

Kapitel 3

Modelle für Zähldaten

“Everybody speaks of probability, but no one is able to say what it is, in a way which is satisfactory for others.”

Garrett Birkhoff

Lernziele

- Sie kennen die drei Axiome der Wahrscheinlichkeitsrechnung.
- Sie kennen den Begriff der Unabhängigkeit und können einfache Rechenaufgaben lösen.
- Sie kennen den Begriff der bedingten Wahrscheinlichkeit und können damit umgehen. Insbesondere kennen Sie den Satz von Bayes und den Satz der totalen Wahrscheinlichkeit. Sie wissen, wie man $P(A|B)$ und $P(B|A)$ mit Formeln in einen Zusammenhang bringen kann und können einfache Aufgaben damit lösen.
- Sie kennen die Begriffe odds und odds-Ratio. Sie können sie interpretieren und einfache Rechenaufgaben mit ihnen lösen.
- Sie kennen den Begriff der Zufallsvariable, der Wahrscheinlichkeitsverteilung und kumulativen Verteilungsfunktion.
- Sie kennen die Binomial- und die Poissonverteilung.

- Sie kennen die Begriffe Erwartungswert, Varianz und Standardabweichung. Sie können diese Werte für einfache Verteilungen berechnen.

3.1 Wahrscheinlichkeitsmodelle (Stahel, Kap. 4.1, 4.2)

Wir betrachten **Zufallsexperimente**, bei denen der Ausgang nicht exakt vorhersagbar ist. Ein **Wahrscheinlichkeitsmodell** beschreibt, welche Ereignisse in einem solchen Experiment möglich sind und welche Chancen die verschiedenen Ergebnisse haben. Ein Wahrscheinlichkeitsmodell erlaubt mittels Simulation mögliche Ergebnisse zu erzeugen und so eine Vorstellung der zufälligen Variabilität zu gewinnen.

Ein Wahrscheinlichkeitsmodell hat die folgenden Komponenten:

- **Grundraum** Ω , bestehend aus den **Elementarereignissen** ω ,
- **Ereignisse** A, B, C, \dots ,
- **Wahrscheinlichkeiten** P .

Elementarereignisse sind mögliche Ergebnisse oder Ausgänge des Experiments, die zusammen den Grundraum bilden:

$$\Omega = \underbrace{\{\text{mögliche Elementarereignisse } \omega\}}_{\text{mögliche Ausgänge/Resultate}}$$

Bei der Durchführung des Experiments wird ein Elementarereignis zufällig gewählt.

Beispiel: 2-maliges Werfen einer Münze

$\Omega = \{KK, KZ, ZK, ZZ\}$ wobei $K = \text{“Kopf”}$ und $Z = \text{“Zahl”}$ bezeichnet. Ein Elementarereignis ist zum Beispiel $\omega = KZ$.

Unter einem **Ereignis** A versteht man eine Teilmenge von Ω :

$$\text{Ereignis } A \subset \Omega$$

“Ein Ereignis A tritt ein“ bedeutet, dass das Ergebnis ω des Experiments zu A gehört.

Beispiel (Forts.): $A = \{\text{genau 1-mal Kopf}\} = \{KZ, ZK\}$.

Die Operationen der Mengenlehre (Komplement, Vereinigung, Durchschnitt) werden für Ereignisse verwendet:

$$\begin{aligned} A \cup B &\Leftrightarrow A \text{ oder } B, \text{ wobei das “oder” nicht-exklusiv ist (“oder/und”)} \\ A \cap B &\Leftrightarrow A \text{ und } B \\ A^c &\Leftrightarrow \text{nicht } A \end{aligned}$$

Beispiel: $A = \text{morgen scheint die Sonne}$, $B = \text{morgen regnet es}$.

$A \cup B$ bedeutet: morgen scheint die Sonne oder morgen regnet es (und dies kann auch bedeuten, dass morgen die Sonne scheint und dass es morgen regnet); $A \cap B$ bedeutet: morgen scheint die Sonne und morgen regnet es; A^c bedeutet: morgen scheint die Sonne nicht.

Jedem Ereignis A wird schliesslich eine **Wahrscheinlichkeit** $P(A)$ zugeordnet. Dabei müssen die folgenden drei grundlegenden Regeln (**Axiome von Wahrscheinlichkeitsrechnung**) erfüllt sein:

1. Die Wahrscheinlichkeiten sind immer nicht-negativ: $P(A) \geq 0$
2. Das Ereignis Ω hat Wahrscheinlichkeit eins: $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$ falls $A \cap B = \emptyset$, d.h. für alle Ereignisse, die sich gegenseitig ausschliessen.

Beispiel (Forts.) Beim Wurf zweier Münzen ist es plausibel, dass alle 4 Elemente von Ω gleich wahrscheinlich sind. Wegen $P(\Omega) = 1$ müssen sich die Wahrscheinlichkeiten zu Eins addieren:

$$P(KK) = P(KZ) = P(ZK) = P(ZZ) = \frac{1}{4}.$$

Weitere Regeln können daraus abgeleitet werden. *Beispiel:*

$$\begin{aligned} P(A^c) &= 1 - P(A), \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad (\text{Additionssatz}). \end{aligned}$$

Im Wesentlichen werden in der Wahrscheinlichkeitstheorie die Wahrscheinlichkeiten gewisser Ereignisse A festgelegt (auf Grund von Plausibilitäten, Symmetrieüberlegungen, wissenschaftlichen Theorien, Fachwissen und Daten) und daraus die Wahrscheinlichkeiten von gewissen anderen Ereignissen B aus den obigen Gesetzen hergeleitet.

Die Statistik geht umgekehrt vor: aus Daten, d.h. aus der Information, dass gewisse Ereignisse eingetreten sind, versucht man Rückschlüsse auf ein unbekanntes Wahrscheinlichkeitsmodell (unbekannte Wahrscheinlichkeiten) zu machen.

