

Diskrete Verteilungen

Dipl.-Ing. Hubert Schölnast, BSc
Stand: 29. Juni 2022

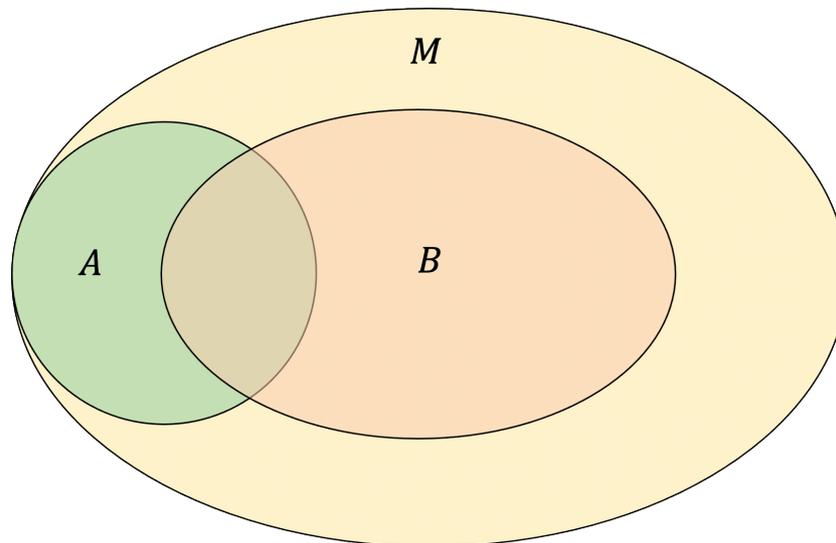
Inhaltsverzeichnis

1	Hypergeometrische Verteilung	3
1.1	Allgemeine Aufgabenstellung:	3
1.1.1	Beispiel	3
1.2	Allgemeine Formel	5
1.3	Erwartungswert und Varianz	6
2	Binomialverteilung	7
2.1	Bernoulli-Experiment; Bernoulli-Kette	7
2.1.1	Bernoulli-Experiment	7
2.1.2	Bernoulli-Kette	8
2.1.3	Beispiel	8
2.2	Allgemeine Formel	10
2.3	Erwartungswert und Varianz	10
2.4	Binomialverteilung als Näherung für die hypergeometrische Verteilung	11
3	Poisson-Verteilung	14
3.1	Poisson-Prozess	14
3.1.1	Der Parameter λ	14
3.2	Allgemeine Formel	15
3.2.1	Beispiel	15
3.3	Erwartungswert und Varianz	17
3.4	Poisson-Verteilung als Näherung für die Binomialverteilung	17
3.4.1	Beispiel	17

1 Hypergeometrische Verteilung

1.1 Allgemeine Aufgabenstellung:

Sie haben eine Menge M , wovon eine Teilmenge A existiert, in der alle Elemente eine bestimmte Eigenschaft haben, die jenen Elementen von M , die nicht in A sind, fehlt. Die Mächtigkeiten von M und A sind bekannt. Sie bilden durch ein Zufallsexperiment eine neue Teilmenge B von M mit einer vorgegebenen Größe. Dabei wird die Menge B Elemente enthalten, die zu A gehören, und es werden Elemente in B sein, die nicht zu A gehören. Gesucht sind die Wahrscheinlichkeiten für die möglichen Zusammensetzungen von B .



1.1.1 Beispiel

Die Menge M sind 30 Kugeln, die in einem Behälter liegen. Davon sind 10 Kugeln grün (das ist die Menge A), alle anderen Kugeln sind nicht grün (gehören also zu M , aber nicht zu A)

Nun nimmt man 7 zufällig ausgewählte Kugeln aus dem Behälter, das ist die Menge B .

Wichtig (Unter welchen Bedingungen liegt eine hypergeometrische Verteilung vor?)

Sie ziehen entweder alle Kugeln zugleich, oder, wenn Sie die Kugeln hintereinander ziehen, legen Sie sie nicht mehr zurück in den Behälter.

Wie groß ist die Wahrscheinlichkeit dafür, dass von den 7 gezogenen Kugeln 0 grün sind? Ebenso für 1 grüne, 2 grüne usw.

Berechnen wir zuerst die Wahrscheinlichkeit für einen speziellen Fall. Wie groß ist die Wahrscheinlichkeit dafür, dass unter den 7 gezogenen Kugeln genau 3 grüne sind?

Bei 30 verschiedenen Kugeln kann man

$$\binom{30}{7} = \frac{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7} = 2.035.800$$

verschiedene Stichproben der Größe 7 ziehen.

In wie vielen dieser Stichproben gibt es 3 grüne und 4 nichtgrüne Kugeln?

Es gibt insgesamt 10 grüne Kugeln, davon muss man genau 3 erwischen, und von den 20 nichtgrünen Kugeln müssen es genau 4 sein:

$$\binom{10}{3} \cdot \binom{20}{4} = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} \cdot \frac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4} = 120 \cdot 4845 = 581.400$$

Unter den 2.035.800 überhaupt möglichen Stichproben sind also 581.400 bei denen wir genau 3 grüne und 4 nichtgrüne erwischt haben.

