

Wahrscheinlichkeitstheorie und Statistik

Franz Hofbauer

Diese Vorlesung bringt eine elementare Einführung in die Wahrscheinlichkeitstheorie ohne Verwendung der Maßtheorie. Sie besteht aus drei Teilen. Im ersten Teil wird Wahrscheinlichkeit und bedingte Wahrscheinlichkeit definiert und grundlegende Formeln zum Rechnen mit Wahrscheinlichkeiten hergeleitet. Im zweiten Teil geht es um Zufallsvariable und deren Verteilungen. Unabhängigkeit von Zufallsvariablen, Erwartungswert, Varianz und Korrelationskoeffizient werden behandelt sowie Formeln zum Rechnen mit Wahrscheinlichkeitsdichten. Dabei wird mit dem Riemannintegral gearbeitet. Es werden jedoch nicht immer alle für das Rechnen mit dem Riemannintegral notwendigen Voraussetzungen formuliert (diese wären manchmal umständlich), sodass manches Integral eigentlich als Lebesgueintegral interpretiert werden müsste. Der dritte Teil bringt einige Resultate über momentenerzeugende und charakteristische Funktionen (Laplace- und Fouriertransformation) und deren Anwendungen in der Wahrscheinlichkeitstheorie. Insbesondere werden die Konvergenz in Verteilung und der zentrale Grenzwertsatz behandelt. Der vierte Teil gibt eine Einführung in die Statistik. Es werden Parameterschätzer, Konfidenzintervalle, statistische Tests, Varianzanalyse und lineare Regression behandelt.

I. Rechnen mit Wahrscheinlichkeiten

1. Ereignisse und deren Wahrscheinlichkeit

Ein Zufallsexperiment hat verschiedene möglichen Ausfälle. Beispiele für Zufallsexperimente sind das Werfen eines Würfels oder die Lottoziehung. Die möglichen Ausfälle beim Würfeln sind die Augenzahlen 1 bis 6.

Ein Ereignis kann man durch Worte beschreiben oder als Menge darstellen. Wählt man die Mengendarstellung, dann fasst man die Menge aller möglichen Ausfälle des Zufallsexperiments zu einer Menge Ω zusammen. Die Teilmengen von Ω sind dann die Ereignisse. Ist $A \subset \Omega$, dann tritt das Ereignis A genau dann ein, wenn das Zufallsexperiment einen Ausfall liefert, der in A liegt. Beim Zufallsexperiment Würfeln ist $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis "gerade Zahl würfeln" ist die Teilmenge $\{2, 4, 6\}$.

Entsprechend kann man auch die Mengenoperationen interpretieren. Sind $A \subset \Omega$ und $B \subset \Omega$ Ereignisse, dann ist $A \cap B$ das Ereignis, dass sowohl A als auch B eintritt. Der Ausfall des Zufallsexperiments liegt ja genau dann in $A \cap B$, wenn er sowohl in A als auch in B liegt. Analog kann man andere Mengenoperationen interpretieren. Wir tun das in folgender Tabelle. Es wird auch die logische Schreibweise angegeben, die man verwendet, wenn man Ereignisse nicht als Mengen darstellt.

logische Schreibweise	Mengenschreibweise	ist das Ereignis, dass
$A \wedge B$	$A \cap B$	A und B eintreten
$A \vee B$	$A \cup B$	A oder B oder beide eintreten
$\neg A$	$A^c = \Omega \setminus A$	A nicht eintritt
$A \wedge \neg B$	$A \setminus B$	A eintritt, aber B nicht

Die Ereignisse A_1, A_2, \dots, A_n heißen unvereinbar, wenn keine zwei dieser Ereignisse gleichzeitig eintreten können. Verwendet man Mengendarstellung, dann bedeutet das, dass keine zwei dieser Ereignisse ein gemeinsames Element haben, somit die Mengen A_1, A_2, \dots, A_n paarweise disjunkt sind.

