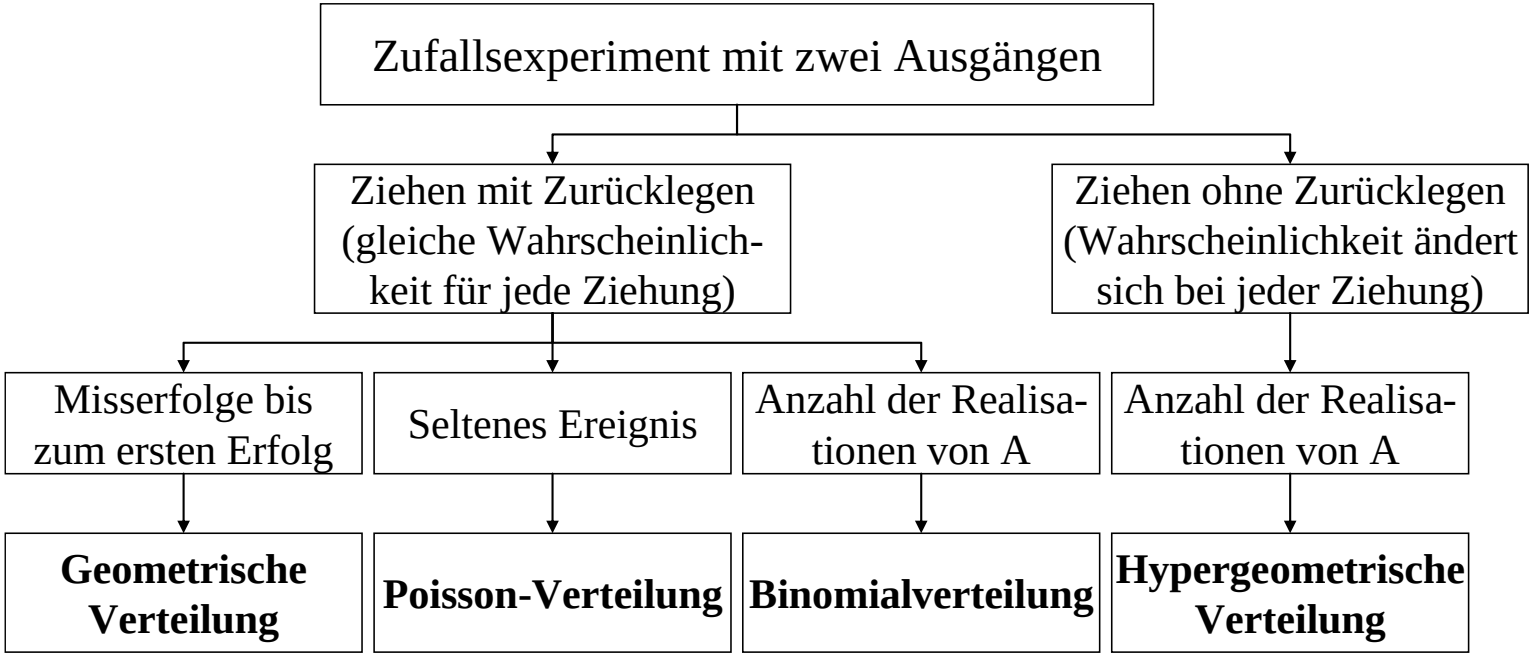


Der Erwartungswert und die Varianz von X betragen:

$$E(X) = p = \frac{1}{3} = 0{,}333$$

$$Var(X) = p(1-p) = \frac{1}{3}\left(1 - \frac{1}{3}\right) = \frac{1}{3}\frac{2}{3} = \frac{2}{9} = 0{,}222$$

Diskrete Verteilungen, die auf der Bernoulli-Verteilung basieren



Bei der hypergeometrischen Verteilung liegen aufgrund der Abhängigkeit der Ziehungen bei jeder Wiederholung unterschiedliche Bernoulli-Verteilungen vor.

→ Die geometrische Verteilung wird in diesem Semester nicht behandelt!

4.3 Binomialverteilung

Bernoulli-Prozess mit n-maliger Wiederholung

Die Binomialverteilung basiert auf einer n-maligen Durchführung eines Bernoulli-Prozesses, bei der in jedem Versuch nur das Ereignis A oder das Komplementärereignis $\overline{A}$ eintreten können.