Wir haben also:

$$P(X = 3) = \frac{581.400}{2.035.800} \cong 0,2856$$

Die Wahrscheinlichkeit dafür, genau 3 grüne in einer Stichprobe der Größe 7 zu haben, liegt bei ungefähr 28,56%.

Versuchen wir nun, daraus eine allgemeine Formel herzuleiten:

$$P(X = 3) = \frac{581400}{2035800}$$

$$P(X = 3) = \frac{\binom{10}{3} \cdot \binom{20}{4}}{\binom{30}{7}}$$

$$P(X = x) = \frac{\binom{10}{n} \cdot \binom{30 - 10}{7 - n}}{\binom{30}{7}}$$

Erinnern wir uns daran, dass die Zahlen 30, 10 und 7 Mächtigkeiten von Mengen waren:

$$P(X = x) = \frac{\binom{|A|}{x} \cdot \binom{|M| - |A|}{|B| - x}}{\binom{|M|}{|B|}}$$

Weil das aber ein wenig umständlich ist, schreibt man das meist ein wenig einfacher.

1.2 Allgemeine Formel

Allgemeine Formel für hypergeometrische Verteilungen:

- N ist Anzahl aller Elemente, aus denen etwas ausgewählt wird
- M ist die Gesamtzahl aller Elemente, die die gewünschte Eigenschaft haben
- n ist die Größe der Stichprobe
- x ist die Anzahl der Elemente in der Stichprobe, die die gewünschte Eigenschaft haben

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}}$$

In Excel gibt es dafür die Funktion

$$=HYPGEOM.VERT(x; n; M; N; k)$$

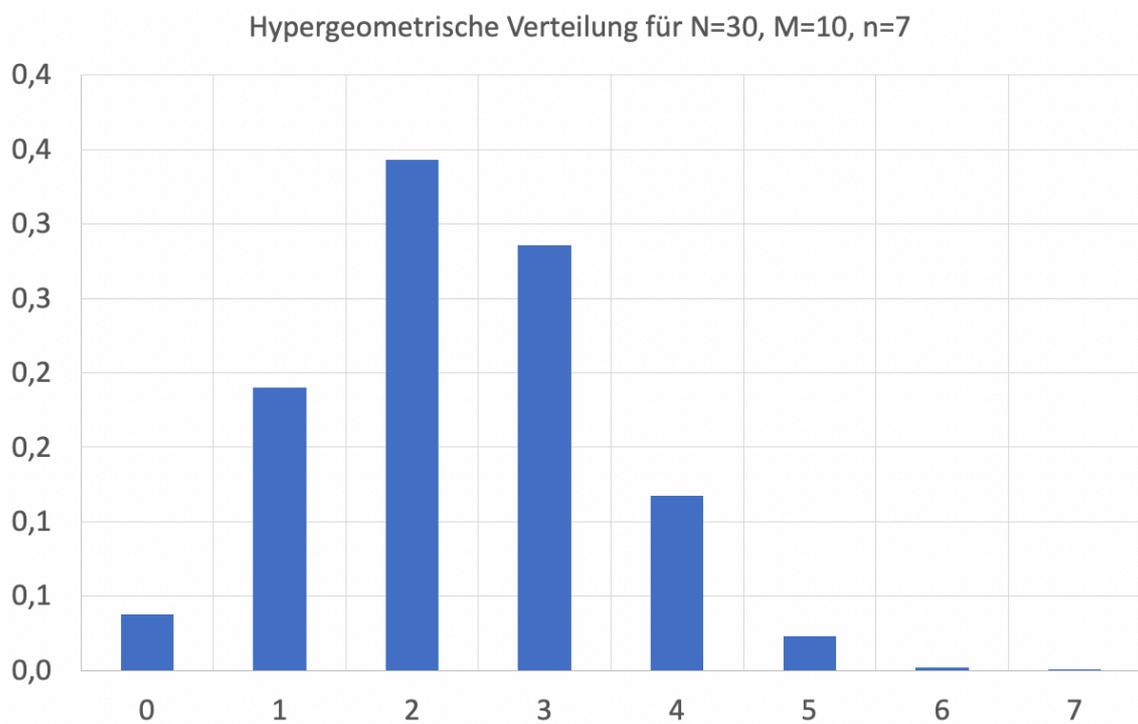
Dabei haben x , n , M und N dieselben Bedeutungen wie im Kasten beschrieben. Der fünfte Parameter k gibt an, ob, so wie hier beschrieben, der Wert für $P(X = x)$ berechnet werden soll (dann muss k gleich 0 sein oder kann weggelassen werden) oder ob der Wert für $P(X \leq x)$ berechnet werden soll (dann muss man bei k den Wert 1 angeben).