Interpretation von Wahrscheinlichkeiten:

- **Frequentistisch:** Idealisierung der relativen Häufigkeiten bei vielen unabhängigen Wiederholungen
- **Bayes'sch:** Mass für den Glauben, dass ein Ereignis eintreten wird

Wir behandeln in diesem Kapitel diskrete Wahrscheinlichkeitsmodelle, bei denen der Grundraum endlich oder "abzählbar" ist (d.h. man kann die Elementarereignisse durchnummerieren). Zum Beispiel ist $\Omega = \{0, 1, \dots, 10\}$ endlich und deshalb diskret; $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ ist zwar unendlich, aber noch abzählbar und daher trotzdem diskret; $\Omega = \mathbb{R}$ ist nicht abzählbar.

Im diskreten Fall ist die Wahrscheinlichkeit eines Ereignisses durch die Wahrscheinlichkeiten der zugehörigen Elementarereignisse $P(\{\omega\})$ festgelegt:

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Beispiel (Forts.) Für $A = \text{genau einmal Kopf} = \{KZ, ZK\}$ hat man also $P(A) = P(KZ) + P(ZK) = 1/4 + 1/4 = 1/2$.

In vielen Fällen ist es plausibel anzunehmen, dass jedes Elementarereignis die gleiche Wahrscheinlichkeit hat. In diesen Fällen gibt es eine besonders einfache Möglichkeit, die Wahrscheinlichkeit eines Ereignisses E , dass aus verschiedenen Elementarereignissen besteht ($E = \{\omega_1, \omega_2, \dots, \omega_g\}$; Grundraum $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$) zu berechnen.

$$P(E) = \frac{g}{m}$$

Man teilt die Anzahl für das Ereignis “**günstigen**“ Elementarereignisse durch die Anzahl der “**möglichen**“ Elementarereignisse.

Beispiel: Es werden zwei Würfel geworfen. Was ist die Wahrscheinlichkeit, dass die Augensumme 7 ergibt? Ein Elementarereignis beschreibt die Augenzahlen auf beiden Würfeln, also z.B. (1, 4), wenn der eine Würfel eine 1 und der andere eine 4 zeigt. Es sind insgesamt 36 Elementarereignisse möglich: $\{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$. Es gibt davon 6 Elementarereignisse, bei denen die Augensumme 7 ist: $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Da alle Elementarereignisse gleich wahrscheinlich sind, ist die Wahrscheinlichkeit für das Ereignis E : $P(E) = \frac{6}{36} = \frac{1}{6}$.

3.2 Der Begriff der Unabhängigkeit (Stahel, Kap. 4.6)

Wenn man die Wahrscheinlichkeiten $P(A)$ und $P(B)$ kennt, lässt sich daraus $P(A \cap B)$ im Allgemeinen nicht berechnen: Es sind alle Werte zwischen 0 und dem Minimum von $P(A)$ und $P(B)$ möglich. Ein wichtiger Spezialfall liegt vor, wenn folgende Produktformel gilt

$$P(A \cap B) = P(A)P(B).$$

Dann heißen A und B **stochastisch unabhängig**.

Beispiel (Forts.): Es sei $A = K$ im 1. Wurf und $B = K$ im 2. Wurf. Dann gilt $P(A) = P(KK) + P(KZ) = \frac{1}{2}$ und analog $P(B) = \frac{1}{2}$. D.h., $P(A)P(B) = \frac{1}{4}$. Wegen $P(A \cap B) = P(KK) = \frac{1}{4}$, sind also A und B unabhängig.

Viel wichtiger ist jedoch der umgekehrte Schluss. Wenn zwischen den Ereignissen A und B kein kausaler Zusammenhang besteht (d.h. es gibt keine gemeinsamen Ursachen oder Ausschlüssungen), dann **postuliert** man stochastische Unabhängigkeit und nimmt damit an, dass obige Produktformel gilt. In diesem Fall kann also $P(A \cap B)$ aus $P(A)$ und $P(B)$ berechnet werden.

Beispiel (Forts.): Es ist plausibel, dass es keinen kausalen Zusammenhang zwischen dem Ergebnis des ersten und des zweiten Wurfs gibt. Die Ereignisse A und B sind also unabhängig. Deshalb kann man $P(A \cap B)$ wie folgt ausrechnen: $P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Beispiel: Angenommen, wir werfen zwei sechsstellige Würfel. Sei E das Ereignis, dass die Augensumme 6 ist. Sei F das Ereignis, dass der erste Würfel die Augenzahl 4 zeigt. Dann gilt

$$P(E \cap F) = P(\{4, 2\}) = \frac{1}{36}$$

aber

$$P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6} = \frac{5}{216}$$

Deshalb sind E und F nicht unabhängig. Das Ergebnis kann man intuitiv einfach verstehen. Wenn wir daran interessiert sind mit beiden Würfeln die Augensumme 6 zu würfeln, sollten wir recht froh sein, wenn der erste Wurf eine 4 zeigt (oder irgendeine Zahl aus 1,2,3,4,5), denn dann haben wir noch eine Chance nach dem zweiten Wurf auf Augensumme 6 zu kommen. Falls aber der erste Wurf schon eine 6 zeigt, gibt es keine Chance mehr auf die Augensumme

6 zu kommen. D.h., die Chance die Augensumme 6 zu erzielen hängt von der Augenzahl des ersten Wurfs ab. Daher können E und F nicht unabhängig sein.

Bei mehr als zwei Ereignissen A_1, \dots, A_n bedeutet Unabhängigkeit, dass die entsprechende Produktformel für alle k -Tupel von Ereignissen gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

für jedes $k = 2, 3, \dots, n$ und jedes $1 \leq i_1 < \dots < i_k \leq n$. Insbesondere muss also gelten

$$P(A_1 \cap A_2) = P(A_1)P(A_2), \quad P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3), \text{ etc.}$$

3.3 Bedingte Wahrscheinlichkeit (Stahel, Kap. 4.7)

Was ist die Wahrscheinlichkeit, dass ein sechseckiger Würfel die Zahl 3 zeigt, wenn wir schon wissen, dass eine ungerade Zahl geworfen wurde? Es gibt drei ungerade Zahlen auf dem Würfel (1,3,5). Die 3 ist eine davon. Wenn man davon ausgehen kann, dass jede Zahl auf dem Würfel gleich wahrscheinlich ist, ist das Ergebnis $\frac{1}{3}$. Fragen von dieser Art kann man auch in komplizierteren Sachverhalten mit dem Begriff der **bedingten Wahrscheinlichkeit** leicht beantworten.

