Stochastik für Wirtschaftswissenschaftler

Hartmut Lanzinger

Wintersemester 2011/12

Inhaltsverzeichnis

Kapitel 1

Wahrscheinlichkeiten

1 Einführung

Definition
Ein **Zufallsexperiment** ist ein Vorgang

- mit unbestimmtem Ergebnis,
- der im Prinzip unter den gleichen Bedingungen beliebig oft wiederholt werden kann, ohne dass sich die einzelnen Wiederholungen gegenseitig beeinflussen.

Die Menge Ω aller Ergebnisse eines Zufallsexperiments heißt **Grundmenge** (bzw. **Merkmalsraum**).

Beispiele

1. Ein Würfel wird einmal geworfen. Das Ergebnis ist hierbei die geworfene Augenzahl.
   $\Omega = \{1, 2, 3, 4, 5, 6\}$
2. Ein Würfel wird so lange geworfen, bis zum ersten Mal die Augenzahl 1 auftritt. Ergebnis des Zufallsexperiments ist dabei die Anzahl der dafür nötigen Würfe.
   $\Omega = \{1, 2, 3, \ldots\} = \mathbb{N}$
3. Auf einer Wegstrecke befinden sich drei Ampeln. An jeder Ampel muss ein Autofahrer entweder anhalten (A) oder er kann weiterfahren (W). Fährt ein Autofahrer diese Strecke, so ergibt sich die Grundmenge
   $$\Omega = \{WWW, WWA, WAW, AWW, AAW, AWA, WAA, AAA\}$$
4. Auf einer Wegstrecke befindet sich eine Ampel. Ergebnis des Zufallsexperiments ist die Zeit in Sekunden, die ein zufällig ausgewählter Autofahrer an dieser Ampel warten muss. Hier ist also die Grundmenge
   $$\Omega = [0, \infty)$$

Definition
Eine Menge $A \subseteq \Omega$ heißt ein **Ereignis**.
Wir sagen, dass A **eintritt**, falls für den Ausgang $\omega \in \Omega$ eines Zufallsexperiments $\omega \in A$ gilt. Im Fall $\omega \notin A$ sagen wir, dass A **nicht eintritt**.

In den oben genannten Beispielen könnten also folgende Ereignisse interessant sein:
Beispiele

1. $A =$ 'gerade Augenzahl' $= \{2, 4, 6\}$

2. $A =$ 'mehr als drei Würfe sind nötig' $= \{4, 5, 6, \dots\}$
3. $A =$ 'Der Fahrer kann immer weiterfahren' $= \{WWW\}$
4. $A =$ 'Die Wartezeit beträgt mehr als eine Minute' $= (60, \infty)$

Definition
Ein Zufallsexperiment werde n-mal durchgeführt, wobei die einzelnen Durchführungen sich gegenseitig nicht beeinflussen sollen. Die Ergebnisse seien $\omega_1, \dots, \omega_n \in \Omega$.
Dann gibt die **absolute Häufigkeit** des Ereignisses $A \subseteq \Omega$ an, wie oft das Ereignis A bei den n Versuchen eintritt, d.h. die absolute Häufigkeit ist die Anzahl der Indizes k mit $\omega_k \in A$.
Die **relative Häufigkeit** des Ereignisses $A \subseteq \Omega$ ist definiert als dessen absolute Häufigkeit dividiert durch die Gesamtanzahl n der Durchführungen des Zufallsexperiments.

Definition
Ein **Wahrscheinlichkeitsraum** ist ein Tripel bestehend aus

(a) einer Menge $\Omega \neq \emptyset$,

(b) einer Menge $\mathcal{F}$ von Teilmengen von Ω mit den Eigenschaften

- (A1) $\emptyset, \Omega \in \mathcal{F}$
- (A2) Wenn $A \in \mathcal{F}$ ist, dann ist auch $A^c \in \mathcal{F}$
- (A3) Wenn $A_n \in \mathcal{F}$ ist, dann ist auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$,

(c) einer Funktion $\mathrm{P}\colon \mathcal{F} \to [0, 1]$ mit den Eigenschaften

- (P1) $P(\Omega) = 1$ und $\mathrm{P}(\emptyset) = 0$
- (P2) ("σ-Additivität")
  Für paarweise disjunkte Mengen $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, d.h. $A_j \cap A_k = \emptyset$ für $j \neq k$ gilt

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Die Elemente von $\mathcal{F}$ heißen **Ereignisse**.
Für ein $A \in \mathcal{F}$ heißt $\mathrm{P}(A)$ die **Wahrscheinlichkeit** des Ereignisses A.
Die Funktion P heißt ein **Wahrscheinlichkeitsmaß** auf $(\Omega, \mathcal{F})$

Bemerkungen

1. Eine Menge $\mathcal{F}$ von Teilmengen einer nichtleeren Menge Ω mit den Eigenschaften (A1)-(A3) nennt man eine **σ-Algebra**. $\mathcal{F}$ enthält die Mengen, denen man eine Wahrscheinlichkeit zuordnen kann.

   Für uns ist daran nur wichtig, dass in einer σ-Algebra bestimmte Mengeoperationen erlaubt sind.

   Wenn $\mathcal{F}$ eine σ-Algebra ist mit $A, B, A_k \in \mathcal{F}$, $n \in \mathbb{N}$, dann gehören folgende Mengen ebenfalls zu $\mathcal{F}$:

$$A \setminus B, A \cup B, A \cap B, \bigcup_{k=1}^{n} A_k, \bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{n} A_k, \bigcap_{k=1}^{\infty} A_k$$

   Dies wird im Folgenden stets benutzt werden.

2. Ω entspricht der Menge der möglichen Ausgänge des entsprechenden Zufallsexperiments. Wenn für ein $\omega \in \Omega$ die einelementige Menge $\{\omega\}$ zu $\mathcal{F}$ gehört, so spricht man von einem **Elementarereigis**.

Proposition 1.1 (Eigenschaften von Wahrscheinlichkeiten)
$(\Omega, \mathcal{F}, \mathrm{P})$ sei ein Wahrscheinlichkeitsraum, $A, B_1, \ldots, B_n \in \mathcal{F}$ $(n \in \mathbb{N})$.
Dann gilt:

(a) Wenn $B_1, \ldots, B_n$ disjunkt sind, dann folgt

$$\mathrm{P}\left(\bigcup_{k=1}^{n} B_k\right) = \sum_{k=1}^{n} \mathrm{P}(B_k)$$

(b) $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

(c) $A \subseteq B \implies \mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A)$ (insb. $\mathrm{P}(A) \leq \mathrm{P}(B)$)

(d) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

(e) $\mathrm{P}\left(\bigcup_{k=1}^{n} B_k\right) \leq \sum_{k=1}^{n} P(B_k)$

2 Laplacesche Wahrscheinlichkeitsräume

Definition
Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ heißt **Laplacescher Wahrscheinlichkeitsraum**, falls gilt.

(a) Ω ist endlich, etwa $\Omega = \{\omega_1, \ldots, \omega_n\}$

(b) $\mathcal{F} = \mathcal{P}(\Omega)$ $\mathrm{P}(\{\omega_k\}) = \dfrac{1}{N}$ für $k = 1, \ldots, N$.

Bemerkung
In einem Laplaceschen Wahrscheinlichkeitsraum gilt für alle $A \in \Omega$:

$$\mathrm{P}(A) = \frac{|A|}{\Omega}$$

Beispiele

1. Werfen eines Würfels (s.o.)

2. Zweifaches Werfen eines Würfels
   $\Omega = \{(1,1); (1,2); \ldots; (1,6); (2,1); \ldots; (6,6)\}$ ($\implies |\Omega| = 36$)
   $\mathrm{P}(\{(j,k)\} = \frac{1}{36}$ für $1 \leq j, k \leq 6$
   Mit welcher Wahrscheinlichkeit ist die Augensumme 7?
   Gesucht ist $\mathrm{P}(A)$ für $A = \{(1,6); (2,5); \ldots; (6,1)\}$, also ist $Prob(A) = \frac{6}{36} = \frac{1}{6}$

3. Zwei Personen werfen eine Münze. Fällt Wappen, so bekommt Person A einen Punkt, bei Zahl erhält dagegen Person B einen Punkt. Sieger ist, wer zuerst drei Punkte auf dem Konto hat. Bei den ersten beiden Würfen ist jeweils Wappen gefallen.

   Wie verhalten sich jetzt die Siegchancen? Bezeichnen wir der Einfachheit halber Wappen mit W und Zahl mit Z, so ergibt sich folgender Wahrscheinlichkeitsraum zur Modellierung der nächsten 3 Würfe:

   $\Omega = \{WWW, WWZ, WZW, WZZ, ZWW, ZWZ, ZZW, ZZZ\}$ ($\implies |\Omega| = 8$)

   Sei A das Ereignis "Person B gewinnt". Dann ist $A = \{ZZZ\}$ und somit $P(A) = \frac{1}{8}$.

Für die Berechnung von Wahrscheinlichkeiten in Laplaceschen Wahrscheinlichkeitsräumen ist es also erforderlich, für gewisse Klassen von Mengen deren Elementezahl zu bestimmen.
Hierzu benutzt man die sogenannten **Urnenmodelle**.
Man stellt sich dabei eine Urne vor, die mit n Kugeln gefüllt ist. Die Kugeln seien von 1 bis n durchnummeriert.
Jetzt werden nacheinander k Kugeln zufällig aus der Urne entnommen. Als Ergebnis erhält man ein k-Tupen $(j_1, \ldots, j_k) \in \{1, \ldots, n\}^k$.
Der Ziehungsvorgang kann dann mit oder ohne Zurücklegenerfolgen und auch mit oder ohne Beachtung der Reihenfolge.
Dadurch ergeben sich insgesamt 4 Möglichkeiten, die wir im Folgenden nacheinander durchgehen werden.
Zur Erinnerung verwenden wir folgende Bezeichnungen aus der Mathematikvorlesung:

A Ziehen mit Zurücklegen unter Beachtung der Reihenfolge

Bei jeden Zug gibt es n Möglichkeiten, also insgesamt n^k Möglichkeiten.

Beispiel
100-maliges Werfen eines Würfels. Hier ist $n = 6$ und $k = 100$, somit ist $|\Omega| = 6^{100}$.

B Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge

Beim ersten Zug hat man n Möglichkeiten, beim zweiten noch $n - 1$, usw.

Also gibt es insgesamt $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ Möglichkeiten.

Schreibweise
$(n)_k = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ für $1 \leq k \leq n$, $k, n \in \mathbb{N}$.

Bemerkung
Im Fall $k = n$ ist $(n)_k = n!$

Mit anderen Worten ist $n!$ die Anzahl der Möglichkeiten, n Objekte anzuordnen.

C Ziehen ohne Zurücklegen und ohne Beachtung der Reihenfolge

Wieder hat man im ersten Zug n Möglichkeiten, beim zweiten noch $n-1$, usw. Jetzt gibt es aber für jedes gezogene k-Tupel noch $k!$ Möglichkeiten für die Reihenfolge.

Insgesamt gibt es also $\frac{n()_k}{k!} = \binom{n}{k}$ Möglichkeiten.

$\binom{n}{k}$ ist also die Anzahl der Möglichkeiten, k Objekte aus einer Menge mit n Objekten auszuwählen.

D Ziehen mit Zurücklegen und ohne Beachtung der Reihenfolge

Dieser Fall tritt relativ selten auf. Eine ähnliche, aber etwas komplizierter Argumentation als in A-C zeigt, dass es $\binom{n+k-1}{k}$ Möglichkeiten gibt.

1 Zusammenfassung

Ziehen von k Kugeln aus einer Urne mit n Kugeln

	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
mit Zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne Zurücklegen	$(n)_k$	$\binom{n}{k}$

Beispiele

1. Mit welcher Wahrscheinlichkeit tritt beim dreimaligen Werfen eines Würfels keine Augenzahl mehrfach auf?

   Es ist $|\Omega| = 6^3$. Für das Ereignis $A =$'keine Augenzahl mehrfach' ist $|A| = (6)_3$ und daher

$$\mathrm{P}(A) = \frac{6 \cdot 5 \cdot 4}{6^3} = \frac{5}{9}$$

2. Wie groß ist die Wahrscheinlichkeit, dass unter k zufällig ausgewählten Personen mindestens zwei den selben Geburtstag haben?

   Wir nehmen vereinfachend an, dass das Jahr 365 Tage hat, die als Geburtstage alle gleich wahrscheinlich sind. Dann ergibt sich für das Ereignis $A =$'alle Personen haben an verschiedenen Tagen Geburtstag

$$\mathrm{P}(A) = \frac{(365)_k}{365^k} = \frac{365 \cdot \ldots \cdot (365-k+1)}{365^k}$$

   Einige Zahlenwerte ergibt sich aus folgender Tabelle:

n	Wahrscheinlichkeit für mindestens zwei gleiche Geburtstage
2	0.00273972602739725
4	0.0163559124665502
7	0.0562357030959754
10	0.116948177711078
12	0.167024788838064
15	0.252901319763686
18	0.346911417871789
20	0.41143838358058
21	0.443688335165206
22	0.47569530766255
23	0.507297234323985
24	0.538344257914529
25	0.568699703969464
26	0.598240820135939
27	0.626859282263242
28	0.654461472342399
29	0.680968537477777
30	0.706316242719269
35	0.814383238874715
40	0.891231809817949
45	0.940975899465775
50	0.970373579577988

   Bemerkung
   Die Annahme, dass alle Tage als Geburtstag gleich wahrscheinlich sind, ist unrealistisch. Man kann aber zeigen, dass die Wahrscheinlichkeit dafür, dass mindestens zwei Personen am selben Tag Geburtstag haben, bei unterschiedlichen Wahrscheinlichkeiten für die einzelnen Tage, höchstens größer werden kann.

3. In einer Urne befinden sich 15 weiße und 5 schwarze Kugeln. Es werden 4 Kugeln ohne Zurücklegen und ohne Beachtung der Reihenfolge gezogen. Mit welcher Wahrscheinlichkeit sind

   (a) alle Kugeln weiß?

Es sei A das Ereignis 'alle Kugeln weiß'. Dann gilt

$$\mathrm{P}(A) = \frac{\binom{15}{4}}{\binom{20}{4}} = \frac{15 \cdot 14 \cdot 13 \cdot 12}{20 \cdot 19 \cdot 18 \cdot 17} = \frac{32760}{116280} \approx 0.28$$

(b) genau zwei Kugeln weiß und zwei Kugeln schwarz?

Es sei B das Ereignis 'genau zwei Kugeln weiß und zwei Kugeln schwarz'. Dann gilt

$$\mathrm{P}(B) = \frac{\binom{15}{3}\binom{5}{2}}{\binom{20}{4}} = \frac{1050}{116280} \approx 0.01$$

(c) mindestens drei Kugel weiß?

Es sei C das Ereignis 'mindestens drei Kugeln weiß'. Dann gilt

$$\mathrm{P}(C) = \frac{\binom{15}{3}\binom{5}{1} + \binom{15}{4}}{\binom{20}{4}} = \frac{35035}{116280} \approx 0.30$$

Proposition 1.2 (Hypergeometrische Verteilung)
Aus einer Urne mit m weißen und n schwarzen Kugeln werden k Kugeln ohne Zurücklegen und ohne Beachtung der Reihenfolge gezogen.
Die Wahrscheinlichkeit, dass unter den gezogenen Kugeln genau r weiß sind, beträgt (für $0 \leq r \leq k$:

$$\frac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}}$$

3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Definition
A, B seien Ereignisse aus $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{P}(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit von A gegeben B.**

Beispiel
Wie oben sei A das Ereignis 'weiße Kugel im zweiten Zug' und B das Ereignis 'weiße Kugel im ersten Zug'. Dann gilt:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\frac{10 \cdot 9}{20 \cdot 19}}{\frac{10}{20}} = \frac{9}{19}$$

Bemerkungen

1. $\mathrm{P}(A \mid B)$ gibt also die Wahrscheinlichkeit für das Ereignis A unter der Zusatzinformation an, dass das Ereignis B eingetreten ist.

2. Wenn $\mathrm{P}(B) > 0$ ist, folgt also $\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B) \cdot \mathrm{P}(B)$.
   Ist sogar $\mathrm{P}(A), \mathrm{P}(B) > 0$, so folgt

$$\mathrm{P}(A \mid B) \cdot \mathrm{P}(B) = \mathrm{P}(A \cap B) = \mathrm{P}(B \cap A) = \mathrm{P}(B \mid A) \cdot \mathrm{P}(A)$$

Lemma 1.1
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und Ereignisse $A, A_k \in \mathcal{F}$, $k \in \mathbb{N}$, sowie $B \in \mathcal{F}$ mit $\mathrm{P}(B) > 0$.
Dann gilt:

(a) $\mathrm{P}(\Omega \mid B) = 1$, $\mathrm{P}(\emptyset \mid B) = 0$.

(b) $\mathrm{P}(A^c \mid B) = 1 - \mathrm{P}(A \mid B)$

(c) Wenn die A_k disjunkt sind, so folgt

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k \mid B\right) = \sum_{k=1}^{\infty} \mathrm{P}(A_k \mid B)$$

(d) Aus $A_1 \subseteq A_2$ folgt $\mathrm{P}(A_1 \mid B) \leq \mathrm{P}(A_2 \mid B)$

(e) $\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k \mid B\right) \leq \sum_{k=1}^{\infty} \mathrm{P}(A_k \mid B)$

Proposition 1.3 (Multiplikationsregel)
Gegeben seien Ereignisse $A_1, \ldots, A_n$ in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{P}(A_1 \cap \ldots \cap A_n) > 0$.
Dann gilt:

$$\mathrm{P}(A_1 \cap \ldots \cap A_n) = \mathrm{P}(A_1) \cdot \mathrm{P}(A_2 \mid A_1)\mathrm{P}(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot \mathrm{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})$$

Beispiel
Gegeben sei eine Urne mit 10 Kugeln (3 schwarz, 7 weiß). Bei jedem Zug wird eine Kugel gezogen, die dann mit zwei weiteren Kugeln der selben Farbe in die Urne zurückgegelegt wird.
Mit welcher Wahrscheinlichkeit wird bei den erten drei Zügen stets eine schwarze Kugel gezogen?

Sei A_ν das Ereignis 'schwarze Kugel im ν-ten Zug' $(\nu = 1, 2, 3)$. Dann gilt

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2 \mid A_1)\mathrm{P}(A_3 \mid A_1 \cap A_2) = \frac{3}{10} \cdot \frac{5}{12} \cdot \frac{7}{14} = \frac{1}{16}$$

Definition
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ sowie $B_1, \ldots, B_n \in \mathcal{F}$, $n \in \mathbb{N}$.
$B_1, \ldots, B_n$ heißen eine **Partition** bzw. **Zerlegung** von Ω, wenn gilt:

(i) $B_1, \ldots, B_n$ sind paarweise disjunkt mit $\bigcup_{k=1}^{n} B_k = \Omega$.

(ii) $\mathrm{P}(B_k) > 0$ für $k = 1, \ldots, n$.

Satz 1.1 (Satz von der totalen Wahrscheinlichkeit, Satz von Bayes)
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, eine Zerlegung $B_1, \ldots, B_n$ von Ω und ein Ereignis $A \in \mathcal{F}$.
Dann gilt:

(a) $\mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A \mid B_k) \cdot \mathrm{P}(B_k)$

(b) Ist zusätzlich $\mathrm{P}(A) > 0$, so folgt für $\nu = 1, \ldots, n$:

$$\mathrm{P}(B_\nu \mid A) = \frac{\mathrm{P}(A \mid B_\nu) \cdot \mathrm{P}(B_\nu)}{\sum_{k=1}^n \mathrm{P}(A \mid B_k) \cdot \mathrm{P}(B_k)} = \frac{\mathrm{P}(A \mid B_\nu) \cdot \mathrm{P}(B_\nu)}{\mathrm{P}(A)}$$

Beispiele

1. Gegeben seien drei Urnen U_1, U_2 und U_3. U_1 enthalte 13 weiße und 7 schwarze Kugeln, U_2 enthalte 9 weiße und 1 schwarze Kugeln und U_3 enthalte 3 weiße und 12 schwarze Kugeln.

   Es werde zunächst eine Urne zufällig ausgewählt und dann aus dieser Urne eine Kugel gezogen.

   (a) Mit welcher Wahrscheinlichkeit ist diese Kugel schwarz?