Probe (Fortsetzung des Beispiels):

Wenn diese Formel stimmt, und wir x von 0 bis 7 durchlaufen lassen, berechnen wir ja die Wahrscheinlichkeiten aller 8 Elementarereignisse, und weil diese Elementarereignisse unvereinbar sind, und in ihrer Vereinigung das sichere Ereignis darstellen, müssen sich diese 8 Wahrscheinlichkeiten zu genau 1 addieren:

$$\begin{aligned}P(X = 0) &= 0,0380784 \\P(X = 1) &= 0,1903920 \\P(X = 2) &= 0,3427056 \\P(X = 3) &= 0,2855880 \\P(X = 4) &= 0,1175950 \\P(X = 5) &= 0,0235190 \\P(X = 6) &= 0,0020631 \\P(X = 7) &= 0,0000589 \\ \text{Summe} &= 1,0000000\end{aligned}$$

Die Probe geht auf. Damit ist zwar nicht wirklich bewiesen, dass die Formel immer korrekt ist, aber die Richtigkeit erscheint wenigstens plausibel.



1.3 Erwartungswert und Varianz

Es gilt wieder:

- N ist Anzahl aller Elemente, aus denen etwas ausgewählt wird
- M ist die Gesamtzahl aller Elemente, die die gewünschte Eigenschaft haben
- n ist die Größe der Stichprobe

Der Erwartungswert einer hypergeometrischen Verteilung ist:

$$\mu = E(X) = n \cdot \frac{M}{N}$$

Für die Varianz gilt:

$$\sigma^2 = \text{VAR}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

2 Binomialverteilung

2.1 Bernoulli-Experiment; Bernoulli-Kette

2.1.1 Bernoulli-Experiment

Ein Bernoulli-Experiment ist ein Zufallsexperiment, das nur zwei Elementarereignisse hat:

- Ein Ereignis mit einer bestimmten Eigenschaft tritt ein (erwünschtes Ereignis)
- Ein Ereignis mit einer bestimmten Eigenschaft tritt nicht ein (unerwünschtes Ereignis)

Beispiel 1:

Sie werfen einen gewöhnlichen Spielwürfel und hoffen auf die Augenzahl 3. Die beiden Elementarereignisse sind:

- Sie werfen die Augenzahl 3 (erwünschtes Ereignis)
- Sie werfen eine andere Augenzahl (welche ist egal) (unerwünschtes Ereignis)

Die Wahrscheinlichkeit für eine erfolgreiche Durchführung ist $p_1 = \frac{1}{6}$.

Beispiel 2:

Sie werfen eine Münze und hoffen auf Kopf:

- Die Münze zeigt Kopf (erwünschtes Ereignis)
- Die Münze zeigt Zahl (unerwünschtes Ereignis)

Die Wahrscheinlichkeit für eine erfolgreiche Durchführung ist $p_2 = \frac{1}{2}$.

Beispiel 3:

Sie ziehen aus einem Sack, in dem sich 10 grüne und 20 andersfarbige Kugeln befinden, eine Kugel, die sie anschließend wieder in den Sack zurückgeben und hoffen darauf, eine grüne Kugel zu erwischen:

- Sie ziehen eine grüne Kugel (erwünschtes Ereignis)
- Sie ziehen eine Kugel, die nicht grün ist (unerwünschtes Ereignis)

Die Wahrscheinlichkeit für eine erfolgreiche Durchführung ist $p_3 = \frac{10}{30} = \frac{1}{3}$.

Man unterscheidet bei Bernoulli-Experimenten immer zwischen Erfolg und Misserfolg. Andere Möglichkeiten für den Ausgang eines Bernoulli-Experiments gibt es nicht.

2.1.2 Bernoulli-Kette

Wenn Sie n gleiche und voneinander unabhängige Bernoulli-Experimente ausführen, bilden diese Experimente zusammen eine Bernoulli-Kette der Länge n .

Bernoulli-Verteilung = Binomialverteilung

Wenn man sich nun dafür interessiert, wie hoch die Wahrscheinlichkeiten für verschiedene Anteile von erfolgreichen Durchführungen in einer Bernoulli-Kette sind, erhält man eine Bernoulli-Verteilung, die aber eher unter dem Namen Binomialverteilung bekannt ist.

2.1.3 Beispiel

Sie haben 8 gewöhnliche Spielwürfel, die Sie alle zugleich werfen. Sie wetten mit einer Freundin, dass von diesen Würfeln genau 2 die Augenzahl 3 zeigen werden. Wie groß ist die Wahrscheinlichkeit, dass Sie diese Wette gewinnen werden?

