

Technische Universität Ilmenau
Institut für Mathematik
Fachgebiet
Wahrscheinlichkeitsrechnung und mathematische Statistik

Silvia Vogel

Stochastik

Wintersemester 2009/2010

Literatur

- [1] F. Beichelt: Stochastik für Ingenieure. Teubner Stuttgart 1995, Taschenbuch 2003.
- [2] O. Beyer, H. Hackel, V. Pieper, J. Tiedge: Wahrscheinlichkeitsrechnung und mathematische Statistik. MINÖL Bd. 17, 8. Auflage, Teubner Stuttgart, Leipzig 1999.
- [3] H. Dehling, B. Haupt: Einführung in die Wahrscheinlichkeitstheorie und Statistik. 2. Auflage; Springer 2007.
- [4] L. Dümbgen: Stochastik für Informatiker. Springer 2003, Taschenbuch 2007.
- [5] M. Fisz: Wahrscheinlichkeitsrechnung und mathematische Statistik. 11. Auflage, Deutscher Verlag der Wissenschaften 1988.
- [6] M. Greiner, G. Tinhofer: Stochastik für Studienanfänger der Informatik. Carl Hanser Verlag 1996.
- [7] G. Hübner: Stochastik. 4. Auflage, Vieweg 2003.
- [8] J. Lehn, H. Wegmann: Einführung in die Statistik. 5. Auflage, Teubner Stuttgart 2006.
- [9] R. Mathar, D. Pfeifer: Stochastik für Informatiker. 2. Auflage, Teubner Stuttgart 2001.
- [10] R. Storm: Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle. 12. Auflage 2007.
- [11] H. Weber: Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Ingenieure. 3. Auflage, Teubner Stuttgart 1992.

Einführung

Die **Stochastik** umfasst die Wahrscheinlichkeitsrechnung und mathematische Statistik sowie die darauf aufbauenden Theorien. Sie beschäftigt sich mit der Untersuchung zufälliger Erscheinungen.

Das Rechnen mit „Unsicherheiten“ kann nicht umgangen werden, da es einerseits prinzipiell nicht möglich ist, alle Eingangsgrößen eines Modells ganz genau zu erfassen (Heisenbergsche Unschärferelation, Beschränktheit der menschlichen Möglichkeiten) und andererseits die Gewinnung und Bearbeitung eines hinreichend angepassten deterministischen Modells häufig einen unvermeidbar hohen Aufwand erfordern würde.

Die Stochastik stellt eine Vielzahl von mathematischen Modellen für Aufgabenstellungen mit „Unsicherheiten“ bereit. Eine Reihe von Anwendungen der Stochastik hat bereits eigenständige Bedeutung erlangt, z.B. Warteschlangentheorie, Zuverlässigkeitstheorie, statistische Qualitätskontrolle, Informationstheorie, Monte-Carlo-Methoden.

Beispiel: Warteschlangentheorie (Bedienungstheorie)

Gegeben ist Bedienungssystem (Rechnernetz), das in folgender Weise charakterisiert werden kann:

- Es gibt eine oder mehrere Bedieneinheiten (Rechner).
- Die Kunden (Jobs) kommen zu vorher nicht bekannten Zeitpunkten im Bedienungssystem an.
- Die Bedienungszeiten sind nicht deterministisch.

In der Regel soll das Bedienungssystem „optimal“ gestaltet werden (kostengünstig, Einhaltung von Leistungsgarantien, geringe Durchlaufzeit ...).

Dazu werden Aussagen über die Entwicklung der „Warteschlange“, die mittlere Verweilzeit der Kunden im System, die mittlere Auslastung der Bediengeräte usw. benötigt.

Die Warteschlangentheorie bietet wahrscheinlichkeitstheoretische Modelle zur Behandlung dieser Fragestellungen an. Oft liefert folgendes Modell eine gute Annäherung an die Wirklichkeit:

- Der Kundeneingangstrom ist ein sogenannter Poisson-Prozess.
- Die Bedienungszeiten werden als unabhängige, exponentialverteilte Zufallsgrößen aufgefasst.

Die Parameter der in den Modellen zugrunde gelegten Verteilungen müssen i. Allg. geschätzt werden (\rightarrow Statistik). Es ist möglich, gewisse Modellannahmen mit Hilfe statistischer Test zu überprüfen (\rightarrow Statistik).

Teil I: Wahrscheinlichkeitstheorie

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Zur historischen Entwicklung der Wahrscheinlichkeitsrechnung | 5 |
| 2 | Grundbegriffe | 6 |
| 2.1 | Zufallsexperiment (zufälliger Versuch) | 6 |
| 2.2 | Die relative Häufigkeit | 7 |
| 2.3 | Die Axiome der Wahrscheinlichkeitstheorie | 8 |
| 2.4 | Spezialfälle der allgemeinen Wahrscheinlichkeitsdefinition | 9 |
| 2.5 | Bedingte Wahrscheinlichkeiten und Unabhängigkeit | 11 |
| 2.6 | Die Formel der totalen Wahrscheinlichkeit und die Bayessche Formel | 14 |
| 3 | Zufallsvariablen und Verteilungen | 16 |
| 3.1 | Zufallsgrößen | 16 |
| 3.2 | Die Verteilungsfunktion | 18 |
| 3.3 | Diskrete und stetige Zufallsgrößen | 19 |
| 3.4 | Erwartungswert und Varianz einer Zufallsgröße | 20 |
| 3.5 | Spezielle Verteilungen | 22 |
| 3.5.1 | Binomialverteilung | 22 |
| 3.5.2 | Hypergeometrische Verteilung | 24 |
| 3.5.3 | Poisson-Verteilung | 25 |
| 3.5.4 | Weitere diskrete Verteilungen | 26 |
| 3.5.5 | Gleichmäßige stetige Verteilung | 26 |
| 3.5.6 | Exponentialverteilung | 26 |
| 3.5.7 | Normalverteilung | 28 |
| 3.5.8 | Cauchy-Verteilung | 30 |
| 3.5.9 | Weitere wichtige stetige Verteilungstypen | 30 |
| 3.6 | Die Verteilung einer Funktion einer Zufallsgröße | 30 |
| 3.7 | Weitere Kenngrößen von Verteilungen | 31 |
| 3.8 | Mehrdimensionale Zufallsvariablen | 32 |
| 3.8.1 | Diskrete Zufallsvektoren | 33 |
| 3.8.2 | Stetige Zufallsvektoren | 35 |
| 3.8.3 | Erwartungswertvektor und Kovarianzmatrix eines Zufallsvektors | 37 |
| 3.9 | Unabhängigkeit von Zufallsgrößen | 38 |
| 3.10 | Funktionen eines Zufallsvektors | 39 |
| 3.10.1 | Der Erwartungswert einer Funktion eines Zufallsvektors | 39 |
| 3.10.2 | Die Verteilung der Summe unabhängiger Zufallsgrößen | 40 |
| 3.10.3 | Die Verteilung weiterer Funktionen unabhängiger Zufallsgrößen | 42 |

| | | |
|----------|-------------------------------------|-----------|
| 4 | Grenzwertsätze | 43 |
| 4.1 | Gesetze der großen Zahlen | 43 |
| 4.2 | Zentrale Grenzwertsätze | 44 |

1 Zur historischen Entwicklung der Wahrscheinlichkeitsrechnung

Die Grundlagen der Wahrscheinlichkeitstheorie wurden im Zusammenhang mit der Betrachtung von Glücksspielen geschaffen.

Der Briefwechsel von **B. Pascal**, **P. de Fermat** (1654) über die 2. Aufgabe des Chevaliers de Méré gilt als der Beginn der Entwicklung der Wahrscheinlichkeitstheorie zur selbstständigen Wissenschaft.

2. Aufgabe von de Méré: Zwei Spieler vereinbaren, eine Reihe von Partien zu spielen. Gewinner soll derjenige sein, der als erster s Partien gewonnen hat. Das Spiel wird vorzeitig beim Stand von $a:b$ abgebrochen. Wie ist der Gewinn aufzuteilen?
(z.B. $s=3$, $a=2$, $b=1$)

Weitere wichtige Beiträge zur Entwicklung der Wahrscheinlichkeitstheorie leisteten die Wissenschaftler

- **Ch. Huygens** 1657: Über Berechnungen beim Würfelspiel,
- **J. Bernoulli** 1713: Ars conjectandi,
- **P. S. Laplace** 1812: Théorie analytique des probabilités (klassische Definition der Wahrscheinlichkeit),
- **S. D. Poisson** (1781 - 1840).

In der zweiten Hälfte des 19. Jahrhunderts kam es zur Stagnation der Entwicklung der Wahrscheinlichkeitstheorie in Westeuropa.

In Russland wurde die Wahrscheinlichkeitstheorie vor allem durch **P.L. Tschebyschev** (1821 - 1894) und seine Schüler (unter ihnen **A. A. Markov** (1856 - 1922)) weiterentwickelt.

Anwendungen der Wahrscheinlichkeitstheorie im 19. Jahrhundert gab es in der Bevölkerungsstatistik, der Fehlerrechnung, im Messwesen, im Versicherungswesen, in der Ballistik und der kinetischen Gastheorie.

Auf dem Mathematikerkongress in Paris im Jahre 1900 sagte D. Hilbert in seiner berühmten Rede:

„Durch die Untersuchungen über die Grundlagen der Geometrie wird uns die Aufgabe nahegelegt, nach diesem Vorbilde diejenigen physikalischen Disziplinen axiomatisch zu behandeln, in denen schon heute die Mathematik eine hervorragende Rolle spielt, dies sind in erster Linie die Wahrscheinlichkeitsrechnung und die Mechanik.“ (6. Hilbertsches Problem)

Die axiomatische Begründung der Wahrscheinlichkeitsrechnung wurde 1933 von **A. N. Kolmogorov** (1903 - 1987) in der Arbeit „Grundbegriffe der Wahrscheinlichkeitsrechnung“ geschaffen. Seither hat sich die Wahrscheinlichkeitsrechnung stürmisch entwickelt.

2 Grundbegriffe

2.1 Zufallsexperiment (zufälliger Versuch)

Definition:

Ein **Zufallsexperiment** (zufälliger Versuch) ist ein Experiment, das unter gleichen äußeren Bedingungen (zumindest gedanklich) beliebig oft wiederholbar ist, und dessen Ergebnis im Rahmen verschiedener Möglichkeiten ungewiss ist.

Beispiele:

- B1: Werfen eines Würfels,
- B2: Messen der Zeit für die Abarbeitung eines Jobs,
- B3: Feststellen der Anzahl der Kunden in einem Bedienungssystem zu einem festen Zeitpunkt.

Die Menge aller möglichen, sich gegenseitig ausschließenden **Ergebnisse** eines Zufallsexperimentes wird mit Ω bezeichnet.

Beispiele:

- B1: $\Omega = \{1,2,3,4,5,6\}$,
- B2: $\Omega = (0;\infty)$ (Maßeinheit angeben),
- B3: $\Omega = \{0,1,2,\dots\}$.

Teilmengen von Ω heißen (**zufällige**) **Ereignisse** (i. Allg. nicht alle). Ereignisse werden in der Regel mit den Buchstaben $A, B, C\dots$ bezeichnet.

Die einelementigen Teilmengen $\{\omega\}$ heißen auch **Elementarereignisse**. Man sagt „Das Ereignis A tritt ein“, wenn ein $\omega \in A$ beobachtet wird.

Beispiele:

- B1: $A\dots$ Es wird eine gerade Zahl gewürfelt.
 $A = \{2,4,6\}$.
- B2: $B\dots$ Die Bearbeitung eines Jobs dauert mindestens 1 Minute.
 $B = [60; \infty)$ (in Sekunden).
- B3: $C\dots$ Es befinden sich höchstens 3 Kunden im Bedienungssystem.
 $C = \{0,1,2,3\}$.

Beziehungen und Operationen zwischen Ereignissen

Es seien $A, B, A_i \subset \Omega$ $i = 1, 2\dots$. Die folgenden Verknüpfungen von Ereignissen werden häufig betrachtet:

| <u>Schreibweise</u> | <u>Sprechweise</u> |
|------------------------------|-----------------------------------|
| $A \cup B$ | A oder B |
| $\bigcup_{i=1}^{\infty} A_i$ | mindestens ein A_i |
| $A \cap B$ | A und B |
| $\bigcap_{i=1}^{\infty} A_i$ | alle A_i |
| \bar{A} (auch A^c) | das zu A komplementäre Ereignis |

\emptyset wird **unmögliches Ereignis** und Ω **sicheres Ereignis** genannt. A und B heißen „**unvereinbar**“, wenn $A \cap B = \emptyset$ gilt.

Falls Ω abzählbar ist, kann man alle Teilmengen als Ereignisse zulassen, d.h., Ereignisse sind i. Allg. die Elemente der Potenzmenge $\mathcal{P}(\Omega)$.

Falls Ω überabzählbar ist, führt die Verwendung der Potenzmenge zu Schwierigkeiten schon bei der Definition einfacher Verteilungen. Man beschränkt sich daher auf ein „kleineres“ Mengensystem, das eine σ -Algebra bildet.

Definition:

Es sei Ω eine nichtleere Menge. Ein System \mathcal{A} von Teilmengen von Ω heißt **σ -Algebra** über Ω , wenn

- (1) $\Omega \in \mathcal{A}$,
- (2) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$,
- (3) $A_i \in \mathcal{A}, i = 1, 2, \dots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Ist neben der Menge der Ergebnisse Ω eine σ -Algebra über Ω gegeben, so heißen alle $A \in \mathcal{A}$ **Ereignisse**. \mathcal{A} wird auch **Ereignisfeld** genannt.

2.2 Die relative Häufigkeit

Die Eigenschaften eines Wahrscheinlichkeitsmaßes werden den Eigenschaften der relativen Häufigkeit nachgebildet. Aus diesem Grund wird an dieser Stelle kurz auf die relative Häufigkeit eingegangen.

Ein Zufallsexperiment werde n -mal unabhängig voneinander wiederholt. Wird dabei m -mal das Ereignis A beobachtet, so heißt

$$h_n(A) := \frac{m}{n}$$

relative Häufigkeit von A in n Versuchen.

Die relative Häufigkeit hat die folgenden Eigenschaften:

- (1) $0 \leq h_n(A) \leq 1$,

$$(2) h_n(\Omega) = 1,$$

$$(3) A \cap B = \emptyset \Rightarrow h_n(A \cup B) = h_n(A) + h_n(B).$$

Konkrete Beispiele zeigen, dass die relative Häufigkeit eine gewisse Stabilität zeigt, wenn n genügend groß ist. Es sei hierbei auf die Münzwürfe verwiesen, die G.L.L. Buffon und K. Pearson durchführten. Sie ermittelten folgende relative Häufigkeiten für das Ereignis

A ...Wappen liegt oben:

| | Anzahl der Würfe | relative Häufigkeit |
|---------|------------------|---------------------|
| Buffon | 4040 | 0,5069 |
| Pearson | 12000 | 0,5016 |
| Pearson | 24000 | 0,5005 |

2.3 Die Axiome der Wahrscheinlichkeitstheorie

Nur den Elementen der σ -Algebra \mathcal{A} werden Wahrscheinlichkeiten zugeordnet.