   A sei das Ereignis 'Kugel ist schwarz', B_ν das Ereignis 'ν-te Kugel wurde gewählt. Dann gilt:

$$\begin{aligned}\mathrm{P}(A) &= \mathrm{P}(A \mid B_1) \cdot \mathrm{P}(B_1) + \mathrm{P}(A \mid B_2) \cdot \mathrm{P}(B_2) + \mathrm{P}(A \mid B_3) \cdot \mathrm{P}(B_3) \\ &= \frac{7}{20} \cdot \frac{1}{3} + \frac{1}{10} \cdot \frac{1}{3} + \frac{12}{15} \cdot \frac{1}{3} = \frac{7+2+16}{60} = \frac{25}{60} = \frac{5}{12}\end{aligned}$$

   (b) Die gezogene Kugel ist schwarz. Mit welcher Wahrscheinlichkeit stammt sie aus U_3?

   Mit den obigen Bezeichnungen gilt:

$$\mathrm{P}(B_3 \mid A) = \frac{\mathrm{P}(A \mid B_3) \cdot \mathrm{P}(B_3)}{\mathrm{P}(A)} = \frac{\frac{12}{15} \cdot \frac{1}{3}}{\frac{5}{12}}$$

2. Bei einem Test auf eine bestimmte Krankheit ist die Wahrscheinlichkeit, dass eine kranke Person richtig als krank erkannt wird, 99% ("Sensitivität"). Die Wahrscheinlichkeit, dass eine nicht kranke Person richtig als nicht krank erkannt wird, liegt bei 95% ("Spezifität").

   Wie groß ist die Wahrscheinlichkeit, dass eine Person mit positivem Testergebnis tatsächlich Krankheit hat, wenn man weiß, dass von dieser Krankheit etwa 1% der Bevölkerung betroffen ist?

   A sei das Ereignis 'Test ist positiv und B das Ereignis 'Person ist krank'.

   Nach den obigen Angaben gilt dann:

$$\mathrm{P}(B) = 0.01, \quad \mathrm{P}(B^c) = 0.99, \quad \mathrm{P}(A \mid B) = 0.99, \quad \mathrm{P}(A \mid B^c) = 0.05$$

   Aus Satz 1.1 folgt dann:

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \mid B) \cdot \mathrm{P}(B)}{\mathrm{P}(A \mid B) \cdot \mathrm{P}(B) + \mathrm{P}(A \mid B^c) \cdot \mathrm{P}(B^c)} = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.05 \cdot 9.99} = \frac{1}{6}$$

Definition
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ sowie $A, B, A_1, \ldots, A_n \in \mathcal{F}$, $n \in \mathbb{N}$.

(a) A und B heißen **unabhängig**, falls $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$ ist.

(b) $A_1, \ldots, A_n$ heißen **unabhängig**, falls für alle $k = 1, \ldots, n$ und beliebige Indizes $i_1, \ldots, i_k$ mit $1 \leq i_1 < \ldots < i_k \leq n$ gilt:

$$\mathrm{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathrm{P}(A_{i_1}) \cdot \ldots \cdot \mathrm{P}(A_{i_k})$$

Beispiele

1. Zweimaliges Werfen eines Würfels

A	sei das Ereignis 'Augensumme ungerade'	$\Longrightarrow \mathrm{P}(A) = \frac{1}{2}$
B	sei das Ereignis 'Eins im ersten Wurf'	$\Longrightarrow \mathrm{P}(B) = \frac{1}{6}$
C	sei das Ereignis 'Augensumme sieben'	$\Longrightarrow \mathrm{P}(B) = \frac{1}{6}$

   Dann gilt

   $$\mathrm{P}(A \cap B) = \mathrm{P}(\{(1,2),(1,4),(1,6)\}) = \frac{3}{36} = \frac{1}{12} = \mathrm{P}(A) \cdot \mathrm{P}(B) \qquad \Longrightarrow A, B \text{ sind unabhängig}$$

   $$\mathrm{P}(A \cap C) = \mathrm{P}(C) = \frac{1}{6} \neq \mathrm{P}(A) \cdot \mathrm{P}(C) \qquad \Longrightarrow A, C \text{ sind nicht unabhängig}$$

   $$\mathrm{P}(B \cap C) = \mathrm{P}(\{(1,6)\}) = \frac{1}{36} = \mathrm{P}(B) \cdot \mathrm{P}(C) \qquad \Longrightarrow B, C \text{ sind unabhängig}$$

Bemerkung
A, B unabhängig $\Longrightarrow$ A^c, B unabhängig und A^c, B^c unabhängig

Kapitel 2

Zufallsvariablen

1 Einführung

Definition
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.
Eine Funktion $X : \Omega \to \mathbb{R}$ heißt eine **Zufallsvariable**, falls gilt:

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F} \quad \text{für alle } x \in \mathbb{R}$$

Bemerkungen

1. Wenn $\mathcal{F}$ aus allen Teilmengen von Ω besteht, dann ist jede Funktion $X : \Omega \to \mathbb{R}$ automatisch eine Zufallsvariable.
2. Anstelle von $\{\omega \in \Omega \mid X(\omega) \leq x\}$ schreibt man meistens $\{X \leq x\}$.

   Entsprechendes gilt für Mengen wie $\{X = x\}$, $\{X \in [a, b]\}$ usw.

Beispiel
Zweimaliges Werfen eines Würfels
$\Omega = \{(j, k) \mid 1 \leq j, k \leq n\}$
$X((j, k)) = j + k$ 'Augensumme'
$Y((j, k)) = \max\{j, k\}$ 'Maximum der Augenzahlen'

Hier ergeben sich folgende Werte:

ν	2	3	4	5	6	7	8	9	10	11	12
$36\mathrm{P}(X = \nu)$	1	2	3	4	5	4	3	2	1		

ν	1	2	3	4	5	6
$36\mathrm{P}(Y = \nu)$	1	3	5	7	9	11

Definition
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und eine Zufallsvariable X.
Dann heißt die Funktion

$$F : \mathbb{R} \to [0, 1], F(x) = \mathrm{P}(X \leq x)$$

die **Verteilungsfunktion** von X.
Insbesondere dann, wenn mehrere Zufallsvariablen im Spiel sind, bezeichnen wir die Verteilungsfunktion von X auch mit F_X.

Beispiel
Zweimaliges Werfen eines Würfels, X, Y wie oben.

$$F_X(x) = \begin{cases} 0 & \text{für } x < 2 \\ \frac{1}{36} & \text{für } 2 \leq x < 3 \\ \frac{3}{36} & \text{für } 3 \leq x < 4 \\ \frac{6}{36} & \text{für } 4 \leq x < 5 \\ \frac{10}{36} & \text{für } 5 \leq x < 6 \\ \frac{15}{36} & \text{für } 6 \leq x < 7 \\ \frac{21}{36} & \text{für } 7 \leq x < 8 \\ \frac{26}{36} & \text{für } 8 \leq x < 9 \\ \frac{30}{36} & \text{für } 9 \leq x < 10 \\ \frac{33}{36} & \text{für } 10 \leq x < 11 \\ \frac{35}{36} & \text{für } 11 \leq x < 12 \\ 1 & \text{für } x \geq 12 \end{cases} \qquad F_Y(x) = \begin{cases} 0 & \text{für } x < 1 \\ \frac{1}{36} & \text{für } 1 \leq x < 2 \\ \frac{4}{36} & \text{für } 2 \leq x < 3 \\ \frac{9}{36} & \text{für } 3 \leq x < 4 \\ \frac{16}{36} & \text{für } 4 \leq x < 5 \\ \frac{25}{36} & \text{für } 5 \leq x < 6 \\ 1 & \text{für } x \geq 6 \end{cases}$$

Proposition 2.1 (Eigenschaften einer Verteilungsfunktion)
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und eine Zufallsvariable X mit Verteilungsfunktion F.
Dann gilt:

(a) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

(b) F ist monoton wachsend.

(c) F ist rechtsseitig stetig.

Bemerkung
Zu jeder Funktion $F\colon \mathbb{R} \to [0,1]$ mit den obigen Eigenschaften existiert ein Wahrscheinlichkeitsraum und eine zugehörige Zufallsvariable, die F als Verteilungsfunktion hat.

Definition
Gegeben sei eine Menge Ω und $A \subseteq \Omega$.
Dann heißt die Funktion

$$\mathbb{1}_A\colon \Omega \to \{0,1\}, \mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

die **Indikatorfunktion von** A (in Ω).

Lemma 2.1
Für einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $A \in \mathcal{F}$ ist $\mathbb{1}_A$ eine Zufallsvariable.

Lemma 2.2
Es sei Ω eine nichtleere Menge und $A, B, A_1, \ldots, A_n \subseteq \Omega$.
Dann gilt:

(a) $\mathbb{1}_\emptyset \equiv 0$, $\mathbb{1}_\Omega \equiv 1$.

(b) $\mathbb{1}_A^2(\omega) = \mathbb{1}_A(\omega)$ für alle $\omega \in \Omega$.

(c) $\mathbb{1}_{A_1\cap\ldots\cap A_n}(\omega) = \mathbb{1}_{A_1}(\omega)\cdot\ldots\cdot\mathbb{1}_{A_n}(\omega)$ für alle $\omega\in\Omega$.
Speziell ist $\mathbb{1}_{A\cap B}(\omega) = \mathbb{1}_A(\omega)\cdot\mathbb{1}_B(\omega)$ für alle $\omega\in\Omega$.

(d) $\mathbb{1}_{A_1\cup\ldots\cup A_n}(\omega) = \max\{\mathbb{1}_{A_1}(\omega),\ldots,\mathbb{1}_{A_n}(\omega)\}$ für alle $\omega\in\Omega$.

(e) $\mathbb{1}_{A\cup B}(\omega) = \mathbb{1}_A(\omega)+\mathbb{1}_B(\omega)-\mathbb{1}_{A\cap B}(\omega)$ für alle $\omega\in\Omega$.

(f) $\mathbb{1}_{\Omega\setminus A}(\omega) = 1-\mathbb{1}_A(\omega)$ für alle $\omega\in\Omega$.

(g) $A\subseteq B \iff \mathbb{1}_A(\omega)\le\mathbb{1}_B(\omega)$ für alle $\omega\in\Omega$.

2 Diskrete Zufallsvariablen

Definition
Eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$ heißt **diskret**, wenn es eine endliche oder abzählbare Menge M gibt, so dass $\mathrm{P}(X\in M)=1$ ist.

Beispiele

1. Bisher waren alle Beispiele für Zufallsvariablen diskret.
2. Wartezeit X bis zur ersten Eins beim fortgesetzten Werfen eines Würfels
   Hier ist offensichtlich $M=\mathbb{N}$.

Definition
Gegeben sei eine diskrete Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$. Dann heißt die Funktion

$$f\colon\mathbb{R}\to[0,\infty), f(x)=\mathrm{P}(X=x)$$

die **Wahrscheinlichkeitsfunktion** bzw. **Zähldichte** von X.

Beispiele

1. X Augensumme beim zweimaligen Werfen eines Würfels

ν	2	3	4	5	6	7	8	9	10	11	12
$f(\nu)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$		

   $f(x)=0$ für $x\notin\{2,3,\ldots,12\}$

2. Wartezeit bis zur ersten Eins beim fortgesetzten Werfen eines Würfels
   $f(k)=\frac{1}{6}\left(\frac{5}{6}\right)^{k-1}$ für $k\in\mathbb{N}$
   $f(x)=0$ für $x\notin\mathbb{N}$

Bemerkung
Wenn $M=\{x_1,x_2,x_3,\ldots\}$ ist, dann kann die Zähldichte durch Angabe der Wahrscheinlichkeiten $p_k=\mathrm{P}(X=x_k)$ $(k\in\mathbb{N})$ beschrieben werden.
Man gibt also gewissermaßen nur die Werte von x an, für die $\mathrm{P}(X=x)>0$ ist.
Dann ist automatisch $\sum\limits_{k=1}^{\infty}\mathrm{P}(X=x_k)=1$.

Definition
Gegeben sei eine diskrete Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Es gelte $\mathrm{P}(X \in M) = 1$, wobei $M = \{x_1, x_2, x_3, \ldots\}$.
Wenn die Reihe $\sum_{k=1}^{\infty} |x_k| \mathrm{P}(X = x_k)$ konvergiert, dann heißt

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} x_k \mathrm{P}(X = x_k)$$

der **Erwartungswert** der Zufallsvariablen X.

Bemerkung
Der Erwartungswert einer Zufallsvariablen gibt an, welchen Wert die Zufallsvariable 'im Schnitt' annimmt.
Wir werden Erwartungswerte auch häufig mit μ bezeichnen.

Beispiele

1. Werfen eines Würfels, X Augenzahl
   Dann ist $\mathrm{E}(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$
2. Zweimaliges Werfen eines Würfels, X Augensumme, Y Maximum der Augenzahlen
   Dann ist

$$\mathrm{E}(X) = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \ldots + \frac{1}{36} \cdot 12 = 7$$
$$\mathrm{E}(Y) = \frac{1}{36} \cdot 1 + \frac{3}{36} \cdot 2 + \ldots + \frac{11}{36} \cdot 6 = \frac{161}{36} \approx 4.47$$

Bemerkung
Ist die oben genannte Menge M endlich, so existiert der Erwartungswert immer.

Definition
Gegeben sei eine diskrete Zufallsvariable X mit Erwartungswert μ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Es gelte $\mathrm{P}(X \in \{x_1, x_2, x_3, \ldots\}) = 1$. Wenn die Reihe

$$\sum_{k=1}^{\infty} (x_k - \mu)^2 \mathrm{P}(X = x_k)$$

konvergiert, dann heißt der Wert dieser Reihe die **Varianz** von X und wird mt $\mathrm{Var}(X)$ (bzw. σ^2) bezeichnet.
Die Wurzel aus der Varianz bezeichnet man als **Standardabweichung** (Bezeichnung: σ)

Bemerkungen

1. Die Varianz beinhaltet Information darüber, wie stark die Zufallsvariable um ihren Erwartungswert streut.
2. Ist die oben genannte Menge M endlich, so existiert die Varianz immer.

Beispiele

1. Werfen eines Würfels, X Augenzahl ($\implies \mu = \frac{7}{2}$)

$$\begin{aligned}
\operatorname{Var}(X) &= \frac{1}{6} \cdot \left(\left(1-\frac{7}{2}\right)^2 + \left(2-\frac{7}{2}\right)^2 + \left(3-\frac{7}{2}\right)^2 + \left(4-\frac{7}{2}\right)^2 + \left(5-\frac{7}{2}\right)^2 + \left(6-\frac{7}{2}\right)^2 \right) \\
&= \frac{1}{6} \cdot \left(\left(-\frac{5}{2}\right)^2 + \left(-\frac{3}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{5}{2}\right)^2 \right) \\
&= \frac{35}{12} \approx 2.92
\end{aligned}$$

2. Zweimaliges Werfen eines Würfels, X Augensumme ($\implies \mu = 7$)

$$\operatorname{Var}(X) = \frac{35}{6}$$

Lemma 2.3
Gegeben sei eine diskrete Zufallsvariable X mit Erwartungswert μ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{P}\left(X \in \{x_1, x_2, x_3, \ldots\}\right) = 1$.
Dann gilt:

(a) $\operatorname{Var}(X)$ existiert genau dann, wenn $\sum\limits_{k=1}^{\infty} x_k^2 \mathrm{P}(X = x_k)$ konvergiert

(b) Wenn $\sum\limits_{k=1}^{\infty} x_k^2 \mathrm{P}(X = x_k)$ konvergiert, dann ist mit $\mu = \mathrm{E}(X)$:

$$\sum_{k=1}^{\infty} (x_k - a)^2 \mathrm{P}(X = x_k) = \operatorname{Var}(X) + (\mu - a)^2 \geq \operatorname{Var}(X)$$

(c) Wenn die Varianz von X existiert, dann gilt $\mu = \mathrm{E}(X)$:

$$\operatorname{Var}(X) = \sum_{k=1}^{\infty} x_k^2 \mathrm{P}(X = x_k) - \mu^2$$

Beispiel
Zweimaliges Werfen eines Würfels, Y Maximum der Augenzahlen (d.h. $\mu = \frac{161}{36}$)

$$\begin{aligned}
\operatorname{Var}(Y) &= \frac{1}{36} \cdot \left(1^2 \cdot 1 + 2^2 \cdot 3 + 3^2 \cdot 5 + 4^2 \cdot 7 + 5^2 \cdot 9 + 6^2 \cdot 11\right) - \left(\frac{161}{36}\right)^2 \\
&= \frac{2555}{1296} \approx 1.97
\end{aligned}$$

3 Spezielle Verteilungen diskreter Zufallsvariablen

Im Folgenden sei stets X eine diskrete Zufallsvariable auf einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $M \subseteq \mathbb{R}$ abzählbar mit $\mathrm{P}(X \in M) = 1$.

A Diskrete Gleichverteilung

Definition
X heißt **gleichverteilt** auf $\{1,2,\ldots,N\}$, wenn $M=\{1,2\ldots,N\}$ und $\mathrm{P}(X=k)=\frac{1}{N}$ für alle $k\in M$ ist.

Bemerkung
Eine auf $\{1,2,\ldots,N\}$ gleichverteilte Zufallsvariable X besitzt die Zähldichte

$$f(x)=\begin{cases}\frac{1}{N} & \text{für } x\in\{1,2,\ldots,N\}\\ 0 & \text{sonst}\end{cases}=\frac{1}{N}\mathbb{1}_{\{1,2,\ldots,N\}}(x)$$

Proposition 2.2
Für eine auf $\{1,2,\ldots,N\}$ gleichverteilte Zufallsvariable X gilt

$$\mathrm{E}(X)=\frac{N+1}{2}\quad\text{und}\quad\mathrm{Var}(X)=\frac{N^2-1}{12}$$

Beispiel
Augenzahl X beim Werfen eines Würfels

B Binomialverteilung

Definition
X heißt **binomialverteilt** mit den Paramtern $n\in\mathbb{N}$ und $p\in(0,1)$, wenn $M=\{0,\ldots,n\}$ und

$$\mathrm{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\qquad\text{für } k\in\mathbb{N}$$

Schreibweise: $X\sim B(n,p)$

Beispiel
120-maliges Werfen eines Würfels, X Anzahl der Einsen

$$\mathrm{P}(X=k)=\binom{120}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}\qquad(\implies\mathrm{E}(X)=20?)$$

Bemerkungen

1. Eine Zufallsvariable $X\sim B(n,p)$ besitzt die Zähldichte

$$f(x)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x} & \text{für } x\in\{0,1,\ldots,n\}\\ 0 & \text{sonst}\end{cases}=\binom{n}{x}p^x(1-p)^{n-x}\mathbb{1}_{\{0,1,\ldots,n\}}(x)$$

2. Gilt $X\sim B(1,p)$ mit einem $p\in(0,1)$, dann heißt X auch *Bernoulli-verteilt*. Für einen Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$ und ein $A\in\mathcal{F}$ gilt $\mathrm{P}(\mathbb{1}_A=1)=\mathrm{P}(A)$ und $\mathrm{P}(\mathbb{1}_A=0)=1-\mathrm{P}(A)$. Somit folgt $\mathbb{1}_A\sim B(1,p)$.

3. Beachte, dass nach dem Binomialsatz gilt:

$$\sum_{k=0}^{n}\mathrm{P}(X=k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=(p+1-p)^n=1$$

4. Die Binomialverteilung tritt stets in folgender Situation auf:

   Ein Zufallsexperiment wird n-mal durchgeführt. Die einzelnen Versuche sind hierbei unabhängig. In jedem Versuch wird festgestellt, ob ein bestimmtes Ereignis A eingetreten ist oder nicht. Die Zufallsvariable X gibt an, wie oft bei diesen n Versuchen das Ereignis A eingetreten ist. Wenn A bei einem einzelnen Versuch die Wahrscheinlichkeit p hat, dann gilt $X \sim B(n, p)$.

Proposition 2.3
Für eine Zufallsvariable $X \sim B(n, p)$ gilt

$$\mathrm{E}(X) = np \quad \text{und} \quad \mathrm{Var}(X) = np(1-p)$$

Bemerkung
Wir werden später (vgl. Satz 2.2 und Proposition 2.11) den allgemeinen Fall einfacher herleiten.

Beispiele

1. Ein Würfel werde n-mal geworfen. X sei die Anzahl der dabei geworfenen Einsen.

   Dann ist $X \sim B\left(n, \frac{1}{6}\right)$.

2. Ein Multiple-Choice-Test besteht aus 20 Fragen mit jeweils 4 vorgegebenen Antworten. Bei jeder Frage ist genau eine Antwort richtig.

   Die Anzahl der richigen Antworten bei zufälliger Auswahl ist dann $B(20, \frac{1}{4})$. Im Durchschnitt wird man in diesem Fall also 5 Fragen richtig beantworten.