Nehmen wir an, die Würfel wären durchnummeriert, so dass man sie in eine bestimmte Reihenfolge bringen kann. Dann wäre 33□□□□□□ ein erfolgreicher Wurf, wobei □ »eine andere Augenzahl als 3« bedeutet. Die Wahrscheinlichkeit dafür ist:

$$P(33□□□□□□) = \frac{1}{6} \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right) \cdot \left(1 - \frac{1}{6}\right)$$

$$P(33□□□□□□) = \left(\frac{1}{6}\right)^2 \cdot \left(1 - \frac{1}{6}\right)^6 = \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^6 = \frac{5^6}{6^8} \cong 0,009303$$

Allerdings sind auch 3□3□□□□□ und □□3□□□3□ erfolgreiche Würfe. Insgesamt gibt es

$$\binom{8}{2} = \frac{8 \cdot 7}{1 \cdot 2} = 28$$

erfolgreiche Würfe. Also:

$$P(2\text{mal } 3) = \binom{8}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(1 - \frac{1}{6}\right)^6 \cong 0,2605$$

Ihre Erfolgchancen liegen also bei 26,05%.

Versuchen wir auch hier die Verallgemeinerung. Wir hatten:

$$P(X = 2) = \binom{8}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(1 - \frac{1}{6}\right)^6$$

Wenn es nicht genau 2, sondern eine beliebige Anzahl von Dreieren sein soll:

$$P(X = x) = \binom{8}{x} \cdot \left(\frac{1}{6}\right)^x \cdot \left(1 - \frac{1}{6}\right)^{8-x}$$

Das können wir gleich mal für alle 9 möglichen Fälle durchrechnen. Die Summe müsste dann wieder 1 ergeben:

$$P(X = 0) = 0,23256804$$

$$P(X = 1) = 0,37210886$$

$$P(X = 2) = 0,26047620$$

$$P(X = 3) = 0,10419048$$

$$P(X = 4) = 0,02604762$$

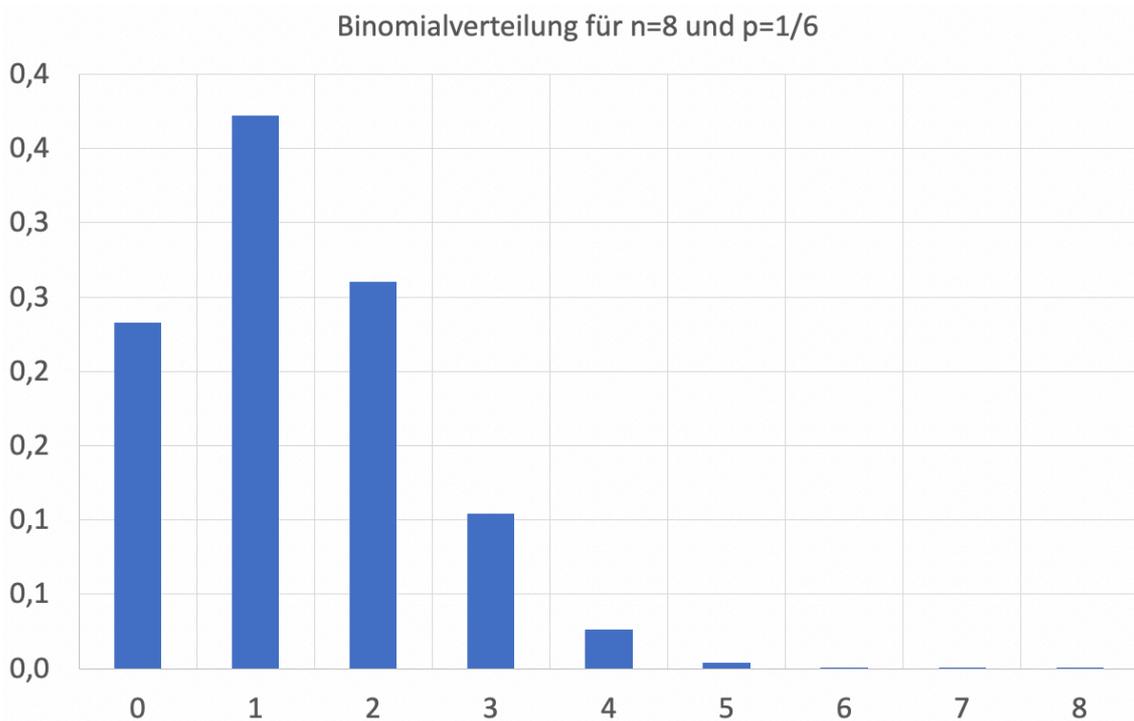
$$P(X = 5) = 0,00416762$$

$$P(X = 6) = 0,00041676$$

$$P(X = 7) = 0,00002381$$

$$P(X = 8) = 0,00000060$$

$$\text{Summe} = 1,00000000$$



Ersetzen wir auch die Größe der Stichprobe (n) und die Wahrscheinlichkeit für das Eintreten des gewünschten Ereignisses (p) als Variable in der Formel, so erhalten wir die allgemeine Formel für Binomialverteilungen.

2.2 Allgemeine Formel

Allgemeine Formel für Binomialverteilungen:

Gegeben ist eine Bernoulli-Kette der Länge n , wobei bei jedem einzelnen Bernoulli-Experiment das Ereignis A mit der Wahrscheinlichkeit p eintritt.