Definition:

Eine auf \mathcal{A} definierte reellwertige Funktion P heißt **Wahrscheinlichkeitsmaß**, wenn

$$(A1) \quad P(A) \geq 0 \quad \forall A \in \mathcal{A},$$

$$(A2) \quad P(\Omega) = 1,$$

$$(A3) \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \text{ falls } A_i \cap A_j = \emptyset \text{ für } i \neq j$$

gilt.

Das Tripel $[\Omega, \mathcal{A}, P]$ heißt **Wahrscheinlichkeitsraum**.

$P(A)$ heißt **Wahrscheinlichkeit** des Ereignisses A .

Gilt $P(A) = 1$, wird A **fast sicheres Ereignis** genannt. Falls $P(A) = 0$ erfüllt ist, nennt man A **fast unmögliches Ereignis**.

Für Wahrscheinlichkeiten gelten die folgenden **Rechenregeln**:

$$(1) P(\emptyset) = 0,$$

$$(2) A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B),$$

$$(3) P(\bar{A}) = 1 - P(A),$$

$$(4) A \subset B \Rightarrow P(A) \leq P(B),$$

$$(5) 0 \leq P(A) \leq 1,$$

$$(6) P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$(7) P(A \cup B) \leq P(A) + P(B),$$

$$(8) P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < l \leq n} P(A_i \cap A_l) + \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right) \\ = \sum_{k=1}^n (-1)^{k+1} \sum_{(i_1, \dots, i_k) \in C_n^k} P\left(\bigcap_{j=1}^k A_{i_j}\right),$$

C_n^k bezeichnet dabei die Menge der Kombinationen ohne Wiederholung von n Elementen zur k -ten Klasse.

Die Rechenregel (8) wird auch Siebformel von Poincaré-Sylvester genannt.

Beweis der Rechenregel (1)

In (A3) werden $A_1 = \Omega$ und $A_i = \emptyset$ für $i = 2, 3, \dots$ gesetzt. Dann folgt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(\Omega) \stackrel{(A2)}{=} 1 = \sum_{i=1}^{\infty} P(A_i) = 1 + \sum_{i=2}^{\infty} P(\emptyset). \text{ Aus der Beziehung } \sum_{i=2}^{\infty} P(\emptyset) = 0 \text{ ergibt sich } P(\emptyset) = 0. \quad \square$$

2.4 Spezialfälle der allgemeinen Wahrscheinlichkeitsdefinition

a) Klassische Wahrscheinlichkeitsdefinition

Der Wahrscheinlichkeitsraum wird in der folgenden Weise spezifiziert:

- Ω bestehe aus n Elementen: $\Omega = \{\omega_1 \dots \omega_n\}$,
- $\mathcal{A} = \mathcal{P}(\Omega)$,
- Alle Elementarereignisse sind gleichwahrscheinlich, d.h., $P(\{\omega_i\}) = \frac{1}{n}$, $i = 1, \dots, n$.

Ein beliebiges Ereignis lässt sich nun in der Form $A = \bigcup_{i \in I} \{\omega_i\}$ mit einer geeigneten Indexmenge I darstellen. Daraus folgt

$$P(A) = P\left(\bigcup_{i \in I} \{\omega_i\}\right) = \sum_{i \in I} P(\{\omega_i\}) = \sum_{i \in I} \frac{1}{n} = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega} \\ = \frac{\text{Anzahl der „günstigen Fälle“}}{\text{Anzahl der „möglichen Fälle“}}.$$

Ein wichtiges Hilfsmittel bei der Berechnung „klassischer Wahrscheinlichkeiten“ ist die Kombinatorik.

Beispiele:

1. Urnenmodell mit Zurücklegen

In einer Urne befinden sich N Kugeln, M weiße und $N - M$ schwarze. Aus der Urne wird zufällig eine Kugel entnommen und ihre Farbe notiert. Anschließend wird die Kugel wieder zurückgelegt. Dieser Vorgang wird n -mal unabhängig voneinander wiederholt.

Gesucht ist die Wahrscheinlichkeit des Ereignisses

A_{nk} ... Unter den n gezogenen Kugeln befinden sich genau k weiße.

Lösung: Die N Kugeln seien von 1 bis N durchnummeriert. Die Menge der möglichen, gleichwahrscheinlichen Ziehungsergebnisse (Elementarereignisse) kann dann folgendermaßen beschrieben werden:

$$\Omega = \{(1, \dots, 1, 1), (1, \dots, 1, 2), \dots, (N, \dots, N)\}.$$

Ω besitzt N^n Elemente (= Anzahl der „möglichen Fälle“).

Die Anzahl der „günstigen Fälle“ lässt sich auf folgende Weise ermitteln:

Es gibt $\binom{n}{k}$ verschiedene Möglichkeiten, k weiße Kugeln in einer Serie von n Kugeln einzuordnen. Weiterhin gibt es M^k Möglichkeiten, weiße Kugeln auf die k gewählten Plätze für weiße Kugeln zu verteilen und entsprechend

$(N - M)^{n-k}$ Möglichkeiten für die schwarzen Kugeln. Insgesamt ergeben sich damit $\binom{n}{k} M^k (N - M)^{n-k}$ „günstige Fälle“. Somit folgt:

$$P(A_{nk}) = \frac{\binom{n}{k} M^k (N - M)^{n-k}}{N^n} = \binom{n}{k} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}.$$

2. Urnenmodell ohne Zurücklegen

In einer Urne befinden sich N Kugeln, M weiße, und $N - M$ schwarze. Aus der Urne werden n Kugeln zufällig entnommen und nicht zurückgelegt.

Gesucht ist die Wahrscheinlichkeit des Ereignisses

B_{nk} ... Unter den n gezogenen Kugeln befinden sich genau k weiße.

Lösung: Ω bestehe aus allen n -elementigen Teilmengen von $\{1, 2, \dots, N\}$;

Ω besitzt somit $\binom{N}{n}$ Elemente (= Anzahl der „möglichen Fälle“).

Weiterhin gibt es $\binom{M}{k}$ verschiedene Möglichkeiten, k Kugeln aus M weißen Kugeln herauszugreifen, und es gibt $\binom{N-M}{n-k}$ verschiedene Möglichkeiten, aus $N - M$ schwarzen Kugeln $n - k$ Kugeln auszuwählen. Die Anzahl der „günstigen Fälle“ ist daher $\binom{M}{k} \binom{N-M}{n-k}$,

und es gilt $P(B_{nk}) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$.

Man beachte, dass eine Aufgabenstellung nicht immer eindeutig festlegt, was unter zufälliger Auswahl zu verstehen ist. Das folgende Beispiel soll das verdeutlichen. Entsprechende Beispiele findet man auch bei anderen Spezifikationen des Wahrscheinlichkeitsraumes.

Beispiel (aus: J. Lehn, H. Wegmann: Einführung in die Statistik)

In einem Speisewagen gibt es 5 Tische mit je 4 Plätzen. Bevor der Speisewagen geöffnet wird, geht der Kellner durch den Zug und nimmt die Platzreservierungswünsche der Fahrgäste entgegen. Gleich die ersten beiden Fahrgäste, die er unabhängig voneinander anspricht, lassen sich einen Platz reservieren. Wie groß

ist (unter geeigneter Annahme bez. der Gleichberechtigung der Auswahlmöglichkeiten) die Wahrscheinlichkeit, dass die beiden am gleichen Tisch sitzen werden, wenn der Kellner die Reservierung zufällig vornimmt?

1. Lösung: Es wird die Situation nach der ersten Reservierung betrachtet. Der Kellner wählt für den zweiten Fahrgast mit gleicher Wahrscheinlichkeit einen der fünf Tische aus. Mit Wahrscheinlichkeit $\frac{1}{5}$ wird dies der Tisch sein, an dem auch der erste Fahrgast sitzen wird.

2. Lösung: Auch hier wird die Situation nach der ersten Reservierung betrachtet. Der Kellner wählt für den zweiten Fahrgast mit gleicher Wahrscheinlichkeit einen der noch freien 19 Plätze aus. Mit Wahrscheinlichkeit $\frac{3}{19}$ wird dies einer der drei noch freien Plätze am Tisch des ersten Fahrgastes sein.

b) Geometrische Wahrscheinlichkeit

Ω sei ein ebenes Flächenstück (Teilmenge des \mathbb{R}^2) mit endlichem Flächeninhalt. Für jedes Ereignis $A \subset \Omega$ wird

$$P(A) := \frac{\text{Flächeninhalt von } A}{\text{Flächeninhalt von } \Omega}$$

gesetzt. Damit ist P ein Wahrscheinlichkeitsmaß.

Analog kann man im \mathbb{R}^1 und allgemein im \mathbb{R}^n vorgehen.

Anwendung der geometrischen Wahrscheinlichkeit

Das Modell der geometrischen Wahrscheinlichkeit kann angewendet werden, wenn ein Punkt Q in Ω „zufällig“ ausgewählt werden soll und die folgenden Annahmen gelten:

- Genau die Punkte von Ω können ausgewählt werden.
- Die Wahrscheinlichkeit dafür, dass Q zu einer vorgegebenen Teilmenge $A \subset \Omega$ gehört, hängt lediglich vom Flächeninhalt von A und nicht von der Lage oder Gestalt von A ab.

2.5 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Beispiel:

Anna möchte ihren Freund Bert besuchen, hat aber vergessen, in welchem der sechs Eingänge 1, 2, ..., 6 eines Wohnblocks Bert wohnt. Für Anna sind alle Eingänge gleichwahrscheinlich.

Betrachtet wird das Ereignis

E ... Bert wohnt im Eingang 1.

- a) Man berechne die Wahrscheinlichkeit des Ereignisses E .

- b) Man berechne die Wahrscheinlichkeit des Ereignisses E , wenn Anna von einem Passanten erfahren hat, dass Bert nicht im Eingang 6 wohnt.

Lösung:

- a) Die Menge der möglichen, gleichberechtigten Ergebnisse des Zufallsexperiments „Auswahl eines Einganges“ ist $\Omega := \{1, 2, \dots, 6\}$.
Damit erhält man
 $E = \{1\}$, $P(E) = \frac{1}{6}$.
- b) Da der Eingang 6 ausgeschlossen wurde, ist die Menge der möglichen, gleichberechtigten Ergebnisse des Zufallsexperimentes „Auswahl eines Einganges“ nun nur noch $\tilde{\Omega} := \{1, 2, \dots, 5\}$, also ergibt sich $\tilde{P}(E) = \frac{1}{5}$.

Man kann bei b) auch von $\Omega = \{1, 2, \dots, 6\}$ ausgehen, muss dann aber beachten, dass den Elementarereignissen $\{i\}$, $i = 1, \dots, 6$, jetzt andere Wahrscheinlichkeiten, nämlich

$$P_6(\{i\}) = \begin{cases} \frac{1}{5} & \text{für } i = 1, \dots, 5, \\ 0 & \text{für } i = 6, \end{cases}$$

zugeordnet sind. Diese Wahrscheinlichkeiten nennt man bedingte Wahrscheinlichkeiten unter der Bedingung, dass das Ereignis F ... Bert wohnt nicht im Eingang 6 eingetreten ist (d.h., nur noch die Ergebnisse $i \in F$ zugelassen sind).

Um die zugrunde liegende Bedingung deutlich zu machen, verwendet man die Schreibweise $P(E|F)$ anstelle von $P_6(E)$.

Offenbar kann man $P(E|F)$ im Beispiel wie folgt berechnen:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(\{1\})}{P(\{1, \dots, 5\})} = \frac{1}{5}.$$

Definition:

Es seien $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum, $A \in \mathcal{A}$, $B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B .

Die bedingte Wahrscheinlichkeit hat die folgenden **Eigenschaften**:

- (1) $P(\cdot|B)$ ist ein Wahrscheinlichkeitsmaß auf \mathcal{A} (d.h., die Axiome (A1)–(A3) und damit die aus den Axiomen abgeleiteten Rechenregeln gelten auch für bedingte Wahrscheinlichkeiten).
- (2) $A \cap B = \emptyset \Rightarrow P(A|B) = 0$.

$$(3) B \subset A \Rightarrow P(A|B) = 1.$$

Das folgende Beispiel zeigt noch einmal die beiden Möglichkeiten beim Rechnen mit bedingten Wahrscheinlichkeiten auf. Beim Lösungsvorschlag 1 wird der Wahrscheinlichkeitsraum der Bedingung angepasst. Beim Lösungsvorschlag 2 wird mit einem Wahrscheinlichkeitsraum gearbeitet, in dem alle vorkommenden Ereignisse beschrieben werden können. Dieser Weg ist vor allem dann vorteilhaft, wenn komplexere Aufgaben zu bearbeiten sind.

Beispiel:

Aus einem Skatenspiel werden 2 Karten gezogen. Wie groß ist die Wahrscheinlichkeit dafür, dass die zweite Karte ein Ass ist, wenn zuvor ein Ass gezogen wurde?

Die folgenden Ereignisse spielen in der Aufgabenstellung eine Rolle:

B ... Die erste Karte ist ein Ass.

A ... Die zweite Karte ist ein Ass.

1. Lösung: Die nach dem Ziehen der ersten Karte verbliebenen Karten werden durchnummeriert, wobei die Asse die Nummern 1,2,3 erhalten. Man verwendet die klassische Wahrscheinlichkeitsdefinition mit $\Omega = \{1, \dots, 31\}$.

Für das Ereignis A gibt es dann 31 mögliche und 3 günstige Fälle, somit $P(A|B) = \frac{3}{31}$.

2. Lösung: Die Karten seien durchnummeriert, wobei die Asse die Zahlen 1,2,3,4 erhalten. Die folgenden gleichwahrscheinlichen Versuchsausgänge bei 2 Ziehungen unter Beachtung der Reihenfolge sind möglich:

$$\Omega = \{(1, 2), \dots, (1, 32), (2, 1), \dots, (32, 31)\},$$

d.h., es gibt $32 \cdot 31$ mögliche Fälle. In diesem Wahrscheinlichkeitsraum lassen sich nun die Ereignisse $A \cap B$ sowie B und ihre Wahrscheinlichkeiten wie folgt beschreiben:

$$\begin{aligned} A \cap B &= \{(1, 2) \dots (4, 3)\} && \Rightarrow \text{card}(A \cap B) = 4 \cdot 3 \\ &&& \Rightarrow P(A \cap B) = \frac{4 \cdot 3}{32 \cdot 31}, \\ B &= \{(1, 2) \dots (1, 32), (2, 1) \dots (4, 32)\} && \Rightarrow \text{card}(B) = 4 \cdot 31 \\ &&& \Rightarrow P(B) = \frac{4 \cdot 31}{32 \cdot 31}. \end{aligned}$$

Daraus ergibt sich nach $P(A|B) = \frac{P(A \cap B)}{P(B)}$ die Wahrscheinlichkeit $P(A|B) = \frac{3}{31}$.

