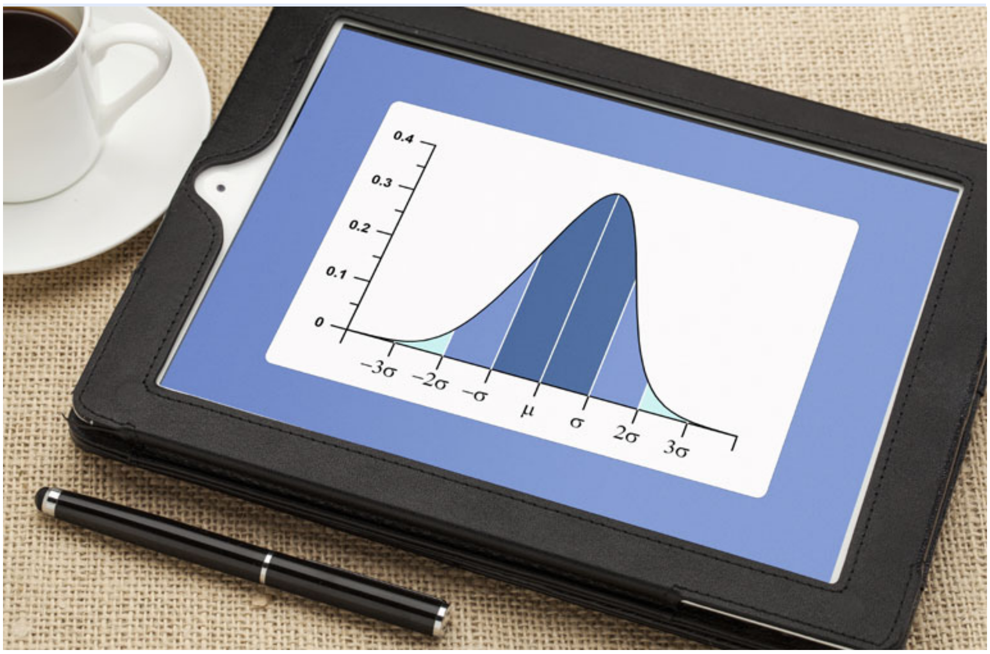
Wahrscheinlichkeitsrechnung – fundiert Prognosen erstellen

17



Was ist eigentlich eine Wahrscheinlichkeit?

Wozu braucht man Binomialverteilung und Normalverteilung?

Was ist der zentrale Grenzwertsatz?

Teil IV

L. Göllmann et al., *Mathematik für Ingenieure: Verstehen – Rechnen – Anwenden*, DOI 10.1007/978-3-662-53867-8_17

In diesem Kapitel verlassen wir den Bereich der deskriptiven Statistik und wenden uns dem theoretischen Konzept der Wahrscheinlichkeiten zu. Es geht ab jetzt also nicht mehr um beobachtbare Häufigkeiten, sondern um das Konzept, das dahinter steckt.

Zuerst beschäftigen wir uns mit Ereignissen, ihrer Modellierung und ihren Wahrscheinlichkeiten. Ein besonderer Schwerpunkt liegt hierbei auf bedingten Wahrscheinlichkeiten, da dieses Konzept erfahrungsgemäß besondere Schwierigkeiten verursacht.

Danach betrachten wir Zufallsvariablen und ihre Verteilungen samt den zugehörigen Kenngrößen näher. Den meist verbreiteten Verteilungen sind jeweils eigene Abschnitte gewidmet.

Nach einem Einblick in die Rechenregeln für Kenngrößen sehen wir uns Möglichkeiten der Wahrscheinlichkeitsabschätzung bei unbekannten Wahrscheinlichkeitsverteilungen an. Hier werden wir insbesondere eine Version des Gesetzes der großen Zahlen und den zentralen Grenzwertsatz kennenlernen.

17.1 Wahrscheinlichkeitsraum, Ereignisse, Wahrscheinlichkeiten

In der Wahrscheinlichkeitsrechnung werden Modelle für Zufallsexperimente entwickelt und untersucht. Dazu muss man natürlich wissen, was das ist.

Definition: Zufallsexperiment

Ein **Zufallsexperiment** E ist durch drei Eigenschaften gekennzeichnet:

- Die Menge Ω aller möglichen Ergebnisse ω ist (prinzipiell) vor dem Experiment bekannt.
- Das tatsächlich realisierte Ergebnis $\omega \in \Omega$ ist nicht mit Sicherheit vorhersagbar.
- Das Experiment kann (zumindest prinzipiell) beliebig oft und unabhängig wiederholt werden.

Beispiel

Betrachten wir ein Standardbeispiel. Es hat den Vorteil, dass es einfach ist und dass wir es alle kennen: den einfachen Würfelwurf, also den einmaligen Wurf eines Spielwürfels mit den Augenzahlen eins bis sechs auf den Seiten. Dabei handelt es sich um ein Zufallsexperiment, denn schon im Vorhinein kennen wir die Menge aller möglichen Ergebnisse $\Omega = \{1, 2, 3, 4, 5, 6\}$. Wenn der Würfel nicht manipuliert ist, können wir aber nicht mit Sicherheit vorhersagen, welches $\omega \in \Omega$ realisiert wird. Und man kann natürlich beliebig oft würfeln. ◀

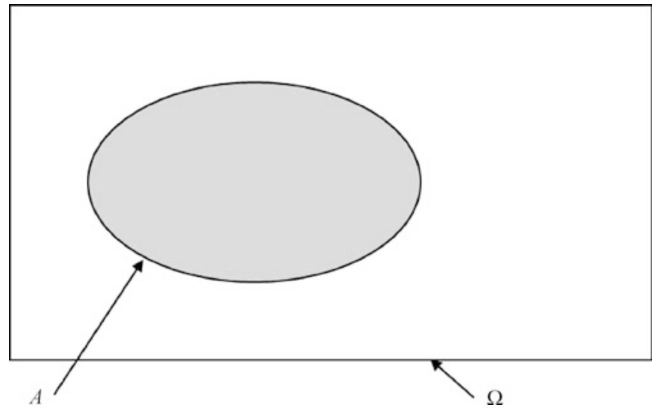


Abb. 17.1 Venn-Diagramm einer Ergebnismenge und eines Ereignisses

Im Zusammenhang mit Zufallsexperimenten muss man einige weitere Begriffe kennen.

Definition: Ergebnis(-menge) und Ereignis

Die Menge Ω aller bei Durchführung des Zufallsexperiments E möglichen Ergebnisse ω heißt **Grundraum** oder **Ergebnismenge**.

Jedes $\omega \in \Omega$ heißt **Ergebnis** oder **Elementarereignis**.

Ein **Ereignis** A ist eine Teilmenge von Ω, d. h. $A \subseteq \Omega$.

Die Gesamtheit $\mathcal{A} = \{A | A \subset \Omega \text{ ist Ereignis}\}$ heißt **Ereignisraum**.

Sprechweise: A tritt ein bedeutet: Das realisierte ω liegt in A: $\omega \in A$.

In der Praxis entspricht jedes Ereignis A einer verbalen Aussage. Das Ereignis „Augenzahl ist gerade" entspricht beispielsweise $A = \{2, 4, 6\}$ beim einfachen Würfelwurf.

Die Ergebnismenge Ω und ein Ereignis A kann man in einem Venn-Diagramm darstellen, wie die Abb. 17.1 zeigt.

Beispiel

Das Ereignis, eine gerade Augenzahl zu würfeln, entspricht dem Ereignis $A = 2, 4, 6$ beim einfachen Würfelwurf. Das zugehörige Venn-Diagramm sieht folgendermaßen aus:

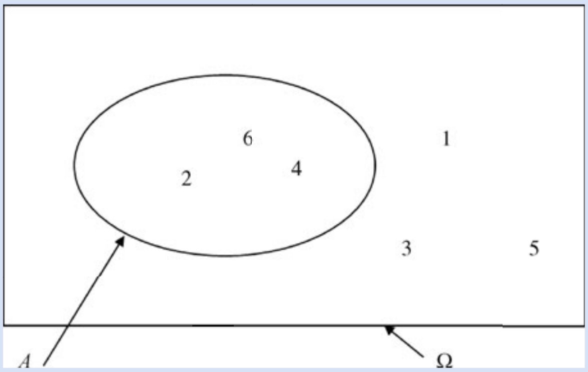


Abb. 17.2 Venn-Diagramm für das Ereignis, eine gerade Zahl zu würfeln ◀

Wichtig sind auch einige spezielle Ereignisse und Notationen.

Definition: Sicheres Ereignis, unmögliches Ereignis, Komplementärereignis

$A = \Omega$ heißt **sicheres Ereignis** ($\omega \in \Omega$ gilt immer!).

$A = \emptyset$ heißt **unmögliches Ereignis** ($\omega \in \emptyset$ gilt nie!).

$\overline{A} := \Omega \setminus A = \{\omega | \omega \in \Omega, \omega \notin A\}$ heißt **komplementäres Ereignis** zu A bzw. Gegenteil von A. Verbal gesehen spricht man auch davon, dass Ereignis A nicht eintritt.

Das zugehörige Venn-Diagramm sieht wie in Abb. 17.3 aus.

Definition: Vereinigung zweier Ereignisse

Seien A und B zwei Ereignisse. Dann heißen $A \cup B := \{\omega | \omega \in A \text{ oder } \omega \in B\}$ (A oder B tritt ein oder beides) die **Vereinigung** der Ereignisse A und B.

Das zur Vereinigungsmenge gehörige Venn-Diagramm sieht wie in Abb. 17.4 aus.

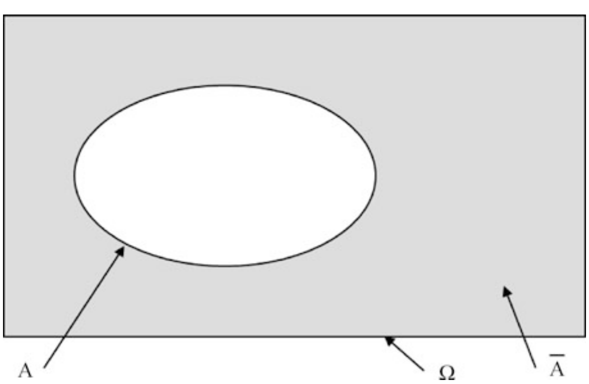


Abb. 17.3 Venn-Diagramm für das Komplementärereignis

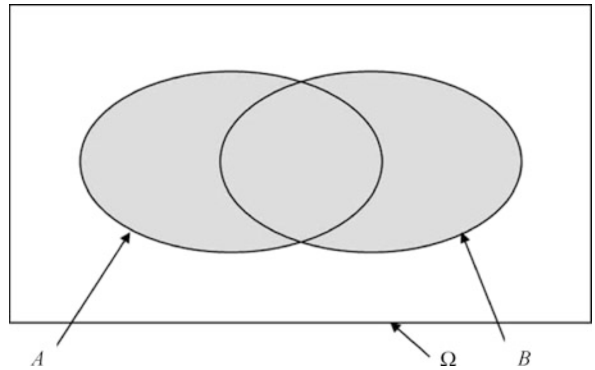


Abb. 17.4 Venn-Diagramm zur Vereinigung von Ereignissen

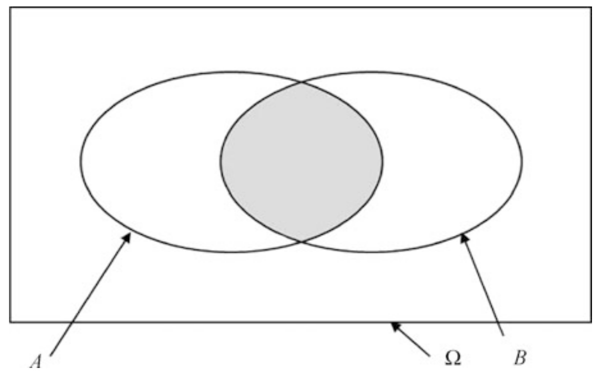


Abb. 17.5 Venn-Diagramm zur Schnittmenge zweier Ereignisse

Definition: Durchschnitt zweier Ereignisse

$A \cap B := \{\omega | \omega \in A \text{ und } \omega \in B\}$ (A und B treten gleichzeitig ein) der **Durchschnitt** der Ereignisse A und B. Man sagt, dass A und B gleichzeitig eintreten.

Das zur Schnittmenge gehörige Venn-Diagramm sieht wie in Abb. 17.5 aus.

Definition: Disjunktheit

Ist $A \cap B = \emptyset$, so heißen die Ereignisse A und B **disjunkt**. Die Ereignisse können nicht gleichzeitig eintreten.

Das zugehörige Venn-Diagramm sieht wie in Abb. 17.6 aus.

Definition: Wichtige Ereignisse

- Die Vereinigung von $k \geq 2$ Ereignissen $A_1, A_2, \ldots, A_k$ wird mit $\bigcup_{i=1}^{k} A_i := A_1 \cup A_2 \cup \ldots \cup A_k$ bezeichnet (mindestens eines der A_i tritt ein).
- Der Durchschnitt von $k \geq 2$ Ereignissen $A_1, A_2, \ldots, A_k$ wird mit $\bigcap_{i=1}^{k} A_i := A_1 \cap A_2 \cap \ldots \cap A_k$ bezeichnet (Alle A_i treten gleichzeitig ein).

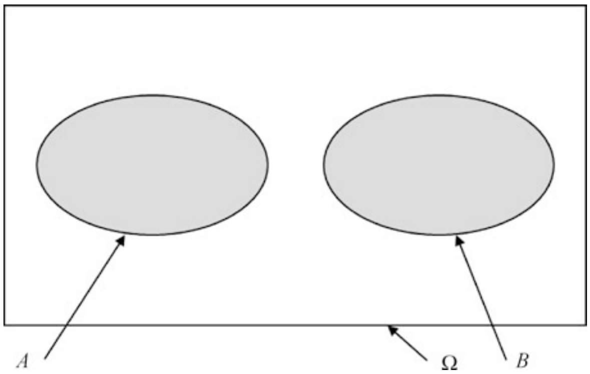


Abb. 17.6 Venn-Diagramm zweier disjunkter Ereignisse

- Die Ereignisse $A_1, A_2, \ldots, A_k$ heißen paarweise disjunkt, wenn für alle Paare $i, j \in 1, \ldots, k$ mit $i \neq j$ gilt: $A_i \cap A_j = \emptyset$.
- Die Notation $A + B$ bzw. $\sum_{i=1}^{k} A_i$ bezeichnet die Vereinigung paarweise disjunkter Ereignisse A, B bzw. $A_1, A_2, \ldots, A_k$.

Beispiel

Betrachten wir noch einmal den einfachen Würfelwurf. Das Ereignis A, eine 1, 2, 3, 4, 5 oder 6 zu würfeln, ist alles, was bei einem einfachen Würfelwurf herauskommen kann, also das sichere Ereignis $A = \Omega$.

Wenn wir uns für das Ereignis B interessieren, gleichzeitig eine gerade und eine ungerade Zahl zu würfeln, dann haben wir es mit dem unmöglichen Ereignis $B = \emptyset$ zu tun, denn wenn man nur einmal würfelt, kann man nicht gleichzeitig eine gerade und eine ungerade Zahl würfeln. Genauer gesagt handelt es sich um zwei komplementäre Ereignisse. Wenn C das Ereignis ist, eine gerade Zahl zu würfeln, dann ist das Ereignis, eine ungerade Zahl zu würfeln, hiervon das Gegenteil, also $\overline{C}$. Daher sind C und $\overline{C}$ immer disjunkt.

Sei D das Ereignis, eine Zahl höchstens 4 zu würfeln. Dann ist das Ereignis, eine gerade Zahl oder höchstens eine 4 zu würfeln, $C \cup D = \{1, 2, 3, 4, 6\}$. Und das Ereignis, gleichzeitig eine gerade Zahl und höchstens eine 4 zu würfeln, ist $C \cap D = \{2, 4\}$. ◀

Zur Berechnung der Wahrscheinlichkeit von vielen Ereignissen muss man über kombinatorische Kenntnisse verfügen. Da **Kombinatorik** bereits Thema in Abschn. 1.7 war, werden an dieser Stelle nur die wichtigsten Berechnungsformeln wiederholt. Dabei geht es immer um die Fragestellung, wie viele Möglichkeiten es gibt, aus einer Urne mit n Dingen k Dinge herauszuziehen. Hierbei muss man unterscheiden, wie man zieht. Einerseits kann man mit oder ohne Zurücklegen der bereits gezogenen Dinge, oder anders ausgedrückt, mit oder ohne Wiederholungen ziehen. Es gibt aber noch eine zweite Dimension der Ziehungsarten, nämlich die Dimension, ob die Reihenfolge der gezogenen Dinge interessiert oder nicht. Das kann man auch interpretieren als: Die einzelnen Dinge sind unterscheidbar oder nicht. Die entsprechende Anzahl der Möglichkeiten, aus n Dingen k Dinge zu ziehen, kann man am besten tabellarisch ausdrücken:

	Mit Reihenfolge/ unterscheidbar	Ohne Reihenfolge/nicht unterscheidbar
Mit Zurücklegen/ mit Wiederholung	n^k	$\binom{n+k-1}{k}$
Ohne Zurücklegen/ ohne Wiederholung	$(n)_k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-s) \cdot \ldots \cdot (n-k+1)$	$\binom{n}{k}$

Beispiel

Als Erstes berechnen wir die Anzahl der Möglichkeiten, aus 49 Kugeln sechs Kugeln zu ziehen, und zwar ohne Zurücklegen und ohne dass uns die Reihenfolge der gezogenen Kugeln interessiert. Wir interessieren uns also für die Anzahl der Möglichkeiten, die beim Samstagslotto gezogen werden können. Das muss gemäß der obigen Tabelle die Situation unten rechts sein, also müssen wir berechnen:

$$\binom{49}{6} = 13\,983\,816$$

(Taschenrechner haben hierfür übrigens eine „nCr"-Taste). Das sind also sehr viele Möglichkeiten.