Die Wahrscheinlichkeit eines Ereignisses A ist ein Maß für die Häufigkeit, mit der das Ereignis eintritt. Wir definieren die Wahrscheinlichkeit als Grenzwert der relativen Häufigkeiten. Wir wiederholen das Zufallsexperiment k Mal und zählen, wie oft das Ereignis A eintritt. Die Anzahl der Wiederholungen des Zufallsexperiments, bei denen A eintritt, bezeichnen wir mit $N_k(A)$. Der Quotient $\frac{N_k(A)}{k}$ ist dann die relative Häufigkeit des Ereignisses A bei k Wiederholungen. Als Wahrscheinlichkeit des Ereignisses A definieren wir

$$P(A) = \lim_{k \rightarrow \infty} \frac{N_k(A)}{k}$$

wobei wir die Existenz des Grenzwertes einfach annehmen. Wegen $0 \leq N_k(A) \leq k$ folgt $0 \leq \frac{N_k(A)}{k} \leq 1$ und daraus $0 \leq P(A) \leq 1$. Die leere Menge ist das Ereignis, das nie eintritt. Daher gilt $N_k(\emptyset) = 0$, woraus $P(\emptyset) = 0$ folgt. Die Menge Ω ist das Ereignis, das immer eintritt. Daher gilt $N_k(\Omega) = k$, woraus $P(\Omega) = 1$ folgt.

Satz 1 (Additionssatz) Die Ereignisse A_1, A_2, \dots, A_n seien unvereinbar (disjunkt), das heißt keine zwei dieser Ereignisse können gleichzeitig eintreten. Dann gilt

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

wobei wir die Mengenschreibweise verwendet haben. Genauso könnte man statt \cup auch \vee schreiben.

Beweis: Wir wiederholen das Zufallsexperiment k Mal. Die Anzahl der Wiederholungen, bei denen A_1 eintritt, ist $N_k(A_1)$. Die Anzahl der Wiederholungen, bei denen A_2 eintritt, ist $N_k(A_2)$ und so weiter. Daher ist $N_k(A_1) + N_k(A_2) + \dots + N_k(A_n)$ die Anzahl der Wiederholungen, bei denen eines der Ereignisse A_1, A_2, \dots, A_n eintritt. Das kann man als mindestens eines oder genau eines verstehen. Beides ist richtig, da wir voraussetzen, dass bei keiner Wiederholung mehr als eines der Ereignisse A_1, A_2, \dots, A_n eintreten kann.

Andererseits ist $A_1 \cup A_2 \cup \dots \cup A_n$ gerade das Ereignis, dass mindestens eines der Ereignisse A_1, A_2, \dots, A_n eintritt. Daraus folgt

$$N_k(A_1 \cup A_2 \cup \dots \cup A_n) = N_k(A_1) + N_k(A_2) + \dots + N_k(A_n)$$

Dividiert man durch k und lässt k gegen ∞ gehen, so folgt das gewünschte Resultat. \square

Satz 2: Seien A und B Ereignisse. Dann gilt

- (a) $P(A \setminus B) = P(A) - P(A \cap B)$
- (b) $B \subset A \Rightarrow P(A \setminus B) = P(A) - P(B)$
- (c) $B \subset A \Rightarrow P(B) \leq P(A)$
- (d) $P(A^c) = 1 - P(A)$
- (e) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (f) $P(A \cup B) \leq P(A) + P(B)$

Beweis: (a) Die Ereignisse $A \setminus B$ und $A \cap B$ sind disjunkt. Ihre Vereinigung ist A . Aus dem Additionssatz folgt daher $P(A) = P(A \setminus B) + P(A \cap B)$ und (a) ist gezeigt.

(b) Ist $B \subset A$, dann gilt $A \cap B = B$. Aus (a) folgt $P(A \setminus B) = P(A) - P(B)$. Das ist (b).

(c) Wegen (b) gilt $P(A) - P(B) = P(A \setminus B)$. Wegen $P(A \setminus B) \geq 0$ ist (c) gezeigt.

(d) Die Mengen A und A^c sind disjunkt und ihre Vereinigung ist Ω . Der Additionssatz ergibt $P(A) + P(A^c) = P(\Omega)$. Wegen $P(\Omega) = 1$ folgt $P(A^c) = 1 - P(A)$ und (d) ist gezeigt.