Es gibt Aufgabenstellungen, in denen die Information, dass ein Ereignis B eingetreten ist, nichts an der Wahrscheinlichkeit, mit der ein anderes Ereignis A eintritt, ändert. Zum Beispiel hat beim zweimaligen Würfeln mit einem Würfel das Ergebnis des ersten Wurfs keinen Einfluss auf die Wahrscheinlichkeit, beim 2. Wurf eine „6“ zu würfeln.

Das gibt Anlass zu der folgenden Definition:

Definition:

Es seien $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{A}, B \in \mathcal{A}$.
Die Ereignisse A und B heißen (voneinander) **unabhängig**, wenn
 $P(A \cap B) = P(A) \cdot P(B)$
gilt.

Man kann zeigen, dass die folgenden Aussagen richtig sind:

- (1) Es sei $P(B) > 0$. Dann sind A, B genau dann unabhängig, wenn $P(A|B) = P(A)$ gilt.
- (2) A, B unabhängig $\Leftrightarrow \bar{A}$ und B unabhängig.

Bei mehr als zwei Ereignissen muss man zwei Unabhängigkeitsbegriffe unterscheiden:

Definition:

- a) Die Ereignisse A_1, \dots, A_n heißen **paarweise unabhängig**, wenn
 $P(A_i \cap A_k) = P(A_i) \cdot P(A_k) \quad \forall i \neq k$ gilt.
- b) Die Ereignisse A_1, \dots, A_n heißen (in ihrer Gesamtheit oder stochastisch) **unabhängig**, wenn
 $P(A_{j_1} \cap \dots \cap A_{j_k}) = P(A_{j_1}) \cdot \dots \cdot P(A_{j_k})$
für jedes k -Tupel (j_1, \dots, j_k) mit $1 \leq j_1 < \dots < j_k \leq n; k = 2, 3, \dots, n$, gilt.

Bemerkung:

Aus der paarweisen Unabhängigkeit folgt i. Allg. nicht die Unabhängigkeit.
Darüber hinaus kann man aus $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$
nicht auf $P(A \cap B) = P(A) \cdot P(B)$ schließen, auch nicht im Fall
 $P(C) > 0$.

2.6 Die Formel der totalen Wahrscheinlichkeit und die Bayesche Formel

Gegeben seien ein Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$ sowie Ereignisse $A \in \mathcal{A};$
 $B_1, \dots, B_n, \dots \in \mathcal{A}$ mit $\Omega = \bigcup_{i=1}^{\infty} B_i$ und $P(B_i) > 0 \forall i, B_i \cap B_j = \emptyset$ für $i \neq j$.

Bekannt sind die Wahrscheinlichkeiten $P(A|B_i), P(B_i) \forall i$.

Gesucht werden die Wahrscheinlichkeiten $P(A)$ und/oder $P(B_i|A), i \in \{0, 1, 2, \dots\}$.

Es gilt:

Formel der totalen Wahrscheinlichkeit: $P(A) = \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i),$

Bayessche Formel: $P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}$

Beweis der Formel der totalen Wahrscheinlichkeit:

$$P(A) = P(A \cap \Omega) = P(A \cap (\bigcup_{i=1}^{\infty} B_i)) = P(\bigcup_{i=1}^{\infty} (A \cap B_i)) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i) \quad \square.$$

Beweis der Bayesschen Formel:

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}. \quad \square$$

Beispiel:

3 Maschinen stellen gleiche Teile her. Ihre Anteile an der Gesamtproduktion und die Ausschussquoten sind in der folgenden Tabelle zusammengestellt:

| | Maschine 1 | Maschine 2 | Maschine 3 |
|--------------------------------|------------|------------|------------|
| Anteil an der Gesamtproduktion | 20 % | 50 % | 30 % |
| Ausschussquote | 1 % | 4 % | 5 % |

- a) Gesucht ist die Wahrscheinlichkeit dafür, dass ein im Lager zufällig ausgewähltes Teil Ausschuss ist.

Zur mathematischen Beschreibung der Aufgabenstellung werden die Ereignisse

A ... Ein zufällig ausgewähltes Teil ist Ausschuss

B_i ... Ein zufällig ausgewähltes Teil stammt von Maschine i eingeführt. Die gegebenen Anteilswerte und Ausschussquoten liefern dann die Wahrscheinlichkeiten

$$P(B_1) = 0,2; \quad P(B_2) = 0,5; \quad P(B_3) = 0,3;$$

$$P(A|B_1) = 0,01; \quad P(A|B_2) = 0,04; \quad P(A|B_3) = 0,05.$$

Gesucht ist die Wahrscheinlichkeit $P(A)$. Sie ergibt sich zu

$$P(A) = \sum_{i=1}^3 P(A|B_i) \cdot P(B_i) = 0,2 \cdot 0,01 + 0,5 \cdot 0,04 + 0,3 \cdot 0,05 = 0,037.$$

- b) Ein im Lager zufällig ausgewähltes Teil ist Ausschuss. Wie groß ist die Wahrscheinlichkeit dafür, dass es von Maschine 2 stammt?

Gesucht ist die bedingte Wahrscheinlichkeit $P(B_2|A)$. Nach der Bayesschen Formel gilt $P(B_2|A) = \frac{P(A|B_2) \cdot P(B_2)}{P(A)} = \frac{0,5 \cdot 0,04}{0,037} \approx 0,54$.

3 Zufallsvariablen und Verteilungen

3.1 Zufallsgrößen

Beispiel 3.1:

Es seien 3 gleichartige Bauelemente BT_1, BT_2, BT_3 gegeben, die unabhängig voneinander ausfallen können. Die Wahrscheinlichkeit, mit der ein Bauteil im Zeitintervall $[0, T]$ ausfällt, betrage 0,1. Gesucht ist die Wahrscheinlichkeit der Ereignisse

$A_k \dots$ Im Zeitraum $[0, T]$ fallen genau k Bauteile aus ($k = 0, 1, 2, 3$).

Als **Zufallsexperiment** kann man die Betrachtung des Ausfallverhaltens der 3 Bauteile auffassen. Dieses Zufallsexperiment hat die

möglichen Ergebnisse

$$\begin{aligned} \omega_1 &= (b_1, b_2, b_3), & \omega_5 &= (\bar{b}_1, b_2, b_3), \\ \omega_2 &= (b_1, b_2, \bar{b}_3), & \omega_6 &= (\bar{b}_1, b_2, \bar{b}_3), \\ \omega_3 &= (b_1, \bar{b}_2, b_3), & \omega_7 &= (\bar{b}_1, \bar{b}_2, b_3), \\ \omega_4 &= (b_1, \bar{b}_2, \bar{b}_3), & \omega_8 &= (\bar{b}_1, \bar{b}_2, \bar{b}_3), \end{aligned}$$

wobei z.B. die Abkürzung (b_1, b_2, \bar{b}_3) das Ergebnis

„Die Bauteile BT_1 und BT_2 fallen nicht aus; Bauteil BT_3 fällt aus.“

beschreibt.

Ein geeigneter **Wahrscheinlichkeitsraum** kann wie folgt eingeführt werden:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_8\}, \quad \mathcal{A} = \mathcal{P}(\Omega).$$

Zur Bestimmung des Wahrscheinlichkeitsmaßes werden die Ereignisse

$B_i \dots$ Das Bauteil BT_i fällt nicht aus

definiert. Da die Bauteile unabhängig voneinander ausfallen, muss dann zum Beispiel

$$P(\{\omega_2\}) = P(B_1 \cap B_2 \cap \bar{B}_3) = P(B_1) \cdot P(B_2) \cdot P(\bar{B}_3) = 0,9 \cdot 0,9 \cdot 0,1 = 0,081$$

gelten. Auf diese Weise ergibt sich die folgende Tabelle:

| Ergebnis | ω_1 | ω_2 | ω_3 | ω_4 | ω_5 | ω_6 | ω_7 | ω_8 |
|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|
| $P(\{\omega_i\})$ | 0,729 | 0,081 | 0,081 | 0,009 | 0,081 | 0,009 | 0,009 | 0,001 |

Die Wahrscheinlichkeiten $P(A_k)$, $k = 0, 1, 2, 3$, können mit Hilfe dieser Tabelle berechnet werden, indem die Ereignisse A_k als Vereinigungen geeigneter $\{\omega_i\}$ dargestellt werden, z.B.

$$\begin{aligned} P(A_2) &= P(\{\omega_4, \omega_6, \omega_7\}) = P(\{\omega_4\}) + P(\{\omega_6\}) + P(\{\omega_7\}) \\ &= 3 \cdot 0,009 = 0,027. \end{aligned}$$

Im Folgenden soll noch ein weiterer Weg zur Bestimmung der $P(A_k)$ aufgezeigt werden. Dazu wird eine Funktion $X|\Omega \rightarrow \{0, 1, 2, 3\} \subset \mathbb{R}^1$ eingeführt, die jedem Element aus Ω die Anzahl der ausfallenden Bauteile zuordnet.

Mit Hilfe dieser Funktion können die betrachteten Ereignisse und ihre Wahrscheinlichkeiten folgendermaßen beschrieben werden:

$$A_k = \{\omega : X(\omega) = k\} =: (X = k) \text{ und } P(A_k) = P(\{\omega : X(\omega) = k\}) =: P(X = k).$$

Da die Wahrscheinlichkeiten $P(\{\omega_i\})$ bekannt sind, ergibt sich

$$\begin{aligned} P(X = 0) &= P(\{\omega_1\}) = 0,729, \\ P(X = 1) &= P(\{\omega_2\} \cup \{\omega_3\} \cup \{\omega_5\}) = P(\{\omega_2\}) + P(\{\omega_3\}) + P(\{\omega_5\}) \\ &= 3 \cdot 0,081 = 0,243, \\ P(X = 2) &= P(\{\omega_4\} \cup \{\omega_6\} \cup \{\omega_7\}) = P(\{\omega_4\}) + P(\{\omega_6\}) + P(\{\omega_7\}) \\ &= 3 \cdot 0,009 = 0,027, \\ P(X = 3) &= P(\{\omega_8\}) = 0,001. \end{aligned}$$

Ausgehend von diesen Wahrscheinlichkeiten kann ein Wahrscheinlichkeitsmaß P_X auf der Menge $\Omega_X := \{0, 1, 2, 3\}$ der für X möglichen Werte, versehen mit der Potenzmenge als σ -Algebra, definiert werden:

$$P_X(\{k\}) := P(X = k) = P(\{\omega : X(\omega) = k\}).$$

Die betrachtete Funktion X ist eine sogenannte Zufallsgröße, und die Werte

$P_X(\{k\}), k = 0, 1, 2, 3$, beschreiben ihre „Wahrscheinlichkeitsverteilung“.

Für die Berechnung von $P(A_k)$ und der Wahrscheinlichkeiten von Verknüpfungen der Ereignisse A_k (z.B. „Mindestens 2 Bauteile fallen aus“) ist die Kenntnis der Wahrscheinlichkeiten $P(X = k) = P_X(\{k\}), k = 0, 1, 2, 3$, ausreichend.

Bemerkungen:

- 1.) Man kann X auch als Abbildung in die Menge \mathbb{R}^1 , versehen mit der σ -Algebra \mathcal{B}^1 , die von allen Intervallen und Halbachsen erzeugt wird, auffassen. Die betrachtete σ -Algebra \mathcal{B}^1 heißt σ -Algebra der **Borel-Mengen**. Ausgehend von P_X kann dann ein Wahrscheinlichkeitsmaß \tilde{P}_X auf \mathcal{B}^1 durch $\tilde{P}_X(A) = P_X(A \cap \{0, 1, 2, 3\}), A \in \mathcal{B}^1$, definiert werden.

In analoger Weise kann man vorgehen, wenn Abbildungen in eine andere Teilmenge der reellen Zahlen betrachtet werden. Im Interesse einer einheitlichen Beschreibungsweise erklärt man

Zufallsgrößen daher in der Regel als Abbildungen in \mathbb{R}^1 . Um ausgehend von P eine Wahrscheinlichkeitsverteilung P_X definieren zu können, müssen die Urbilder der Elemente von \mathcal{B}^1 zu \mathcal{A} gehören. Diese Eigenschaft nennt man **($\mathcal{A}, \mathcal{B}^1$)-Messbarkeit**.

- 2.) Manchmal ist es nötig, Abbildungen X in allgemeinere Räume als $[\mathbb{R}^1, \mathcal{B}^1]$ zu betrachten. Man spricht dann allgemein von Zufallsvariablen.
- 3.) Es wäre möglich, im betrachteten Beispiel gleich mit dem Wahrscheinlichkeitsraum $[\Omega_X, \mathcal{P}(\Omega_X), P_X]$ zu arbeiten. Müssen aber weitere Ereignisse (die sich z.B. auf die Nummer des ausgefallenen Bauteils beziehen können) in die Berechnungen einbezogen werden, wäre es erforderlich, einen neuen Wahrscheinlichkeitsraum einzuführen. Deshalb ist es günstig, einen hinreichend allgemeinen Wahrscheinlichkeitsraum „im Hintergrund“ zu haben und darauf aufbauend, Abbildungen in geeignete Räume (z.B. $\mathbb{R}^1, \mathbb{R}^n \dots$) zu betrachten.

Definition:

Eine Funktion $X|\Omega \rightarrow \mathbb{R}^1$, die die Eigenschaft besitzt, dass die Urbilder von Borel-Mengen zu \mathcal{A} gehören, heißt **Zufallsgröße**.

Definition:

Durch die Vorschrift $P_X(B) := P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\})$, $B \in \mathcal{B}^1$, wird auf \mathcal{B}^1 ein Wahrscheinlichkeitsmaß P_X erzeugt (induziert).

P_X heißt die **Wahrscheinlichkeitsverteilung** (oder kurz Verteilung) der Zufallsgröße X .

3.2 Die Verteilungsfunktion

Definition:

Die Funktion $F_X|\mathbb{R}^1 \rightarrow [0, 1]$, die durch $F_X(x) := P(X \leq x)$ erklärt ist, heißt **Verteilungsfunktion** der Zufallsgröße X .

Satz: Die Verteilung P_X ist durch F_X eindeutig bestimmt.

