



ulm university universität
uulm

Vorlesungsskript

Angewandte Stochastik I

Dr. Katharina Best

Sommersemester 2011

1 Einleitung

»Im Grunde ist Wahrscheinlichkeitsrechnung nur normaler gesunder Menschenverstand, ausgedrückt durch Mathematik. Sie versetzt uns in die Lage, Dinge exakt abzuschätzen, die wir in einer Art Instinkt fühlen, aber nicht erklären können.«
Pierre-Simon Laplace

1.1 Der Begriff *Stochastik*

Die Stochastik gliedert sich in die Wahrscheinlichkeitstheorie und die Statistik. Der Begriff *Stochastik* stammt aus dem Griechischen *στωχαστική* (die Kunst des Vermutens) und ist abgeleitet von *στωχωξ* (Vermutung, Ahnung, Ziel). Der Begriff wurde von Jacob Bernoulli in seinem Buch *Ars conjectandi* 1713 geprägt, in dem das erste Gesetz der großen Zahlen bewiesen wurde.

Die Stochastik beschäftigt sich mit den Ausprägungen und quantitativen Merkmalen des Zufalls. Dabei ist noch zu klären, was Zufall ist. Gibt es Zufall überhaupt? Und was würde man als Zufall betrachten? Mit diesem Aspekt beschäftigt sich die Philosophie¹.

Für die moderne Mathematik ist der Zufall eine Arbeitshypothese, die es unter anderem ermöglicht, Vorgänge in der Natur, Technik oder Wirtschaft, deren Ausgang ungewiß ist, zu beschreiben und zu analysieren. Das Konzept des Zufalls eröffnet einen Weg, Vorgänge mit Wahrscheinlichkeiten zu belegen, diese zu berechnen, zu quantifizieren und zu vergleichen.

Allerdings muss man an dieser Stelle erneut nachhaken: Was ist Wahrscheinlichkeit? Wie können wir Wahrscheinlichkeit interpretieren? Bei einem Spiel, für das wir die Regeln festlegen, ist die Antwort noch vergleichsweise einfach. *Zwei Ausgänge sind gleich wahrscheinlich, wenn keiner bevorzugt wird.* Dazu muss man aus einer weitergehenden Quelle wissen, dass eben keiner bevorzugt ist. Bei vielen Fragestellungen steht der Begriff der Wahrscheinlichkeit für das Maß an Sicherheit resp. Unsicherheit². Diese, der Wahrscheinlichkeit innewohnende, Ambivalenz sollte man stets im Hinterkopf behalten.

1.2 Geschichtliche Entwicklung

Die Stochastik wurde zu Beginn über das Gebiet der Glücksspiele entwickelt. Die ersten Würfelspiele sind in Ägypten (ca. 3500 v. Chr.) nachgewiesen, und wurden in Griechenland und vor allem im römischen Reich fortgesetzt. Zugleich gab es im Handel in Bezug auf Versicherungen von Schiffstranspor-

¹siehe hierzu <http://plato.stanford.edu/entries/chance-randomness/>

²siehe hierzu <http://plato.stanford.edu/entries/probability-interpret/>
und <http://de.wikipedia.org/wiki/Wahrscheinlichkeit>

1 Einleitung



Abbildung 1.1: Von links nach rechts: Gerolamo Cardano, Jakob Bernoulli, Pierre-Simon Laplace und Andrej Nikolajewitsch Kolmogorow.

ten das Bedürfnis nach Berechenbarkeit. Die älteste bekannte Form der Versicherungsverträge stammt aus Babylon (4-3 T. Jahre v. Chr.). Die ersten Sterbetafeln in der Lebensversicherung stammen von dem römischen Juristen Ulpian (220 v. Chr.), die erste datierte Police aus dem Jahr 1347 aus Genua.

Der erste Wissenschaftler, der sich mit diesen Fragestellungen aus mathematischer Sicht beschäftigt, ist Gerolamo Cardano. In seinem *Buch der Glücksspiele* (*Liber de Ludo Aleae*) beschreibt er 1524 zum ersten Mal für den Spieler vorteilhafte Ereignisse und deren Kombinationen beim Würfelspiel, verwendet Binomialkoeffizienten und auch als erster den Anteil

$$\frac{\text{Anzahl vorteilhafter Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

als Maß für die Wahrscheinlichkeit.

Außerdem ist er auch ein Praktiker: Sein Universitätsgehalt bessert er mit dem durch sein Wissen beim Glücksspiel verdienten Geld auf. Etwa hundert Jahre später diskutieren Blaise Pascal und Pierre de Fermat die Wahrscheinlichkeit betreffende Probleme in ihrem Briefwechsel. Wieder etwa hundert Jahre später gelingt Pierre-Simon Laplace 1812 in seinem Buch *Théorie Analytique des Probabilités* eine mathematische Behandlung der Wahrscheinlichkeit.

Allerdings erlaubten diese Betrachtungen noch keine Aussage bei Problemen, bei denen die Gesamtheit der Ereignisse gar nicht beschreibbar ist, und das Problem sich nicht auf eine Betrachtung gleichwahrscheinlicher Ereignisse reduzieren lässt.

An dieser Stelle wurde von David Hilbert ein axiomatischer Ansatz, ähnlich wie er in anderen Gebieten der Mathematik bereits verwendet wurde, gefordert. Basierend auf der Maß- und Integrationstheorie von Émile Borel und Henri Léon Lebesgue, führte Andrei Nikolajewitsch Kolmogorow in seinem Werk *Grundbegriffe der Wahrscheinlichkeitsrechnung* 1933 Axiome der Wahrscheinlichkeitstheorie ein.

1.3 Wahrscheinlichkeitsbegriffe

1.3.1 Klassisches Vorgehen

Im obigen Abschnitt wurde schon die Beschreibung der Wahrscheinlich von Laplace als das Verhältnis der günstigen Ausgänge zu allen,

$$\frac{\text{Anzahl vorteilhafter Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

vorge stellt. Dieses Vorgehen ist jedoch einigen Einschränkungen unterworfen.

So verschwindet zum Einen der Bruch, falls die mögliche Anzahl aller Ereignisse unendlich ist. Zum anderen funktioniert die Definition nur, wenn die Ereignisse, auf die man sich bezieht, alle gleich wahrscheinlich sind. Manchmal ist eine solche Aufspaltung gar nicht möglich.

1.3.2 Statistisches Vorgehen

Das Konzept der *relative Häufigkeit* kann auf Beobachtungen ausgeweitet werden. Die Wahrscheinlichkeit wird dann als

$$\lim_{\text{Anzahl der Beobachtungen} \rightarrow \infty} \frac{\text{Anzahl der Beobachtungen mit gesuchtem Ausgang}}{\text{Anzahl aller Beobachtungen}}$$

verstanden. Dass diese Betrachtung mit der klassischen konform geht, ist dem Gesetz der großen Zahlen geschuldet. Das Problem, welches sich hier ergibt, ist, dass die Beobachtungen tatsächlich gemacht werden müssen, dass also der Zufallsvorgang durchgeführt wird. Bevor man beginnt, kann keine Aussage gemacht werden. Eine weitere Schwierigkeit ist die Frage nach der Durchführbarkeit. Ist es tatsächlich möglich, einen Zufallsvorgang mehrmals, vielleicht sogar sehr oft, unter gleichen Bedingungen durchzulaufen? Selbst wenn es nur zeitlich oder finanziell nicht geht, kommt dieses Verfahren schnell an seine Grenzen.

1.3.3 Subjektive Wahrscheinlichkeitssicht

Bei der subjektiven Betrachtung werden Ereignisse unter Kenntnis externer Informationen bewertet. Diese basieren auf Sachkenntnis, persönlicher Erfahrung und vergangener Beobachtung.

Die Wahrscheinlichkeit reflektiert somit die Unsicherheit resp. Unkenntnis.

Bei diesem Vorgehen besteht die Gefahr, dass die Wahrscheinlichkeit nicht bedacht genug oder nicht unvoreingenommen vergeben werden. Auch werden unterschiedliche Personen zu unterschiedlichen Einschätzungen kommen. Zu beachten ist ebenfalls, dass sich diese Einschätzungen mit dem Eintreffen neuer Informationen immer wieder ändern.

In der Praxis ist die subjektive Wahrscheinlichkeit nicht unbedeutend, denn sowohl das Aufstellen aller möglichen Ausgänge, die gleichwahrscheinlich sind als auch die Wiederholung des Zufallsvorgangs sind selten möglich.

1.4 Problemstellungen der Stochastik

Die typischen Fragestellungen der Stochastik ergeben sich direkt aus den Wahrscheinlichkeitsbegriffen. Zum Einen müssen Vorgänge, die zu betrachten sind, modelliert werden. Dazu muss ein geeignetes Zufallsexperiment erstellt und beschrieben werden.

Des Weiteren müssen Ereignisse formuliert und deren Wahrscheinlichkeiten bestimmt werden. Darüber hinaus sollen interessante Funktionen der Ereignisse, sogenannte Zufallsvariablen, formuliert werden. Ihre Kennzahlen, wie Mittelwert oder Varianz, sind weitere interessante Details. Manchmal werden aber diese nicht ausreichen und Verteilungsgesetze werden benötigt.

Sollten auch diese zu schwierig sein, so helfen oft Grenzwertsätze bei der Näherung.

Sind bereits einige Daten und ein Modell vorhanden, können sich Fragen nach den Modellparametern, die aus den Daten zu schätzen sind, ergeben oder die Prüfung statistischer Hypothesen gewollt sein.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

In der Wahrscheinlichkeitsrechnung interessieren wir uns für Ausgänge von Zufallsexperimenten. Um sie beschreiben zu können, werden die verschiedenen Ausgänge als Mengen betrachtet. Diesen Mengen wird danach die Wahrscheinlichkeit ihres Eintretens als Zahl zugeordnet. Damit wir mit ihnen rechnen können, müssen wir zuerst Mengen selbst genauer betrachten.

2.1 Mengen

Im folgenden definieren/wiederholen wir einige Grundbegriffe der Mengenlehre und betrachten zugehörige Mengenoperationen.

An die Wahrscheinlichkeiten voraus denkend ist es manchmal sinnvoller, sich die gesuchten Ausgänge des Zufallsvorgangs aus anderen zusammenzustellen anstatt direkt vorzugehen.

Definition 2.1 (Menge): Eine *Menge* ist eine Zusammenfassung verschiedener Objekte zu einem Ganzen. Die einzelnen Objekte werden *Elemente* genannt.

Mengen werden entweder durch eine Auflistung ihrer Elemente oder durch eine definierende Eigenschaft angegeben.

Beispiel 2.2: Sei A eine Menge der Zahlen auf einem Würfel. Wir können A als

$$A = \{1, 2, 3, 4, 5, 6\} \quad \text{oder} \quad A = \{n : n \text{ ist eine natürliche Zahl mit } 1 \leq n \leq 6\}$$

schreiben.

Notation: Die *leere* Menge wird mit \emptyset dargestellt.

Definition 2.3 (Grundbegriffe der Mengenlehre): (i) Die Eigenschaft x ist ein Element von A wird mit $x \in A$ notiert, das Gegenteil mit $x \notin A$.

(ii) Für $A \subset B$, d. h. A ist *Teilmenge* von B , gilt dass für alle $x \in A$ auch $x \in B$ gilt.

(iii) Die *Schnittmenge* $A \cap B$ ist die Menge aller Elemente, die sowohl in A als auch in B sind. Für alle $x \in A \cap B$ gilt $x \in A$ und $x \in B$.

(iv) Die *Vereinigungsmenge* $A \cup B$ ist die Menge aller Elemente, die in A oder in B sind. Für alle $x \in A \cup B$ gilt $x \in A$ oder $x \in B$.

(v) Die *Differenzmenge* $A \setminus B$ ist die Menge aller Elemente, die in A aber nicht in B sind. Für alle $x \in A \setminus B$ gilt $x \in A$ und $x \notin B$.

(vi) Die *Komplementärmenge* A^c von $A \subset \Omega$ wird bezüglich einer Grundmenge Ω definiert und ist die Menge aller Elemente, die in Ω sind, aber nicht in A .

(vii) Die *Potenzmenge* $\mathcal{P}(A)$ ist die Menge aller Teilmengen von A . Für alle $M \in \mathcal{P}(A)$ gilt $M \subset A$.

(viii) Die *Mächtigkeit* von A bezeichnet die Anzahl der Elemente von A . Notiert wird sie mit $|A| = \#\{x : x \in A\}$.

(ix) Die *Gleichheit* von zwei Mengen A und B ist gegeben, wenn $A \subset B$ und $B \subset A$ gilt.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

(x) Zwei Mengen A und B sind *disjunkt*, wenn sie keine gemeinsamen Elemente haben, d. h. $A \cap B = \emptyset$.

(xi) Die Mengen A_1, A_2, \dots heißen *paarweise disjunkt*, falls für alle Paare i und j mit $i \neq j$ die Mengen A_i und A_j disjunkt sind.

Zur Darstellung eignen sich häufig Venn-Diagramme.

Beispiel 2.4: (a) Sei $\Omega = \{1, 3, 4\}$. Dann ist $\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{3\}, \{4\}, \{1, 3\}, \{1, 4\}, \{3, 4\}, \Omega\}$.

(b) Sei $\Omega = \{1, 2, 3, 4, 5\}$ und seien $A_1 = \{2, 4, 5\}$ und $A_2 = \{x \in \Omega : x \text{ ungerade}\}$. Dann gilt

$$A_1 \cup A_2 = \{1, 2, 3, 4, 5\} = \Omega \quad A_1 \cap A_2 = \{5\}$$

$$A_1 \setminus A_2 = \{2, 4\} \quad A_2 \setminus A_1 = \{1, 3\} \quad A_1^c = \{1, 3\}$$

(c) Seien $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$ und $C = \{1, 3, 5\}$. Dann gilt

$$A \cap B \cap C = \emptyset,$$

aber

$$A \cap B = \{2\} \quad A \cap C = \emptyset \quad B \cap C = \{1, 3\},$$

die Mengen sind somit nicht paarweise disjunkt, obwohl der Gesamtschnitt leer ist.

Definition 2.5 (Symmetrische Differenz): Die *symmetrische Differenz* zweier Mengen A_1 und A_2 ist definiert als

$$A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

Um kompliziertere Mengen und Mengenbeziehungen beschreiben zu können, benötigen wir noch einige Rechenregeln.

Satz 2.6 (Rechenregeln für Mengen): Seien A, B und C Mengen.

(i) *Eindeutigkeitsgesetze*

$$A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset \quad \text{und} \quad A \cup \Omega = \Omega, \quad A \cap \Omega = A$$

(ii) *Kommutativgesetze*

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

(iii) *Assoziativgesetze*

$$(A \cap B) \cap C = A \cap (B \cap C) \quad \text{und} \quad (A \cup B) \cup C = A \cup (B \cup C)$$

(iv) *Distributivgesetze*

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad \text{und} \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

(v) *De Morgansche Regeln*

$$(A \cap B)^c = A^c \cup B^c \quad \text{und} \quad (A \cup B)^c = A^c \cap B^c$$

(vi) Aus $A \subset B$ folgt für die Inklusion bei Negation $B^c \subset A^c$.

(vii) Die Differenzmenge $A \setminus B$ lässt sich als Schnitt darstellen durch $A \cap B^c$.

2.2 Mengensysteme

Erneut an die Wahrscheinlichkeiten denkend, werden wir benötigen, dass betrachtete Familien von Mengen bezüglich der Operationen \cap , \cup und \setminus abgeschlossen sind.

Definition 2.7 (Algebra): Eine nichtleere Familie \mathcal{F} von Teilmengen von Ω heißt *Algebra*, wenn
 (A1) aus $A \in \mathcal{F}$ folgt $A^c \in \mathcal{F}$,
 (A2) aus $A_1, A_2 \in \mathcal{F}$ folgt $A_1 \cup A_2 \in \mathcal{F}$
 gilt.

Beispiel 2.8: (a) Sei Ω gegeben. Dann ist $\mathcal{F} = \{\emptyset, \Omega\}$ eine Algebra.
 (b) Sei $\Omega = \{1, 2, 3, 4, 5\}$ und $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4, 5\}, \Omega\}$. Dann ist \mathcal{F} eine Algebra. Allgemein ist zu einer Menge $A \subset \Omega$ das Mengensystem $\{\emptyset, A, A^c, \Omega\}$ eine Algebra.
 (c) Sei $\Omega = \{0, 1\}^{\mathbb{N}}$. Dann ist $\mathcal{P}(\Omega)$ eine Algebra. Allgemein ist die Potenzmenge eine Algebra.

Lemma 2.9 (Eigenschaften einer Algebra): Sei \mathcal{F} eine Algebra und $A_1, A_2, \dots, A_n \in \mathcal{F}$. Dann gilt

- (i) $\emptyset, \Omega \in \mathcal{F}$,
- (ii) $A_1 \cap A_2 \in \mathcal{F}$,
- (iii) $A_1 \setminus A_2 \in \mathcal{F}$,
- (iv) $\bigcup_{i=1}^n A_i \in \mathcal{F}$ und $\bigcap_{i=1}^n A_i \in \mathcal{F}$.

Beispiel 2.10: Wir betrachten das Beispiel 2.8 (c) erneut, sei $A_1 = \{1, 2\}$ und $A_2 = \{3, 4, 5\}$. Dann ist auch $A_1 \cap A_2 = \emptyset \in \mathcal{F}$.

Um Grenzwerte von Mengen bilden zu können, benötigen wir ein Mengensystem \mathcal{F} , das nicht nur bezüglich Vereinigung und Durchschnitt *endlich* vieler Mengen, sondern bezüglich Vereinigung und Durchschnitt von *abzählbar unendlich* vielen Mengen abgeschlossen ist. Dazu nimmt man die folgende Bedingung hinzu.

Definition 2.11 (σ -Algebra): Sei \mathcal{F} eine Algebra.

(A3) Aus $A_1, A_2, \dots \in \mathcal{F}$ folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Beispiel 2.12:

Sei Ω gegeben. Dann ist $\mathcal{P}(\Omega)$ eine σ -Algebra.

Lemma 2.13: Sei \mathcal{F} eine σ -Algebra und $A_1, A_2, \dots \in \mathcal{F}$. Dann gilt $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

2.3 Zufallereignisse

Die Ausgänge, die bei einem Zufallsexperiment eintreten können, werden als Mengen modelliert. Sei E ein Grundraum.

Definition 2.14 (Ergebnisse): Die Menge $\Omega \subset E$ aller möglichen Ausgänge eines Zufallsexperiments nennen wir *Ergebnisraum* oder *Stichprobenraum*. Die Elemente $\omega \in \Omega$ heißen *Ergebnisse* oder *Elementarereignisse*. Eine Vereinigung von Elementarereignissen ist ein *Ereignis*.

Beispiel 2.15: (a) *Bernoulli-Experiment:* Einmaliger Münzwurf. $\Omega = \{\text{Kopf, Zahl}\}$ oder $E = \mathbb{N}$ und $\Omega = \{0, 1\}$.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

- (b) n -maliger Würfelwurf. $E = \mathbb{N}^n$ und $\Omega = \{1, 2, 3, 4, 5, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, 2, 3, 4, 5, 6\} \text{ für } 1 \leq i \leq n\}$. Zu beachten ist, dass hier nicht jeweils ein ω_i das Ergebnis ist, sondern $\omega = (\omega_1, \dots, \omega_n)$.
- (c) Die Strahlenwerte in Fukushima sollen zehn Jahre lang gemessen werden. Das benutzte Messgerät misst sehr häufig, so dass die Zeit als stetig angenommen werden kann. Dann ist

$$\Omega = \{f(t) : 0 \leq t \leq 3653, f \in C([0, 3653])\},$$

wenn $t = 0$ genau dem 11. März 2011 um 15:41 entspricht und eine Zeiteinheit einem Tag.

Die letzten zwei Beispiele zeigen, dass als Ergebnis eines Zufallsexperiments auch Folgen und Funktionen auftreten können und Ω deswegen nicht nur endlich sein kann, sondern auch unendlich abzählbar oder sogar überabzählbar.