Die bedingte Wahrscheinlichkeit von Ereignis A gegeben B , d.h., die Wahrscheinlichkeit von A , wenn wir wissen, dass B eingetreten ist, wird mit $P(A|B)$ bezeichnet und so definiert:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Beispiel: Wir rechnen nochmals das Beispiel aus dem Anfang dieses Unterkapitels. Sei A = “Die 3 wird geworfen” und B = “Eine ungerade Zahl wird geworfen”. Für die bedingte Wahrscheinlichkeit $P(A|B)$ gilt also

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Im allgemeinen ist $P(A|B)$ **nicht** das gleiche wie $P(B|A)$. Dieser Fehler wird vor Gericht (und nicht nur da) so oft gemacht, dass er schon einen eigenen Namen hat: Prosecutor’s fallacy¹.

Beispiel: Ein medizinischer Test für eine Krankheit (D falls vorhanden, D^c falls nicht vorhanden) kann positiv (+) oder negativ (−) sein. Die Wahrscheinlichkeiten sind in folgender Tabelle gegeben. Z.B. ist die Wahrscheinlichkeit, dass die Krankheit vorhanden ist und der Test positiv ausfällt $P(D \cap +) = 0.009$.

	D	D^c
+	0.009	0.099
−	0.001	0.891

¹Suchen Sie in Wikipedia danach!

Daraus folgt

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9$$

und

$$P(-|D^c) = \frac{P(- \cap D^c)}{P(D^c)} = \frac{0.891}{0.891 + 0.099} = 0.9.$$

Offensichtlich ist dieser Test recht genau. Kranke Personen werden zu 90% als positiv eingestuft, und gesunde Personen werden zu 90% als negativ eingestuft. Angenommen, Sie gehen zu einem Test und werden als positiv eingestuft. Wie gross ist die Wahrscheinlichkeit, dass Sie die Krankheit wirklich haben? Die meisten Leute würden 0.9 antworten. Die richtige Antwort ist aber

$$P(D|+) = \frac{P(+ \cap D)}{P(+)} = \frac{0.009}{0.009 + 0.099} = 0.08.$$

Die Lektion ist: Bei bedingten Wahrscheinlichkeiten darf man seiner Intuition nicht trauen, sondern muss das Ergebnis ausrechnen.

Das **Bayes Theorem** liefert einen oft nützlichen Zusammenhang zwischen $P(A|B)$ und $P(B|A)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Beispiel (Forts.): Das Bayes Theorem liefert die gleiche Lösung wie unsere obige Rechnung:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.9 \cdot (0.009 + 0.001)}{0.009 + 0.099} = \frac{0.009}{0.009 + 0.099} = 0.08.$$

Eine zweite nützliche Formel ist das **Gesetz der totalen Wahrscheinlichkeit**: Betrachte Ereignisse A_1, \dots, A_k , die miteinander keine Schnittmenge haben und zusammen alle möglichen Fälle erfassen². Dann gilt für jedes beliebige Ereignis B :

$$P(B) = \sum_{i=1}^k P(B|A_i)P(A_i)$$

Beispiel (fktiv): Ich teile meine emails in drei Kategorien ein: A_1 = "spam", A_2 = "niedrige Priorität" und A_3 = "hohe Priorität". Von früheren Beobachtungen weiss ich, dass $P(A_1) = 0.7$, $P(A_2) = 0.2$ und $P(A_3) = 0.1$ ³. Sei B das Ereignis, dass das Wort ffreein der email auftaucht. Von früheren Beobachtungen weiss ich, dass $P(B|A_1) = 0.9$, $P(B|A_2) = 0.01$ und $P(B|A_3) = 0.01$ ⁴. Angenommen, ich erhalte eine email, die das Wort ffreeenthält. Wie gross

²Eine solche Aufteilung nennt man eine **Partitionierung**. Z.B. bei einem Münzwurf: A_1 = Kopf, A_2 = Zahl; bei einem Würfelwurf: A_1 = gerade, A_2 = ungerade; oder $A_1 = \{1\}$, $A_2 = \{2,4\}$, $A_3 = \{3,5,6\}$

³Beachte, dass $P(A_1) + P(A_2) + P(A_3) = 1$, wie es bei einer Partitionierung auch sein sollte.

⁴Beachte, dass hier die Summe nicht 1 ergibt.

ist die Wahrscheinlichkeit, dass es sich um spam handelt? Das Bayes Theorem zusammen mit dem Gesetz der totalen Wahrscheinlichkeit liefern die Lösung:

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} = \\ &= \frac{0.9 \cdot 0.7}{(0.9 \cdot 0.7) + (0.01 \cdot 0.2) + (0.01 \cdot 0.1)} = 0.995 \end{aligned}$$

Viele Spamfilter basieren tatsächlich auf diesem Prinzip.

Um zu beschreiben, wie wahrscheinlich ein Ereignis E ist, kann man die Wahrscheinlichkeit $P(E)$ angeben. In den Life Sciences ist eine zweite Beschreibung der Wahrscheinlichkeit sehr verbreitet: Die **odds** des Ereignisses E , $odds(E)$. Die odds von einem Ereignis und die Wahrscheinlichkeit von einem Ereignis können immer ineinander umgerechnet werden. Sie drücken also ein- und denselben Sachverhalt auf zwei verschiedene Arten aus. Die $odds(E)$ geben an, wieviel mal wahrscheinlicher das Eintreten von E als das Eintreten von E^c ("nicht E ") ist.

$$odds(E) = \frac{P(E)}{1 - P(E)}$$

Beispiel: Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person Krankheit K hat, ist $P(K) = 0.1$. Die odds, dass eine zufällig ausgewählte Person Krankheit K hat, sind also $odds(K) = 0.1/0.9 = \frac{1}{9}$. Es ist also neunmal so wahrscheinlich, dass eine Person Krankheit K nicht hat als dass sie K hat.