Im Urnenmodell seien die beiden Ereignisse wie folgt definiert:

Ereignis A: rote Kugel und Ereignis $\overline{A}$: nicht-rote Kugel

Die Ziehungen erfolgen dabei mit Zurücklegen, so dass sie unabhängig voneinander sind. Dadurch ist sichergestellt, dass die Wahrscheinlichkeiten der beiden betrachteten Ereignisse in jedem Versuch unverändert bleiben:

$P(A) = p$ und $P(\overline{A}) = 1 - p$

Definition einer binomialverteilten Zufallsvariablen X:

Anzahl der Realisationen des Ereignisses A (Urnenmodell: Anzahl der gezogenen roten Kugeln)

Mögliche Werte der Zufallsvariablen X: 0, 1, 2, …, n

X = 0: Keine rote Kugel wird gezogen, d.h. A tritt 0-mal auf
X = n: Alle gezogenen Kugeln sind rot, d.h. A tritt n-mal auf

Herleitung der Wahrscheinlichkeitsfunktion:

Wenn das Ereignis A x-mal in der Stichprobe enthalten ist, dann muss das Komplementärereignis $\bar{A}$ (n-x)-mal eingetreten sein. Eine mögliche Stichprobe vom Umfang n, in der x-mal A und (n-x)-mal $\bar{A}$ enthalten sind, ist also z.B.:

$$\underbrace{A\ A \ldots A}_{\text{x-mal}} \qquad \underbrace{\bar{A}\ \bar{A} \ldots \bar{A}}_{\text{n-x-mal}}$$

Die Wahrscheinlichkeit für das Eintreten dieser konkreten Stichprobe, in der x-mal das Ereignis A und dann (n-x)-mal das Ereignis $\bar{A}$ enthalten sind, ist folgendermaßen gegeben:

$$P\left(\underbrace{A \cap A \cap \ldots \cap A}_{\text{x-mal}} \quad \cap \quad \underbrace{\bar{A} \cap \bar{A} \cap \ldots \cap \bar{A}}_{(\text{n-x})\text{-mal}}\right)$$

Aufgrund der Unabhängigkeit der Ziehungen kann der Multiplikationssatz für unabhängige Ereignisse angewandt werden. Somit erhält man:

$$P\left(A \cap A \cap \ldots \cap A \cap \bar{A} \cap \bar{A} \cap \ldots \cap \bar{A}\right) = \underbrace{P(A)P(A)\cdots P(A)}_{\text{x-mal}}\ \underbrace{P(\bar{A})P(\bar{A})\cdots P(\bar{A})}_{(\text{n-x})\text{-mal}}$$

$$= \underbrace{p \cdot p \cdots p}_{\text{x-mal}}\ \underbrace{(1\text{-}p)(1\text{-}p)\cdots(1\text{-}p)}_{(\text{n-x})\text{-mal}} = p^x (1\text{-}p)^{\text{n-x}}$$

Diese Wahrscheinlichkeit bezieht sich zunächst auf eine Stichprobe, in der A die ersten x und $\overline{A}$ die nächsten n-x Positionen belegt. Wie viele Stichproben des x-maligen Eintreffens von A bei n Zufallsexperimenten gibt es? Nach den Regeln der Kombinatorik erhalten wir die gesuchte Anzahl unterschiedlicher Stichproben mit folgender Formel (für Permutationen mit Wiederholung):

$$P_{n,x} = \frac{n!}{x!(n-x)!} = \binom{n}{x}$$

Es gibt also $\binom{n}{x}$ mögliche Stichproben, die an beliebigen Positionen x rote und n-x nicht-rote Kugeln enthalten. Jede dieser Stichproben hat dabei die gleiche Wahrscheinlichkeit von $p^x(1-p)^{n-x}$, sich zu realisieren.

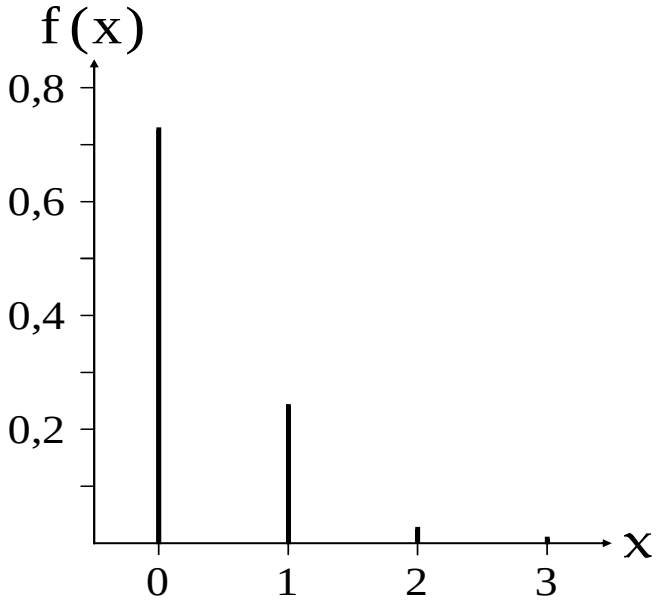
Da die Stichproben disjunkt sind, gilt für die Wahrscheinlichkeit, dass bei n unabhängigen Durchführungen eines Zufallsvorgangs x-mal das Ereignis A und (n-x)-mal das Ereignis $\overline{A}$ eintritt:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$$