(e) Die Ereignisse B und $A \setminus B$ sind disjunkt. Ihre Vereinigung ist $A \cup B$. Der Additionssatz ergibt $P(A \cup B) = P(B) + P(A \setminus B)$. In (a) wurde $P(A \setminus B) = P(A) - P(A \cap B)$ gezeigt. Setzt man das ein, so hat man bereits (e).

(f) Wegen $P(A \cap B) \geq 0$ folgt (f) aus (e). \square

2. Gleichwahrscheinliche Ausfälle

Bei vielen Zufallsexperimenten haben alle Ausfälle die gleiche Wahrscheinlichkeit. Beispiele dafür sind das Werfen eines fairen Würfels und die Lottoziehung. Für eine endliche Menge X sei $|X|$ die Anzahl der Elemente von X . Der folgende Satz führt das Berechnen der Wahrscheinlichkeit auf das Abzählen der Elemente von Mengen zurück.

Satz 3: Die Menge Ω aller Ausfälle eines Zufallsexperiments sei endlich. Sind alle Ausfälle gleich wahrscheinlich, dann gilt $P(A) = |A|/|\Omega|$ für alle $A \subset \Omega$.

Beweis: Sei q die Wahrscheinlichkeit, mit der jeder der Ausfälle eintritt, oder genauer, mit der jedes Ereignis, das nur aus einem Ausfall besteht, eintritt. Sei $A \subset \Omega$ ein beliebiges Ereignis und $k = |A|$ die Anzahl der Elemente von A . Seien A_1, A_2, \dots, A_k die einelementigen

Teilmengen von A . Da diese disjunkt sind und ihre Vereinigung A ist, folgt aus dem Additionssatz, dass $P(A) = P(A_1) + P(A_2) + \dots + P(A_k)$ gilt. Da die Ereignisse A_j einelementig sind, gilt $P(A_j) = q$ für $1 \leq j \leq k$. Es folgt $P(A) = qk = q|A|$.

Für $A = \Omega$ heißt das $P(\Omega) = q|\Omega|$. Wegen $P(\Omega) = 1$ folgt $q = \frac{1}{|\Omega|}$. Damit ist $P(A) = q|A| = |A|/|\Omega|$ gezeigt. \square

Mit Hilfe dieses Satzes und Formeln aus der Kombinatorik kann man Beispiele rechnen.

Beispiel 1: Es wird mit 2 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit, Augensumme 5 zu erhalten? Wie groß ist die Wahrscheinlichkeit, dass mindestens eine 6 auftritt?

Wirft man zwei Würfel gleichzeitig, so nimmt man unterscheidbare Würfel (einen roten und einen grünen). Die Ausfälle des Zufallsexperiments sind dann Paare von Augenzahlen, wobei die Augenzahl des roten Würfels an die erste Stelle und die Augenzahl des grünen Würfels an die zweite Stelle geschrieben wird. Dadurch erhält man gleichwahrscheinliche Ausfälle. Wir können Satz 3 anwenden. Die Ausfallsmenge Ω ist dann $\{(i, j) : 1 \leq i, j \leq 6\}$. Sie hat $6^2 = 36$ Elemente.

Das Ereignis "Augensumme 5" ist die Menge aller Paare von Augenzahlen, deren Summe 5 ist. Wir erhalten $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$. Das ergibt $P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36}$.

Die Menge $A = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5)\}$ stellt das Ereignis "mindestens eine 6" dar. Das ergibt $P(A) = \frac{|A|}{|\Omega|} = \frac{11}{36}$.

Beispiel 2: Wie groß ist die Wahrscheinlichkeit, dass beim Lotto 6 aus 45 von den sechs Zahlen, die ich getippt habe, genau fünf gezogen werden?