Stellen wir uns kurz vor, Franziska Reichenbacher wirft jedes Mal, nachdem sie eine Lottokugel gezogen hat, diese wieder in die Urne zurück. Wie viele Möglichkeiten gibt es dann für die Ziehung von sechs Lottokugeln? Gemäß der Situation oben rechts (denn die Reihenfolge, in der sie die Kugeln zieht, interessiert uns immer noch nicht) sind es

$$\binom{49+6-1}{6} = \binom{54}{6} = 25\,827\,165 \text{ Möglichkeiten.}$$

Als nächstes hätte ich gerne gewusst, wie viele sinnhafte und sinnlose vierbuchstabige Wörter wir aus den Buchstaben des Wortes „Blumentopf" (zehn Buchstaben) bilden können (ohne Wiederholung). Da in diesem Wort jeder Buchstabe nur einmal vorkommt und die Reihenfolge der gezogenen Buchstaben wichtig ist, müssen wir nach der Situation unten links vorgehen. Wir berechnen:

$$(10)_4 = \frac{10!}{6!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040.$$

(Dafür haben Taschenrechner eine „nPr"-Taste.)

Und zum Schluss noch: Wie viele sinnlose und sinnhafte vierbuchstabige Wörter können wir aus dem Wort „Blumentopf" bilden, wenn wir jeden Buchstaben so oft verwenden dürfen, wie wir möchten? Dann ziehen wir die Buchstaben mit Wiederholung und unter Beachtung der Reihenfolge, also gibt es $10^4 = 10\,000$ Möglichkeiten. ◀

Um Wahrscheinlichkeiten von komplizierteren Ereignissen berechnen zu können, muss man noch einige Zusammenhänge beherrschen:

Regeln von de Morgan

Sind A und B zwei Ereignisse, so gelten für die Komplemente von Vereinigung und Durchschnitt:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{und} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Entsprechendes gilt für die Komplemente der Vereinigung und des Durchschnitts von k Ereignissen $A_1, A_2, \ldots, A_k$:

$$\overline{\bigcup_{i=1}^{k} A_i} = \bigcap_{i=1}^{k} \overline{A_i}$$

und

$$\overline{\bigcap_{i=1}^{k} A_i} = \bigcup_{i=1}^{k} \overline{A_i}.$$

Beispiel

Das Ereignis, nicht eine gerade Zahl oder höchstens eine 4 zu würfeln, ist also dasselbe wie das Ereignis, keine gerade Zahl und gleichzeitig mindestens eine fünf zu würfeln.

Betrachten wir lieber die Formeln: Es seien $A = \{2.4.6\}$ und $B = \{1, 2, 3, 4\}$. Dann ist $\overline{A \cup B} = \overline{\{1, 2, 3, 4, 6\}} = \{5\}$.

Auf der anderen Seite ist $\overline{A} \cup \overline{B} = \overline{\{2, 4, 6\}} \cap \overline{\{1, 2, 3, 4\}} = \{1, 3, 5\} \cup \{5, 6\} = \{5\}$. Also handelt es sich tatsächlich um dasselbe Ereignis. ◀

Ebenfalls nützlich ist die Kenntnis der Distributivgesetze:

Distributivgesetze

Für Mengen/Ereignisse A, B, C gelten die Rechenregeln

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ und $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Beispiel

Betrachten wir das Ereignis, dass man eine gerade Zahl (A) oder gleichzeitig höchstens eine 4 (B) und mindestens eine 3 (C) würfelt. Dann ist das Ereignis $A \cup (B \cap C) = \{2, 4, 6\} \cup (\{1, 2, 3, 4\} \cap \{3, 4, 5, 6\}) = \{2, 3, 4, 6\}$ gesucht. Andererseits sind $A \cup B = \{1, 2, 3, 4, 6\}$ und $A \cup C = \{2, 3, 4, 5, 6\}$. Deren Schnittmenge ist $(A \cup B) \cap (A \cup C) = \{2, 3, 4, 6\}$. Die Ereignisse sind also gleich, wie es nach dem Distributivgesetz zu erwarten war. ◀

Bislang haben wir uns noch nicht mit Wahrscheinlichkeiten beschäftigt. Alle Zusammenhänge bezogen sich lediglich auf Ereignisse. Nun beziehen wir Wahrscheinlichkeiten mit in unsere Überlegungen ein. Allerdings müssen wir hierfür erst einmal festhalten, was eigentlich eine Wahrscheinlichkeit ist. Ehrlich gesagt ist das gar nicht so einfach, und die Gelehrten streiten sich hierüber schon lange. Wen es interessiert, findet bei Bamberg, G., Baur, F. (2006): „Statistik", 12. Auflage (München) einen Überblick über die Diskussion.

Wir beschränken uns hier auf den axiomatischen Ansatz. Dazu halten wir fest, dass alles Wahrscheinlichkeit ist, was die folgenden Kolmogorov-Axiome erfüllt:

Kolmogorov-Axiome

Ein Wahrscheinlichkeitsmaß auf Ω ist eine Funktion $P : \mathcal{A} \longrightarrow \mathcal{R}$, die jedem Ereignis $A \in \mathcal{A}$ eine Zahl $P(A)$ zuordnet und die die Kolmogorov-Axiome erfüllt:

- $P(A) \geq 0$ für alle Ereignisse $A \in \mathcal{A}$ (Nicht-Negativität). Wir fordern also, dass eine Wahrscheinlichkeit nicht negativ werden darf. Das sollte bekannt sein.
- $P(\Omega) = 1$ (Normierungseigenschaft). Die Wahrscheinlichkeit, dass irgendein Ergebnis eintritt, beträgt 100 %, also 1.
- Für paarweise disjunkte Ereignisse $A_1, A_2, \ldots, A_n \in \mathcal{A}$ gilt (Additivität):

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Diese Bedingung muss auch für $n = \infty$ erfüllt sein. In diesem Fall spricht man von σ-Additivität. Das bedeutet, dass man bei Ereignissen, die nicht gleichzeitig eintreten können, die Einzelwahrscheinlichkeiten addieren darf.

Jetzt haben wir alle Bestandteile des Modells für Zufallsexperimente beisammen:

Definition: Wahrscheinlichkeitsraum

Das Paar (Ω, P) bzw. das Tripel $(\Omega, \mathcal{A}, P)$ heißt **Wahrscheinlichkeitsraum**.

Teil IV

Für unsere Zwecke genügt es, dass wir uns mit der ersten Version eines Wahrscheinlichkeitsraums, also mit (Ω, P) beschäftigen: Ein Wahrscheinlichkeitsraum enthält tatsächlich alle Informationen, die man über ein Zufallsexperiment braucht: Was kann herauskommen? Mit welcher Wahrscheinlichkeit kommen die Ergebnisse heraus?

Um kompliziertere Wahrscheinlichkeiten berechnen zu können, muss man einige Rechenregeln für Wahrscheinlichkeiten kennen, die man auch sehr schön mit Hilfe von Venn-Diagrammen begründen kann. Denn die Wahrscheinlichkeit eines Ereignisses wird im Venn-Diagramm durch den Flächeninhalt des Ereignisses dargestellt:

Eigenschaften von Wahrscheinlichkeiten

Es seien A, B Ereignisse. Dann gilt:

- $P(\emptyset) = 0$. Die Wahrscheinlichkeit des unmöglichen Ereignisses beträgt 0.
- $P(\overline{A}) = 1 - P(A)$. Die Wahrscheinlichkeit, dass A nicht eintritt, beträgt 100 % bzw. 1 minus der Wahrscheinlichkeit, dass A eintritt.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  Um die Wahrscheinlichkeit zu berechnen, dass A oder B eintreten, addiert man zuerst die Wahrscheinlichkeiten von A und von B getrennt. Dann stellt man fest, dass man die Schnittmenge zweimal gezählt hat, also muss man sie einmal wieder abziehen. Das kann man an der Abb. 17.4 nachvollziehen.
- Falls A, B disjunkt sind (aber auch nur dann), gilt sogar: $P(A \cup B) = P(A) + P(B)$. Dies ist in Abb. 17.7 dargestellt.

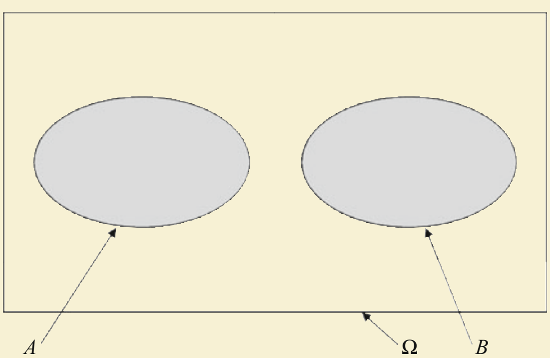


Abb. 17.7 Venn-Diagramm zur Wahrscheinlichkeit zweier disjunkter Ereignisse

  Wenn man das obige Venn-Diagramm betrachtet, sollte das auch unmittelbar einleuchten, denn hier gibt es keine Schnittmenge, die man doppelt gezählt hat, also muss man auch nichts abziehen.
- $P(B \setminus A) = P(B) - P(B \cap A)$
  Auch diese Formel lässt sich mit dem obigen Venn-Diagramm erläutern: Die Wahrscheinlichkeit, dass B eintritt, A aber nicht, entspricht der Wahrscheinlichkeit, dass B eintritt abzüglich der Wahrscheinlichkeit, dass A und B gleichzeitig eintreten.

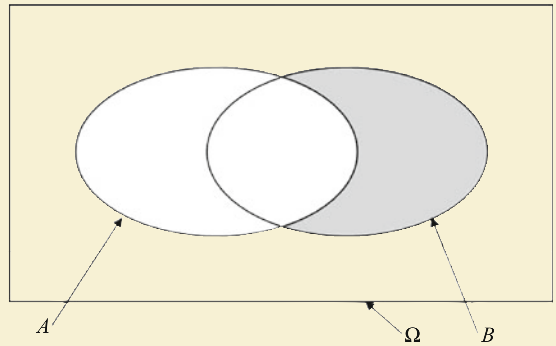


Abb. 17.8 Venn-Diagramm zum Ereignis, dass B eintritt, A aber nicht

17.2 Bedingte Wahrscheinlichkeiten

Die Idee der bedingten Häufigkeiten kennen wir ja bereits. Die gleiche Idee gibt es auch bei Wahrscheinlichkeiten, nur sieht es da etwas anders aus (zumindest auf den ersten Blick).

Definition: Bedingte Wahrscheinlichkeit

Es sei (Ω, P) ein Wahrscheinlichkeitsraum. A, B seien zwei Ereignisse. Dann ist die Wahrscheinlichkeit, dass A eintritt, wenn bereits feststeht, dass B eingetreten ist, definiert durch:

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

Natürlich muss $P(B) > 0$ gelten. Wir lesen den senkrechten Strich „|" also als „unter der Bedingung, dass …", „wenn", „falls" oder als eine ähnliche Formulierung, die den konditionalen Charakter zum Ausdruck bringt.

Bedingte Wahrscheinlichkeiten $P(A|B)$ muss man sorgfältig von gemeinsamen Wahrscheinlichkeiten $P(A \cap B)$ unterscheiden: **Gemeinsame Wahrscheinlichkeiten** geben an, wie wahrscheinlich es ist, dass zwei Ereignisse gleichzeitig eintreten. Bei bedingten Wahrscheinlichkeiten steht ein Ereignis immer bereits fest! Und das halten wir schon mal als erste Merkregel fest:

Merkregel für den Umgang mit bedingten Wahrscheinlichkeiten

Das Ereignis, das bereits eingetreten ist bzw. das bereits feststeht, gehört immer hinter den Strich. Das Ereignis, für dessen Wahrscheinlichkeit man sich unter dieser Bedingung interessiert, gehört vor den Bedingungsstrich.

Menschen haben echte Schwierigkeiten mit bedingten Wahrscheinlichkeiten. Das wird man merken, wenn es darum geht, einen verbalen Zusammenhang als bedingte Wahrscheinlichkeit zu formulieren. Dieses Phänomen, dass Menschen Ursache und Wirkung systematisch vertauschen, wird in der Psychologie unter dem Begriff „conditional-probability-fallacy" untersucht. Es sei hier bereits verraten, dass auch die immer wieder genannte Hypothese, Frauen seien die besseren Autofahrer, auf diesem Fehler beruht.

Beispiel

Immer wieder liest man, dass Frauen die besseren Autofahrer sind. Die Richtigkeit dieser Aussage ist sicherlich unstrittig, lediglich die Argumentation ist oftmals falsch. Man argumentiert nämlich anhand der Unfallstatistiken. Und da gibt es tatsächlich einen riesigen Unterschied: Etwa 75 % aller Unfälle werden von Männern verursacht, lediglich 25 % der Unfälle von Frauen. Ganz klares Indiz, oder?

Schauen wir uns das doch mal näher an. In der Sprache der bedingten Wahrscheinlichkeiten ausgedrückt brauchen wir zwei Ereignisse. Man kann sie A und B nennen, sprechende Bezeichnungen sind hilfreicher. Sei U also das Ereignis, dass ein Unfall passiert und F sei das Ereignis, dass eine Frau am Steuer eines Autos sitzt. Dann besagen die 75 %, dass $P(\overline{F}|U) = 0.75$ ist. Da man sich ja die Unfallstatistiken ansieht, steht ja fest, dass ein Unfall definitiv passiert ist, U muss also hinter den Strich. Analog beträgt also $P(F|U) = 0.25$. Aber ist diese Wahrscheinlichkeit wirklich aussagekräftig? Aussagekräftig wäre, wenn die Wahrscheinlichkeit, dass Frauen einen Unfall verursachen, deutlich geringer wäre als die Wahrscheinlichkeit, dass Männer einen Unfall verursachen. Das wäre überzeugend, um zu behaupten, dass Frauen zumindest weniger Unfälle verursachen. Wir müssen also stattdessen $P(U|F)$ und $P(U|\overline{F})$ vergleichen. Das machen wir auch gleich, aber dazu brauchen wir noch ein paar Rechenregeln. Hier genügt es, wenn klar ist, dass wir über zwei vollkommen unterschiedliche bedingte Wahrscheinlichkeiten sprechen. ◀

Eigentlich muss man zum Rechnen mit bedingten Wahrscheinlichkeiten nur drei Dinge kennen: einige Rechenregeln, den Satz von der totalen Wahrscheinlichkeit und die Formel von Bayes.

Beginnen wir mit den Rechenregeln:

Rechenregeln für bedingte Wahrscheinlichkeiten

Folgende Rechenregeln gelten für bedingte Wahrscheinlichkeiten:

1. $P(A \cap B) = P(B) \cdot P(A|B)$
2. $P(\overline{A}|B) = 1 - P(A|B)$ Die bekannte Komplementregel gilt also weiterhin.
3. **Aber!** Im Allgemeinen gilt auch: $P(A|\overline{B}) \neq 1 - P(A|B)$. Das bedeutet, dass man sehr genau schauen muss, wo das Komplementzeichen steht! Wenn wir hinter dem Bedingungsstrich mit dem Komplement arbeiten wollen, müssen wir zuerst die Formel von Bayes (die lernen wir gleich kennen) anwenden, bevor wir mit „1−" weiterrechnen dürfen.

Beispiel

Diese Rechenregeln haben wir gerade schon angewendet. Es galt: $P(\overline{F}|U) = 0.75$ und $P(F|U) = 1 - 0.75 = 0.25$. Und nehmen wir einmal an, dass die Unfallwahrscheinlichkeit generell in Deutschland ungefähr 5 % beträgt, dass also $P(U) = 0.05$, dann können wir die Wahrscheinlichkeit, dass gleichzeitig eine Frau fährt und ein Unfall passiert, berechnen zu $P(F \cap U) = P(F|U) \cdot P(U) = 0.25 \cdot 0.05 = 0.0125$. ◀

Den **Satz von der totalen Wahrscheinlichkeit** braucht man immer dann, wenn man die bedingten Wahrscheinlichkeiten kennt, aber die unbedingte (= totale) Wahrscheinlichkeit benötigt:

Satz von der totalen Wahrscheinlichkeit

Es seien $B_1, B_2, \ldots, B_n$ eine Partition von Ω, d. h., die B_i sind disjunkt und ihre Vereinigung ergibt ganz Ω. (Synonyme für „Partition" sind auch „Zerlegung" oder „Parkettierung". Gerade die letzte Bezeichnung beschreibt genau, was das Besondere ist. Denn diese Ereignisse sehen aus wie ein Parkettboden.)

Dann gilt für ein Ereignis $A \subset \Omega$:

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

Diesen Satz kann man sich einfach erklären, wenn man im Hinterkopf behält, dass $P(A|B_i) \cdot P(B_i) = P(A \cap B_i)$. Das Venn-Diagramm in Abb. 17.9 sollte dabei helfen.

Gesucht ist die Wahrscheinlichkeit, also der Flächeninhalt von A. Dann sagt der Satz von der totalen Wahrscheinlichkeit, dass man hierzu alle kleinen Flächen $P(A \cap B_i)$ addieren muss. Wir drücken das nicht über die gemeinsamen Wahrscheinlichkeiten, sondern über die bedingten Wahrscheinlichkeiten aus. Gemäß der obigen Rechenregel gilt aber $P(A \cap B_i) = P(A|B_i) \cdot P(B_i)$. Damit ist der Satz von der totalen Wahrscheinlichkeit bewiesen.

Diesen Satz gibt es noch in einer Spezialfallformulierung, denn auch B und $\overline{B}$ bilden eine Partition.