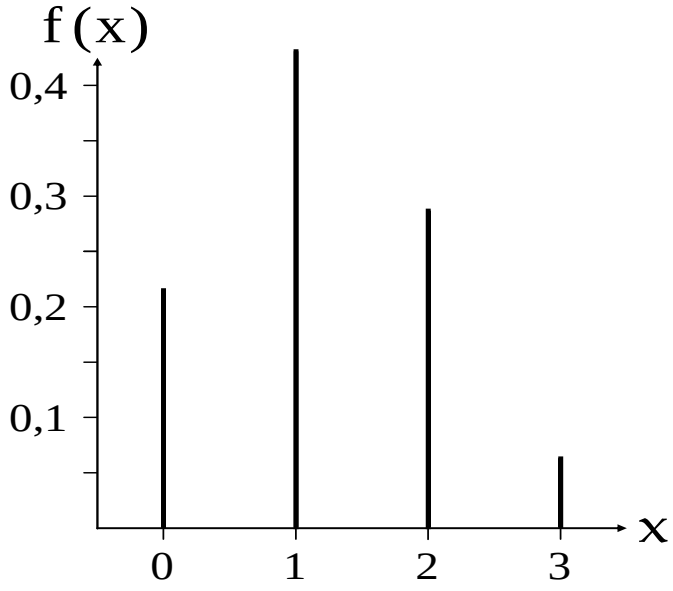
Eine Zufallsvariable X folgt einer Binomialverteilung mit den Parametern n und p, wenn die Wahrscheinlichkeitsfunktion von X folgendermaßen gegeben ist:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{sonst} \end{cases}$$

Wahrscheinlichkeitsfunktionen von binomialverteilten Zufallsvariablen X mit n = 3

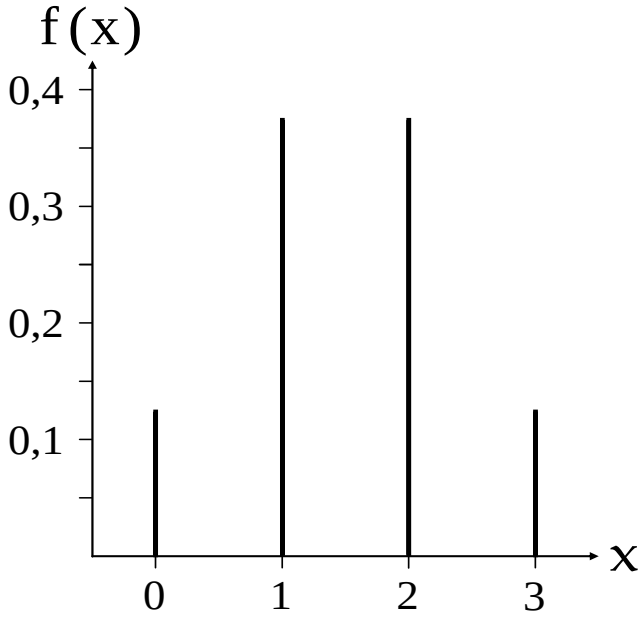


p = 0,1

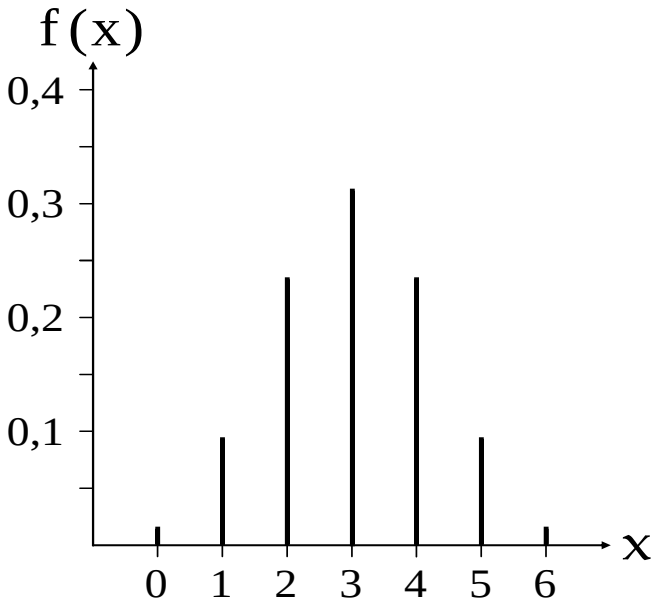


p = 0,4

Wahrscheinlichkeitsfunktionen von binomialverteilten Zufallsvariablen X mit p = 0,5



n = 3



n = 6

Binomialverteilte Zufallsvariable X

Erwartungswert	Varianz
$E(X) = n \cdot p$	$Var(X) = n \cdot p \cdot (1-p)$

Beweis für die Formeln zu Erwartungswert und Varianz

Für den Beweis wird X als Summe von n Bernoulli-verteilten Zufallsvariablen X_i mit $E(X_i) = p$ betrachtet:

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$$E(X) = \sum_{i=1}^{n} p = np$$

Aufgrund der Unabhängigkeit der X_i lässt sich die Varianz von X folgendermaßen darstellen:

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

$$Var(X) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Beispiel:

Die Zufallsvariable X bezeichne die Anzahl der geworfenen Köpfe beim dreimaligen Werfen einer Münze (d.h. n = 3). Da beide Seiten der Münze mit gleicher Wahrscheinlichkeit geworfen werden, beträgt p = 0,5. Wie groß ist die Wahrscheinlichkeit, dass 2 x Kopf geworfen wird?

Unter Verwendung der Wahrscheinlichkeitsfunktion der Binomialverteilung erhält man:

$$P(X = 2) = f(2) = \binom{3}{2} 0{,}5^2 0{,}5^1 = \frac{3!}{2!1!} 0{,}25 \cdot 0{,}5 = 0{,}375$$

Der Erwartungswert und die Varianz der betrachteten binomialverteilten Zufallsvariablen betragen:

$$E(X) = np = 3 \cdot 0{,}5 = 1{,}5$$

$$\mathrm{Var}(X) = np(1\text{-}p) = 3 \cdot 0{,}5 \cdot 0{,}5 = 0{,}75$$

Beispiel:

Bei der Produktion eines Gutes sind 10% Ausschuss. Wie groß ist die Wahrscheinlichkeit, dass von 20 produzierten Stücken mindestens 4 Ausschuss sind?

Auch hier ist die Binomialverteilung anzuwenden, weil ein Bernoulli-Prozess mit n = 20 Wiederholungen vorliegt. Die Wahrscheinlichkeit für ein Ausschussstück ist bei jeder Ziehung konstant und beträgt 0,1 (= p). Ob ein defektes oder brauchbares Gut ausgewählt wurde, verändert nicht die Wahrscheinlichkeit dafür, dass das nächste entnommene Teil Ausschuss ist. Die Zufallsvariable X ist als Anzahl der Ausschussstücke definiert.

Damit lautet die Wahrscheinlichkeitsfunktion der Zufallsvariablen X:

$$f(x) = \begin{cases} \binom{20}{x} 0{,}1^{x}\, 0{,}9^{20-x} & \text{für } x = 0, 1, 2, \ldots, 20 \\ 0 & \text{sonst} \end{cases}$$

Daraus folgt:

$$P(X \geq 4) = f(4) + f(5) + f(6) + f(7) + \ldots + f(18) + f(19) + f(20)$$

Einfachere Berechnung mit Hilfe der Komplementärwahrscheinlichkeit:

$$P(X \geq 4) = 1 - P(X < 4)$$
$$= 1 - \left[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\right]$$

Für die Einzelwahrscheinlichkeiten gilt:

$$P(X = 0) = f(0) = \binom{20}{0} 0{,}1^0 0{,}9^{20} = \frac{20!}{0!20!} 0{,}1^0 0{,}9^{20} = 0{,}122$$

$$P(X = 1) = f(1) = \binom{20}{1} 0{,}1^1 0{,}9^{19} = \frac{20!}{1!19!} 0{,}1^1 0{,}9^{19} = 0{,}270$$

$$P(X = 2) = f(2) = \binom{20}{2} 0{,}1^2 0{,}9^{18} = \frac{20!}{2!18!} 0{,}1^2 0{,}9^{18} = 0{,}285$$

$$P(X = 3) = f(3) = \binom{20}{3} 0{,}1^3 0{,}9^{17} = \frac{20!}{3!17!} 0{,}1^3 0{,}9^{17} = 0{,}190$$

Daraus folgt:

$$P(X \geq 4) = 1 - (0{,}122 + 0{,}270 + 0{,}285 + 0{,}190)$$
$$= 1 - 0{,}867 = 0{,}133$$