Jetzt ist die Ausfallsmenge Ω die Menge aller ungeordneten Stichproben vom Umfang 6, also die Menge aller 6-elementigen Teilmengen aus der Menge der ersten 45 natürlichen Zahlen. Somit ist $|\Omega| = \binom{45}{6}$. Das Ereignis "genau 5 meiner getippten Zahlen werden gezogen" ist die Menge A aller 6-elementigen Teilmengen, die fünf der 6 getippten und eine der 39 nicht getippten Zahlen enthalten. Es gibt $\binom{6}{5}$ fünfelementige Teilmengen aus den getippten Zahlen und $\binom{39}{1}$ einelementige Teilmengen aus den nicht getippten Zahlen. Daraus folgt $|A| = \binom{6}{5} \binom{39}{1}$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{6}{5} \binom{39}{1}}{\binom{45}{6}}$.

Beispiel 3: In einer Schachtel sind 7 rote, 5 grüne und 8 blaue Kugeln. Es werden zufällig 9 Kugeln ohne Zurücklegen (eine 9-elementige Teilmenge) gezogen. Wie groß ist die Wahrscheinlichkeit, 2 rote, 3 grüne und 4 blaue Kugeln zu ziehen?

Die Ausfallsmenge Ω ist die Menge aller 9-elementigen Teilmengen einer 20-elementigen Menge. Somit ist $|\Omega| = \binom{20}{9}$. Das gefragte Ereignis ist die Menge A aller 9-elementigen Teilmengen, die 2 rote, 3 grüne und 4 blaue Kugeln enthalten. Da es $\binom{7}{2}$ 2-elementige Teilmengen aus den 7 roten Kugeln, $\binom{5}{3}$ 3-elementige Teilmengen aus den 5 grünen Kugeln und $\binom{8}{4}$ 4-elementige Teilmengen aus den 8 blauen Kugeln gibt, erhalten wir $|A| = \binom{7}{2} \binom{5}{3} \binom{8}{4}$. Daraus ergibt sich dann $P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{7}{2} \binom{5}{3} \binom{8}{4}}{\binom{20}{9}}$.

3. Bedingte Wahrscheinlichkeit

Wir führen ein Zufallsexperiment durch und interessieren uns dafür, ob das Eintreten eines Ereignisses B ein anderes Ereignis A begünstigt. Dazu wiederholen wir das Zufallsexperiment k Mal. Wir berücksichtigen jedoch nur die Wiederholungen, bei denen das Ereignis B

eintritt. Unter diesen zählen wir die Häufigkeit des Ereignisses A . Diese ist die Anzahl der Wiederholungen, wo A und B eintreten, also $N_k(A \cap B)$. Die bedingte relative Häufigkeit des Eintretens von A unter B ist dann $N_k(A \cap B)/N_k(B)$, da ja nur die Wiederholungen des Zufallsexperiments berücksichtigt werden, bei denen B eintritt. Lässt man k gegen ∞ gehen, so erhält man wieder eine Wahrscheinlichkeit, diesmal die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B (oder gegeben B), die mit $P(A|B)$ bezeichnet wird. Wir definieren

$$P(A|B) = \lim_{k \rightarrow \infty} \frac{N_k(A \cap B)}{N_k(B)}$$

Mit Hilfe der Definition der Wahrscheinlichkeit erhalten wir dann

$$P(A|B) = \lim_{k \rightarrow \infty} \frac{\frac{1}{k} N_k(A \cap B)}{\frac{1}{k} N_k(B)} = \frac{P(A \cap B)}{P(B)}$$

Die bedingte Wahrscheinlichkeit $P(A|B) = \frac{P(A \cap B)}{P(B)}$ ist eine Maßzahl für die Häufigkeit des Eintretens des Ereignisses A unter der Bedingung, dass auch das Ereignis B eintritt. Dabei wird immer vorausgesetzt, dass $P(B) > 0$ gilt. In den Anwendungsbeispielen werden Wahrscheinlichkeiten oft mit Hilfe von bedingten Wahrscheinlichkeiten berechnet.

Satz 4 (Multiplikationssatz) Für Ereignisse A_1, A_2, \dots, A_n gilt

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

Es wird dabei vorausgesetzt, dass die Ereignisse, die als Bedingungen auftreten, Wahrscheinlichkeit > 0 haben.