Beispiel 2.16: (a) Zum Beispiel 2.15 (d) ist *Die ersten 3 Würfe ergeben Zahl* ein Ereignis, $A_2 = \{(\omega_i)_{1 \leq i \leq n} : \omega_i = 1 \text{ für } 1 \leq i \leq 3 \text{ und } \omega_i \in \{0, 1\} \text{ für } 4 \leq i \leq n\}$.

(b) Zum Beispiel 2.15 (e) ist *Der gemessene Wert bleibt unter 700 mSv* ein Ereignis,

$$A_3 = \left\{ f(t) : 0 \leq t \leq 3653, f \in C([0, 3653]) \text{ und } \sup_t f(t) < 700 \right\}.$$

Mit dieser Modellierung ist es klar, wieso wir vorher die Mengensysteme definiert haben. Bevor wir uns der Wahrscheinlichkeit zuwenden, schauen wir uns die Formulierung der Ereignisse als Mengen noch etwas genauer an.

Anmerkung (Terminologie): (a) Wird bei einem Zufallsexperiment das Ergebnis ω erzielt, so sagen wir: Das Ereignis A tritt ein, falls $\omega \in A$.

(b) Interessiert man sich für das Ereignis A^c , so sagt man, dass A nicht eintritt.

(c) Ist $A = \emptyset$, so wird A das *unmögliche Ereignis* genannt.

(d) Ist $A = \Omega$, so wird A das *sichere* oder *wahre Ereignis* genannt.

(e) Gilt $A \subset B$, so folgt aus dem Eintreten von A auch, dass B eintritt.

2.4 Wahrscheinlichkeiten

Den auf diese Weise definierten Ereignissen wollen wir jetzt eine Kennzahl zuordnen, die wir Wahrscheinlichkeit nennen. Eine solche Zahl sollte nicht-negativ sein. Sie sollte normiert sein, damit wir die Ereignisse bewerten können. Außerdem wollen wir Ereignisse sinnvoll verbinden können.

2.4.1 Axiome von Kolmogorov

Die oben genannten Forderungen wurden von Kolmogorov in drei Axiomen festgehalten, die für den allgemeinen Fall wie folgt aussehen:

Definition 2.17 (Kolmogorov-Axiome): Ein *Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbb{P})$ ist ein Tripel bestehend aus einer Menge Ω , einer σ -Algebra \mathcal{F} und einer Funktion $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$, *Wahrscheinlichkeitsmaß* genannt, mit den Eigenschaften

- (i) *Nichtnegativität* $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{F}$.
- (ii) *Normiertheit* $\mathbb{P}(\Omega) = 1$.

(iii) σ -Additivität Sind die Mengen $A_1, A_2, \dots \in \mathcal{F}$ paarweise disjunkt, so gilt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Anmerkung: Im abzählbar unendlichen Fall kann man die geforderte σ -Algebra einfach durch die Potenzmenge $\mathcal{P}(\Omega)$ ersetzen. Im endlichen Fall kann man das Axiom der σ -Additivität vereinfachend noch durch die Additivität ersetzen, da sich jede unendliche Vereinigung in eine endliche Vereinigung umschreiben lässt.

Aus diesen Axiomen lassen sich nun Rechenregeln ableiten.

Satz 2.18 (Rechenregeln für Wahrscheinlichkeiten): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, A_1, A_2, \dots \in \mathcal{F}$. Dann gilt

- (i) $\mathbb{P}(\emptyset) = 0$,
- (ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- (iii) Aus $A_1 \subset A_2$ folgt $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$,
- (iv) $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$,
- (v) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.

BEWEIS (i) Wir wissen, dass $\Omega = \Omega \cup \emptyset$ und $\Omega \cap \emptyset = \emptyset$. Dann ist

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset),$$

und damit $\mathbb{P}(\emptyset) = 0$.

(ii) Auch hier wissen wir, dass $A \cap A^c = \emptyset$ und $A \cup A^c = \Omega$ und erhalten

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$$

(iii) Da $A_1 \subset A_2$ ist, lässt sich A_2 schreiben als $A_2 = (A_2 \setminus A_1) \cup A_1$, wobei $A_2 \setminus A_1$ und A_1 disjunkt sind und damit folgt

$$\mathbb{P}(A_2) = \mathbb{P}((A_2 \setminus A_1) \cup A_1) = \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_1) \geq \mathbb{P}(A_1)$$

(iv) Wegen $A_1 \cup A_2 = (A_1 \setminus (A_1 \cap A_2)) \cup (A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2))$, wobei die drei Mengen disjunkt sind, gilt

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1 \setminus (A_1 \cap A_2)) + \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_2 \setminus (A_1 \cap A_2)) \\ &= (\mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2)) + \mathbb{P}(A_1 \cap A_2) + (\mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \end{aligned}$$

mit den Erkenntnissen aus (iii).

(v) Dies folgt durch Induktion aus (iii). □

Eine Verallgemeinerung des Satzes 2.18 (iv) ist der *Allgemeine Additionssatz*, auch *Siebformel von Poincaré-Sylvester* genannt, bei dem für n Mengen alle k -weisen Schnittmengen (Paare, Tripel, usw.) betrachtet und entsprechend addiert oder subtrahiert werden.

Satz 2.19 (Siebformel von Poincaré-Sylvester): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots, A_n \in \mathcal{F}$. Dann gilt

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} \mathbb{P}(A_{k_1} \cap \dots \cap A_{k_i}) \quad (2.1)$$

BEWEIS durch Induktion mit Hilfe des Satzes 2.18 (iv). □

2.4.2 Laplacescher Wahrscheinlichkeitsraum

Sei Ω endlich, also $|\Omega| < \infty$. Dann gilt natürlich auch für alle $A \subset \Omega$, dass $|A| < \infty$. Dann heißt ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ ein *endlicher Wahrscheinlichkeitsraum*.

Definition 2.20 (Laplacescher Wahrscheinlichkeitsraum): Sei Ω endlich und $\mathcal{F} = \mathcal{P}(\Omega)$. Ein endlicher Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, bei dem alle Elementarereignisse gleichwahrscheinlich sind, d. h. $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$, heißt *Laplacescher Wahrscheinlichkeitsraum*.

Anmerkung: Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Laplacescher Wahrscheinlichkeitsraum. Für alle $A \subset \Omega$ gilt dann $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ wegen der σ -Additivität von Wahrscheinlichkeitsmaßen. Die so definierte Wahrscheinlichkeit heißt *Laplace-Wahrscheinlichkeit*.

Beispiel 2.21 (Zweimaliges Würfeln): Als erstes muss man sich über Ω klar werden:

$$\Omega = \{\omega = (i, j) : 1 \leq i, j \leq 6, i = \text{Augenzahl beim 1. Wurf}, j = \text{Augenzahl beim 2. Wurf}\}$$

Dann ist $|\Omega| = 36$ und wir nehmen $\mathcal{F} = \mathcal{P}(\Omega)$. Sei $\mathbb{P} : \Omega \mapsto [0, 1]$ ein Laplace-Wahrscheinlichkeitsmaß, d. h. für alle $\omega_1, \omega_2 \in \Omega$ ist $\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\})$ und $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$, somit $\mathbb{P}(\{\omega\}) = \frac{1}{36}$ für alle $\omega \in \Omega$. Das Tripel $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ bildet einen Laplaceschen Wahrscheinlichkeitsraum.

Sei nun $A = \{\text{Gesamtaugenzahl} \geq 10\} = \{(4, 6), (5, 6), (6, 6), (5, 5), (6, 5), (6, 4)\}$. Dann ist $|A| = 6$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{6}$.

2.4.3 Kombinatorische Wahrscheinlichkeiten

Oft lässt sich die Größe des Laplaceschen Wahrscheinlichkeitsraum wie auch diejenige eines gesuchten Ereignisses durch kombinatorische Überlegungen herausfinden. Dazu benötigen wir noch einige Begriffe.

Definition 2.22 (Fakultät): Für eine natürliche Zahl n definieren wir *n Fakultät* als

$$n! = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1,$$

weiterhin ist $0! = 1$ (als leeres Produkt).

Definition 2.23 (Binomialkoeffizient): Für zwei natürliche Zahlen n_1 und n_2 mit $n_1 > n_2$ definieren wir den *Binomialkoeffizienten* als

$$\binom{n_1}{n_2} := \frac{n_1!}{n_2! \cdot (n_1 - n_2)!}.$$

Lemma 2.24 (Multiplikationsregel): Werden n Zufallsexperimente ausgeführt und hat das Experiment k immer (ohne Beachtung der Ergebnisse in den vorherigen Experimenten) n_k mögliche Ergebnisse, so ist die Gesamtanzahl der Ergebnisse $n_1 \cdot n_2 \cdot \dots \cdot n_m$.

Damit kommen wir jetzt zu den Urnenmodellen.

Definition und Lemma 2.25 (Urnenmodell): In einer Urne liegen N durchnummerierte Kugeln. Es werden n Kugeln zufällig gezogen. Die Ziehung kann mit oder ohne Zurücklegen und mit oder ohne Beachtung der Reihenfolge geschehen.

(i) Ziehen mit Reihenfolge und mit Zurücklegen: Damit ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\}\}$$

mit der Mächtigkeit $|\Omega| = N^n$.

(ii) Ziehen mit Reihenfolge und ohne Zurücklegen: Hier ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ mit } \omega_i \neq \omega_j \text{ für } i \neq j\}$$

mit der Mächtigkeit $|\Omega| = \frac{N!}{(N-n)!}$.

(iii) Ziehen ohne Reihenfolge und ohne Zurücklegen: Der Ergebnisraum ist

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ und } \omega_1 < \omega_2 < \dots < \omega_n\}$$

mit der Mächtigkeit $|\Omega| = \frac{N!}{n!(N-n)!} = \binom{N}{n}$.

(iv) Ziehen ohne Reihenfolge und mit Zurücklegen: In diesem Fall ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ und } \omega_1 \leq \omega_2 \leq \dots \leq \omega_n\}$$

mit der Mächtigkeit $|\Omega| = \frac{(N+n-1)!}{n!(N-1)!} = \binom{N+n-1}{n}$.

Beispiel 2.26: (a) Ein Ruderverein hat 23 (ein Vorsitzender und 22 weitere) Mitglieder und plant sein jährliches Sommerfest. Dabei wird der Vorsitzende von 5 Mitgliedern einmal über den See gerudert. Der Verein könnte $\binom{22}{5} = 26334$ verschiedene Rudermannschaften aufstellen. Wir haben hier den Fall ohne Reihenfolge und ohne Zurücklegen.

(b) Bei einer Tombola machen 87 Personen mit und es gibt 10 Preise zu gewinnen. Für den ersten Preis wird der Name des Gewinners aus einer Urne gezogen, dann wird aus der gleichen Urne der Name des Gewinners des zweiten Preises gezogen usw. bis alle 10 Preise vergeben sind. Wir haben $\frac{87!}{77!}$ mögliche Gewinner-10-tupel. Unter den Teilnehmern ist das Ehepaar Müller. Wie groß ist die Wahrscheinlichkeit, dass beide Müllers unter den Gewinnern sind? Wir haben $|\Omega| = \frac{87!}{77!}$ und betrachten wieder das Gegenereignis. Seien

$$\begin{aligned} A &= \{\text{kein Herr Müller}\} & \text{mit} & \quad |A| = \frac{86!}{76!} \\ B &= \{\text{keine Frau Müller}\} & \text{mit} & \quad |A| = \frac{86!}{76!} \\ C &= \{\text{gar kein Müller}\} & \text{mit} & \quad |A| = \frac{85!}{75!} \end{aligned}$$

Dann ist

$$\begin{aligned} \mathbb{P}(\text{beide Müller}) &= 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(C)) \\ &= 1 - \frac{\frac{86!}{76!} + \frac{86!}{76!} - \frac{85!}{75!}}{\frac{87!}{77!}} \approx 0,012 = 1,2\% \end{aligned}$$

(c) In einer Gruppe sind 30 Personen. Wie groß ist die Wahrscheinlichkeit, dass zwei Personen mit dem Geburtstag am gleichen Tag darunter sind.

Zuerst wird jeder Person ein Geburtstag zugeordnet. Wir ziehen also mit Reihenfolge und mit Zurücklegen, da jede Person eindeutig einen Geburtstag bekommt und diese sich auch wiederholen können.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

Mathematisch formuliert ist $N = 365$ die Anzahl der möglichen Geburtstage und $n = 30$ die Anzahl der Leute. Dann ist $\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\}\}$ und $|\Omega| = N^n$. Das Ereignis

$$A_n = \{\text{mindestens 2 Personen haben am gleichen Tag Geburtstag}\}$$

lässt sich schreiben als

$$= \{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \text{es gibt } 1 \leq i, j \leq n \text{ mit } i \neq j : \omega_i = \omega_j\}.$$

Zur Berechnung der Wahrscheinlichkeit wählen wir den Ansatz über das Gegenereignis $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c)$, d. h. wir interessieren uns für das Ereignis unter Beachtung der Reihenfolge, da wieder jede einzelne Person betrachtet wird, und ohne Zurücklegen, weil wir mehrfache Geburtstage nicht zulassen und haben

$$A_n^c = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_i \neq \omega_j \text{ für alle } 1 \leq i, j \leq n \text{ mit } i \neq j\}$$

Die Mächtigkeit des Ereignisses ist dann

$$|A_n^c| = \underbrace{N}_{\text{frei Wahl}} \cdot \underbrace{(N-1)}_{\text{nicht am gewählten Tag}} \cdot \underbrace{(N-2)}_{\text{nicht an den 2 gewählten Tagen}} \cdot \dots \cdot \underbrace{(N-(n-1))}_{\text{nicht an den } n-1 \text{ gewählten Tagen}}$$

und somit

$$\mathbb{P}(A_n^c) = \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{N^n}$$

Für die obige Gruppe von oben folgt $\mathbb{P}(A_{30}) = 1 - \mathbb{P}(A_{30}^c) \approx 0,7063162$, also in etwa 71% der Fälle sind zwei Personen mit Geburtstag am gleichen Tag zu finden. Diese Fragestellung ist unter dem Begriff *Geburtstagsproblem* oder *Geburtstagsparadoxon* bekannt.

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

Wir wollen uns nicht nur mit den Ereignissen selbst beschäftigen, sondern auch ihre Zusammenhänge beschreiben. Als erstes schauen wir uns den Fall an, dass es keine Verbindung zwischen ihnen gibt.

Wir setzen den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ voraus.

2.5.1 Unabhängige Ereignisse

Wir nennen zwei Ereignisse A und B unabhängig voneinander, wenn das Eintreten von A die Wahrscheinlichkeit des Eintretens von B nicht verändert.

Definition 2.27 (Unabhängigkeit zweier Ereignisse): Zwei Ereignisse $A, B \in \mathcal{F}$ heißen *unabhängig*, wenn gilt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B). \quad (2.2)$$

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

Wir sehen, dass diese Eigenschaft symmetrisch ist.

Anmerkung: Beim Laplaceschen Wahrscheinlichkeitsraum sehen wir zur Interpretation, dass sich die relative Häufigkeit nicht ändert, wenn wir die Bezugsmenge von Ω zu A ändern

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

das Ereignis B in der Menge A anteilig also genauso oft auftritt, wie in der Grundgesamtheit Ω .

Lemma 2.28: Das sichere und das unmögliche Ereignis sind von allen anderen unabhängig, insbesondere sind sie von sich selbst unabhängig. Umgekehrt gilt für ein Ereignis, welches von sich selbst unabhängig ist, dass seine Wahrscheinlichkeit entweder 0 oder 1 ist.

BEWEIS Sei $A \in \mathcal{F}$. Dann gilt

$$\begin{aligned} x\mathbb{P}(A \cap \Omega) &= \mathbb{P}(A) = \mathbb{P}(A) \cdot 1 = \mathbb{P}(A) \cdot \mathbb{P}(\Omega), \\ \mathbb{P}(A \cap \emptyset) &= \mathbb{P}(\emptyset) = 0 = 0 \cdot \mathbb{P}(A) = \mathbb{P}(\emptyset) \cdot \mathbb{P}(A), \end{aligned}$$

wobei A keinerlei Einschränkung unterworfen war, insbesondere gilt die Argumentation für $A = \Omega$ und $A = \emptyset$. Sei nun $B \in \mathcal{F}$ von sich selbst unabhängig, also

$$\mathbb{P}(B) = \mathbb{P}(B \cap B) = \mathbb{P}(B) \cdot \mathbb{P}(B)$$

falls $\mathbb{P}(B) \neq 0$, können wir beide Seiten durch $\mathbb{P}(B)$ teilen

$$1 = \mathbb{P}(B),$$

andernfalls ist gerade $\mathbb{P}(B) = 0$. □

Wir können die Definition auch auf Familien von Ereignissen ausweiten.

Definition 2.29: Seien $A_1, \dots, A_n \in \mathcal{F}$. Die Ereignisse A_1, \dots, A_n heißen unabhängig, wenn für jedes $1 \leq k \leq n$ und jede Auswahl $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ gilt

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}). \quad (2.3)$$

Anmerkung: (a) Wir müssen die Eigenschaft der Unabhängigkeit für jede Auswahl der Mengen der Familie fordern, denn sonst wäre jede Familie unabhängig, wenn ein $A_i = \emptyset$ für $1 \leq i \leq n$ wäre.

(b) Betrachten wir das Beispiel des 2-maligen Wurfs mit einer fairen Münze, also

$$\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$$

und dabei die Ereignisse

$$\begin{aligned} A &= \{\text{beim ersten Wurf Kopf}\} = \{(0,0), (0,1)\} \\ B &= \{\text{beim zweiten Wurf Kopf}\} = \{(0,0), (1,0)\} \\ C &= \{\text{Anzahl der Würfe mit Kopf ist gerade}\} = \{(0,0), (1,1)\} \end{aligned}$$

Dann gilt

$$\begin{aligned} \mathbb{P}(A) &= \frac{1}{2} & \mathbb{P}(A \cap B) &= \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B) \\ \mathbb{P}(B) &= \frac{1}{2} & \mathbb{P}(A \cap C) &= \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(C) \\ \mathbb{P}(C) &= \frac{1}{2} & \mathbb{P}(B \cap C) &= \frac{1}{4} = \mathbb{P}(B) \cdot \mathbb{P}(C) \end{aligned}$$

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

aber

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

Deswegen ist paarweise Unabhängigkeit nicht ausreichend für die Unabhängigkeit einer Familie von Mengen.

2.5.2 Bedingte Wahrscheinlichkeit

Beispiel 2.30: (a) Eine Familie mit zwei Kindern hat mindestens einen Jungen. Was ist die Wahrscheinlichkeit, dass sie zwei Jungen haben?

Alle Möglichkeiten mit mindestens einem Jungen sind $(M, J), (J, M), (J, J)$. Von diesen drei ist eine die gesuchte, d. h. die Antwort ist $\frac{1}{3}$.

(b) Das jüngere Kind einer Familie mit zwei Kindern ist ein Junge. Was ist die Wahrscheinlichkeit, dass sie zwei Jungen haben?

Hier sind die Möglichkeiten $(J, M), (J, J)$, also nur zwei und damit ist die Antwort $\frac{1}{2}$.

Definition 2.31 (Bedingte Wahrscheinlichkeit): Seien A und B zwei Ereignisse mit $\mathbb{P}(A) > 0$. Dann ist

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad (2.4)$$

die *bedingte Wahrscheinlichkeit* von B gegeben A .

Anmerkung (bedingte Wahrscheinlichkeit ist ein Wahrscheinlichkeitsmaß): Sei $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$. Dann gilt, dass die Abbildung

$$\begin{aligned} \mathbb{P}(\cdot|A) : \mathcal{F} &\rightarrow [0, 1] \\ B &\mapsto \mathbb{P}(B|A) \end{aligned}$$

ein Wahrscheinlichkeitsmaß ist. Dazu überprüfen wir die Axiome von Kolmogorov.