Verteilungsfunktion

Für die Verteilungsfunktion F(x) einer binomialverteilten Zufallsvariablen X gibt es keinen kompakten Ausdruck. Man erhält sie durch Kumulierung der Wahrscheinlichkeiten f(y) mit $y \leq x$:

$$F(x) = P(X \leq x) = \sum_{y \leq x} \binom{n}{y} p^y (1-p)^{n-y}$$

4.4 Hypergeometrische Verteilung

Gegeben ist eine Urne, die mit N Kugeln gefüllt ist. Es seien M dieser Kugeln rot und N-M Kugeln nicht rot. Wir entnehmen n Kugeln, d.h. Stichproben vom Umfang n. Dabei wird im Unterschied zur Binomialverteilung ohne Zurücklegen gezogen. Dadurch ändert sich nach jeder Ziehung die Zusammensetzung der Kugeln, die noch in der Urne sind und damit die Wahrscheinlichkeit, eine rote oder nicht-rote Kugel zu ziehen:

1. Ziehung (Ereignis A_1 bedeutet das Ziehen einer roten Kugel):

$$P(A_1) = \frac{M}{N} \quad \text{und} \quad P(\bar{A}_1) = \frac{N\text{-}M}{N}$$

2. Ziehung:

$$P(A_2 | A_1) = \frac{M\text{-}1}{N\text{-}1} \text{ und } P(\bar{A}_2 | A_1) = \frac{N\text{-}M}{N\text{-}1}$$ falls rote Kugel in 1. Ziehung

$$P(A_2 | \bar{A}_1) = \frac{M}{N\text{-}1} \text{ und } P(\bar{A}_2 | \bar{A}_1) = \frac{N\text{-}M\text{-}1}{N\text{-}1}$$ falls nicht-rote Kugel in 1. Ziehung

Die Ziehungen sind nicht unabhängig. Damit liegt kein Bernoulli-Prozess vor.

Das Ereignis A tritt ein, wenn eine rote Kugel gezogen wird. Entsprechend ist $\bar{A}$ das Ereignis, dass die gezogene Kugel nicht rot ist. Die Zufallsvariable X bezeichnet die Anzahl der Realisationen von Ereignis A, d.h. die Anzahl der gezogenen roten Kugeln.

Herleitung der Wahrscheinlichkeitsfunktion

Wir erhalten die gesuchte Wahrscheinlichkeit, indem wir die Anzahl der Auswahlmöglichkeiten mit x roten und n-x nicht-roten Kugeln auf die Anzahl aller möglichen Stichproben von n aus N Kugeln beziehen. Da das Ziehen der Kugeln aus der Urne ohne Zurücklegen erfolgt und die Reihenfolge irrelevant ist, lassen sich die Auswahlmöglichkeiten über folgende Formeln (für Kombinationen ohne Wiederholung) berechnen.

Auswahl von x aus M roten Kugeln:

$$\binom{M}{x}$$ Möglichkeiten

Auswahl von n-x aus N-M nicht-roten Kugeln:

$$\binom{N\text{-}M}{n\text{-}x}$$ Möglichkeiten

Auswahl von x aus M roten Kugeln und n-x aus N-M nicht-roten Kugeln:

$$\binom{M}{x}\binom{N\text{-}M}{n\text{-}x}$$ Möglichkeiten

Auswahl von n aus N Kugeln:

$$\binom{N}{n}$$ Möglichkeiten

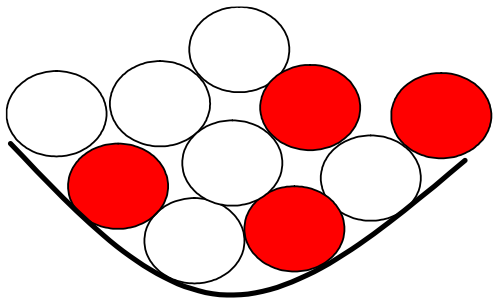
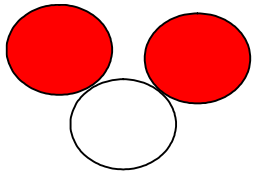
Eine Zufallsvariable X folgt einer hypergeometrischen Verteilung mit den Parametern N, M und n, wenn die Wahrscheinlichkeitsfunktion von X folgendermaßen gegeben ist:

$$f(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{für } x = 0, 1, \ldots, n \\ 0 & \text{sonst} \end{cases}$$

Dabei ist $x \leq M, (n-x) \leq N-M$ und $n \leq N$. Sofern diese Bedingungen nicht erfüllt sind, nimmt die Wahrscheinlichkeitsfunktion den Wert 0 an.

Beispiel:

In einer Urne befinden sich N = 10 Kugeln, von denen M = 4 Kugeln rot und N-M = 6 nicht rot (hier weiß) sind. Wie groß ist die Wahrscheinlichkeit, bei einer Stichprobe vom Umfang n = 3 genau x = 2 rote Kugeln zu ziehen, wenn die Kugeln nach dem Ziehen nicht wieder zurückgelegt werden?