Beispiel: Angenommen, die odds eines Ereignisses sind $odds(E) = 3$. Dann ist $\frac{P(E)}{1-P(E)} = 3$. Auflösen nach $P(E)$ liefert: $P(E) = \frac{odds(E)}{1+odds(E)} = \frac{3}{4}$.

Angenommen, wir haben zwei Gruppen $G = 1$ und $G = 2$. Die odds für das Ereignis E in Gruppe 1 sind

$$odds(E|G = 1) = \frac{P(E|G = 1)}{1 - P(E|G = 1)}.$$

Analog sind die odds für das Ereignis E in Gruppe 2

$$odds(E|G = 2) = \frac{P(E|G = 2)}{1 - P(E|G = 2)}.$$

Das Verhältnis dieser odds wird **odds-Ratio** genannt (OR).⁵

Beispiel: Personen können gegen eine Krankheit K geimpft werden. Gegeben man ist geimpft ($I = 1$), ist die Wahrscheinlichkeit Krankheit K zu bekommen nur 0.0001, d.h., $P(K|I = 1) = 0.0001$. Die odds für die Krankheit gegeben man ist geimpft sind also $odds(K|I = 1) = \frac{P(K|I=1)}{1-P(K|I=1)} = \frac{0.0001}{0.9999} \approx 0.0001$. Wenn man nicht geimpft ist ($I = 0$), ist die Wahrscheinlichkeit

⁵Falls Sie es später mal brauchen: Mit der Funktion `fisher.test` können Sie in R odds-Ratio und Vertrauensintervalle dazu berechnen. Das werden wir in dieser Vorlesung aber nicht behandeln.

Krankheit K zu bekommen 0.2, d.h., $P(K|I = 0) = 0.2$. Die odds sind in diesem Fall also $odds(K|I = 0) = \frac{0.2}{0.8} = 0.25$. Das odds-Ratio OR ist dann:

$$OR = \frac{odds(K|I = 0)}{odds(K|I = 1)} = \frac{0.25}{0.0001} = 2500$$

(Ob das odds-Ratio als $\frac{odds(K|I=0)}{odds(K|I=1)}$ oder als $\frac{odds(K|I=1)}{odds(K|I=0)}$ definiert wird, muss angegeben werden; beides ist möglich, solange die Definition klar gemacht wird.). Die odds zu erkranken werden durch die Impfung um dem Faktor 2500 reduziert. Die Impfung ist also sehr wirksam.

3.4 Zufallsvariable (Stahel, Kap. 4.3, 4.4)

Oft sind mit einem Zufallsexperiment Zahlenwerte verknüpft, d.h. zu jedem Elementarereignis ω gehört ein Zahlenwert $X(\omega) = x$.

Beispiel: Wert einer gezogenen Jass-Karte.

$$\begin{aligned} \omega = \text{As} &\mapsto X(\omega) = 11 \\ \omega = \text{König} &\mapsto X(\omega) = 4 \\ &\vdots \\ \omega = \text{Sechs} &\mapsto X(\omega) = 0 \end{aligned}$$

Wie man in diesem Beispiel sieht, ist X eine Funktion auf dem Grundraum Ω . Wir halten fest:

Eine **Zufallsvariable** X ist eine **Funktion**:

$$\begin{aligned} X : \quad \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

Die Notation X (oder auch Y, Z, \dots) ist eher ungewohnt für die Bezeichnung einer Funktion, ist aber üblich in der Wahrscheinlichkeitsrechnung. Sie haben hoffentlich in der Analysis gesehen, dass man Funktionen wie Zahlen addieren oder multiplizieren kann (man addiert oder multipliziert einfach die Funktionswerte). Summen und Produkte von Zufallsvariablen sind also wieder Zufallsvariablen.

Konvention: Eine Zufallsvariable wird durch einen **Grossbuchstaben** (z.B. X) dargestellt. Der gleiche **Kleinbuchstabe** (z.B. x) stellt einen konkreten Wert dar, den die Zufallsvariable annehmen kann. Das Ereignis, dass die Zufallsvariable X den Wert x annimmt, können wir dann so schreiben: $X = x$.

Bei einer Zufallsvariable ist nicht die Funktion $X(\cdot)$ zufällig, sondern nur das Argument ω : Je nach Ausgang des Zufallsexperiments (d.h. von ω) erhalten wir einen anderen Wert $x = X(\omega)$, x ist **eine Realisierung** der Zufallsvariablen X . Wenn wir aber das Experiment zweimal durchführen und zwei Mal das gleiche Ergebnis ω herauskommt, dann sind auch die realisierten Werte von X gleich.

Wenn der Grundraum Ω diskret ist, dann muss auch der Wertebereich $W = W_X$ (Menge der möglichen Werte von X) diskret sein, d.h. endlich oder abzählbar. Wir werden in diesem Kapitel bloss diskrete Zufallsvariablen genauer diskutieren. Insbesondere sind Anzahlen stets diskret, während Messungen besser als kontinuierlich, d.h. mit dem Wertebereich \mathbb{R} modelliert werden (obwohl man praktisch nur mit endlicher Genauigkeit messen kann).

Wahrscheinlichkeitsverteilung einer Zufallsvariablen

Die Werte einer Zufallsvariablen X (die möglichen Realisationen von X) treten mit gewissen Wahrscheinlichkeiten auf. Die Wahrscheinlichkeit, dass X den Wert x annimmt, berechnet sich wie folgt:

$$P(X = x) = P(\{\omega; X(\omega) = x\}) = \sum_{\omega; X(\omega)=x} P(\omega).$$

Beispiel (Forts): X = Wert einer gezogenen Jass-Karte.

$$\begin{aligned} & \text{Wahrscheinlichkeit für Zahl 4} = P(X = 4) \\ &= P(\{\omega; \omega = \text{ein König}\}) \\ &= P(\text{Eicheln-König}) + P(\text{Rosen-König}) + P(\text{Schellen-König}) + P(\text{Schilten-König}) \\ &= 4/36 = 1/9. \end{aligned}$$

Die “Liste” von $P(X = x)$ für alle möglichen Werte x heisst (diskrete) **(Wahrscheinlichkeits-) Verteilung** der (diskreten) Zufallsvariablen X . Dabei gilt immer

$$\sum_{\text{alle möglichen } x} P(X = x) = 1.$$

Beispiel (Forts): X = Wert einer gezogenen Jass-Karte.