(i) $0 \leq \mathbb{P}(B|A) \leq 1$ wegen $0 \leq \mathbb{P}(B \cap A) \leq \mathbb{P}(A)$,

(ii) $\mathbb{P}(\Omega|A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)}$,

(iii) Seien B_1, B_2, \dots paarweise disjunkt, so sind auch $B_i \cap A$ paarweise disjunkt. Damit gilt

$$\left(\bigcup_{i=1}^{\infty} B_i \right) \cap A = \bigcup_{i=1}^{\infty} (B_i \cap A)$$

und wir haben

$$\mathbb{P} \left(\bigcup_{i=1}^{\infty} B_i | A \right) = \frac{\mathbb{P} \left(\bigcup_{i=1}^{\infty} (B_i \cap A) \right)}{\mathbb{P}(A)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \sum_{i=1}^{\infty} \mathbb{P}(B_i | A). \quad (2.5)$$

Lemma 2.32: Sind zwei Ereignisse A und B unabhängig, so ist $\mathbb{P}(B|A) = \mathbb{P}(B)$

BEWEIS Da $\mathbb{P}(A) > 0$ ist, gilt

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(B). \quad \square$$

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

Beispiel 2.33: Wir betrachten zweimaliges Ziehen ohne Zurücklegen aus einer Urne mit N Kugeln, R roten und $N - R$ blauen, und definieren die Ereignisse

$$A = \{\omega = (\omega_1, \omega_2) : \omega_1 = \text{rot}\},$$

$$B = \{\omega = (\omega_1, \omega_2) : \omega_2 = \text{rot}\}$$

Wir kennen $\mathbb{P}(A) = \frac{R}{N}$ und $\mathbb{P}(A \cap B) = \frac{R(R-1)}{N(N-1)}$ und somit $\mathbb{P}(B|A) = \frac{R-1}{N-1}$.

Es folgen drei wichtige Aussagen über bedingte Wahrscheinlichkeiten.

Definition 2.34 (Partition): Die Mengenfamilie $B_1, \dots, B_n \subset \Omega$ ist eine Partition von Ω , falls $B_1 \cup \dots \cup B_n = \Omega$ und $B_i \cap B_j = \emptyset$ für alle $1 \leq i, j \leq n$ mit $i \neq j$ gilt.

Satz 2.35 (Satz von der totalen Wahrscheinlichkeit): Es seien $B_1, \dots, B_n \in \mathcal{F}$ eine Partition von Ω und gelte $\mathbb{P}(B_i) > 0$ für alle $0 \leq i \leq n$. Dann gilt

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i) \quad (2.6)$$

für alle $A \in \mathcal{F}$.

BEWEIS Definiere $C_i = A \cap B_i$ für $1 \leq i \leq n$. Dann sind die Mengen C_1, \dots, C_n disjunkt und $A = C_1 \cup \dots \cup C_n$ und es gilt

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(C_i) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i). \quad \square$$

Der Satz von der totalen Wahrscheinlichkeit wird vor allem zur Modellierung mehrstufiger Experimente verwendet.

Beispiel 2.36 (Zweistufiges Experiment): Betrachte 10 Urnen mit jeweils 9 Kugeln. Die erste Urne enthält 9 rote Kugeln, die zweite 8 rote und eine blaue, usw., die zehnte Urne enthält 9 blaue Kugeln. Es wird zufällig eine Urne ausgewählt. Anschließend wird zweimalig aus dieser Urne mit Zurücklegen gezogen. Dann ist

$$\Omega = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 \in \{\text{Urne } 1, \dots, \text{Urne } 10\}, \omega_2, \omega_3 \in \{\text{rot, blau}\}\}$$

die betrachtete Grundgesamtheit. Wir definieren die Ereignisse

$$B_i = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 = \text{Urne } i\},$$

$$A_1 = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_2 = \text{rot}\},$$

$$A_2 = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_3 = \text{rot}\}$$

und interessieren uns für $\mathbb{P}(A_1)$ und $\mathbb{P}(A_2|A_1)$. Wir kennen $\mathbb{P}(B_i) = \frac{1}{10}$ und $\mathbb{P}(A_1|B_i) = \frac{10-i}{9}$. Damit berechnen wir

$$\mathbb{P}(A_1) = \sum_{i=1}^{10} \mathbb{P}(A_1|B_i)\mathbb{P}(B_i) = \frac{1}{10} \sum_{i=1}^{10} \frac{10-i}{9} = \frac{1}{90} \sum_{i=0}^9 i = \frac{1}{2}.$$

Weiter gilt

$$\begin{aligned} \mathbb{P}(A_2|A_1) &= \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = 2 \cdot \mathbb{P}(A_1 \cap A_2) = 2 \cdot \sum_{i=1}^{10} \mathbb{P}(A_1 \cap A_2|B_i)\mathbb{P}(B_i) \\ &= \frac{2}{10} \sum_{i=1}^{10} \left(\frac{10-i}{9}\right)^2 = \frac{2}{10 \cdot 9 \cdot 9} \sum_{i=0}^9 i^2 = \frac{2}{10 \cdot 9 \cdot 9} \cdot \frac{9 \cdot 10 \cdot 19}{6} = \frac{19}{27}. \end{aligned}$$

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

Satz 2.37 (Bayes-Formel): Es seien $B_1, \dots, B_n \in \mathcal{F}$ eine Partition von Ω und es gelte $\mathbb{P}(B_i) > 0$ für alle $1 \leq i \leq n$. Dann gilt

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)} \quad (2.7)$$

für alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$ und alle $1 \leq k \leq n$.

BEWEIS Mit zweimaliger Anwendung der Definition der bedingten Wahrscheinlichkeit wissen wir, dass

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)},$$

wobei die letzte Gleichheit aus dem Satz 2.35 folgt. \square

Beispiel 2.38: Wir betrachten das Modell aus dem Beispiel 2.36 und interessieren uns für $\mathbb{P}(B_k|A_1)$:

$$\mathbb{P}(B_k|A_1) = \frac{\mathbb{P}(A_1|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{10} \mathbb{P}(A_1|B_i)\mathbb{P}(B_i)} = \frac{\frac{10-k}{9} \frac{1}{10}}{\sum_{i=1}^{10} \frac{10-i}{9} \frac{1}{10}} = \frac{10-k}{\sum_{i=0}^9 i} = \frac{10-k}{45}$$

Beispiel 2.39 (Mammografie):] Etwa 1% aller Frauen zwischen 40 und 50 haben Brustkrebs. Eine Frau mit Brustkrebs bekommt mit 90% Wahrscheinlichkeit ein positives Ergebnis, eine Frau ohne Brustkrebs erhält mit 10% Wahrscheinlichkeit ein fälschlicherweise positives Ergebnis. Wir definieren die Ereignisse

$$\begin{aligned} A &= \{\text{Frau hat Brustkrebs}\} \\ B &= \{\text{positives Testergebnis}\} \end{aligned}$$

Wir interessieren uns für $\mathbb{P}(A|B)$:

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)} \\ &= \frac{0,9 \cdot 0,01}{0,9 \cdot 0,01 + 0,1 \cdot 0,99} = \frac{0,009}{0,009 + 0,099} = \frac{9}{108} \approx 8,3\%, \end{aligned}$$

d. h. von allen Personen, bei denen der Test positiv ausfällt, sind nur 8.3% erkrankt.

Beispiel 2.40: Wir betrachten die Zinsentwicklung einer Anlageform innerhalb eines vorgegebenen Zeitraumes und definieren die Ereignisse

$$\begin{aligned} A_1 &= \{\text{Der Zins fällt um } 0,5\%\} \\ A_2 &= \{\text{Der Zins bleibt unverändert}\} \\ A_3 &= \{\text{Der Zins steigt um } 0,5\%\} \end{aligned}$$

der Einfachheit halber sei $\Omega = A_1 \cup A_2 \cup A_3$ und

$$B = \{\text{Anlageberater sagt, der Zins steige}\}$$

Wir kennen die folgenden Einschätzungen

$$\mathbb{P}(A_1) = 0,1 \qquad \mathbb{P}(A_2) = 0,6 \qquad \mathbb{P}(A_3) = 0,3$$

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

und die Erfahrungswerte

$$\mathbb{P}(B|A_1) = 0,15 \qquad \mathbb{P}(B|A_2) = 0,3 \qquad \mathbb{P}(B|A_3) = 0,75$$

Wir interessieren uns für $\mathbb{P}(A_1|B)$, $\mathbb{P}(A_2|B)$ und $\mathbb{P}(A_3|B)$.

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \mathbb{P}(B|A_3)\mathbb{P}(A_3) \\ &= 0,15 \cdot 0,1 + 0,3 \cdot 0,6 + 0,75 \cdot 0,3 = 0,42 \end{aligned}$$

und damit

$$\mathbb{P}(A_1|B) = \frac{0,15 \cdot 0,1}{0,42} \approx 0,036 \qquad \mathbb{P}(A_2|B) \approx 0,429 \qquad \mathbb{P}(A_3|B) \approx 0,536$$

Unsere Unsicherheit bezüglich des Eintretens des uns interessierenden Ereignisses A_3 ist also mit Hilfe der Information B gestiegen. (Wir haben noch keine Aussage, was passiert, wenn der Anlageberater etwas anderes vermutet.)

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

3 Zufallsvariablen und Verteilungen

Zufallsvariablen sind ein weiteres Instrument bei der Beschäftigung mit Zufallsexperimenten. Sie kommen immer dann zum Einsatz, wenn wir uns nicht für die Einzelheiten der Ergebnisse und Ereignisse interessieren, sondern nur eine Zusammenfassung brauchen.

3.1 Grundbegriffe

Eine Zufallsvariable ordnet jedem Elementarergebnis eine Zahl zu.

Beispiel 3.1: Ein Unternehmen hat drei neue Produkte A, B und C entwickelt, deren Erfolgswahrscheinlichkeiten auf 70, 90 und 60% eingeschätzt werden. Die Produkte selbst sind voneinander unabhängig und deren Erfolg wäre etwa gleich anzusetzen. Die Frage ist, mit wie vielen erfolgreichen Produkten zu rechnen ist. Wir können die Ereignisse

$$A = \{\text{Produkt } A \text{ ist erfolgreich}\}, B = \{\text{Produkt } B \text{ ist erfolgreich}\}, C = \{\text{Produkt } C \text{ ist erfolgreich}\}$$

definieren und betrachten die 8 Elementarereignisse:

Elementarereignis	Wahrscheinlichkeit	Anzahl der Erfolge	deren Wahrscheinlichkeit
$A^c \cap B^c \cap C^c$	0.012	0 Erfolge	0.012
$A \cap B^c \cap C^c$	0.028		
$A^c \cap B \cap C^c$	0.108	1 Erfolg	0.154
$A^c \cap B^c \cap C$	0.018		
$A \cap B \cap C^c$	0.252		
$A^c \cap B \cap C$	0.162	2 Erfolge	0.456
$A \cap B^c \cap C$	0.042		
$A \cap B \cap C$	0.378	3 Erfolge	0.378

Wir haben die Elementarereignisse zu interessierenden Ereignissen sortiert. Diese Sortierung, also die fragebezogene Zuordnung der Elementarereignisse zu den Zahlen 0, 1, 2, 3 ist eine Zufallsvariable. Diese ist eine Funktion auf den Elementarereignissen.

Anmerkung: Zur Definition der Zufallsvariable benötigen wir eine Regularitätsbedingung, die *Messbarkeit*. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt bezüglich der σ -Algebra \mathcal{F} *messbar*, wenn

$$\{\omega : X(\omega) \leq y\} \in \mathcal{F} \tag{3.1}$$

für alle $y \in \mathbb{R}$ gilt.

Definition 3.2 (Zufallsvariable): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine *Zufallsvariable* ist eine messbare Funktion $X : \Omega \rightarrow \mathbb{R}$.

3 Zufallsvariablen und Verteilungen

Anmerkung: Eine Zufallsvariable heißt zwar Variable, ist aber eine Funktion.

Beispiel 3.3: Die Zufallsvariable aus Beispiel 3.1 ist $X : \Omega \rightarrow \{0, 1, 2, 3\} \subset \mathbb{R}$.

Anmerkung: Oft interessiert man sich nicht für einen speziellen Wert, den X annimmt, sondern ob sie in einen Bereich fällt, also für $X \in B$, wobei $B \subset \mathbb{R}$ ist, z. B. $B = (-\infty, c]$ oder $B = [c_1, d_1] \cup [c_2, d_2]$. Deshalb betrachten wir auch im Bildraum \mathbb{R} von X eine σ -Algebra. Dazu nehmen wir die sogenannte *Borel- σ -Algebra*, die als die kleinste σ -Algebra von Teilmengen von \mathbb{R} , die alle offenen Intervalle von \mathbb{R} enthält, definiert ist, d. h.

$$\mathcal{B}(\mathbb{R}) = \sigma \left(\{(a, b) : -\infty < a < b < \infty\} \right)$$

Durch Bildung von Schnitten über unendlich viele Mengen enthält $\mathcal{B}(\mathbb{R})$ auch alle halboffenen und abgeschlossenen Intervalle, denn

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + 1/n) \in \mathcal{B}(\mathbb{R})$$

$$[a, b) = \bigcap_{n=1}^{\infty} (a - 1/n, b) \in \mathcal{B}(\mathbb{R})$$

$$[a, b] = \bigcap_{n=1}^{\infty} (a - 1/n, b + 1/n) \in \mathcal{B}(\mathbb{R})$$

und damit sind auch einelementige Mengen

$$\{c\} = \bigcap_{n=1}^{\infty} (c - 1/n, c + 1/n) \in \mathcal{B}(\mathbb{R})$$

enthalten. Damit können wir Gleichung (3.1) schreiben als

$$\{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{für alle } B \in \mathcal{B}(\mathbb{R}), \quad (3.2)$$

was äquivalent ist.

Notation: Bei der Definition von $\mathcal{B}(\mathbb{R})$ haben wir eine Notation verwendet, die der Erläuterung bedarf. Zu einer Grundmenge Ω und einer Menge von Mengen \mathcal{G} aus Ω , also für alle $G \in \mathcal{G}$ gilt $G \subset \Omega$ bezeichnet $\sigma(\mathcal{G})$ eine σ -Algebra, die von \mathcal{G} erzeugt wird. Sie enthält alle Mengen aus \mathcal{G} , d. h. für alle $G \in \mathcal{G}$ gilt $G \in \sigma(\mathcal{G})$ und darüber hinaus enthält sie alle Mengen, die sich durch Anwendung der Komplementbildung und (unendlicher) Vereinigung aus diesen zusammenstellen lassen. Damit erfüllt $\sigma(\mathcal{G})$ die Eigenschaften (A1), (A2) und (A3) aus den Definitionen 2.7 und 2.11 und ist eine σ -Algebra, und zwar, da bei dieser Konstruktion zusätzlich nichts „überflüssiges“ hinzukam, die kleinste σ -Algebra, welche \mathcal{G} enthält.

Eine alternative Interpretation ist, dass $\sigma(\mathcal{G})$ der Schnitt aller σ -Algebren ist, die \mathcal{G} enthalten. Damit fallen auch alle nicht „notwendigen“ Mengen heraus und sie ist die kleinste σ -Algebra, welche \mathcal{G} enthält.

Ein Spezialfall ist die von einer Menge erzeugte σ -Algebra. Für eine Menge $A \subset \Omega$ ist $\sigma(A) := \sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$ und damit die kleinste σ -Algebra, die A enthält.

Notation: Es ist üblich eine Zufallsvariable mit einem großen lateinischen Buchstaben zu bezeichnen. Der Wert $X(\omega)$ einer Zufallsvariablen X wird mit dem entsprechenden Kleinbuchstaben, hier x bezeichnet, und *Realisierung von X* genannt.

Beispiel 3.4 (Indikatorfunktion): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben und $A \in \mathcal{F}$ ein Ereignis. Die *Indikatorfunktion* ist definiert als

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A. \end{cases} \quad (3.3)$$

und wir betrachten die Zufallsvariable $X(\omega) = \mathbb{1}(\omega)$ und haben

$$\{\omega : X(\omega) \in B\} = \begin{cases} A, & \text{falls } 1 \in B \text{ und } 0 \notin B, \\ A^c, & \text{falls } 0 \in B \text{ und } 1 \notin B, \\ \Omega, & \text{falls } 0 \in B \text{ und } 1 \in B, \\ \emptyset, & \text{falls } 0 \notin B \text{ und } 1 \notin B \end{cases}$$

für alle $B \in \mathcal{B}(\mathbb{R})$.

Definition 3.5 (Verteilung): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable.

(i) Dann heißt die Funktion

$$F_X(y) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq y\})$$

für alle $y \in \mathbb{R}$ *Verteilungsfunktion* von X . Es gilt $F_X : \mathbb{R} \rightarrow [0, 1]$.

(ii) Die Funktion

$$P_X(B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

für alle Mengen B der σ -Algebra $\mathcal{B}(\mathbb{R})$ heißt *Verteilung* von X und es gilt $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$.

Satz 3.6 (Eigenschaften der Verteilungsfunktion): Sei X eine Zufallsvariable und F_X ihre Verteilungsfunktion. Dann gilt

- (i) *Asymptotisches Verhalten* $\lim_{y \rightarrow -\infty} F_X(y) = 0$ und $\lim_{y \rightarrow \infty} F_X(y) = 1$,
- (ii) *Rechtsstetigkeit* $\lim_{y \rightarrow z, y > z} F_X(y) = F_X(z)$ und $\lim_{y \rightarrow z, y < z} F_X(y)$ existiert,
- (iii) *Monotonie* $F_X(y) \leq F_X(z)$ für $y \leq z$.

Notation: Die folgenden Schreibweisen sind zur Vereinfachung üblich:

$$\begin{aligned} F_X(y) &= \mathbb{P}(X \leq y) \\ P_X(B) &= \mathbb{P}(X \in B) \end{aligned}$$

Anmerkung: Die Funktion P_X ist ein Wahrscheinlichkeitsmaß, auch *von X induziertes Wahrscheinlichkeitsmaß* genannt, und wir erhalten den Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

Anmerkung: Wir haben zuerst eine neue Grundmenge genommen, uns dann eine σ -Algebra definiert und im Anschluss ein Wahrscheinlichkeitsmaß auf demselben mit Hilfe von X gebastelt. Die Abbildung $\mathbb{P} \rightarrow P_X$ heißt auch *Maßtransport*.

Arten von Zufallsvariablen Wir unterscheiden zwischen *diskreten* und *stetigen* Zufallsvariablen. Diskret ist eine Zufallsvariable, wenn die Realisation nur einzelne bestimmte Werte eines Intervalls annehmen kann. Stetig ist sie hingegen, wenn im Prinzip jeder beliebige Wert eines Intervalls angenommen werden kann. Besteht die Realisation aus nur einer Größe, so nennen wir die Zufallsvariable *ein-dimensional*, setzt sie sich aus mehreren Zahlen zusammen, nennen wir sie *mehr-dimensional*.

Wir wollen zuerst diskrete Zufallsvariablen untersuchen.

3.2 Diskrete Zufallsvariablen

Dazu benötigen wir die folgende

3.2.1 Grundbegriffe

Definition 3.7 (Diskrete Zufallsvariable): Eine Zufallsvariable heißt *diskret*, wenn es eine endliche oder abzählbar unendliche Teilmenge $W \subset \mathbb{R}$ gibt mit $\mathbb{P}(X \in W) = 1$.

Im Abschnitt 2.4 haben wir diskrete Wahrscheinlichkeitsverteilungen definiert. Eine Zufallsvariable ist genau dann diskret, wenn ihre Verteilung diskret ist. Analog gilt für diskrete Zufallsvariablen, dass sie vollständig durch ihre Wahrscheinlichkeitsfunktion beschrieben werden.

Definition 3.8 (Wahrscheinlichkeitsfunktion): Sei X eine diskrete Zufallsvariable mit Wertebereich $W = \{x_1, x_2, \dots\}$. Dann heißt die Funktion $p : W \rightarrow [0, 1]$ definiert durch

$$p(x_i) := \mathbb{P}(\{\omega : X(\omega) = x_i\}) = \mathbb{P}(X = x_i), \quad (3.4)$$

die *Wahrscheinlichkeitsfunktion* von X , auch *Zähldichte* genannt.

Anmerkung: Gelegentlich wird p zu einer Funktion auf ganz \mathbb{R} ausgedehnt, indem $p(x) = 0$ für $x \in \mathbb{R} \setminus X(\Omega)$ gesetzt wird.