Urne	Stichprobe
	

Nach Einsetzen der Parameter N = 10, M = 4 und n = 3 ergibt sich:

$$f(2) = \frac{\binom{4}{2}\binom{10-4}{3-2}}{\binom{10}{3}} = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{\frac{4!}{2!2!}\frac{6!}{1!5!}}{\frac{10!}{3!7!}} = \frac{6\cdot 6}{120} = 0{,}3$$

Hypergeometrisch verteilte Zufallsvariable X

Erwartungswert	Varianz
$E(X) = n\frac{M}{N}$	$Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

Vergleich zwischen hypergeometrischer und Binomialverteilung

Erwartungswert:

Die Erwartungswerte bei der hypergeometrischen und Binomialverteilung stimmen überein, wenn man p = M/N setzt.

Varianz:

Die Varianz bei der hypergeometrischen Verteilung ist für n > 1 um den Faktor (N-n)/(N-1) (sog. Endlichkeitskorrektur) kleiner als die Varianz bei der Binomial-verteilung. Der Unterschied nimmt mit wachsendem Stichprobenumfang n zu.

Grund: Informationsgewinn beim Ziehen ohne Zurücklegen

Je mehr Kugeln ohne Zurücklegen gezogen werden, desto genauere Informationen liegen über die restlichen noch in der Urne enthaltenen Kugeln vor. Die Streuung der Zufallsvariablen X verringert sich dadurch. Wenn dagegen mit Zurücklegen gezogen wird, bleibt die Zusammensetzung der Urne stets unverändert. Die Streuung der Zufallsvariablen X bleibt dann gleich.

Approximation der hypergeometrischen Verteilung durch die Binomialverteilung

Bei endlichem Stichprobenumfang n geht die Endlichkeitskorrektur (N-n)/(N-1) gegen 1, wenn N über alle Grenzen wächst. Die Varianz bei der hypergeometrischen Verteilung geht dann in die Varianz bei der Binomialverteilung über.

Allgemein lässt sich zeigen, dass die Wahrscheinlichkeitsfunktion $f_H(x|\ N,M,n)$ einer hypergeometrisch verteilten Zufallsvariablen X für $N \to \infty$ und $M \to \infty$ in die Wahrscheinlichkeitsfunktion $f_B(x\ I\ n,p)$ einer binomialverteilten Zufallsvariablen X übergeht, sofern $M/N \to p$:

$$\lim_{\substack{M \to \infty \\ N \to \infty \\ M/N \to p}} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

Faustregel:

Wenn der Auswahlsatz $n/N \leq 0{,}05$, lassen sich Wahrscheinlichkeiten bei hypergeometrisch verteilten Zufallsvariablen approximativ mit Wahrscheinlichkeiten bei einfacher handhabbaren binomialverteilten Zufallsvariablen berechnen.

Beispiel:

Auf einem Markt mit 100 Unternehmen befinden sich 10 innovative Unternehmen. Wie groß ist die Wahrscheinlichkeit, dass in einem Kartell von 4 Unternehmen mindestens die Hälfte der Unternehmen innovativ ist?

Da ein Unternehmen, das dem Kartell beigetreten ist, nicht nochmals für einen Beitritt in Betracht kommt, liegt das Auswahlmodell “Ziehen ohne Zurücklegen“ vor. Die Zufallsvariable X gibt die Anzahl der innovativen Unternehmen in dem Kartell an. Die gesuchte Wahrscheinlichkeit $P(X \geq 2)$ lässt sich mit der hypergeometrischen Verteilung bestimmen. Mit den Parametern N = 100, M = 10 und n = 4 erhält man:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{\binom{10}{2}\binom{100-10}{4-2}}{\binom{100}{4}} + \frac{\binom{10}{3}\binom{100-10}{4-3}}{\binom{100}{4}} + \frac{\binom{10}{4}\binom{100-10}{4-4}}{\binom{100}{4}}$$

$$= \frac{\frac{10!}{2!8!}\frac{90!}{2!88!}}{\frac{100!}{4!96!}} + \frac{\frac{10!}{3!7!}\frac{90!}{1!89!}}{\frac{100!}{4!96!}} + \frac{\frac{10!}{4!6!}\frac{90!}{0!90!}}{\frac{100!}{4!96!}} = \frac{45 \cdot 4005}{3921225} + \frac{120 \cdot 90}{3921225} + \frac{210 \cdot 1}{3921225}$$

$$= 0{,}049$$

Da der Auswahlsatz n/N = 4/100 = 0,04 kleiner als 0,05 ist, kann die gesuchte Wahrscheinlichkeit approximativ auch mit der Binomialverteilung bestimmt werden. Mit den Parametern n = 4 und p = M/N = 10/100 = 0,1 ergibt sich:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$\approx \binom{4}{2} 0{,}1^2 0{,}9^2 + \binom{4}{3} 0{,}1^3 0{,}9^1 + \binom{4}{4} 0{,}1^4 0{,}9^0$$

$$= \frac{4!}{2!2!} 0{,}1^2 0{,}9^2 + \frac{4!}{3!1!} 0{,}1^3 0{,}9 + \frac{4!}{4!0!} 0{,}1^4 1$$

$$= 0{,}052$$

Mit zunehmender Zahl der Unternehmen auf dem Markt wird die Approximation durch die Binomialverteilung genauer.