Fortsetzung des Beispiels 3.1:

Die Zufallsgröße X bezeichne wieder die Anzahl der ausfallenden Bauteile. Die Wahrscheinlichkeiten

$P(X = 3) = 0,001$, $P(X = 2) = 0,027$, $P(X = 1) = 0,243$, $P(X = 0) = 0,729$

sind gegeben. Es soll die Verteilungsfunktion von X bestimmt werden. Man erhält

$$\begin{aligned}
x < 0 : F_X(x) &= P(X \leq x) = P\{\omega : X(\omega) \leq x\} = P(\emptyset) = 0, \\
0 \leq x < 1 : F_X(x) &= P(X \leq x) = P(X = 0) = 0,729, \\
1 \leq x < 2 : F_X(x) &= P(X = 0) + P(X = 1) = 0,729 + 0,243 = 0,972, \\
2 \leq x < 3 : F_X(x) &= 0,729 + 0,243 + 0,27 = 0,999, \\
x \geq 3 : F_X(x) &= 1.
\end{aligned}$$

Eine Verteilungsfunktion hat folgende **Eigenschaften**:

- 1.) $\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$
- 2.) $F_X(x)$ ist monoton wachsend, d.h. $F_X(x_1) \leq F_X(x_2)$ für $x_1 < x_2.$
- 3.) $F_X(x)$ ist an jeder Stelle x_0 (mindestens) rechtsseitig stetig, d.h. $\lim_{x \rightarrow x_0+0} F_X(x) = F_X(x_0).$

Die folgenden **Rechenregeln** erweisen sich häufig als nützlich:

- (1) $P(a < X \leq b) = F_X(b) - F_X(a),$
denn $F_X(b) = P((X \leq a) \cup (a < X \leq b)) = P(X \leq a) + P(a < X \leq b),$
- (2) $P(X = a) = \lim_{h \rightarrow 0+0} P(a - h < X \leq a) = F_X(a) - \lim_{h \rightarrow 0+0} F_X(a - h),$
- (3) $P(X > a) = 1 - P(X \leq a) = 1 - F_X(a),$
- (4) $P(a \leq X \leq b) = P(a < X \leq b) + P(X = a),$
- (5) $P(a \leq X < b) = P(a \leq X \leq b) - P(X = b).$

3.3 Diskrete und stetige Zufallsgrößen

Definition:

Eine Zufallsgröße X heißt **diskret (verteilt)**, wenn sie höchstens abzählbar viele Werte x_i annehmen kann.

Die Wahrscheinlichkeitsverteilung einer diskreten Zufallsgröße X ist durch die Angabe der möglichen Werte x_i und ihrer Wahrscheinlichkeiten $P(X = x_i) =: p_i$ eindeutig bestimmt. Dabei muss $\sum p_i = 1$ erfüllt sein.

Es gilt

$$\begin{aligned}
P(X \in A) &= \sum_{\{i: x_i \in A\}} P(X = x_i) = \sum_{\{i: x_i \in A\}} p_i, \\
F_X(x) = P(X \leq x) &= \sum_{\{i: x_i \leq x\}} P(X = x_i) = \sum_{\{i: x_i \leq x\}} p_i \quad (\text{Treppenfunktion}).
\end{aligned}$$

Definition:

Eine Zufallsgröße X heißt **stetig (verteilt)**, wenn es eine nichtnegative Funktion f_X derart gibt, dass

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

gilt. Die Funktion $f_X|_{\mathbb{R}^1} \rightarrow \mathbb{R}^+$ heißt **Dichte** der Zufallsgröße $X.$

Beispiel für eine Dichte und die dazugehörige Verteilungsfunktion:

$$f_X(t) = \begin{cases} 1, & \text{falls } 0 \leq t \leq 1, \\ 0 & \text{sonst.} \end{cases} \quad F_X(x) = \int_0^x f_X(t)dt = \begin{cases} 0, & \text{falls } x \leq 0, \\ x, & \text{falls } 0 < x \leq 1, \\ 1, & \text{falls } x > 1. \end{cases}$$

Für stetige Zufallsgrößen gelten die folgenden **Beziehungen**:

$$(1) \int_{-\infty}^{\infty} f_X(t)dt = 1.$$

(2) F_X ist stetig.

(3) $f_X(x) = \frac{d}{dx}F_X(x)$ in allen Stetigkeitspunkten x von f_X .

$$(4) P(a < X \leq b) = F(b) - F(a) = \int_a^b f_X(t)dt$$

(5) $P(X = x_0) = 0$, denn $P(X = x_0) = \lim_{h \rightarrow 0+0} F_X(x_0 + h) - F_X(x_0) = 0$.

$$(6) P(X \in A) = \int_{t \in A} f_X(t)dt.$$

Bemerkung:

Dichtefunktionen, die sich in höchstens endlich vielen Punkten unterscheiden, beschreiben dieselbe Verteilung.

3.4 Erwartungswert und Varianz einer Zufallsgröße

Definition:

a) Ist X eine diskrete Zufallsgröße mit den Werten x_i und den Einzelwahrscheinlichkeiten $p_i = P(X = x_i)$ $i = 1, 2, \dots$, und gilt $\sum_{i=1}^{\infty} |x_i|p_i < \infty$, so heißt

$$E(X) := \sum_{i=1}^{\infty} x_i p_i$$

Erwartungswert der Zufallsgröße X .

b) Ist X eine stetige Zufallsgröße mit der Dichtefunktion f_X und gilt $\int_{-\infty}^{\infty} |t|f_X(t)dt < \infty$, so heißt

$$E(X) := \int_{-\infty}^{\infty} t f_X(t)dt$$

Erwartungswert der Zufallsgröße X .

Falls $\sum_{i=1}^{\infty} |x_i| p_i = \infty$ bzw. $\int_{-\infty}^{\infty} |t| f_X(t) dt = \infty$ gilt, existiert $E(X)$ nicht. Die Forderung nach Existenz des Erwartungswertes wird häufig durch $E(|X|) < \infty$ ausgedrückt.

Definition:

$$\text{var}(X) := E[(X - E(X))^2]$$

heißt **Varianz** (oder Streuung) der Zufallsgröße X , falls der Erwartungswert von $(X - E(X))^2$ existiert.

$\sqrt{\text{var}(X)}$ wird als **Standardabweichung** bezeichnet.

Bemerkung:

Anstelle von $\text{var}(X)$ wird auch die Schreibweise $D^2(X)$ verwendet.

Häufig wird der Erwartungswert einer Zufallsgröße Y der Gestalt $Y = g(X)$ benötigt, wobei die Verteilung der Zufallsgröße X bekannt ist.

$E(g(X))$ lässt sich (außer über die Berechnung der Verteilung von Y) auf folgende Weise bestimmen:

$$E(Y) = E(g(X)) = \begin{cases} \sum_{i=1}^{\infty} g(x_i) \cdot p_i, & \text{falls } X \text{ diskret verteilt ist,} \\ \int_{-\infty}^{\infty} g(t) f_X(t) dt, & \text{falls } X \text{ stetig verteilt ist.} \end{cases}$$

Bemerkung:

Damit $Y = g(x)$ wieder eine Zufallsgröße darstellt, muss g folgende Eigenschaft haben:

$$g^{-1}(B) = \{x \in \mathbb{R}^1 | g(x) \in B\} \in \mathcal{B}^1 \quad \forall B \in \mathcal{B}^1.$$

Diese Eigenschaft heißt **Borel-Messbarkeit**. Sie ist für stetige Funktionen erfüllt.

Beispiel

X besitze die Dichtefunktion $f_X(t) = \begin{cases} 1 & \text{falls } 0 \leq t \leq 1, \\ 0 & \text{sonst.} \end{cases}$

$$\text{Dann gilt } E(X^2) = \int_{-\infty}^{\infty} t^2 f_X(t) dt = \int_0^1 t^2 dt = \left[\frac{t^3}{3} \right]_0^1 = \frac{1}{3}.$$

Wichtige Aussagen und Rechenregeln:

X und Y seien Zufallsgrößen, und die im Folgenden auftretenden Erwartungswerte sollen existieren.

(1) $E(aX + b) = aE(X) + b$, $a \in \mathbb{R}^1$, $b \in \mathbb{R}^1$, denn es gilt

$$E(aX + b) = \sum_{i=1}^{\infty} (ax_i + b)p_i = \sum_{i=1}^{\infty} ax_i p_i + \sum_{i=1}^{\infty} bp_i = a \underbrace{\sum_{i=1}^{\infty} x_i p_i}_{E(X)} + b \underbrace{\sum_{i=1}^{\infty} p_i}_1$$

$$\text{bzw. } E(aX + b) = \int_{-\infty}^{\infty} (at + b)f_X(t)dt = a \int_{-\infty}^{\infty} t f_X(t)dt + b \int_{-\infty}^{\infty} f_X(t)dt.$$

(2) $E(X + Y) = E(X) + E(Y)$ (Die Begründung erfolgt im Abschnitt „Zufallsvektoren“).

(3) $\text{var}(X) = E(X^2) - (E(X))^2$, denn

$$\begin{aligned} \text{var}(X) &= E[(X - E(X))^2] = E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

(4) $\text{var}(aX + b) = a^2 \text{var}(X)$, $a, b \in \mathbb{R}^1$.

(5) $\text{var}(X) = 0 \Leftrightarrow \exists a \in \mathbb{R}^1 : P(X = a) = 1$.

(6) $\text{var}X \leq E(X - c)^2 \quad \forall c \in \mathbb{R}^1$.

(7) **Tschebyschevsche Ungleichung:**

Es sei X eine Zufallsgröße, deren Varianz existiert. Dann gilt für jedes $\varepsilon > 0$

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{var}(X)}{\varepsilon^2}.$$

Beweis :

Der Beweis wird hier nur für eine stetige Zufallsgröße mit Dichtefunktion f angegeben. Im diskreten Fall kann der Beweis analog erfolgen.

Zur Abkürzung wird $\mu := E(X)$ gesetzt. Dann erhält man

$$\begin{aligned} \text{var}(X) &= E[(X - E(X))^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \geq \int_{|x-\mu| \geq \varepsilon} (x - \mu)^2 f(x)dx \\ &\geq \int_{|x-\mu| \geq \varepsilon} \varepsilon^2 f(x)dx = \varepsilon^2 \int_{|x-\mu| \geq \varepsilon} f(x)dx = \varepsilon^2 P(|X - \mu| \geq \varepsilon). \square \end{aligned}$$

3.5 Spezielle Verteilungen

3.5.1 Binomialverteilung

Eine Zufallsgröße X heißt **binomialverteilt** mit den Parametern n ($n = 1, 2, \dots$) und p ($0 < p < 1$), wenn X die Werte $0, 1, 2, \dots, n$ mit den Wahrscheinlichkeiten

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

annimmt.

Es gilt $E(X) = np$, $\text{var}(X) = np(1 - p)$.

Beweis der Gleichung für den Erwartungswert:

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n \frac{(n-1) \cdots (n-k+1)}{(k-1)!} p^{(k-1)} (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \stackrel{l:=k-1}{=} np \sum_{l=0}^{n-1} \binom{n-1}{l} p^l (1-p)^{n-1-l} = np. \square \end{aligned}$$

Bemerkung:

Eine binomialverteilte Zufallsgröße mit p und $n = 1$ heißt auch **Bernoulli-verteilt** (zweipunktverteilt auf 0 und 1).

Das im Folgenden dargestellte Modell kann häufig ausgenutzt werden, um eine Binomialverteilung zu erkennen. Es wird **Bernoulli-Schema** genannt.

Ausgangspunkt ist ein Zufallsexperiment, bei dem nur unterschieden wird, ob ein Ereignis A eintritt oder nicht. Die Wahrscheinlichkeit $P(A)$ muss bekannt oder (z.B. über die klassische Wahrscheinlichkeitsdefinition) berechenbar sein.

Das Zufallsexperiment wird n -mal unabhängig voneinander wiederholt.

Die Zufallsgröße X bezeichne die **absolute Häufigkeit von A** in den n Wiederholungen des Zufallsexperimentes. Es wird gezeigt, dass die Verteilung von X gerade eine Binomialverteilung darstellt. Dies geschieht mit Hilfe der Ereignisse

$A_i \dots A$ tritt im i -ten Versuch ein.

Es gilt

$$\begin{aligned} P(X = k) &= P\left(\bigcup_{(i_1, \dots, i_k) \in C_n^k} (A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap \left(\bigcap_{j \in \{1 \dots n\} \setminus \{i_1, \dots, i_k\}} \bar{A}_j\right))\right) \\ &= \sum_{(i_1, \dots, i_k) \in C_n^k} P(A_{i_1} \cap \dots \cap A_{i_k} \cap \left(\bigcap_{j \in \{1 \dots n\} \setminus \{i_1, \dots, i_k\}} \bar{A}_j\right)). \end{aligned}$$

Dabei bezeichnet C_n^k die Menge der Kombinationen ohne Wiederholung von n Elementen zur k -ten Klasse (d.h. die Menge der Möglichkeiten, k Elemente aus einer Klasse von n Elementen auszuwählen).

Es gibt $\binom{n}{k}$ Möglichkeiten, k Elemente aus einer Menge von n Elementen auszuwählen. Damit und der Beziehung

$$\begin{aligned} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap \left(\bigcap_{j \in \{1 \dots n\} \setminus \{i_1, \dots, i_k\}} \bar{A}_j\right)) &= P(A_{i_1}) \cdot \dots \cdot P(A_{i_k}) \cdot \prod_{j \in \{1 \dots n\} \setminus \{i_1, \dots, i_k\}} P(\bar{A}_j) \\ &= p^k (1-p)^{n-k} \end{aligned}$$

ergibt sich schließlich

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

Aus der Verteilung der absoluten Häufigkeit kann die Verteilung der **relativen Häufigkeit** abgeleitet werden:

Es sei X die absolute Häufigkeit eines Ereignisses A in n Versuchen. Die Zufallsgröße $H_n := \frac{X}{n}$ heißt dann relative Häufigkeit von A in n Versuchen.

$$\begin{aligned} \text{Es gilt } P(H_n = \frac{k}{n}) &= P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \\ E(H_n) &= E(\frac{X}{n}) = \frac{1}{n} E(X) = \frac{1}{n} np = p, \\ \text{var}(H_n) &= \text{var}(\frac{X}{n}) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}. \end{aligned}$$

3.5.2 Hypergeometrische Verteilung

Die hypergeometrische Verteilung steht in enger Beziehung zum Urnenmodell ohne Zurücklegen.

Eine Zufallsgröße X heißt **hypergeometrisch verteilt** mit den Parametern N, M und n

($0 \leq M \leq N, 0 \leq n \leq N$), wenn sie die Werte k mit $\max\{0, n + M - N\} \leq k \leq \min\{M, n\}$ mit den Einzelwahrscheinlichkeiten

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

annimmt.

$$\text{Es gilt } E(X) = n \cdot \frac{M}{N} \text{ und } \text{var}(X) = n \cdot \frac{M}{N} (1 - \frac{M}{N}) \frac{N-n}{N-1}.$$

Ein wichtiges Anwendungsgebiet ist die statistische Qualitätskontrolle.

Satz:

Für $N \rightarrow \infty$ und $\frac{M}{N} \rightarrow p$ konvergieren die Einzelwahrscheinlichkeiten der hypergeometrischen Verteilung gegen die der Binomialverteilung mit den Parametern n und p .