- n ist die Größe der Stichprobe = Länge der Bernoulli-Kette
- p ist die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses
- x ist die Anzahl der Elemente in der Bernoulli-Kette, die die gewünschte Eigenschaft haben

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

In Excel gibt es dafür die Funktion

$$=BINOM.VERT(x; n; p; k)$$

Dabei haben x , n und p dieselben Bedeutungen wie im Kasten beschrieben. Der vierte Parameter k gibt an, ob, so wie hier beschrieben, der Wert für $P(X = x)$ berechnet werden soll (dann muss k gleich 0 sein oder kann weggelassen werden) oder ob der Wert für $P(X \leq x)$ berechnet werden soll (dann muss man bei k den Wert 1 angeben).

2.3 Erwartungswert und Varianz

n und p haben dieselben Bedeutungen wie soeben beschrieben:

- n ist die Größe der Stichprobe = Länge der Bernoulli-Kette
- p ist die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses

Dann ist der Erwartungswert:

$$\mu = E(X) = n \cdot p$$

Und die Varianz:

$$\sigma^2 = VAR(X) = n \cdot p \cdot (1 - p)$$

2.4 Binomialverteilung als Näherung für die hypergeometrische Verteilung

Rechnen wir nochmals das Beispiel aus der hypergeometrischen Verteilung durch, diesmal aber mit einer kleinen Veränderung: Im Behälter liegen nicht 30 Kugeln, von denen 10 grün sind, sondern 300 von denen 100 grün sind. Das Verhältnis ist dasselbe, aber die Menge ist zehnmal so groß.

Wir hatten in der ersten Version des Beispiels diese Zahlen:

$$P_{h_{30}}(X = 3) = \frac{\binom{10}{3} \cdot \binom{20}{4}}{\binom{30}{7}} = \frac{\frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} \cdot \frac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}}{\frac{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}}$$

$$P_{h_{30}}(X = 3) = \frac{10 \cdot 9 \cdot 8 \cdot 20 \cdot 19 \cdot 18 \cdot 17 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24}$$

$$P_{h_{30}}(X = 3) = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3 \cdot 4} \cdot \frac{10 \cdot 9 \cdot 8 \cdot 20 \cdot 19 \cdot 18 \cdot 17}{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24}$$

$$P_{h_{30}}(X = 3) = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} \cdot \frac{10 \cdot 9 \cdot 8}{30 \cdot 29 \cdot 28} \cdot \frac{20 \cdot 19 \cdot 18 \cdot 17}{27 \cdot 26 \cdot 25 \cdot 24}$$

$$P_{h_{30}}(X = 3) = \binom{7}{3} \cdot \frac{10}{30} \cdot \frac{9}{29} \cdot \frac{8}{28} \cdot \frac{20}{27} \cdot \frac{19}{26} \cdot \frac{18}{25} \cdot \frac{17}{24}$$

$$P_{h_{30}}(X = 3) = \binom{7}{3} \cdot \frac{10}{30} \cdot \frac{9}{29} \cdot \frac{8}{28} \cdot \left(1 - \frac{7}{27}\right) \cdot \left(1 - \frac{7}{26}\right) \cdot \left(1 - \frac{7}{25}\right) \cdot \left(1 - \frac{7}{24}\right)$$

Der erste Bruch (gemeint ist $\frac{10}{30}$) spiegelt noch exakt die Wahrscheinlichkeit wider, aus dem vollen Behälter eine grüne Kugel zu ziehen. Die nachfolgenden Brüche weichen davon immer weiter ab.

Und das sind die Zahlen aus dem modifizierten Beispiel:

$$P_{h_{300}}(X = 3) = \frac{\binom{100}{3} \cdot \binom{200}{4}}{\binom{300}{7}} = \frac{\frac{100 \cdot 99 \cdot 98}{1 \cdot 2 \cdot 3} \cdot \frac{200 \cdot 199 \cdot 198 \cdot 197}{1 \cdot 2 \cdot 3 \cdot 4}}{\frac{300 \cdot 299 \cdot 298 \cdot 297 \cdot 296 \cdot 295 \cdot 294}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}}$$

$$P_{h_{300}}(X = 3) = \frac{100 \cdot 99 \cdot 98 \cdot 200 \cdot 199 \cdot 198 \cdot 187 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 300 \cdot 299 \cdot 298 \cdot 297 \cdot 296 \cdot 295 \cdot 294}$$

$$P_{h_{300}}(X = 3) = \binom{7}{3} \cdot \frac{100}{300} \cdot \frac{99}{299} \cdot \frac{98}{298} \cdot \frac{200}{297} \cdot \frac{199}{296} \cdot \frac{198}{295} \cdot \frac{197}{294}$$

$$P_{h_{300}}(X = 3) = \binom{7}{3} \cdot \frac{100}{300} \cdot \frac{99}{299} \cdot \frac{98}{298} \cdot \left(1 - \frac{97}{297}\right) \cdot \left(1 - \frac{97}{296}\right) \cdot \left(1 - \frac{97}{295}\right) \cdot \left(1 - \frac{97}{294}\right)$$