Beweis: Aus der Definition folgt $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$, $P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$ und so fort bis $P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) = \frac{P(A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n)}{P(A_1 \cap A_2 \cap \dots \cap A_{n-1})}$. Setzt man das in die rechte Seite der Formel ein, so kürzt sich alles weg, es bleibt nur die linke Seite übrig. \square

Typische Anwendungsbeispiele für den Multiplikationssatz sind geordnete Stichproben. Dazu gehört auch wiederholtes Würfeln und Münzenwerfen.

Beispiel 4: Aus der Buchstabenmenge ANANAS werden der Reihe nach drei Buchstaben ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, ANA zu ziehen?

Sei A_1 = "erster Zug ist A", A_2 = "zweiter Zug ist N" und A_3 = "dritter Zug ist A". Gesucht ist $P(A_1 \cap A_2 \cap A_3)$. Aus Satz 4 folgt $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$.

$P(A_1)$ = Wahrscheinlichkeit, aus ANANAS ein A zu ziehen = $\frac{3}{6}$

$P(A_2|A_1)$ = Wahrscheinlichkeit, dass der zweite Zug N ist, wenn der erste A war

= Wahrscheinlichkeit, aus NANAS ein N zu ziehen = $\frac{2}{5}$

$P(A_3|A_1 \cap A_2)$ = Wahrscheinlichkeit, dass dritter Zug A ist, wenn vorher AN gezogen wurde

= Wahrscheinlichkeit, aus ANAS ein A zu ziehen = $\frac{2}{4}$

Wir erhalten somit $P(A_1 \cap A_2 \cap A_3) = \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{2}{4} = \frac{1}{10}$.

Beispiel 5: Aus einer Menge von drei schwarzen, zwei blauen und einer roten Kugel wird solange ohne Zurücklegen gezogen, bis eine schwarze Kugel kommt. Wie groß ist die Wahrscheinlichkeit, die rote Kugel zu ziehen?

Das gesuchte Ereignis lässt sich in drei Ereignisse zerlegen, die den Zugfolgen R, BR und BBR entsprechen. Nach Satz 4 haben diese Zugfolgen die Wahrscheinlichkeiten $\frac{1}{6}$, $\frac{2}{6} \cdot \frac{1}{5} = \frac{2}{30}$ und $\frac{2}{6} \cdot \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{60}$. Die Summe dieser Wahrscheinlichkeiten ist nach den Additionssatz die Wahrscheinlichkeit des gefragten Ereignisses. Sie ist somit $\frac{1}{6} + \frac{2}{30} + \frac{1}{60} = \frac{1}{4}$.

Beispiel 6: Wie groß ist beim Lotto die Wahrscheinlichkeit, dass 5 der 6 von mir getippten Zahlen gezogen werden, und die sechste getippte Zahl als Zusatzzahl gezogen wird?

Sei A das Ereignis “5 der 6 getippten Zahlen werden gezogen” und B das Ereignis “die sechste getippte Zahl wird als Zusatzzahl gezogen”. Gesucht ist $P(A \cap B) = P(A)P(B|A)$. In Beispiel 2 wurde $P(A) = \binom{6}{5} \binom{39}{1} / \binom{45}{6}$ berechnet. Weiters ist $P(B|A)$ die Wahrscheinlichkeit, die sechste getippte Zahl zu ziehen, wenn bei den vorhergehenden sechs Zügen bereits fünf getippte Zahlen gezogen wurden, das heißt aus 39 Zahlen die sechste getippte Zahl zu ziehen. Diese Wahrscheinlichkeit ist $\frac{1}{39}$. Somit haben wir $P(A \cap B) = P(A) \cdot \frac{1}{39} = \binom{6}{5} / \binom{45}{6}$.

Weitere wichtige Formeln zum Rechnen von Beispielen sind die Formel für die totale Wahrscheinlichkeit und die Formel von Bayes.

Definition: Man sagt, die Ereignisse B_1, B_2, \dots, B_n bilden eine Zerlegung der Ausfallsmenge Ω wenn sie paarweise disjunkt sind und wenn $\bigcup_{j=1}^n B_j = \Omega$ gilt.