Die Wahrscheinlichkeitsverteilung von X ist

x	0	2	3	4	10	11
$P(X = x)$	4/9	1/9	1/9	1/9	1/9	1/9

Umgekehrt ist jede Liste von nichtnegativen Zahlen, die sich zu eins addieren, die Verteilung einer gewissen Zufallsvariablen. Wenn man nur an der Zufallsvariablen X interessiert ist, kann man den zu Grunde liegenden Raum Ω vergessen, man braucht nur die Verteilung von X . Zufallsvariablen sind einfach Zufallsexperimente, bei denen die Ergebnisse Zahlen sind.

3.5 Binomialverteilung (Stahel Kap. 5.1)

Wir betrachten die Situation, wo es um das Zählen der Anzahl Erfolge (oder Misserfolge) geht. Solche Anwendungen treten z.B. auf bei der Qualitätskontrolle, Erfolg/Misserfolg bei Behandlungen (medizinisch, biologisch) oder auch bei Glücksspielen.

Die Verteilung einer Zufallsvariable X mit Werten in $W = \{0, 1\}$ kann durch einen einzelnen Parameter π beschrieben werden:

$$P(X = 1) = \pi, \quad P(X = 0) = 1 - \pi, \quad 0 \leq \pi \leq 1.$$

Diese Verteilung heisst **Bernoulli(π)-Verteilung**. Sie beschreibt einfach das Eintreffen oder Nicht-Eintreffen eines bestimmten Ereignisses, z.B. das Ergebnis “Kopf“ beim Werfen einer Münze. Falls die Münze fair ist, so ist $\pi = 1/2$.

Etwas interessanter wird es, wenn wir das Experiment n Mal wiederholen, also z.B. die Münze n -mal werfen. Der Grundraum Ω besteht dann aus allen “Wörtern“ der Länge n , welche man mit den Buchstaben K (für “Kopf“) und Z (für “Zahl“) schreiben kann. Ω hat also 2^n Elemente. Wir betrachten die Zufallsvariablen

$$X_i = \begin{cases} 1 & \text{falls K im } i\text{-ten Wurf} \\ 0 & \text{falls Z im } i\text{-ten Wurf.} \end{cases}$$

$$X = \sum_{i=1}^n X_i = \text{Gesamtzahl von Würfeln mit K}$$

Um die Verteilung von X bestimmen zu können, müssen wir eine Wahrscheinlichkeit auf Ω festlegen. Wir postulieren, dass die Ereignisse $X_i = 1$ (also “K im i -ten Wurf“) alle die Wahrscheinlichkeit π haben und unabhängig sind. Dann gilt zum Beispiel:

$$\begin{aligned} P(X = 0) &= P(X_1 = \dots = X_n = 0) = (1 - \pi)^n, \\ P(X = 1) &= P(\text{ein } X_i = 1 \text{ und alle anderen } X_j = 0) = \\ &= P(X_1 = 1, X_2 = 0, \dots, X_n = 0) + P(X_1 = 0, X_2 = 1, X_3 = 0, \dots, X_n = 0) + \dots \\ &= n\pi(1 - \pi)^{n-1}. \end{aligned}$$

Um $P(X = x)$ zu berechnen, muss man offenbar bestimmen, auf wieviele Arten man x Einer auf n Plätze anordnen kann. Die Antwort ist gegeben durch den Binomialkoeffizienten

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

So kommt man auf die Binomial-Verteilung.

Binomial(n, π)-Verteilung:

Eine Zufallsvariable X mit Werten in $W = \{0, 1, \dots, n\}$ heisst Binomial(n, π)-verteilt, falls

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, \dots, n.$$

Dabei ist $0 \leq \pi \leq 1$ der Erfolgsparameter der Verteilung.

Wie in obigem Beispiel motiviert, ist die Binomialverteilung angebracht für die Zufallsvariable “Anzahl Erfolge/Misserfolge“ (Eintreten eines bestimmten Ereignis) bei n **unabhängigen** Versuchen. Das Prädikat “unabhängig“ ist wesentlich für die Korrektheit der Binomialverteilung.

Konvention: Wenn man notieren will, dass die Zufallsvariable X einer gewissen Wahrscheinlichkeitsverteilung F folgt, schreibt man abgekürzt: $X \sim F$. Dabei kann F von Parametern

abhängen, also z.B. $X \sim F(\theta)$. Wenn also X einer Binomial-Verteilung mit Parametern n und π folgt, schreibt man abgekürzt $X \sim \text{Binomial}(n, \pi)$ oder einfach nur $X \sim \text{Bin}(n, \pi)$.

Beispiel Bei einer Losbude steht: “Jedes 5. Los gewinnt!“, d.h., die Gewinnwahrscheinlichkeit ist bei jedem Los $\pi = 0.2$. Nehmen wir weiter an, dass das Ziehen von einem Los keinen Einfluss auf das Ziehen des nächsten Loses hat (z.B. gibt es eine riesige Anzahl Lose und die Lostrommel wird nach jedem Verkauf eines Loses gut gemischt). Wir kaufen 100 Lose und bezeichnen mit X die Anzahl Gewinne unter den 100 Losen. Dann ist X Binomial($n = 100$, $\pi = 0.2$) verteilt. Abgekürzt: $X \sim \text{Binomial}(100, 0.2)$.