Beispiel 3.9: Wir betrachten wieder die Indikatorfunktion aus dem Beispiel 3.4. Sie ist eine diskrete Zufallsvariable mit dem Wertebereich $\{0, 1\}$. Die zugehörige Wahrscheinlichkeitsfunktion ist gegeben durch $p(1) = \mathbb{P}(A)$ und $p(0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Beispiel 3.10 (Münzwurf): Wir betrachten 4-maliges Werfen einer fairen Münze. Sei X definiert durch $X(\omega) = \omega_1 + \omega_2 + \omega_3 + \omega_4$ die interessierende Zufallsvariable. Sie ist diskret mit dem Wertebereich $\{0, 1, 2, 3, 4\}$. Wir erhalten dann die Wahrscheinlichkeitsfunktion

k	0	1	2	3	4
$p(k)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

Damit lässt sich ganz leicht die Wahrscheinlichkeit aller Ereignisse der Form $\{X \in A\}$ berechnen, z. B. ist $A = (-\infty, 2]$ und $\mathbb{P}(A) = \mathbb{P}(X \leq 2) = p(0) + p(1) + p(2) = 11/16$.

3.2.2 Charakterisierung diskreter Zufallsvariablen

Es gibt eine Vielzahl praktisch auftretender Verteilungen, deren „Form“ sich stark unterscheidet. Wir wollen uns im Folgenden Möglichkeiten ansehen, diese „Form“ zu charakterisieren. Dazu führen wir die Begriffe des Erwartungswertes und der Varianz ein und beschreiben deren Eigenschaften.

Der Erwartungswert einer Zufallsvariablen ist der Wert, der sich gemittelt über alle Realisierungen, gewichtet mit der Wahrscheinlichkeit ihres Auftretens, ergibt. Er sollte nicht mit dem wahrscheinlichsten Einzelwert verwechselt werden. Das Konzept stammt aus der Untersuchung von Glücksspielen, wo der Erwartungswert den durchschnittlichen Gewinn oder Verlust bei praktisch unendlicher Wiederholung des Spiels beschreibt.

Definition 3.11 (Erwartungswert): Sei X eine diskrete Zufallsvariable mit Wertebereich W und p ihre Wahrscheinlichkeitsfunktion. Der Erwartungswert von X existiert, falls $\sum_x |x| \cdot p(x) < \infty$. Der Erwartungswert ist dann definiert als

$$\mathbb{E}(X) := \sum_{x \in W} x \cdot p(x), \quad (3.5)$$

als Symbol wählt man manchmal μ oder μ_X .

Beispiel 3.12: Betrachten wir die Zufallsvariable aus Beispiel 3.10. Dann ist

$$\mathbb{E}(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 2$$

der Erwartungswert in dem Spiel, d. h. von vier Würfeln mit einer fairen Münze werde zweimal *Kopf* erwartet. Das entspricht auch dem common sense.

Lemma 3.13 (Erwartungswerte von Indikatorfunktionen): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, seien $A, B \in \mathcal{F}$. Dann gilt für die Indikatorfunktionen $\mathbb{1}_A$ und $\mathbb{1}_B$, dass

$$\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A) \quad \text{und} \quad \mathbb{E}(\mathbb{1}_A \mathbb{1}_B) = \mathbb{P}(A \cap B). \quad (3.6)$$

BEWEIS Aus der Definition der Indikatorfunktion 3.4 wissen wir, dass $\mathbb{1}_A(\omega) = 1$, falls $\omega \in A$, und $\mathbb{1}_A(\omega) = 0$, falls $\omega \notin A$. Dann ist

$$\mathbb{E}(\mathbb{1}_A) = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

Weiter ist

$$(\mathbb{1}_A \mathbb{1}_B)(\omega) = \mathbb{1}_A(\omega) \cdot \mathbb{1}_B(\omega) = \begin{cases} 1, & \omega \in A, \omega \in B, \\ 0, & \omega \in A, \omega \notin B, \\ 0, & \omega \notin A, \omega \in B, \\ 0, & \omega \notin A, \omega \notin B, \end{cases} = \begin{cases} 1, & \omega \in A \cap B, \\ 0, & \omega \notin A \cap B, \end{cases}$$

und mit dem ersten Teil folgt die Behauptung. \square

Anmerkung: Das vorige Lemma gibt uns ein Werkzeug an die Hand, wie die Mengenoperation Schnittbildung durch eine Multiplikation ersetzt wird. Zusammen mit der Eigenschaft der Komplementärmenge und ihrer Berechnung, siehe Beispiel 3.9, und damit auch der Berechnung von Vereinigungen, können wir Mengenoperationen durch die einfacher handhabbaren Multiplikation und Addition ersetzen. Nicht zuletzt in der Programmierung erweist sich das als sinnvoll.

Satz 3.14 (Transformationformel für den Erwartungswert): Sei X eine diskrete Zufallsvariable mit Wertebereich W , p ihre Wahrscheinlichkeitsfunktion, sei weiter $u : W \rightarrow \mathbb{R}$ eine Abbildung, für die $\sum_{x \in W} |u(x)|p(x) < \infty$ gilt. Wir definieren $Y := u(X)$ und es gilt

$$\mathbb{E}(Y) = \mathbb{E}(u(X)) = \sum_{x \in X(\Omega)} u(x) \cdot p(x) \quad (3.7)$$

für den Erwartungswert der transformierten Zufallsvariablen Y .

BEWEIS Wir untersuchen die zugehörige Wahrscheinlichkeitsfunktion

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(u(X) = y) = \sum_{x:u(x)=y} p(x),$$

mit der wir den Erwartungswert

$$\mathbb{E}(Y) = \sum_y y \cdot p_Y(y) = \sum_y y \sum_{x:u(x)=y} p(x) = \sum_y \sum_{x:u(x)=y} u(x)p(x) = \sum_x u(x)p(x)$$

berechnen. \square

3 Zufallsvariablen und Verteilungen

Satz 3.15 (Dreiecksungleichung für den Erwartungswert): Sei X eine diskrete Zufallsvariable mit Wertebereich W , deren Erwartungswert existiert, d. h. $\sum_{x \in W} |x|p(x) < \infty$. Dann gilt

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|). \quad (3.8)$$

BEWEIS Die Transformationsformel aus Satz 3.14 mit $u(x) = |x|$ liefert

$$\mathbb{E}(|X|) = \sum_{x \in W} |x|p(x) = \sum_{x \in W} |xp(x)| \geq \left| \sum_{x \in W} xp(x) \right| = |\mathbb{E}(X)|$$

wobei bei der Ungleichheit die Dreiecksungleichung für Summen reeller Zahlen verwendet wurde. \square

Satz 3.16 (Linearität des Erwartungswertes): Seien X und Y zwei Zufallsvariablen auf dem gleichen Wahrscheinlichkeitsraum, deren Erwartungswerte existieren. Dann gilt für $a, b \in \mathbb{R}$

- (i) $\mathbb{E}(aX) = a\mathbb{E}(X)$,
- (ii) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
- (iii) $\mathbb{E}(b) = b$.

BEWEIS (i) Aus der Transformationsformel haben wir mit der Funktion $u(x) = a \cdot x$

$$\mathbb{E}(aX) = \sum_x (ax)p(x) = a \sum_x xp(x) = a\mathbb{E}(X)$$

(ii) Eine hinreichende Bedingung für die Gleichheit von Summen ist die Gleichheit und das Auftreten aller Summanden, somit

$$\mathbb{E}(X + Y) = \sum_z z \mathbb{P}(X + Y = z),$$

aus der Summe wir mit dem Satz von der totalen Wahrscheinlichkeit (Satz 2.35) eine Doppelsumme bilden

$$= \sum_z \sum_x z \mathbb{P}(X + Y = z, X = x),$$

die wir umstellen können

$$= \sum_x \sum_z ((z - x) + x) \mathbb{P}(Y = z - x, X = x)$$

und jetzt $y := z - x$ definieren und wegen $x, z \in \mathbb{R}$ gilt auch $y \in \mathbb{R}$, weswegen wir nun nur die Reihenfolge der Summanden ändern

$$\begin{aligned} &= \sum_x \sum_y (y + x) \mathbb{P}(Y = y, X = x) \\ &= \sum_x \sum_y x \mathbb{P}(Y = y, X = x) + \sum_x \sum_y y \mathbb{P}(Y = y, X = x), \\ &= \sum_x x \sum_y \mathbb{P}(Y = y, X = x) + \sum_y y \sum_x \mathbb{P}(Y = y, X = x), \end{aligned}$$

der Satz von der totalen Wahrscheinlichkeit liefert

$$\begin{aligned} &= \sum_x x \mathbb{P}(X = x) + \sum_y y \mathbb{P}(Y = y), \\ &= \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

(iii) Die konstante Zufallsvariable $X \equiv b$ hat genau eine Realisierung, und zwar mit Wahrscheinlichkeit 1. Die Wahrscheinlichkeitsfunktion ist dann $p(b) = 1$ und $p(x) = 0$ falls $x \neq b$. \square

Anmerkung: Wir haben in diesem Beweis verwendet, dass $X + Y$ eine Zufallsvariable ist. Allgemeiner formuliert sagen wir, dass $u(X, Y)$ eine Zufallsvariable ist, wenn u messbar ist. Das ist eine Generalisierung des Transformationsatzes 3.14, die wir im Abschnitt 3.4 für mehrdimensionale Zufallsvariablen kennenlernen werden. Die Aussage selbst werden hin und wieder vorwegnehmen.

Neben der Erwartung benötigen wir noch die Varianz einer Zufallsvariablen.

Definition 3.17 (Varianz): Sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann ist die *Varianz* von X gegeben durch

$$\text{Var}X := \mathbb{E}(X - \mathbb{E}(X))^2. \quad (3.9)$$

Als Symbol wird oft σ^2 oder σ_X^2 verwendet. Die positive Wurzel aus der Varianz nennt man die *Standardabweichung* von X .

Anmerkung: Aus der Bedingung $\mathbb{E}(X^2) < \infty$ folgt, dass $\mathbb{E}(|X|) < \infty$, also die Existenz des Erwartungswertes.

Mit dieser Definition als mittlere quadratische Abweichung ist die Varianz ein Maß für die Streuung.

Satz 3.18 (Rechenregeln für die Varianz): Sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann gilt

- (i) $\text{Var}(aX + b) = a^2 \text{Var}(X)$ für $a, b \in \mathbb{R}$,
- (ii) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

BEWEIS (i) Wegen der Linearität des Erwartungswertes gilt $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ und es folgt

$$\text{Var}(aX + b) = \mathbb{E}(aX + b - (a\mathbb{E}(X) + b))^2 = \mathbb{E}(a(X - \mathbb{E}(X)))^2 = a^2 \mathbb{E}(X - \mathbb{E}(X))^2 = a^2 \text{Var}(X).$$

(ii) Ebenfalls aus der Linearität des Erwartungswertes folgt

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2 - 2(\mathbb{E}(X))X + (\mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \quad \square$$

Die Varianz und den Erwartungswert in Beziehung setzend sehen wir, dass der Erwartungswert eine Minimumeigenschaft besitzt.

Satz 3.19 (Minimiereigenschaft des Erwartungswertes): Sei X eine Zufallsvariable, für die $\mathbb{E}(X^2) < \infty$ gilt, und sei $a \in \mathbb{R}$. Dann gilt

$$\mathbb{E}(X - a)^2 = \text{Var}(X) + (\mathbb{E}(X) - a)^2 \quad (3.10)$$

und damit

$$\mathbb{E}(X - a)^2 \geq \text{Var}(X). \quad (3.11)$$

BEWEIS Wegen der Linearität des Erwartungswertes ist $\mathbb{E}(X - \mathbb{E}(X)) = 0$ und damit folgt

$$\begin{aligned} \mathbb{E}(X - a)^2 &= \mathbb{E}(X - \mathbb{E}(X) + \mathbb{E}(X) - a)^2 \\ &= \mathbb{E}\left((X - \mathbb{E}(X))^2 + 2\mathbb{E}(X - \mathbb{E}(X))(\mathbb{E}(X) - a) + (\mathbb{E}(X) - a)^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) + 2\mathbb{E}\left((X - \mathbb{E}(X))(\mathbb{E}(X) - a)\right) + \mathbb{E}\left((\mathbb{E}(X) - a)^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) + 2(\mathbb{E}(X) - a) \cdot \mathbb{E}(X - \mathbb{E}(X)) + \mathbb{E}\left((\mathbb{E}(X) - a)^2\right) \\ &= \mathbb{E}(X - \mathbb{E}(X))^2 + (\mathbb{E}(X) - a)^2, \end{aligned}$$

weiterhin ist der Term für $a = \mathbb{E}(X)$ minimal. \square

3.3 Stetige Zufallsvariablen

Für viele Zufallsexperimente, die wir modellieren wollen, benötigen wir ein Kontinuum an möglichen Werten für die verwendete Zufallsvariable. Beispielsweise sind dies die Lebensdauer einer Person, eine Zahl aus dem Intervall $[0, 1]$ oder der zufällige Zeitpunkt an dem eine Option auf ein Wertpapier ausgeübt wird usw. Der Wertebereich einer (absolut) stetigen¹ Zufallsvariablen ist überabzählbar.

Definition 3.20: Eine integrierbare, nicht-negative Funktion f heißt *Wahrscheinlichkeitsdichte*, *Dichtefunktion* oder *Dichte* der Zufallsvariablen X , falls für alle $a, b \in \mathbb{R}$ mit $a < b$ gilt

$$\mathbb{P}(a < X \leq b) = P_X((a, b]) = \int_a^b f(x) dx. \quad (3.12)$$

Verteilungen mit einer Dichtefunktion heißen (absolut) stetige Verteilungen. Die Verteilungsfunktion F_X besitzt die Darstellung

$$F_X(y) = \int_{-\infty}^y f(x) dx. \quad (3.13)$$

Anmerkung: (i) Für eine Dichtefunktion f gilt immer $\int f(x) dx = 1$. Umgekehrt wird von einer Funktion f mit dieser Eigenschaft eine Wahrscheinlichkeitsverteilung auf \mathbb{R} definiert. Daher ist jede Funktion f mit $\int f(x) dx = 1$ eine Wahrscheinlichkeitsdichte.

(ii) Zwei Verteilungen P_X und P_Y sind gleich, wenn ihre Dichtefunktionen bis auf endlich oder abzählbar unendlich viele Werte übereinstimmen.

Beispiel 3.21: Sei $f_c(x) = cx^3 \cdot \mathbb{1}_{[0,1]}(x)$ mit $c \in \mathbb{R}$ eine Funktionenschar. Für welche Werte von c ist f_c eine Dichtefunktion?

Wir halten fest, dass $f_c(x) \geq 0$ für alle $x \in \mathbb{R}$ und alle $c \in \mathbb{R}$ gilt. Als nächstes bestimmen wir das Integral

$$\int cx^3 \cdot \mathbb{1}_{[0,1]}(x) dx = \int_0^1 cx^3 dx = c \cdot \frac{1}{4} x^4 \Big|_0^1 = \frac{c}{4},$$

d. h. für $c = 4$ haben wir eine Dichte. Sei X eine Zufallsvariable mit der Verteilung mit Dichte f_4 . Dann gilt

$$\mathbb{P}(1/4 < X \leq 1/2) = \int_{1/4}^{1/2} 4x^3 \mathbb{1}_{[0,1]}(x) dx = \int_{1/4}^{1/2} 4x^3 dx = x^4 \Big|_{1/4}^{1/2} = \frac{1}{16} - \frac{1}{256} = \frac{15}{256}$$

und

$$\begin{aligned} \mathbb{P}(X \leq a) &= \int_{-\infty}^a 4x^3 \mathbb{1}_{[0,1]}(x) dx = \int_0^{\min\{\max\{0,a\}, 1\}} 4x^3 dx = x^4 \Big|_0^{\min\{\max\{0,a\}, 1\}} \\ &= \begin{cases} 0 & = \text{falls } 0 > a, \\ a^4 & = \text{falls } 0 \leq a \leq 1, \\ 1 & = \text{falls } 1 < a. \end{cases} \end{aligned}$$

¹Im allgemeinen Sprachgebrauch wird der Begriff *stetig* verwendet, mathematisch spricht man in dem Fall von *absolut stetig*. Trotz unterschiedlicher Bezeichnungen wird in beiden Kontexten eine Zufallsvariable betrachtet, für deren Verteilung eine Dichte (siehe Definition 3.20) existiert.

Satz 3.22 (Eigenschaften der Verteilungsfunktion einer (absolut) stetigen Zufallsvariablen): Sei X eine (absolut) stetige Zufallsvariable und sei F_X ihre die Verteilungsfunktion. Dann gilt

- (i) Für jeden Punkt $x \in \mathbb{R}$ gilt $\mathbb{P}(X = x) = 0$.
- (ii) Für Intervalle erhalten wir

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

und $\mathbb{P}(X \geq a) = 1 - F_X(a)$.

3.3.1 Charakterisierung stetiger Zufallsvariablen

Wir gehen hier analog zum Abschnitt 3.2.2 für diskrete Zufallsvariablen vor.

Definition 3.23 (Erwartungswert): Sei X eine stetige Zufallsvariable mit Dichte f . Der *Erwartungswert* existiert, falls $\int |x|f(x)dx < \infty$ und ist dann definiert durch

$$\mathbb{E}X = \int xf(x)dx. \quad (3.14)$$

Beispiel 3.24: Sei X eine Zufallsvariable mit der Dichte $f_X(x) = 1/2 \cdot \mathbb{1}_{(0,2)}(x)$. Dann ist

$$\mathbb{E}(X) = \int x \cdot \frac{1}{2} \cdot \mathbb{1}_{(0,2)}(x)dx = \frac{1}{2} \int_0^2 xdx = \frac{1}{4}x^2 \Big|_0^2 = 1$$

Satz 3.25 (Transformationsformel für den Erwartungswert): Sei X eine stetige Zufallsvariable mit Dichte f und sei $u : \mathbb{R} \rightarrow \mathbb{R}$ eine messbare Abbildung. Dann gilt

$$\mathbb{E}(u(X)) = \int u(x)f(x)dx \quad (3.15)$$

falls $\int |u(x)|f(x)dx < \infty$ ist.

Anmerkung (Analogie von Eigenschaften stetiger Zufallsvariablen): Die anderen Sätze zum Erwartungswert aus dem Abschnitt 3.2.2 wie die Dreiecksungleichung (Satz 3.15) und die Linearität (Satz 3.16) gelten analog.

Für die Definition der Varianz (Definition 3.17) als auch ihre Eigenschaften (Satz 3.18) haben wir nur die Definition des Erwartungswertes verwendet und können sie deswegen problemlos übernehmen.

3.4 Mehrdimensionale Verteilungen

Wir hatten es schon bei der Linearität des Erwartungswertes mit mehreren Zufallsvariablen zu tun und waren davon ausgegangen, dass auch ihre Summe eine Zufallsvariable ist. Das müssen wir uns noch genauer ansehen.

3.4.1 Gemeinsame und marginale Verteilungen

Wenn wir mehrere Zufallsvariablen X_1, \dots, X_n betrachten, so können wir diese als einen Zufallsvektor $X = (X_1, \dots, X_n)$ auffassen.

Eigentlich sollten wir X als Spaltenvektor schreiben und damit konsequenterweise $X = (X_1, \dots, X_n)^T$. Wir lassen das der besseren Lesbarkeit zu Grunde oft ungeschrieben.

Definition 3.26 (Zufallsvektor): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X_1, \dots, X_n Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$ für $1 \leq i \leq n$. Die Abbildung $X : \Omega \rightarrow \mathbb{R}^n$ heißt dann *n-dimensionaler Zufallsvektor* mit den Komponenten (oder Koordinaten) X_1, \dots, X_n . Die Funktion $F_X : \mathbb{R}^n \rightarrow [0, 1]$ definiert durch

$$F_X(x_1, \dots, x_n) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (3.16)$$

heißt *gemeinsame (oder multivariate) Verteilungsfunktion* des Zufallsvektors $X = (X_1, \dots, X_n)$.

Anmerkung: Dann ist $X : \Omega \rightarrow \mathbb{R}^n$ eine messbare Funktion. Dazu müssten wir uns eigentlich die Regularitätsbedingung der Messbarkeit auf \mathbb{R}^n genauer ansehen. Man kann zeigen, dass die Messbarkeit von $X = (X_1, \dots, X_n)$ zur Messbarkeit aller Koordinaten X_1, \dots, X_n ist.