Auch hier gibt der erste Bruch $\left(\frac{100}{300}\right)$ noch die genaue Wahrscheinlichkeit wieder, und alle weiteren Brüche weichen nach wie vor ein wenig davon ab, aber sie haben alle Werte, die deutlich näher an diesem Wert liegen.

$$\frac{100}{300} \cong 0,33333 \qquad \frac{97}{294} \cong 0,32993$$

Wenn wir nun diese Abweichung ignorieren, und für den ursprünglichen Anteil der grünen Kugel die Variable p verwenden ($p = \frac{10}{30} = \frac{100}{300} = \frac{1}{3}$), sieht unsere Formel so aus:

$$P_b(X = 3) = \binom{7}{3} \cdot p \cdot p \cdot p \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) \cdot (1 - p)$$

$$P_b(X = 3) = \binom{7}{3} \cdot p^3 \cdot (1 - p)^4$$

oder Allgemein:

$$P_b(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

Und das ist exakt die Formel für die Binomialverteilung.

Und hier sind die Ergebnisse in absoluten Zahlen:

- $n=7, N=30, M=10 \quad P_{h_{30}}(X = 3) = 0,285587975$
- $n=7, N=300, M=100 \quad P_{h_{300}}(X = 3) = 0,258653041$
- $n=7, p=1/3 \quad P_b(X = 3) = 0,256058528$

Wie man sieht, eignet sich die Binomialverteilung sehr als Näherung für die hypergeometrische Verteilung, und zwar umso besser, je größer die Mengen sind, aus denen etwas gezogen wird.

Als Faustregel gilt, dass es ausreicht eine hypergeometrische Verteilung mit der einfacheren Formel für die Binomialverteilung zu rechnen, wenn

$$\frac{N}{n} > 20$$

In der ersten Version des Beispiels war $\frac{N}{n} = \frac{30}{7} \cong 4,29 < 20$, daher war es da nicht ratsam, die Näherung zu verwenden. Der relative Fehler lag hier bei

$$\frac{0,285587975 - 0,256058528}{0,285587975} \cong 0,1043$$

also bei rund 10%.

In der zweiten Version hatten wir aber $\frac{N}{n} = \frac{300}{7} \cong 42,86 > 20$, und da liegt der Fehler bereits in einem Bereich, der für die meisten Anwendungen völlig akzeptabel ist, nämlich bei nur mehr rund 1 %:

$$\frac{0,258653041 - 0,256058528}{0,258653041} \cong 0,01003$$

3 Poisson-Verteilung¹

3.1 Poisson-Prozess

Poisson-Prozesse treten in der Praxis meist als Fehler, Unfälle, Störfälle oder andere seltene Ereignisse auf. Folgende drei Kriterien müssen erfüllt sein, damit ein Prozess ein Poisson-Prozess ist:

1. Die Wahrscheinlichkeit, dass ein Ereignis innerhalb einer festen Zeitspanne auftritt, ist konstant, egal wann genau diese Zeitspanne beginnt. An die Stelle der Zeitspanne kann auch eine bestimmte Länge (z.B., wenn es um Beschädigungen bei einem Seil oder Kabel geht), eine bestimmte Fläche (z.B. Lackschäden), ein bestimmtes Volumen (z.B. Keime in Trinkwasser) usw. treten. Mit anderen Worten: Heute gelten dieselben Wahrscheinlichkeiten wie morgen oder in 1 Jahr, bzw. auf dem ersten Meter des Kabels ist die Wahrscheinlichkeit für einen Fehler gleich groß wie auf einem 1 Meter langen Teilstück bei Kilometer 23.
2. Die Wahrscheinlichkeit dafür, dass einer bestimmten Stelle bzw. zu einem bestimmten Zeitpunkt ein Ereignis auftritt, ist unabhängig davon, ob an einer bestimmten anderen Stelle oder zu einem bestimmten anderen Zeitpunkt ein anderes Ereignis auftritt oder nicht.
3. Für eine genügend kleine Zeitspanne Δt oder einen genügend kleinen Ausschnitt Δx ist Wahrscheinlichkeit dafür, dass dort 2 oder mehr Ereignisse auftreten, vernachlässigbar klein. (Genauer gesagt: Die Wahrscheinlichkeit dafür, in Δt 2 Ereignisse zu finden sollte proportional zur Wahrscheinlichkeit sein, in Δt^2 1 Ereignis zu finden.)

3.1.1 Der Parameter λ

Betrachten wir zwei Straßenkreuzungen, an denen es gelegentlich zu Unfällen kommt. An Kreuzung A ereignen sich im Schnitt 2 Unfälle pro Jahr, an Kreuzung B sind es 4 Unfälle im selben Zeitraum. Die Unfallhäufigkeit ist bei beiden Kreuzungen im Sommer gleich hoch wie im Winter, und ob es am Tag X zu einem Unfall kommt, hängt nicht davon ab, ob es auch am Tag Y einen Unfall gab.