Beispiel: Spermasexing (Tages-Anzeiger 6.12.2000)

Geschlechts-Beeinflussung von Kuhkälbern mit einer Methode, die Spermasexing genannt wird. Ziel ist es, ein weibliches Kalb zu züchten. In einem Testlauf wurden zwölf Kühe mit Spermien besamt, die optisch nach dem Y-Chromosom sortiert wurden (d.h. mit der Spermasexing-Methode). Da die Methode nicht hundertprozentig sicher ist, können wir das als Zufallsexperiment auffassen. Sei X = Anzahl weiblicher gezüchteter Kuhkälber. Ein vernünftiges Modell ist dann:

$$X \sim \text{Binomial}(12, \pi),$$

wobei π unbekannt ist. Effektiv beobachtet wurden $x = 11$ weiblich gezüchtete Kuhkälber: d.h. $X = x = 11$ wurde tatsächlich **realisiert**. Später mehr dazu.

Eigenschaften der Binomialverteilung (siehe Abb. 3.1): $P(X = x)$ ist maximal wenn x gleich dem ganzzahligen Teil von $(n + 1)\pi$ ist, und auf beiden Seiten von diesem Wert nehmen die Wahrscheinlichkeiten monoton ab. Wenn $n\pi(1 - \pi)$ nicht allzu klein ist, ist die Verteilung praktisch symmetrisch und hat die Form einer Glocke. Wenn n gross ist, sind die meisten Wahrscheinlichkeiten $P(X = x)$ verschwindend klein, d.h. grosse Abweichungen von $(n + 1)\pi$ sind extrem unwahrscheinlich.

3.6 Kennzahlen einer Verteilung (Stahel Kap. 5.3)

Eine beliebige (diskrete) Verteilung kann vereinfachend zusammengefasst werden durch 2 Kennzahlen, den **Erwartungswert** $E(X)$ und die **Standardabweichung** $\sigma(X)$.

Der Erwartungswert beschreibt die mittlere Lage der Verteilung und ist wie folgt definiert:

$$E(X) = \sum_{x \in W_X} xP(X = x), \quad W_X = \text{Wertebereich von } X.$$

Die Standardabweichung beschreibt die Streuung der Verteilung. Rechnerisch ist das Quadrat der Standardabweichung, die sogenannte **Varianz** bequemer:

$$\begin{aligned} \text{Var}(X) &= \sum_{x \in W_X} (x - E(X))^2 P(X = x) \\ \sigma(X) &= \sqrt{\text{Var}(X)}. \end{aligned}$$

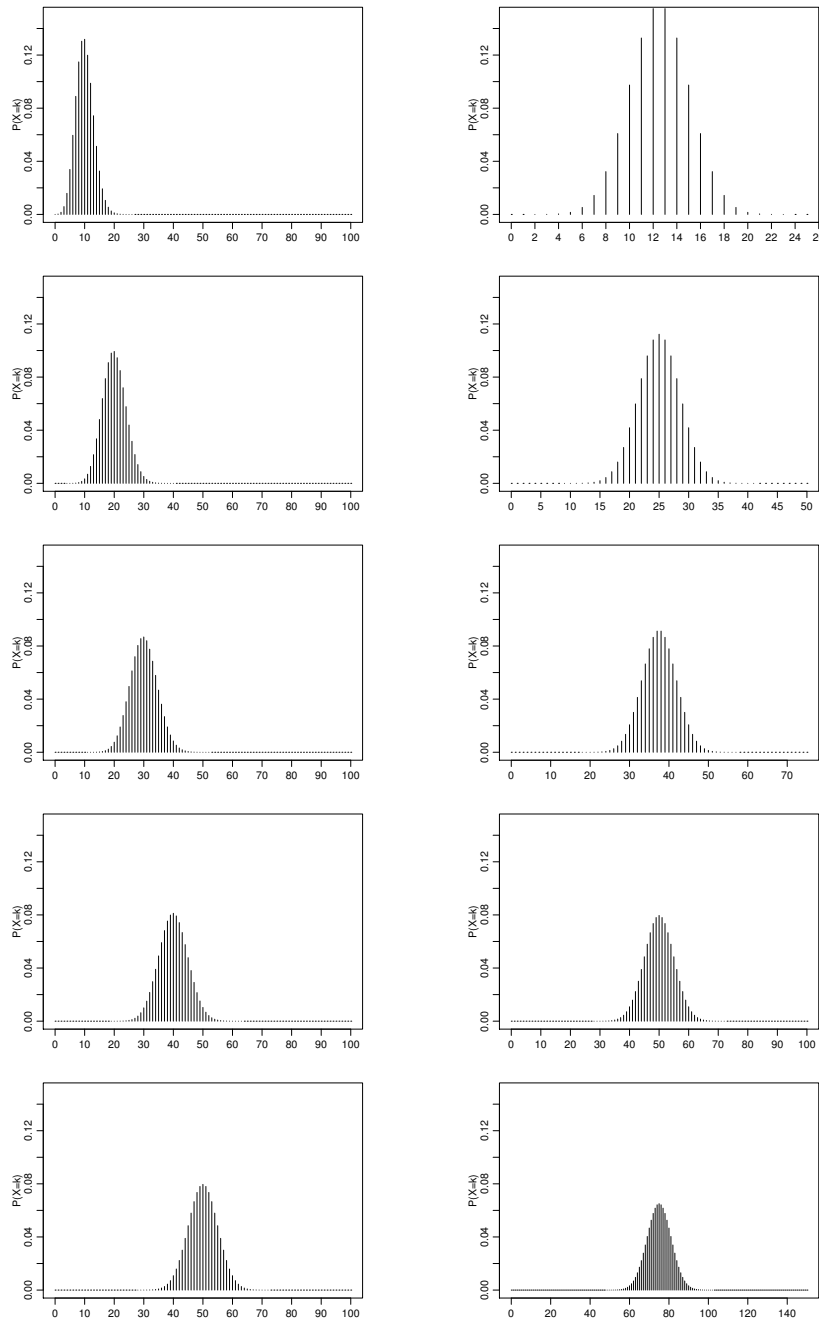


Abbildung 3.1: Die Binomialwahrscheinlichkeiten $P(X = x)$ als Funktion von x für verschiedene n 's und π 's. Links ist $n = 100$ und $\pi = 0.1, 0.2, 0.3, 0.4, 0.5$ (von oben nach unten) und rechts ist $\pi = 0.5$ und $n = 25, 50, 75, 100, 150$ (von oben nach unten).