Beispiel 3.27: Wir betrachten zweifachen Würfelwurf. Sei $X = (X_1, X_2, X_3) : \Omega \rightarrow \mathbb{R}^3$ mit Wertebereich $W = \{2, 3, \dots, 12\} \times \{1, \dots, 6\} \times \{0, 1\}$. Dabei bezeichnet X_1 die Summe der Augenzahlen der zwei Würfel, X_2 die Augenzahl des ersten Würfels und X_3 die Kodierung ob die Augenzahl des zweiten Würfels gerade ist.

Satz 3.28 (Eigenschaften multivariater Verteilungsfunktionen): Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor mit der Verteilungsfunktion F_X . Dann gilt

(i) *Asymptotisches Verhalten:* Für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $1 \leq i \leq n$ gilt

$$\lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = 0,$$

weiterhin

$$\lim_{x_1, \dots, x_n \rightarrow \infty} F_X(x_1, \dots, x_n) = 1.$$

Seien $1 \leq i_1 < \dots < i_k \leq n$ Indizes und $\{j_1, \dots, j_{n-k}\} := \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$ die komplementäre Indexmenge. Dann ist

$$\lim_{x_{j_1}, \dots, x_{j_{n-k}} \rightarrow \infty} F_X(x_1, \dots, x_n) = \lim_{x_{j_1}, \dots, x_{j_{n-k}} \rightarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k})$$

die Verteilungsfunktion der *k-dimensionalen Marginalverteilung* der gemeinsamen Verteilung.

(ii) *Rechtsstetigkeit:* F_X ist stetig von rechts mit linksseitigen Grenzwerten, formal

$$F_X(x_1, \dots, x_n) = \lim_{h_1, \dots, h_n \rightarrow 0} F_X(x_1 + h_1, \dots, x_n + h_n)$$

für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $h_1, \dots, h_n \geq 0$.

(iii) *Monotonie:* F_X ist monoton wachsend mit Werten im Intervall $[0, 1]$, formal

$$F_X(x_1, \dots, x_n) \leq F_X(x_1 + h_1, \dots, x_n + h_n)$$

für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $h_1, \dots, h_n \geq 0$.

Beispiel 3.29: Nehme einen Zufallsvektor $X = (X_1, X_2)$ mit Wertebereich $\{0, 1, 2\} \times \{5, 10\}$ und sei die gemeinsame Wahrscheinlichkeitsfunktion p_{X_1, X_2} gegeben durch

	0	1	2	
5	1/4	1/6	1/3	3/4
10	1/12	1/12	1/12	1/4
5	4/12	3/12	5/12	

Dann ist $F_{X_1, X_2}(1, 5) = 1/4 + 1/6 = 5/12$ und $\lim_{x_2 \rightarrow \infty} F_{X_1, X_2}(1, x_2) = F_{X_1, X_2}(1, 5) + F_{X_1, X_2}(1, 10) = 7/12$, da für größere x_2 nichts mehr hinzukommt.

Anmerkung: (i) Ein Zufallsvektor heißt diskret, wenn jede Koordinate X_i für $a \leq i \leq n$ diskret ist. Dann existiert eine Wahrscheinlichkeitsfunktion $p_X = p_{X_1, \dots, X_n}$ von X .

(ii) Ist jede Koordinate X_i für $a \leq i \leq n$ stetig, so nennen wir auch den Zufallsvektor stetig. Es existiert dann eine integrierbare Funktion $f_X = f_{X_1, \dots, X_n}$ von X .

(iii) Analog zur Verteilungsfunktion einer Marginalverteilung können wir die zugehörige Wahrscheinlichkeitsfunktion oder Dichte durch aufsummieren resp. integrieren erhalten. Seien $\{i_1, \dots, i_k\}$ und $\{j_1, \dots, j_{n-k}\}$ zwei komplementäre Indexmengen mit Indizes aus $\{1, \dots, n\}$. Dann gilt

$$\sum_{j_1, \dots, j_{n-k}} p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k})$$

und

$$\int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_{j_1} \cdots dx_{j_{n-k}} = f_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}).$$

3.4.2 Erwartungswert

Definition 3.30 (Erwartungswert): (i) Sei $X = (X_1, \dots, X_n)^T$ ein diskreter Zufallsvektor mit gemeinsamer Wahrscheinlichkeitsfunktion p_X bei dem alle Koordinaten X_i mit $1 \leq i \leq n$ einen Erwartungswert besitzen. Dann ist

$$\begin{aligned} \mathbb{E}X &= \mathbb{E} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \sum_{x_1, \dots, x_n} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} p_X(x_1, \dots, x_n) = \begin{pmatrix} \sum_{x_1, \dots, x_n} x_1 p_X(x_1, \dots, x_n) \\ \vdots \\ \sum_{x_1, \dots, x_n} x_n p_X(x_1, \dots, x_n) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{x_1} x_1 p_{X_1}(x_1) \\ \vdots \\ \sum_{x_n} x_n p_{X_n}(x_n) \end{pmatrix} = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{pmatrix} \end{aligned} \quad (3.17)$$

der Erwartungswert von X .

(ii) Sei $Y = (Y_1, \dots, Y_n)^T$ ein stetiger Zufallsvektor mit gemeinsamer Wahrscheinlichkeitsdichte f_Y bei dem alle Koordinaten Y_i mit $1 \leq i \leq n$ integrierbar sind. Dann ist der Erwartungswert von Y gegeben durch

$$\mathbb{E}Y = \mathbb{E} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \int_{\mathbb{R}^n} \cdots \int (y_1, \dots, y_n)^T f_Y(y_1, \dots, y_n) dy_1 \cdots dy_n = \begin{pmatrix} \mathbb{E}Y_1 \\ \vdots \\ \mathbb{E}Y_n \end{pmatrix}. \quad (3.18)$$

3 Zufallsvariablen und Verteilungen

Beispiel 3.31: Wir beziehen uns auf das Beispiel 3.27. Der Erwartungswert ist

$$\mathbb{E}X = \mathbb{E} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mathbb{E}X_1 \\ \mathbb{E}X_2 \\ \mathbb{E}X_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 7/2 \\ 1/2 \end{pmatrix},$$

da $\mathbb{E}X_1 = 7, \mathbb{E}X_2 = 7/2, \mathbb{E}X_3 = 1/2$ gilt.

Satz 3.32 (Transformation von Zufallsvektoren): Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor und $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $m \geq 1$. Dann ist $u(X)$ ein m -dimensionaler Zufallsvektor.

Korollar 3.33 (Allgemeine Transformationsformel für den Erwartungswert): Analog zum Satz (3.14) und Satz (3.25) können wir die Transformationsformel für den Erwartungswert angeben.

(i) Sei $X = (X_1, \dots, X_n)$ ein diskreter Zufallsvektor mit der gemeinsamen Wahrscheinlichkeitsfunktion p und sei $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Funktion. Dann gilt

$$\mathbb{E}(u(X)) = \mathbb{E}(u(X_1, \dots, X_n)) = \sum_{x_1, \dots, x_n} u(x_1, \dots, x_n) p(x_1, \dots, x_n). \quad (3.19)$$

(ii) Sei $Y = (Y_1, \dots, Y_n)$ ein stetiger Zufallsvektor mit der gemeinsamen Wahrscheinlichkeitsdichte f und sei $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Funktion. Dann gilt

$$\mathbb{E}(u(Y)) = \mathbb{E}(u(Y_1, \dots, Y_n)) = \int_{y_1, \dots, y_n} u(y_1, \dots, y_n) f(y_1, \dots, y_n) dy_1 \dots dy_n. \quad (3.20)$$

Satz 3.34 (Eigenschaften des Erwartungswertes): Seien $X = (X_1, \dots, X_n)^T$ und $Y = (Y_1, \dots, Y_n)^T$ beide n -dimensionale Zufallsvektoren, $a, b \in \mathbb{R}$ und $A \in \mathbb{R}^{k \times n}$ eine Matrix. Dann gilt

$$(i) \mathbb{E}(aX + bY) = a \mathbb{E}X + b \mathbb{E}Y,$$

$$(ii) \mathbb{E}(AX) = A \mathbb{E}X,$$

der Erwartungswert ist also linear.

3.4.3 Unabhängige Zufallsvariablen

Wir wollen das Konzept der Stochastischen Unabhängigkeit aus Abschnitt 2.5 auf Zufallsvariablen ausweiten.

Definition 3.35 (Unabhängige Zufallsvariablen): Seien X_1, \dots, X_n Zufallsvariablen. Dann heißen X_1, \dots, X_n (stochastisch) unabhängig, falls

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i) \quad (3.21)$$

für alle $B_1, \dots, B_n \subset \mathbb{R}$ gilt.

Lemma 3.36: Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn die Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ unabhängig sind für alle $B_1, \dots, B_n \subset \mathbb{R}$.

Satz 3.37: Für zwei unabhängige Zufallsvariablen X und Y gilt

$$\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y, \quad (3.22)$$

falls $\mathbb{E}X$ und $\mathbb{E}Y$ existieren. Seien $X = (X_1, \dots, X_n)^T$ und $Y = (Y_1, \dots, Y_n)^T$ unabhängige Zufallsvektoren, so gilt

$$\mathbb{E}(XY^T) = \mathbb{E}(X) \mathbb{E}(Y)^T. \quad (3.23)$$

BEWEIS Diese Aussage zeigen wir für diskrete Zufallsvariablen durch Nachrechnen, für stetige folgt sie analog

$$\begin{aligned}
 \mathbb{E}(XY) &= \sum_z z \mathbb{P}(XY = z) = \sum_z \sum_x z \mathbb{P}(XY = z, X = x) \\
 &\stackrel{x \neq 0}{=} \sum_x \sum_y \frac{z}{x} x \mathbb{P}\left(Y = \frac{z}{x}, X = x\right) \\
 &\stackrel{y = \frac{z}{x}}{=} \sum_x \sum_y y x \mathbb{P}(Y = y, X = x) \\
 &= \sum_x \sum_y x y \mathbb{P}(X = x) \mathbb{P}(Y = y) \\
 &= \sum_x \left(x \mathbb{P}(X = x) \sum_y y \mathbb{P}(Y = y) \right) \\
 &= \left(\sum_x x \mathbb{P}(X = x) \right) \cdot \left(\sum_y y \mathbb{P}(Y = y) \right) \\
 &= \mathbb{E}X \cdot \mathbb{E}Y \quad \square
 \end{aligned}$$

Satz 3.38 (Faltungformel für Wahrscheinlichkeitsfunktionen): Seien X und Y unabhängige Zufallsvariablen. Wir betrachten ihre Summe $Z = X + Y$.

(i) Sind X und Y diskret mit den Wahrscheinlichkeitsfunktionen p_X resp. p_Y , so hat Z die Wahrscheinlichkeitsfunktion

$$p_Z(z) = \sum_x p_X(x) p_Y(z-x) = \sum_y p_X(z-y) p_Y(y). \quad (3.24)$$

(ii) Sind X und Y stetig mit den Wahrscheinlichkeitsdichten f_X resp. f_Y , so hat Z die Wahrscheinlichkeitsfunktion

$$f_Z(z) = \int_{x \in \mathbb{R}} f_X(x) f_Y(z-x) dx = \int_{y \in \mathbb{R}} f_X(z-y) f_Y(y) dy. \quad (3.25)$$

BEWEIS Das Ereignis $\{X + Y = z\}$ schreiben wir unter Verwendung von x und y mit $x + y = z$ als disjunkte Vereinigung der Ereignisse $\{X = x\} \cap \{Y = y\} = \{X = x, Y = y\}$ mit $x \in \mathbb{R}$. Dann gilt $y = z - x$ und es folgt mit dem Satz von der totalen Wahrscheinlichkeit und der Definition der Unabhängigkeit im diskreten Fall, dass

$$\begin{aligned}
 \mathbb{P}(X + Y = z) &= \sum_x \sum_y \mathbb{P}(X = x, Y = y) = \sum_x \mathbb{P}(X = x, Y = z - x) \\
 &= \sum_x \mathbb{P}(X = x) \mathbb{P}(Y = z - x) = \sum_x p_X(x) p_Y(z - x),
 \end{aligned}$$

im stetigen analog. □

Notation: Für die Faltung zweier Wahrscheinlichkeitsfunktionen schreiben wir $p_X * p_Y$.

3.4.4 Varianz und Kovarianz

Definition 3.39 (Kovarianz und Korrelationskoeffizient): Für zwei Zufallsvariablen X und Y definieren wir die *Kovarianz* durch

$$\mathbb{C}_{\text{ov}}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) \quad (3.26)$$

3 Zufallsvariablen und Verteilungen

und den *Korrelationskoeffizienten* durch

$$\rho_{X,Y} = \frac{\mathbb{C}\text{ov}(X,Y)}{\sqrt{\text{Var}X}\sqrt{\text{Var}Y}}. \quad (3.27)$$

Die Zufallsvariablen heißen unkorreliert, falls $\rho_{X,Y} = 0$ ist.

Satz 3.40 (Rechnen mit Kovarianzen): Seien X und Y Zufallsvariablen. Dann gilt

- (i) $\mathbb{C}\text{ov}(X, X) = \text{Var}X$,
- (ii) $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$,
- (iii) $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2\mathbb{C}\text{ov}(X, Y)$,
- (iv) sind X und Y unabhängig, so sind sie unkorreliert und es gilt

$$\text{Var}(X + Y) = \text{Var}X + \text{Var}Y, \quad (3.28)$$

falls alle Ausdrücke existieren.

BEWEIS Teil (i) folgt direkt aus den Definitionen 3.17 und 3.39, (ii) erhält man analog zum Satz 3.18 (ii) aus der Linearität des Erwartungswertes. Für Teil (iii) gilt

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}\left((X + Y - \mathbb{E}(X + Y))^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}X + Y - \mathbb{E}Y)^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}X)^2 + (Y - \mathbb{E}Y)^2 + 2 \cdot ((X - \mathbb{E}X)(Y - \mathbb{E}Y))\right) \\ &= \text{Var}X + \text{Var}Y + 2\mathbb{C}\text{ov}(X, Y), \end{aligned}$$

und Teil (iv) folgt hieraus dann direkt. □

Anmerkung: Aus Unkorreliertheit folgt nicht Unabhängigkeit.

Beispiel 3.41 (Gegenbeispiel zur Umkehrung): Betrachte die Zufallsvariable X mit Wertebereich $W_X = \{-1, 0, 1\}$ und der Wahrscheinlichkeitsfunktion $p_X(-1) = 1/4$, $p_X(0) = 1/2$, $p_X(1) = 1/4$. Sei weiter $Y := X^2$. Dann gilt $W_Y = \{0, 1\}$ und Y besitzt die Wahrscheinlichkeitsfunktion $p_Y(0) = 1/2$, $p_Y(1) = 1/2$. Dann gilt $\mathbb{E}X = 1/4 \cdot (-1) + 1/2 \cdot 0 + 1/4 \cdot 1 = 0$ und $\mathbb{E}Y = 1/2 \cdot 0 + 1/2 \cdot 1 = 1/2$ und die Kovarianz ist

$$\begin{aligned} \mathbb{C}\text{ov}(X, Y) &= \mathbb{E}\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) = \mathbb{E}\left(X(Y - 1/2)\right) = \mathbb{E}(XY) - 1/2\mathbb{E}X \\ &= \mathbb{E}(X^3) = 1/4 \cdot (-1)^3 + 1/2 \cdot 0^3 + 1/4 \cdot 1^3 = 0, \end{aligned}$$

und damit ist auch $\rho_{X,Y} = 0$, also sind X und Y unkorreliert. Sie sind aber nicht unabhängig, da z. B. $\mathbb{P}(X = 0, Y = 1) = \mathbb{P}(X = 0, X^2 = 1) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1)$.

Definition 3.42 (Kovarianzmatrix): Sei $X = (X_1, \dots, X_n)^T$ ein Zufallsvektor und definiere die Kovarianzen $s_{ij} := \mathbb{C}\text{ov}(X_i, X_j)$ für $1 \leq i, j \leq n$. Dann ist

$$\Sigma := \mathbb{C}\text{ov}X = \mathbb{E}\left((X - \mathbb{E}X)(X - \mathbb{E}X)^T\right) = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix} \quad (3.29)$$

die Kovarianzmatrix von X .

Anmerkung: Zu beachten ist, dass auf der Diagonalen der Kovarianzmatrix die Varianzen zu finden sind. Deswegen haben wir den Zusammenhang $s_{ii} = \sigma_i^2$ für die Einträge der Kovarianzmatrix mit der Varianz.

Satz 3.43 (Eigenschaften der Kovarianzmatrix): Die Kovarianzmatrix Σ ist

(i) symmetrisch, d. h. $s_{ij} = s_{ji}$,

(ii) nicht-negativ definit, d. h. $x^T \Sigma x \geq 0$ für alle $x \in \mathbb{R}^n$.

Gilt sogar $x^T \Sigma x > 0$, so ist Σ positiv definit.

3.5 Spezielle/Wichtige diskrete Verteilungen

Im Folgenden wollen wir uns einige in der Praxis häufig vorkommende Verteilungen ansehen.

Notation: Ist eine Zufallsvariable X einer Verteilung $F(x) := \mathbb{P}(X \leq x)$ entsprechend verteilt, so sagt man $X \sim F$.

Bernoulli-Verteilung

Sei $A \in \Omega$ mit $\mathbb{P}(A) = \eta$ mit $\eta \in (0, 1)$ gegeben. Sei X eine Zufallsvariable, definiert durch

$$X(\omega) = \mathbb{1}_A(\omega)$$

wobei wir das Eintreten von A mit *Erfolg*, dasjenige von A^c mit *Misserfolg* identifizieren. Der Wertebereich $W = \{0, 1\}$ und X ist diskret mit der Wahrscheinlichkeitsfunktion

$$p(k) = \begin{cases} \eta & \text{für } k = 1, \\ 1 - \eta & \text{für } k = 0. \end{cases}$$

Die Verteilung heißt dann *Bernoulli-Verteilung* mit Parameter η .

Als Erwartungswert erhalten wir

$$\mathbb{E}X = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = \eta$$

und als Varianz

$$\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = (1^2 \cdot \mathbb{P}(X = 1) + 0^2 \cdot \mathbb{P}(X = 0)) - \eta^2 = \eta(1 - \eta).$$

Binomial-Verteilung

Beispiel 3.44: Typisches Szenario als einführendes Beispiel.

Seien X_i unabhängige Zufallsvariablen für $i = 1, \dots, n$ Bernoulli-verteilt mit Parameter η . Dann definiert $S_n := X_1 + \dots + X_n$ eine Zufallsvariable, die die Anzahl der Erfolge in n Experimenten angibt. Die Wahrscheinlichkeitsfunktion ist

$$p(k) = \mathbb{P}(S_n = k) = \binom{n}{k} \eta^k (1 - \eta)^{n-k} \quad \text{für } k = 0, \dots, n,$$

3 Zufallsvariablen und Verteilungen

was aus der Überlegung folgt, dass wir, um genau k Erfolge zu erhalten, k Erfolge und $n - k$ Misserfolge benötigen, wobei deren Reihenfolge unwichtig ist.

Notation: Die Binomialverteilung mit den Parametern n und η wird mit $\text{Bin}(n, \eta)$ abgekürzt.

Bevor wir den Erwartungswert ausrechnen, müssen wir uns noch folgende Identitäten kennen.