Teil IV

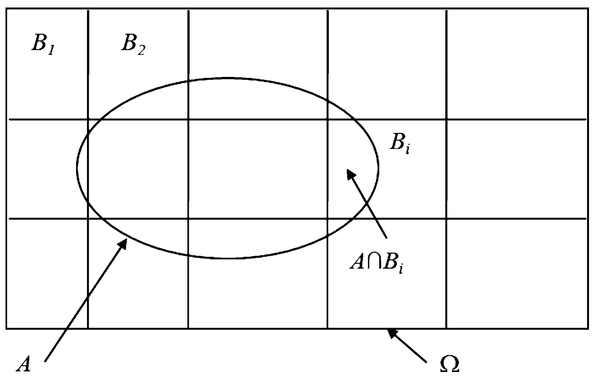


Abb. 17.9 Venn-Diagramm zum Satz von der totalen Wahrscheinlichkeit

Satz von der totalen Wahrscheinlichkeit (Spezialfall)

Es seien A und B zwei Ereignisse. Dann gilt für das Ereignis $A \subset \Omega$:

$$P(A) = P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B})$$

Beispiel

Wir betrachten ein anderes Beispiel: eine Mathematikklausur, zu der man Vorleistungen in Form von Übungsaufgaben erbringen kann. Nehmen wir an, dass von denjenigen, die die Übungsaufgaben immer (selbst) bearbeitet haben, 80 % die Klausur bestanden haben. Von denjenigen Studenten hingegen, die die Übungsaufgaben nicht (selbst) bearbeitet haben, haben lediglich 50 % die Mathematikklausur bestanden. Insgesamt haben 70 % der Studierenden die Übungsaufgaben selbst bearbeitet. Hieraus können wir mithilfe des Satzes von der totalen Wahrscheinlichkeit die generelle Bestehenswahrscheinlichkeit bei der genannten Mathematikklausur berechnen. Definieren wir auch hierzu die beiden relevanten Ereignisse: B soll das Ereignis sein, die Mathematikklausur zu bestehen, $\ddot{U}$ soll das Ereignis sein, die Übungsaufgaben (selbst) zu bearbeiten. Dann wissen wir, dass $P(B|\ddot{U}) = 0.8$ und $P(B|\overline{\ddot{U}}) = 0.5$. Gesucht ist aber $P(B)$. Das berechnet sich gemäß der obigen Formel als

$$\begin{aligned} P(B) &= P(B|\ddot{U}) \cdot P(\ddot{U}) + P(B|\overline{\ddot{U}}) \cdot P(\overline{\ddot{U}}) \\ &= 0.8 \cdot 0.7 + 0.5 \cdot 0.3 = 0.71. \end{aligned}$$

Die Bestehenswahrscheinlichkeit beträgt bei dieser Klausur also 71 %. Da lohnt sich das Bearbeiten der Aufgaben doch. Man erhöht die Bestehenswahrscheinlichkeit durch das Bearbeiten der Aufgaben von 71 % auf immerhin 80 %. ◀

Es gibt eine Möglichkeit, den Satz von der totalen Wahrscheinlichkeit grafisch darzustellen, nämlich mit Wahrscheinlichkeitsbäumen. Diese Bäume helfen aber nur beim Satz von der totalen Wahrscheinlichkeit und nicht bei der Formel von Bayes!

Vorgehensweise beim Ablesen bedingter Wahrscheinlichkeiten aus Wahrscheinlichkeitsbäumen

Schritt 1: Man zeichnet einen Knoten (die Wurzel des Baumes) mit abgehenden Ästen, für jedes Element der gegebenen Partition von Ω, also für $B_1, B_2, \ldots, B_n$, je einen Ast. Jeder Ast endet in einem Knoten, der mit $B_1, B_2, \ldots, B_n$ beschriftet wird. Man beschriftet die Äste mit den gegebenen Wahrscheinlichkeiten der Partitionselemente $P(B_1)), P(B_2), \ldots, P(B_n)$.

Schritt 2: Man hängt an jeden Ast zwei Äste und beschriftet einen mit A und einen mit dem Gegenteil $\overline{A}$. Sie werden mit den zugehörigen bedingten Wahrscheinlichkeiten $P(A|B_i)$ und $P(\overline{A}|B_i)$ beschriftet.

Schritt 3: Man liest die unbedingte Wahrscheinlichkeit $P(A)$ ab, indem man alle Wege betrachtet, die von der Wurzel zum Ereignis A führen. Die Wahrscheinlichkeiten, die auf einem Weg liegen, werden miteinander multipliziert. Diese Ergebnisse werden für alle unterschiedlichen Wege addiert.

Beispiel

Im obigen Beispiel, in dem wir die Bestehenswahrscheinlichkeit für eine Mathematik-Klausur berechnet haben, sieht der Wahrscheinlichkeitsbaum wie folgt aus:

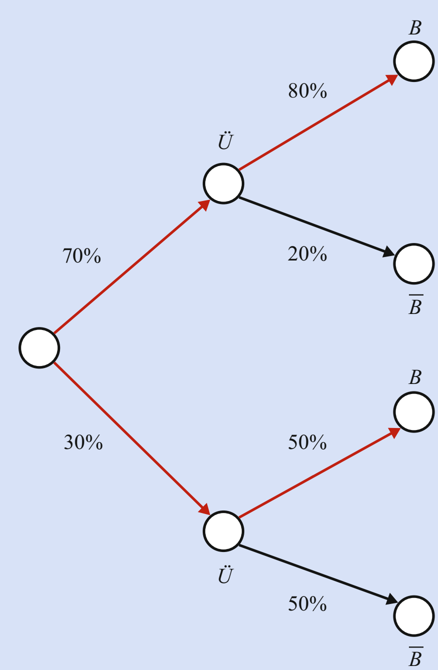


Abb. 17.10 Beispiel eines Wahrscheinlichkeitsbaums zur Illustration des Satzes von der totalen Wahrscheinlichkeit

Wenn man nun die Wahrscheinlichkeiten an einem Ast, der zum Ergebnis B führt, erst multipliziert und diese Wahrscheinlichkeiten anschließend addiert, erhält man

$$P(B) = 0.7 \cdot 0.80 + 0.3 \cdot 0.5 = 0.56 + 0.15 = 0.71.$$

Auch auf diesem Weg erhalten wir eine generelle Bestehenswahrscheinlichkeit von 71 %. ◄

Als Letztes lernen wir die **Formel von Bayes** kennen. Diese Formel braucht man immer dann, wenn man die eine bedingte Wahrscheinlichkeit $P(B|A)$ kennt, aber genau die umgedrehte Wahrscheinlichkeit $P(A|B)$ benötigt:

Formel von Bayes

Seien A, B zwei Ereignisse. Dann gilt

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

Auch das kann man mit wenig Aufwand nachrechnen: Auf der linken Seite der Gleichung steht gemäß Definition:

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

Auf der rechten Seite der Gleichung steht

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

Beide Seiten der Gleichung liefern also dasselbe. Die Gleichung stimmt.

Beispiel

Das ist die Formel, die uns gefehlt hat, um endlich die Diskussion zu beenden, ob Frauen die besseren Autofahrer sind. Mithilfe der Formel von Bayes können wir die Wahrscheinlichkeit berechnen, dass ein Unfall passiert, wenn eine Frau bzw. ein Mann am Steuer sitzt. Also, es gilt:

$$P(U|F) = \frac{P(F|U) \cdot P(U)}{P(F)}.$$

Um das auszurechnen, müssen wir noch schätzen, wie groß die Wahrscheinlichkeit ist, dass eine Frau generell am Steuer sitzt. Auch diese Zahl ist nur eine Vermutung, wir gehen von 30 % aus. Also gilt: $P(F) = 0.3$. Damit gilt:

$$P(U|F) = \frac{P(F|U) \cdot P(U)}{P(F)} = \frac{0.25 \cdot 0.05}{0.3} \approx 0.0417.$$

Andererseits gilt:

$$P(U|\overline{F}) = \frac{P(\overline{F}|U) \cdot P(U)}{P(\overline{F})} = \frac{0.75 \cdot 0.05}{0.7} \approx 0.0534$$

Das ist nur geringfügig höher. Männer und Frauen scheinen also mit ähnlich hoher bzw. niedriger Wahrscheinlichkeit Unfälle zu verursachen. ◄

Diese Sätze, Formeln und Rechenregeln kann man sehr gut miteinander kombinieren.

Damit wir es ein bisschen einfacher haben, gibt es im Folgenden ein Kochrezept, an dem man sich orientieren kann:

Vorgehensweise beim Berechnen bedingter Wahrscheinlichkeiten

Schritt 1: Definieren der beiden Ereignisse/Ereignisarten, um die es in der Aufgabenstellung geht, und Bezeichnen mit Buchstaben unserer Wahl. Hierbei darauf achten, keine Buchstaben zu „verschwenden“, also nicht zwei Komplementärereignisse als A und B zu bezeichnen, sondern als A und $\overline{A}$.

Schritt 2: Aus dem Aufgabentext suchen wir alle gegebenen Wahrscheinlichkeiten heraus und benennen sie mit unseren Bezeichnungen. Generell fängt es immer mit „P(irgendwas) =“ an.

Schritt 3: Wir formulieren mit unseren Bezeichnungen, welche Wahrscheinlichkeit wir gemäß Aufgabenstellung ermitteln sollen.

Schritt 4: Wir suchen diejenige Formel heraus, die auf unser Problem am besten passt.

Teil IV

17.3 Stochastische Unabhängigkeit von Ereignissen

Die Idee der Unabhängigkeit, die wir im empirischen Fall schon kennen gelernt haben, gibt es bei Wahrscheinlichkeiten:

Definition: Stochastische Unabhängigkeit von Ereignissen

Seien (Ω, P) ein Wahrscheinlichkeitsraum; A, B zwei Ereignisse. Dann heißen A und B **stochastisch unabhängig**, wenn gilt: $P(A|B) = P(A)$ bzw. $P(B|A) = P(B)$

Diese Definition ist auch sehr verständlich: Wenn das Eintreten des einen Ereignisses nichts damit zu tun hat, ob das andere Ereignis eintritt oder nicht, dann darf es keinen Einfluss auf die

Eintrittswahrscheinlichkeit haben, ob das besagte Ereignis nun eingetreten ist oder nicht.

Wenn man die obige Definition gemäß der Definition der bedingten Wahrscheinlichkeiten umformt, dann erhält man die folgende Bedingung, die erfüllt sein muss, damit zwei Ereignisse stochastisch unabhängig sind.

Produktregel

Seien (Ω, P) ein Wahrscheinlichkeitsraum; A, B zwei Ereignisse. Dann sind A und B stochastisch unabhängig, wenn gilt: $P(A \cap B) = P(A) \cdot P(B)$ (Produktregel)

Mit diesem Zusammenhang kann man meistens einfacher überprüfen, ob zwei Ereignisse unabhängig voneinander sind oder nicht.

Wenn man mehr als zwei Ereignisse auf stochastische Unabhängigkeit untersucht, gibt es zwei Arten der stochastischen Unabhängigkeit:

Definition: Paarweise und gemeinsame stochastische Unabhängigkeit

- Die n Ereignisse $A_1, A_2, \ldots, A_n$ heißen **paarweise stochastisch unabhängig**, wenn die Produktregel für je zwei Ereignisse gilt, d. h. $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für alle $i \neq j$.
- Die n Ereignisse $A_1, A_2, \ldots, A_n$ heißen **gemeinsam stochastisch unabhängig**, wenn die Produktregel für alle Teilmengen der Ereignisse gilt, d. h. $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für alle $i \neq j$, $P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cdot P(A_k)$ für alle $i \neq j \neq k$ etc.

Da ja alle Bedingungen, die erfüllt sein müssen, damit die Ereignisse paarweise stochastisch unabhängig sind, auch erfüllt sein müssen (und noch mehr), damit die Ereignisse gemeinsam stochastisch unabhängig sind, gilt der folgende Zusammenhang:

Zusammenhang zwischen paarweiser und gemeinsamer stochastischer Unabhängigkeit

Wenn die Ereignisse $A_1, A_2, \ldots, A_k$ gemeinsam stochastisch unabhängig sind, dann sind sie auch paarweise stochastisch unabhängig. Aus gemeinsamer Unabhängigkeit folgt also paarweise Unabhängigkeit. Die Umkehrung gilt aber nicht. Wenn die Ereignisse $A_1, A_2, \ldots, A_k$ paarweise stochastisch unabhängig sind, müssen sie noch lange nicht gemeinsam stochastisch unabhängig sein.

An dieser Stelle wird es Zeit, dass wir die wichtigsten Zusammenhänge der Wahrscheinlichkeitsrechnung mal übersichtlich hinschreiben, die braucht man nämlich häufig in der Zuverlässigkeitsrechnung. Zusammenfassend gilt:

- Wenn zwei Ereignisse A und B nicht gleichzeitig eintreten können, also disjunkt sind, dann gilt:

$$P(A \cup B) = P(A) + P(B)$$

- Wenn zwei Ereignisse A und B (stochastisch) unabhängig sind, sich also wechselseitig nicht beeinflussen, dann gilt:

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

- Wenn zwei Ereignisse A und B (stochastisch) unabhängig sind, sich also wechselseitig nicht beeinflussen, dann sind sie niemals disjunkt, und es gilt:

$$P(A \cup B) = 1 - (1 - P(A)) \cdot (1 - P(B))$$

17.4 Zufallsvariablen

Mit Ereignissen kennen wir uns jetzt gut aus. Hinzu kommt nun der Begriff der **Zufallsvariablen**.

Intuitiv versteht man unter einer Zufallsvariablen X einen Sachverhalt, den man stochastisch untersucht. Allerdings dürfen für diesen Sachverhalt nur Zahlen als Ergebnis herauskommen. X ist sozusagen eine Abkürzung für diesen zu untersuchenden Sachverhalt.

Formal bedeutet das:

Definition: Zufallsvariable

Es sei Ω die Ergebnismenge eines Zufallsexperiments. Dann heißt die Abbildung $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable.

Eine Zufallsvariable besitzt eine Wahrscheinlichkeitsverteilung, die jedem möglichen Wert x, den die Zufallsvariable annehmen kann, seine Wahrscheinlichkeit zuordnet.

Natürlich interessiert man sich auch für die Wahrscheinlichkeit, dass eine Zufallsvariable höchstens den Wert x annimmt. Diese Problematik haben wir bereits in Abschn. 15.4 diskutiert. Analog zum empirischen Fall gibt es auch in diesem Fall eine Verteilungsfunktion, die wie folgt definiert ist:

Definition: Verteilungsfunktion

Die **Verteilungsfunktion** $F^X(x)$ der Zufallsvariablen X ist definiert durch:

$$F^X(x) = P(X \leq x)$$

Anwendung: Wahrscheinlichkeitsverteilung beim zweifachen Würfelwurf

Wir betrachten den zweifachen Würfelwurf. Hierbei modellieren wir, dass $\Omega = \{(1.1), (1.2), (1.3), \dots, (6.6)\}$ ist, wobei jedes ω dieselbe Wahrscheinlichkeit haben soll. Es gilt also: $P(\omega) = \frac{1}{|\Omega|} = \frac{1}{36}$. Wir können nun eine Abbildung definieren, die jedem $\omega \in \Omega$ die Augensumme des Ergebnisses zuordnet, also $X((1,1)) = 2, /X((1,2)) = 3$ etc. Da jedes Ergebnis eine Augensumme besitzt, können wir die Wahrscheinlichkeit für eine beliebige Augensumme dadurch berechnen, dass wir die Wahrscheinlichkeiten der zugehörigen Ergebnisse heranziehen. Beispielsweise erhält man die Augensumme 5, wenn man $(1,4), (2,3), (3,2)$ oder $(4,1)$ würfelt. Jedes Ergebnis besitzt die Wahrscheinlichkeit $\frac{1}{36}$, also beträgt die Wahrscheinlichkeit, die Augensumme 5 zu würfeln, $\frac{4}{36}$. Die folgende Tabelle listet alle möglichen Augensummen und ihre Wahrscheinlichkeiten auf.

x	2	3	4	5	6	7	8	9	10	11	12
ω	$(1,1)$	$(1,2)$, $(2,1)$	$(1,3)$, $(2,2)$, $(3,1)$	$(1,4)$, $(2,3)$, $(3,2)$, $(4,1)$	$(1,5)$, $(2,4)$, $(3,3)$, $(4,2)$, $(5,1)$	$(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$, $(6,1)$	$(2,6)$, $(3,5)$, $(4,4)$, $(5,2)$, $(6,2)$	$(3,6)$, $(4,5)$, $(5,4)$, $(6,3)$	$(4,6)$, $(5,5)$, $(6,4)$	$(5,6)$, $(6,5)$	$(6,6)$
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Als kleine Kontrollmöglichkeit gilt: Die Summe über alle Wahrscheinlichkeiten muss natürlich gerade 1 betragen. Das tut sie auch.

Die Verteilungsfunktion hat ähnliche Eigenschaften wie die empirische Verteilungsfunktion:

Eigenschaften der Verteilungsfunktion

- F^X ist monoton wachsend.
- F^X geht für kleine x-Werte gegen 0, für große x-Werte gegen 1 bzw. $\lim_{x\to-\infty} F^X(x) = 0$ und $\lim_{x\to\infty} F^X(x) = 1$.
- F^X ist rechtsseitig stetig, d. h., wenn man von rechts nach links auf der Funktion langwandert, gehört der „linkeste“ Punkt immer mit zur Gerade.