Bemerkung: Die beiden folgenden Aussagen sind äquivalent

- (1) Die Ereignisse B_1, B_2, \dots, B_n bilden eine Zerlegung von Ω .
- (2) Genau eines der Ereignisse B_1, B_2, \dots, B_n tritt ein.

Satz 5: Seien B_1, B_2, \dots, B_n Ereignisse, die Wahrscheinlichkeit > 0 haben und eine Zerlegung von Ω bilden. Für ein Ereignis $A \subset \Omega$ gilt dann

- (a) $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$
- (b) $P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)}$ für $1 \leq j \leq n$, wenn $P(A) > 0$

Man nennt (a) die Formel für die totale Wahrscheinlichkeit und (b) die Formel von Bayes.

Beweis: Für (a) verwenden wir zuerst die Definition der bedingten Wahrscheinlichkeit und erhalten $\sum_{j=1}^n P(A|B_j)P(B_j) = \sum_{j=1}^n P(A \cap B_j)$. Da die Ereignisse B_j paarweise disjunkt sind, sind es auch die Ereignisse $A \cap B_j$ und es folgt $\sum_{j=1}^n P(A \cap B_j) = P(\bigcup_{j=1}^n (A \cap B_j))$ aus dem Additionssatz. Wegen $\bigcup_{j=1}^n B_j = \Omega$ erhalten wir $\bigcup_{j=1}^n (A \cap B_j) = A \cap \bigcup_{j=1}^n B_j = A$ und (a) ist bewiesen.

Es gilt $P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}$ und $P(A \cap B_j) = P(A|B_j)P(B_j)$ nach der Definition der bedingten Wahrscheinlichkeit. Setzt man die zweite Formel in die erste ein, dann hat man (b). \square

Beispiel 7: Eine Versicherung teilt die Autofahrer in zwei Typen ein, in Risikofahrer und in Sicherheitsfahrer. Die Wahrscheinlichkeit, dass ein Sicherheitsfahrer in einem Jahr (Versicherungsperiode) einen Unfall hat, ist 0.06. Die Wahrscheinlichkeit, dass ein Risikofahrer in einem Jahr einen Unfall hat, ist 0.6. Die Versicherung weiß aus Erfahrung, dass $\frac{5}{6}$ der Autofahrer Sicherheitsfahrer und $\frac{1}{6}$ der Autofahrer Risikofahrer sind. Ein Autofahrer schließt eine Versicherung ab (man sieht ihm natürlich nicht an, von welchem Typ er ist). Wie groß ist die Wahrscheinlichkeit, dass er nächstes Jahr einen Unfall haben wird?

Sei B_1 das Ereignis, dass der Autofahrer ein Sicherheitsfahrer ist, und B_2 das Ereignis, dass der Autofahrer ein Risikofahrer ist. Es tritt genau eines dieser beiden Ereignisse ein, sodass B_1 und B_2 eine Zerlegung von Ω bilden. Es gilt $P(B_1) = \frac{5}{6}$ und $P(B_2) = \frac{1}{6}$.

Gefragt ist nach der Wahrscheinlichkeit des Ereignisses A , dass der Autofahrer im nächsten Jahr einen Unfall hat. Wenn der Autofahrer ein Sicherheitsfahrer ist, dann ist diese Wahrscheinlichkeit 0.06, das heißt $P(A|B_1) = 0.06$. Wenn der Autofahrer ein Risikofahrer ist, dann ist diese Wahrscheinlichkeit 0.6, das heißt $P(A|B_2) = 0.6$. Die Formel für die totale Wahrscheinlichkeit ergibt $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0.06 \cdot \frac{5}{6} + 0.6 \cdot \frac{1}{6} = 0.15$. Die Wahrscheinlichkeit für einen Unfall im nächsten Jahr beträgt also 0.15.

Beispiel 8: Seit sich der Autofahrer aus dem letzten Beispiel versichern ließ, ist ein Jahr vergangen. Es hat sich herausgestellt, dass er einen Unfall hatte. Wie groß ist die Wahrscheinlichkeit, dass der Autofahrer ein Risikofahrer ist?