Lemma 3.45: (i) Für $k, n \in \mathbb{N}$ mit $k > n$ gilt

$$\binom{n}{k} = 0. \quad (3.30)$$

(ii) Für $k, n \in \mathbb{N}$ mit $0 < k < n$ gilt

$$k \binom{n}{k} = n \binom{n-1}{k-1}. \quad (3.31)$$

(iii) Für die natürlichen Zahlen k, l, m gilt

$$\binom{l+m}{k} = \sum_{i=0}^k \binom{l}{i} \binom{m}{k-i}.$$

Der Erwartungswert von S_n ist dann

$$\mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i = \sum_{i=1}^n \eta = n\eta$$

und weiter

$$\begin{aligned} \mathbb{E}(S_n(S_n - 1)) &= \sum_{k=0}^n k(k-1) \binom{n}{k} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n k(k-1) \binom{n}{k} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n n(k-1) \binom{n-1}{k-1} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n n(n-1) \binom{n-2}{k-2} \eta^k (1-\eta)^{n-k} \\ &= n(n-1) \cdot \eta^2 \sum_{k=2}^n \binom{n-2}{k-2} \eta^{k-2} (1-\eta)^{(n-2)-(k-2)} \\ &= n(n-1) \cdot \eta^2 \underbrace{\sum_{j=0}^{n-2} \binom{n-2}{j} \eta^j (1-\eta)^{(n-2)-j}}_{=1, \text{ da Bin}(n-2, \eta)} \\ &= n(n-1) \cdot \eta^2 \end{aligned}$$

und damit die Varianz

$$\begin{aligned} \text{Var}(S_n) &= \mathbb{E}(S_n^2) - (\mathbb{E}S_n)^2 \\ &= \mathbb{E}(S_n(S_n - 1)) + \mathbb{E}S_n - (\mathbb{E}S_n)^2 \\ &= n(n-1)\eta^2 + n\eta - (n\eta)^2 = n\eta - n\eta^2 = n\eta(1-\eta) \end{aligned}$$

Beispiel 3.46: Im Beispiel 3.10 hatten wir es mit einer Binomialverteilung mit $n = 4$ und $\eta = 0,5$ zu tun. Damit ist $\mathbb{E}S_4 = 4 \cdot 0,5 = 2$ (was wir schon hatten) und $\text{Var}S_4 = 4 \cdot 0,5 \cdot 0,5 = 1$.

Satz 3.47 (Symmetrieeigenschaft): Sei $X \sim \text{Bin}(n, \eta)$ und sei $Y := n - X$. Dann gilt

$$Y \sim \text{Bin}(n, 1 - \eta). \quad (3.32)$$

BEWEIS Seien p_X und p_Y die Wahrscheinlichkeitsfunktion von X resp. Y . Dann ist

$$\begin{aligned} p_Y(k) &= \mathbb{P}(Y = k) = \mathbb{P}(n - X = k) = \mathbb{P}(X = n - k) = p_X(n - k) \\ &= \binom{n}{n - k} \eta^{n - k} (1 - \eta)^{n - (n - k)} \\ &= \binom{n}{k} (1 - \eta)^k (1 - (1 - \eta))^{n - k} \end{aligned}$$

was die Wahrscheinlichkeitsfunktion von $\text{Bin}(n, 1 - \eta)$ ist. \square

Faltung binomialverteilter Zufallsvariablen

Seien X und Y unabhängige Zufallsvariablen mit $\text{Bin}(n, \eta)$ bzw. $\text{Bin}(m, \eta)$ -Verteilung. Dann lässt sich die Wahrscheinlichkeitsfunktion von $Z = X + Y$ berechnen

$$\begin{aligned} p_Z(k) &= \sum_{i=0}^k \left(\binom{n}{i} \eta^i (1 - \eta)^{n - i} \right) \cdot \left(\binom{m}{k - i} \eta^{k - i} (1 - \eta)^{m - (k - i)} \right) \\ &= \eta^k (1 - \eta)^{n + m - k} \sum_{i=0}^k \binom{n}{i} \binom{m}{k - i} \\ &= \eta^k (1 - \eta)^{n + m - k} \binom{n + m}{k} \end{aligned}$$

was die Wahrscheinlichkeitsfunktion der $\text{Bin}(n + m, \eta)$ -Verteilung ist.

Beispiel 3.48: Betrachte den fünffachen und den siebenfachen Münzwurf, die unabhängig voneinander stattfinden. Sei X die Anzahl der Erfolge beim fünffachen Wurf, also $X \sim \text{Bin}(5, 1/2)$ und sei Y die Anzahl der Erfolge beim siebenfachen Wurf, also $Y \sim \text{Bin}(7, 1/2)$. Dann gilt für $Z = X + Y$, dass $Z \sim \text{Bin}(12, 1/2)$ und Z beschreibt die Anzahl der Erfolge beim zwölffachen Münzwurf.

Hypergeometrische Verteilung

Beispiel 3.49: Typisches Szenario als einführendes Beispiel.

Wir betrachten eine Urne mit N Kugeln, von denen R rot und $N - R$ weiß sind. Daraus ziehen wir ohne Zurücklegen und ohne Beachtung der Reihenfolge n Kugeln. Wir bezeichnen mit X die Anzahl der gezogenen roten Kugeln. Dann ist der Wertebereich $W = \{\max\{n - (N - R), 0\}, \dots, \min\{n, R\}\}$ und die Wahrscheinlichkeitsfunktion von X ist

$$p(r) = \mathbb{P}(X = r) = \frac{\binom{R}{r} \binom{N - R}{n - r}}{\binom{N}{n}} \quad \text{für } r = 0, \dots, n.$$

3 Zufallsvariablen und Verteilungen

Der Nenner ist klar, da er sich aus dem Urnenmodell ohne Zurücklegen und ohne Reihenfolge ergibt. Für den Zähler teilen wir die Urne im Geiste in die roten und in die weißen Kugeln. Aus der rotkugeligen Urne mit R Kugeln ziehen wir r Kugeln, aus der weißkugeligen Urne mit $N - R$ Kugeln ziehen wir $n - r$ Kugeln. Damit stellen wir sicher, dass genau r der n gezogenen Kugeln rot sind.

Diese Verteilung heißt hypergeometrische Verteilung mit den Parametern N, R und n und wir sagen $X \sim \text{Hyper}(N, R, n)$.

Der Erwartungswert ist

$$\begin{aligned}\mathbb{E}X &= \sum_{r=0}^n r \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N}{n}} \\ &= \frac{1}{\binom{N}{n}} \sum_{r=0}^n R \binom{R-1}{r-1} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \sum_{r=0}^n \binom{R-1}{r-1} \binom{N-R}{(n-1)-(r-1)} \\ &= \frac{R}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{R}{N} n.\end{aligned}$$

Dabei entspricht $\frac{R}{N}$ dem Anteil der roten Kugeln in der Urne vor dem Ziehen, die Wahrscheinlichkeit, als erste Kugel eine rote Kugel zu ziehen ist $\frac{R}{N}$ und damit ist der Erwartungswert analog dem einer Binomial-verteilten Zufallsvariable $Y \sim \text{Bin}(n, \frac{R}{N})$.

Zur Berechnung der Varianz betrachten wir

$$\begin{aligned}\mathbb{E}X^2 &= \frac{1}{\binom{N}{n}} \sum_{r=0}^n r^2 \binom{R}{r} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \sum_{r=0}^n r \binom{R-1}{r-1} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \left(\sum_{r=0}^n (r-1) \binom{R-1}{r-1} \binom{N-R}{n-r} + \sum_{r=0}^n \binom{R-1}{r-1} \binom{N-R}{n-r} \right) \\ &= \frac{R(R-1)}{\binom{N}{n}} \sum_{k=0}^n \binom{R-2}{k} \binom{N-R}{(n-2)-(k)} + \frac{R}{N} n \\ &= \frac{R(R-1)}{\binom{N}{n}} \binom{N-2}{n-2} + \frac{R}{N} n = \frac{R(R-1)}{N(N-1)} n(n-1) + \frac{R}{N} n\end{aligned}$$

und somit ist

$$\text{Var}X = \frac{R}{N} n \left(\frac{R-1}{N-1} (n-1) + 1 - \frac{R}{N} n \right) = \frac{R}{N} n \left(1 - \frac{R}{N} \right) \frac{N-n}{N-1}.$$

Auch hier sehen wir eine Ähnlichkeit zur Binomialverteilung. Wir haben $\text{Var}X = \text{Var}Y \cdot \frac{N-n}{N-1}$, d. h. für $n \geq 2$ wird die Varianz kleiner. Da wir ohne Zurücklegen ziehen, wissen wir schon etwas mehr über den neuen Urneninhalt und „nutzen diese Information“.

Geometrische Verteilung

Die geometrische Verteilung wurde bereits auf dem Übungszettel 5 in der Aufgabe 3 vorgestellt.

Zu beachten ist, dass manchmal die Anzahl der Misserfolge bis zum ersten Erfolg gezählt wird, manchmal andererseits die Anzahl der Versuche gezählt wird. Die entsprechenden Zufallsvariablen unterscheiden sich dann um 1.

Anmerkung: Die geometrische Verteilung wird angewendet, wenn es um Lebensdauer geht. Dabei wird eine Grundmenge an baugleichen Produkten oder eine Population modelliert, die mit der Wahrscheinlichkeit η ausfallen resp. sterben.

Negative Binomial-Verteilung

Die Negative Binomial-Verteilung ist eine Verallgemeinerung der geometrischen Verteilung. Sie wurde bereits auf dem Übungszettel 7 in der Aufgabe 3 vorgestellt.

Poisson-Verteilung

Beispiel 3.50: Typisches Szenario als einführendes Beispiel.

Die Poisson-Verteilung wird als Grenzverteilung der Binomialverteilung bei seltenen Ereignissen motiviert. Wir werden dies noch näher beim Thema der Approximationen und Grenzwertsätze kennenlernen.

Wir betrachten eine Zufallsvariable X , die die Anzahl der Erfolge angibt, die innerhalb einer Zeiteinheit eintreten. Die Wahrscheinlichkeitsfunktion ist

$$p(k) = (X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{für } k = 0, 1, 2, \dots$$

und sagen, dass X Poisson-verteilt ist mit Parameter λ bezeichnen die Verteilung mit $\text{Poisson}(\lambda)$.

Auch hier benötigen wir eine Reihe, bevor wir den Erwartungswert ausrechnen können.

Lemma 3.51: Für ein $x \in \mathbb{R}$ ist

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

die *Exponentialreihe* mit dem Exponenten x .

Dann ist der Erwartungswert

$$\mathbb{E}X = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

und

$$\mathbb{E}X^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 + \lambda$$

und damit die Varianz

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda.$$

3 Zufallsvariablen und Verteilungen

Faltung der Poisson-Verteilung Seien $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$, weiterhin sei $S_n := X_1 + \dots + X_n$. Dann ist $S_n \sim \text{Poisson}(n\lambda)$.

Multinomialverteilung

Die Multinomialverteilung ist das multivariate Analogon der Binomialverteilung. Dabei betrachten wir eine Experiment mit k möglichen Ausgängen $1, \dots, k$ und den Einzelwahrscheinlichkeiten η_1, \dots, η_k mit $\sum_{i=1}^k \eta_i = 1$. Wir ziehen n -mal mit Zurücklegen.

Bevor wir uns den Zufallsvariablen und der Wahrscheinlichkeitsfunktion zuwenden, benötigen wir noch die folgende Definition:

Definition 3.52: Seien $n, n_1, \dots, n_k \in \mathbb{N}_0$ mit $\sum_{i=1}^k n_i = n$. Dann ist

$$\binom{n}{n_1, \dots, n_k} = \frac{n!}{n_1! \dots n_k!} \quad (3.33)$$

der Multinomialkoeffizient. Der Vollständigkeit halber sagen wir $\binom{n}{n_1, \dots, n_k} = 0$, falls $\sum_{i=1}^k n_i \neq n$ oder $n_i < 0$ für ein $1 \leq i \leq k$.

Lemma 3.53: Für den Multinomialkoeffizienten gelten folgende Identitäten

$$\binom{n}{m, n-m} = \frac{n!}{m!(n-m)!} = \binom{n}{m} \quad (3.34)$$

und

$$\binom{n}{n_1, \dots, n_k} = \binom{n}{n_i} \cdot \binom{n-n_i}{n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k} \quad (3.35)$$

für alle $1 \leq i \leq k$.

Bezeichne nun N_i die Anzahl der Experimente mit dem Ausgang i . Die gemeinsame Wahrscheinlichkeitsfunktion ist

$$p_{N_1, \dots, N_k}(n_1, \dots, n_k) = \binom{n}{n_1, \dots, n_k} \eta_1^{n_1} \cdot \dots \cdot \eta_k^{n_k}.$$

Wir müssen zuerst nachprüfen, dass p_{N_1, \dots, N_k} tatsächlich eine Wahrscheinlichkeitsfunktion ist.

(i) $p_{N_1, \dots, N_k}(n_1, \dots, n_k)$ für alle $n_1, \dots, n_k \in \mathbb{N}_0$ wegen der Definition des Binomialkoeffizienten.

(ii) Die Normiertheit ergibt sich induktiv aus der Gleichung (3.35) durch sukzessive Berechnung der Marginalverteilungen.

Insbesondere folgt, dass die 1-dimensionalen Marginalverteilungen Binomialverteilungen mit dem entsprechenden Parameter η_i sind.

3.6 Spezielle/Wichtige stetige Verteilungen

Stetige Gleichverteilung

Die Gleichverteilung auf dem Intervall $[a, b] \subset \mathbb{R}$ ist gegeben durch

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \quad (3.36)$$

für alle $x \in \mathbb{R}$. Dann ist f eine Dichtefunktion da

(i) $f(x) \geq 0$ für alle $x \in \mathbb{R}$ und

$$(ii) \int \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b 1 dx = \frac{1}{b-a} \cdot x \Big|_a^b = \frac{1}{b-a} (b-a) = 1 \text{ gilt.}$$

Die Verteilungsfunktion ergibt sich dann als

$$F(x) = \begin{cases} 0 & \text{falls } x < a, \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b, \\ 1 & \text{falls } b < x. \end{cases} \quad (3.37)$$

Als Symbol verwenden wir $\mathcal{U}(a, b)$ und sagen, dass eine Zufallsvariable mit dieser Verteilung auf $[a, b]$ gleichverteilt ist. Eine auf $[0, 1]$ gleichverteilte Zufallsvariable nennen wir standardgleichverteilt.

Die stetige Gleichverteilung ist ein stetiges Analogon der Laplace-Wahrscheinlichkeit.

Sei nun X eine auf $[a, b]$ gleichverteilte Zufallsvariable. Dann ist $\mathbb{E}X = \frac{a+b}{2}$, da

$$\mathbb{E}X = \int x \cdot \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \cdot \frac{1}{2} x^2 \Big|_a^b = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{b+a}{2}$$

gilt. Für die Varianz berechnen wir zuerst

$$\mathbb{E}X^2 = \int x^2 \cdot \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \cdot \frac{1}{3} x^3 \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

und damit die Varianz

$$\begin{aligned} \text{Var}X &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{1}{12} (4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2) = \frac{1}{12} (a^2 - 2ab + b^2) = \frac{(a-b)^2}{12} \end{aligned}$$

Exponentialverteilung

Die Exponentialverteilung ist das stetige Analogon der geometrischen Verteilung. Siehe Übungszettel 7 mit Aufgabe 2 bezüglich der Definition.

Normalverteilung

Die Normalverteilung mit den Parametern μ und σ^2 , wobei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ sind, wird definiert durch die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.38)$$

Bereits zu zeigen, dass f eine Dichte ist, ist nicht ganz trivial und deswegen übergehen wir diesen Schritt. Als Symbol verwenden wir $\mathcal{N}(\mu, \sigma^2)$. Es gilt $\mathbb{E}X = \mu$ und $\text{Var}X = \sigma^2$, was wir ebenfalls nicht nachrechnen.

3 Zufallsvariablen und Verteilungen

Anmerkung: Die Normalverteilung hat eine große Bedeutung, da viele Messgrößen zumindest approximativ normalverteilt sind. Eine Erklärung werden wir mit dem *zentralen Grenzwertsatz* kennenlernen.

Anmerkung: Zuerst wurde die Normalverteilung von de Moivre definiert als Approximation der Binomialverteilung für große n . Gauß hat die Wichtigkeit der Normalverteilung gezeigt, weshalb die Dichtefunktion auch Gaußsche Glockenkurve genannt wird.

Die Standardnormalverteilung ist der Spezialfall mit $\mu = 0$ und $\sigma^2 = 1$, ihre Dichte ist

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3.39)$$

und ihre Verteilungsfunktion $\Phi(x)$, die sich nicht durch elementare Funktionen darstellen lässt.

Aus der (Achsen-)Symmetrie von φ bezüglich $x = 0$, d. h.

$$\varphi(-x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-x)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$$

folgt die Punktsymmetrie von Φ bezüglich des Punktes $(0, 1/2)$ und es gilt

$$\Phi(-x) = 1 - \Phi(x). \quad (3.40)$$

Standardisierung Die Verteilungsfunktion F einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen X lässt sich durch Umrechnen aus der Verteilungsfunktion einer Standardnormal-verteilten Zufallsvariablen ausdrücken. Es gilt

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right); \quad (3.41)$$

die Zufallsvariable $Z = \frac{X - \mu}{\sigma}$ ist $\mathcal{N}(0, 1)$ -verteilt.

Additivität

Satz 3.54: Seien X und Y unabhängige Zufallsvariablen, die $\mathcal{N}(\mu_X, \sigma_X^2)$ resp. $\mathcal{N}(\mu_Y, \sigma_Y^2)$ -verteilt sind. Dann ist deren Summe $Z = X + Y$ wieder normalverteilt, und zwar $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Diese Eigenschaft wird *Faltungsstabilität* genannt.

Multivariate Normalverteilung Ein Zufallsvektor $X = (X_1, \dots, X_n)^T$ mit der Dichte

$$f_X(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n} \cdot \sqrt{\det(\Sigma)}} \cdot e^{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (3.42)$$

heißt multivariat normalverteilt mit den Parametern $\mu = (\mu_1, \dots, \mu_n)^T$ und

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \dots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & \ddots & \vdots \\ \rho_{1n}\sigma_n\sigma_1 & \dots & \rho_{n,n-1}\sigma_n\sigma_{n-1} & \sigma_n^2 \end{pmatrix}, \quad (3.43)$$

3.6 Spezielle/Wichtige stetige Verteilungen

wobei $\sigma_1^2, \dots, \sigma_n^2$ die Varianzen der einzelnen Koordinaten sind und $\rho_{ij} < 0$ die Korrelationskoeffizienten von X_i und X_j für $1 \leq i, j \leq n$ und $i \neq j$.

- Anmerkung:** (a) Die Marginalverteilungen von X_1 und X_2 sind $\mathcal{N}(\mu_1, \sigma_1^2)$ resp. $\mathcal{N}(\mu_2, \sigma_2^2)$.
(b) Zu beachten ist, dass X_1 und X_2 im Allgemeinen nicht unabhängig sind. Dies ist den Marginalverteilungen nicht anzusehen.
(c) Gilt hingegen Unabhängigkeit, so ist Σ eine Diagonalmatrix. Unter der zusätzlichen Bedingung $\sigma_1 = \sigma_2$ ist die Dichte rotationsinvariant.
(d) Wie im eindimensionalen Fall durch μ und σ^2 wird im mehrdimensionalen Fall die Verteilung eines normalverteilten Zufallsvektors eindeutig durch $(\mu_1, \dots, \mu_n)^T$ und Σ bestimmt.

3 Zufallsvariablen und Verteilungen

4 Abschätzungen und Grenzwertsätze

Oftmals lässt sich die Wahrscheinlichkeit eines bestimmten Ereignisses nicht in geschlossener Form angeben bzw. die Berechnung ist sehr aufwendig. In diesem Fall helfen vielfach Ungleichungen, um Schranken für die Wahrscheinlichkeit zu bestimmen.

4.1 Ungleichungen

Satz 4.1 (Markov-Ungleichung): Sei X eine Zufallsvariable mit $\mathbb{E}|X| < \infty$. Dann gilt

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a} \mathbb{E}|X| \quad (4.1)$$

für alle reellen $a > 0$.