Beweis:

$$\begin{aligned} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} &= \frac{n!}{k!(n-k)!} \frac{M(M-1)\cdots(M-k+1) \cdot (N-M)(N-M-1)\cdots(N-M-(n-k)+1)}{N(N-1)\cdots(N-n+1)} \\ &= \binom{n}{k} \frac{\frac{M}{N} \frac{M-1}{N} \cdots \frac{M-k+1}{N} \cdot \frac{N-M}{N} \frac{N-M-1}{N} \cdots \frac{N-M-(n-k)+1}{N}}{\frac{N}{N} \frac{N-1}{N} \cdots \frac{N-n+1}{N}} \\ &= \binom{n}{k} \frac{\frac{M}{N} (\frac{M}{N} - \frac{1}{N}) \cdots (\frac{M}{N} - \frac{k-1}{N}) \cdot (1 - \frac{M}{N}) (1 - \frac{M+1}{N}) \cdots (1 - \frac{M+n-k-1}{N})}{1 \cdot (1 - \frac{1}{N}) \cdots (1 - \frac{n-1}{N})} \end{aligned}$$

Die Faktoren im Nenner des Bruches konvergieren alle gegen 1. Die

ersten k Faktoren im Zähler des Bruches konvergieren gegen p ; die übrigen $n - k$ Faktoren konvergieren gegen $1 - p$. Also ergibt sich

$$\lim_{N \rightarrow \infty, \frac{M}{N} \rightarrow p} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}. \quad \square$$

3.5.3 Poisson-Verteilung

Eine Zufallsgröße X heißt **Poisson-verteilt** mit dem Parameter $\lambda > 0$, wenn X die Werte $0, 1, 2, \dots$ mit den Wahrscheinlichkeiten

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

annimmt.

Es gilt: $E(X) = \lambda$, $\text{var}(X) = \lambda$.

Die erste Gleichung ergibt sich wie folgt:

$$E(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda} = \lambda.$$

Die Poisson-Verteilung findet unter anderem Anwendung in der Warteschlangentheorie (Anzahl der Kunden, die in einer festen Zeiteinheit in einem Bedienungssystem ankommen) und der Zuverlässigkeitstheorie (Anzahl von Ausfällen in einer festen Zeiteinheit).

Satz (Poisson 1837):

Es sei $np_n = \lambda = \text{const}$. Dann konvergieren die Einzelwahrscheinlichkeiten der Binomialverteilung mit wachsendem n gegen die der Poisson-Verteilung, d.h.

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Beweis: Es gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square \end{aligned}$$

Beispiel

X bezeichne die Anzahl der Personen aus einem vorgegebenen Personenkreis von 500 Personen, die an einem bestimmten Tag Geburtstag haben. Es wird angenommen, dass alle Tage eines Jahres gleichberechtigt sind, d.h., die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Person an einem vorgegebenen Tag des Jahres

Geburtstag hat, sei $\frac{1}{365}$.

Die Zufallsgröße X ist dann binomialverteilt mit den Parametern $n = 500$ und $p = \frac{1}{365}$. Die folgende Tabelle gibt gerundete Einzelwahrscheinlichkeiten dieser Binomialverteilung und der approximierenden Poisson-Verteilung mit dem Parameter $\lambda = \frac{500}{365}$ an. Man beachte, dass die X nur Werte zwischen 0 und 500 annehmen kann, während die Einzelwahrscheinlichkeiten der Poisson-Verteilung für jedes $n \in \{0, 1, \dots\}$ größer als 0 sind.

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---------------|--------|--------|--------|--------|--------|--------|--------|-----|
| Binomialvert. | 0,2537 | 0,3484 | 0,2385 | 0,1089 | 0,0872 | 0,0101 | 0,0023 | ... |
| Poisson-Vert. | 0,2541 | 0,3481 | 0,2385 | 0,1089 | 0,0373 | 0,0102 | 0,0023 | ... |

3.5.4 Weitere diskrete Verteilungen

Weitere wichtige diskrete Verteilungen sind die

- **geometrische Verteilung** $P(X = k) = (1 - p)p^k$, $k = 0, 1, 2, \dots$, ($0 < p < 1$) und die
- **negative Binomialverteilung** $P(X = k) = \binom{-v}{k}(-p)^k(1 - p)^v$, $k = 0, 1, 2, \dots$ ($0 < p < 1$, $v > 0$).
- Von einer **Einpunktverteilung** spricht man im Fall $P(X = x_0) = 1$.

3.5.5 Gleichmäßige stetige Verteilung

Eine Zufallsgröße X heißt **gleichmäßig stetig verteilt** auf dem Intervall $[a, b]$, wenn sie die Dichte

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b, \\ 0 & \text{sonst} \end{cases}$$

besitzt.

Aufgrund der Gestalt der Dichtefunktion spricht man auch von einer Rechteckverteilung und verwendet die abkürzende Schreibweise $X \sim R[a, b]$.

Es gilt $E(X) = \frac{a+b}{2}$, $\text{var}(X) = \frac{(b-a)^2}{12}$.

Die erste Gleichung ergibt sich wie folgt

$$E(X) = \int_{-\infty}^{\infty} t f_X(t) dt = \int_a^b \frac{t}{b-a} dt = \frac{1}{b-a} \left[\frac{t^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

3.5.6 Exponentialverteilung

Eine Zufallsgröße X heißt **exponentialverteilt** mit dem Parameter $\lambda > 0$, wenn sie die Dichte

$$f_X(x) = \begin{cases} 0 & \text{für } x < 0, \\ \lambda e^{-\lambda x} & \text{für } x \geq 0 \end{cases}$$

besitzt.

$$\text{Es gilt } F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } x \geq 0 \end{cases}, \quad E(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}.$$

Anwendungen findet die Exponentialverteilung unter anderem in der Warteschlangentheorie (Bedienungszeit, Zeit zwischen der Ankunft zweier Kunden) und der Zuverlässigkeitstheorie (Lebensdauer).

Das folgende Beispiel soll deutlich machen, unter welchen Voraussetzungen die Exponentialverteilung gut zur Beschreibung der Lebensdauer eines Gerätes geeignet ist, und gleichzeitig einige Grundbegriffe aus der Zuverlässigkeitstheorie bereitstellen.

Beispiel

Die Zufallsgröße X bezeichne die Lebensdauer eines technischen Gerätes. Es ist bekannt, dass das Gerät bereits t Zeiteinheiten arbeitet. Welche Verteilung hat die verbleibende Lebensdauer des Gerätes?

Es wird angenommen, dass die Lebensdauer X stetig verteilt mit der Verteilungsfunktion F_X ist.

Als **Zuverlässigkeit** (Überlebenswahrscheinlichkeit) wird die Funktion R_X bezeichnet, die durch $R_X := P(X \geq x) = 1 - F_X(x)$ definiert ist.

Sei $h > 0$ beliebig. Dann gilt für ein t mit $P(X \geq t) > 0$

$$P(X \leq t+h | X \geq t) = \frac{P(t \leq X \leq t+h)}{P(X \geq t)} = \frac{F_X(t+h) - F_X(t)}{1 - F_X(t)} = \frac{R_X(t) - R_X(t+h)}{R_X(t)}.$$

Die im Folgenden definierte Funktion λ heißt **Ausfallrate**:

$$\begin{aligned} \lambda(t) &:= \lim_{h \rightarrow 0+0} \frac{P(X \leq t+h | X \geq t)}{h} = \lim_{h \rightarrow 0+0} \frac{F_X(t+h) - F_X(t)}{h} \cdot \frac{1}{1 - F_X(t)} \\ &= \frac{1}{1 - F_X(t)} f_X(t) = -\frac{(R_X)'(t)}{R_X(t)}. \end{aligned}$$

Die Wahrscheinlichkeit für den Ausfall eines Gerätes, das bereits t Zeiteinheiten arbeitet, in einem kurzen Zeitintervall $(t, t+h]$ ist also näherungsweise gleich $\lambda(t) \cdot h$.

Die Ausfallrate ist häufig bekannt, z.B. aufgrund von Schätzungen. Sie hat in der Regel als Funktion von t einen charakteristischen Verlauf (Badewannenkurve).

Im Weiteren soll deshalb $R(t) := R_X(t)$ aus einer bekannten Ausfallrate ermittelt werden. Die Gleichung

$$-\frac{R'(t)}{R(t)} = \lambda(t) \quad (\text{mit der Anfangsbedingung } R(0) = 1)$$

ist eine Differentialgleichung mit trennbaren Variablen.

Es wird nun die **zusätzliche Annahme** $\lambda(t) = \lambda = \text{const.}$ getroffen. Von dieser Annahme kann man im mittleren Bereich der Badewannenkurve ausgehen.

R kann dann wie folgt berechnet werden:

$$\begin{aligned} -\frac{R'(t)}{R(t)} &= \lambda \\ \int \frac{dR}{R} &= -\int \lambda dt \\ \ln R &= -\lambda t + c \quad (R > 0) \\ R(t) &= C_1 e^{-\lambda t}. \end{aligned}$$

Wegen $R(0) = 1$ erhält man $R(t) = e^{-\lambda t}$ und $F_X(t) = 1 - e^{-\lambda t}$ ($t \geq 0$),

d.h., die Lebensdauer X besitzt eine Exponentialverteilung.

Weiterhin erhält man für $h \geq 0$

$$P(X \leq t + h | X \geq t) = \frac{(1 - e^{-\lambda(t+h)}) - (1 - e^{-\lambda t})}{e^{-\lambda t}} = 1 - e^{-\lambda h},$$

die verbleibende Lebensdauer besitzt also die gleiche Verteilung wie die Lebensdauer X . Diese Eigenschaft der Exponentialverteilung wird **Gedächtnislosigkeit** genannt.

Bemerkung:

Für die Ausfallrate $\lambda(t) = \alpha \cdot \beta \cdot t^{\beta-1}$ ($\alpha > 0$, $\beta > 0$) ergibt sich eine **Weibull-Verteilung**. Die Klasse der Weibull-Verteilungen enthält die Exponentialverteilung als Spezialfall ($\beta = 1$).

3.5.7 Normalverteilung

Eine Zufallsgröße X heißt **normalverteilt** mit den Parametern μ und σ^2 ($\mu \in \mathbb{R}^1, \sigma > 0$) (in Kurzschreibweise $X \sim N(\mu, \sigma^2)$), wenn sie die Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (x \in \mathbb{R}^1)$$

besitzt.

Es gilt $E(X) = \mu$, $\text{var}(X) = \sigma^2$.

Die Verteilungsfunktionen der Normalverteilungen sind nicht geschlossen angebar. Ihre Werte können aber aus den Werten der Verteilungsfunktion der

sogenannten standardisierten Normalverteilung $N(0, 1)$, für die Tabellen zur Verfügung stehen, berechnet werden.

Für Dichte und Verteilungsfunktion der **standardisierten Normalverteilung** gibt es spezielle Bezeichnungen:

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

$$\text{Verteilungsfunktion: } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Da φ eine gerade Funktion ist, gilt

$$\varphi(x) = \varphi(-x),$$

$$\Phi(-x) = 1 - \Phi(x) \quad (\Phi \text{ ist deshalb i. Allg. nur für } x \geq 0 \text{ tabelliert}),$$

$$\Phi(0) = \frac{1}{2}.$$

Satz:

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X-\mu}{\sigma} \sim N(0, 1).$$

Beweis:

Es seien $X \sim N(\mu, \sigma^2)$, $Y := \frac{X-\mu}{\sigma}$. Dann gilt

$$F_Y(x) = P\left(\frac{X-\mu}{\sigma} \leq x\right) = P(X \leq \sigma x + \mu) = F_X(\sigma x + \mu)$$

$$= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \Phi(x).$$

Die umgekehrte Richtung kann durch analoge Überlegungen gezeigt werden. \square

Zur **Berechnung von $P(X \leq t)$** , $X \sim N(\mu, \sigma^2)$, kann man nun folgendermaßen vorgehen:

$$P(X \leq t) = P\left(\underbrace{\frac{X-\mu}{\sigma}}_{\sim N(0,1)} \leq \frac{t-\mu}{\sigma}\right) = \Phi\left(\frac{t-\mu}{\sigma}\right).$$

Beispiel

Es sei $X \sim N(\mu, \sigma^2)$. Wie groß ist die Wahrscheinlichkeit dafür, dass X einen Wert zwischen $\mu - \sigma$ und $\mu + \sigma$ (d.h. innerhalb der „ $1 - \sigma$ -Grenzen“) annimmt?

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq \frac{X-\mu}{\sigma} \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0,6827.$$

Weiterhin gilt: $P(|X - \mu| \leq 2\sigma) \approx 0,9545$,

$$P(|X - \mu| \leq 3\sigma) \approx 0,9973.$$

Die Normalverteilung entsteht beispielsweise durch Überlagerung (Addition) sehr vieler voneinander unabhängiger Einzeleffekte (\rightarrow zentraler Grenzwertsatz) und eignet sich gut zur Beschreibung von Messfehlern.

3.5.8 Cauchy-Verteilung

Die Zufallsgröße X heißt **Cauchy-verteilt**, wenn sie die Dichte

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad (x \in \mathbb{R}^1)$$

besitzt.

(Es handelt sich hierbei um einen Spezialfall der Funktionen $f_X(x) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x-\mu)^2}$, die ebenfalls als Dichten Cauchy-verteilter Zufallsgrößen bezeichnet werden.)

Für die Cauchy-Verteilung gilt

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \lim_{A \rightarrow \infty} \frac{1}{\pi} [\ln(1+x^2)]_0^A = \infty,$$

d.h., der Erwartungswert existiert nicht.

Die Cauchy-Verteilung besitzt praktische Bedeutung vor allem in der Physik. Sie heißt dort auch Breit-Wigner-Verteilung.

3.5.9 Weitere wichtige stetige Verteilungstypen

- Die **Weibull-Verteilung** ist eine Lebensdauerverteilung, die die Exponentialverteilung als Spezialfall enthält.
- Die **logarithmische Normalverteilung** entsteht durch multiplikative Überlagerung vieler unabhängiger Einzeleffekte. Sie hat Bedeutung unter anderem als Lebensdauerverteilung.
- Die **Erlang-Verteilung** entsteht als Summe von unabhängigen, identisch exponentialverteilten Zufallsgrößen und ist vor allem in der Warteschlangentheorie von Bedeutung.
- Die **Pareto-Verteilung** wird in der Ökonomie und im Versicherungswesen benötigt.
- Die **Prüfverteilungen** χ^2 -Verteilung, t -Verteilung und F -Verteilung werden in der mathematischen Statistik, zum Beispiel beim Testen von Hypothesen, benötigt. Sie werden im 2. Teil des Vorlesungsskriptes erläutert.

3.6 Die Verteilung einer Funktion einer Zufallsgröße

Gegeben seien eine Zufallsgröße X mit der Verteilungsfunktion F_X und eine Borel-messbare Funktion $g: \mathbb{R}^1 \rightarrow \mathbb{R}^1$. Gesucht ist die Verteilung der Zufallsgröße $Y := g(X)$.