4.5 Poisson-Verteilung

Die Poisson-Verteilung ist eine Wahrscheinlichkeitsverteilung, die bei der Modellierung seltener Ereignisse von Relevanz ist.

Beispiele für entsprechende Fragestellungen:

- Unfälle in einem Betrieb pro Tag
- Telefonanrufe in einer Vermittlungsstelle in einer bestimmten Zeit
- Entwicklung von Marktneuheiten in einer Branche pro Jahr
- Tödliche Betriebsunfälle in einer Periode
- Ankünfte von Flugzeugen auf einem Flughafen pro Minute
- Druckfehler auf einer Buchseite

Eine Zufallsvariable X folgt einer Poisson-Verteilung mit dem Parameter $\lambda > 0$, wenn die Wahrscheinlichkeitsfunktion von X folgendermaßen gegeben ist:

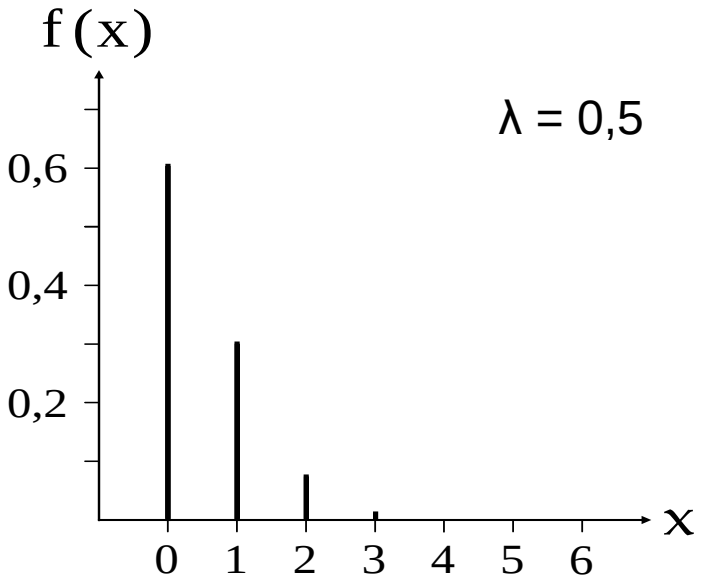
$$f(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, \ldots \\ 0 & \text{sonst} \end{cases}$$

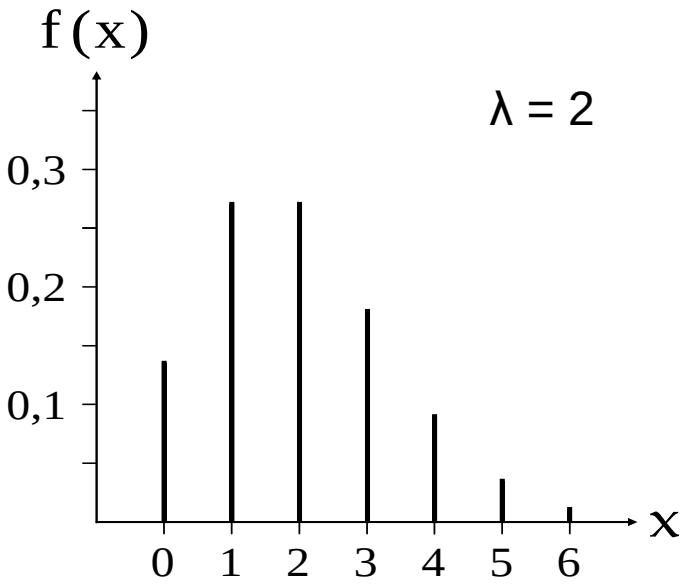
Poisson-verteilte Zufallsvariable X

Erwartungswert	Varianz
$E(X) = \lambda$	$Var(X) = \lambda$

Erwartungswert und Varianz stimmen also überein.

Wahrscheinlichkeitsfunktionen von Poisson-verteilten Zufallsvariablen X





Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung verläuft linkssteil (rechtsschief). Mit zunehmendem Wert des Parameters λ nehmen die Wahrscheinlichkeiten für größere x-Werte zu.