17.5 Verteilungen von Zufallsvariablen und ihre Kenngrößen

Genau wie bei quantitativen Merkmalen unterscheidet man bei Zufallsvariablen stetige und diskrete Zufallsvariablen. Auch hier gilt, dass bei einer diskreten Zufallsvariablen bzw. Zufallsvariablen mit diskreter Verteilung nur endlich viele oder höchstens abzählbar unendlich viele Werte herauskommen können, während bei stetigen Zufallsvariablen bzw. Zufallsvariablen mit stetiger Verteilung zumindest in einem Intervall alle möglichen Werte herauskommen können. Je nachdem, ob man es mit einer stetigen oder mit einer diskreten Verteilung zu tun hat, muss man die Zufallsvariable unterschiedlich behandeln. In diesem Abschnitt behandeln wir, wie man Kenngrößen, den Erwartungswert, der praktisch dem Mittelwert bei empirischen Verteilungen entspricht, und die Varianz sowie die Standardabweichung berechnet. Außerdem schauen wir uns die wichtigsten Verteilungen etwas genauer an.

Kenngrößen von diskreten Verteilungen charakterisieren diskrete Verteilungen

Definition: Wahrscheinlichkeitsfunktion

Bei diskreten Zufallsvariablen X kann man für jeden Wert x, der herauskommen kann, die Wahrscheinlichkeit berechnen, dass er herauskommt, also $P(X = x)$. Das macht man zumindest bei endlich vielen Werten gerne tabellarisch. Die Übersicht über alle Wahrscheinlichkeiten heißt dann **Wahrscheinlichkeitsfunktion** oder **Zähldichte** von X.

Genau wie im empirischen Fall kann man für Zufallsvariablen Kenngrößen berechnen. Die gebräuchlichsten sind der Erwartungswert und die Varianz bzw. die Standardabweichung. Im Fall einer diskreten Verteilung werden die Größen wie folgt berechnet:

Definition: Erwartungswert und k-tes Moment sowie Varianz und Standardabweichung

Der **Erwartungswert** μ einer diskreten Zufallsvariablen X mit den möglichen Ausprägungen x_i berechnet sich durch

$$\mu = \mathrm{E}[X] = \sum_{i \geq 1} x_i \cdot P(X = x_i).$$

Das ***k*-te Moment** einer diskreten Zufallsvariablen X mit den möglichen Ausprägungen x_i berechnet sich durch

$$\mathrm{E}[X^k] = \sum_{i \geq 1} x_i^k \cdot P(X = x_i).$$

Eigentlich interessiert uns immer nur das zweite Moment, also schreiben wir sicherheitshalber die Formel für das zweite Moment extra auf:

$$\mathrm{E}[X^2] = \sum_{i \geq 1} x_i^2 \cdot P(X = x_i).$$

Die **Varianz** σ^2 einer diskreten Zufallsvariablen X mit den möglichen Ausprägungen x_i berechnet sich durch

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

Die **Standardabweichung** σ einer diskreten Zufallsvariablen X berechnet sich durch

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Die Formel für die Varianz benutzt praktisch niemand. Stattdessen verwendet man die folgende viel einfachere Formel:

Die Varianz wird berechnet durch

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

Wir schauen uns das an einem Beispiel an: dem einfachen Würfelwurf.

Beispiel

Die Wahrscheinlichkeitsfunktion des einfachen Wurfes eines unverfälschten Würfels lautet:

Ausprägung x_i	1	2	3	4	5	6
Wahrscheinlichkeit P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Somit beträgt der Erwartungswert $\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$.

Das zweite Moment beträgt dann $E[X^2] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$.

Daraus berechnet sich die Varianz zu $\sigma^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$, somit beträgt die Standardabweichung $\sigma = 1.708$.

Man kann das folgendermaßen interpretieren: Im Durchschnitt würfelt man also die Augenzahl 3.5. Das erscheint auf den ersten Blick irritierend, vielleicht ist die Formulierung, dass man man im Durchschnitt bei zehn Würfen die Augensumme 35 erhält leichter vorstellbar. Die Standardabweichung von 1.708 kann interpretiert werden als die durchschnittliche Abweichung von der durchschnittlichen Augenzahl. Wichtig ist hierbei, sich klarzumachen, dass es nicht um wirkliche Würfelwürfe geht, sondern um ein Modell eines idealen Würfels. Um wirkliche Würfelergebnisse zu analysieren, verwendet man die Methoden der deskriptiven Statistik. ◀

Einige der diskreten Verteilungen sind in der Praxis so wichtig, dass man ihnen einen Namen gegeben hat. Diese sind (u.a.) Gleichverteilung, Poisson-Verteilung, geometrische Verteilung, hypergeometrische Verteilung und Binomialverteilung. Diese fünf schauen wir uns näher an. Zu diesen Verteilungen sollte man wissen, in welchen Situationen man sie üblicherweise einsetzt, was der Wertebereich ist, wie die Wahrscheinlichkeitsfunktion lautet, wie groß Erwartungswert und Varianz sind und ggf. welche Hilfsmittel man benutzen kann. Diese Informationen kommen jetzt im Folgenden:

Die diskrete Gleichverteilung beschreibt gleich wahrscheinliche Ereignisse

Die **diskrete Gleichverteilung** wird auch **Laplace-Verteilung** genannt. Sie wird immer dann als Modell angenommen, wenn es keine Begründung für unterschiedliche Wahrscheinlichkeiten der einzelnen Ergebnisse gibt.

Eine Zufallsvariable, die gleichverteilt ist, kann als Werte z. B. die Zahlen $1, \ldots, s$ annehmen. Somit besitzt die Ergebnismenge s Elemente bzw. $|\Omega| = s$. Jeder Wert wird dementsprechend mit der Wahrscheinlichkeit $P(X = k) = \frac{1}{s}$ angenommen.

Man kann leicht berechnen, dass dann gilt:

$\mu = \mathrm{E}[X] = \frac{s+1}{2}$ und $\sigma^2 = \mathrm{Var}(X) = \frac{s^2-1}{12}$.

Denn es gilt:

$$\mu = \mathrm{E}[X] = \sum_{i \geq 1} x_i \cdot P(X = x_i) = \sum_{i=1}^{s} i \cdot \frac{1}{s}$$
$$= \frac{1}{s} \cdot \sum_{i=1}^{s} i = \frac{1}{s} \cdot \frac{s \cdot (s+1)}{2} = \frac{s+1}{2}.$$

Teil IV

Außerdem gilt:

$$\begin{aligned}\mathrm{E}[X^2] &= \sum_{i\geq 1} x_i^2 \cdot P(X=x_i) = \sum_{i=1}^{s} i^2 \cdot \frac{1}{s} = \frac{1}{s} \cdot \sum_{i=1}^{s} i^2 \\ &= \frac{1}{s} \cdot \frac{s\cdot(s+1)\cdot(2s+1)}{6} = \frac{(s+1)\cdot(2s+1)}{6}\end{aligned}$$

Damit gilt:

$$\begin{aligned}\sigma^2 = \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \\ &= \frac{(s+1)\cdot(2s+1)}{6} - \left(\frac{s+1}{2}\right)^2 \\ &= \frac{2(s+1)\cdot(2s+1)}{12} - \frac{3(s+1)^2}{12} \\ &= \frac{[s+1)\,[2(2s+1)-3(s+1)]}{12} \\ &= \frac{(s+1)(s-1)}{12} = \frac{s^2-1}{12}.\end{aligned}$$

Beispiel

Der einfache Würfelwurf ist ein Beispiel für eine diskrete Gleichverteilung: Die Ergebnismenge ist $\Omega = \{1,2,3,4,5,6\}$. Damit beträgt $|\Omega| = 6$. Die Wahrscheinlichkeit für jedes Ergebnis beträgt also $P(X=k) = \frac{1}{6}$. Der Erwartungswert $E[X] = \frac{6+1}{2} = 3.5$, die Varianz $\mathrm{Var}(X) = \frac{36-1}{12} = \frac{35}{12}$. ◀

Die Poisson-Verteilung beschreibt seltene Ereignisse

Die **Poisson-Verteilung** wird immer dann als Modell angenommen, wenn X die Anzahl der eingetretenen Fälle pro Zeiteinheit, Flächeneinheit etc. darstellt. Sie passt besonders gut, wenn man die Wahrscheinlichkeiten für seltene Ereignisse berechnen will. Sie wurde „erfunden“, als man in der preußischen Armee die Todesfälle durch Pferdetritte untersuchte.

Eine Zufallsvariable, die Poisson-verteilt ist, kann als Werte die Zahlen $0, 1, \ldots$ annehmen. Ein Wert k wird dann mit der Wahrscheinlichkeit

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

angenommen, wobei $\lambda > 0$ sein muss.

Man kann (nicht mehr ganz so) leicht berechnen, dass dann gilt:

$$\mu = \mathrm{E}[X] = \lambda \text{ und } \sigma^2 = \mathrm{Var}(X) = \lambda.$$

Man schreibt abkürzend: $X \sim \mathrm{po}(\lambda)$. Durch die Angabe von λ ist die Verteilung eindeutig bestimmt.

Die geometrische Verteilung modelliert Wartezeiten

Die **geometrische Verteilung** wird immer dann als Modell angenommen, wenn man die Wahrscheinlichkeiten für die Wartezeit bis zum ersten Treffer berechnen möchte. Zum Beispiel versteht man hierunter die Wahrscheinlichkeit, dass man eine symmetrische Münze genau dreimal werfen muss, bis zum ersten Mal „Zahl“ geworfen wird. Wichtig ist, dass die Trefferwahrscheinlichkeit in jeder einzelnen Ziehung identisch sein muss, nämlich p, und dass bei jeder Durchführung des Zufallsexperiments nur zwei Ergebnisse möglich sind, nämlich Treffer oder Nichttreffer.

Hierbei gibt es zwei unterschiedliche Formulierungen, je nachdem, ob die Zufallsvariable X die Anzahl der Ziehungen inklusive oder exklusive des Treffers bezeichnet. Wir sehen uns hier den Fall an, dass X die Anzahl der Ziehungen *bis zum* ersten Treffer (*inkl.* Treffer) beschreibt.

Eine Zufallsvariable, die geometrisch verteilt ist, kann als Werte die Zahlen $0, 1, \ldots$ annehmen. Ein Wert k wird dann mit der Wahrscheinlichkeit $P(X=k) = (1-p)^{k-1} \cdot p$ angenommen.

Man kann (nicht mehr ganz so) leicht berechnen, dass dann gilt:

$$\mu = E[X] = \frac{1}{p} \text{ und } \sigma^2 = \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Man schreibt abkürzend: $X \sim \mathrm{geo}(p)$. Wenn man p kennt, ist die Verteilung eindeutig bestimmt.

Beispiel

Wir stellen uns einen angeheiterten Studenten vor, der einen sehr gelungenen Abend auf einer Fachschaftsparty verbracht hat, aber leider etwas zu viel Alkohol getrunken hat. Nun steht er vor seiner Haustür und starrt mit verschwommenem Blick auf seinen Schlüsselbund mit acht Schlüsseln. Da er leider keine Ahnung mehr hat, welcher der Schlüssel sein Haustürschlüssel ist, probiert er einen aus. Er passt nicht. Und noch schlimmer, der Schlüsselbund fällt ihm aus der Hand. Da die Schlüssel (zumindest für den Studenten in seinem aktuellen Zustand) alle gleich aussehen, muss er dasselbe Spiel noch einmal spielen. Er ist ziemlich betrunken, deswegen lässt er jedes Mal wieder den Schlüssel fallen.

Wenn wir nun wissen wollen, wie groß die Wahrscheinlichkeit ist, dass der Student z.B. zehn Versuche braucht, bis er bei diesem Verfahren den richtigen Schlüssel gefunden hat, dann müssen wir die geometrische Verteilung bemühen. Hierzu sei X die Anzahl der Versuche, die der Student bis zum ersten Treffer benötigt, wobei der Trefferversuch selbst mitgezählt wird. Da es bei jedem Versuch nur zwei mögliche Ergebnisse gibt, nämlich „Schlüssel passt“ oder „Schlüssel passt nicht“, und da wir uns für die Wartezeit bis zum ersten Treffer interessieren,

Anwendung: Poisson-Verteiung

Ein Nutzfahrzeughändler hat in den letzten 24 Monaten notiert, wie viele Fahrzeuge er pro Monat verkauft hat. In der folgenden Tabelle finden wir die entsprechenden Zahlen:

Anzahl k der in einem Monat verkauften Autos	0	1	2	3	4	5	6	≥ 7
Anzahl der Monate, in denen k Autos verkauft wurden	2	5	6	5	3	2	1	0

Der Nutzfahrzeughändler möchte nun ein geeignetes Prognosemodell entwickeln, das es ihm erlaubt, die Wahrscheinlichkeiten für bestimmte Abverkaufszahlen zu prognostizieren. Da es sich beim Verkauf von Baggern offensichtlich um seltene Ereignisse handelt und X die Anzahl der verkauften Bagger pro Monat darstellt, erscheint die Poisson-Verteilung als geeignetes Modell. Der Händler geht also davon aus, dass $P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ gilt.

Allerdings fehlt ihm noch der Parameter λ. Gemäß den obigen Ausführungen wissen wir, dass bei der Poisson-Verteilung der Erwartungswert $E[X] = \lambda$ ist. Wie wir in Abschn. 18.1 sehen werden, kann man den Erwartungswert einer Zufallsvariablen am besten schätzen, indem man das arithmetische Mittel der Stichprobe, also $\bar{x}$, berechnet. Also berechnen wir: $\bar{x} = \frac{0\cdot2+1\cdot5+2\cdot6+3\cdot5+4\cdot3+5\cdot2+6\cdot1}{24} = \frac{60}{24} = 2.5$. Daher sollte der Händler die folgende Poisson-Verteilung benutzen: $P(X = k) = \frac{2.5^k}{k!} \cdot e^{-2.5}$. Damit würde er beispielsweise die Wahrscheinlichkeit, zwei Bagger in einem Monat zu verkaufen, als $P(X = 2) = \frac{2.5^2}{2!} \cdot e^{-2.5} \approx 0.2565$ schätzen. Tatsächlich hat er in sechs von 24 Monaten zwei Bagger verkauft, also in 25 % der Monate. Die Schätzung ist somit ziemlich genau.

müssen wir nur noch die Trefferwahrscheinlichkeit in jedem Zug berechnen. Nun ja, bei acht Schlüsseln beträgt die Trefferwahrscheinlichkeit in jeden Zug ein Achtel, also $p = \frac{1}{8}$.

Damit haben wir alles beisammen, um die Wahrscheinlichkeit für 10 Versuche auszurechnen: $P(X = 10) = (1 - \frac{1}{8})^{(10-1)} \cdot \frac{1}{8} = (\frac{7}{8})^9 \cdot \frac{1}{8} \approx 0.0376$. Es ist also relativ unwahrscheinlich, dass der Student so viele Versuche benötigt. Trotzdem wäre es besser, sich die Schlüssel zu merken, dann wäre nämlich bei acht Versuchen definitiv Schluss. ◀

Binomialverteilung und hypergeometrische Verteilung beschreiben Urnenexperimente

Stellen wir uns das folgende Zufallsexperiment vor: Vor uns steht eine (undurchsichtige) Urne, in der sich r rote und s schwarze Kugeln (also insgesamt $r + s$ Kugeln) befinden. Wir sollen nun n Kugeln aus dieser Urne ziehen (wie wir die genau ziehen sollen, wird gleich erläutert) und interessieren uns für die Wahrscheinlichkeit, dass unter den n gezogenen Kugeln genau k rote Kugeln sind.

Bei diesen Urnenexperimenten, die auch für ganz andere Sachverhalte angewendet werden können, muss man selbstständig in der Lage sein zu entscheiden, ob man die Binomialverteilung oder die hypergeometrische Verteilung anwenden muss. Hierzu muss man drei Fragen beantworten. Zuerst die folgenden zwei:

1. Soll die Zufallsvariable X die Anzahl der Treffer darstellen?
2. Können bei jeder Ziehung genau zwei Ergebnisse rauskommen, z. B. gut und schlecht, schwarz oder rot, Kopf oder Zahl?

Wenn wir beide Fragen mit Ja beantwortet haben, können wir uns sicher sein, dass wir es mit einem Urnenexperiment zu tun haben. Jetzt muss die Entscheidung für eine der beiden Verteilungen gefüllt werden. Hierzu müssen wir uns die dritte Frage stellen:

3. Erfolgt die Ziehung mit oder ohne Zurücklegen, bzw., wenn man kein Zurücklegen erkennt, ist in jeder Ziehung die Trefferwahrscheinlichkeit konstant?

Beim Ziehen mit Zurücklegen bzw. beim Vorliegen einer konstanten Trefferwahrscheinlichkeit in jedem Zug stellt die **Binomialverteilung** das geeignete Modell dar.

Abb. 17.11 zeigt noch einmal die Vorgehensweise, wie wir zwischen hypergeometrischer und Binomialverteilung wählen sollten.

Eine Zufallsvariable X, die binomialverteilt ist, kann als Werte die Zahlen $0, 1, \ldots, n$ annehmen, wobei n die Anzahl der Ziehungen ist. Ein Wert k wird dann mit der Wahrscheinlichkeit

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

angenommen, wobei $p = \dfrac{r}{r+s}$ die als konstant angenommene Trefferwahrscheinlichkeit in einer Ziehung bezeichnet. Hierbei bezeichnet (wie oben schon erwähnt) r die Anzahl der roten Kugeln bzw. die Anzahl der Treffer und s die Anzahl der schwarzen Kugeln, also die Anzahl der Nichttreffer.