BEWEIS Sei X diskret mit der Wahrscheinlichkeitsfunktion p . Für $x \in \mathbb{R}$ mit $|x| \geq a$ gilt $\frac{|x|}{a} \geq 1$ und damit

$$\mathbb{P}(|X| \geq a) = \sum_{x:|x| \geq a} p(x) \leq \sum_{x:|x| \geq a} \frac{|x|}{a} p(x) \leq \sum_x \frac{|x|}{a} p(x) = \frac{1}{a} \sum_x |x| p(x) = \frac{1}{a} \mathbb{E}|X|.$$

Sei nun X stetig mit der Dichte f_X . Dann folgt mit analogen Argumenten zu oben

$$\begin{aligned} \mathbb{P}(|X| \geq a) &= \int_{-\infty}^{\infty} \mathbb{1}_{\{|x| \geq a\}} f_X(x) dx \leq \int_{-\infty}^{\infty} \frac{|x|}{a} \mathbb{1}_{\{|x| \geq a\}} f_X(x) dx \\ &\leq \int_{-\infty}^{\infty} \frac{|x|}{a} f_X(x) dx = \frac{1}{a} \int_{-\infty}^{\infty} |x| f_X(x) dx \\ &= \frac{1}{a} \mathbb{E}|X|. \quad \square \end{aligned}$$

Beispiel 4.2: Betrachte den n -fachen Wurf mit einer fairen Münze, sei X die Zufallsvariable für die Anzahl der Würfe mit dem Ausgang Kopf. Dann ist $X \sim \text{Bin}(n, 1/2)$. Die möglichen Ausgänge sind $X(\Omega) = \{0, 1, \dots, n\}$. Damit ist $|X| = X$ und wir kennen auch $\mathbb{E}X = n/2$.

Die Wahrscheinlichkeit, in mindestens 75% der Würfe Kopf zu erzielen, ist beschränkt durch

$$\mathbb{P}(X \geq \frac{3}{4} \cdot n) \leq \frac{\mathbb{E}X}{\frac{3}{4} \cdot n} = \frac{n \cdot \frac{1}{2}}{\frac{3}{4} \cdot n} = \frac{2}{3}$$

wobei die Markov-Ungleichung mit $a = \frac{3}{4} \cdot n$ verwendet wurde.

4 Abschätzungen und Grenzwertsätze

Satz 4.3 (Tschebyschev-Ungleichung): Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann gilt

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{1}{a^2} \text{Var}X \quad (4.2)$$

für alle reellen $a > 0$.

BEWEIS Wir betrachten die Zufallsvariable $|X - \mathbb{E}X|^2$ und wenden auf diese die Markov-Ungleichung (4.1) an. Dann ist

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2) \leq \frac{1}{a^2} \mathbb{E}(|X - \mathbb{E}X|^2) = \frac{1}{a^2} \text{Var}X. \quad \square$$

Anmerkung: Die Tschebyschev-Ungleichung gibt eine einfache Abschätzung für die Wahrscheinlichkeit an, dass die Zufallsvariable vom Erwartungswert abweicht.

Korollar 4.4: Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann gilt $\text{Var}X = 0$ genau dann, wenn $\mathbb{P}(X = \mathbb{E}X) = 1$.

BEWEIS Falls $\text{Var}X = 0$ gilt, so folgt aus der Tschebyschev-Ungleichung (4.2), dass

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = 0$$

für alle $\varepsilon > 0$ und damit

$$1 - \mathbb{P}(X = \mathbb{E}X) = \mathbb{P}(|X - \mathbb{E}X| > 0) = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = 0,$$

woraus über das Gegenereignis $\mathbb{P}(X = \mathbb{E}X) = 1$ folgt.

Andererseits ergibt sich aus $\mathbb{P}(X = \mathbb{E}X) = 1$, dass X diskret ist. Dann gilt

$$\begin{aligned} (\mathbb{E}X)^2 &= \left(\sum_x x \mathbb{P}(X = x) \right)^2 \\ &= \mathbb{E}X \cdot \mathbb{P}(X = \mathbb{E}X) \cdot \mathbb{E}X \cdot \mathbb{P}(X = \mathbb{E}X) \\ &= (\mathbb{E}X)^2 \mathbb{P}(X = \mathbb{E}X) = \sum_x x^2 \mathbb{P}(X = x) \\ &= \mathbb{E}X^2 \end{aligned}$$

wobei die Transformationsformel für Erwartungswerte (Satz 3.14) verwendet wurde. Nun haben wir

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0, \quad \square$$

also die Aussage.

Beispiel 4.5: Wir betrachten eine Messmethode, die fehlerhafte Ergebnisse liefert. Dabei sei $\mu \in \mathbb{R}$ zu messen, X_i der Messfehler und $Z_i := \mu + X_i$ das Messergebnis der i -ten Messung für $1 \leq i \leq n$. Wir gehen davon aus, dass X_1, \dots, X_n unabhängig und identisch verteilt sind (i.i.d. Folge).

Wir berechnen die Anzahl der erforderlichen Messungen, die notwendig sind, damit das arithmetische Mittel $Y_n := \frac{1}{n} \sum_{i=1}^n Z_i$ der zufälligen Messwerte mit einer Wahrscheinlichkeit von höchstens 0,1 um mehr als 1 von dem unbekanntem Wert μ abweicht.

(a) Wir wissen über die Verteilung der X_i nur, dass $\mathbb{E}X_i = 0$ und $\text{Var}X_i = 1$. Wegen der Linearität des Erwartungswertes und der Unabhängigkeit wissen wir, dass

$$\mathbb{E}Y_n = \mu \quad \text{und} \quad \text{Var}Y_n = \frac{1}{n}$$

und damit folgt aus der Tschebyschev-Ungleichung (4.2)

$$\mathbb{P}(|Y_n - \mu| > 1) \leq \frac{1}{n},$$

also $\mathbb{P}(|Y_n - \mu| > 1) \leq 0,1$, falls $1/n \leq 0,1$ ist, d. h. $n \geq 10$.

(b) Haben wir hingegen Information über die Verteilung der Fehler, so können wir diese nutzen. Seien hier die Messfehler $X_i \sim \mathcal{N}(0,1)$. Dann gilt

$$Y_n \sim \mathcal{N}(\mu, 1/n) \quad \text{und normiert} \quad \sqrt{n}(Y_n - \mu) \sim \mathcal{N}(0,1).$$

Dann gilt

$$\begin{aligned} \mathbb{P}(|Y_n - \mu| \geq 1) &= \mathbb{P}(\sqrt{n}|Y_n - \mu| \geq \sqrt{n}) \\ &= \mathbb{P}(\sqrt{n}(Y_n - \mu) \leq -\sqrt{n}) + \mathbb{P}(\sqrt{n}(Y_n - \mu) \geq \sqrt{n}) \\ &= \Phi(-\sqrt{n}) + (1 - \Phi(\sqrt{n})) = (1 - \Phi(\sqrt{n})) + (1 - \Phi(\sqrt{n})) \\ &= 2(1 - \Phi(\sqrt{n})), \end{aligned}$$

wobei wir die Punktsymmetrie der Verteilungsfunktion Φ der Normalverteilung verwendet haben. Es gilt

$$2(1 - \Phi(\sqrt{n})) \leq 0,1 \quad \iff \quad \Phi(\sqrt{n}) \geq 0,95 \quad \iff \quad \sqrt{n} \geq 1,645$$

d. h. $n \geq 3$.

Wir sehen hier, dass Verteilungsinformation zu einer viel höheren Genauigkeit führt.

Anmerkung: Sowohl die Markov- als auch die Tschebyschev-Ungleichung kommen ohne Verteilungsannahme aus. Das macht sie zu einem beliebten Instrument. Der Nachteil davon ist allerdings, dass die Abschätzungen, die sie liefern, relativ grob sind. Sind zusätzlich Annahmen über die Verteilung der Zufallsvariablen X möglich, so lassen sich genauere Abschätzungen erreichen.

4.2 Grenzwertsätze

Grenzwertsätze sind Aussagen, die es uns ermöglichen, Annäherungen zur Berechnung von Wahrscheinlichkeiten zu verwenden. Das ist immer dann sinnvoll, wenn die exakte Berechnung zu kompliziert ist. Der erkaufte Nachteil ist, dass wir eine gewisse Ungenauigkeit bekommen.

Betrachte eine $\text{Bin}(n, \eta)$ -verteilte Zufallsvariable. Die Berechnung der Binomialkoeffizienten ist nicht einfach, wenn n groß ist. Die Poisson-Approximation wird für große Werte von n und sehr kleine Werte von p angewendet, also seltene Ereignisse.

Satz 4.6 (Poisson-Grenzwertsatz): Es sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, wobei $X_n \sim \text{Bin}(n, p_n)$ -verteilt ist. Falls ein $\lambda > 0$ existiert so dass $n p_n \rightarrow \lambda$ für $n \rightarrow \infty$, so gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (4.3)$$

also bekommen wir eine Poisson-Verteilung.

4 Abschätzungen und Grenzwertsätze

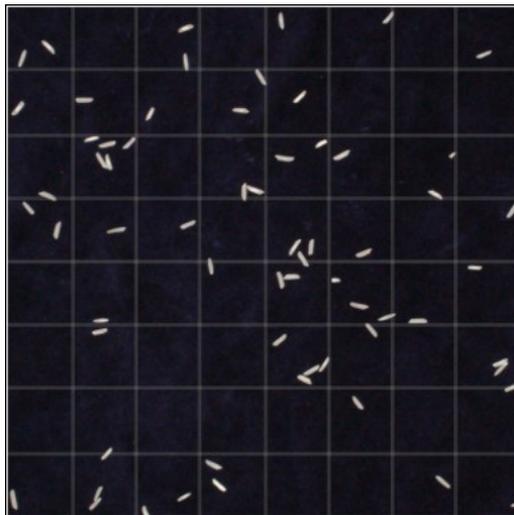


Abbildung 4.1: Reiskörner auf einem Schachbrett zur Darstellung der Poisson-Approximation aus dem Beispiel 4.7 (b).

Anmerkung: Zur Genauigkeit betrachten wir die Vergleichstabelle

Verteilung	Wahrscheinlichkeit				
	0	1	2	3	≥ 4
Bin(5, 0.3)	0.1681	0.3602	0.3087	0.1323	0.0308
Bin(50, 0.03)	0.2181	0.3372	0.2555	0.1264	0.0628
Bin(500, 0.003)	0.2226	0.3349	0.2515	0.1256	0.0654
Bin(5000, 0.0003)	0.2231	0.3347	0.2511	0.1255	0.0656
Poi(1.5)	0.2231	0.3347	0.2510	0.1255	0.0656

Beispiel 4.7 (Gesetz der kleinen Zahlen): (a) Betrachte eine Folge X_1, \dots, X_n identisch verteilter Zufallsvariablen, die voneinander unabhängig sind und n mögliche Ausgänge haben, d. h. $|W| = n$, die gleichverteilt angenommen werden. Dabei interessieren wir uns dafür, einen bestimmten Ausgang k -mal zu erzielen, sei Z_n die zugehörige Zufallsvariable, d. h. $Z_n(k) := \sum_{i=1}^n \mathbb{1}_{\{X_i=k\}}$. Dann ist $Z_n \sim \text{Bin}(n, 1/n)$ -verteilt. Mit dem Satz 4.6 gilt, dass Z_n ungefähr Poi(1)-verteilt ist.

(b) Wir beschäftigen wir uns mit einem Schachbrett und 64 Reiskörnern, die wir auf das Schachbrett werfen. Dazu untersuchen wir die Zufallsvariable Z_{64} aus dem Teil (a), die angibt, auf wie vielen Schachbrettfeldern jeweils die Anzahl k der Reiskörner zu liegen kommt. (Siehe hierzu Bild 4.1.)

k	Z_{64}	$Z_{64}/64$	Poi(1)
0	23	0.3594	0.3679
1	25	0.3906	0.3679
2	12	0.1875	0.1839
3	2	0.0312	0.0613
4	1	0.0156	0.0153
5	1	0.0156	0.0031
Σ	64	1	0.9994

(c) Wir gehen (virtuell) ins Casino und spielen Roulette. Hier sind 37 Ausgänge möglich. Betrachte immer einen Satz von 37 Spielen und schaue wie viele Ausgänge 0 mal, 1 mal, usw. vorkommen.

Wir simulieren 10000 solcher Sätze, bezeichne mit \bar{Z} das arithmetische Mittel der Zufallsvariablen Z aus den 10000 Läufen. Dann erhalten wir

k	\bar{Z}	$\bar{Z}/37$	Poi(1)
0	13.4364	0.3631	0.3679
1	13.7757	0.3723	0.3679
2	6.9008	0.1865	0.1839
3	2.3293	0.0630	0.0613
4	0.5310	0.0144	0.0153
5	0.0268	0.0007	0.0031
Σ	37	1	0.9994

Jedes Mal wird im Schnitt ein Drittel der möglichen Ausgänge nicht erreicht. Dies wird von Casino-Gängern auch das *Gesetz des Drittels* genannt. Allerdings ist damit *nicht* gesagt, dass es sich immer um die Gleichen handle. Im Gegensatz: Zählen wir alle Vorkommnisse zusammen, so bekommen wir

k	0	1	2	3	4	5	6
Y	9969	10021	9816	9921	10019	9833	9918
k	7	8	9	10	11	12	
Y	9850	9914	10041	10097	10184	9935	
k	13	14	15	16	17	18	
Y	10026	10068	10283	10165	10169	10216	
k	19	20	21	22	23	24	
Y	10014	9943	9938	9932	9817	10159	
k	25	26	27	28	29	30	
Y	9879	10053	9904	10141	9912	10005	
k	31	32	33	34	35	36	
Y	10001	10049	9945	9830	10011	10022	

Beachte dabei, dass wir insgesamt $10000 \cdot 37$ Roulette-Spiele simuliert haben. Da jeder Ausgang gleich wahrscheinlich ist, also $p = 1/37$ folgt für die $\text{Bin}(370000, 1/37)$ -verteilte Zufallsvariable Y , dass $\mathbb{E}Y = 370000 \cdot 1/37 = 10000$ ist.

4.2.1 Gesetze der großen Zahlen

Den Effekt der Mittelung, den wir im letzten Beispiel zum Schluss gesehen haben, wollen wir hier genauer untersuchen.

Fragestellung Welcher Zusammenhang besteht allgemein zwischen dem arithmetischen Mittel und dem Erwartungswert? Zuvor benötigen wir jedoch einige Begriffe zur Konvergenz.

Definition 4.8 (Stochastische Konvergenz): Seien $X_1, X_2 \dots$ und X Zufallsvariablen auf dem gleichen Wahrscheinlichkeitsraum. Wir sagen, dass X_n gegen X stochastisch konvergiert, geschrieben

$$X_n \xrightarrow{st} X,$$

falls

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 \quad (4.4)$$

für alle $\varepsilon > 0$.

Satz 4.9 (Schwaches Gesetz der großen Zahlen): Sei $X_1, X_2 \dots$ eine Folge von Zufallsvariablen mit dem gleichen Erwartungswert $\mu := \mathbb{E}X_i$ und mit $\mathbb{E}X_i^2 < \infty$ für alle $i \geq 1$ und sei $Z_n := 1/n \sum_{i=1}^n X_i$. Falls

$$\lim_{n \rightarrow \infty} \text{Var}Z_n = 0,$$

dann konvergiert die Folge $X_1, X_2 \dots$ stochastisch gegen μ , i. e.

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Z_n - \mu| > \varepsilon) = 0 \quad (4.5)$$

für alle $\varepsilon > 0$.

BEWEIS Da $\mathbb{E}X_i < \infty$ für alle $i \geq 1$ gilt, haben wir

$$\mathbb{E}Z_n = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

Mit der Tschebyschevschen Ungleichung aus Satz 4.3 gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\text{Var}Z_n}{\varepsilon^2} = \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} \text{Var}Z_n = 0. \quad (4.6)$$

für alle $\varepsilon > 0$. □

Anmerkung: Die Bedingungen des Satzes 4.9 sind insbesondere erfüllt, wenn die Zufallsvariablen X_1, X_2, \dots unabhängig sind und die gleiche Varianz σ^2 haben, da dann

$$\lim_{n \rightarrow \infty} \text{Var}Z_n = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var}X_i = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

gilt.

Definition 4.10 (Fast sichere Konvergenz): Seien $X_1, X_2 \dots$ und X Zufallsvariablen auf dem gleichen Wahrscheinlichkeitsraum. Dann konvergiert X_n gegen X fast sicher, geschrieben

$$X_n \xrightarrow{f.s.} X,$$

falls

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1 \quad (4.7)$$

ist.

Satz 4.11 (Starkes Gesetz der großen Zahlen): Sei X_1, X_2, \dots eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $\mathbb{E}|X_1| < \infty$ und sei $\mu = \mathbb{E}X_1$. Sei weiter $Z_n := 1/n \sum_{i=1}^n X_i$ das arithmetische Mittel. Dann konvergiert die Folge X_1, X_2, \dots fast sicher gegen μ , i. e.

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} Z_n = \mu \right) = 1. \quad (4.8)$$

Anmerkung (Unterschied zwischen den Konvergenzarten): Das schwache Gesetz sagt aus, dass für große n das arithmetische Mittel nah an μ ist. Deswegen gilt, dass $|Z_n(\omega) - \mu| > \varepsilon$ für beliebige ω unendlich oft möglich ist, aber in unregelmäßigen Abständen (und auch immer seltener); die Wahrscheinlichkeit wird immer kleiner.

Das starke Gesetz sagt aus, dass wir zu jedem $\varepsilon > 0$ ein N finden, das groß genug ist, so dass für alle $n \geq N$ die Schranke $|Z_n - \mu| < \varepsilon$ eingehalten wird. Dies gilt dann für fast alle ω ¹, so dass wir sagen können, dass es mit Wahrscheinlichkeit 1 gilt.

Anmerkung: Aus der fast sicheren Konvergenz folgt die stochastische Konvergenz. Das Gegenteil gilt im Allgemeinen nicht.

Beispiel 4.12: Sei X_1, X_2, \dots eine Folge unabhängiger Zufallsvariablen, wobei $X_n \sim \text{Bin}(1, 1/n)$ -verteilt ist. Es gilt

$$X_n \xrightarrow{st} 0.$$

Andererseits nehme an, es gäbe ein N so, dass $\mathbb{P} \left(\{\omega : X_n(\omega) = 0 \text{ für alle } n \geq N\} \right) = 1$. Dann würde auch $X_n \xrightarrow{f.s.} 0$ gelten. Wir betrachten das obere Ereignis genauer

$$\mathbb{P} \left(\{\omega : X_n(\omega) = 0 \text{ für alle } n \geq N\} \right) = \frac{N-1}{N} \cdot \frac{N}{N+1} \cdot \frac{N+1}{N+2} \cdot \dots = 0,$$

d. h. es gibt *immer wieder* ein (zufälliges) i , für welches X_i nicht 0 annimmt.

Beispiel 4.13 (Monte-Carlo-Integration im Buffonschen Nadelexperiment): Hier wird das starke Gesetz der großen Zahlen zur numerischen Berechnung angewandt. Betrachte hierzu die Menge

$$K = \{(x, y) : x \in \mathbb{Z} \text{ und } y \in \mathbb{R}\} \subset \mathbb{R}^2$$

der parallelen vertikalen Geraden mit Abstand 1 in der euklidischen Ebene.

Werfe nun eine Nadel der Länge 1 zufällig in die Ebene, d. h. betrachte zwei Zufallsvariablen S und T zur Beschreibung der Lage der Nadel, wobei

- (a) S der senkrechte Abstand des Nadelmittelpunktes zur nächsten linken Geraden aus K ist,
- (b) T für den Winkel steht, den die Nadel zum Lot auf die Geraden aus K bildet,
- (c) S und T unabhängig sind,
- (d) und $S \sim \mathcal{U}[0, 1]$ und $T \sim \mathcal{U}[-\pi/2, \pi/2]$ gilt, d. h. wir kennen die Dichten $f_S(x) = \mathbb{1}_{[0,1]}(x)$ und $f_T(x) = \frac{1}{\pi} \mathbb{1}_{[-\pi/2, \pi/2]}(x)$.

Wir interessieren uns für die Wahrscheinlichkeit, dass die Nadel eine der Geraden aus K schneidet, also für das Ereignis

$$A = \{0 < S < 1/2 \cdot \cos T\} \cup \{1 - 1/2 \cdot \cos T < S < 1\}.$$

¹Mit *fast alle* meinen wir, dass einzelne Werte z aus der Wertemengen von Z_n , für die $\mathbb{P}(Z_n = z) = 0$ gilt, aus der Reihe tanzen dürfen. Gedacht sei dabei z. B. an stetige Zufallsvariablen.