C Geometrische Verteilung

Definition
Eine Zufallsvariable X heißt **geometrisch verteilt** mit Parameter $p \in (0, 1)$, wenn $M = \mathbb{N}$ und $P(X = k) = p(1-p)^{k-1}$ ist.

Bemerkungen

1. Eine geometrisch verteilte Zufallsvariable besitzt die Zähldichte.

$$f(x) = \begin{cases} p(1-p)^{k-1} & \text{für } x \in \mathbb{N} \\ 0 & \text{sonst} \end{cases}$$

2. Beachte, dass

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \cdot \frac{1}{1-(1-p)} = 1$$

3. Die geometrische Verteilung tritt in folgender Situation auf:

   Ein Zufallsexperiment wird so lange durchgeführt, bis ein bestimmtes Ereignis A eintritt. Die Zufallsvariable X gibt die Anzahl der dafür notwendigen Versuche an.

   Wenn die einzelnen Versuche voneinander unabhängig sind und A bei einem einzelnen Versuch die Wahrscheinlichkeit p besitzt, dann ist X geometrisch verteilt mit Parameter p.

Proposition 2.4
Eine Zufallsvariable X sei geometrisch verteilt mit Parameter $p \in (0, 1)$.
Dann ist $\mathrm{E}(X) = \dfrac{1}{p}$ und $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$.

Beispiele

1. Werfen eines Würfels so lange, bis die erste Eins auftritt. Insbesondere sind im Durchschnitt 6 Würfe nötig, bis eine Eins fällt.

2. Eine Serie von Überraschungseiern enthält eines von vier unterschiedlichen Spielzeugen. Die Spielzeuge seien dabei gleich häufig vorhanden.

   Wie viele Eier muss man im Durchschnitt kaufen, bis man alle vier Spielzeuge hat?

   Beim Kauf des ersten Eis erhält man auf jeden Fall ein neues Spielzeug, da man vorher ja noch keines hatte. Die Wartezeiten X_2, X_3, X_4 auf das zweite, dritte und vierte Spielzeug sind dann geometrisch verteilt, und zwar X_2 mit Parameter $\frac{3}{4}$, X_3 mit Parameter $\frac{1}{2}$ und X_4 mit Parameter $\frac{1}{4}$.

   Die durchschnittliche Zeit, bis man alle vier Spielzeuge hat, ist dann

$$1 + \mathrm{E}(X_2) + \mathrm{E}(X_3) + \mathrm{E}(X_4) = 1 + \frac{4}{3} + 2 + 4 = \frac{25}{3}$$

D Hypergeometrische Verteilung

Definition

X heißt **hypergeometrisch verteilt** mit Parametern $k, m, n \in \mathbb{N}$ $(k \leq m+n)$, wenn $M = \{0, \ldots, k\}$ und

$$\mathrm{P}(X = \nu) = \frac{\binom{m}{\nu}\binom{n}{k-\nu}}{\binom{m+n}{k}}$$

Bemerkungen

1. Eine hypergeometrisch verteilte Zufallsvariable besitzt die Zähldichte.

$$f(x) = \begin{cases} \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} & \text{für } x = 0, \ldots, k \\ 0 & \text{sonst} \end{cases}$$

2. Hypergeometrische verteilte Zufallsvariablen treten z.B. in folgender Situation auf:

   Aus einer Urne mit m weißen und n schwarzen Kugeln werden k Kugeln ohne Zurücklegen und ohne Beachten der Reihenfolge gezogen.

   Die Zufallsvariable X gebe die Anzahl der weißen Kugeln unter den k geozogenen Kugeln an. Dann ist X hypergeometrisch verteilt mit Parameter $k, m, n \in \mathbb{N}$.

Beispiel

Die Wahrscheinlichkeit für vier Richtige Zahlen beim Lotto '6 aus 49' beträgt

$$\frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} = \frac{15 \cdot 903}{13983816} \approx 0.001$$

Proposition 2.5

Die Zufallsvariable X sei hypergeometrisch verteilt mit Parameter $k, m, n \in \mathbb{N}$.
Dann gilt:

$$\mathrm{E}(X) = \frac{km}{m+n} \quad \text{und} \quad \mathrm{Var}(X) = \frac{km}{m+n} \cdot \frac{n}{m+n} \cdot \frac{m+n-k}{m+n-1}$$

Bemerkung

Für große Werte von m und n und einen im Vergleich dazu kleinen Wert von k, lässt sich die hypergeometrische Verteilung durch die $B(k,p)$-verteilung mit $p = \frac{m}{m+n}$ annähern.
Als Faustregel sollte $n + m \geq 60$ und $\frac{k}{m+n} < 0.1$ sein.

E Poisson-Verteilung

Definition
X heißt **Poisson-verteilt** mit Parameter $\lambda > 0$, wenn $M = \mathbb{N}_0$ ist und

$$\mathrm{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{für alle } k \in \mathbb{N}_0$$

Schreibweise $X \sim P(\lambda)$

Bemerkungen

1. Eine Zufallsvariable $X \sim P(\lambda)$ besitzt die Zähldichte
$$f(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{für } x \in \mathbb{N}_0 \\ 0 & \text{sonst} \end{cases}$$

2. Beachte, dass
$$\sum_{k=0}^{\infty} \mathrm{P}(X = k) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = 1$$

3. Die Poisson-Verteilung tritt in Anwendungen z.B. in folgenden Situationen auf:
   - Anzahl der Zerfälle eines radioaktiven Präparats in einer festen (kurzen) Zeitspanne.
   - Anzahl der an einem Mobilfunkmasten in einem festen Zeitraum eintreffenden Gespräche
   - Anzahl der in einer festen Zeitspanne von einer Versicherung zu regulierenden Schadensfälle
   - Anzahl der Bruchstellen bei 100 km Eisenbahngleisen usw.

Proposition 2.6
Für eine Zufallsvariable $X \sim P(\lambda)$ $(\lambda > 0)$ gilt:

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$$

Beispiele

1. Die Anzahl X der Druckfehler auf einer Zeitungsseite lasse sich als $P(0.2)$-verteilte Zufallsvariable darstellen. Dann gilt $\mathrm{P}(X = 0) \approx 0.8187$, $\mathrm{P}(X = 1) \approx 0.1637$ und $\mathrm{P}(X = 2) \approx 0.0.0164$.

2. Bei einer Telefonzentrale kommen im Schnitt 240 Anrufe pro Stunde an. Dann lässt sich die Zahl der Anrufe X pro Minute als $P(4)$-verteilte Zufallsvariable darstellen.

   Dann ist beispielsweise
$$\mathrm{P}(X \leq 2) = e^{-4} \cdot \left(\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!}\right) \approx 0.2381$$

Satz 2.1 (Poisson-Approximation der Binomialverteilung)
Es sei $(p_n)_{n=1}^{\infty}$ eine Folge in $(0, 1)$ mit $\lim\limits_{n\to\infty} np_n = \lambda \in (0, \infty)$.
Dann gilt für alle $k \in \mathbb{N}_0$:

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \to e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad (n \to \infty)$$

Bemerkung
Dieses Ergebnis lässt sich so interpretieren, dass für 'kleine Werte' von p die Binomialverteilung $B(n, p)$ durch die Poisson-Verteilung mit Parameter $\lambda = np$ angenähert werden kann. (*Gesetz der seltenen Ereignisse*)

4 Absolut stetige Zufallsvariablen

Definition
Eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ heißt **absolut stetig**, wenn eine Funktion $f\colon \mathbb{R} \to [0,\infty)$ existiert, so dass $\int_{-\infty}^{\infty} f(t)\,dt$ konvergiert und für alle $x \in \mathbb{R}$ gilt:

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

Die Funktion f heißt die **Dichte** von X.

Schreibweise
Falls in einer Fragestellung mehrere Zufallsvariablen auftreten, werden wir die Dichte von X typischerweise mit f_X bezeichnen.

Bemerkungen

1. Wenn F die Verteilungsfunktion von X ist, gilt also

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

   Insb. ist, wenn f stetig ist, F eine Stammfunktion von f.

2. Offensichtlich ist $\int_{-\infty}^{\infty} f(t)\,dt = 1$, da $\lim_{x\to\infty} F(x) = 1$.

Lemma 2.4
Die Zufallsvariable X auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ sei absolut stetig mit Verteilungsfunktion F.
Dann ist F stetig und für alle $x \in \mathbb{R}$ gilt $\mathrm{P}(X = x) = 0$.

Bemerkung
Offensichtlich ist also

$$\mathrm{P}(X \leq x) = \mathrm{P}(X < x) \quad \text{und} \quad \mathrm{P}(X > x) = \mathrm{P}(X \geq x) \qquad \text{für } x \in \mathbb{R}$$

und

$$\mathrm{P}(X \in (a,b]) = \mathrm{P}(X \in [a,b)) = \mathrm{P}(X \in (a,b)) = \mathrm{P}(X \in [a,b]) = \int_a^b f(t)\,dt \qquad \text{für } a < b$$

Beispiele

1. $f(x) = \frac{1}{2}e^{-|x|}$. Beachte dazu

$$\frac{1}{2}\int_{-\infty}^{\infty} e^{-|x|}\,dx = \int_0^{\infty} e^{-x}\,dx = 1$$

2. $f(x) = \begin{cases} x^{-2} & x \geq 1 \\ 0 & x < 1 \end{cases} = x^{-2}\mathbb{1}_{[1,\infty)}(x)$. Dann ist

$$\int_{-\infty}^{\infty} e^{-|x|} f(x)\,dx = \int_1^{\infty} \frac{1}{x^2}\,dx = 1$$

Definition
Gegeben sei eine absolut stetige Zufallsvariable X mit Dichte f auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Wenn $\int_{-\infty}^{\infty} |x| f(x)\, dx$ konvergiert, dann heißt

$$\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

der **Erwartungswert** von X.

Beispiele

1. $f(x) = \frac{1}{2} e^{-|x|}$. Dann ist

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} |x| e^{-|x|}\, dx = \int_{0}^{\infty} x e^{-x}\, dx = -(x+1)e^{-x}\Big|_0^{\infty} = 1$$

Außerdem ist $x f(x)$ ungerade, daher ist $\mathrm{E}(X) = 0$.

2. $f(x) = \begin{cases} x^{-2} & x \geq 1 \\ 0 & x < 1 \end{cases} = x^{-2} \mathbb{1}_{[1,\infty)}(x)$. Dann ist

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \int_{1}^{\infty} x \frac{1}{x^2}\, dx$$

Dieses Integral konvergiert nicht, also besitzt eine Zufallsvariable mit Dichte f keinen Erwartungswert.

Bemerkung
Ist f gerade (d.h. $f(-x) = f(x)$ für alle $x \in \mathbb{R}$), so ist der Erwartungswert (falls existent) stets Null.

Definition
Gegeben sei eine absolut stetige Zufallsvariable X mit Dichte f und Erwartungswert μ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Wenn

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

konvergiert, dann heißt der Wert dieses Integrals die **Varianz** von X und wird mit $\mathrm{Var}(X)$ bzw. σ^2 bezeichnet.
Die Wurzel aus der Varianz heißt **Standardabweichung** von X und wird mit σ bezeichnet.

Beispiel
$f(x) = \frac{1}{2} e^{-|x|}$. Dann ist

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_{0}^{\infty} x^2 e^{-x}\, dx = -x^2 e^{-x}\Big|_0^{\infty} + 2 \int_{0}^{\infty} x e^{-x}\, dx = 2$$

Das folgende Lemma ist sozusagen die Version von Lemma 2.3 für absolut stetige Zufallsvariablen:

Lemma 2.5
Gegeben sei eine absolut stetige Zufallsvariable X mit Dichte f auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann gilt:

(a) $\mathrm{Var}(X)$ existiert genau dann, wenn $\int_{-\infty}^{\infty} x^2 f(x)\,dx$ konvergiert.

(b) Wenn $\int_{-\infty}^{\infty} x^2 f(x)\,dx$ dann gilt für alle $a \in \mathbb{R}$ (mit der üblichen Bezeichnung $\mu = \mathrm{E}(X)$):

$$\int_{-\infty}^{\infty} (x-a)^2 f(x)\,dx = \mathrm{Var}(X) + (\mu - a)^2 \geq \mathrm{Var}(X)$$

(c) Wenn die Varianz von X existiert, dann gilt mit $\mu = \mathrm{E}(X)$:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

5 Spezielle Verteilungen absolut stetiger Zufallsvariablen

Im Folgenden sei X stets eine absolut stetige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit Dichte f.

A Stetige Gleichverteilung

Definition
Es seien $a, b \in \mathbb{R}$ mit $a < b$.
Eine Zufallsvariable X mit Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in (a,b) \\ 0 & \text{sonst} \end{cases} = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x)$$

heißt **gleichverteilt** auf (a, b).
Schreibweise $X \sim U(a, b)$.

Proposition 2.7
Für $X \sim U(a, b)$ mit $a, b \in \mathbb{R}$, $a < b$ gilt

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

Beispiele

1. Ein Zeiger wird gedreht, bis er zufällig an einem Punkt eines Ziffernblatts stehen bleibt. Der Winkel, an dem der Zeiger stehen bleibt, kann als eine auf dem Intervall $[0, 2\pi]$ gleichverteilte Zufallsvariable modelliert werden.

2. Die Ziehung einer Zufallszahl entspricht im Idealfall einer $U(0, 1)$-verteilten Zufallsvariable. Bei dieser Anwendung ist es eine wichtige statistische Fragestellung, in wie weit ein konkretes Verfahren dieser Idealvorstellung entspricht.

B Normalverteilung

Definition
Es seien $\mu \in \mathbb{R}$ und $\sigma^2 > 0$.
Eine Zufallsvariable X heißt **normalverteilt** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, wenn X die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \qquad (\textit{Gaußsche Glockenkurve})$$

besitzt.
Im Fall $\mu = 0$ und $\sigma^2 = 1$ heißt X **standardnormalverteilt**.
Schreibweise: $X \sim N(\mu, \sigma^2)$.

Bemerkungen

1. Die Normalverteilung kann häufig dazu verwendet werden, um Wahrscheinlichkeiten bei Zufallsvariablen, die einer anderen Verteilung folgen, näherungsweise zu bestimmen (siehe dazu auch den sog. *Zentralen Grenzwertsatz*).

   In dieser Möglichkeit liegt die große Bedeutung der Normalverteilung begründet.

2. Die Verteilungsfunktion einer standardnormalverteilten Zufallsvariablen wird häufig mit Φ bezeichnet, deren Dichte mit φ, d.h. man setzt

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{und} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt \qquad \text{für alle } t \in \mathbb{R}$$

Lemma 2.6
Es gilt

$$\int_{-\infty}^{\infty} e^{-t^2/2}\,dt = \sqrt{2\pi} \qquad \text{bzw. allgemeiner} \int_{-\infty}^{\infty} e^{-(t-\mu)^2/(2\sigma^2)}\,dt = \sqrt{2\pi\sigma^2}$$

Lemma 2.7
Ist $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$, so ist die Zufallsvariable $\dfrac{X-\mu}{\sigma}$ standardnormalverteilt.

Bemerkungen

1. Allgemeiner folgt für $X \sim N(\mu, \sigma^2)$ mit $a > 0$, $b \in \mathbb{R}$, dass $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

2. Das in der Definition der Normalverteilung auftretende Integral kann nicht ohne weiteres berechnet werden. Insbesondere ist es nicht möglich, eine Stammfunktion von φ in elementarer Weise anzugeben. Daher werden Berechnungen zur Normalverteilung mit Hilfe einer Tabelle bzw. durch Statistiksoftware vorgenommen. Aus Lemma 2.7 folgt nun, dass hier eine Tabelle für die Standardnormalverteilung genügt bzw. dass auch die Statistiksoftware die Berechnungen nur für die Standardnormalverteilung durchführen muss.

Beispiele

1. Die Punktzahl eines zufällig ausgewählten Teilnehmers eines Eignungstests lasse sich als eine Zufallsvariable $X \sim N(120, 400)$ darstellen.

   Dann ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Teilnehmer eine Punktzahl zwischen 90 und 120 besitzt, gegeben durch

$$\mathrm{P}(90 \le X \le 120) = \mathrm{P}\left(-1.5 \le \frac{X-100}{20} \le 1\right) = \Phi(1) - \Phi(-1.5) \approx 0.7745$$

2. Für jede Zufallsvariable $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$ gilt:

$$\begin{aligned}
\mathrm{P}(\mu - \sigma \le X \le \mu + \sigma) &= \Phi(1) - \Phi(-1) && \approx 0.8627 \\
\mathrm{P}(\mu - 2\sigma \le X \le \mu + 2\sigma) &= \Phi(2) - \Phi(-2) && \approx 0.9545 \\
\mathrm{P}(\mu - 3\sigma \le X \le \mu + 3\sigma) &= \Phi(3) - \Phi(-3) && \approx 0.9973
\end{aligned}$$

Proposition 2.8
Für $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$ gilt

$$\mathrm{E}(X) = \mu \quad \text{und} \quad \mathrm{Var}(X) = \sigma^2$$

Bemerkung
Speziell gilt

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\,dx = 0 \qquad \text{und} \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx = 1$$

C Exponentialverteilung

Definition
Es sei $\lambda > 0$.
Eine Zufallsvariable X heißt **exponentialverteilt** mit Parameter λ, wenn X die Dichte $f(x) = \lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$ besitzt.
Schreibweise $X \sim \operatorname{Exp}(\lambda)$

Bemerkung
Offensichtlich ist

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$$

Lemma 2.8
Für $X \sim \operatorname{Exp}(\lambda)$ mit $\lambda > 0$ gilt:

$$\mathrm{E}(X) = \frac{1}{\lambda} \quad \text{und} \quad \operatorname{Var}(X) = \frac{1}{\lambda^2}$$

Beispiel
Die Exponentialverteilung tritt häufig bei sogenannten Wartezeitproblemen auf:
Die mittlere Lebensdauer einer speziellen Sorte von Glühbirnen wird mit 2000 Stunden angegeben. Dann ist die Lebensdauer einer zufällig ausgewählten Birne dieser Sorte als Zufallsvariable $X \sim \operatorname{Exp}(1/2000)$ darstellbar. Somit gilt:

$$\begin{aligned} \mathrm{P}(X \leq 1000) &= 1 - \exp(-0.5) \approx 0.3935 \\ \mathrm{P}(X \geq 2000) &= \exp(-2) \approx 0.1353 \end{aligned}$$

D Cauchy-Verteilung

Definition
Es sei $\alpha \in \mathbb{R}$, $\beta > 0$.
Dann heißt X **Cauchy-verteilt** mit den Parametern α und β, wenn X die Dichte

$$f(x) = \frac{1}{\pi\beta}\frac{1}{1 + \left(\frac{x-\alpha}{\beta}\right)^2}$$

besitzt.

Bemerkungen

1. Sei etwa $\alpha = 0$ und $\beta = 1$. Dann ist

$$\int_0^T \frac{1}{\pi}\frac{x}{1+x^2}\,dx = \frac{1}{2\pi}\log(1+x^2)\Big|_0^T \to \infty \quad (T \to \infty)$$

   Die Cauchy-Verteilung mit Parametern 0 und 1 besitzt also keinen Erwartungswert. Das selbe gilt für jede andere Wahl von Parametern.

2. Weitere für die Anwendungen wichtige absolut stetige Verteilungen erfordern Vorkenntnisse, die wir uns erst im Lauf der Vorlesung erarbeiten werden. Diese Verteilungen werden daher später eingeführt.

6 Momente von Zufallsvariablen

Lemma 2.9
Es sei X stets eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $g\colon \mathbb{R} \to \mathbb{R}$ stetig. Dann ist $g(X)$ ebenfalls eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathrm{P})$ und es gilt:

(a) Wenn X diskret ist mit $\mathrm{P}(X \in \{x_1, x_2, \ldots\}) = 1$, dann gilt

$$\mathrm{E}(g(X)) = \sum_{k=1}^{\infty} g(x_k) \mathrm{P}(X = x_k)$$

falls diese Reihe absolut konvergiert.

(b) Wenn X absolut stetig ist mit Dichte f, dann gilt

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

falls dieses Integral absolut konvergiert.

Bemerkung
Im Fall $g(x) = x$ ergibt sich der Erwartungswert von X und im Fall $g(x) = (x - \mu)^2$ ($\mu = \mathrm{E}(X)$) die Varianz von X (jeweils im Falle der Existenz).

Definition
Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ (diskret oder absolut stetig).

(a) Falls existent, heißt $\mathrm{E}(X^r)$ für $r \in \mathbb{N}$ das **r-te Moment** von X.