Man kann zeigen, dass dann gilt:

$$\mu = E[X] = n \cdot p \text{ und } \sigma^2 = \mathrm{Var}(X) = n \cdot p \cdot (1-p).$$

Teil IV

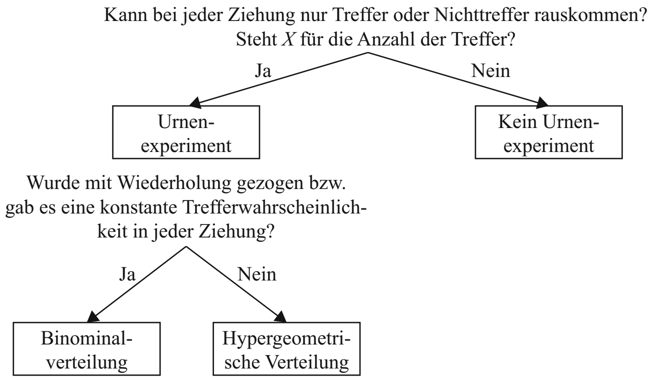


Abb. 17.11 Unterscheidung zwischen Binomialverteilung und hypergeometrischer Verteilung

Man schreibt übrigens abkürzend $X \sim \text{Bin}(n; p)$. Durch die Angabe der beiden Parameter ist die Verteilung eindeutig bestimmt.

Wenn man die Verteilungsfunktion der Binomialverteilung, die $F_{n,p}(x)$ heißt und wie immer $P(X \leq x)$ ist, berechnen will, kann man dies bei kleinen Werten von n und x durch Addition der Wahrscheinlichkeiten $P(X = k)$ für $k \leq x$ tun. Bei großen Werten für n und x wird das aber sehr aufwändig. Für diese Fälle gibt es Tabellen, in denen wir die Werte der Verteilungsfunktion direkt und ohne Rechnen ablesen können.

Vorgehensweise Ablesen in der Binomialverteilungstabelle

Schritt 1: Wir prüfen, ob wir tatsächlich $P(X \leq x)$ berechnen wollen.

Schritt 2: Wir prüfen, ob es sich bei unserer Problemstellung tatsächlich um die Binomialverteilung handelt.

Schritt 3: Wir benennen x, n und p.

Schritt 4: Wir wählen die **Binomialverteilungstabelle** zu unserem p aus.

Schritt 5: Wir suchen in der ersten Spalte das richtige n.

Schritt 6: Wir suchen in der ersten Zeile das richtige x.

Schritt 7: Wir finden den Wert, der sich im Kreuzungspunkt der betreffenden Zeile und der Spalte befindet. Dieser Wert ist die gesuchte Wahrscheinlichkeit $P(X \leq x)$ bzw. $F_{n,p}(x)$.

Bevor wir dazu ein Beispiel betrachten, stellen wir noch die Eigenschaften der hypergeometrischen Verteilung vor.

Die **hypergeometrische Verteilung** wird immer dann als Modell angenommen, wenn man ein Urnenexperiment ohne Zurücklegen durchführt.

Eine Zufallsvariable, die hypergeometrisch verteilt ist, kann als Werte die Zahlen von $\max(n-s, 0)$ bis $\min(n, r)$ annehmen. Ein Wert k wird dann mit der Wahrscheinlichkeit

$$P(X = k) = \frac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{r+s}{n}}$$

angenommen.

Man kann zeigen (muss aber nicht unbedingt), dass dann gilt

$$\begin{aligned} \mu = \text{E}[X] = n \cdot \frac{r}{r+s} \text{ und } \sigma^2 &= \text{Var}(X) \\ &= \frac{n \cdot r \cdot s \cdot (r+s-n)}{(r+s)^2 \cdot (r+s-1)}. \end{aligned}$$

Die Fragen, die wir uns stellen müssen, finden wir noch einmal in Abb. 17.11.

Man schreibt übrigens abkürzend $X \sim \text{hyp}(n; r, s)$. Durch die Angabe der drei Parameter ist die Verteilung eindeutig bestimmt.

Kenngrößen von stetigen Verteilungen charakterisieren stetige Verteilungen

Die große Umstellung bei stetigen Zufallsvariablen besteht darin, dass die Wahrscheinlichkeitsfunktion keinen Sinn mehr macht. Aufgrund der Tatsache, dass unendlich viele beliebig nahe beieinander liegende Werte von der Zufallsvariablen angenommen werden können, beträgt die Wahrscheinlichkeit $P(X = x)$ immer 0. Aus diesem Grund wird für stetige Zufallsvariablen ein „Ersatzkonstrukt" herangezogen: die Dichte. Diese Funktion hat tatsächlich ähnliche Eigenschaften wie die Wahrscheinlichkeitsfunktion. Trotzdem ist sie nicht dasselbe wie eine Wahrscheinlichkeit.

Definition: Dichtefunktion

Eine Funktion $f^X(x)$ mit den Eigenschaften

- $f^X(x) \geq 0 \; \forall x \in \mathbb{R}$
- $\displaystyle\int_{-\infty}^{\infty} f^X(x)\, dx = 1$

heißt **Dichtefunktion** der Zufallsvariablen X. Zur Erinnerung: $F^X(x)$ heißt Verteilungsfunktion von X (bitte nicht verwechseln).

An dieser Definition ist gut zu erkennen, dass eine Dichte so etwas ähnliches wie eine Wahrscheinlichkeit ist, denn Wahrscheinlichkeiten müssen ja auch immer positiv sein, und wenn man alle Wahrscheinlichkeiten aufaddiert, erhält man 1. Eine stetige Summe, also eine unendlich feine Summe, ist gerade das Integral.

Wie immer lässt man das ^{X} weg, wenn aus dem Kontext klar ist, welche Zufallsvariable gemeint ist.

Anwendung: Binomialverteilung und hypergeometrische Verteilung

Zur Qualitätsprüfung der Wareneingangskontrolle eines Automobilherstellers werden einem laufenden Produktionsprozess von 100 Getrieben eines Zulieferers zufällig zehn Getriebe entnommen. Die Liefervereinbarung des Automobilzulieferers mit dem Automobilhersteller sieht vor, dass der Automobilhersteller die Getriebe abnimmt, wenn in der Stichprobe vom Umfang 10 höchstens zwei defekte Motoren vorkommen.

Der Produktionsleiter weiß ziemlich genau, dass sich unter den 100 Getrieben fünf defekte Getriebe befinden. Er fragt sich daher, wie groß die Wahrscheinlichkeit ist, dass der Automobilhersteller die Lieferung zurückweist, dass sich also in der Stichprobe mehr als zwei defekte Getriebe befinden. Um seine Frage zu beantworten, müssen wir zunächst die Verteilung erkennen. Ganz klar haben wir es mit einem Urnenexperiment zu tun, denn bei jeder Ziehung kann nur „intakt" oder „defekt", also 0 oder 1, herauskommen, und wir interessieren uns für die Anzahl der Treffer, bei uns nämlich die Zahl der defekten Getriebe. Diese nennen wir X, und wir haben unsere Zufallsvariable definiert.

Im zweiten Schritt müssen wir noch festlegen, ob die zehn Getriebe ohne oder mit Zurücklegen gezogen wurden, ob es also möglich ist, dass ein Getriebe mehrfach geprüft wird oder nicht. Falls jedes Getriebe nur höchstens einmal gezogen wird, haben wir es mit dem Fall „ohne Zurücklegen" zu tun. In diesem Fall ist die hypergeometrische Verteilung das Modell der Wahl. Wenn nach der Prüfung eines Getriebes dieses mit gleicher Wahrscheinlichkeit noch einmal gezogen werden kann, ist die Binomialverteilung die passende Verteilung. In beiden Fällen interessieren wir uns auf jeden Fall für $P(X > 2)$.

Beginnen wir mit dem ersten Fall: Für die hypergeometrische Verteilung benötigen wir die Parameter n, r, s und natürlich das k. n ist ganz offensichtlich 10, nämlich der Stichprobenumfang. r sind die Treffer insgesamt, d. h. die defekten Getriebe insgesamt, also $r = 5$. s sind die Nichttreffer, also alle intakten Getriebe. Das sind logischerweise 95. Wenn wir jetzt wissen wollen, wie groß die Wahrscheinlichkeit ist, dass wir genau zwei defekte Getriebe finden, also $P(X = 2)$, müssen wir

$$P(X = 2) = \frac{\binom{5}{2}\binom{95}{10-2}}{\binom{5+95}{10}} = \frac{\binom{5}{2}\binom{95}{8}}{\binom{100}{10}} \approx 0.0702$$

berechnen.

Leider interessiert uns aber $P(X > 2)$, was sich als $P(X = 3) + P(X = 4) + P(X = 5)$ berechnen lässt. Mehr defekte Getriebe können wir nicht ziehen, weil mehr nicht vorhanden sind. Ein bisschen Arbeit können wir sparen, indem wir nicht dieses Ereignis betrachten, sondern das Gegenteil, also das Komplement. Das berechnet sich als $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$, wobei wir die letztere Wahrscheinlichkeit sogar schon berechnet haben. Fehlen noch

$$P(X = 1) = \frac{\binom{5}{1}\binom{95}{10-1}}{\binom{5+95}{10}} = \frac{\binom{5}{1}\binom{95}{9}}{\binom{100}{10}} \approx 0.339$$

und

$$P(X = 0) = \frac{\binom{5}{0}\binom{95}{10-0}}{\binom{5+95}{10}} = \frac{\binom{5}{0}\binom{95}{10}}{\binom{100}{10}} \approx 0.584.$$

Wenn man die drei Wahrscheinlichkeiten addiert, erhält man $P(X \le 2) = 0.584 + 0.339 + 0.07 = 0.993$. Damit können wir auch $P(X > 2)$ berechnen, denn $P(X > 2) = 1 - P(X \le 2) = 1 - 0.993 = 0.007$. Also beträgt die Wahrscheinlichkeit, dass die Getriebelieferung zurückgeht, bei diesem Verfahren der Stichprobenziehung ungefähr 0.7 %.

Wenn hingegen die Getriebe mehrfach ausgewählt werden können, müssen wir die Binomialverteilung verwenden. Hier brauchen wir neben n, das weiterhin 10 ist, die Trefferwahrscheinlichkeit p, die $p = 5/(5 + 95) = 0.05$ beträgt. Auch hier berechnen wir erst einmal die Wahrscheinlichkeit dafür, dass genau zwei defekte Motoren entdeckt werden:

$$\begin{aligned} P(X = 2) &= \binom{10}{2} \cdot 0.05^2 \cdot (1 - 0.05)^{10-2} \\ &= \binom{10}{2} \cdot 0.05^2 \cdot 0.95^8 \approx 0.0746. \end{aligned}$$

Mit derselben Argumentation wie soeben berechnen wir $P(X \le 2)$ als Summe aus $P(X = 2)$, $P(X = 1)$ und $P(X = 0)$. Hier wäre es noch viel aufwendiger, $P(X > 2)$ direkt zu berechnen, weil sogar bis zu zehn defekte Getriebe gezogen werden könnten. Es gilt:

$$\begin{aligned} P(X = 1) &= \binom{10}{1} \cdot 0.05^1 \cdot (1 - 0.05)^{10-1} \\ &= \binom{10}{1} \cdot 0.05^1 \cdot 0.95^9 \approx 0.315 \end{aligned}$$

und

$$\begin{aligned} P(X = 0) &= \binom{10}{0} \cdot 0.05^0 \cdot (1 - 0.05)^{10-0} \\ &= \binom{10}{0} \cdot 0.05^0 \cdot 0.95^{10} \approx 0.599. \end{aligned}$$

Teil IV

Dies bedeutet, dass $P(X \leq 2) = 0.599 + 0.315 + 0.07 = 0.988$. Nach dieser Vorgehensweise beträgt also die Wahrscheinlichkeit, dass der Automobilhersteller, die Lieferung zurückweist, $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.988 = 0.012$, also ca. 1.2 %. Es fällt Ihnen vielleicht auf, dass der Unterschied zwischen den beiden Verteilungen eher gering ist. Ist ja auch kein Wunder. Ob bei 100 Getrieben die zehn Getriebe mit oder ohne Zurücklegen gezogen werden, ist ja fast egal. Die beiden Verteilungen nähern sich immer mehr einander an, je kleiner n im Vergleich zu $r + s$ ist.

Zum Schluss benutzen wir eine der Binomialverteilungstabellen im Anhang. Wie wir leicht erkennen können, sind diese nach der Trefferwahrscheinlichkeit p geordnet. Wir brauchen also die Tabelle mit $p = 0.05$. Um den richtigen Eintrag zu finden, benötigen wir n und x. Also müssen wir in der ersten Spalte, in der ja n steht, die 10 suchen, in der ersten Zeile hingegen die 2. Dann suchen wir den Wert, der sich im Kreuzungspunkt der Zeile und der Spalte befindet, und treffen auf den Wert 0.988.

Definition: Stetige Verteilungsfunktion

Die Funktion $F^X(x)$ mit der Eigenschaft

$$F^X(x) = P(X \leq x) = \int_{-\infty}^{x} f^X(t)\,dt$$

heißt **Verteilungsfunktion** der Zufallsvariablen X. Auch hier wird oft das ^{X} weggelassen.

Dies passt zu den Berechnungen im diskreten Fall. Dort musste man die Wahrscheinlichkeiten bis zu dem Wert, an dem man die Verteilungsfunktion berechnen wollte, aufaddieren. Jetzt müssen wir die Dichte (als Ersatz für die Wahrscheinlichkeiten) bis zum gesuchten x integrieren, und integrieren ist ja nichts anderes als eine unendlich feine Summe.

Definition: Erwartungswert

Der **Erwartungswert** μ einer stetigen Zufallsvariablen X berechnet sich durch

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

Auch hier erkennen wir die Analogie zur Berechnung des Erwartungswerts einer diskreten Zufallsvariablen. Statt das, Ergebnis mit der Wahrscheinlichkeit, des Eintretens, zu multiplizieren, müssen wir x mit der Dichte multiplizieren. Statt zu addieren, wird integriert.

Definition: k-tes Moment

Das **k-te Moment** einer stetigen Zufallsvariable X berechnet sich durch

$$\mathrm{E}[X^k] = \int_{-\infty}^{\infty} x^k \cdot f(x)dx.$$

Auch hier interessiert man sich meistens für das zweite Moment, weil man es für die Varianz braucht. Dessen Formel lautet:

$$\mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

Und schon wieder gilt dasselbe Analogieprinzip wie oben: Statt der Wahrscheinlichkeit nehmen wir die Dichte, statt zu addieren, wird integriert.

Definition: Varianz und Standardabweichung

Die **Varianz** σ^2 einer stetigen Zufallsvariablen X berechnet sich wie im diskreten Fall durch

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

Die **Standardabweichung** σ einer stetigen Zufallsvariablen X ist wie im diskreten Fall

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Diese beiden Formeln sehen im diskreten Fall selbstverständlich exakt genau so aus. Aber die Berechnung der einzelnen Bestandteile ist eben etwas anders.

Auch im stetigen Fall gibt es wieder einige Verteilungen, die so wichtig sind, dass sie einen Namen bekommen haben. Genau gesagt handelt es sich um die Rechteckverteilung, die Exponentialverteilung und die Normalverteilung mit dem Spezialfall der Standardnormalverteilung. Diese vier schauen wir uns wieder genauer an und betrachten Einsatzgebiet, Wertebereich, Dichte und Verteilungsfunktion, Erwartungswert und Varianz sowie ggf. praktische Hilfsmittel (Das bedeutet übrigens, dass wir bei diesen Verteilungen nicht integrieren müssen, denn die Integration ist bereits erledigt. Wir müssen nur noch einsetzen):

Die Rechteckverteilung ist eine stetige Gleichverteilung

Die **Rechteckverteilung** bzw. **stetige Gleichverteilung** wird immer dann unterstellt, wenn in einem Intervall $[a; b]$ jeder beliebige Wert als Ergebnis möglich ist und es keinen Grund für die Annahme unterschiedlicher Wahrscheinlichkeiten einzelner Ergebnisse gibt.

Dementsprechend ist der Wertebereich für die Zufallsvariable X das Intervall $[a; b]$. Die Verteilungsdichte $f^X(x)$ lautet:

$$f^X(x) = \begin{cases} \frac{1}{b-a}, & \text{wenn } x \in [a; b] \\ 0, & \text{sonst} \end{cases}$$

Diese Dichtefunktion sieht wie in Abb. 17.12 aus.

Dann lässt sich die Verteilungsfunktion $F^X(x) = P(X \leq x)$ berechnen zu:

$$F^X(x) = \begin{cases} 0, & \text{wenn } x < a \\ \frac{x-a}{b-a}, & \text{wenn } a \leq x \leq b \\ 1, & \text{wenn } x > b \end{cases}$$

Diese Verteilungsfunktion sieht wie in Abb. 17.13 aus.

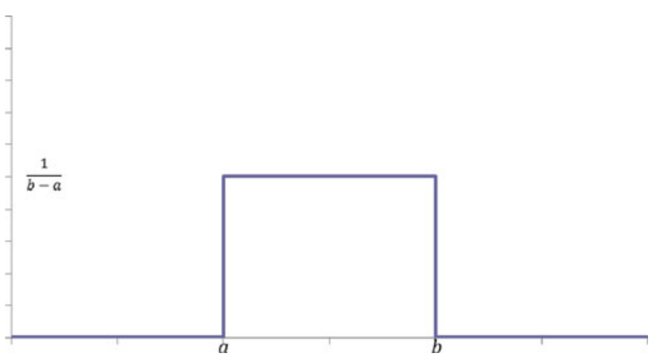


Abb. 17.12 Dichtefunktion der Rechteckverteilung

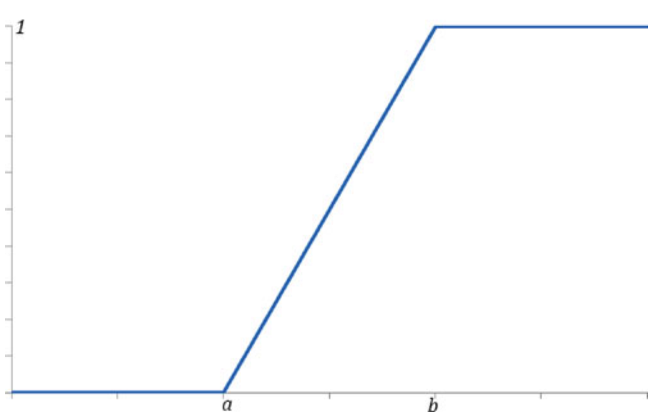


Abb. 17.13 Verteilungsfunktion der Rechteckverteilung

Man kann zeigen, dass

$$E[X] = \frac{a+b}{2} \text{ und } \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Man schreibt abkürzend $X \sim \mathcal{R}[a; b]$. Damit ist klar, dass man das Intervall $[a; b]$ kennen muss.