¹ Siméon Denis Poisson (1781-1840) war ein sehr produktiver französischer Mathematiker und Physiker und ein Schüler von Pierre Simon Laplace. In der Physik arbeitete er gemeinsam mit seinem Lehrer an Akustik, Elektrizität und der Wärmeleitung in festen Körpern. In der Optik sagte er den Poisson-Fleck voraus, der später auch experimentell bestätigt wurde. (Die Beugung des Lichts führt zu einem hellen Fleck im Zentrum des Schattens einer Kugel.) In der Mathematik beschäftigte er sich mit der Differentialgeometrie, der Infinitesimalrechnung und der Wahrscheinlichkeitstheorie.

Die Unfallereignisse beider Kreuzungen sind also poissonverteilt, aber sie unterscheiden sich in der Häufigkeit. Diese Häufigkeit wird durch den Parameter λ beschrieben. Der Wert für λ ist bei Kreuzung B doppelt so groß wie bei Kreuzung A.

3.2 Allgemeine Formel

Allgemeine Formel für Poissonverteilungen:

Gegeben ist ein Poisson-Prozess, bei dem pro Untersuchungseinheit im Mittel λ Ereignisse stattfinden.

- λ ist der Erwartungswert für die Anzahl der Ereignisse pro Untersuchungseinheit
 $\lambda \in \mathbb{R} \wedge \lambda > 0$
- x ist die Anzahl der Ereignisse, die tatsächlich innerhalb einer bestimmten Untersuchungseinheit auftreten. $x \in \mathbb{N}$

Die Wahrscheinlichkeit, innerhalb einer Untersuchungseinheit genau x Ereignisse zu finden, ist dann:

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

In Excel gibt es dafür die Funktion

$$=POISSON.VERT(x;\lambda;k)$$

Dabei haben x und λ dieselben Bedeutungen wie im Kasten beschrieben. Der dritte Parameter k gibt an, ob, so wie hier beschrieben, der Wert für $P(X = x)$ berechnet werden soll (dann muss k gleich 0 sein) oder ob der Wert für $P(X \leq x)$ berechnet werden soll (dann muss man bei k den Wert 1 angeben).

3.2.1 Beispiel

Eine Website im Internet wird im Schnitt 400-mal pro Stunde aufgerufen. Dieser Wert ist zu jeder Uhrzeit gleich und es gibt auch keine Unterschiede zwischen den verschiedenen Wochentagen. Wie groß ist die Wahrscheinlichkeit dafür, dass innerhalb einer Minute gar kein Aufruf der Seite erfolgt? Mit welcher Wahrscheinlichkeit kommt es zu genau 10 Aufrufen innerhalb einer Minute?

400 Aufrufe pro Stunde sind $\frac{400}{60} = \frac{20}{3} \cong 6,6667$ Aufrufe pro Minute. Das ist der Wert für λ :

$$\lambda = 6,\bar{6}$$

Berechnen wir zuerst den Wert für $X = 0$:

$$P(X = 0) = \frac{6,6^0}{0!} \cdot e^{-\frac{20}{3}} = \frac{1}{1} \cdot e^{-6,6} = e^{-6,6} \cong 0,001272634$$

Die Wahrscheinlichkeit dafür liegt bei ca. 1,27‰.

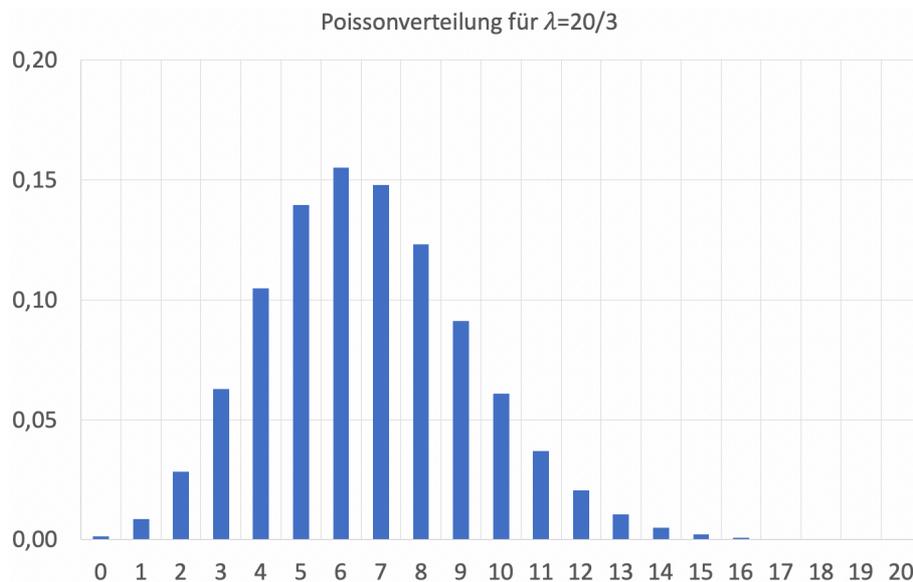
$$e^{-6,6} = \frac{1}{e^{6,6}} \cong \frac{1}{785,8}$$

Unter ungefähr 786 Minuten (das sind 13 Stunden und 6 Minuten) findet man eine Minute, in der niemand diese Website aufruft.