Entsprechend heißt im Falle der Existenz $\mathrm{E}(|X|^r)$ für $r \in \mathbb{N}$ das **r-te absolute Moment** von X.

(b) Es sei $\mu = \mathrm{E}(X)$.

Dann heißt im Falle der Existenz $\mathrm{E}((X - \mu)^r)$ für $r \in \mathbb{N}$ das **r-te zentrierte Moment** von X.

Beispiel
$X \sim N(0,1) \implies \mathrm{E}(X^3) = 0$ und

$$\mathrm{E}(X^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} x^3 (-e^{-x^2/2}) \Big|_{-\infty}^{\infty} + 3 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx = 3$$

Definition
Gegeben sei eine diskrete oder absolut stetige Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Wenn ein $\varepsilon > 0$ existiert, so dass für alle $t \in (-\varepsilon, \varepsilon)$ der Erwartungswert

$$m(t) = \mathrm{E}\left(e^{tX}\right)$$

existiert, dann heißt die Funktion $m(t)$ auf $(-\varepsilon, \varepsilon)$ die **momenterzeugende Funktion** von X.

Beispiele

1. $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} m(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\Big(x^2 - 2x(\mu + t\sigma^2) + \mu^2\Big)\right\} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu-t\sigma^2)^2}{2\sigma^2}\right\} dx \cdot \exp\left\{-\frac{1}{2\sigma^2}\Big(\mu^2 - (\mu + t\sigma^2)^2\Big)\right\} \\ &= \exp\left\{t\mu + t^2\sigma^2/2\right\} \qquad \text{für alle } t \in \mathbb{R} \end{aligned}$$

2. $X \sim P(\lambda)$

$$m(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = \exp\{\lambda(e^t - 1)\} \qquad \text{für alle } t \in \mathbb{R}$$

3. $X \sim B(n, p)$

$$m(t) = \sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k (1-p)^{n-k} = (1 - p + e^t p)^n \qquad frallet \in \mathbb{R}$$

4. $X \sim \operatorname{Exp}(\lambda)$

$$m(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - t} \qquad \text{für alle } t \in \mathbb{R}$$

Definition
Für eine diskrete Zufallsvariable X mit $\mathrm{P}(X \in \mathbb{N}_0) = 1$ heißt

$$g(z) = \sum_{k=0}^{\infty} \mathrm{P}(X = k) \cdot z^k$$

die **erzeugende Funktion** von X.

Bemerkung
Die g definierende Potenzreihe besitzt einen Konvergenzradius ≥ 1.

Beispiele

1. $X \sim B(n, p)$

$$g(z) = \sum_{k=0}^{n} \binom{n}{k} z^k p^k (1-p)^{n-k} = (1 - p + zp)^n$$

2. $X \sim P(\lambda)$

$$g(z) = \sum_{k=0}^{\infty} z^k e^{-\lambda} \frac{\lambda^k}{k!} = \exp\{\lambda(z - 1)\}$$

Definition
Gegeben sei eine Zufallsvariable $X \sim F$ (d.h. X besitzt die Verteilungsfunktion F) und ein $q \in (0, 1)$. Dann heißt eine Zahl x_q ein **q-Quantil** von X (bzw. von F), wenn $\mathrm{P}(X \leq x_q) \geq q$ und $\mathrm{P}(X \geq x_q) \geq 1 - q$ ist.
Im Fall $q = \frac{1}{4}$ oder $q = \frac{3}{4}$ nennt man x_q auch ein **Quartil**, im Fall $q = \frac{1}{2}$ einen **Median**.

Beispiele

1. $X \sim U(0,1)$

   Dann ist $\mathrm{P}(X \leq x) = x$ und $\mathrm{P}(X \geq x) = 1 - x$ für alle $x \in (0,1)$. Daher ist $q \in (0,1)$ stets ein q-Quantil.

2. $X \sim \mathrm{Exp}(\lambda)$

   Dann ist $\mathrm{P}(X \leq x) = 1 - e^{-\lambda x}$ und $\mathrm{P}(X \geq x) = e^{-\lambda x}$ für alle $x > 0$. Somit ist

$$\mathrm{P}(X \leq x) \geq q \iff x \geq \frac{1}{\lambda} \log \frac{1}{1-q} \quad \text{und} \quad \mathrm{P}(X \geq x) \geq q \iff x \leq \frac{1}{\lambda} \log \frac{1}{1-q}$$

3. $X \sim P(\lambda)$

   Sei $n \in \mathbb{N}_0$ und q beliebig mit $\sum_{k=0}^{n-1} e1{-}\lambda\frac{\lambda^k}{k!} < q < \sum_{k=0}^{n} e1{-}\lambda\frac{\lambda^k}{k!}$ Dann ist $\mathrm{P}(X \leq n) \geq q$ und $\mathrm{P}(X \geq n) = 1 - \sum_{k=0}^{n-1} e1{-}\lambda\frac{\lambda^k}{k!} \geq 1 - q$.

   Also ist n ein q-Quantil für jedes solche q.

   Sei nun $q = \sum_{k=0}^{n-1} e1{-}\lambda\frac{\lambda^k}{k!}$ für ein $n \in \mathbb{N}_0$.

   Dann gilt für alle $x_q \in [n, n+1]$;

$$\mathrm{P}(X \leq x_q) \geq q \quad \text{und} \quad \mathrm{P}(X \geq x_q) = 1 - \sum_{k=0}^{n-1} e1{-}\lambda\frac{\lambda^k}{k!} \geq 1 - q$$

   Also ist jedes solche x_q ein q-Quantil.

Quantile treten häufig in Anwendungsaufgaben auf (siehe nachfolgendes Beispiel) und sie werden uns später beispielsweise auch bei statistischen Tests begegegnen.

Beispiel
Die Ergebnisse bei einem Eignungstest sind $N(120, 400)$-verteilt. Es sollen die 25% besten Teilnehmer berücksichtigt werden.
Welcher Punktzahl entspricht das?
Ist $X \sim N(120, 400)$, so ist $\dfrac{X-120}{20} \sim N(0,1)$. Gesucht ist ein Wert q mit $\Phi(q) = 0.75$, d.h. ein 75%-Quantil der Standardnormalverteilung. Wir erhalten $q = 0.6745$. Insgesamt ergibt sich

$$0.75 \approx \Phi(0.6745) = \mathrm{P}\left(\frac{X-120}{20} \leq 0.6745\right) = \mathrm{P}(X \leq 120 + 0.6745 \cdot 20) = \mathrm{P}(X \leq 133.49)$$

7 Zufallsvektoren

Definition
$X_1, \ldots, X_d$ seien Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann heißt die Funktion

$$X\colon \Omega \to \mathbb{R}^d, X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_d(\omega) \end{pmatrix}$$

ein d-**dimensionaler Zufallsvektor** auf $(\Omega, \mathcal{F}, \mathrm{P})$.
Die Funktion

$$F\colon \mathbb{R}^d \to [0,1], F(x_1, \ldots, x_d) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

heißt **(gemeinsame) Verteilungsfunktion** von $X_1, \ldots, X_d$.

Bemerkungen

1. Beachte, dass für $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$ gilt:

$$\{\omega \in \Omega \mid X_1(\omega) \leq x_1, \ldots, X_d(\omega) \leq x_d\} = \{X_1 \leq x_1\} \cap \ldots \cap \{X_d \leq x_d\}$$

2. Die Funktionen

$$F_{X_k}(x) = \mathrm{P}(X_k \leq x) = \lim_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_d)\to\infty} F(x_1,\ldots,x_{k-1},x,x_{k+1},\ldots,x_d)$$

heißen **Randverteilungsfunktionen** von X.

Definition
Ein Zufallsvektor $X = (X_1,\ldots,X_d)$ auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$ heißt **diskret**, wenn $X_1,\ldots,X_d$ diskrete Zufallsvariablen sind.
Für einen diskreten Zufallsvektor X heißt die Funktion

$$f\colon \mathbb{R}^d \to [0,1], f(x) = \mathrm{P}(X = x)$$

die **Wahrscheinlichkeitsfunktion** oder **Zähldichte** von X.

Beispiele

1. Zweimaliges Werfen eines Würfels

   X Augenzahl im ersten Wurf, Y Maximum der Augenzahlen

$\mathrm{P}(X=x,Y=y)$	1	2	3	4	5	6	$\mathrm{P}(X=x)$
1	1/36	1/36	1/36	1/36	1/36	1/36	6/36
2	0	2/36	1/36	1/36	1/36	1/36	6/36
3	0	0	3/36	1/36	1/36	1/36	6/36
4	0	0	0	4/36	1/36	1/36	6/36
5	0	0	0	0	6/36	1/36	6/36
6	0	0	0	0	0	6/36	6/36
$\mathrm{P}(Y=y)$	1/36	3/36	5/36	7/36	9/36	11/36	1

   Für alle nicht in der Tabelle aufgeführten Werte ist $\mathrm{P}(X=x,Y=y) = 0$.

2. Die Verteilung von $\binom{X}{Y}$ sei gegeben durch folgende Tabelle (mit einem beliebigen Wert $c \in \left(0,\frac{1}{2}\right)$):

$\mathrm{P}(X=x,Y=y)$	1	2	$\mathrm{P}(X=x)$
1	c	$\frac{1}{2}-c$	$\frac{1}{2}$
2	$\frac{1}{2}-c$	c	$\frac{1}{2}$
$\mathrm{P}(Y=y)$	$\frac{1}{2}$	$\frac{1}{2}$	1

   Unabhängig vom konkreten Wert gilt hier immer

$$\mathrm{P}(X=1) = \mathrm{P}(X=2) = \mathrm{P}(Y=1) = \mathrm{P}(Y=2) = \frac{1}{2}$$

   Die Verteilungen der einzelnen Zufallsvariablen, die einen Zufallsvektor bilden, legen also die Verteilung des Zufallsvektors noch nicht fest. Allgemein gilt, dass die Verteilungsfunktion eines Zufallsvektors durch die Randverteilungsfunktionen noch nicht festgelegt ist.

Definition
Ein Zufallsvektor $X = (X_1,\ldots,X_n)$ auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$ mit gemeinsamer Verteilungsfunktion F heißt **absolut stetig**, wenn eine Funktion $f\colon \mathbb{R}^n \to [0,\infty)$ existiert, so dass $\int_{\mathbb{R}^n} f(x)\,dx$ existiert und für alle $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ gilt:

$$F(x_1,\ldots,x_n) = \underset{(-\infty,x_n]\times\ldots\times(-\infty,x_1]}{\int\cdots\int} f(t_1,\ldots,t_n)\,dt_1\ldots dt_n$$

Diese Funktion f heißt die **(gemeinsame) Dichte** von X.

Bemerkungen

1. Offensichtlich gilt $\int\limits_{\mathbb{R}^n} f(x)\,dx = 1$.

2. Durch Angabe einer Dichte ist die Verteilung eines Zufallsvektors eindeutig festgelegt.

Beispiel
Definiere

$$f\colon \mathbb{R}^2 \to [0,\infty), f(x,y) = \begin{cases} c(x+y) & \text{für } 0 < x, y < 1 \\ 0 & \text{sonst} \end{cases}$$

f ist genau dann Dichte eines Zufallsvektors X, wenn gilt

$$\begin{aligned} 1 &= \int\limits_{\mathbb{R}^2} f(x,y)\,d(x,y) = \int_0^1 \int_0^1 c(x+y)\,dy\,dx = \int_0^1 c\left(xy + \tfrac{y^2}{2}\right)\Big|_0^1 dx = \int_0^1 c(x+1/2)\,dx \\ &= c\frac{x^2+x}{2}\Big|_0^1 = c \end{aligned}$$

Dann gilt z.B.

$$\begin{aligned} \mathrm{P}(X_1 \le X_2) &= \int\limits_{0<x\le y<1} (x+y)\,d(x,y) = \int_0^1 \int_0^y (x+y)\,dx\,dy = \int_0^1 c\left(xy + \tfrac{y^2}{2}\right)\Big|_0^y dx \\ &= \int_0^1 \tfrac{3}{2}y^2\,dy = \frac{1}{2}y^3\Big|_0^1 = \frac{1}{2} \end{aligned}$$

8 Bedingte Verteilungen und Unabhängigkeit von Zufallsvariablen

Definition
$\binom{X}{Y}$ sei ein diskreter Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit Zähldichte f. Gegeben sein ein $x \in \mathbb{R}$ mit $\mathrm{P}(X = x) > 0$.
Dann heißt die durch

$$f_{Y|X}(y \mid x) = \frac{\mathrm{P}(Y = y, X = x)}{\mathrm{P}(X = x)}$$

definierte Funktion $f_{Y|X}(\cdot \mid x)$ die **bedingte Zähldichte** von Y gegeben $X = x$.
Weiter heißt die durch

$$F_{Y|X}(y \mid x) = \frac{\mathrm{P}(Y \le y, X = x)}{\mathrm{P}(X = x)}$$

definierte Funktion $F_{Y|X}(\cdot \mid x)$ die **bedingte Verteilungsfunktion** von Y gegeben $X = x$.

Beispiel
Zweimaliges Werfen eines Würfels, X Anzahl der Einsen, Y Anzahl der Sechsen

$\mathrm{P}(Y=y, X=x)$	0	1	2	$\mathrm{P}(X=x)$
0	16/36	8/36	1/36	25/36
1	8/36	2/36	0	10/36
2	1/36	0	0	1/36
$\mathrm{P}(Y=y)$	25/36	10/36	1/36	1

Dann ist

$P(Y = y \mid X = x)$	0	1	2
0	16/25	8/10	1
1	8/25	2/10	0
2	1/25	0	0

Definition
$\binom{X}{Y}$ sei ein absolut stetiger Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit Dichte f. Weiter sei $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$ die Dichte von X. Gegeben sei ein $x \in \mathbb{R}$ mit $f_X(x) > 0$.
Dann heißt die durch

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

definierte Funktion $f_{Y|X}(\cdot \mid x)$ die **bedingte Dichte** von Y gegeben $X = x$.
Weiter heißt die durch

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} f_{Y|X}(u \mid x)\,du$$

definierte Funktion $F_{Y|X}(\cdot \mid x)$ die **bedingte Verteilungsfunktion** von Y gegeben $X = x$.

Beispiel
Definiere

$$f\colon \mathbb{R}^2 \to [0,\infty), f(x,y) = \begin{cases} x + y & \text{für } 0 < x, y < 1 \\ 0 & \text{sonst} \end{cases}$$

Dann ist f eine Dichte (siehe Abschnitt 7). Außerdem gilt

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} \int_0^1 (x+y)\,dy = x + \frac{1}{2} & \text{für } x \in (0,1) \\ 0 & \text{sonst} \end{cases}$$

Für $x \in (0,1)$ ist daher

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{x+y}{x+1/2} & \text{für } 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

$$\Longrightarrow F_{Y|X}(y \mid x) = \int_0^y \frac{x+u}{x+1/2}\,du = \frac{xy + y^2/2}{x+1/2} \qquad \text{für } 0 < y < 1$$

Definition
Gegeben sei ein Folge $(X_k)_{k=1}^{\infty}$ von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$, $n \in \mathbb{N}$.

(a) $X_1, \dots, X_n$ heißen **unabhängig**, wenn für alle $x_1, \dots, x_n \in \mathbb{R}$ gilt:

$$\mathrm{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathrm{P}(X_1 \le x_1) \cdot \ldots \cdot \mathrm{P}(X_n \le x_n)$$

(b) $X_1, X_2, X_3, \dots$ heißen **unabhängig**, wenn $X_1, \dots, X_n$ für alle $n \in \mathbb{N}$ unabhängig sind.

Proposition 2.9
$X = (X_1, \dots, X_n$ sei ein Zufallsvektor $(\Omega, \mathcal{F}, \mathrm{P})$.
Es gelte eine der folgenden Bedingungen:

(i) X ist diskret mit Zähldichte f.

(ii) X ist absolut stetig mit Dichte f.

Dann sind $X_1, \dots, X_n$ unabhängige Zufallsvariablen genau dann, wenn für alle $x_1, \dots, x_n$ gilt:

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

(wobei f_{X_k} die Zähldichte bzw. Dichte von X_k ist).

Beispiele

1. Zweimaliges Werfen eines Würfels
   X Augenzahl im ersten Wurf, Y Maximum der Augenzahlen

$\mathrm{P}(X = x, Y = y)$	1	2	3	4	5	6	$\mathrm{P}(X = x)$
1	1/36	1/36	1/36	1/36	1/36	1/36	6/36
2	0	2/36	1/36	1/36	1/36	1/36	6/36
3	0	0	3/36	1/36	1/36	1/36	6/36
4	0	0	0	4/36	1/36	1/36	6/36
5	0	0	0	0	6/36	1/36	6/36
6	0	0	0	0	0	6/36	6/36
$\mathrm{P}(Y = y)$	1/36	3/36	5/36	7/36	9/36	11/36	1

   Dann ist z.B. $\mathrm{P}(X = 2, Y = 1) = 0 \neq \mathrm{P}(X = 2)\mathrm{P}(Y = 1)$.

2. Es sei

$$f\colon \mathbb{R}^2 \to [0, \infty), f(x, y) = \begin{cases} x + y & \text{für } 0 < x, y < 1 \\ 0 & \text{sonst} \end{cases}$$

   Dann besitzt X_1 die Dichte $f_1\colon \mathbb{R} \to [0, \infty), f_1(x) = \left(x + \frac{1}{2}\right) \mathbb{1}_{(0,1)}(x)$ und X_2 die Dichte $f_2\colon \mathbb{R} \to [0, \infty), f_2(y) = \left(y + \frac{1}{2}\right) \mathbb{1}_{(0,1)}(y)$. Für $x, y \in (0, 1)$ ist daher

$$f_1(x) f_2(y) = \left(x + \frac{1}{2}\right) \cdot \left(y + \frac{1}{2}\right) = xy + \frac{x + y}{2} + \frac{1}{4}$$

   Somit existierten $x, y \in (0, 1)$ mit $f_1(x) f_2(y) \neq x + y = f(x, y)$.

   Also sind X_1 und X_2 nicht unabhängig.

Bemerkung
Wenn $X_1, \dots, X_n$ unabhängige Zufallsvariablen sind und $g_k\colon \mathbb{R} \to \mathbb{R}$ $(k = 1, \dots, n)$ stetig, dann sind auch $g_1(X_1), \dots, g_n(X_n)$ unabhängig.

9 Summen von Zufallsvariablen

Proposition 2.10
X, Y seien absolut stetige Zufallsvariablen mit gemeinsamer Dichte f.
Dann ist $X + Y$ ebenfalls absolut stetig mit Dichte

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f(t, x - t)\, dt \qquad \text{für alle } x \in \mathbb{R}$$

Lemma 2.10
X und Y seien unabhängige Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann gilt:

(a) Wenn X, Y absolut stetig sind mit Dichte f_X bzw. f_Y, dann ist $X + Y$ absolut stetig mit Dichte

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_1(t) f_2(x-t)\, dt \qquad \textit{(Faltung von } f_1 \textit{ und } f_2)$$

(b) Wenn $\mathrm{P}(X \in \mathbb{N}_0) = \mathrm{P}(Y \in \mathbb{N}_0) = 1$ ist, dann folgt $\mathrm{P}(X + Y \in \mathbb{N}_0) = 1$ und für alle $n \in \mathbb{N}_0$ gilt:

$$\mathrm{P}(X + Y = n) = \sum_{k=0}^{n} \mathrm{P}(X = k)\mathrm{P}(Y = n-k)$$

Definition
Eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ heißt **symmetrisch**, wenn $\mathrm{P}(X \leq x) = \mathrm{P}(X \geq -x)$ für alle $x \in \mathbb{R}$ ist, d.h. wenn $-X$ die selbe Verteilungsfunktion (und daher die selbe Verteilung) wie X besitzt.

Bemerkungen

1. Für $X \sim N(0,1)$ ist

$$\mathrm{P}(-X \leq x) = \mathrm{P}(X \geq -x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{\infty} e^{-t^2/2}\, dt \underset{v=-t}{=\!=\!=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-v^2/2}\, dv = \mathrm{P}(X \leq x)$$

Eine standardnormalverteilte Zufallsvariable X ist also symmetrisch. Allgemeiner gilt das für jede Zufallsvariable $X \sim N(0, \sigma^2)$ mit $\sigma^2 > 0$.