Die Standardabweichung hat dieselbe Einheit wie X , während die Einheit der Varianz deren Quadrat ist: Wird z.B. X in Metern (m) gemessen, so besitzt $\text{Var}(X)$ die Dimension Quadratmeter (m^2) und $\sigma(X)$ wiederum die Dimension Meter (m).

Beispiel: Sei $X \sim \text{Bernoulli}(\pi)$.

Dann:

$$\begin{aligned} E(X) &= 0 \cdot P(X=0) + 1 \cdot P(X=1) = \pi, \\ \text{Var}(X) &= (0 - E(X))^2 P(X=0) + (1 - E(X))^2 P(X=1) = (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi \\ &= \pi(1 - \pi), \\ \sigma(X) &= \sqrt{\pi(1 - \pi)}. \end{aligned}$$

Für die Binomial-Verteilung erhält man mit einigen Rechnungen

$$X \sim \text{Binomial}(n, \pi) \rightarrow E(X) = n\pi, \quad \text{Var}(X) = n\pi(1 - \pi), \quad \sigma(X) = \sqrt{n\pi(1 - \pi)}.$$

(Weil $\text{Bernoulli}(\pi) = \text{Binomial}(1, \pi)$, stimmt das mit obigen Formeln überein). Die Kennzahlen fassen also sehr gut zusammen, was wir in der Abbildung 3.1 gesehen haben: Die Verteilung ist um den Erwartungswert konzentriert, die Streuung wächst mit n , aber langsamer als n . Für festes n ist die Streuung maximal, wenn $\pi = 1/2$.

Beispiel (Forts.) Wir sind wieder bei der Losbude, bei der wir (nach dem vierten Bier) 100 Lose gekauft hatten. Um die Freundin zu beeindrucken, kramen wir unser Statistikwissen hervor und berechnen im Kopf den Erwartungswert und die Standardabweichung der Anzahl Gewinne unter 100 Losen.

$$\begin{aligned} E(X) &= n \cdot \pi = 100 \cdot 0.2 = 20 \\ \sigma(X) &= \sqrt{n\pi(1 - \pi)} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4 \end{aligned}$$

Wir erinnern uns, dass Beobachtungen typischerweise ein bis zwei Standardabweichungen vom Erwartungswert entfernt liegen und prophezeien der Freundin mit Stolz geschwellter Brust, dass wir wohl zwischen 16 und 24 Gewinnen zu erwarten haben. Sie solle sich schonmal einen Teddybär aussuchen. Hundert nervenaufreibende Öffnungsversuche später stehen wir mit nur 8 Gewinnen und hängenden Schultern da. Kann das Zufall sein? Ja, aber wir beobachten ein sehr unwahrscheinliches Ereignis (später werden wir solche Überlegungen mit einem statistischen Test präzise formulieren können). Wir beschliessen, den hühnerhaften und grimmig drein schauenden Losbudenbesitzer nicht mit unserer Erkenntnis zu verärgern und denken uns: “Pech im Spiel, Glück in der Liebe“

3.6.1 Kumulative Verteilungsfunktion

Manchmal ist es für Rechnungen nützlicher, statt der “Liste” $P(X = x)$ (für alle x) die sukzessiven Summen

$$\sum_{y \in W_X; y \leq x} P(X = y) = P(X \leq x)$$

anzugeben. Dabei läuft x ebenfalls über den Wertebereich W_X von X . Man kann in dieser Definition aber auch beliebige reelle Werte x betrachten und erhält dann eine Funktion

$$F(x) = P(X \leq x) = \sum_{y \in W_X; y \leq x} P(X = y),$$

die sogenannte **kumulative Verteilungsfunktion**. Diese springt an den Stellen, die zum Wertebereich gehören, und ist dazwischen konstant. Siehe auch Abbildung 3.2.

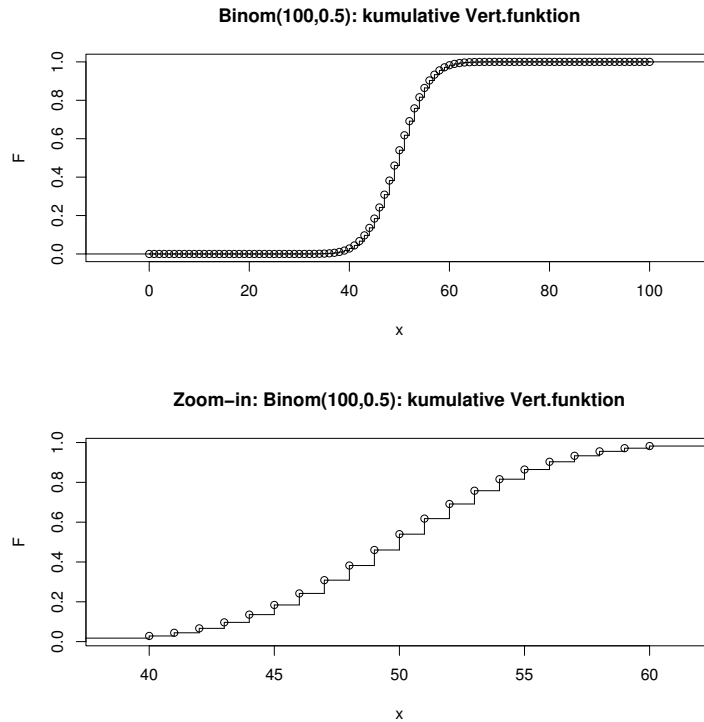


Abbildung 3.2: Kumulative Verteilungsfunktion $F(\cdot)$ für $X \sim \text{Binomial}(100, 0.5)$. Unten: zoom-in für die Werte $x \in [40, 60]$. Die Kreise zeigen an, dass an den Sprungstellen der obere Wert gilt.