Nun für 10 Aufrufe pro Minute:

$$P(X = 10) = \frac{6,6^{10}}{10!} \cdot e^{-\frac{20}{3}} \cong \frac{173.415.299,2}{3.628.800} \cdot 0,001272634 \cong 0,060817397$$

Die Wahrscheinlichkeit dafür liegt bei ca. 6,08%.



3.3 Erwartungswert und Varianz

Die Formeln für Erwartungswert und Varianz einer Poissonverteilung sind besonders einfach:

Erwartungswert:

$$\mu = E(X) = \lambda$$

Varianz:

$$\sigma^2 = VAR(X) = \lambda$$

3.4 Poisson-Verteilung als Näherung für die Binomialverteilung

3.4.1 Beispiel

Ein Hersteller produziert elektronische Schaltkreise. Der Anteil der schadhafte Chips beträgt 0,5% ($= \frac{1}{200}$). Die Chips werden in Paketen zu je 500 Stück abgepackt. Wie groß ist die Wahrscheinlichkeit dafür, dass in einer Packung genau 2 Chips defekt sind?

Das ist eigentlich ein typisches Beispiel für eine Binomialverteilung. Also rechnen wir es auch nach dieser Formel:

$$P_b(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

- $n = 500$ ist die Größe der Stichprobe = Länge der Bernoulli-Kette
- $p = \frac{1}{200}$ ist die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses
- $x = 2$ ist die Anzahl der Elemente in der Bernoulli-Kette, die die gewünschte Eigenschaft haben

$$P_b(X = 2) = \binom{500}{2} \cdot \left(\frac{1}{200}\right)^2 \cdot \left(1 - \frac{1}{200}\right)^{500-2}$$

$$P_b(X = 2) = \frac{500 \cdot 499}{1 \cdot 2} \cdot \left(\frac{1}{200}\right)^2 \cdot \left(\frac{199}{200}\right)^{498}$$

$$P_b(X = 2) \cong 0,256965473$$

Die exakte Wahrscheinlichkeit dafür liegt bei ungefähr 25,69%.

Führen wir nun die Berechnung mit der Formel für die Poisson-Verteilung durch.

- λ ist der Erwartungswert für die Anzahl der Ereignisse pro Untersuchungseinheit. Diesen Wert müssen wir erst berechnen. Die Untersuchungseinheit ist eine Packung á 500 Stück, und wenn die Fehlerwahrscheinlichkeit bei $\frac{1}{200}$ liegt, sollten pro Packung im Mittel ungefähr $\frac{500}{200} = 2,5$ schadhafte Chips enthalten sein.
- $x = 2$ ist die Anzahl der Ereignisse, die tatsächlich innerhalb einer bestimmten Untersuchungseinheit auftreten.

Hier ist noch einmal die Formel:

$$P_p(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Einsetzen:

$$P_p(X = 2) = \frac{2,5^2}{2!} \cdot e^{-2,5}$$

$$P_p(X = 2) \cong \frac{6,25}{2} \cdot 0,082085$$

$$P_p(X = 2) \cong 0,256515621$$

Die angenäherte Wahrscheinlichkeit dafür liegt bei ungefähr 25,65%.

Der relative Fehler beträgt

$$\frac{0,256965473 - 0,256515621}{0,256965473} = \frac{0,000449853}{0,256965473} \cong 0,00175$$

Der relative Fehler liegt also bei ca. 1,75‰.

Faustregel:

Wenn $n > 50$ und $p < \frac{1}{10}$, dann kann man eine Binomialverteilung durch eine Poissonverteilung approximieren, mit $\lambda = n \cdot p$.

Begründung für die Approximation:

$$p = \frac{\lambda}{n}$$

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

$$P(X = x) = \binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$P(X = x) = \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{x!} \cdot \frac{\lambda^x}{n^x} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$P(X = x) = \frac{\lambda^x}{x!} \cdot \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Zwischenschritt:

$$\frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-x+2}{n} \cdot \frac{n-x+1}{n}$$

$$\lim_{n \rightarrow \infty} \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-x+2}{n} \cdot \frac{n-x+1}{n} = \lim_{n \rightarrow \infty} 1 \cdot 1 \cdot 1 \cdot \dots \cdot 1 \cdot 1 = 1$$

$$\lim_{n \rightarrow \infty} \frac{n \cdot (n-1) \cdot \dots \cdot (n-x+1)}{n^x} = 1$$

Ein anderer Zwischenschritt:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-x} = \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-x} = e^{-\lambda}$$

Einsetzen der Zwischenergebnisse

$$\lim_{n \rightarrow \infty} P(X = x) = \frac{\lambda^x}{x!} \cdot 1 \cdot e^{-\lambda}$$

$$\lim_{n \rightarrow \infty} P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Wenn n über alle Schranken wächst, werden die Binomialverteilung und die Poissonverteilung identisch.

□