2. Es sei $X \sim N(\mu, \sigma^2)$ mit $a < 0$, $b \in \mathbb{R}$.

Nach der vorhergehenden Bemerkung gilt

$$\frac{X-\mu}{\sigma} \sim N(0,1) \implies -\frac{X-\mu}{\sigma} = \frac{\mu - X}{\sigma} \sim N(0,1)$$

In einer Bemerkung nach Lemma 2.7 hatten wir festgestellt, dass für $X \sim N(\mu, \sigma^2)$ mit $a > 0$, $b \in \mathbb{R}$ stets $aX + b \sim N(a\mu + b, a^2\sigma^2)$ gilt. Nun zeigt sich, dass diese Behauptung auch für $a < 0$ gilt, denn es folgt:

$$aX + b = (-a)\sigma \frac{\mu - X}{\sigma} + a\mu + b \sim N(a\mu + b, a^2\sigma^2)$$

Satz 2.2
X_1, X_2 seien unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann gilt:

(a) Aus $X_k \sim P(\lambda_k)$ mit $\lambda_k > 0$ $(k = 1, 2)$ folgt

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$$

(b) Aus $X_k \sim B(n_k, p)$ mit $n_k \in \mathbb{N}$, $p \in (0,1)$ $(k = 1, 2)$ folgt

$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

(c) Aus $X_k \sim N(\mu_k, \sigma_k^2)$ mit $\mu_k \in \mathbb{R}$, $\sigma_k^2 > 0$ $(k = 1, 2)$ folgt

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Bemerkung
Mittels vollständiger Induktion folgt für unabhängige Zufallsvariablen $X_1, \ldots, X_n$:

(a) Aus $X_k \sim P(\lambda_k)$ mit $\lambda_k > 0$ $(k = 1, \ldots, n)$ folgt $\sum_{k=1}^n X_k \sim P\left(\sum_{k=1}^n \lambda_k\right)$

(b) Aus $X_k \sim B(n_k, p)$ mit $n_k \in \mathbb{N}$, $p \in (0, 1)$ $(k = 1, \ldots, n)$ folgt $\sum_{k=1}^n X_k \sim B\left(\sum_{k=1}^n n_k, p\right)$.

(c) Aus $X_k \sim N(\mu_k, \sigma_k^2)$ mit $\mu_k \in \mathbb{R}$, $\sigma_k^2 > 0$ $(k = 1, \ldots, n)$ folgt $\sum_{k=1}^n X_k + \sim N\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right)$. Insbesondere folgt im Spezialfall $X_k \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$ $(k = 1, \ldots, n)$ dann $\sum_{k=1}^n X_k + \sim N\left(n\mu, n\sigma^2\right)$.

Lemma 2.11
$X = (X_1, \ldots, X_n)$ sei ein Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $g\colon \mathbb{R}^d \to \mathbb{R}$ stetig.
Dann gilt:

(a) Wenn X diskret ist mit $\mathrm{P}(X \in \{x_1, x_2, \ldots\}) = 1$, dann gilt

$$\mathrm{E}(g(X)) = \sum_{k=1}^{\infty} g(x_k)\mathrm{P}(X = x_k)$$

falls diese Reihe absolut konvergiert.

(b) Wenn X absolut stetig ist mit Dichte f, dann gilt

$$\mathrm{E}(g(X)) = \int_{\mathbb{R}^d} g(x) f(x)\, dx$$

falls dieses Integral absolut konvergiert.

Proposition 2.11
X_1, X_2 seien Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren Erwartungswert existiert. Außerdem sei $a_1, a_2 \in \mathbb{R}$.
Dann existiert auch der Erwartungswert von $a_1X_1 + a_2X_2$ und es gilt:

$$\mathrm{E}(a_1X_1 + a_2X_2) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2)$$

Bemerkung
Durch vollständige Induktion folgt für Zufallsvariablen $X_1, \ldots, X_n$ auf $(\Omega, \mathcal{F}, \mathrm{P})$, deren Erwartungswert existieren, und Konstanten $a_1, \ldots, a_n \in \mathbb{R}$, dass auch $\mathrm{E}\left(\sum_{k=1}^n a_k X_k\right)$ existiert und dass gilt:

$$\mathrm{E}\left(\sum_{k=1}^n a_k X_k\right) = \sum_{k=1}^n a_k \mathrm{E}\left(X_k\right)$$

Lemma 2.12
Gegeben seien Zufallsvariablen X, Y auf $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann gilt:

(a) Wenn $\mathrm{E}(X^2)$ und $\mathrm{E}(Y^2)$ existieren. dann existiert auch $\mathrm{E}(XY)$.

(b) Wenn $\mathrm{E}(XY)$ existiert und X und Y unabhängig sind, dann gilt

$$\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$$

Bemerkung
Entsprechend zeigt man für Zufallsvariablen $X_1, \ldots, X_n$ auf $(\Omega, \mathcal{F}, \mathrm{P})$:

(a) Wenn $\mathrm{E}(|X_k|^n)$ für $1 \leq k \leq n$ existiert, dann existiert auch $\mathrm{E}(X_1 \cdot \ldots \cdot X_n)$.

(b) Wenn zusätzlich $X_1, \ldots, X_n$ unabhängig sind, dann gilt

$$\mathrm{E}(X_1 \cdot \ldots \cdot X_n) = \mathrm{E}(X_1) \cdot \ldots \cdot \mathrm{E}(X_n)$$

Definition
X und Y seien Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X^2) < \infty$ und $\mathrm{E}(Y^2) < \infty$.

(a) Dann heißt

$$\mathrm{Cov}(X, Y) = \mathrm{E}\Big((X - \mathrm{E}(X)) \cdot (Y - \mathrm{E}(Y))\Big)$$

die **Kovarianz** zwischen X und Y.

(b) Wenn zusätzlich $\mathrm{Var}(X) > 0$ und $\mathrm{Var}(Y) > 0$ ist, dann heißt

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

der **Korrelationskoeffizient** von X und Y.

(c) X und Y heißen **unkorreliert**, wenn $\mathrm{Cov}(X, Y) = 0$ ist.

Bemerkung
Offensichtlich ist $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
Wenn $\mathrm{Var}(X) > 0$ ist, so folgt $\rho(X, X) = 1$.

Proposition 2.12
X und Y seien Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X^2) < \infty$ und $\mathrm{E}(Y^2) < \infty$, sowie $a, b, c, d \in \mathbb{R}$.
Dann gilt:

(a) $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X) \cdot \mathrm{E}(Y)$

(b) $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$
Speziell gilt $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

Beispiele

1. Ein Tetraeder werde zweimal geworfen, d.h. wir es treten in jedem Wurf die Ausgänge 1,2,3,4 mit der Wahrscheinlichkeit $\frac{1}{4}$ auf. Nun sei X die Augenzahl beim ersten Wurf und Y das Maximum der Augenzahlen

$\mathrm{P}(X=x, Y=y)$	1	2	3	4	$\mathrm{P}(X=x)$
1	1/16	1/16	1/16	1/16	1/4
2	0	2/16	1/16	1/16	1/4
3	0	0	3/16	1/16	1/4
4	0	0	0	4/16	1/4
$\mathrm{P}(Y=y)$	1/16	3/16	5/16	7/16	1

Dann gilt

$$\begin{aligned}
\mathrm{E}(X) &= \frac{1+2+3+4}{4} = \frac{5}{2} \\
\mathrm{E}(Y) &= \frac{1+6+15+28}{16} = \frac{50}{16} = \frac{25}{8} \\
\mathrm{E}(XY) &= \frac{1}{16}(1+2+3+4+8+6+8+27+12+64) = \frac{135}{16} \\
\mathrm{Cov}(X,Y) &= \frac{135}{16} - \frac{5}{2}\cdot\frac{25}{8} = \frac{5}{8}
\end{aligned}$$

Weiter ist

$$\begin{aligned}
\mathrm{E}(X^2) &= \frac{1+4+9+16}{4} = \frac{30}{4} \\
\mathrm{E}(Y^2) &= \frac{1+12+45+112}{16} = \frac{170}{16}
\end{aligned}$$

und daher $\mathrm{Var}(X) = \frac{5}{4}$ und $\mathrm{Var}(Y) = \frac{55}{64}$.

Somit ist $\rho(X,Y) = \frac{2}{\sqrt{11}}$.

2. Es sei

$$f\colon \mathbb{R}^2 \to [0,\infty), f(x,y) = \begin{cases} x+y & \text{für } 0 < x,y < 1 \\ 0 & \text{sonst} \end{cases}$$

Dann besitzt X die Dichte $f_X\colon \mathbb{R} \to [0,\infty), f_X(x) = \left(x+\frac{1}{2}\right)\mathbb{1}_{(0,1)}(x)$ und Y die Dichte $f_Y\colon \mathbb{R} \to [0,\infty), f_Y(y) = \left(y+\frac{1}{2}\right)\mathbb{1}_{(0,1)}(y)$.

Dann gilt:

$$\begin{aligned}
\mathrm{E}(X) = \mathrm{E}(Y) &= \int_0^1 x\cdot\left(x+\frac{1}{2}\right)dx = \left(\frac{x^3}{3}+\frac{x^2}{4}\right)\Big|_0^1 = \frac{7}{12} \\
\mathrm{E}(XY) &= \int_0^1\int_0^1 xy(x+y)\,dx\,dy = \int_0^1 \left(\frac{x^3y}{3}+\frac{x^2y12}{2}\right)\Big|_0^1 dy = \left(\frac{y^2}{6}+\frac{y^3}{6}\right)\Big|_0^1 = \frac{1}{3} \\
\mathrm{Cov}(X,Y) &= \frac{1}{3} - \frac{7}{12}\cdot\frac{7}{12} = -\frac{1}{144}
\end{aligned}$$

Schließlich ist

$$
\begin{aligned}
\mathrm{E}(X^2) = \mathrm{E}(Y^2) &= \int_0^1 x^2 \cdot \left(x + \frac{1}{2}\right)\, dx = \left(\frac{x^4}{4} + \frac{x^3}{6}\right)\Bigg|_0^1 = \frac{5}{12} \\
\mathrm{Var}(X) = \mathrm{Var}(Y) &= \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144} \\
\rho(X,Y) &= -\frac{1}{11}
\end{aligned}
$$

Proposition 2.13
X und Y seien Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X^2) < \infty$ und $\mathrm{E}(Y^2) < \infty$.
Dann gilt:

(a) $\left|\mathrm{E}(XY)\right| \leq \sqrt{\mathrm{E}(X^2) \cdot \mathrm{E}(Y^2)}$ und daher

$$\mathrm{Cov}(X,Y) \leq \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)} \qquad (\textit{Cauchy-Schwarz-Ungleichung})$$

(b) Wenn $\mathrm{Var}(X) > 0$ und $\mathrm{Var}(Y) > 0$ ist, dann folgt $-1 \leq \rho(X,Y) \leq 1$.

(c) Wenn X und Y unabhängig sind, dann sind X und Y auch unkorreliert.

Bemerkungen

1. Bei der Cauchy-Schwarz-Ungleichung gilt Gleichheit genau dann, wenn ein $t \in \mathbb{R}$ existiert mit $\mathrm{P}(tX = Y) = 1$.

2. Die Umkehrung der letzten Behauptung gilt i.A. nicht, d.h. unkorrelierte Zufallsvariablen müssen nicht notwendigerweise unabhängig sein (Gegenbeispiel siehe unten).

Beispiel
X mit $\mathrm{P}(X = -1) = \mathrm{P}(X = 0) = \mathrm{P}(X = 1) = \frac{1}{3}$ und $Y = X^2$. Dann ist $\mathrm{Cov}(X,Y) = 0$.

Satz 2.3
X, Y seien Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X^2) < \infty$, $\mathrm{E}(Y^2) < \infty$.
Dann gilt:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

Falls X und Y unkorreliert sind, folgt $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Bemerkungen

1. Allgemeiner gilt für Zufallsvariablen $X_1, \ldots, X_n$ mit $\mathrm{E}(X_k^2) < \infty$ für $1 \leq k \leq n$:

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \mathrm{Var}(X_j) + 2 \sum_{1 \leq j < k \leq n} \mathrm{Cov}(X_j, X_k)$$

Falls $X_1, \ldots, X_n$ unkorreliert sind, dann gilt $\mathrm{Var}(X_1 + \ldots + X_n) = \displaystyle\sum_{j=1}^{n} \mathrm{Var}(X_j)$.

2. Insbesondere gilt für unabhängige Zufallsvariablen $X_1, \dots, X_n$:

$$\operatorname{Var}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \operatorname{Var}(X_j)$$

3. Die selbe Rechnung zeigt für Zufallsvariablen $X_1, \dots, X_n, Y_1, \dots, Y_n$ auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X_k^2) < \infty$, $\mathrm{E}(Y_k^2) < \infty$ für $1 \le k \le n$ sowie $a_1, \dots, a_n, b_1, \dots, b_n$:

$$\operatorname{Cov}\left(\sum_{j=1}^{n} a_j X_j, \sum_{k=1}^{n} b_k Y_k\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j b_k \operatorname{Cov}(X_j, Y_k)$$

Ein wichtiger Spezielfall davon ist

$$\operatorname{Var}\left(\sum_{j=1}^{n} a_j X_j\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \operatorname{Cov}(X_j, X_k) = (a_1\ a_2\ \ldots a_n) \cdot \begin{pmatrix} \operatorname{Cov}(X_1, X_1) & \ldots & \operatorname{Cov}(X_1, X_n) \\ \vdots & & \vdots \\ \operatorname{Cov}(X_n, X_1) & \ldots & \operatorname{Cov}(X_n, X_n) \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Definition
$X = (X_1, \dots, X_n)$ sei ein Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$.

(a) Wenn $\mathrm{E}(X_k)$ existiert für alle $k \in \{1, \dots, n\}$, dann heißt der Vektor $\mathrm{E}(X) = \big(\mathrm{E}(X_1), \dots, \mathrm{E}(X_n)\big)$ der **Erwartungswertvektor** von X.

(b) Wenn $\mathrm{E}(X_k^2)$ existiert für alle $k \in \{1, \dots, n\}$, dann heißt die Matrix

$$\operatorname{Cov}(X) = \Big(\operatorname{Cov}(X_j, X_k)\Big)_{j=1}^{n}{}_{k=1}^{n} = \begin{pmatrix} \operatorname{Cov}(X_1, X_1) & \ldots & \operatorname{Cov}(X_1, X_n) \\ \vdots & & \vdots \\ \operatorname{Cov}(X_n, X_1) & \ldots & \operatorname{Cov}(X_n, X_n) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

die **Kovarianzmatrix** von X.

Zur Erinnerung:

Definition
Eine Matrix $C \in \mathbb{R}^{n \times n}$ heißt

- **positiv definit**, wenn $a^t \operatorname{Cov}(X) a > 0$ für alle $a \in \mathbb{R}^n \setminus \{0\}$ gilt.
- **positiv semidefinit**, wenn $a^t \operatorname{Cov}(X) a \ge 0$ für alle $a \in \mathbb{R}^n$ gilt.

Lemma 2.13
$X = (X_1, \dots, X_n)$ sei ein Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(X_k^2)$ existiert für alle $k \in \{1, \dots, n\}$. Dann ist die Kovarianzmatrix von X symmetrisch (d.h. für $j, k \in \{1, \dots, n\}$ gilt $\operatorname{Cov}(X_j, X_k) = \operatorname{Cov}(X_k, X_j)$) und positiv semidefinit.

Beispiel
Es sei $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ sowie $\rho \in [0, 1)$. Dann kann man zeigen, dass

$$f\colon \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \exp\left\{-\tfrac{1}{2(1-\rho^2)}\left(\tfrac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\tfrac{x_1-\mu_1}{\sigma_1}\tfrac{x_2-\mu_2}{\sigma_2} + \tfrac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}$$

eine Dichte ist und dass für einen Zufallsvektor $X = (X_1, X_2)$ mit Dichte f gilt:

$$\mathrm{E}(X) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \text{und} \qquad \mathrm{Cov}(X) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} =: C$$

Dann ist für $x \in \mathbb{R}^2$, etwa $x = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$:

$$f(x) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{\det(C)}} \exp\left\{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right\}$$

Definition
Es sei $\mu \in \mathbb{R}^n$, etwa $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_2 \end{pmatrix}$, und $C \in \mathbb{R}^{n\times n}$ eine positiv definite Matrix.
Dann heißt ein Zufallsvektor X mit der Dichte

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{1}{\sqrt{\det(C)}} \exp\left\{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right\}$$

normalverteilt mit Erwartungswertvektor μ und Kovarianzmatrix C.
Im Fall $\mu = 0$ und $C = I_n$ heißt X **standardnormalverteilt**.

Bemerkungen

1. Ein Zufallsvektor X mit obiger Dichte besitzt tatsächlich den Erwartungswertvektor μ und die Kovarianzmatrix C.
2. $X = (X_1, \ldots, X_n)$ ist standardnormalverteilt genau dann, wenn $X_1, \ldots, X_n$ unabhängig sind mit $X_k \sim N(0,1)$ $(k = 1, \ldots, n)$.
3. Speziell beschreibt die Dichte im obigen Beispiel die zweidimensionale Normalverteilung.

Kapitel 3

Grenzwertsätze

1 Gesetze der großen Zahlen

Satz 3.1 (Tschebyscheff-Ungleichung)
X sei eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, für die $\mathrm{E}(X^2)$ existiert. Dann gilt für alle $\varepsilon > 0$:

$$\mathrm{P}\Big(|X - \mathrm{E}(X)| > \varepsilon\Big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Bemerkung
Die Tschebyscheff-Ungleichung liefert häufig eine recht grobe Abschätzung für $\mathrm{P}\Big(|X - \mu| > \varepsilon\Big)$. Ihre Stärke liegt eher in ihrer Allgemeinheit.

Lemma 3.1
Für $X \sim N(0,1)$ und $\varepsilon > 0$ gilt:

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{\varepsilon} - \frac{1}{\varepsilon^3}\right) e^{-\varepsilon^2/2} \leq \mathrm{P}(X > \varepsilon) \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-\varepsilon^2/2}}{\varepsilon}$$

Bemerkung
Wegen der Symmetrie der $N(0,1)$-Verteilung gilt für $X \sim N(0,1)$ und $\varepsilon > 0$:

$$\mathrm{P}(|X| > \varepsilon) = 2\mathrm{P}(X > \varepsilon) \leq \frac{2}{\sqrt{2\pi}} \cdot \frac{e^{-\varepsilon^2/2}}{\varepsilon}$$

Beispiel

$$\begin{aligned} & X \sim N(0,1) \\ \text{Tschebyscheff:}\quad & \mathrm{P}(|X| > 2) \leq \frac{\mathrm{Var}(X)}{2^2} = \frac{1}{4} \\ \text{Lemma 3.1:}\quad & \mathrm{P}(|X| > 2) = 2\mathrm{P}(X > 2) \leq \frac{2}{\sqrt{2\pi}} \cdot \frac{e^{-2}}{2} \approx 0.054 \end{aligned}$$

Satz 3.2 (Schwaches Gesetz der großen Zahlen)
$(X_k)_{k=1}^{\infty}$ sei eine Folge unabhängiger Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$. Für alle $k \in \mathbb{N}$ gelte $\mu = \mathrm{E}(X_k)$ und $\sigma^2 = \mathrm{Var}(X_k)$ (d.h. alle X_k besitzen den selben Erwartungswert und die selbe Varianz).
Dann gilt für alle $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \varepsilon\right) = 0$$

Bemerkungen

1. Die Bedingung der Gleichheit aller Erwartungswerte und Varianzen aller X_k ist insbesondere dann erfüllt, wenn alle X_k die selbe Verteilung besitzen.
2. Die Voraussetzungen von Satz 3.2 können erheblich abgeschwächt werden, allerdings funktioniert dann die dargestellte einfache Beweistechnik nicht mehr.

Definition
$(Y_n)_{n=1}^{\infty}$ sei eine Folge von Zufallsvariablen, Z eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathrm{P})$. Wenn für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}(|Y_n - Z| > \varepsilon) = 0$$

ist, dann sagt man, dass $(Y_n)_{n=1}^{\infty}$ **stochastisch** bzw. **in Wahrscheinlichkeit** gegen Z konvergiert.
Schreibweise: $Y_n \xrightarrow{p} Z$ $(n \to \infty)$

Bemerkung
Von besonderer Bedeutung ist der Fall in dem $Z \equiv a \in \mathbb{R}$ (d.h. der Grenzwert eine Konstante) ist, wie z.B. in Satz 3.2.

Proposition 3.1
$(Y_n)_{n=1}^{\infty}$ sei eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $Y_n \sim B(n, p_n)$. Hierbei gelte $np_n \to 0$ $(n \to \infty)$.
Dann folgt $Y_n \xrightarrow{p} 0$ $(n \to \infty)$.