Aus der kumulativen Verteilungsfunktion kann man die “Liste” $P(X = x)$ zurückgewinnen: $P(X = x)$ ist einfach die Höhe des Sprungs an der Stelle x . Insbesondere gilt für X mit Wertebereich in den ganzen Zahlen und ganzzahliges x

$$P(X = x) = F(x) - F(x - 1), \quad P(X \geq x) = 1 - P(X \leq x - 1) = 1 - F(x - 1).$$

3.7 Poissonverteilung (Stahel Kap. 5.2)

Der Wertebereich der $\text{Binomial}(n, \pi)$ -Verteilung ist $W = \{0, 1, \dots, n\}$. Falls eine Zufallsvariable nicht im vornherein einen beschränkten Wertebereich hat, so bietet sich für Zählraten die Poisson-Verteilung an.

Eine Zufallsvariable X mit Werten in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ heisst Poisson(λ)-verteilt, falls

$$P(X = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad (x = 0, 1, 2, \dots)$$

wobei $\lambda > 0$ ein Parameter der Verteilung ist.

Die Poisson-Verteilung ist die Standardverteilung für unbeschränkte **Zählraten**.

Beispiele: Die Poisson(λ)-Verteilung kann bei folgenden Anwendungen als Modell gebraucht werden:

Anzahl Schadenmeldungen eines Versicherten pro Jahr,

Anzahl spontaner Ereignisse in einer Nervenzelle während einer Sekunde via Transmitterfreisetzung an einer Synapse.

Die Kennzahlen sind wie folgt: für $X \sim \text{Poisson}(\lambda)$:

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda, \quad \sigma(X) = \sqrt{\lambda}.$$

3.7.1 Poisson-Approximation der Binomial-Verteilung

Betrachte $X \sim \text{Binomial}(n, \pi)$ und $Y \sim \text{Poisson}(\lambda)$. Falls n gross und π klein mit $\lambda = n\pi$, dann:

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \approx P(Y = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad (x = 0, 1, \dots, n).$$

Das heisst: für grosse n und kleine π : $\text{Binomial}(n, \pi) \approx \text{Poisson}(\lambda)$ für $\lambda = n\pi$. Mit anderen Worten: die Poisson-Verteilung kann interpretiert werden als Verteilung für **seltenere Ereignisse bei vielen unabhängigen Versuchen** (selten für einen einzelnen Fall, die Gesamt-Anzahl kann trotzdem gross sein).

3.7.2 Summen von Poisson-verteilten Zufallsvariablen

Die Poisson-Verteilung hat die folgende Additionseigenschaft: Wenn $X \sim \text{Poisson}(\lambda_X)$ und $Y \sim \text{Poisson}(\lambda_Y)$ unabhängig sind, dann ist $X + Y \sim \text{Poisson}(\lambda_X + \lambda_Y)$. Wenn also zum Beispiel die Anzahlen spontaner Ereignisse in einer Nervenzelle in zwei disjunkten Zeitintervallen Poisson-verteilt und unabhängig sind, dann ist auch das Total wieder Poisson-verteilt. Wir erhalten also eine Poisson-Verteilung für alle Intervalle. Weil sich bei der Addition der Zufallsvariablen die Parameter der Poisson-Verteilung addieren, ist üblicherweise λ proportional zur Länge des betrachteten Zeitintervalls.

Oft entstehen Poisson verteilte Daten von einem sog. Poisson-Prozess auf einer Menge S (z.B. Zeit, Fläche oder Raum) mit Parameter λ . Das Modell sagt folgendes aus: Wenn S_1, S_2, \dots, S_n nicht überlappende Teilmengen von S sind (z.B. Zeitintervalle, Teile der Gesamtfläche oder Teile vom Raum), dann sind die Anzahl Ereignisse N_1, N_2, \dots, N_n in jeder Teilmenge unabhängige Zufallsvariablen, die jeweils einer Poisson-Verteilung mit den Parametern $\lambda \cdot |S_1|, \lambda \cdot |S_2|, \dots, \lambda \cdot |S_n|$ folgen. Dabei ist $|S_i|$ die "Grösse" von der Teilmenge S_i (z.B. Zeitdauer in Sekunden, Fläche in m^2 oder Volumen in m^3).

Beispiel Angenommen ein Büro erhält Telefonanrufe als ein Poisson-Prozess mit $\lambda = 0.5$ pro Minute⁶. Die Anzahl Anrufe in einem 5 Minuten Intervall folgt dann einer Poisson-Verteilung mit Parameter $\rho = 5 \cdot \lambda = 2.5$. Die Wahrscheinlichkeit, dass es in einem 5 Minuten Intervall

⁶Beachte, dass hier "pro Minute" steht. Der Parameter des Poisson-Prozesses misst immer eine Anzahl geteilt durch etwas (z.B. Zeitdauer, Fläche, Volumen, etc.)

keine Anrufe gibt, ist daher $\exp(-2.5) = 0.082$. Die Wahrscheinlichkeit, dass es genau einen Anruf gibt ist $2.5 \cdot \exp(-2.5)$.

3.8 Software

In der Statistiksoftware **R** sind sehr viele Wahrscheinlichkeitsverteilungen schon vorprogrammiert. Wenn man eine Zufallsvariable X hat, die einer Verteilung namens “xxx” (xxx steht für eine Abkürzung des Namens der Verteilung, z.B. “binom” oder “pois”) folgt, kann man folgende drei Dinge bequem erledigen:

dxxx Berechnet $P[X = x]$

pxxx Berechnet $P[X \leq x]$

rxxx Liefert eine Zufallszahl gemäss der Verteilung von X

Konkrete Beispiele sind: **dbinom**, **pbinom**, **rbinom** für die Binomialverteilung und **dpois**, **ppois**, **rpois** für die Poissonverteilung. Schauen Sie sich die Hilfeseite dieser Funktionen an, in dem Sie vor den Befehl ein “?” tippen. Also z.B. `?dbinom`.

In **R** gibt es noch weitere Modelle für Zähldaten: Geometrische Verteilung (Abkürzung **geom**), Negative Binomialverteilung (Abkürzung **nbinom**), Hypergeometrische Verteilung (Abkürzung **hyper**), usw. Die genannten Verteilungen werden in der Praxis oft gebraucht. Wenn es Sie interessiert, schauen Sie doch mal auf Wikipedia nach, in welchen Fällen diese drei Verteilungen Verwendung finden.