Beispiel Es seien $Y = aX + b$, $a, b \in \mathbb{R}^1$, $a \neq 0$.

$$\alpha) \text{ Im Fall } a > 0 \text{ gilt } F_Y(y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a}).$$

$$\begin{aligned} \beta) \text{ Im Fall } a < 0 \text{ gilt } F_Y(y) &= P(aX + b \leq y) = P(X \geq \frac{y-b}{a}) \\ &= 1 - F_X(\frac{y-b}{a}) + P(X = \frac{y-b}{a}). \end{aligned}$$

Falls X stetig verteilt ist, ist auch Y stetig verteilt, und es folgt

$$\begin{aligned} f_Y(y) &= f_X(\frac{y-b}{a}) \cdot \frac{1}{a} && \text{im Fall } \alpha, \\ f_Y(y) &= -f_X(\frac{y-b}{a}) \cdot \frac{1}{a} && \text{im Fall } \beta. \end{aligned}$$

Diese beiden Fälle können zu $f_Y(y) = f_X(\frac{y-b}{a}) \cdot \frac{1}{|a|}$ zusammengefasst werden.

Im Folgenden sei g stetig differenzierbar, und es gelte $g'(t) \neq 0 \quad \forall t \in \mathbb{R}^1$. Es werden wiederum zwei Fälle unterschieden:

- $\alpha)$ Es sei $g'(t) > 0 \quad \forall t$. Dann gilt für $y \in g(\mathbb{R}^1)$
 $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$
- $\beta)$ Es sei $g'(t) < 0 \quad \forall t$. Dann gilt für $y \in g(\mathbb{R}^1)$

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) = P(X \geq g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)) + P(X = g^{-1}(y)). \end{aligned}$$

Ist X stetig verteilt, folgt weiterhin

$$\left. \begin{aligned} \alpha) f_Y(y) &= f_X(g^{-1}(y)) \cdot (g^{-1}(y))' \\ \beta) f_Y(y) &= -f_X(g^{-1}(y)) \cdot (g^{-1}(y))' \end{aligned} \right\} = f_X(g^{-1}(y)) \cdot |(g^{-1}(y))'| \text{ für } y \in g(\mathbb{R}^1)$$

Für alle anderen Werte von y hat die Dichtefunktion den Wert 0.

Beispiel

Gegeben sei eine stetig verteilte Zufallsgröße X . Gesucht sind Verteilungsfunktion und Dichtefunktion von $Y = X^2$.

Für $y \leq 0$ folgt $F_Y(x) = 0$.

Für $y > 0$ ergibt sich

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) = P(|X| \leq \sqrt{y}) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \text{ und} \\ f_Y(y) &= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})). \end{aligned}$$

3.7 Weitere Kenngrößen von Verteilungen

1. Momente

- $m_k := E(X^k)$ heißt - im Falle der Existenz - **(gewöhnliches) Moment k -ter Ordnung**.
- $\mu_k := E((X - E(X))^k)$ heißt - im Falle der Existenz - **zentrales Moment k -ter Ordnung**.

Für die Normalverteilung existieren alle Momente.

2. Quantile

Sei $p \in (0, 1)$. Die Zahl

$$Q_p := \sup\{x \in \mathbb{R} : F_X(x) < p\} = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}$$

heißt **Quantil der Ordnung p** der Zufallsgröße X .

Das Quantil der Ordnung $\frac{1}{2}$ wird als **Median m** der Zufallsgröße X bezeichnet.

Falls F_X^{-1} existiert, gilt $Q_p = F_X^{-1}(p)$.

3. Modalwert

Der **Modalwert** einer Zufallsgröße ist der Wert, für den die Einzelwahrscheinlichkeit im diskreten Fall bzw. die Dichtefunktion im stetigen Fall ein lokales Maximum annimmt. Hat eine Verteilung nur einen Modalwert, heißt sie **unimodal**.

4. Schiefe (skewness) Die Kenngröße

$$\gamma_1 = \frac{E(X - E(X))^3}{\sqrt{(\text{var}(X))^3}}$$

heißt **Schiefe** der Zufallsgröße X bzw. ihrer Verteilung (Definition nach Charlier).

Für bezüglich $E(X)$ symmetrische Verteilungen gilt $\gamma_1 = 0$. Die Schiefe dient zur Charakterisierung der Asymmetrie einer Verteilung.

5. Exzess (kurtosis) Die Kenngröße

$$\gamma_2 := \frac{E(X - E(X))^4}{(\text{var}(X))^2}$$

heißt **Exzess** einer Zufallsgröße bzw. ihrer Verteilung. Eine normalverteilte Zufallsgröße hat den Exzess 3.

Sowohl Schiefe als auch Exzess werden in der Praxis dazu häufig herangezogen, um die Abweichung einer Verteilung von der Normalverteilung zu beurteilen.

3.8 Mehrdimensionale Zufallsvariablen

Es ist häufig notwendig, mehrere Komponenten einer zufälligen Erscheinung, die zusammenhängen, gleichzeitig in die Betrachtungen einzubeziehen.

Definition:

Ein geordnetes p -Tupel $\mathcal{X} = (X_1, \dots, X_p)^T$ von Zufallsgrößen heißt **p -dimensionale Zufallsvariable** (auch Zufallsvektor, zufälliger Vektor).

Zufallsvektoren werden im Folgenden als Spaltenvektoren aufgefasst.

3.8.1 Diskrete Zufallsvektoren

Die Verteilung von diskreten Zufallsvektoren $\mathcal{X} = (X_1, \dots, X_p)^T$ kann durch die Einzelwahrscheinlichkeiten

$p_{i_1 i_2 \dots i_p} = P(X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_p = x_{i_p})_{i_1=1,2,\dots,i_p=1,2,\dots}$ charakterisiert werden.

Es gilt $P(\mathcal{X} \in A) = \sum_{\{i_1, \dots, i_p: (x_{i_1}, \dots, x_{i_p}) \in A\}} p_{i_1 i_2 \dots i_p}$.

Beispiel D1:

Es seien $p = 2$, $\mathcal{X} = (X, Y)^T$.

Die folgende Tabelle enthält die Einzelwahrscheinlichkeiten $P(X = x_i, Y = y_j)$ sowie die Zeilen- und Spaltensummen. (Die Bedeutung der Zeilen- und Spaltensummen wird unten erläutert.)

| $X \setminus Y$ | $y_1 = 2$ | $y_2 = 3$ | $y_3 = 4$ | Σ |
|-----------------|------------------------|------------------------|------------------------|------------------------|
| $x_1 = 1$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $p_{1.} = \frac{1}{3}$ |
| $x_2 = 2$ | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $p_{2.} = \frac{1}{3}$ |
| $x_3 = 3$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{6}$ | $p_{3.} = \frac{1}{3}$ |
| Σ | $p_{.1} = \frac{1}{3}$ | $p_{.2} = \frac{1}{3}$ | $p_{.3} = \frac{1}{3}$ | |

Die Verteilung von X kann nun wie folgt bestimmt werden:

$$\begin{aligned}
 P(X = x_i) &= P(X = x_i, Y < \infty) = P((X = x_i) \cap \bigcup_{j=1}^3 (Y = y_j)) \\
 &= P(\bigcup_{j=1}^3 (X = x_i, Y = y_j)) = \sum_{j=1}^3 P(X = x_i, Y = y_j) =: p_{i.} \quad (\text{Zeilensummen}).
 \end{aligned}$$

Analog ergibt sich aus den Spaltensummen

$$P(Y = y_j) = \sum_{i=1}^3 P(X = x_i, Y = y_j) = p_{.j}.$$

Aus den Zeilen- bzw. Spaltensummen kann also die Verteilung von X bzw. die Verteilung von Y abgelesen werden. Die Verteilungen von X und Y heißen auch **Randverteilungen** des Zufallsvektors $(X, Y)^T$.

Die **Verteilungsfunktion eines Zufallsvektors** hat folgende **Eigenschaften** (auch im stetigen Fall):

$$(1) \lim_{t_i \rightarrow -\infty} F_{\mathcal{X}}(t_1, \dots, t_p) = 0 \quad i \in \{1, 2, \dots, p\}.$$

$$\lim_{\substack{t_i \rightarrow +\infty \\ \forall i \in \{1, \dots, p\}}} F_{\mathcal{X}}(t_1, \dots, t_p) = 1.$$

$$(2) F_{\mathcal{X}} \text{ ist monoton wachsend in jeder Variablen.}$$

(3) $F_{\mathcal{X}}$ ist rechtsseitig stetig in jeder Variablen.

Fortsetzung des Beispiels D1:

Gesucht sind (zum Beispiel) die bedingten Wahrscheinlichkeiten $P(Y = y_i | X = 2)$, $j = 1, 2, 3$. Sie können wie folgt berechnet werden:

$$P(Y = 2 | X = 2) = \frac{P(Y=2, X=2)}{P(X=2)} = \frac{\frac{1}{12}}{\frac{1}{3}} = \frac{1}{4},$$

$$\text{und analog } P(Y = 3 | X = 2) = \frac{1}{2}, \quad P(Y = 4 | X = 2) = \frac{1}{4}.$$

Die Einzelwahrscheinlichkeiten $P(Y = y_j | X = 2)$, $j = 1, 2, 3$, definieren eine diskrete Verteilung auf den Werten y_1, y_2, y_3 . Diese Verteilung heißt **bedingte Verteilung** von Y unter der Bedingung ' $X = 2$ '. Der Erwartungswert bezüglich dieser Verteilung heißt **bedingter Erwartungswert** von Y unter der Bedingung ' $X = 2$ '. Für den bedingten Erwartungswert wird in der Regel die Schreibweise $E(Y | X = 2)$ benutzt.

$$\text{Es gilt } E(Y | X = 2) = \sum_j y_j P(Y = y_j | X = 2).$$

Beispiel D2: Polynomialverteilung (Multinomialverteilung)

Die Polynomialverteilung ergibt sich aus einer **Verallgemeinerung des Bernoulli-Schemas**. Ein Versuch, bei dem die Ereignisse A_i , $i = 1, \dots, l$, mit Wahrscheinlichkeit p_i auftreten können (wobei $\sum_{i=1}^l p_i = 1$ gelten muss), wird n -mal unabhängig voneinander wiederholt. Die Zufallsgrößen $X^{(i)}$, $i = 1, \dots, l$ bezeichnen jeweils die (absolute) **Häufigkeit von A_i** in diesen n Wiederholungen.

Es gilt

$$P(X^{(1)} = k_1, X^{(2)} = k_2, \dots, X^{(l-1)} = k_{l-1}) = \frac{n!}{k_1! k_2! \dots k_{l-1}! k_l!} p_1^{k_1} p_2^{k_2} \dots p_l^{k_l} \quad (*)$$

$$\text{mit } 0 \leq k_i \leq n, \quad \sum_{i=1}^{l-1} k_i \leq n; \quad k_l := n - \sum_{i=1}^{l-1} k_i.$$

Dabei ist zu beachten, dass die Ereignisse $(X^{(1)} = k_1, X^{(2)} = k_2, \dots, X^{(l-1)} = k_{l-1})$ und $(X^{(1)} = k_1, X^{(2)} = k_2, \dots, X^{(l-1)} = k_{l-1}, X^{(l)} = k_l)$ übereinstimmen.

Ein $(l-1)$ -dimensionaler Zufallsvektor, der die Werte $(k_1, k_2, \dots, k_{l-1})$ mit $0 \leq k_i \leq n$, $\sum_{i=1}^{l-1} k_i \leq n$ mit den Wahrscheinlichkeiten $(*)$

annimmt, heißt **polynomialverteilt** mit den Parametern n und p_1, \dots, p_{l-1}

$$(0 < p_i \forall i \in \{1, \dots, l-1\}, \sum_{i=1}^{l-1} p_i < 1).$$

3.8.2 Stetige Zufallsvektoren

Die Verteilung eines stetigen Zufallsvektors kann durch die Dichtefunktion $f_{\mathcal{X}}$ charakterisiert werden.

Es gilt
$$P(\mathcal{X} \in A) = \int_{\{(t_1, \dots, t_p): (t_1, \dots, t_p) \in A\}} f_{\mathcal{X}}(t_1, \dots, t_p) dt_1 \dots dt_p,$$

$$F_{\mathcal{X}}(x_1, \dots, x_p) = P(X_1 \leq x_1, \dots, X_p \leq x_p) \\ \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f_{\mathcal{X}}(t_1, \dots, t_p) dt_1 \dots dt_p.$$

Beispiel S2: Zweidimensionale (nichtsinguläre) Normalverteilung

Eine zweidimensionale (nichtsinguläre) Normalverteilung besitzt die Dichtefunktion

$$f_{\mathcal{X}}(t_1, t_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2} \frac{1}{1-\rho^2} \left(\frac{(t_1-\mu_1)^2}{\sigma_1^2} + \frac{(t_2-\mu_2)^2}{\sigma_2^2} - 2\rho \frac{(t_1-\mu_1)(t_2-\mu_2)}{\sigma_1\sigma_2} \right)},$$

$$\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1; \mu_1, \mu_2 \in \mathbb{R}^1.$$

Die Verteilungsfunktion von X kann wie folgt bestimmt werden:

$$F_X(x_1) = P(X \leq x_1) = P(X \leq x_1, Y < \infty) = \int_{t_1=-\infty}^{x_1} \int_{t_2=-\infty}^{\infty} f_{\mathcal{X}}(t_1, t_2) dt_2 dt_1.$$

Daraus ergibt sich die Dichtefunktion

$$f_X(x_1) = \frac{d}{dx_1} F_X(x_1) = \frac{d}{dx_1} \int_{t_1=-\infty}^{x_1} \int_{t_2=-\infty}^{\infty} f_{\mathcal{X}}(t_1, t_2) dt_2 dt_1 \\ = \int_{t_2=-\infty}^{\infty} f_{\mathcal{X}}(x_1, t_2) dt_2 \quad (\text{Randverteilungsdichte}).$$

Das Integral wird folgendermaßen berechnet: Es gilt

$$f_X(x_1) = \int_{t_2=-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2} \frac{1}{1-\rho^2} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(t_2-\mu_2)^2}{\sigma_2^2} - 2\rho \frac{(x_1-\mu_1)(t_2-\mu_2)}{\sigma_1\sigma_2} \right)} dt_2.$$

Mit der Substitution $\frac{t_2-\mu_2}{\sqrt{1-\rho^2}\sigma_2} =: s$ und der Abkürzung $\frac{x_1-\mu_1}{\sqrt{1-\rho^2}\sigma_1} =: r$ erhält man

$$f_X(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}r^2} \int_{s=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(s^2-2\rho rs)} ds$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(r^2 - \rho^2 r^2)} \underbrace{\int_{s=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(s - \rho r)^2} ds}_{=1 \quad (\text{setze } s - \rho r =: u)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(r^2(1 - \rho^2))} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2} \frac{(x_1 - \mu_1)^2}{\sigma_1^2}},
\end{aligned}$$

also ist X nach $N(\mu_1, \sigma_1^2)$ -verteilt.