Satz 3.3 (Starkes Gesetz der großen Zahlen)
$(X_k)_{k=1}^{\infty}$ sei eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$. Die Zufallsvariablen seien unabhängig und besitzen alle die Verteilung F. Außerdem existiere der Erwartungswert, etwa $\mu = \mathrm{E}(X_1)$.
Dann gilt:

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \mu\right) = 1$$

Definition
$(Y_n)_{n=1}^{\infty}$ sei eine Folge von Zufallsvariablen, Z eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathrm{P})$. Wenn für jedes $\mathrm{P}(\lim_{n\to\infty} Y_n = Z) = 1$ ist, dann sagt man, dass $(Y_n)_{n=1}^{\infty}$ **fast sicher** gegen Z konvergiert.
Schreibweise: $Y_n \xrightarrow{\text{f.s.}} Z$ $(n \to \infty)$

2 Der zentrale Grenzwertsatz

Definition
$(X_n)_{n=1}^{\infty}$ sei eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann heißen $X_1, X_2, X_3, \ldots$ **unabhängig und identisch verteilt** bzw. abgekürzt **i.i.d.** (für *independent and identically distributed*), wenn $X_1, X_2, X_3, \ldots$ unabhängig sind und eine Verteilungsfunktion F existiert mit $X_k \sim F$ für alle $k \in \mathbb{N}$.

Satz 3.4 (Zentraler Grenzwertsatz)
$X_1, X_2, X_3, \ldots$ seien unabhängige und identisch verteilte Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit endlicher Varianz. Es gelte $\mu = \mathrm{E}(X_k)$ und $\sigma^2 = \mathrm{Var}(X_k)$ für alle $k \in \mathbb{N}$.
Dann gilt für alle $x \in \mathbb{R}$:

$$\mathrm{P}\left(\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \qquad (n \to \infty)$$

Bemerkungen

1. Für unabhängige Zufallsvariablen $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ gilt nach Lemma 2.7 und Satz 2.2:
$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

2. Für $a, b \in \mathbb{R}$, $a < b$ gilt dann
$$\mathrm{P}\left(a < \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \to \Phi(b) - \Phi(a) \qquad (n \to \infty)$$

3. Wichtiger Spezialfall:
$(S_n)_{n=1}^{\infty}$ sei eine Folge $B(n,p)$-verteilte Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$.
Dann gilt für alle $x \in \mathbb{R}$:
$$\mathrm{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) \to \Phi(x) \qquad (n \to \infty) \quad \text{(Satz von de Moivre-Laplace)}$$
Beachte, dass für eine Folge $(X_k)_{k=1}^{\infty}$ unabhängiger $B(1,p)$-verteilter Zufallsvariablen gilt $\sum_{k=1}^{n} X_k \sim B(n,p)$.
Satz 3.4 beinhaltet jedoch keine Information darüber, ab welchen Werten von n die Approximation von $\mathrm{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right)$ durch $\Phi(x)$ gute Ergebnisse liefert.
Hier hat sich die Faustregel $np(1-p) \geq 9$ durchgesetzt.

Beispiele

1. Ein Würfel werde 600-mal geworfen.
Mit welcher Wahrscheinlichkeit ist die Anzahl der dabei geworfenen Einsen zwischen 190 und 210 (einschließlich)?
Wir definieren unabhängige Zufallsvariablen $X_1, \ldots, X_{600}$ durch
$$X_k = \begin{cases} 1 & \text{falls im } k\text{-ten Wurf eine Eins fällt} \\ 0 & \text{falls im } k\text{-ten Wurf keine Eins fällt} \end{cases}$$
Dann ist $X_k \sim B(1, 1/6)$ und daher $S = \sum_{k=1}^{600} X_k \sim B(600, 1/6)$.

Dann folgt $\mathrm{E}(S) = 100$ und $\mathrm{Var}(S) = \frac{500}{6} = \frac{250}{3}$. Somit ergibt sich

$$\begin{aligned}
\mathrm{P}(90 \leq S \leq 110) &= \mathrm{P}\left(-\frac{10}{\sqrt{250/3}} \leq \frac{S-\mathrm{E}(S)}{\sqrt{\mathrm{Var}(S)}} \leq \frac{10}{\sqrt{250/3}}\right) \approx \Phi\left(\frac{10}{\sqrt{250/3}}\right) - \Phi\left(-\frac{10}{\sqrt{250/3}}\right) \\
&= 2\cdot\Phi\left(\frac{10}{\sqrt{250/3}}\right) - 1 = 2\Phi\left(\sqrt{\frac{6}{5}}\right) - 1 \approx 2\cdot 0.86433 - 1 = 0.72866
\end{aligned}$$

Die gesuchte Wahrscheinlichkeit beträgt also ca. 73%.

2. Bei einer Wahl gibt es zwei Kandidaten A und B.

   In der Gesamtbevölkerung sind 48% für Kandidat A und 52% für Kandidat B.

   Mit welcher Wahrscheinlichkeit geben bei einer Befragung von 1000 Personen mehr als die Hälfte der Befragten an, sie würden für A stimmen?

   Wir modelieren die Befragung durch unabhängige $B(1, 0.46)$-verteilte Zufallsvariablen $X_1, \ldots, X_{1000}$. Dann ist $S = \sum_{k=1}^{1000} X_k \sim B(1000, 0.48)$, (jedenfalls wenn wir davon ausgehen, dass mit Zurücklegen gezogen wird) und daher $\mathrm{E}(S) = 480$ und $\mathrm{Var}(S) = 1000\cdot 0.48\cdot 0.52 = 249.6 \geq 9$. Somit ergibt sich

$$\begin{aligned}
\mathrm{P}(S > 500) &= 1 - \mathrm{P}(S \leq 500) = 1 - \mathrm{P}\left(\frac{S-480}{\sqrt{249.6}} \leq \frac{20}{\sqrt{249.6}}\right) \approx 1 - \Phi\left(\frac{20}{\sqrt{249.6}}\right) \\
&\approx 1 - \Phi(1.27) \approx 1 - 0.89796 = 0.10204
\end{aligned}$$

Bemerkung
Bei binomialverteilten Zufallsvariablen benutzt man häufig die sogenannten *Stetigkeitskorrektur*, d.h. man rechnet für $S_n \in B(n,p)$ mit $np(1-p) \geq 9$ sowie $\alpha, \beta \in \mathbb{Z}$ mit $\alpha < \beta$:

$$\mathrm{P}(\alpha \leq S_n \leq \beta) = \mathrm{P}\left(\frac{\alpha - np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{\beta - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{\beta - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{\alpha - np - \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

Damit ergeben sich oft genauere Näherungswerte.

In den Beispielen von oben ist etwa

1. Mit Stetigkeitskorrektur: $\mathrm{P}(90 \leq S \leq 110) \approx 2\Phi(1.15) - 1 \approx 2\cdot 0.8749 - 1 \approx 0.7498$
   Mit exakter Verteilung: $\mathrm{P}(90 \leq S \leq 110) \approx 0.7501$

2. Mit Stetigkeitskorrektur: $\mathrm{P}(S > 500) = 1 - \mathrm{P}(S \leq 500) \approx 1 - \Phi\left(\frac{20.5}{\sqrt{249.6}}\right) = 1 - \Phi(1.30) \approx 1 - 0.9032 = 0.0968$
   Mit exakter Verteilung: $\mathrm{P}(S > 500) \approx 0.0973$

Definition
$(Y_n)_{n=1}^{\infty}$ sei eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathrm{P})$. Weiter sei Z eine Zufallsvariable mit Verteilungsfunktion F.
Wenn $\mathrm{P}(Y_n \leq x) \to F(x)$ $(n \to \infty)$ für jede Stelle x gilt, an der F stetig ist, dann sagt man, dass $(Y_n)_{n=1}^{\infty}$ **in Verteilung** gegen Z konvergiert.
Schreibweise: $Y_n \xrightarrow{d} Z$ $(n \to \infty)$

Bemerkung
Für Zufallsvariablen Z, Y_n $(n \in \mathbb{N})$ sowie $a \in \mathbb{R}$ gilt:

(a) Aus $Y_n \xrightarrow{p} Z$ $(n \to \infty)$ folgt stets $Y_n \xrightarrow{d} Z$ $(n \to \infty)$.

(b) Aus $Y_n \xrightarrow{d} a$ $(n \to \infty)$ folgt $Y_n \xrightarrow{p} a$ $(n \to \infty)$

Kapitel 4

Parameterschätzung

1 Stichproben

Definition
$X_1, \dots, X_n$ seien unabhängige, identisch verteilte Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Es gelte $X_k \sim F$ für $k = 1, \dots, n$.
Dann nennt man $X_1, \dots, X_n$ eine **einfache Zufallsstichprobe** vom Umfang n zur Verteilung F.
Werte $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ heißen eine **Realisierung der Zufallsstichprobe** bzw. eine **konkrete Stichprobe** vom Umfang n.
Die Menge aller potentiell möglichen Realisierungen einer Stichprobe nennt man den **Stichprobenraum**.

Bemerkung
Die Verteilungsfunktion F ist in Anwendungen typischerweise unbekannt. Mit Hilfe der Stichprobenwerte $x_1, \dots, x_n$ sollen über F Informationen gewonnen werden.
Häufig versucht man dabei zunächst, einige Kenngrößen wie z.B. Erwartungswert oder Varianz zu bestimmen.

Definition
Gegeben sei eine einfach Zufallsstichprobe $X_1, \dots, X_n$ sowie eine Funktion $T\colon \mathbb{R}^n \to \mathbb{R}$. Unter gewissen technischen Voraussetzungen an T nennt man die Funktion

$$T(X_1, \dots, X_n)\colon \Omega \to \mathbb{R}$$

eine **Statistik**.

Definition
Es sei $X_1, \dots, X_n$ eine einfach Zufallsstichprobe auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ zur Verteilung F.
Definiere

$$\bar{X} = \bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

und

$$S^2 = S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

Dann heißt $\bar{X}$ das **Stichprobenmittel** und S^2 die **Stichprobenvarianz**. $S = \sqrt{S^2}$ heißt **Stichprobenstandardabweichung**.

Bemerkung
Die konkreten, auf einer Realisierung der Stichprobe basierenden, Werte bezeichnet man mit

$$\bar{x} = \bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

und

$$s^2 = s_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$$

Proposition 4.1
Es sei $X_1, \ldots, X_n$ eine einfache Zufallsstichprobe auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ zur Verteilung F mit $\mathrm{E}(X_k) = \mu$ und $\mathrm{Var}(X_k) = \sigma^2$ für $k = 1, \ldots, n$. Insbesondere existieren also Erwartungswert und Varianz.
Dann gilt:

(a) $\mathrm{E}(\bar{X}) = \mu$

(b) $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$

(c) $\mathrm{E}(S^2) = \sigma^2$

Bemerkungen

1. Beachte, dass nach dem zentralen Grenzwertsatz gilt:

$$\sqrt{n}\frac{\bar{X} - \mu}{\sigma} = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \qquad (n \to \infty)$$

2. Es gilt $\mathrm{Var}(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$, wobei $\mu_4 = \mathrm{E}((X - \mu)^4)$.

3. Allgemein nennt man für $r \in \mathbb{N}$:

$$\frac{1}{n}\sum_{k=1}^{n} X_k^r \quad \text{das } r\text{-te Stichprobenmoment}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^r \quad \text{das } r\text{-te zentrale Stichprobenmoment}$$

$$\frac{1}{n}\sum_{k=1}^{n} |X_k|^r \quad \text{das } r\text{-te absolute Stichprobenmoment (sogar für } r \in (0,\infty))$$

Definition
Es sei $X_1, \ldots, X_n$ eine einfache Zufallsstichprobe auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ zur Verteilung F. Dann heißt die Funktion

$$\hat{F}_n \colon \mathbb{R} \times \Omega \to [0,1], \hat{F}_n(x, \omega) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{(-\infty,x]}\Big(X_k(\omega)\Big) = \frac{\text{Anzahl aller } k \text{ mit } X_k \leq x}{n}$$

die **empirische Verteilungsfunktion** der Zufallsstichprobe.

Proposition 4.2
$(X_k)_{k=1}^{\infty}$ sei eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $X_k \sim F$ $(k \in \mathbb{N})$. $\hat{F}_n$ sei für jedes $n \in \mathbb{N}$ die empirische Verteilungsfunktion von $X_1, \ldots, X_n$.
Dann gilt für alle $x \in \mathbb{R}$:

(a) $n\hat{F}_n(x) \sim B(n, F(x))$ und daher $\mathrm{E}(\hat{F}_n(x)) = F(x)$ und $\mathrm{Var}(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$

(b) $\hat{F}_n(x) \xrightarrow{p} F(x)$ für $n \to \infty$.

(c) $\sqrt{n}\frac{\hat{F}_n(x)-F(x)}{\sqrt{F(x)(1-F(x))}} \xrightarrow{d} N(0,1)$ $(n \to \infty)$, falls $F(x) \in (0,1)$.

Satz 4.1 (Satz von Glivenko-Cantelli)
$(X_k)_{k=1}^{\infty}$ sei eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $X_k \sim F$ $(k \in \mathbb{N})$. Für jedes $n \in \mathbb{N}$ sei $\hat{F}_n$ die empirische Verteilungsfunktion von $X_1, \ldots, X_n$ und $D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|$.
Dann ist D_n für alle $n \in \mathbb{N}$ eine Zufallsvariable und es gilt:

$$D_n \xrightarrow{p} 0(n \to \infty)$$

2 Schätzfunktionen

1 Parametrische Modelle

Definition
Es sei $X_1, \ldots, X_n$ eine einfach Zufallsstichprobe zur Verteilung $F \in \{F_\theta \mid \theta \in \Theta\}$ mit einer Parametermenge $\Theta \subseteq \mathbb{R}^p$. Weiter sei $g\colon \Theta \to \mathbb{R}^d$ eine Funktion.
Eine Statistik T (bzw. ein Vektor von Statistiken), deren Wert als Näherung für $g(\theta)$ verwendet wird, heißt eine **Schätzfunktion** bzw. ein **Schätzer** von $g(\theta)$.

Bemerkungen

1. Im Folgenden werden wir meistens reellwertige Schätzer betrachten. Am Beispiel der von den beiden Parametern μ und σ^2 abhängigen Normalverteilung sieht man aber unmittelbar, dass auch der vektorwertige Fall von Belang ist.

2. Die Definition eines Schätzers ist sehr allgemein. Wir benötigen daher noch Kriterien um die Güte eines Schätzers zu beurteilen.

Schreibweise
$(X_k)_{k=1}^{\infty}$ bzw. $X_1, \ldots, X_n$ seien unabhängige Zufallsvariablen mit $X_k \sim F \in \{F_\theta \mid \theta \in \Theta\}$ ($k \in \mathbb{N}$ bzw. $k = 1, \ldots, n$).
Dann schreiben wir $\mathrm{P}_\theta(A)$ oder $\mathrm{E}_\theta(Y)$, wenn die Wahrscheinlichkeit des Ereignisses A bzw. der Erwartungswert der Zufallsvariablen Y unter der Annahme $X_k \sim F_\theta$ berechnet wird, d.h. wir nehmen an, dass θ der 'wahre Parameter' ist.
Ähnliches gilt für die Notation $\mathrm{Var}_\theta(Y)$.

Definition
$(X_k)_{k=1}^{\infty}$ seien unabhängige Zufallsvariablen mit $X_k \sim F \in \{F_\theta \mid \theta \in \Theta\}$ $(k \in \mathbb{N})$. Weiter sei $T_n = T(X_1, \ldots, X_n)$ ein Schätzer für θ.

(a) T_n heißt **erwartungstreu** für θ, falls $\mathrm{E}_\theta(T_n) = \theta$ für alle $\theta \in \Theta$ ist.

(b) **asymptotisch erwartungstreu**, falls für alle $\theta \in \Theta$ gilt:

$$\mathrm{E}_\theta(T_n) \to \theta \qquad (n \to \infty)$$

(c) **schwach konstistent**, falls für alle $\theta \in \Theta$ gilt:

$$\mathrm{P}_\theta\Big(|T_n - \theta| > \varepsilon\Big) \to 0 \qquad (n \to \infty) \qquad \text{für alle } \varepsilon > 0$$

Beispiel
$(X_k)_{k=1}^\infty$ seien unabhängige Zufallsvariablen mit $X_k \sim F \in \{N(\mu, 1) \mid \mu \in \mathbb{R}\}$
Es sei $T_1 = \bar{X}$, $T_2 = S^2$, $T_3 = X_1$
Es gilt $\mathrm{E}_\theta(T_1) = \mathrm{E}_\theta(T_3) = \theta$, d.h. T_1 und T_3 sind erwartungstreu. T_2 ist nicht erwartungstreu, da $\mathrm{E}_\theta(T_2) = 1$ für alle $\theta \in \mathbb{R}$.
Obwohl T_1 und T_3 beide erwartungstreu sind, hat man den Eindruck, dass T_1 ein besserer Schätzer ist, etwa weil T_3 nur einen einzigen Stichprobenwert verwendet und die restliche Information nicht nutzt.
Beachte, dass T_1 (im Gegensatz zu T_3) schwach konsistent ist, denn es gilt:

$$\mathrm{P}_\mu\left(|T_1 - \mu| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}_\mu(T_1) = \frac{1}{\varepsilon^2} \cdot \frac{1}{n} \to 0 \qquad (n \to \infty)$$

3 Methoden zur Konstruktion von Schätzern

A Momentenmethode

Zur Erinnerung:
Für $r \in \mathbb{N}$ ist $m_r = \mathrm{E}(X^r)$ das r-te Moment (falls existent) und $\hat{m}_r = \frac{1}{n}\sum_{k=1}^n X_k^r$ das r-te Stichprobenmoment.
Falls $\mathrm{E}(X^r)$ existiert, dann gilt nach dem schwachen Gesetz der großen Zahlen:

$$\hat{m}_r \xrightarrow{p} m_r \quad (n \to \infty)$$

Es sei $\{F_\theta \mid \theta \in \Theta\}$ eine Menge von Verteilungsfunktionen, $\Theta \subseteq \mathbb{R}^d$. Dann lässt sich in manchen Fällen θ als eine Funktion der Momente ausdrücken, d.h. es existiert ein $l \in \mathbb{N}$ und eine Funktion $g\colon \mathbb{R}^l \to \mathbb{R}^d$, so dass für alle $\theta \in \Theta$ gilt: $\theta = g(m_1, \ldots, m_l)$. In diesem Fall erhält man einen Schätzer $\hat{\theta}$ für θ durch

$$\hat{\theta} = g(\hat{m}_1, \ldots, \hat{m}_l)$$

Beispiele

1. $\{P(\lambda) \mid \lambda > 0\}$
   Dann ist $\lambda = m_1$ (d.h. $g\colon \mathbb{R} \to \mathbb{R}, g(x) = x$) und daher $\hat{\lambda} = \hat{m}_1 = \bar{X}$.
2. $\{\mathrm{Exp}(\lambda) \mid \lambda > 0\}$
   Dann ist $m_1 = \frac{1}{\lambda}$ und daher $\hat{\lambda} = \frac{1}{\bar{X}}$.

Bemerkungen

1. Die Anwendung der Momentenmethode setzt natürlich voraus, dass entsprechendende Moment existieren. Das ist aber nicht immer der Fall, z.B. bei der Cauchy-Verteilung nicht.
2. Die Darstellung des Parameters als Funktion geeigneter Momente ist nicht immer eindeutig, z.B. gilt für die Poissonverteilungen $\{P(\lambda) \mid \lambda > 0\}$ auch $\mathrm{Var}(X) = \lambda$, d.h. $\lambda = m_2 - m_1^2$.

B Maximum–Likelihood–Methode

Definition

$X_1, \ldots, X_n$ sei eine einfache Zufallsstichprobe zur Verteilung $F \in \{F_\theta \mid \theta \in \Theta\}$ mit einer Parametermenge $\Theta \subseteq \mathbb{R}$. Es gelte eine der Bedingungen

- Für jedes $\theta \in \Theta$ ist F_θ diskret mit Zähldichte $f(\cdot, \theta)$.
- Für jedes $\theta \in \Theta$ ist F_θ absolut stetig mit Dichte $f(\cdot, \theta)$.

Dann heißt

$$L\colon \mathbb{R}^n \times \Theta \to \mathbb{R}, L(x_1, \ldots, x_n; \theta) = f(x_1, \theta) \cdot \ldots \cdot f(x_n, \theta)$$

die zugehörige **Likelihood-Funktion**.

Sei nun $\hat{\theta}\colon \mathbb{R}^n \to \Theta$ eine Funktion mit $L(x_1, \ldots, x_n; \theta) \leq L\Big(x_1, \ldots, x_n, \hat{\theta}(x_1, \ldots, x_n)\Big)$ für alle $x_1, \ldots, x_n \in \mathbb{R}$ und alle $\theta \in \Theta$. Dann heißt $\hat{\theta}(x_1, \ldots, x_n)\Big)$ ein **Maximum–Likelihood–Schätzer** für θ.