Faustregel:

Mit $p \leq 0{,}1$ und $n \geq 50$ lassen sich Wahrscheinlichkeiten bei binomialverteilten Zufallsvariablen approximativ durch Poisson-verteilte Zufallsvariablen mit $\lambda = np$ berechnen.

Beispiel:

Die Wahrscheinlichkeit, dass eine Person ein Medikament nicht verträgt, sei 0,001. Insgesamt wurden 2000 Personen mit diesem Medikament behandelt. Die Anzahl X der behandelten Personen, die das Medikament nicht vertragen, ist somit binomialverteilt mit den Parametern n = 2000 und p = 0,001, so dass die Wahrscheinlichkeiten folgendermaßen ermittelt werden können:

$$P(X = x) = \binom{2000}{x} 0{,}001^{x}\, 0{,}999^{2000-x}$$

Die Berechnung mit dieser Formel ist jedoch umfangreich. Da $p = 0{,}001 < 0{,}1$ und $n = 2000 > 50$, ist das betrachtete Ereignis, das Medikament nicht zu vertragen, ein seltenes Ereignis, das mit der Poisson-Verteilung modelliert werden kann.

Mit dem Parameter $\lambda = np = 2000 \cdot 0{,}001 = 2$ erhält man die entsprechende Wahrscheinlichkeitsfunktion der Poisson-verteilten Zufallsvariablen X:

$$f(x) = \begin{cases} \dfrac{2^{x}}{x!} e^{-2} & x = 0, 1, \ldots \\ 0 & \text{sonst} \end{cases}$$

Mit der Wahrscheinlichkeitsfunktion der Poisson-verteilten Zufallsvariablen X erhält man z.B.:

$$P(X = 0) = f(0) = \frac{2^0}{0!} e^{-2} = e^{-2} = 0{,}135$$

$$P(X = 1) = f(1) = \frac{2^1}{1!} e^{-2} = 2e^{-2} = 0{,}271$$

Der Vergleich zwischen Wahrscheinlichkeiten mit der Binomialverteilung und der Poisson-Verteilung ergibt nur geringe Abweichungen ab der vierten Dezimalstelle:

	Binomialverteilung	Poisson-Verteilung
P(X = 0)	0,1352	0,1353
P(X = 1)	0,2707	0,2707
P(X = 2)	0,2708	0,2707
P(X = 3)	0,1805	0,1804
P(X = 4)	0,0902	0,0902

Beispiel:

Eine Versicherung hat für einen Zeitraum von einem Jahr in 200 Bauunternehmen 122 schwere Unfälle bei Hochbauarbeiten ermittelt.

Wie groß ist die Wahrscheinlichkeit, dass es in einem Bauunternehmen in einem Jahr zu 0, 1, 2 und mehr als 2 schweren Unfällen bei Hochbauarbeiten kommt?

Ein schwerer Unfall ist ein seltenes Ereignis, da sich durchschnittlich im Jahr und pro Bauunternehmen nur 122 / 200 = 0,61 schwere Unfälle ereignet haben. In vielen Bauunternehmen hat sich im betrachteten Jahr kein oder ein Unfall ereignet. Wir nehmen deshalb einmal an, dass die Wahrscheinlichkeiten mit der Poissonverteilung berechnet werden können.

Hierzu muss der Parameter λ festgelegt werden, der den Erwartungswert für die Anzahl der schweren Unfälle in einem Bauunternehmen pro Jahr angibt. Unter Ausnutzung der verfügbaren Informationen verwenden wir für den Parameter λ die durchschnittliche Zahl der schweren Unfälle $\lambda = 122 / 200 = 0{,}61$.

Zur Bestimmung der gesuchten Wahrscheinlichkeiten betrachten wir daher folgende Wahrscheinlichkeitsfunktion für die Anzahl X der schweren Unfälle bei Hochbau-arbeiten:

$$f(x) = \begin{cases} \dfrac{0{,}61^x}{x!} e^{-0{,}61} & x = 0, 1, \ldots \\ 0 & \text{sonst} \end{cases}$$

Daraus ergeben sich folgende Wahrscheinlichkeiten:

$$P(X = 0) = f(0) = \frac{0{,}61^0}{0!} e^{-0{,}61} = e^{-0{,}61} = 0{,}543$$

$$P(X = 1) = f(1) = \frac{0{,}61^1}{1!} e^{-0{,}61} = 0{,}61 e^{-0{,}61} = 0{,}331$$

$$P(X = 2) = f(2) = \frac{0{,}61^2}{2!} e^{-0{,}61} = 0{,}186 e^{-0{,}61} = 0{,}101$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - [f(0) + f(1) + f(2)]$$

$$= 1 - (0{,}543 + 0{,}331 + 0{,}101) = 1 - 0{,}975 = 0{,}025$$