Beispiel

In einen Blumenkasten werden Sonnenblumen gesät, deren Größe (mehr oder weniger) gleichverteilt im Intervall $[0.8\,\text{m}; 1.5\,\text{m}]$ ist. Um einen geeigneten Platz für den Blumenkasten zu finden, interessiert sich der Besitzer des Blumenkastens für die Wahrscheinlichkeit dafür, dass die Sonnenblumen größer als 1 m werden. Er sucht also $P(X > 1)$, wenn X die zufällige Blumengröße ist. Da die Blumengrößen im Intervall von 0.8 m bis 1.5 m gleichverteilt sein sollen, können wir annehmen, dass $X \sim \mathcal{R}[0.8; 1.5]$.

Da wir die Verteilungsfunktion, also $P(X \leq x)$, kennen, müssen wir die gesuchte Wahrscheinlichkeit so umformen, dass wir die Verteilungsfunktion benutzen können. Hierbei gilt:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - F(1) = 1 - \frac{1-0.8}{1.5-0.8} \\ &= 1 - \frac{0.2}{0.7} = \frac{5}{7} \approx 0.714. \end{aligned}$$

Da diese Wahrscheinlichkeit doch recht groß ist, sollten hinter dem Blumenkasten besser keine Fenster sein, durch die man unbedingt hinaussehen muss. Außerdem beträgt der Erwartungswert der Blumengröße $E[X] = \frac{0.8+1.5}{2} = 1.15$, was auch für einen Bewuchs vor dem Fenster spricht. ◀

Die Exponentialverteilung beschreibt Lebensdauern

Die **Exponentialverteilung** wird vorrangig dann unterstellt, wenn Lebensdauern bzw. Dauern sonstiger zufälliger Ereignisse modelliert werden.

Dementsprechend ist der Wertebereich für die Zufallsvariable X das Intervall $[0; \infty)$. Die Verteilungsdichte $f^X(x)$ lautet

$$f^X(x) = \begin{cases} 0, & \text{wenn } x < 0 \\ \lambda \cdot e^{-\lambda \cdot x}, & \text{wenn } x \geq 0 \end{cases}$$

Abb. 17.14 zeigt die Dichte für verschiedene Werte von λ.

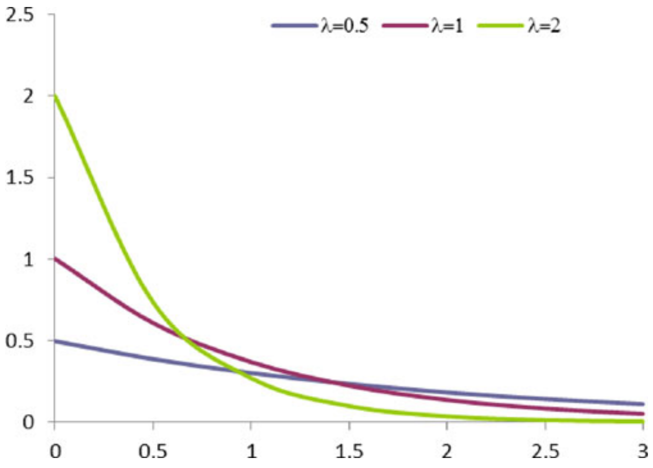


Abb. 17.14 Dichtefunktion der Exponentialverteilung für verschiedene Werte von λ

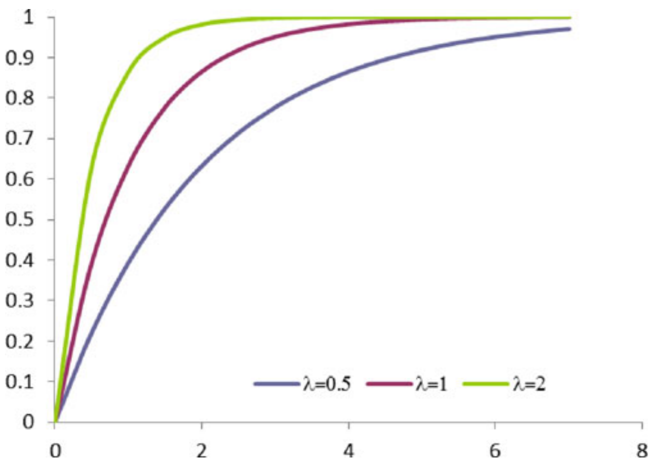


Abb. 17.15 Verteilungsfunktion der Exponentialverteilung für verschiedene Werte von λ

Damit lässt sich die Verteilungsfunktion $F^X(x) = P(X \leq x)$ berechnen zu

$$F^X(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda \cdot x}, & \text{wenn } x > 0 \\ 0 \quad , & \text{sonst} \end{cases}$$

Auch die Verteilungsfunktion ist für einige Werte von λ im Folgenden skizziert:

Mit etwas Aufwand ergibt sich dann

$$\mathrm{E}[X] = \frac{1}{\lambda} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Man schreibt abkürzend $X \sim \exp(\lambda)$. Damit ist klar, dass der Parameter λ bekannt sein muss.

Beispiel

Die Reparaturzeit eines Kühlschranks soll als exponentialverteilt angesehen werden. Es ist bekannt, dass die mittlere Reparaturzeit (also der Erwartungswert) 15 min bzw. 0.25 h beträgt. Mit dieser Information kann man schlussfolgern, dass $\lambda = 4$ sein muss, denn der Erwartungswert bei Exponentialverteilung ist immer der Kehrwert des gesuchten Parameters. Also lautet die Dichtefunktion

$$f^X(x) = \begin{cases} 0 \quad , & \text{wenn } x < 0 \\ 4 \cdot \mathrm{e}^{-4 \cdot x}, & \text{wenn } x \geq 0 \end{cases}$$

und die Verteilungsfunktion

$$F^X(x) = \begin{cases} 1 - \mathrm{e}^{-4 \cdot x}, & \text{wenn } x > 0 \\ 0 \quad , & \text{sonst} \end{cases}$$

Nun kann man die Wahrscheinlichkeit berechnen, dass z. B. eine Reparatur länger als eine Stunde dauert:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - F^X(1) \\ &= 1 - (1 - \mathrm{e}^{-4 \cdot 1}) = \mathrm{e}^{-4} = 0.0183 = 1.83\,\%. \end{aligned}$$

Es ist also nicht sehr wahrscheinlich. ◂

Die Normalverteilung modelliert zufällige Abweichungen von einem erwarteten Wert

Die **Normalverteilung** wird vorrangig dann unterstellt, wenn eigentlich ein bestimmter Wert erwartet wird, der aber zufällig nach oben und unten abweichen kann. Dies ist z. B. bei Produktionsprozessen der Fall, wo eine Solllänge vorgegeben ist, die Werkstücke aber nicht genau die geplante Länge besitzen.

Dementsprechend ist der Wertebereich für die Zufallsvariable X das Intervall $(-\infty; \infty)$. Die Verteilungsdichte $f^X(x)$ lautet

$$f^X(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Einige Dichtefunktionen von Normalverteilungen sind in Abb. 17.16 abgebildet.

Die Verteilungsfunktion $F^X(x) = P(X \leq x)$ lässt sich nicht geschlossen angeben.

Dafür muss man den Erwartungswert und die Varianz/die Standardabweichung auch nicht berechnen; diese sind als Parameter vorgegeben:

$$\mathrm{E}[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2$$

Man schreibt übrigens abkürzend $X \sim \mathcal{N}(\mu; \sigma)$. Manche Autoren schreiben auch $X \sim \mathcal{N}(\mu; \sigma^2)$. Man muss genau hinsehen, was gemeint ist.

Teil IV

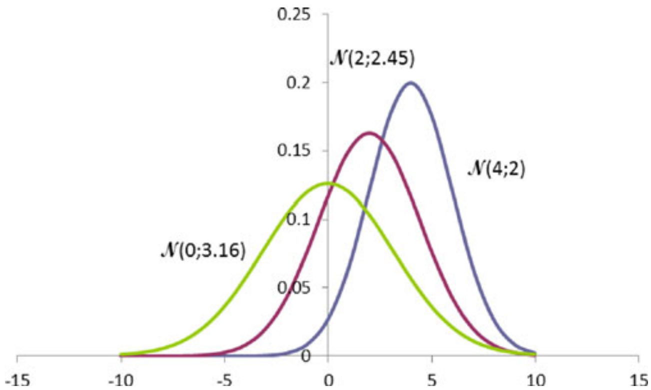


Abb. 17.16 Dichtefunktionen einiger Normalverteilungen

Die Normalverteilung mit Erwartungswert 0 und Standardabweichung 1 heißt Standardnormalverteilung

Es gibt eine Normalverteilung, die besonders wichtig ist, die **Standardnormalverteilung**. Dabei handelt es sich um die Normalverteilung mit Erwartungswert 0 und Standardabweichung 1, d. h. $\mathrm{E}[X] = 0$ und $\mathrm{Var}(X) = 1$. Man schreibt auch: $X \sim \mathcal{N}(0; 1)$. Diese Verteilung ist so wichtig, dass ihre Dichte und ihre Verteilungsfunktion einen eigenen Namen bekommen haben. Die Dichte der Standardnormalverteilung heißt $\varphi(x)$, die Verteilungsfunktion dementsprechend $\Phi(x)$.

Es gilt (das kann man durch Einsetzen nachrechnen)

$$\varphi = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \mathrm{e}^{-\frac{x^2}{2}}$$

Abb. 17.17 zeigt die Dichtefunktion der Standardnormalverteilung:

Die Verteilungsfunktion $\Phi(x)$ kann man nicht geschlossen darstellen. Sie liegt aber in tabellarischer Form vor, und mit ein bisschen Übung kann man jeden benötigten Wert ablesen. Nur die Standardnormalverteilung liegt in Tabellenform vor. Das genügt auch, wenn man das Folgende weiß:

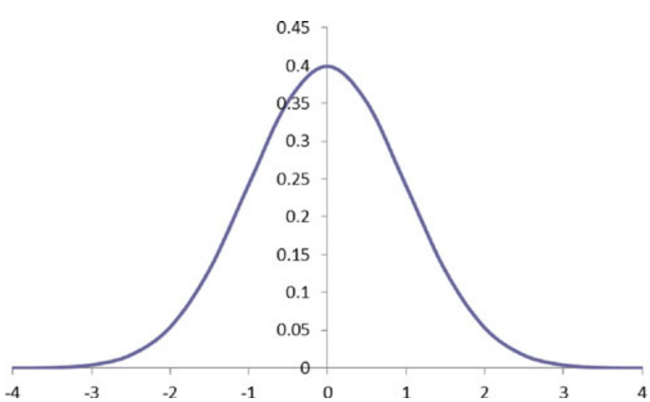


Abb. 17.17 Dichtefunktion der Standardnormalverteilung

Standardisierung

Es sei X eine normalverteilte Zufallsvariable, d. h. $X \sim \mathcal{N}(\mu; \sigma^2)$. Dann ist die Zufallsvariable $Y = \frac{X-\mu}{\sigma}$ standardnormalverteilt, d. h. es gilt:

$$F^X(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Das bedeutet, dass man bei einer beliebigen Normalverteilung nur den Wert verändern muss und dann den Wert der Verteilungsfunktion in der Standardnormalverteilungstabelle ablesen kann.

In Tab. A.11 findet man streng genommen zwei Tabellen. Die obere große Tabelle brauchen wir, wenn wir einen Wert x gegeben haben und den Wert $\Phi(x)$ suchen. Dann müssen wir die erste Vor- und die erste Nachkommastelle in der ersten Spalte suchen, die zweite Nachkommastelle in der ersten Zeile, und im Schnittpunkt der beiden steht der gesuchte Wert.

Bei Tests und Konfidenzintervallen, die wir im nächsten Kapitel kennen lernen, sind wir aber in der umgekehrten Situation. Hier kennen wir immer den Wert von $\Phi(x) = \alpha$, und wir suchen das zugehörige x_α, also das α-Quantil. Hierbei hilft uns die untere Tabelle. In der ersten Zeile findet man die Werte für Φ, die man am häufigsten sucht, und in der zweiten Zeile die zugehörigen Quantile. Falls wir ein Quantil suchen, das in der unteren Tabelle nicht aufgeführt ist, müssen wir in der Mitte der oberen Tabelle ungefähr den Wert für Φ suchen und den entsprechenden Wert für x am Rand der Tabelle ablesen.

Beispiel

$\Phi(2.36) = 0.99086253$. Das 0.95-Quantil der Standardnormalverteilung, also das x mit $\Phi(x) = 0.95$, heißt $x_{0.95} = 1.64485363$. ◀

Außerdem gibt es noch einige nützliche Rechenregeln für die Verteilungsfunktion der Standardnormalverteilung.

Rechenregeln für die Standardnormalverteilung

Sei X eine standardnormalverteilte Zufallsvariable. Dann gilt für die Verteilungsfunktion $\Phi(x)$:

- $\Phi(-x) = 1 - \Phi(x)$.
- Es sei x_p das p-Quantil der Standardnormalverteilung, d. h. $\Phi(x_p) = p$. Dann gilt $x_{1-p} = -x_p$.
- Als Näherung gilt: $\Phi(x) = 1$, wenn $x \geq 4.5$, und $\Phi(x) = 0$, wenn $x \leq -4.5$.

Wenn man das Modell der Normalverteilung unterstellen darf, ergeben sich interessante Wahrscheinlichkeitsabschätzungen, die in der Praxis weit verbreitet sind.

Eigenschaften von $k\sigma$-Bereichen

Wenn eine Zufallsvariable $\mathcal{N}(\mu, \sigma)$-verteilt ist, gilt:

- Im Intervall $[\mu-\sigma; \mu+\sigma]$ liegen etwa 68.3 % der Ausprägungen.
- Im Intervall $[\mu - 2\sigma; \mu + 2\sigma]$ liegen etwa 95.4 % der Ausprägungen.
- Im Intervall $[\mu - 3\sigma; \mu + 3\sigma]$ liegen etwa 99.9 % der Ausprägungen, also praktisch alle.

Das kann man leicht nachrechnen. Beginnen wir damit, die Wahrscheinlichkeit auszurechnen, dass eine $\mathcal{N}(\mu, \sigma)$-verteilte Zufallsvariable im Intervall $[\mu - \sigma; \mu + \sigma]$ liegt:

$$\begin{aligned} P(X \in [\mu - \sigma; \mu + \sigma]) &= P(X \leq \mu + \sigma) - P(X < \mu - \sigma) \\ &= \Phi\left(\frac{\mu + \sigma - \mu}{\sigma}\right) \\ &\quad - \Phi\left(\frac{\mu - \sigma - \mu}{\sigma}\right) \\ &= \Phi(1) - \Phi(-1) = 2 \cdot \Phi(1) - 1 \\ &= 2 \cdot 0.84134 - 1 = 68.3\,\% \end{aligned}$$

Die Herleitung der anderen Abschätzungen läuft völlig analog.

Das Anwendungsbeispiel „Standardisierung" erläutert die Anwendung.

17.6 Stochastische Unabhängigkeit und Unkorreliertheit von Zufallsvariablen

Ebenso wie Ereignisse können auch Zufallsvariablen stochastisch unabhängig und unkorreliert sein. Das schauen wir uns etwas näher an. Als Erstes müssen wir uns an eine neue Schreibweise gewöhnen. Statt des $\cap$-Zeichens macht man bei Zufallsvariablen ein Komma. Daher wird die stochastische Unabhängigkeit bei Zufallsvariablen wie folgt erklärt:

Definition: Stochastische Unabhängigkeit von Zufallsvariablen

- Zwei diskrete Zufallsvariablen X und Y heißen **stochastisch unabhängig**, wenn gilt: Das Produkt der beiden Randwahrscheinlichkeiten ist immer gleich der gemeinsamen Wahrscheinlichkeit

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \quad \text{für alle } i, j = 1, 2, \ldots$$

- Zwei stetige Zufallsvariablen X und Y heißen **stochastisch unabhängig**, wenn gilt: Das Produkt der beiden Randdichten ist gleich der gemeinsamen Dichte

$$f^{X,Y}(x, y) = f^X(x) \cdot f^Y(y) \text{ für alle } x, y \in \mathbb{R}.$$

- Bei mehr als zwei Zufallsvariablen unterscheidet man wiederum zwischen gemeinsamer und paarweiser stochastischer Unabhängigkeit. Hier müssen die entsprechenden Produktregeln erfüllt sein.

Die Unabhängigkeit und Unkorreliertheit können wir an dieser Stelle nur für den diskreten Fall detailliert besprechen und berechnen. Im stetigen Fall muss man hierzu Doppel- und Dreifachintegrale lösen.

Generell (also egal ob die Zufallsvariablen diskret oder stetig sind) gilt:

Definition: Kovarianz und Unkorreliertheit

- Die **Kovarianz** zweier Zufallsvariablen X, Y ist definiert als

$$\operatorname{Kov}(X, Y) = \mathrm{E}[X \cdot Y] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$$

- Zwei Zufallsvariablen X, Y heißen **unkorreliert**, falls

$$\operatorname{Kov}(X, Y) = 0$$

bzw. falls
$$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y].$$

Es bleibt nur noch zu klären, wie man den Erwartungswert $\mathrm{E}[X \cdot Y]$ berechnet.