Bemerkungen

1. Weder existiert ein Maximum–Likelihood–Schätzer immer noch ist er im Falle der Existenz notwendigerweise eindeutig.

   Es zeigt sich aber, dass in vielen für die Anwendungen wichtigen Fällen ein Maximum–Likelihood–Schätzer einfach bestimmt werden kann.

2. Anstelle der Likelihood-Funktion L kann man auch die sog. Loglikelihoodfunktion $\log L$ betrachten, die manchmal rechnerisch einfacher gehandhabt werden kann. Beachte hierzu, dass für $x_1, \ldots, x_n \in \mathbb{R}$ und $\theta_1, \theta_2 \in \Theta$ gilt:

   $$L(x_1, \ldots, x_n; \theta_1) \leq L(x_1, \ldots, x_n; \theta_2) \iff \log L(x_1, \ldots, x_n; \theta_1) \leq \log L(x_1, \ldots, x_n; \theta_2)$$

Beispiele

1. $\{P(\lambda) \mid \lambda > 0\}$

   $$L(x_1, \ldots, x_n, \lambda) = \prod_{k=1}^{n} e^{-\lambda} \frac{\lambda^{x_k}}{x_k!} \mathbb{1}_{\mathbb{N}_0}(x_k) = \begin{cases} e^{-n\lambda} \frac{\lambda^{\sum_{k=1}^n x_k}}{x_1! \ldots x_n!} & \text{falls } x_1, \ldots, x_n \in \mathbb{N}_0 \\ 0 & \text{sonst} \end{cases}$$

   Somit ist

   $$\log L(x_1, \ldots, x_n, \lambda) = \sum_{k=1}^{n} x_k \log \lambda - n\lambda - \log(x_1! \cdot \ldots \cdot x_n!)$$

   für $x_1, \ldots, x_n \in \mathbb{N}_0$. Somit ist $\bar{x}$ der eindeutige Maximum-Likelihood-Schätzer.

2. $\{B(1, p) \mid p \in (0, 1)\}$

   Die Zähldichte ist dann

   $$f(x, p) = \begin{cases} p^x (1-p)^{1-x} & \text{für } x \in \{0, 1\} \\ 0 & \text{sonst} \end{cases}$$

   Somit ist die Likelihoodfunktion

   $$L(x_1, \ldots, x_n, p) = p^{x_1}(1-p)^{1-x_1} \cdot \ldots \cdot p^{x_n}(1-p)^{1-x_n} = p^{\sum_{k=1}^n x_k}(1-p)^{n - \sum_{k=1}^n x_k}$$

   für $x_1, \ldots, x_n \in \{0, 1\}$ und $L(x_1, \ldots, x_n, p) = 0$ sonst.

   Zur Bestimmung des Maximums genügt es also, den Fall $x_1, \ldots, x_n \in \{0, 1\}$ zu betrachten.

Für $\sum_{k=1}^n x_k = 0$ ist dann $L(x_1,\dots,x_n,p) = (1-p)^n \le L(x_1,\dots,x_n,0)$ und für $\sum_{k=1}^n x_k = n$ ist dann $L(x_1,\dots,x_n,p) = p^n \le L(x_1,\dots,x_n,1)$.

Für $0 < \sum_{k=1}^n x_k < n$ ist schließlich $L(x_1,\dots,x_n,0) = L(x_1,\dots,x_n,1) = 0$, daher besitzt $L(x_1,\dots,x_n,p)$ ein Maximum in $(0,1)$. Beachte hierbei, dass $x_1,\dots,x_n$ fest sind.

Also gilt dort:

$$
\begin{aligned}
&\frac{d}{dp}L(x_1,\dots,x_n,p) = 0\\
\iff &\sum_{k=1}^n x_k \cdot p^{\sum_{k=1}^n x_k - 1}(1-p)^{n-\sum_{k=1}^n x_k} - \left(n - \sum_{k=1}^n x_k\right) p^{\sum_{k=1}^n x_k - 1}(1-p)^{n-\sum_{k=1}^n x_k - 1} = 0\\
\iff &\sum_{k=1}^n x_k \cdot (1-p) - \left(n - \sum_{k=1}^n x_k\right) p = 0\\
\iff &p = \frac{1}{n}\sum_{k=1}^n x_k
\end{aligned}
$$

Alternativ lässt sich diese Rechnung auch mit der log-Likelhoodfunktion $\log L(x_1,\dots,x_n,p) = \sum_{k=1}^n x_k \log p + (n - \sum_{k=1}^n x_k)\log(1-p)$ durchführen. Man erhält dann

$$
0 = \frac{d}{dp}\log L(x_1,\dots,x_n,p) = \frac{1}{p}\sum_{k=1}^n x_k - \frac{1}{1-p}\left(n - \sum_{k=1}^n x_k\right)
$$

3. $\{N(\mu,\sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$

$$
\begin{aligned}
L(x_1,\dots,x_n,\mu,\sigma^2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\exp\left\{-\frac{\sum_{k=1}^n (x_k-\mu)^2}{2\sigma^2}\right\}\\
\implies \log L(x_1,\dots,x_n,\mu,\sigma^2) &= -\frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2 - \frac{n}{2}\log\sigma^2 - -\frac{n}{2}\log(2\pi)\\
\implies \frac{\partial}{\partial\mu}\log L(x_1,\dots,x_n,\mu,\sigma^2) &= -\frac{1}{\sigma^2}\sum_{k=1}^n (x_k-\mu) = -\frac{1}{\sigma^2}\left(\sum_{k=1}^n x_k - n\mu\right)\\
\text{und} \qquad \frac{\partial^2}{\partial\mu^2}\log L(x_1,\dots,x_n,\mu,\sigma^2) &= -\frac{n}{\sigma^2} < 0
\end{aligned}
$$

Also ist

$$
\frac{\partial}{\partial\mu}\log L(x_1,\dots,x_n,\mu,\sigma^2) = 0 \iff \mu = \frac{1}{n}\sum_{k=1}^n = \bar{x}
$$

und daher

$$
\log L(x_1,\dots,x_n,\bar{x},\sigma^2) \ge \log L(x_1,\dots,x_n,\mu,\sigma^2) \qquad \text{für alle } \mu \in \mathbb{R}, \sigma^2 > 0
$$

Wenn $x_1,\dots,x_n$ nicht alle gleich sind, dann ist $\sum_{k=1}^n (x_k-\bar{x})^2 > 0$ und daher gilt:

$$
\frac{\partial}{\partial\sigma^2}\log L(x_1,\dots,x_n,\bar{x},\sigma^2) = -\frac{1}{2\sigma^4}\sum_{k=1}^n (x_k-\bar{x})^2 - \frac{n}{2}\frac{1}{\sigma^2}
$$

und daher

$$
\frac{\partial}{\partial\sigma^2}\log L(x_1,\dots,x_n,\bar{x},\sigma^2) = 0 \iff \sigma^2 = \frac{1}{n}\sum_{k=1}^n (x_k-\bar{x})^2
$$

Außerdem gilt

$$\lim_{\sigma^2\to 0} L(x_1,\dots,x_n,\bar{x},\sigma^2) = \lim_{\sigma^2\to\infty} L(x_1,\dots,x_n,\bar{x},\sigma^2) = -\infty$$

$$\implies \log L\left(x_1,\dots,x_n,\bar{x},\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2\right) \geq L(x_1,\dots,x_n,\bar{x},\sigma^2) \quad \text{für alle } \sigma^2>0$$

Da $\mathrm{P}(X_1=\ldots=X_n)=0$, ist $\left(\bar{X},\frac{n-1}{n}S^2\right)$ der Maximum-Likelihood-Schätzer für (μ,σ^2).

4 Konfidenzintervalle

Definition
$X_1,\dots,X_n$ sei eine Zufallsstichprobe zur Verteilung $F\in\{F_\theta \mid \theta\in\Theta\}$ mit $\Theta\subseteq\mathbb{R}^d$. Weiter sei $g\colon\Theta\to\mathbb{R}$ eine Funktion sowie $\alpha\in(0,1)$.
Für zwei Statistiken T_1 und T_2 mit $T_1\leq T_2$ heißt das Intervall $[T_1,T_2]$ ein $1-\alpha$–**Konfidenzintervall für** $g(\theta)$, falls $\mathrm{P}_\theta(T_1\leq g(\theta)\leq T_2)\geq 1-\alpha$ für alle $\theta\in\Theta$.

Bemerkung
Ein Konfidenzintervall ist somit ein Intervall mit zufälligen Grenzen, das den gesuchten (und nicht zufälligen) Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt.

A Konfidenzintervall für den Erwartungswert bei bekannter Varianz unter Normalverteilung

Es sei $X_1,\dots,X_n$ eine Zufallsstichprobe zu $F\in\{N(\mu,\sigma^2)\mid\mu\in\mathbb{R}\}$. $\sigma^2>0$ sei hierbei eine feste und bekannte Zahl. Wir gehen vor wie im Einleitungsbeispiel.
Es ist $\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\sim N(0,1)$, falls $X_k\sim N(\mu,\sigma^2)$. Für $c>0$ ist dann $\mathrm{P}\left(-c\leq\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\leq c\right)=2\Phi(c)-1$ (eigentlich P_μ). Beachte, dass $2\Phi(c)-1=1-\alpha \iff \Phi(c)=1-\frac{\alpha}{2}$ und

$$-c\leq\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\leq c \iff \bar{X}-c\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{X}+c\frac{\sigma}{\sqrt{n}}$$

Für ein $1-\frac{\alpha}{2}$–Quantil $z_{1-\alpha/2}$ der Standardnormalverteilung ist also

$$\mathrm{P}\left(\bar{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Also ist durch $\left[\bar{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ ein $1-\alpha$–Konfidenzintervall für μ gegeben.

Beispiel
Eine Maschine schneidet Stifte, die eine bestimmte Länge μ haben sollen. Da hierbei Schwankungen entstehen, kann man die tatsächliche Länge als normalverteilt mit Erwartungswert μ mm und Standardabweichung $\sigma=2.4$ mm ansehen.
Bei einer Stichprobe von 9 Stück erhält man die Werte

$$184.2,\ 182.8,\ 185.3,\ 184.5,\ 186.2,\ 183.9,\ 185.0,\ 187.1,\ 184.4 \qquad (\implies \bar{x}=184.8)$$

99%–Konfidenzintervall: $\left[184.8-2.576\cdot\frac{2.4}{3}, 184.8+2.576\cdot\frac{2.4}{3}\right]=[182.74,186.86]$

Bemerkungen

1. Das eben konstruierte Konfidenzintervall hat die Eigenschaft, dass es symmetrisch um $\bar{X}$ ist. Das ist für die meisten Anwendungen so passend, allerdings lassen sich, falls nötig, auch Konfidenzintervalle konstruieren, bei denen dies nicht der Fall ist.

2. Die Annahme, dass die Varianz bekannt ist, ist bei vielen Anwendungen nicht vertretbar. Ist die Varianz unbekannt, treten eine Reihe weiterer Verteilungen auf, die wir im Folgenden kurz einführen wollen.

Proposition 4.3
$X_1, \ldots, X_r$ seien unabhängig und $N(0,1)$-verteilt.
Dann besitzt $X_1^2 + \ldots + X_r^2$ eine Verteilung mit der Dichte

$$f\colon \mathbb{R} \to [0,\infty), f(x) = \frac{1}{\Gamma(r/2)} \left(\frac{1}{2}\right)^{r/2} x^{r/2-1} e^{-x/2} \mathbb{1}_{(0,\infty)}(x)$$

Hierbei ist $\Gamma(y) = \int_0^\infty t^{y-1} e^{-t}\, dt$ für $y > 0$ (*Gammafunktion*).

Bemerkungen

1. Es gilt $\Gamma(n) = (n-1)!$ für alle $n \in \mathbb{N}$.
2. Offensichtlich ist

$$\begin{aligned} \mathrm{E}(X_1^2 + \ldots + X_r^2) &= r\mathrm{E}(X_1^2) = r \qquad \text{und} \\ \mathrm{Var}(X_1^2 + \ldots + X_r^2) &= r\mathrm{Var}(X_1^2) = r\Big(\mathrm{E}(X_1^4) - (\mathrm{E}(X_1^2))^2\Big) = 2r \end{aligned}$$

Definition
Es sei $r \in \mathbb{N}$. Dann heißt eine Verteilung mit der Dichte

$$f\colon \mathbb{R} \to [0,\infty), f(x) = \frac{1}{\Gamma(r/2)} \left(\frac{1}{2}\right)^{r/2} x^{r/2-1} e^{-x/2} \mathbb{1}_{(0,\infty)}(x)$$

eine χ^2**–Verteilung mit** r **Freiheitsgraden**.
Schreibweise: χ_r^2

Bemerkung
Die χ_2^2–Verteilung entspricht der Exp(1/2)–Verteilung.

Definition
Es sei $r, s \in \mathbb{N}$. Weiter seien X, Y, Z unabhängige Zufallsvariablen mit $X \sim N(0,1)$, $Y \sim \chi_r^2$ und $Z \sim \chi_s^2$.
Dann heißt die Verteilung von $\dfrac{X}{\sqrt{Y/r}}$ eine t**-Verteilung mit** r **Freiheitsgraden** (Schreibweise: t_r) und die Verteilung von $\dfrac{Y/r}{Z/s}$ eine F**-Verteilung mit** r **und** s **Freiheitsgraden** (Schreibweise: $F_{r,s}$).

Bemerkungen

1. t_r–verteilte Zufallsvariablen sind symmetrisch.
2. Die Dichte der t_r–Verteilung und der $F_{r,s}$–Verteilung lässt sich explizit angeben.

Satz 4.2
$X_1, \ldots, X_n$ seien unabhängig und $N(\mu, \sigma^2)$–verteilt.
Dann sind $\bar{X}$ und S^2 unabhängig und es gilt:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2 \quad \text{und} \quad \sqrt{n}\frac{\bar{X} - \mu}{S} \sim t_{n-1}$$

B Konfidenzintervall für den Erwartungswert bei unbekannter Varianz unter Normalverteilung

Es sei $X_1, \ldots, X_n$ eine Zufallsstichprobe zu $F \in \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R},\, \sigma^2 > 0\}$.
Es ist $\sqrt{n}\frac{\bar{X}-\mu}{S} \sim t_{n-1}$, falls $X_k \sim N(\mu, \sigma^2)$. Für ein $1-\frac{\alpha}{2}$–Quantil $z_{1-\alpha/2}$ der t_{n-1}–Verteilung ist also

$$1-\alpha = \mathrm{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\frac{\bar{X}-\mu}{S} \leq z_{1-\alpha/2}\right) = \mathrm{P}\left(\bar{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Also ist durch $\left[\bar{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}},\, \bar{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$ ein $1-\alpha$–Konfidenzintervall für μ bei unbekannter Varianz gegeben.
Die für dieses Konfidenzintervall benötigten Quantile der t-Verteilung können durch ein Statistikprogramm bestimmt bzw. aus der in der Vorlesung benutzten Tabelle (siehe Anhang) abgelesen werden.

Beispiel
Bei einer Maschine zum Abfüllen von Bierfässern wird eine Stichprobe vom Umfang 25 zur Bestimmung der tatsächlichen Füllmenge erhoben. Bei den 25 Messungen ergibt sich ein Stichprobenmittel von 49.4 l bei einer Stichprobenstandardabweichung von 0.3 l.
Dann ist $n = 25$. Ein 0.975-Quantil der t_{24}–Verteilung ist gegeben durch $z_{0.975} = 2.064$.
Ein 95%–Konfidenzintervall für die durchschnittliche Füllmenge hat also die Grenzen $49.4 \pm 2.064\frac{0.3}{5}$, wenn man annimmt, dass die Füllmenge normalverteilt ist.

C Asymptotisches Konfidenzintervall für p unter Binomialverteilung

Es sei $X_1, \ldots, X_n$ eine Zufallsstichprobe zu $F \in \{B(1,p) \mid p \in [0,1]\}$.
Dann ist $X_1 + \ldots + X_n \sim B(n,p)$. Nach der Tschebyscheff-Ungleichung gilt für $c > 0$:

$$\mathrm{P}(|\bar{X} - p| > c) = \mathrm{P}(|X_1 + \ldots + X_n - np| > nc) \leq \frac{\mathrm{Var}(X_1 + \ldots + X_n)}{n^2c^2} = \frac{np(1-p)}{n^2c^2} \leq \frac{1}{4nc^2}$$

und daher

$$\mathrm{P}(|\bar{X} - p| \leq c) \geq 1 - \frac{1}{4nc^2}$$

Wählt man nun $\beta \in (0,1)$ und wählt c gemäß

$$\beta = 1 - \frac{1}{4nc^2} \iff \frac{1}{4nc^2} = 1-\beta \iff 4nc^2 = \frac{1}{1-\beta} \iff c = \frac{1}{\sqrt{4n(1-\beta)}}$$

so folgt $\mathrm{P}(|\bar{X} - p| \leq c) \geq \beta$.

Beispiel
In einer Wahlurne befinden sich alle Stimmzettel zu einer Wahl zwischen zwei Kandidaten. Es werden 1000 Stimmzettel mit Zurücklegen gezogen. Unter diesen sind genau 450 Stimmen für A.
In diesem Fall ist $\frac{1}{\sqrt{4n(1-\beta)}} = \frac{1}{2\sqrt{1000\cdot 0.05}} = \frac{1}{10\sqrt{2}} \approx 0.071$. Durch $[0.45 - 0.071, 0.45 + 0.071]$ ist also ein 0.95–Konfidenzintervall für p gegeben.

Angesichts der Tatsache, dass die Tschebyscheff-Ungleichung häufig sehr ungenaue Schätzungen liefert, ist zu erwarten, dass es wesentlich kürzere Konfidenzintervalle für p gibt.

Mit Hilfe des zentralen Grenzwertsatzes erhält man das approximative β-Konfidenzintervall

$$\left[\bar{X} - c\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + c\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

wobei c ein $\frac{1+\beta}{2}$-Quantil der Standardnormalverteilung sei.
Faustregel: Diese Näherungsformel sollte nur angewandt werden, wenn die konkrete Stichprobe die Bedingung $50 \leq n\bar{X} \leq n - 50$ erfüllt.

Beispiel
Im obigen Beispiel ist $n = 1000$, $\bar{X}$, $c = 1.96$ und daher $c\frac{\bar{X}(1-\bar{X})}{n} \approx 0.0308$

Kapitel 5

Statistische Tests

1 Einführung

Definition
Es sei $X_1, \dots, X_n$ eine Zufallsstichprobe zu $F \in \{F_\theta \mid \theta \in \Theta\}$.
Unter einer **Hypothese** verstehen wir eine Annahme über θ. Konkret betrachten wir eine Teilmenge $\emptyset \neq \Theta_0 \subseteq \Theta$ und sagen, die Hypothese treffe zu, wenn $\theta \in \Theta_0$ liegt.
Eine Hypothese heißt **einfach**, wenn Θ_0 aus genau einem Element besteht (d.h. die Hypothese legt die Verteilung vollständig fest), sonst heißt die Hypothese **zusammengesetzt**.
Bei einem **Testproblem** teilt man Θ in zwei nichtleere disjunkte Teilmengen Θ_0 und Θ_1 auf. Die Annahme, dass $\theta \in \Theta_0$ sei, heißt dann die **Nullhypothese** H_0 und die Annahme $\theta \in \Theta_1$ die **Alternativhypothese** H_1.

Beispiele

1. $\{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}\}$, $\sigma^2 > 0$ bekannt.
   $H_0\colon \mu = 0$ gegen $H_1\colon \mu \neq 0$ (d.h. $\Theta_0 = \{0\}$ und $\Theta_1 = \mathbb{R} \setminus \{0\}$)
2. $\{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$
   $H_0\colon \mu = 0$ gegen $H_1\colon \mu \neq 0$ (d.h. $\Theta_0 = \{0\} \times (0, \infty)$ und $\Theta_1 = \mathbb{R} \setminus \{0\} \times (0, \infty)$)
3. $\{B(1, p =\mid p \in [0, 1]\}$
   $H_0\colon p \leq \frac{1}{4}$ gegen $H_1\colon p > \frac{1}{4}$

Definition
Ein (deterministischer) **Test** für das Testproblem $H_0\colon \theta \in \Theta_0$ gegen $H_1\colon \theta \in \Theta_1$ ist eine Entscheidungsregel, die für jede mögliche konkrete Stichprobe $(x_1, \dots, x_n)$ festlegt, ob H_0 oder H_1 gewählt wird.
Hierzu wird die Menge S aller möglichen Stichproben (der sog. *Stichprobenraum*) in zwei nichtleere disjunkte Mengen K_0 und K_1 zerlegt.
Wenn $(x_1, \dots, x_n) \in K_0$ ist, dann wird die Nullhypothese H_0 akzeptiert.
Wenn $(x_1, \dots, x_n) \in K_0$ ist, dann wird die Nullhypothese H_0 abgelehnt und daher für H_1 entschieden.
K_0 heißt der **Annahmebereich** und K_1 der **kritische Bereich** des Tests.