4 Abschätzungen und Grenzwertsätze

Dann gilt

$$\begin{aligned}
 \mathbb{P}(A) &= \mathbb{P}(0 < S < 1/2 \cdot \cos T) + \mathbb{P}(1 - 1/2 \cdot \cos T < S < 1) \\
 &= \int_{-\pi/2}^{\pi/2} P(0 < S < 1/2 \cdot \cos t) f_T(t) dt + \int_{-\pi/2}^{\pi/2} P(1 - 1/2 \cdot \cos t < S < 1) f_T(t) dt \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \left(\int_0^{1/2 \cos t} f_S(s) ds \right) dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \left(\int_{1-1/2 \cos t}^1 f_S(s) ds \right) dt \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t dt = \frac{1}{\pi} \cdot \sin t \Big|_{-\pi/2}^{\pi/2} = \frac{2}{\pi}.
 \end{aligned}$$

Betrachte die Folge von unabhängigen Zufallsvektoren $(S_1, T_1), (S_2, T_2), \dots$ und definiere die Folge

$$X_n := \mathbb{1}_{\{S_n < 1/2 \cos T_n\}} + \mathbb{1}_{\{1 - 1/2 \cos T_n < S_n\}}$$

dann sind X_1, X_2, \dots unabhängig und identisch verteilt mit $\mathbb{E}X_1 = 2/\pi$.

Nach dem starken Gesetz der großen Zahlen gilt für das arithmetische Mittel $Z_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} 2/\pi$. Damit lässt sich für große n eine Näherung der Zahl π durch Simulation berechnen.

4.2.2 Zentraler Grenzwertsatz

Eine weitere Art der Konvergenz öffnet das Schloss zum zentralen Grenzwertsatz

Definition 4.14 (Konvergenz in Verteilung): Sei X_1, X_2, \dots eine Folge von Zufallsvariablen. Sei X eine weitere Zufallsvariable. Seien F_1, F_2, \dots und F ihre jeweiligen Verteilungsfunktionen. Dann konvergiert X_n in Verteilung gegen X , geschrieben

$$X_n \xrightarrow{D} X,$$

wenn

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

für alle $x \in \mathbb{R}$, an denen F stetig ist, gilt. Dabei steht D für *distribution*.

Anmerkung (Zusammenhang zwischen den Konvergenzarten): Für eine Folge X_1, X_2, \dots von Zufallsvariablen und die Zufallsvariable X gelten die Implikationen

$$\left(X_n \xrightarrow{f.s.} X \right) \implies \left(X_n \xrightarrow{st} X \right) \implies \left(X_n \xrightarrow{D} X \right)$$

Die Umkehrung gilt im Allgemeinen nicht.

Satz 4.15 (Satz von Moivre-Laplace): Sei $\eta \in (0,1)$ und sei X_1, X_2, \dots eine Folge von Zufallsvariablen, die unabhängig und $\text{Bin}(1, \eta)$ -verteilt und auf dem gleichen Wahrscheinlichkeitsraum definiert sind. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_1 + \dots + X_n - n\eta}{\sqrt{n\eta(1-\eta)}} \leq x \right) = \Phi(x) \quad (4.9)$$

für alle $x \in \mathbb{R}$, d. h. die Folge $(Y_n)_{n \in \mathbb{N}}$ mit $Y_n := \sum_{i=1}^n X_i \sim \text{Bin}(n, \eta)$ in Verteilung konvergiert; wir normieren Y_n und erhalten Konvergenz gegen die Standardnormalverteilung.

Beispiel 4.16: Betrachte Münzwurf mit einer fairen Münze, sagen wir 200-fachen. Wir fragen uns, wie groß die Wahrscheinlichkeit ist, dass wir mehr als 105-mal das Ereignis Kopf finden. Genau so finden wir die Wahrscheinlichkeit für weniger als 95 indem wir die Rolle von Kopf und Zahl vertauschen. Wir basteln eine Zufallsvariable

$$Z := \sum_{i=1}^{200} \mathbb{1}_{\{\text{i.ter Wurf} = \text{Kopf}\}}$$

und wissen $Z \sim \text{Bin}(200, 1/2)$. Damit suchen wir nach der Wahrscheinlichkeit

$$\mathbb{P}(Z > 105) = 1 - \mathbb{P}(Z \leq 105) = 1 - \mathbb{P}(Z - 100 \leq 5) = 1 - \mathbb{P} \left(\frac{Z - 100}{\sqrt{50}} \leq \frac{5}{\sqrt{50}} \right),$$

was der Form im Satz 4.15 entspricht. Somit ist

$$\mathbb{P}(Z > 105) \approx 1 - \Phi(5/\sqrt{50}) \approx 0.2398^2$$

Satz 4.17 (Zentraler Grenzwertsatz): Sei X_1, X_2, \dots eine Folge von Zufallsvariablen, die unabhängig und auf dem gleichen Wahrscheinlichkeitsraum definiert sind. Weiterhin gelte $\mu := \mathbb{E}X_n$ und $\infty > \sigma^2 := \text{Var} X_n > 0$ für alle $n \geq 1$. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x) \quad (4.10)$$

für alle $x \in \mathbb{R}$.

Korollar 4.18: Nehme die Voraussetzung von Satz 4.17. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < x \right) = \Phi(x)$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a).$$

Beispiel 4.19 (Messungen): Wir kommen zurück auf das Beispiel 4.5 und beschäftigen uns mit einer ungenauen Messmethode. Betrachte für $1 \leq i \leq n$ das Messergebnis der i -ten Messung in Form der Zufallsvariablen $Z_i := \mu + X_i$, wobei μ zu messen ist und X_i der Messfehler ist, σ die Varianz des Messfehlers. Alle Messfehler seien unabhängig und identisch verteilt. Die Wahrscheinlichkeit, dass

²Zu beachten ist, dass die Verteilungsfunktion der Normalverteilung entweder in einem Tabellenwerk nachzusehen oder implementiert in einer Software (z. B. in R) auszulesen ist.

4 Abschätzungen und Grenzwertsätze

das arithmetische Mittel Y_n der Messwerte Z_1, \dots, Z_n vom „echten Wert μ “ um mehr als ε abweicht, ergibt sich mit dem Satz 4.17

$$\begin{aligned}\mathbb{P}(|Y_n - \mu| > \varepsilon) &= \mathbb{P}\left(\left|\frac{Z_1 + \dots + Z_n - n\mu}{n}\right| > \varepsilon\right) \\ &= \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n}\right| > \varepsilon\right) \\ &= \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| > \varepsilon \cdot \frac{\sqrt{n}}{\sigma}\right) \\ &= 1 - \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| \leq \varepsilon \cdot \frac{\sqrt{n}}{\sigma}\right) \\ &= 1 - \mathbb{P}\left(-\varepsilon \cdot \frac{\sqrt{n}}{\sigma} \leq \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \leq \varepsilon \cdot \frac{\sqrt{n}}{\sigma}\right) \\ &\approx 1 - \left(\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \\ &= 2 \left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right).\end{aligned}$$

Für eine genügende Anzahl n von Messungen benötigen wir keine Annahme der Normalverteilung für die Messfehler.

Bis hierher relevanter Stoff für die erste Klausur.

(Für die Nachklausur ist der Stoff der ganzen Vorlesung von Bedeutung.)

5 Nachtrag zu Zufallsvariablen und Verteilungen

Dieser Abschnitt enthält Inhalte aus Kapitel 3, welche wegen des frühen Klausurtermins zurückgehalten wurden, um die Grenzwertsätze rechtzeitig vorstellen zu können.

5.1 Transformation von Dichten

Wir hatten uns Transformationen bei der Berechnung von Erwartungswerten angesehen. In diesem Abschnitt werden wir uns mit Dichten und deren Umrechnungen beschäftigen.

5.1.1 Transformation eindimensionaler Zufallsvariablen

Satz 5.1 (Transformationsformel für Dichten): Sei X eine stetige Zufallsvariable mit Werten in $I \subset \mathbb{R}$, wobei I ein offenes Intervall ist, und der Dichte f_X . Ist $J \subset \mathbb{R}$ ebenfalls ein offenes Intervall und $u : I \rightarrow J$ bijektiv mit u und u^{-1} stetig diffbar, so besitzt $Y := u(X)$ die Dichte

$$f_Y(y) = f_X\left(u^{-1}(y)\right) \left| \frac{d}{dy} u^{-1}(y) \right| \quad (5.1)$$

für $y \in J$ und $f_Y(y) = 0$ für $y \in \mathbb{R} \setminus J$.

Beispiel 5.2: Sei $X \sim \mathcal{U}(0,1]$ und $Y := u(X) = \frac{1}{X}$. Die Funktion $u(x) = \frac{1}{x}$ ist bijektiv und stetig differenzierbar, $u^{-1}(y) = \frac{1}{y}$ ebenso. Außerdem ist $\frac{d}{dy} u^{-1}(y) = -\frac{1}{y^2}$. Dann hat $Y = u(X)$ die Dichte

$$f_Y(y) = f_X\left(\frac{1}{y}\right) \cdot \frac{1}{y^2} = \mathbb{1}_{(0,1]}\left(\frac{1}{y}\right) \cdot \frac{1}{y^2} = \frac{1}{y^2} \mathbb{1}_{[1,\infty)}(y) \quad (5.2)$$

Beispiel 5.3 (Kubische Transformation): Sei X eine stetige Zufallsvariable und $Y := u(X) = X^3$. Dann ist die Funktion $u(x) = x^3$ bijektiv und stetig diffbar, ihre Umkehrung

$$u^{-1}(y) = |y|^{\frac{1}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right)$$

ebenfalls mit

$$\frac{d}{dy} u^{-1}(y) = \frac{1}{3} |y|^{-\frac{2}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right).$$

Die Dichte von Y ist

$$\begin{aligned} f_Y(y) &= f_X\left(|y|^{\frac{1}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right)\right) \cdot \frac{1}{3} |y|^{-\frac{2}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right) \\ &= \begin{cases} f_X\left(y^{\frac{1}{3}}\right) \cdot \frac{1}{3} \cdot y^{-\frac{2}{3}} & \text{falls } y \geq 0, \\ -f_X\left(-|y|^{\frac{1}{3}}\right) \cdot \frac{1}{3} \cdot |y|^{-\frac{2}{3}} & \text{falls } y < 0. \end{cases} \end{aligned}$$

5 Nachtrag zu Zufallsvariablen und Verteilungen

Die Funktionen sehen komplizierter aus als sie sind. Wir müssen einfach unterscheiden, ob y negativ ist oder nicht, da Wurzeln aus negativen Zahlen höchstens über die Betragsschreibweise, wie wir sie verwenden, definiert sind.

Beispiel 5.4 (Quadrierung): Sei X eine stetige Zufallsvariable und wir untersuchen die Transformation $Y := u(X) = X^2$. Dann ist die Funktion $u(x) = x^2$ für $u \geq 0$ bijektiv mit $u^{-1}(y) = \sqrt{y}$ und u und u^{-1} sind stetig diffbar. Dann ist

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}.$$

Bierbei haben wir die Bedingung $X > 0$. Aber: Für eine beliebige stetige Zufallsvariable X haben wir die Verteilungsfunktion

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(|X| \leq \sqrt{y}) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}), \end{aligned}$$

die stetig und differenzierbar ist, und erhalten die Dichte

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

An diesem Beispiel sehen wir, dass wir nicht ohne Nachdenken einfach die Formel verwenden können, sondern uns genau über die Voraussetzungen klar sein sollen.

5.1.2 Transformation von Zufallsvektoren

Im Weiteren wollen wir uns mit Transformationen beschäftigen, die mehrere Zufallsvariablen einbeziehen. Wir können auch sagen, wir haben einen Zufallsvektor und betrachten die mehrdimensionale Transformation. Die Bildung einer Summe ist die Faltung der Dichten, die wir bereits kennengelernt haben.

Satz 5.5 (Transformationsformel für gemeinsame Dichten): Sei $X = (X_1, \dots, X_n)$ ein stetiger Zufallsvektor mit der Dichte f_X . Seien $A, B \subset \mathbb{R}^n$ offene zusammenhängende Teilmengen, wobei X nur Werte in A annimmt. Sei weiter $u : A \rightarrow B$ eine bijektive Abbildung mit u, u^{-1} stetig differenzierbar. Dann besitzt $Y = (Y_1, \dots, Y_n) := u(X)$ die Dichte

$$f_Y(y_1, \dots, y_n) = f_X(u^{-1}(y_1, \dots, y_n)) \left| \det(J_{u^{-1}}(y_1, \dots, y_n)) \right| \cdot \mathbb{1}_B(y_1, \dots, y_n), \quad (5.3)$$

wobei $J_{u^{-1}}$ die Jacobi-Matrix von u^{-1} ist.

Anmerkung (Jacobi-Matrix): Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine differenzierbare Funktion mit den Komponenten f_1, \dots, f_m , die dem Wert $x = (x_1, \dots, x_n)$ den Wert $y = (f_1(x), \dots, f_m(x))$ zuordnet. Die Jacobi-Matrix J_f von f ist die $m \times n$ -Matrix aller erster partieller Ableitungen, d. h.

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Aus Satz 5.5 können wir für Produkte und Quotienten zweier Zufallsvariablen direkt geschlossene Ausdrücke angeben.

Korollar 5.6 (Produkt- und Quotienttransformationen): Seien X_1 und X_2 stetige Zufallsvariablen mit den Dichten f_{X_1} resp. f_{X_2} und seien sie unabhängig. Dann sind die Zufallsvariablen $Y := X_1 \cdot X_2$ und $Z := X_1/X_2$ stetig mit den Dichten

$$f_Y(y) = \int \frac{1}{|t|} f_{X_1}(t) f_{X_2}(y/t) dt \quad \text{für alle } y \in \mathbb{R} \quad (5.4)$$

und

$$f_Z(z) = \int |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \quad \text{für alle } z \in \mathbb{R}. \quad (5.5)$$

Anmerkung: Das gedankliche Vorgehen ist hier genauso wie bei der Faltung. Es wird über alle Werte aufsummiert/integriert, die die eine der beteiligten Zufallsvariablen annehmen kann, kombiniert mit dem Wert, der dann der anderen Zufallsvariable zukommt.

Beispiel 5.7 (Produkttransformation): Seien X_1 und X_2 unabhängig und beide $\mathcal{U}(0,1)$ -verteilt, so haben wir die Dichten $f_{X_1}(x) = f_{X_2}(x) = \mathbb{1}_{(0,1)}(x)$. Die Dichte des Produktes $Y = X_1 \cdot X_2$ ist dann

$$\begin{aligned} f_Y(y) &= \int \frac{1}{|t|} \mathbb{1}_{(0,1)}(t) \mathbb{1}_{(0,1)}(y/t) dt \\ &= \int_0^1 \frac{1}{t} \mathbb{1}_{(0,1)}(y/t) dt \\ &= \int_y^1 \frac{1}{t} dt \quad \text{falls } 0 < y < 1, \\ &= \log t \Big|_y^1 = \log 1 - \log y = -\log y. \end{aligned}$$

Für den Quotienten wenden wir Satz 5.5 zwecks ausführlicher Darstellung an.

Beispiel 5.8: Seien X_1 und X_2 Zufallsvariablen für die Wartezeiten zweier Personen an unabhängigen Service-Schaltern, d. h. X_1 und X_2 sind unabhängig und exponentialverteilt mit den Parametern λ_1 resp. λ_2 .

Wir interessieren uns für den Quotienten $\frac{X_1}{X_2}$, wollen den aber mit dem verallgemeinerten Satz 5.5 angehen. Da wir zur Anwendung eine Funktion benötigen, die beim zwei-dimensionalen Zufallsvektor wieder einen zwei-dimensionalen Vektor liefert, betrachten wir die Funktion

$$u \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1(x_1, x_2) \\ y_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2/x_1 \end{pmatrix} \quad \text{mit} \quad u^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \cdot y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2/x_1 \cdot x_1 \end{pmatrix}.$$

Die entsprechende Jacobi-Matrix ist

$$J_{u^{-1}}(y_1, y_2) = \begin{pmatrix} \frac{\partial y_1}{\partial y_1} & \frac{\partial y_1}{\partial y_2} \\ \frac{\partial (y_2 \cdot y_1)}{\partial y_1} & \frac{\partial (y_2 \cdot y_1)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ y_2 & y_1 \end{pmatrix}$$

5 Nachtrag zu Zufallsvariablen und Verteilungen

mit der Determinante $\det(J_{u^{-1}}(y_1, y_2)) = 1 \cdot y_1 - 0 \cdot y_2 = y_1$. Damit ergibt sich für den Zufallsvektor $Y = (Y_1, Y_2) = u(X_1, X_2)$ die Dichte

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= f_X \begin{pmatrix} y_1 \\ y_2 \cdot y_1 \end{pmatrix} \cdot |y_1| = \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot |y_1| \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_1 y_2) \\ &= \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot y_1 \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_2). \end{aligned}$$

Da wir uns für die Marginaldichte interessieren, integrieren wir y_1 aus:

$$\begin{aligned} f_{Y_2}(y_2) &= \int \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot y_1 \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_2) \\ &= \lambda_1 \lambda_2 \mathbb{1}_{(0, \infty)}(y_2) \int_0^{\infty} y_1 e^{-(\lambda_1 + \lambda_2 y_2) y_1} dy_1, \end{aligned}$$

was wir auch durch die Anwendung von Satz 5.6, Gleichung (5.4) erhalten könnten. Das Integral lösen wir durch partielle Integration, wobei $\alpha := \lambda_1 + \lambda_2 y_2$ ist:

$$\begin{aligned} \int_0^{\infty} y_1 e^{-(\lambda_1 + \lambda_2 y_2) y_1} dy_1 &= \int_0^{\infty} \underbrace{y_1}_{:=v} \underbrace{e^{-\alpha y_1}}_{:=u'} dy_1 = y_1 \left(-\frac{1}{\alpha} \right) e^{-\alpha y_1} \Big|_0^{\infty} - \int_0^{\infty} \left(-\frac{1}{\alpha} \right) e^{-\alpha y_1} dy_1 \\ &= \frac{1}{\alpha} \int_0^{\infty} e^{-\alpha y_1} dy_1 = \frac{1}{\alpha^2} \int_0^{\infty} e^{-z} dz \quad (\text{Substitution } z := \alpha y_1) \\ &= \frac{1}{\alpha^2} = \frac{1}{(\lambda_1 + \lambda_2 y_2)^2}. \end{aligned}$$

Damit haben wir

$$f_{Y_2}(y_2) = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 y_2)^2} \mathbb{1}_{(0, \infty)}(y_2)$$

Als nächstes interessieren wir uns für den Erwartungswert. Wir substituieren $s := \lambda_1 + \lambda_2 y_2$, d. h. $y_2 = \frac{s - \lambda_1}{\lambda_2}$ und $ds = \lambda_2 dy_2$, und haben

$$\begin{aligned} \mathbb{E} \left(\frac{X_2}{X_1} \right) &= \mathbb{E}(Y_2) = \int_0^{\infty} y_2 \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 y_2)^2} dy_2 \\ &= \int_{\lambda_1}^{\infty} \frac{s - \lambda_1}{\lambda_2} \cdot \frac{\lambda_1 \lambda_2}{s^2} \cdot \frac{1}{s^2} ds = \frac{\lambda_1}{\lambda_2} \int_{\lambda_1}^{\infty} \frac{s - \lambda_1}{s^2} ds \\ &= \frac{\lambda_1}{\lambda_2} \left[\int_{\lambda_1}^{\infty} \frac{1}{s} ds - \int_{\lambda_1}^{\infty} \frac{\lambda_1}{s^2} ds \right] = \frac{\lambda_1}{\lambda_2} \log s \Big|_{\lambda_1}^{\infty} + \frac{\lambda_1^2}{\lambda_2} \frac{1}{s} \Big|_{\lambda_1}^{\infty} = \infty \end{aligned}$$

Für $\mathbb{E}(X_1/X_2)$ erhalten wir die exakt analoge Rechnung und damit dasselbe Ergebnis.

Fazit: Zwei Personen mit den exponentialverteilten Wartezeiten X_1 und X_2 haben beide das Gefühl, in der falschen Schlange zu stehen.

Abbildungsverzeichnis

1.1	Stochastiker in der Zeit. (Die Bilder entstammen Wikimedia-Commons und unterliegen der CC-Lizenz.)	4
4.1	Reiskorn Beispiel zur Poisson-Approximation (Das Bild entstammt Wikimedia-Commons und unterliegt der CC-Lizenz.)	48