Berechnung von $\mathrm{E}[X \cdot Y]$

Seien X, Y zwei diskrete Zufallsvariablen mit Werten $x_1, x_2, \ldots, x_n$ bzw. $y_1, y_2, \ldots, y_m$. Dann gilt:

$$\mathrm{E}[X \cdot Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i \cdot y_j \cdot P(X = x_i, Y = y_j)$$

An einem Beispiel sieht das Ganze schon nur noch halb so schlimm aus:

Es muss klar zwischen stochastischer Unabhängigkeit und Unkorreliertheit getrennt werden. Unkorreliertheit bezeichnet lediglich den linearen Zusammenhang zwischen zwei Zufallsvariablen. Zwischen Unabhängigkeit und Unkorreliertheit besteht folgender Zusammenhang:

Zusammenhang von Unkorreliertheit und Unabhängigkeit

- Sind zwei Zufallsvariablen X, Y stochastisch unabhängig, so sind sie auch unkorreliert, d. h. $\operatorname{Kov}(X, Y) = 0$.
- Wenn zwei Zufallsvariablen X, Y unkorreliert sind, sind sie nicht zwangsläufig stochastisch unabhängig.

Anwendung: Standardisierung

In einem Produktionsprozess werden Schrauben mit einem Durchmesser von 5 mm gefertigt. Es ist bekannt, dass die Standardabweichung des Produktionsprozesses bei $\sigma = 0.01$ mm liegt. Man nimmt daher an, dass der zufällige Durchmesser einer Schraube X normalverteilt ist, $X \sim \mathcal{N}(5; 0.01)$.

Der Produktionsleiter interessiert sich nun dafür, wie groß die Wahrscheinlichkeit ist, dass eine Schraube einen Durchmesser von höchstens 4.98 mm besitzt, da solche Schrauben nicht mehr einwandfrei in die zugehörigen Schraubenmuttern passen.

Um diese Frage zu klären, muss $P(X \leq 4.98)$ berechnet werden. Gemäß obiger Formel gilt

$$\begin{aligned} P(X \leq 4.98) &= \Phi\left(\frac{4.98-5}{0.01}\right) \\ &= \Phi(-2) \quad = 1-\Phi(2) \quad = 1-0.9772 \\ &= 0.0228 \end{aligned}$$

Etwa 2.3 % der Schrauben werden also nicht wackelfrei zu den vorgesehenen Muttern passen.

Des Weiteren interessiert sich der Produktionsleiter für die Wahrscheinlichkeit, dass eine zufällig ausgewählte Schraube einen Durchmesser von mehr als 5.02 mm aufweist. Solche Schrauben lassen sich nicht mehr in die zugehörigen Muttern drehen.

Diesmal müssen wir also $P(X > 5.02)$ berechnen. Es gilt:

$$\begin{aligned} P(X < 5.02) &= 1 - P(X \leq 5.02) = 1 - \Phi\left(\frac{5.02-5}{0.01}\right) \\ &= 1-\Phi(2) \quad = 1-0.9772 = 0.0228 \end{aligned}$$

Auch diese Wahrscheinlichkeit liegt also bei 2.3 %.

Nach den obigen Argumentationen betrachtet der Produktionsleiter alle Schrauben, deren Durchmesser zwischen 4.98 und 5.02 mm liegen, als in Ordnung. Dementsprechend möchte er gern wissen, wie groß die Wahrscheinlichkeit ist, dass eine zufällig ausgewählte Schraube den Ansprüchen genügt. Hierzu berechnen wir Folgendes:

$$\begin{aligned} P(4.98 \leq X \leq 5.02) &= P(X \leq 5.02) - P(X < 4.98) \\ &= \Phi\left(\frac{5.02-5}{0.01}\right) - \Phi\left(\frac{4.98-5}{0.01}\right) \\ &= \Phi(2) - \Phi(-2) = 2\cdot\Phi(2) - 1 \\ &= 2\cdot 0.9772 - 1 = 0.9544 \end{aligned}$$

Da in dem Unternehmen, das die Schrauben fertigt, der Qualitätsanspruch dahingehend definiert ist, dass mindestens 98 % der Schrauben innerhalb einer definierten Toleranz liegen sollen, möchte der Produktionsleiter berechnen, wo aktuell die Grenzwerte liegen, die 98 % der Durchmesser einhalten. Er will also berechnen:

$$\begin{aligned} P(5-c \leq X \leq 5+c) &= P(X \leq 5+c) - P(X < 5-c) \\ &= \Phi\left(\frac{c}{0.01}\right) - \Phi\left(\frac{c}{0.01}\right) \\ &= 2\cdot\Phi\left(\frac{c}{0.01}\right) - 1 \end{aligned}$$

Für diese Wahrscheinlichkeit gilt gemäß der Fragestellung, dass sie 98 % betragen soll. Daher müssen wir die Gleichung

$$2\cdot\Phi\left(\frac{c}{0.01}\right) - 1 = 0.98$$

nach c auflösen. Es gilt:

$$\begin{aligned} 2\cdot\Phi\left(\frac{c}{0.01}\right) - 1 &= 0.98 \\ \Phi\left(\frac{c}{0.01}\right) &= 0.99 \\ \frac{c}{0.01} = x_{0.99} &= 2.326 \\ c = 2.326\cdot 0.01 &= 0.0233 \end{aligned}$$

Aktuell liegt der Durchmesser einer zufällig ausgewählten Schraube also mit 98 %iger Wahrscheinlichkeit zwischen $5 - 0.0233 = 4.977$ mm und $5 + 0.0233 = 5.0233$ mm. Das bedeutet, dass die Qualitätsansprüche des Unternehmens nicht erfüllt sind, da wir festgelegt hatten, dass nur Schrauben mit Durchmessern zwischen 4.98 mm und 5.02 mm akzeptabel sind.

Teil IV

Anwendung: Zusammenhänge bei zwei Zufallsvariablen

Wir betrachten ein Bauunternehmen, bei dem aufgrund von schlechter Leitung sowohl der Umsatz X als auch die Kosten Y zufällig verteilt sind. In der folgenden Kontingenztabelle sind die möglichen Werte für Umsatz und Kosten sowie die zugehörigen gemeinsamen Wahrscheinlichkeiten gegeben:

$X\backslash Y$	0	1	2
0	0.1	0.2	0.3
1	0.2	0.1	0.1

Es gilt also z. B., dass die Wahrscheinlichkeit für einen Umsatz von 1 Mio. € und Kosten von 2 Mio. € 0.1 oder 10 % beträgt.

Als Erstes interessiert uns, ob Umsatz und Kosten stochastisch unabhängig sind. Hierzu müssen wir die Tabelle um die Randwahrscheinlichkeiten erweitern:

$X\backslash Y$	0	1	2	$P(X = x)$
0	0.1	0.2	0.3	0.6
1	0.2	0.1	0.1	0.4
$P(Y = y)$	0.3	0.3	0.4	1

Wenn X und Y stochastisch unabhängig sind, dann muss für alle Einträge gelten, dass der Eintrag in der Mitte dem Produkt der beiden zugehörigen Randwahrscheinlichkeiten entspricht. Aber schon der erste Eintrag zeigt, dass dies nicht erfüllt ist: $P(X = 0, Y = 0) = 0.1$, aber $P(X = 0) \cdot P(Y = 0) = 0.6 \cdot 0.3 = 0.18$. Also gibt es einen Zusammenhang zwischen Umsatz und Kosten, jedoch können wir noch nichts über das Ausmaß oder die Richtung des Zusammenhangs aussagen. Vielleicht handelt es sich ja um einen linearen Zusammenhang. Darauf würde eine hohe Korrelation hinweisen. Also berechnen wir zuerst die Kovarianz und dann die Varianzen der Zufallsvariablen. Wir erhalten

$$\mathrm{E}[X] = 0 \cdot 0.6 + 1 \cdot 0.4 = 0.4,$$
$$\mathrm{E}[Y] = 0 \cdot 0.3 + 1 \cdot 0.3 + 2 \cdot 0.4 = 1.1.$$

Als Nächstes berechnen wir

$$\begin{aligned}\mathrm{E}[XY] &= 0 \cdot 0 \cdot 0.1 + 0 \cdot 1 \cdot 0.2 + 0 \cdot 2 \cdot 0.3 \\ &\quad + 1 \cdot 0 \cdot 0.2 + 1 \cdot 1 \cdot 0.1 + 1 \cdot 2 \cdot 0.1 = 0.3.\end{aligned}$$

Hieraus können wir die Kovarianz von X und Y errechnen:

$$\mathrm{Kov}(X, Y) = 0.3 - 0.4 \cdot 1.1 = -0.14.$$

Es gibt also, wenn überhaupt, einen linearen Zusammenhang im Sinne von: Je höher die Umsätze sind, desto geringer sind die Kosten. Um aber wirklich einschätzen zu können, wie stark dieser Zusammenhang ist, brauchen wir die Korrelation. Dazu fehlen noch die beiden Varianzen, die wir nun berechnen:

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - 0.4^2.$$

Das zweite Moment wiederum berechnet sich zu

$$\mathrm{E}[X^2] = 0^2 \cdot 0.6 + 1^2 \cdot 0.4 = 0.4.$$

Also beträgt die Varianz

$$\mathrm{Var}(X) = 0.4 - 0.4^2 = 0.24.$$

Analog ergibt sich für das zweite Moment von Y

$$\mathrm{E}[Y^2] = 0^2 \cdot 0.3 + 1^2 \cdot 0.3 + 2^2 \cdot 0.4 = 1.9$$

und somit für die Varianz

$$\mathrm{Var}(Y) = 1.9 - 1.1^2 = 0.69.$$

Insgesamt erhalten wir:

$$\rho = \frac{-0.14}{\sqrt{0.24} \cdot \sqrt{0.69}} = -0.344.$$

Somit wird klar, dass der lineare Zusammenhang nur schwach ausgeprägt ist.

17.7 Rechenregeln für Erwartungswert, Varianz und Kovarianz

Die wichtigen Kenngrößen von Wahrscheinlichkeitsverteilungen kennen wir nun. Aber wie immer genügt es nicht, die Definitionen zu kennen. In den allermeisten Füllen benötigen wir Rechenregeln, um diese Kenngrößen schnell zu berechnen.

Rechenregeln für Erwartungswert, Varianz und Kovarianz

Es seien X, Y, X_i Zufallsvariablen und a, b, c, d reelle Zahlen. Dann gilt:

- $\mathrm{E}[a + bX] = a + b \cdot \mathrm{E}[X]$
- $\mathrm{E}[a \cdot X + b \cdot Y + c] = a \cdot \mathrm{E}[X] + b \cdot \mathrm{E}[Y] + c$
- $\mathrm{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathrm{E}[X_i]$

Teil IV

Man sagt, der Erwartungswert ist linear, denn man kann Additionen und Multiplikationen mit reellen Zahlen ohne Probleme durchführen.

- Wenn X, Y unabhängig sind, dann gilt

$$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$$

- $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$

Achtung Das bedeutet, dass man Vorfaktoren bei der Varianz immer quadrieren muss, wenn man sie vor die Varianz ziehen will. Das liegt daran, dass die Varianz ja eine quadratische Größe ist. Zudem spielen Verschiebungen, also Additionen und Subtraktionen von Konstanten, keine Rolle, fallen also weg. ◀

- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Kov}(X, Y)$
  Man muss also nicht nur die Einzelvarianzen addieren, sondern auch noch die doppelte Kovarianz hinzufügen.
- Wenn $X_1, X_2, \ldots, X_n$ stochastisch unabhängig oder unkorreliert sind, dann gilt

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

  Nur in diesen beiden Ausnahmesituationen ist die Varianz der Summe zweier Zufallsvariablen also gleich der Summe der Einzelvarianzen.
- $\mathrm{Kov}(X, Y) = \mathrm{Kov}(Y, X)$
- $\mathrm{Kov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Kov}(aX + b, cY + d) = ac \cdot \mathrm{Kov}(X, Y)$

Beispiel

Untersuchen wir noch einmal das Bauunternehmen aus dem obigen Beispiel. Zuerst betrachten wir den Gewinn G des Unternehmens. Logischerweise gilt $G = X - Y$. Wenn wir nun den erwarteten Gewinn berechnen möchten, können wir die Rechenregel anwenden, die besagt:

$$\mathrm{E}[G] = \mathrm{E}[X - Y] = \mathrm{E}[X] - \mathrm{E}[Y] = 0.4 - 1.1 = -0.7.$$

Also macht das Unternehmen im Mittel einen Verlust in Höhe von 700.000 €. Die Schwankungen des Gewinns sind etwas komplizierter zu berechnen:

$$\begin{aligned}\mathrm{Var}(G) &= \mathrm{Var}(X - Y) = \mathrm{Var}(X + (-Y)) \\ &= \mathrm{Var}(X) + \mathrm{Var}(-Y) + 2 \cdot \mathrm{Kov}(X, -Y) \\ &= \mathrm{Var}(X) + (-1)^2 \cdot \mathrm{Var}(Y) + 2 \cdot (-1) \cdot \mathrm{Kov}(X, Y) \\ &= 0.24 + 0.69 - 2 \cdot (-0.14) = 1.21.\end{aligned}$$

Übrigens bedeutet das, dass der Gewinn durchschnittlich um $\rho = \sqrt{1.21} = 1.1$, also um 1.1 Mio. €, um den durchschnittlichen Gewinn schwankt. ◀

17.8 Grenzwertsätze der Wahrscheinlichkeitsrechnung

Die Grundidee der Wahrscheinlichkeitsrechnung besteht darin, dass es einen Zusammenhang zwischen den beobachtbaren Häufigkeiten und empirischen Kenngrößen auf der einen Seite und den Wahrscheinlichkeiten und den Kenngrößen von Zufallsvariablen auf der anderen Seite gibt. Genauer gesagt gilt, dass das arithmetische Mittel gegen den Erwartungswert konvergiert. Aber die Art der Konvergenz ist etwas anders, als wir es gewohnt sind.

Definition: Stochastische Konvergenz

Eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen **konvergiert stochastisch** gegen eine Zahl $a \in \mathbb{R}$, wenn

$$\lim_{n\to\infty} P(|X_n - a| < \varepsilon) = 1 \text{ für alle } \varepsilon > 0$$

bzw., anders ausgedrückt, wenn

$$\lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0 \text{ für alle } \varepsilon > 0.$$

Das bedeutet, dass die Wahrscheinlichkeit, dass X_n nahe an a liegt, gleich 1 sein muss. Streng genommen heißt es aber nicht, dass X_n im klassischen mathematischen Sinne gegen diese Zahl konvergiert.

Man schreibt

$$X_n \xrightarrow{\text{stoch}} a.$$

Wenn man diese Art der Konvergenz akzeptiert, kann man das schwache Gesetz der großen Zahlen formulieren. Es gibt auch noch ein starkes Gesetz der großen Zahlen, das ähnlich lautet, aber noch einen anderen Konvergenzbegriff verwendet.

Schwaches Gesetz der großen Zahlen

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit $\mu = \mathrm{E}[X_i] \quad \forall i \in \mathbb{N}$.

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ bezeichne das arithmetische Mittel der ersten n Zufallsvariablen.

Man sagt, diese Folge erfüllt das schwache Gesetz der großen Zahlen genau dann, wenn

$$\overline{X}_n \xrightarrow{\text{stoch}} \mu.$$

Das schwache Gesetz der großen Zahlen gilt in vielen Zusammenhängen, z. B. für unabhängige identisch verteilte Zufallsvariablen in jedem Fall, also für jede zufällige Stichprobe. Daher kann man schlussfolgern, dass sich bei einer Zufallsstichprobe

Anwendung: Ungleichung von Tschebyscheff

Ein Logistikunternehmen betreibt in einer Region automatisierte Hochregale. Jedes Regal hat (unabhängig von den anderen) mit einer Wahrscheinlichkeit von $\frac{1}{20}$ eine Störung pro Woche. Für die Entscheidung über die personelle Stärke eines ständigen Wartungsdienstes ist die Wahrscheinlichkeit von Interesse, dass pro Woche die Anzahl der defekten Regale mehr als fünf und weniger als 15 beträgt.

Wir nennen daher die Anzahl der defekten Hochregale pro Woche X. Damit ist X binomialverteilt mit den Parametern $n = 200$ und $p = \frac{1}{20}$. Gesucht ist $P(5 < X < 15)$. Exakt errechnet sich diese Wahrscheinlichkeit zu

$$\begin{aligned} P(5 < X < 15) &= P(X < 15) - P(X \leq 5) \\ &= P(X \leq 14) - P(X \leq 5) \\ &= F_{200;0.05}(14) - F_{200;0.05}(5) \\ &= 0.922 - 0.062 = 0.86. \end{aligned}$$

Da man diese Werte aber schon nicht mehr aus den einschlägigen Tabellen ablesen kann, sondern in Excel berechnen muss, nähern wir dieselbe Wahrscheinlichkeit einmal mit der Ungleichung von Tschebyscheff an. Hierfür benötigen wir den Erwartungswert und die Varianz von X, was aber kein Problem ist, da wir ja die Verteilung von X kennen.

Wenn $X \sim Bin(200; 0.05)$, dann ist $\mathrm{E}[X] = 200 \cdot 0.05 = 10$ und $\mathrm{Var}(X) = 200 \cdot 0.05 \cdot 0.95 = 9.5$. Nun müssen wir die gesuchte Wahrscheinlichkeit noch so umformen, dass sie wie die Wahrscheinlichkeit aussieht, die mithilfe der Ungleichung von Tschebyscheff angenähert werden kann.