Beispiel
$\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$
$H_0\colon \mu \leq 0$ gegen $H_1\colon \mu > 0$

Lehne ab, falls $\frac{x_1+...+x_n}{n} \geq c$, d.h.

$$K_0 = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \frac{x_1 + \dots + x_n}{n} < c\right\}$$
$$K_1 = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \frac{x_1 + \dots + x_n}{n} \geq c\right\}$$

Bei einem Testproblem können folgende Situationen auftreten:

tatsächliche Situation / Entscheidung	H_0 ist richtig	H_0 ist falsch
H_0 wird akzeptiert	richtige Entscheidung	Fehler 2. Art
H_0 wird abgelehnt	Fehler 1. Art	richtige Entscheidung

Leider ist es im Allgemeinen nicht möglich, die Wahrscheinlichkeiten für beide Fehlerraten gleichzeitig zu minimieren. Im obigen Beispiel etwa wird die Wahrscheinlichkeit für einen Fehler 1. Art mit wachsendem c kleiner, während die Wahrscheinlichkeit für einen Fehler 2. Art steigt.

Definition
Gegeben sei ein Testproblem $H_0 \colon \theta \in \Theta_0$ gegen $H_1 \colon \theta \in \Theta_1$ mit dem Annahmebereich K_0 und dem kritischen Bereich K_1.
Dann heißt die Funktion

$$g \colon \Theta \to [0,1], g(\theta) = \mathrm{P}_\theta\Big((X_1, \dots, X_n) \in K_1\Big)$$

die **Gütefunktion** des Tests.
Wenn ein $\alpha \in (0,1)$ existiert mit $g(\theta) \leq \alpha$ für alle $\theta \in \Theta_0$, dann heißt der Test ein **Test zum Signifikanzniveau** α.
Ein Test zum Signifikanzniveau α heißt **unverfälscht**, wenn $g(\theta) \geq \alpha$ für alle $\theta \in \Theta_1$ ist.

Bemerkung
Bei einem Test zum Signifikanzniveau α ist die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen $\leq \alpha$. Es ist üblich, für α kleine Werte anzusetzen ($\alpha \leq 0.1$, oft $\alpha = 0.05$). Wenn H_0 richtig ist, wird also ein Test zum Signifikanzniveau α nur mit einer Wahrscheinlichkeit von höchstens α zur Ablehnung von H_0 führen. Das bedeutet im Umkehrschluss, dass dann, wenn ein solcher Test zur Ablehnung der Nullhypothese führt, auch tatsächlich Zweifel an deren Gültigkeit berechtigt sind.
Sprechweise: 'H_0 wird zum Signifikanzniveau α abgelehnt.'
Allerdings ist es nicht zulässig, von einer Bestätigung von H_0 zu sprechen, sollte der Test nicht zu einer Ablehnung der Nullhypothese führen. Die Wahrscheinlichkeit, H_0 fälschlicherweise nicht abzulehnen (also einen Fehler 2. Art zu begehen), wird nicht kontrolliert und könnte unter Umständen sehr groß sein. Für entsprechende Beispiele verweisen wir auf Abschnitt 2.
Ein Test zum Signifikanzniveau α dient also eher zur statistischen Bestätigung von H_1. Aus diesem Grund wird in der Regel diejenige Hypothese, die statistisch bestätigt werden soll, als Alternativhypothese gewählt.

2 Spezielle Tests

A Test für den Erwartungswert bei bekannter Varianz unter Normalverteilungsannahme

Es sei $X_1, \dots, X_n$ eine Zufallsstichprobe zu $F \in \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}\}$. $\sigma^2 > 0$ sei hierbei eine feste und bekannte Zahl. Weiter sei ein Signifikanzniveau $\alpha \in (0,1)$ vorgegeben.

A1 Zweiseitiger Test

Testproblem $H_0\colon \mu = \mu_0$ gegen $H_1\colon \mu \neq \mu_0$
Betrachte die Teststatistik

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

Für ein $1 - \alpha/2$-Quantil c von $N(0,1)$ gilt:

$$\mathrm{P}_{\mu_0}(-c \leq T \leq c) = 1 - \alpha \qquad \text{bzw.} \qquad \mathrm{P}_{\mu_0}(|T| > c) = \alpha$$

Wir lehnen H_0 also ab, wenn $|T| > c$ ist.

Beispiel
Eine Maschine füllt 200-g-Packungen mit Trockenfrüchten ab. Aus Erfahrung weiß man, dass die tatsächlichen Füllungen $N(\mu, 4)$-verteilt sind. Bei einer Stichprobe von 10 Packungen ergibt sich ein durchschnittliches Gewicht von 197 g.
Lässt dies zu einem Signifikanzniveau von 5% darauf schließen, dass $\mu \neq 200$ ist?
Es gilt $T = \frac{197-200}{2}\sqrt{10} = -\frac{3}{2}\sqrt{10} \approx -4.74$. Somit wird H_0 abgelehnt.

A2 Einseitiger Test

Testproblem $H_0\colon \mu \leq \mu_0$ gegen $H_1\colon \mu > \mu_0$
Betrachte wieder die Teststatistik

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

Wir lehnen H_0 ab, wenn $T > c$ ist.

Beispiel
Briefumschläge für Luftpost sollen im Schnitt höchstens 2g wiegen. Bei einer Stichprobe von 20 Umschlägen ergibt sich ein durchschnittliches Gewicht von 2.01g. Erfahrungsgemäß ist das Gewicht eines zufällig ausgewählten Umschlags dieser Sorte $N(\mu, 0.03^2)$-verteilt.
Lässt dies zu einem Signifikanzniveau von 1% darauf schließen, dass das Durchschnittsgewicht der Umschläge im Schnitt höher als 2g ist?
Für $\alpha = 0.01$ ist $c = 2.326$ und es gilt $T = \frac{2.01-2}{0.03}\sqrt{20} \approx 1.49$.
Auf der Basis der oben beschriebenen Stichprobe kann H_0 also nicht verworfen werden.

Bemerkungen

1. Ein Test für das Testproblem $H_0\colon \mu \geq \mu_0$ gegen $H_1\colon \mu < \mu_0$ kann entsprechend konstruiert werden, da für ein α-Quantil c von $N(0,1)$ und alle $\mu \geq \mu_0$ gilt: $\mathrm{P}_\mu(T < c) \leq \alpha$.

2. In den bisher betrachteten Fällen lässt sich die Gütefunktion leicht bestimmen.

   Für den einseitigen Test gilt:

   $$g(\mu) = \mathrm{P}_\mu(T > c) = \mathrm{P}\left(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} > c + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) = 1 - \Phi\left(c + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right)$$

   Somit ist g streng monoton wachsend mit $g(\mu_0) = \alpha$. Insbesondere ist der Test unverfälscht.

   Für den zweiseitigen Test ergibt eine ähnliche Rechnung:

   $$g(\mu) = 1 - \Phi\left(c + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) + \Phi\left(-c + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right)$$

B Test für den Erwartungswert bei unbekannter Varianz unter Normalverteilungsannahme

Es sei $X_1, \ldots, X_n$ eine Zufallsstichprobe zu $F \in \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R},\, \sigma^2 > 0\}$. Weiter sei ein Signifikanzniveau $\alpha \in (0,1)$ vorgegeben.

B1 Zweiseitiger Test

Testproblem $H_0\colon \mu = \mu_0$ gegen $H_1\colon \mu \neq \mu_0$
Betrachte die Teststatistik

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{S}$$

Dann ergibt sich folgender Test zum Signifikanzniveau α:
Lehne H_0 ab, falls $|T| > c$ ist.

Beispiel
In einem Geschäft kaufen Kunden im Schnitt für 70 € ein. Der Betrag, den ein zufällig ausgewählter Kunde ausgibt, kann als normalverteilte Zufallsvariable angesehen werden. Nachdem die Anordnung der Regale in diesem geschäft geändert wurde, stellt man bei einer Stichprobe von 25 Kunden einen durchschnittlichen Umsatz von 65 € bei einer Stichprobenstandardabweichung von 4.5 € fest.
Lässt sich daraus zu einem Signifikanzniveau von 5% schließen, dass sich der durchschnittliche Umsatz pro Kunde verändert hat?
Für $\alpha = 0.05$ und $n = 25$ ergibt sich $c = 2.064$ und es gilt $\left|\sqrt{n}\frac{\bar{x}-\mu_0}{s}\right| = \left|5\frac{65-70}{4.5}\right| = \frac{10}{9} \cdot 5 > 2.064$.
Die Nullhypothese wird also abgelehnt.

B2 Einseitiger Test

Testproblem $H_0\colon \mu \leq \mu_0$ gegen $H_1\colon \mu > \mu_0$
Betrachte wieder die Teststatistik

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{S}$$

Mit einem $1-\alpha$-Quantil c der t_{n-1}-Verteilung ergibt sich ein Test zum Signifikanzniveau α für dieses Testproblem durch die Entscheidungsregel:
Lehne H_0 ab, falls $T > c$ ist.

Beispiel
Die Hörer eines Radiosenders verfolgen dessen Programm im Schnitt 3h wöchentlich. Der Sender nimmt einige Programmänderungen vor, um dies Zeit zu verlängern.
Eine Umfrage unter 30 Hörern ergibt eine durchschnittliche Hördauer von 183 min bei einer Stichprobenstandardabweichung von 20 min.
Lässt dies unter Normalverteilungsannahme zu einem Signifikanzniveau von 5% auf eine verlängerte Hördauer schließen?
Für $\alpha = 0.05$ und $n = 30$ ist $c = 1.699$ und $\sqrt{n}\frac{\bar{x}-\mu_0}{s} = \sqrt{30}\frac{3}{10} \approx 1.644$
Auf der Basis dieser Stichprobe kann H_0 also nicht verworfen werden.

Bemerkung
Der hier beschriebene Test wird auch als *Einstichproben-t-Test* bezeichnet.

B3 t-Test für verbundenene Stichproben

Bei vielen Anwendungen macht man wiederholte Messungen der selben Größe. Wenn beispielsweise zweimal hintereinander der selbe Wert gemessen wird, spricht man von zwei **verbundenen Stichproben**. Häufig kann man die oben dasgestellten Tests dann einfach auf die Differenz der Messwerte anwenden.

Beispiel
Eine Messstation registriert die relative Luftfeuchtigkeit an 5 Tagen an einem bestimmten Ort um 8 Uhr (X) und um 20 Uhr (Y). Es ergeben sich die Werte

X	72.2	56.1	88.4	63.9	65.9
Y	74.5	56.5	87.9	64.2	67.4
Differenz	-2.3	-0.4	0.5	-0.3	-1.5

Wenn man nun davon ausgeht, dass die Differenzen als Realierung von unabhängigen, identisch normalverteilten Zufallsvariablen angesehen werden können, dann kann untersucht werden, ob die erhobenen Daten für oder gegen die Annahme sprechen, dass um 8 Uhr und um 20 Uhr die gleiche Luftfeuchtigkeit herrscht.

Wir testen $H_0\colon \mu = 0$ gegen $H_1\colon \mu \neq 0$.
Bezeichnet man die Differenzdaten mit $z_1, \ldots, z_5$, so ist $\bar{z} = -0.8$ und $s^2 = 1.21$ sowie $n = 5$. Ein 0.975-Quantil der t_4-Verteilung ist gegeben durch $c = 2.776$. Mit diesen Werten ist $T \approx -1.626$.
Die Nullhypothese kann daher nicht abgelehnt werden.

C Test für den Anteilswert p einer Binomialverteilung

Es sei $X_1, \ldots, X_n$ eine Zufallsstichprobe zu $F \in \{B(1, p =| p \in [0,1]\}$. Weiter sei ein Signifikanzniveau $\alpha \in (0,1)$ vorgegeben.

C1 Einseitiger Test

Testproblem $H_0\colon p \geq p_0$ gegen $H_1\colon p < p_0$
Definiere $k(\alpha, n)$ als den maximalen Index mit der Eigenschaft

$$\sum_{\nu=0}^{k(\alpha,n)} \binom{n}{\nu} p_0^\nu (1-p_0)^{n-\nu} \leq \alpha_1$$

Lehne H_0 ab, falls $x_1 + \ldots + x_n \leq k(\alpha, n)$.

Beispiel
Nach Angabe des Herstellers enthalten mindestens 30% aller Chipspackungen ein bestimmtes Spielzeug. Bei einer Stichprobe von 16 Stück enthält eines das entsprechende Spielzeug.
Lässt das zu einem Signifikanzniveau von 5% darauf schließen, dass die Angabe des Herstellers nicht korrekt ist?
Mit den vorgegebenen Daten ergibt sich:

k	$\sum_{\nu=0}^{k} \binom{16}{\nu} 0.3^\nu 0.7^{16-\nu}$
0	0.0033
1	0.0261
2	0.0994

Also besteht der Ablehnungsbereich aus den Werten 0 und 1. In unserem Beispiel wird H_0 also abgelehnt.

Bemerkungen

1. Anders als bei den Tests zur Normalverteilung wird die vorgegebene Irrtumswahrscheinlichkeit beim Binomialtest im Allgemeinen nicht exakt erreicht sondern unterschritten.

2. Analog konstruiert man einen Test für das Testproblem $H_0\colon p \leq p_0$ gegen $H_1\colon p > p_0$

   Definiere dazu $k(\alpha, n)$ als den minimalen Index mit der Eigenschaft

$$\sum_{\nu=k(\alpha,n)}^{n} \binom{n}{\nu} p_0^\nu (1-p_0)^{n-\nu} \leq \alpha$$

   und lehne H_0 ab, falls $x_1 + \ldots + x_n \geq k(\alpha, n)$.

C2 Zweiseitiger Test

Testproblem $H_0: p = p_0$ gegen $H_1: p \neq p_0$
Wähle $\alpha_1, \alpha_2 \geq 0$ mit $\alpha_1 + \alpha_2 = \alpha$.
Definiere $k_1(\alpha_1, n)$ als maximalen Index mit

$$\sum_{\nu=1}^{k_1(\alpha_1,n)} \binom{n}{\nu} p_0^\nu (1-p_0)^{n-\nu} \leq \alpha_1$$

und $k_2(\alpha_2, n)$ als minimalen Index mit

$$\sum_{\nu=k_2(\alpha_2,n)}^{n} \binom{n}{\nu} p_0^\nu (1-p_0)^{n-\nu} \leq \alpha_2$$

Lehne H_0 ab, falls $x_1 + \ldots + x_n \leq k_1(\alpha_1, n)$ oder $x_1 + \ldots + x_n \geq k(\alpha_2, n)$.

Bemerkung
Für kleinere Werte von p_0 wird man im Normalfall $\alpha_1 < \alpha_2$ wählen, für große Werte von p_0 $\alpha_1 > \alpha_2$ wählen. Im Fall $p_0 \approx \frac{1}{2}$ ist $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ eine plausible Wahl.

Beispiel
Bei einer Wahl zwischen zwei Kandidaten um ein politisches Amt haben 47% für A gestimmt und 53% für B. Zwei Wochen nach dem Amtsantritt von B geben in einer Umfrage unter 100 Personen 51 an, sie würden für A stimmen.
Kann man daraus zu einem Signifikanzniveau von 5% schließen, dass sich die Stimmanteil verändert haben?
Es ist also $\alpha = 0.05$, $p_0 = 0.47$ und $n = 100$. Wir setzen $\alpha_1 = \alpha_2 = 0.025$. Dann ergibt sich $k_1(\alpha_1, n) = 36$ und $k_2(\alpha_1, n) = 57$. H_0 kann nicht abgelehnt werden.

C3 Asymptotischer Test

Nach dem zentralen Grenzwertsatz ist für große Stichprobenumfänge die Teststatistik

$$T = \sqrt{n} \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)}}$$

näherungsweise normalverteilt, wenn p_0 der wahre Parameter ist.
Faustregel: Wenn ein solcher asymptotischer Test durchgeführt werden soll, dann sollte bei $\alpha = 0.05$ zumindest $np_0(1-p_0) \geq 10$ sein.
Darauf basierend kann man einen Test wie in Abschnitt A konstruieren.
In den beiden oben betrachteten Beispielen ergibt sich dann:
Beispiele

1. Es gilt $16 \cdot 0.3 \cdot 0.7 < 10$. Ein asymptotischer Test kann also nicht angewandt werden.

2. Es gilt $100 \cdot 0.3 \cdot 0.7 = 21 \geq 10$, also kann auch ein asymptotischer Test angewandt werden.
   $c = 1.96$, $\sqrt{n}\frac{\bar{X}-p_0}{\sqrt{p_0(1-p_0)}} = 10 \cdot \frac{0.51-0.47}{\sqrt{0.47 \cdot 0.7}} \approx 0.80$.

   H_0 kann also nicht abgelehnt werden.

D t-Test für unabhängige Stichproben

Gegeben seien unabhängige Zufallsstichproben $X_1, \ldots, X_n \sim N(\mu_x, \sigma_x^2)$ und $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$. Wenn die Varianzen gleich aber unbekannt sind (d.h. $\sigma_x^2 = \sigma_y^2 = \sigma^2$), dann ist

$$\frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

und $\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$. Außerdem sind die beiden Zufallsvariablen unabhängig. Daher gilt:

$$\sqrt{\frac{nm}{n+m}} \cdot \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t_{n+m-2}$$

Ein Test zum Signifikanzniveau α für das Testproblem $H_0\colon \mu_x = \mu_y$ gegen $H_1\colon \mu_x \neq \mu_y$ ergibt sich durch die Entscheidungsregel:
Es sei c ein $1 - \alpha/2$-Quantil c der t_{n+m-2}-Verteilung und

$$t = \sqrt{\frac{nm}{n+m}} \cdot \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}}$$

Lehne H_0 ab, falls $|t| > c$ ist.

Beispiel
Bei 17 Männchen und 15 Weibchen einer bestimmten Tierart misst man die Länge. Es ergeben sich folgende Werte ($x_1, \ldots, x_{17}$ Männchen, $y_1, \ldots, y_{15}$ Weibchen, jeweils in cm):

$$\bar{x} = 79.4, \bar{y} = 85.7, s_x = 3.55, s_y = 3.37$$

Zu $\alpha = 0.01$ ist $n+m-2 = 30$ und daher $c = 2.75$. Da gleichzeitig $t = 5.129$ ist, wird die Nullhypothese abgelehnt.

Bemerkungen

1. Wir haben im Beispiel einfach vorausgesetzt, dass die Varianzen gleich sind. Dafür existiert ebenfalls ein Test.
2. Wenn die Varianzen nicht gleich sind, benutzt man häufig die Teststatistik

   $$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2/n + S_y^2/m}}$$

   die aber keiner t-Verteilung folgt. (*Behrens-Fisher-Problem*)

E Vergleich der Varianzen zweier normalverteilter Stichproben

Gegeben seien unabhängige Zufallsstichproben $X_1, \ldots, X_n \sim N(\mu_x, \sigma_x^2)$ und $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$. Um nun $H_0\colon \sigma_x^2 = \mu\sigma_y^2$ gegen $H_1\colon \sigma_x^2 \neq \sigma_y^2$ zu testen, betrachtet man die Teststatistik $T = \frac{S_x^2}{S_y^2}$.
Wenn H_0 richtig ist, dann gilt $T \sim F_{n-1,m-1}$. Nun wählt man ein $\alpha/2$-Quantil c_1 von $F_{n-1,m-1}$ und ein $1 - \alpha/2$-Quantil c_2 von $F_{n-1,m-1}$. Ein Test zum Signifikanzniveau α ergibt sich durch die Entscheidungsregel:
Lehne H_0 ab, falls $\frac{s_x^2}{s_y^2} < c_1$ oder $\frac{s_x^2}{s_y^2} > c_2$.

Bemerkung
Bei diesem sogenannten F-Test wählt man α üblicherweise nicht zu klein, z.B. $\alpha = 0.1$.

Beispiel
Vgl. Abschnitt D mit $\alpha = 0.1$. Dann ist $c_1 = 0.42$ und $c_2 = 2.44$.
H_0 kann also nicht abgelehnt werden.