Analog erhält man $f_Y(x_2) = \int_{t_1=-\infty}^{\infty} f_X(t_1, x_2) dt_1$ und $Y \sim N(\mu_2, \sigma_2^2)$.

Die Randverteilungen eines 2-dimensionalen (nichtsingulär) normalverteilten Zufallsvektors sind also Normalverteilungen.

Die Interpretation des sogenannten Korrelationskoeffizienten ρ erfolgt später.

Im Folgenden sollen **bedingte Verteilungen** im Fall stetiger Zufallsvektoren betrachtet werden:

Es sei $x \in \mathbb{R}^1$ mit $P(x - h < X \leq x) > 0 \quad \forall h$ mit $0 < h < h_0$ fest gewählt.

Gesucht sind $F_{Y|X=x}(y) := P(Y \leq y | X = x)$ sowie die bedingte Dichte $f_{Y|X=x}(y)$.

$$\text{Es gilt } P(Y \leq y | x - h < X \leq x) = \frac{P(Y \leq y \wedge x - h < X \leq x)}{P(x - h < X \leq x)} = \frac{\int_{t_1=x-h}^x \int_{t_2=-\infty}^y f_X(t_1, t_2) dt_2 dt_1}{\int_{t_1=x-h}^x \int_{t_2=-\infty}^{\infty} f_X(t_1, t_2) dt_2 dt_1}.$$

Wenn f stetig ist und $f_X(x) > 0$ gilt, existiert $\lim_{h \rightarrow 0+0} P(Y \leq y | x - h < X \leq x)$, und es folgt

$$\begin{aligned}
F_{Y|X=x}(y) &= \lim_{h \rightarrow 0+0} P(Y \leq y | x - h < X \leq x) = \frac{\int_{t_2=-\infty}^y f_{(X,Y)}(x, t_2) dt_2}{f_X(x)} \quad \text{und} \\
f_{Y|X=x}(y) &= \frac{d}{dy} F_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}.
\end{aligned}$$

Für die 2-dimensionale (nichtsinguläre) Normalverteilung erhält man insbesondere

$$f_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} e^{-\frac{1}{2} \frac{(y - \mu_2 - \rho \frac{(x - \mu_1)\sigma_2}{\sigma_1})^2}{(1-\rho^2)\sigma_2^2}},$$

d.h., es liegt eine $N(\mu_2 + \rho \frac{(x - \mu_1)\sigma_2}{\sigma_1}, (1 - \rho^2)\sigma_2^2)$ -Verteilung vor.

Beispiel S3: p-dimensionale (nichtsinguläre) Normalverteilung

Die Dichtefunktion einer p -dimensionalen (nichtsingulären) Normalverteilung kann in der Form

$$f_{\mathcal{X}}(t_1, \dots, t_p) = \frac{1}{\sqrt{(2\pi)^p \det B}} e^{-\frac{1}{2}(\vec{t} - \vec{\mu})^T B^{-1}(\vec{t} - \vec{\mu})}, \quad (*)$$

geschrieben werden, wobei B eine symmetrische, positiv definite Matrix bezeichnet und $\vec{\mu} = (\mu_1, \dots, \mu_p)^T, \vec{t} = (t_1, \dots, t_p)^T$ gilt.

Die 2-dimensionale (nichtsinguläre) Normalverteilung ergibt sich aus (*) mit der Matrix

$$B = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}.$$

Als abkürzende Schreibweise wird $\mathcal{X} \sim N(\vec{\mu}, B)$ benutzt.

3.8.3 Erwartungswertvektor und Kovarianzmatrix eines Zufallsvektors

Erwartungswertvektor, Kovarianz und Kovarianzmatrix sind wie folgt definiert:

Erwartungswertvektor: $E\mathcal{X} := (E(X_1), \dots, E(X_p))^T,$

Kovarianz: $cov(X_i, X_k) := E[(X_i - E(X_i)) \cdot (X_k - E(X_k))] = E(X_i X_k) - E(X_i)E(X_k),$

Kovarianzmatrix: $\Gamma_{\mathcal{X}} := [cov(X_i, X_k)]_{\substack{i=1, \dots, p \\ k=1, \dots, p}}.$

Eine Kovarianzmatrix $\Gamma_{\mathcal{X}}$ ist symmetrisch und positiv semidefinit (d.h., es gilt $\vec{a}^T \Gamma_{\mathcal{X}} \vec{a} \geq 0 \quad \forall \vec{a} \in \mathbb{R}^p$).

Die Beziehung $\vec{a}^T \Gamma_{\mathcal{X}} \vec{a} \geq 0 \quad \forall \vec{a} \in \mathbb{R}^p$ ergibt sich aus den folgenden Überlegungen:

$$\begin{aligned} 0 &\leq \text{var}(\vec{a}^T \mathcal{X}) = E\left[\left(\sum_{i=1}^p a_i X_i - \sum_{i=1}^p a_i E(X_i)\right)^2\right] = E\left[\left(\sum_{i=1}^p a_i (X_i - E(X_i))\right)^2\right] \\ &= \sum_{i=1}^p \sum_{j=1}^p a_i a_j E[(X_i - E(X_i))(X_j - E(X_j))] = \vec{a}^T \Gamma_{\mathcal{X}} \vec{a}. \end{aligned}$$

Der Wert $\rho(X_i, X_k) := \frac{cov(X_i, X_k)}{\sqrt{\text{var}(X_i) \cdot \text{var}(X_k)}}$ heißt **Korrelationskoeffizient** der Zufallsgrößen X_i und X_k .

Für $i = k$ ergibt sich $\rho(X_i, X_i) = 1$.

Falls $\rho(X_i, X_k) = 0$ erfüllt ist, heißen X_i und X_k **unkorreliert**.

Der Korrelationskoeffizient hat die folgenden Eigenschaften:

- a) $-1 \leq \rho(X_i, X_k) \leq 1,$
- b) $\rho(X_i, X_k) = \begin{matrix} +1 \\ -1 \end{matrix} \Leftrightarrow \exists a, b \in \mathbb{R}^1, b \begin{matrix} > \\ < \end{matrix} 0, \text{ mit } P(X_i = a + bX_k) = 1,$

d.h., $\rho(X_i, X_k)$ ist ein **Maß für die lineare Abhängigkeit**.

Die Korrelationskoeffizienten eines Zufallsvektors können zu der **Korrelationsmatrix** $(\rho(X_i, X_k))_{\substack{i=1, \dots, p \\ k=1, \dots, p}}$ zusammengefasst werden.

3.9 Unabhängigkeit von Zufallsgrößen

Definition:

Die Zufallsgrößen X_1, \dots, X_p heißen **(stochastisch) unabhängig**, wenn für beliebige $A_i \in \mathcal{B}^1, i = 1, \dots, p$, die Beziehung

$$P(X_1 \in A_1, \dots, X_p \in A_p) = P(X_1 \in A_1) \cdot \dots \cdot P(X_p \in A_p)$$

gilt.

Wenn X_1, \dots, X_p (stochastisch) unabhängig sind, sind sie auch paarweise unabhängig, d.h. X_i und X_k sind unabhängig für jedes Paar $(i, k), i, k \in \{1, \dots, p\}, i \neq k$.

Die folgenden Beziehungen stellen **äquivalente Charakterisierungen der Unabhängigkeit** im diskreten bzw. im stetigen Fall dar:

$$\begin{aligned} P(X_1 = x_1, \dots, X_p = x_p) &= P(X_1 = x_1) \cdot \dots \cdot P(X_p = x_p), & \text{falls } \mathcal{X} \text{ diskret,} \\ f_{\mathcal{X}}(t_1, \dots, t_p) &= f_{X_1}(t_1) \cdot \dots \cdot f_{X_p}(t_p), & \text{falls } \mathcal{X} \text{ stetig.} \end{aligned}$$

Folgerung:

Die Verteilung eines zufälligen Vektors mit unabhängigen Komponenten ist durch seine Randverteilungen bestimmt.

Fortsetzung des Beispiels D1:

Es gilt (zum Beispiel) $P(X = 1, Y = 2) = \frac{1}{6} \neq P(X = 1) \cdot P(Y = 2) = \frac{1}{9}$, d.h., X, Y sind nicht unabhängig.

Fortsetzung des Beispiels S2:

Im Fall $\rho = 0$ gilt $f_{\mathcal{X}}(t_1, t_2) = f_X(t_1) \cdot f_Y(t_2)$, d.h., X, Y sind unabhängig.

Satz (o.B):

Die Zufallsgrößen X_1, \dots, X_n seien unabhängig, und die Funktionen g_1, \dots, g_n seien Borel-messbar. Dann sind die Zufallsgrößen $g_1(X_1), \dots, g_n(X_n)$ ebenfalls unabhängig.

3.10 Funktionen eines Zufallvektors

3.10.1 Der Erwartungswert einer Funktion eines Zufallsvektors

Es seien $g: R^p \rightarrow R^1$ Borel-messbar und $Y := g(\mathcal{X})$. Dann gilt

$$E(g(X_1, \dots, X_p)) = \begin{cases} \sum_{x_1} \dots \sum_{x_p} g(x_1, \dots, x_p) P(X_1 = x_1, \dots, X_p = x_p), & \text{falls } \mathcal{X} \text{ diskret,} \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_1, \dots, t_p) f_{\mathcal{X}} dt_1 \dots dt_p, & \text{falls } \mathcal{X} \text{ stetig.} \end{cases}$$

Die folgenden Spezialfälle dieser Aussage haben besondere Bedeutung. Die Herleitung wird jeweils nur für den Fall eines stetigen Zufallsvektors angegeben. Im diskreten Fall kann analog argumentiert werden.

a) Es seien $\mathcal{X} = (X, Y)^T$, $g(x, y) = x + y$.

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{(X,Y)}(x, y) dx dy \\ &= \int_{x=-\infty}^{\infty} x \int_{y=-\infty}^{\infty} f_{(X,Y)}(x, y) dy dx + \int_{y=-\infty}^{\infty} y \int_{x=-\infty}^{\infty} f_{(X,Y)}(x, y) dx dy \\ &= E(X) + E(Y). \end{aligned}$$

b) Es werden $\mathcal{X} = (X, Y)^T$, $g(x, y) = x \cdot y$ betrachtet, und es wird zusätzlich angenommen, dass x und Y unabhängig sind. Dann gilt

$$\begin{aligned} E(X \cdot Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y f_{\mathcal{X}}(x, y) dx dy = \int_{x=-\infty}^{\infty} x \cdot f_X(x) \underbrace{\int_{y=-\infty}^{\infty} y f_Y(y) dy}_{E(Y)} dx \\ &= E(X) \cdot E(Y). \end{aligned}$$

Beide Aussagen lassen sich auf endlich viele Zufallsgrößen verallgemeinern:

$$E(X_1 + \dots + X_n) = \sum_{k=1}^n E(X_k),$$

$$E(X_1 \cdot \dots \cdot X_n) = \prod_{k=1}^n (E(X_k)), \quad \text{falls } X_1, \dots, X_n \text{ unabhängig sind.}$$

Weiterhin können die folgenden Aussagen abgeleitet werden:

Folgerung 1:

Wenn zwei Zufallsgrößen X_1, X_2 unabhängig sind, dann sind sie auch unkorreliert.

Beweis: $E[(X_1 - E(X_1)) \cdot (X_2 - E(X_2))] = E(X_1 - E(X_1)) \cdot E(X_2 - E(X_2)) = 0$.

□

Die Umkehrung von Folgerung 1 gilt i. Allg. nicht.

Folgerung 2:

X_1, \dots, X_n seien paarweise unkorreliert. Dann gilt

$$\text{var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{var}(X_i).$$

Beweis:

$$\begin{aligned} \text{var}(X_1 + \dots + X_n) &= \text{E}\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \text{E}(X_i)\right)^2\right] = \text{E}\left[\left(\sum_{i=1}^n (X_i - \text{E}(X_i))\right)^2\right] \\ &= \text{E}\left(\sum_{i=1}^n \sum_{k=1}^n (X_i - \text{E}(X_i))(X_k - \text{E}(X_k))\right) \\ &= \sum_{i=1}^n \text{E}[(X_i - \text{E}(X_i))^2] + \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \underbrace{\text{E}[(X_i - \text{E}(X_i))(X_k - \text{E}(X_k))]}_{\text{E}(X_i - \text{E}(X_i)) \cdot \text{E}(X_k - \text{E}(X_k))} \\ &= \sum_{i=1}^n \text{var}(X_i). \quad \square \end{aligned}$$

Folgerung 3:

Der Zufallsvektor $(X_1, X_2)^T$ sei normalverteilt. Dann gilt

$(X_1, X_2 \text{ unkorreliert}) \Leftrightarrow (X_1, X_2 \text{ unabhängig})$.

Zum Beweis:

Die Richtung \Leftarrow ist Inhalt von Folgerung 1. Die Gültigkeit der Implikation \Rightarrow ergibt sich aus den folgenden Überlegungen:

Durch Nachrechnen kann man zeigen, dass

$$\begin{aligned} \rho(X_1, X_2) &= \frac{1}{\sigma_1 \sigma_2} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_1 t_2 f_{(X_1, X_2)}(t_1, t_2) dt_1 dt_2 - \text{E}(X) \cdot \text{E}(Y) \right] \\ &= \dots = \rho \end{aligned}$$

gilt. Dabei bezeichnet $f_{(X_1, X_2)}$ die Dichtefunktion des Vektors $(X_1, X_2)^T$.

Im Abschnitt 3.9 wurde begründet, dass die Implikation

$(\rho = 0) \Rightarrow (X_1, X_2 \text{ unabhängig})$

für normalverteilte Zufallsvektoren richtig ist. \square

Folgerung 4:

Es sei \mathcal{X} ein p -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertvektor $\vec{\mu}$ und Kovarianzmatrix B . Dann gilt

$(X_i, X_k \text{ unkorreliert } \forall i, k \in \{1, \dots, p\}, i \neq k) \Leftrightarrow ((X_1, \dots, X_n) \text{ stochastisch unabhängig})$.

3.10.2 Die Verteilung der Summe unabhängiger Zufallsgrößen

α) Gegeben sei ein diskreter Zufallsvektor $(X, Y)^T$. Gesucht ist die Verteilung von

$$Z := X + Y.$$

Für die Einzelwahrscheinlichkeiten von Z gilt:

$$\begin{aligned} P(Z = z) &= P(X + Y = z) = \sum_{\{(x_i, y_i): x_i + y_i = z\}} P(X = x_i, Y = y_i) \\ &= \sum_{\{x_i: P(Y = z - x_i) > 0\}} P(X = x_i, Y = z - x_i) = \sum_{x_i} P(X = x_i, Y = z - x_i) \end{aligned}$$

Sind X und Y zusätzlich unabhängig, gilt $P(Z = z) = \sum_{x_i} P(X = x_i) \cdot P(Y = z - x_i)$.

β) Gegeben sei ein stetiger Zufallsvektor $(X, Y)^T$ mit Dichtefunktion $f_{(X, Y)}$. Gesucht ist die Dichtefunktion von $Z := X + Y$.