Hierzu subtrahieren wir zuerst den Erwartungswert auf allen Seiten der Ungleichung: $P(5 < X < 15) = P(5 - 10 < X - 10 < 15 - 10) = P(-5 < X - 10 < 5)$. Wir erhalten also eine Doppelungleichung, bei der auf der linken Seite dieselbe Zahl steht wie auf der rechten Seite, nur mit einem anderen Vorzeichen. Das passt perfekt, denn das können wir mithilfe des Betrags in eine einfache Ungleichung umformen. Schließlich gilt ja die folgende Umformung: $|x| < a \Leftrightarrow -a < x < a$.

Und wenn wir diese Umformung von rechts nach links lesen, erhalten wir eine sehr schöne Umformung für die obige Doppelungleichung: $P(5 < X < 15) = P(5 - 10 < X - 10 < 15 - 10) = P(-5 < X - 10 < 5) = P(|X - 10| < 5)$.

Und genau dafür kann die Ungleichung von Tschebyscheff angewendet werden. Gemäß der Ungleichung von Tschebyscheff gilt $P(5 < X < 15) = P(5-10 < X-10 < 15-10) = P(-5 < X - 10 < 5) = P(|X - 10| < 5) \geq 1 - 9.5/5^2 = 1 - 0.38 = 0.62$.

Hieran erkennt man gut das Problem der Ungleichung von Tschebyscheff: Sie ist oftmals relativ ungenau. Denn zwischen der Näherung 0.62 und dem exakten Wert 0.86 liegen immerhin 24 Prozentpunkte. Die können bei der Entscheidung, wie stark das Wartungsteam sein soll, schon eine gewichtige Rolle spielen.

Anwendung: Zentraler Grenzwertsatz

Betrachten wir noch einmal die Hochregale aus dem letzten Beispiel. Diesmal müssen wir aber anders modellieren, denn beim zentralen Grenzwertsatz gibt es nicht mehr nur eine Zufallsvariable X, sondern wir brauchen eine Folge von Zufallsvariablen X_i.

Hier bietet es sich an, den Zustand des i-ten Hochregals als X_i zu bezeichnen. Und zwar soll $X_i = 1$ bedeuten, dass das Regal i defekt ist, während $X_i = 0$ bedeutet, dass das betreffende Regal intakt ist.

Also ist auch klar, welche Verteilung wir benötigen: $X_i \sim Bin(1; 0.05)$. Auch für den zentralen Grenzwertsatz brauchen wir $\mathrm{E}[X_i] = 0.05$ und $\mathrm{Var}(X_i) = 0.05 \cdot 0.95 = 0.0475$.

Dann lässt sich die gesuchte Wahrscheinlichkeit schreiben als

$$\begin{aligned} P(5 < S_{200} < 15) &= P(S_{200} < 15) - P(S_{200} \leq 5) \\ &= P(S_{200} \leq 14) - P(S_{200} \leq 5). \end{aligned}$$

Diese beiden Wahrscheinlichkeiten können wir nun mithilfe des zentralen Grenzwertsatzes annähern:

$$\begin{aligned} P(5 < S_{200} < 15) &= P(S_{200} < 15) \\ -P(S_{200} \leq 5) &= P(S_{200} \leq 14) - P(S_{200} \leq 5) \\ &\approx \Phi\left(\frac{14 - 200 \cdot 0.05}{\sqrt{0.0475} \cdot \sqrt{200}}\right) \\ &\quad - \Phi\left(\frac{5 - 200 \cdot 0.05}{\sqrt{0.0475} \cdot \sqrt{200}}\right) \\ &= \Phi(1.30) - \Phi(-1.62) \\ &= \Phi(1.30) - 1 + \Phi(1.62) \\ &= 0.9032 - 1 + 0.9474 = 0.8506. \end{aligned}$$

Das ist doch schon mal eine Näherung, die sich sehen lassen kann. Exakt beträgt die Wahrscheinlichkeit immer noch 0.86.

das arithmetische Mittel dem wahren Erwartungswert annähert (im stochastischen Sinn).

Dieses Gesetz schließt sozusagen die Lücke zwischen deskriptiver Statistik und Wahrscheinlichkeitsrechnung: Wenn man die Stichprobe groß genug wählt (im schlimmsten Fall unendlich groß), dann nähert sich das arithmetische Mittel tatsächlich dem (rein theoretischen) Erwartungswert an.

Es gibt noch weitere wichtige Zusammenhänge, die man zur Abschätzung von Wahrscheinlichkeiten benutzt, also in solchen Situationen, in denen man die genaue Wahrscheinlichkeit nicht berechnen kann.

Die erste dieser Abschätzungen ist die **Ungleichung von Tschebyscheff**.

Ungleichung von Tschebyscheff

Sei X eine Zufallsvariable mit $\mu = \mathrm{E}[X]$ und $\sigma^2 = \mathrm{Var}(X)$, $a > 0$ beliebig. Dann gilt

$$P(|X-\mu| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

bzw., anders formuliert,

$$P(|X-\mu| < a) \geq 1 - \frac{\mathrm{Var}(X)}{a^2}$$

Das bedeutet: Die Wahrscheinlichkeit, dass eine Zufallsvariable von ihrem Erwartungswert um mindestens a abweicht, beträgt höchstens ihre Varianz dividiert durch a^2.

Diese Ungleichung benutzt man als Näherung, um Wahrscheinlichkeiten zu berechnen, aber auch, um stochastische Konvergenz nachzuweisen.

Eine weitere Möglichkeit, bestimmte Wahrscheinlichkeiten anzunähern, besteht in der Anwendung des zentralen Grenzwertsatzes.

Zentraler Grenzwertsatz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit identischen Erwartungswerten $\mathrm{E}[X_i] = \mu$ und identischen Varianzen $\mathrm{Var}(X_i) = \sigma^2$.

Ferner sei $S_n = \sum_{i=1}^{n} X_i$ die n-te Partialsumme, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ sei das arithmetische Mittel der ersten n Zufallsvariablen.

Dann gilt

$$P(S_n \leq x) \approx \Phi\left(\frac{x - n\cdot\mu}{\sigma\cdot\sqrt{n}}\right)$$

bzw.

$$P(\overline{X}_n \leq x) \approx \Phi\left(\frac{x-\mu}{\sigma}\cdot\sqrt{n}\right)$$

Und das ist unabhängig von der tatsächlichen Verteilung der einzelnen X_i.

Bereits bei einer Stichprobengröße von 20 bis 50 kann die Approximation als hinreichend genau angesehen werden.

Aufgaben

17.1 Geben Sie in den folgenden Beispielen jeweils eine passende Ergebnismenge Ω an:

a) Zweimaliges Werfen einer Münze, wenn man sich nur für die gefallenen Seiten, nicht aber für die Reihenfolge interessiert.
b) Zweimaliges Werfen einer Münze, wenn man sich zusätzlich für die Reihenfolge interessiert.
c) Familien mit drei Kindern werden nach dem Geschlecht der Kinder und der Reihenfolge, in der die Kinder geboren wurden, befragt.
d) Familien mit drei Kindern werden nach dem Geschlecht der Kinder befragt.
e) In einer Urne befinden sich eine rote, zwei gelbe und zehn schwarze Kugeln. Es werde dreimal ohne Zurücklegen gezogen und die Reihenfolge notiert.
f) Aus der Urne aus Aufgabe e werde dreimal mit Zurücklegen gezogen und die Reihenfolge sei ohne Bedeutung.

17.2 Für die folgenden Zufallsexperimente gebe man ein passendes Ω an und ordne den genannten Ereignissen jeweils Teilmengen von Ω zu. Bestimmen Sie zusätzlich die Wahrscheinlichkeiten der Ereignisse.

a) Beim dreimaligen Werfen eines Würfels ist die Summe ≤ 2.
b) Beim dreimaligen Werfen eines Würfels ist die Summe ≥ 16.
c) Beim zweimaligen Werfen eines Würfels ist die erste Zahl gerade und die zweite ≥ 5.
d) Beim viermaligen Werfen einer Münze werden mindestens drei Wappen beobachtet.

17.3 Eine Firma lässt zwei verschiedene Werbespots (W_1, W_2) im Fernsehen senden. Es ist bekannt, dass ein Fernsehzuschauer den Werbespot W_1 mit Wahrscheinlichkeit 0.1, den Werbespot W_2 mit Wahrscheinlichkeit 0.15 und beide mit Wahrscheinlichkeit 0.05 sieht.

Berechnen Sie die Wahrscheinlichkeit dafür, dass eine Person

a) genau einen,
b) mindestens einen,
c) keinen,
d) höchstens einen

Werbespot sieht.

17.4 Für den Prozess von Herrn K. standen zwei Richter R_1, R_2 zur Verfügung. Die Wahrscheinlichkeit, dass die Wahl auf Richter R_1 fiel, betrug 0.8 und die für Richter R_2 0.2. Es war ferner bekannt, dass Herr K. seinen Prozess mit Wahrscheinlichkeit 0.95 verliert, sollte Richter R_1 den Prozess leiten, und dass er ihn mit Wahrscheinlichkeit 0.9 verliert, wenn Richter R_2 für den Prozess ausgewählt werden sollte.

Wie groß ist die Wahrscheinlichkeit, dass Richter R_1 den Vorsitz hatte, wenn Herr K. seinen Prozess verloren hat?

17.5 Von einem Test zur Diagnose einer bestimmten Krankheit ist bekannt, dass er 98 % der Kranken und 95 % der Nichtkranken richtig anzeigt. Die Wahrscheinlichkeit, dass eine Person diese Krankheit hat, betrage p.

a) Mit welcher Wahrscheinlichkeit hat eine als krank diagnostizierte Person tatsächlich diese Krankheit?
b) Ermitteln Sie die Wahrscheinlichkeit aus Aufgabe a für $p = 0.005; 0.01; 0.05; 0.1$.

17.6 Ein Würfel werde zweimal geworfen, wobei alle Ergebnisse gleich wahrscheinlich seien. Gegeben sind die folgenden Ereignisse:

A: Eine Drei im ersten Wurf

B: Eine Eins im zweiten Wurf

C: Zwei gleiche Augenzahlen

D: Die Augensumme ist ungerade

E: Die Augensumme ist gerade

Untersuchen Sie die folgenden Ereignisse auf Unabhängigkeit:

a) A und B
b) C und D
c) A und C
d) C und E
e) B und D

17.7 Ein Produkt wird in zwei Phasen hergestellt. Zunächst sind für die Grobbearbeitung zwei Maschinen notwendig, wobei drei Produktionsstraßen benutzt werden können. Bezeichnet G_{ij} die i-te Maschine in der j-ten Produktionsstraße ($i = 1, 2$, $j = 1, 2, 3$), so sind die Intaktwahrscheinlichkeiten der einzelnen Maschinen gegeben durch

$$P(G_{11}) = 0.9, \quad P(G_{12}) = 0.8, \quad P(G_{13}) = 0.7,$$
$$P(G_{21}) = 0.8, \quad P(G_{22}) = 0.8, \quad P(G_{23}) = 0.9.$$

Anschließend erfolgen die Feinarbeiten, wofür eine einzige Maschine mit Ausfallwahrscheinlichkeit 0.05 zur Verfügung steht. Alle sieben Komponenten der Gesamtanlage seien bzgl. Ausfall stochastisch unabhängig.

Berechnen Sie die Zuverlässigkeit der Anlage, d. h. die Wahrscheinlichkeit dafür, dass eine Produktion stattfinden kann.

17.8 In einem Reihensystem mit sechs Bauteilen, das nur funktioniert, wenn alle Bauteile in Ordnung sind, sei eines ausgefallen, wobei die Ausfallwahrscheinlichkeit für jedes Bauteil gleich groß sei. Ein herbeigerufener Techniker schlägt zur Ermittlung des defekten Bauteils zwei unterschiedliche Verfahren vor:

a) Jedes Bauteil werde einzeln getestet. Die Kosten hierfür belaufen sich auf 1 500 € pro Bauteil.
b) Es werden zuerst drei Bauteile als Ganzes getestet, danach die drei Bauteile, unter denen sich das defekte Bauteil befindet, einzeln. Die Kosten für den Test der drei Bauteile betragen 1 800 € und für das Testen eines Bauteiles müssen wieder 1 500 € bezahlt werden.

Ermitteln Sie, welches Verfahren langfristig günstiger ist.

17.9 Unter den 50 000 Zuschauern eines Fußballspieles befinden sich 35 000 einheimische Zuschauer und 15 000 Zuschauer von auswärts. Ein Sportreporter will fünf rein zufällig ausgewählte Zuschauer befragen. Wie groß ist die Wahrscheinlichkeit, dass sich unter den fünf rein zufällig gewählten Zuschauern

a) höchstens ein Auswärtiger befindet?
b) mindestens vier einheimische Zuschauer befinden?

Berechnen Sie die Wahrscheinlichkeiten unter der Annahme, (i) dass niemand mehrfach gefragt wird bzw. (ii) dass nicht verzeichnet wird, wer bereits befragt wurde.

17.10 In einem Landkreis werden 70 % der Verkehrsunfälle durch überhöhte Geschwindigkeit verursacht. Geben Sie die Wahrscheinlichkeit dafür an, dass von acht an einem beliebigen (festen) Wochenende vorkommenden (und als unabhängig vorausgesetzten) Verkehrsunfällen

a) genau fünf,
b) wenigstens fünf,
c) höchstens fünf

auf überhöhte Geschwindigkeit zurückzuführen sind.

17.11 Gegeben sei die Funktion

$$g(x) = \begin{cases} 0 \ , x < 1 \\ cx^2, 1 \le x \le 5 \\ 0 \ , x > 5 \end{cases}$$

a) Wählen Sie c so, dass die Funktion g eine Dichtefunktion ist.
b) Bestimmen Sie die Verteilungsfunktion.

17.12 Zeigen Sie, dass die Funktion g, definiert durch

$$g(x) = \begin{cases} \frac{3}{2}\left(x - \frac{1}{2}x^2\right), & \text{für } 0 < x < 2 \\ 0 & \text{, sonst} \end{cases}$$

eine Dichte ist.

Berechnen Sie für die Zufallsvariable X, deren Dichte gerade g ist

1. $\mathrm{E}[X]$,
2. $\mathrm{Var}(X)$.

17.13 Die Zufallsvariable X sei im Intervall $[-7; 5]$ gleichverteilt. Berechnen Sie die folgenden Wahrscheinlichkeiten:

a) $P(X \le 3)$
b) $P(X > 2)$
c) $P(-4 < X \le 1.8)$
d) $P(X \le 1 | X \le 2)$

17.14 Die Zufallsvariable X beschreibe die Wartezeit (in Minuten) der Kunden einer Bank. Nehmen Sie an, dass X exponentialverteilt ist mit Parameter $\lambda = 3$. Berechnen Sie die Wahrscheinlichkeit folgender Ereignisse:

a) Der Kunde muss mehr als eine aber höchstens drei Minuten warten.
b) Die Wartezeit übersteigt 2.5 Minuten.
c) Ein Kunde, der schon zwei Minuten gewartet hat, muss mindestens 2.5 weitere Minuten warten.

17.15 In einem 220-V-Stromkreis werde die Spannung gemessen. Die Messabweichung X (in Volt) sei standardnormalverteilt, d. h., für die gemessene Spannung U (in Volt) gelte

$$U = 220 + X \quad \text{mit } U \sim \mathcal{N}(220;1).$$

a) Wie groß ist die Wahrscheinlichkeit dafür, dass bei einer Messung
   (a) höchstens 221 V gemessen werden?
   (b) der Messwert um max. 0.5 % (vom wahren Wert 220 V) nach unten abweicht?
   (c) eine Abweichung von unter 0.1 % auftritt?
b) Welche Höchstspannung wird mit einer Wahrscheinlichkeit von 90 % gemessen?

17.16 Die Zufallsvariablen $R_1, \ldots, R_5$ seien unabhängig mit $\mathrm{E}[R_i] = 0.1$ und $\mathrm{Var}(R_i) = 0.25$ für $i = 1, \ldots, 5$. Es soll angenommen werden, dass sie die möglichen Renditen von fünf Wertpapieren am 30.12.16 beschreiben. Investiert man heute x € in das Wertpapier 1, so beschreibt die Zufallsvariable $X = x(1 + R_1)$ die möglichen Werte dieser Investition am 30.12.07.

a) Berechnen Sie $\mathrm{E}[X]$ und $\mathrm{Var}(X)$.
b) Investiert man heute in jedes der fünf Wertpapiere $\frac{x}{5}$ €, so gibt die Zufallsvariable $Y = \frac{x}{5}(1 + R_1) + \ldots + \frac{x}{5}(1 + R_5)$ die Werte an, die insgesamt aus den fünf Investitionen zum 30.12.17 resultieren können. Berechnen Sie $\mathrm{E}[Y]$ und $\mathrm{Var}(Y)$.
c) Interpretieren Sie die beiden Ergebnisse.
d) Der Anlagebetrag sei 1 000 €. Schätzen Sie mit Hilfe der Ungleichung von Tschebytscheff die Wahrscheinlichkeit dafür ab, dass
   (a) X um mindestens 50 % von $\mathrm{E}[X]$,
   (b) Y um mindestens 50 % von $\mathrm{E}[Y]$ abweicht.

17.17 Eine Bank schreibt ihren Kunden für Münzrollen mit 1-€-Stücken einen Betrag von 50 € gut, ohne den Inhalt der Münzrollen zu prüfen. Die bisherige Erfahrung zeigt, dass

- 10 % der Münzrollen 48 Münzen,
- 20 % der Münzrollen 49 Münzen,
- 60 % der Münzrollen 50 Münzen,
- 10 % der Münzrollen 51 Münzen

enthalten.

a) Bestimmen Sie näherungsweise die Wahrscheinlichkeit, dass 1000 Rollen mindestens 49 650 € wert sind, sowie die Wahrscheinlichkeit, dass sie zwischen 49 670 € und 49 730 € wert sind.
b) Wie groß ist approximativ die Wahrscheinlichkeit, dass 1000 Rollen mehr als 50 000 € wert sind?