Es gilt

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = \iint_{\{(x, y): x + y \leq z\}} f_{(X, Y)}(x, y) dx dy \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{(X, Y)}(x, y) dy dx. \end{aligned}$$

Daraus ergibt sich die Dichtefunktion

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{(X, Y)}(x, y) dy dx = \int_{x=-\infty}^{\infty} f_{(X, Y)}(x, z - x) dx.$$

Sind X und Y zusätzlich unabhängig, gilt $f_Z(z) = \underbrace{\int_{x=-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) dx}_{\text{Faltung von } f_X \text{ und } f_Y}$

Beispiel:

X und Y seien unabhängige, stetige Zufallsgrößen mit

$$f_X(x) = f_Y(x) = \begin{cases} 1 & \text{für } 0 \leq x \leq 1, \\ 0 & \text{sonst.} \end{cases}$$

Gesucht ist die Dichtefunktion von $Z := X + Y$.

$$\text{Es gilt } f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) dx$$

$$\text{und } f_X(x) \cdot f_Y(z - x) \neq 0 \Leftrightarrow 0 \leq x \leq 1 \wedge \underbrace{0 \leq z - x \leq 1}_{z-1 \leq x \leq z}.$$

Daraus folgt

$$z \leq 0 : f_Z(z) = 0,$$

$$0 < z \leq 1 : f_Z(z) = \int_{x=0}^z 1 \cdot 1 dx = z,$$

$$1 < z \leq 2 : f_Z(z) = \int_{x=z-1}^1 1 dx = [x]_{z-1}^1 = 2 - z,$$

$$2 < z : f_Z(z) = 0.$$

Satz:

- (1) X_1 und X_2 seien unabhängige, mit den Parametern μ_1 und σ_1^2 bzw. μ_2 und σ_2^2 normalverteilte Zufallsgrößen. Dann ist die Summe $X_1 + X_2$ normalverteilt mit den Parametern $\mu_1 + \mu_2$ und $\sigma_1^2 + \sigma_2^2$.
- (2) (Umkehrung): Die Summe $X = X_1 + X_2$ zweier unabhängiger, nicht einpunktverteilter Zufallsgrößen X_1 und X_2 sei normalverteilt. Dann sind auch X_1 und X_2 normalverteilt.
- (3) Der Zufallsvektor \mathcal{X} sei normalverteilt mit dem Erwartungswertvektor $\vec{\mu}$ und der regulären Kovarianzmatrix B . Dann ist der Zufallsvektor $A\mathcal{X} + \vec{b}$, wobei A eine Matrix mit $\det ABA^T \neq 0$ und \vec{b} einen Vektor geeigneter Dimension bezeichnen, (nichtsingulär) normalverteilt mit dem Erwartungswertvektor $A\vec{\mu} + \vec{b}$ und der Kovarianzmatrix ABA^T .
- (4) X_1 und X_2 seien unabhängige, Poisson-verteilte Zufallsgrößen mit den Parametern λ_1 bzw. λ_2 . Dann ist die Summe $X_1 + X_2$ Poisson-verteilt mit dem Parameter $\lambda_1 + \lambda_2$.
- (5) (Umkehrung): Die Summe $X = X_1 + X_2$ zweier unabhängiger, nicht einpunktverteilter Zufallsgrößen sei Poisson-verteilt. Dann sind auch die Zufallsgrößen X_1 und X_2 Poisson-verteilt.
- (6) Die Summe zweier unabhängiger binomialverteilter Zufallsgrößen mit den Parametern (n_1, p) und (n_2, p) ist binomialverteilt mit den Parametern $(n_1 + n_2, p)$.

3.10.3 Die Verteilung weiterer Funktionen unabhängiger Zufallsgrößen

X, Y seien stetige, unabhängige Zufallsgrößen. Ähnlich zum Vorgehen bei der Herleitung der Dichtefunktion einer Summe können Formeln für die Dichten der Differenz, des Produktes und des Quotienten zweier Zufallsgrößen hergeleitet werden. Die Herleitung der Dichte des Produktes wird angegeben, weil hier (wie auch bei der Herleitung der Verteilung des Quotienten) zusätzliche Überlegungen angestellt werden müssen.

$$\alpha) Z_1 := X - Y$$

$$f_{Z_1}(z) = \int_{x=-\infty}^{\infty} f_X(x) \cdot f_Y(x - z) dx.$$

β) $Z_2 := X \cdot Y$

$$F_{Z_2}(z) = P(X \cdot Y \leq z) = \iint_{\{(x,y):x \cdot y < z\}} f_X(x) \cdot f_Y(y) dy dx.$$

$$\{(x, y) : xy \leq z\} = \{(x, y) : x \in (-\infty, 0) \wedge y \geq \frac{z}{x}\} \\ \cup \{(x, y) : x \in (0, \infty) \wedge y \leq \frac{z}{x}\} \cup A \text{ mit } A = \begin{cases} \{0\} \times \mathbb{R}^1, & \text{falls } z > 0, \\ \emptyset & \text{sonst.} \end{cases}$$

$$F_{Z_2}(z) = \int_{x=-\infty}^0 \int_{y=\frac{z}{x}}^{\infty} f_X(x) \cdot f_Y(y) dy dx + \int_{x=0}^{\infty} \int_{y=-\infty}^{\frac{z}{x}} f_X(x) \cdot f_Y(y) dy dx$$

$$f_{Z_2}(z) = - \int_{x=-\infty}^0 f_X(x) \cdot f_Y\left(\frac{z}{x}\right) \cdot \frac{1}{x} dx + \int_{x=0}^{\infty} f_X(x) \cdot f_Y\left(\frac{z}{x}\right) \cdot \frac{1}{x} dx \\ = \int_{x=-\infty}^{\infty} f_X(x) \cdot f_Y\left(\frac{z}{x}\right) \cdot \frac{1}{|x|} dx.$$

γ) $Z_3 := \frac{X}{Y}$

$$f_{Z_3}(z) = \int_{y=-\infty}^{\infty} f_X(z \cdot y) \cdot f_Y(y) |y| dy.$$

4 Grenzwertsätze

4.1 Gesetze der großen Zahlen

Gesetze der großen Zahlen sind im wesentlichen Aussagen über das Konvergenzverhalten des arithmetischen Mittels $\frac{1}{n} \sum_{i=1}^n X_i$ einer Folge von Zufallsgrößen.

Beispiel:

Ausgangspunkt der Betrachtungen ist ein Bernoulli-Schema, bei dem das Eintreten eines Ereignisses A mit $P(A) = p$ untersucht wird. Durch die Vorschrift

$$X_i = \begin{cases} 1, & \text{falls } A \text{ im } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst} \end{cases}$$

werden Zufallsgrößen X_1, \dots, X_n definiert.

Es gilt $P(X_i = 1) = P(A) = p$, $E(X_i) = p$, $\text{var}(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1 - p)$.

Die Zufallsgröße $S_n := \sum_{i=1}^n X_i$ gibt dann die absolute Häufigkeit von A in n Versuchen an; sie ist binomialverteilt mit den Parametern n und p . Somit folgt $E(S_n) = np$, $\text{var}(S_n) = np(1 - p)$.

Für die relative Häufigkeit $H_n := \frac{S_n}{n}$ von A in n Versuchen ergibt sich

$$E(H_n) = p, \text{ var}(H_n) = \text{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \text{var}(S_n) = \frac{p(1-p)}{n}.$$

Nach der Tschebyschevschen Ungleichung erhält man daraus

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(|H_n - p| > \varepsilon) = 0. \quad (*)$$

(*) wird **Gesetz der großen Zahlen von Bernoulli** genannt.

Definition:

Eine Folge $(X_k)_{k \in \mathbb{N}}$ von Zufallsgrößen heißt

konvergent in Wahrscheinlichkeit gegen $a \in \mathbb{R}^1$, wenn

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(|X_n - a| > \varepsilon) = 0 \text{ gilt}$$

(Abkürzung: $X_n \xrightarrow{\text{Wkt.}} a$),

bzw. **fast sicher konvergent gegen $a \in \mathbb{R}^1$** , wenn

$$P(\lim_{n \rightarrow \infty} X_n = a) = 1 \text{ gilt}$$

(Abkürzung: $X_n \xrightarrow{f.s.} a$).

Es gilt: $(X_n \xrightarrow{f.s.} a) \Rightarrow (X_n \xrightarrow{\text{Wkt.}} a)$.

Das „Gesetz der großen Zahlen von Bernoulli“ besagt somit, dass die relative Häufigkeit eines Ereignisses A in Wahrscheinlichkeit gegen die Wahrscheinlichkeit $P(A)$ konvergiert.

Gesetze der großen Zahlen, die Aussagen über die Konvergenz in Wahrscheinlichkeit (bzw. die fast sichere Konvergenz) machen, werden **schwache Gesetze der großen Zahlen** (bzw. **starke Gesetze der großen Zahlen**) genannt. Die Tschebyschevsche Ungleichung ist ein wichtiges Hilfsmittel zur Herleitung von schwachen Gesetzen der großen Zahlen.

Satz: (Starkes Gesetz der großen Zahlen von Kolmogorov)

Es sei $(X_k)_{k \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen, deren Erwartungswert existiert. Dann gilt $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} E(X_1)$.

4.2 Zentrale Grenzwertsätze

Eine Aussage über die Konvergenz einer Folge von Verteilungen gegen eine Grenzverteilung wird Grenzwertsatz (GWS) genannt.

Ein bereits bekanntes Beispiel ist der Grenzwertsatz von Poisson.

Zentrale Grenzwertsätze (ZGWS) machen Aussagen über die Konvergenz gegen eine Normalverteilung.

Im Weiteren sei $(X_k)_{k \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen, deren Varianzen existieren und größer als Null sind.

Außerdem seien $E(X_1) =: m$, $\text{var}(X_1) =: \sigma^2$, $S_n := X_1 + \dots + X_n$.
 Dann gilt $E(S_n) = nm$, $\text{var}(S_n) = n\sigma^2$.

Die Zufallsgrößen Z_n werden standardisiert: $Z_n := \frac{S_n - nm}{\sigma\sqrt{n}}$.

Satz: (ZGWS von Lindeberg-Levy, 1925)

Für eine Folge $(X_k)_{k \in \mathbb{N}}$ unabhängiger, identisch verteilter Zufallsgrößen, deren Varianzen existieren, gilt

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) \quad \forall x \in \mathbb{R}^1.$$

$\sum_{i=1}^n X_i$ ist daher unter den oben angegebenen Voraussetzungen für hinreichend großes n näherungsweise normalverteilt mit Erwartungswert nm und Varianz $n\sigma^2$.

Folgerung: Satz von Moivre-Laplace

Die Zufallsgrößen $Y_n, n \in \mathbb{N}$, seien binomialverteilt mit den Parametern n und p . Dann gilt für $Z_n := \frac{Y_n - np}{\sqrt{np(1-p)}}$ die Beziehung

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) \quad \forall x \in \mathbb{R}^1,$$

d.h., für hinreichend großes n kann die Verteilung einer mit n, p binomialverteilten Zufallsgröße näherungsweise durch eine $N(np, np(1-p))$ -Verteilung ersetzt werden.

Zum Beweis: Es gilt $Y_n = \sum_{i=1}^n X_i$, wobei die X_i unabhängige, mit dem Parameter p Bernoulli-verteilte Zufallsgrößen bezeichnen. □

Zur näherungsweisen Berechnung der Verteilung von Summen unabhängiger, identisch verteilter Zufallsgrößen $X_i, i = 1, \dots, n$, deren Varianzen existieren und größer als Null sind, kann man wie folgt vorgehen:

$$P(a \leq \sum_{i=1}^n X_i \leq b) = P\left(\frac{a-nm}{\sigma\sqrt{n}} \leq \underbrace{\frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}}}_{\approx N(0,1)} \leq \frac{b-nm}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{b-nm}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-nm}{\sigma\sqrt{n}}\right).$$

Sind die Zufallsgrößen X_i diskret verteilt mit ganzzahligen Werten und sind a und b ebenfalls ganzzahlig, so erhält man i. Allg. einen besseren Näherungswert durch die sogenannte

Stetigkeitskorrektur: $F_{\sum_{i=1}^n X_i}(a) \approx \Phi\left(\frac{a+0,5-nm}{\sigma\sqrt{n}}\right)$.

Mit der Stetigkeitskorrektur ergibt sich

$$\begin{aligned} P(a \leq \sum_{i=1}^n X_i \leq b) &= P(a-1 < \sum_{i=1}^n X_i \leq b) = F_{\sum_{i=1}^n X_i}(b) - F_{\sum_{i=1}^n X_i}(a-1) \\ &\approx \Phi\left(\frac{b+0,5-nm}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-0,5-nm}{\sigma\sqrt{n}}\right). \end{aligned}$$

Beispiel:

Wie groß ist die Wahrscheinlichkeit dafür, dass bei 100 Münzwürfen mindestens 65-mal die Zahl oben liegt?

Die Zufallsgröße X bezeichne die absolute Häufigkeit des Ereignisses „Zahl“ bei 100 Würfeln. X ist binomialverteilt mit $n = 100$ und $p = \frac{1}{2}$. Also gilt $E(X) = 50$, $\text{var}(X) = 25$.

Somit ergibt sich $P(X \geq 65) = \sum_{k=65}^{100} \binom{100}{k} \frac{1}{2^{100}} \approx 0,00176$.

Bei Approximation der Einzelwahrscheinlichkeiten der Binomialverteilung durch die Einzelwahrscheinlichkeiten der Poisson-Verteilung erhält man

(mit $(\lambda = np = 50)$)

$$P(X \geq 65) = \sum_{k=65}^{\infty} \frac{50^k}{k!} e^{-50} \approx 0,0236.$$

Bei Approximation mittels Normalverteilung (und Anwendung der Stetigkeitskorrektur) ergibt sich

$$P(X \geq 65) = 1 - P(X \leq 64) \approx 1 - \Phi\left(\frac{64+0,5-50}{\sqrt{25}}\right) = 1 - \Phi(2,9) \approx 0,001866.$$

Mit Hilfe der folgenden Ungleichungen kann der Fehler bei der Approximation durch die Normalverteilung abgeschätzt werden:

Für unabhängige, identisch verteilte X_i , $\text{var}(X_i) =: \sigma^2 > 0$ und $Z_n := \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}}$ gilt

$$\sup_{x \in \mathbb{R}} |F_{Z_n}(x) - \Phi(x)| \leq C \cdot \frac{E(|X_1 - E(X_1)|^3)}{\sigma^3 \sqrt{n}} \quad (\text{Ungleichung von Berry-Esseen})$$

und

$$|F_{Z_n}(x) - \Phi(x)| \leq K \cdot \frac{E(|X_1 - E(X_1)|^3)}{\sigma^3 \sqrt{n}(1+|x|^3)}.$$

Für die Konstanten wurden die Abschätzungen $C < 0,7655$ (1982), $K < 31$ (1981) bewiesen.