Wir wissen, dass der Autofahrer einen Unfall hatte. Gefragt ist daher die bedingte Wahrscheinlichkeit dafür, dass er ein Risikofahrer ist. Verwendet man die Bezeichnung aus dem letzten Beispiel, dann ist das die bedingte Wahrscheinlichkeit $P(B_2|A)$.

Die Formel von Bayes besagt $P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)}$. Aus dem letzten Beispiel wissen wir $P(A|B_2)P(B_2) = 0.6 \cdot \frac{1}{6} = 0.1$ und $P(A) = 0.15$. Wir erhalten $P(B_2|A) = \frac{0.1}{0.15} = \frac{2}{3}$. Nach diesem Unfall ist der Autofahrer mit Wahrscheinlichkeit $\frac{2}{3}$ ein Risikofahrer.

Zwei Ereignisse A und B sind unabhängig, wenn das eine das andere nicht beeinflusst. Unter den Wiederholungen des Zufallsexperiments, bei denen B eintritt, wird A genauso häufig auftreten, wie unter allen Wiederholungen insgesamt. Das führt zu

Definition: Zwei Ereignisse A und B heißen unabhängig, wenn $P(A|B) = P(A)$ gilt, das heißt $P(A \cap B) = P(A)P(B)$.

Definition: Sei $n \geq 2$. Dann werden die Ereignisse A_1, A_2, \dots, A_n unabhängig genannt, wenn $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$ für alle Teilmengen $\{i_1, i_2, \dots, i_k\}$ von $\{1, 2, \dots, n\}$ mit $k \geq 2$ gilt.

Satz 6: Die Ereignisse A_1, A_2, \dots, A_n seien unabhängig. Dann sind auch die Ereignisse $A_1^c, A_2^c, \dots, A_n^c$ unabhängig. Die Ereignisse A_1, A_2, \dots, A_n bleiben unabhängig, wenn man einige dieser Ereignisse oder auch alle durch ihre Komplemente ersetzt.

Beweis: Sei $\{i_1, i_2, \dots, i_k\}$ eine Teilmenge von $\{1, 2, \dots, n\}$, die 1 nicht enthält. Nach Voraussetzung gilt dann $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$. Ebenso gilt $P(A_1 \cap A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_1)P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$. Aus Satz 2 erhalten wir $P(A_1^c \cap A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) - P(A_1 \cap A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$ und $P(A_1^c) = 1 - P(A_1)$, sodass $P(A_1^c \cap A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_1^c)P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$ mit Hilfe obiger Gleichungen folgt. Damit haben wir alle für die Unabhängigkeit der Ereignisse $A_1^c, A_2^c, \dots, A_n^c$ notwendigen Gleichungen erhalten. Die zweite Aussage des Satzes folgt durch wiederholtes Anwenden der ersten. \square

4. Axiome für die Wahrscheinlichkeit

Mit der Wahrscheinlichkeit, die wir eingeführt haben, hat man zwei Probleme. Der Additionssatz für endlich viele Ereignisse, wie er im ersten Kapitel vorgekommen ist, ist nicht allgemein genug, um darauf eine Wahrscheinlichkeitstheorie aufzubauen. Dazu kommt noch, dass man bei überabzählbarem Ω eine Wahrscheinlichkeit, die die für eine mathematische Theorie notwendigen Eigenschaften hat, nicht mehr auf der ganzen Potenzmenge, das heißt für alle Teilmengen von Ω , definieren kann. Um solchen Schwierigkeiten zu entkommen, haben die Mathematiker den axiomatischen Zugang erfunden. Man listet die Eigenschaften auf, die man braucht – diese heißen Axiome – und nennt dann alles, was diese Eigenschaften erfüllt, eine Wahrscheinlichkeit.

Für die Wahrscheinlichkeitstheorie haben sich folgende Axiome eingebürgert.

Definition: Eine Teilmenge \mathcal{A} der Potenzmenge von Ω heißt σ -Algebra, wenn gilt

- (a) $\emptyset \in \mathcal{A}$
- (b) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- (c) $A_j \in \mathcal{A}$ für $j \geq 1 \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$

Auf der Ausfallsmenge Ω legt man eine σ -Algebra \mathcal{A} fest. Die Mengen in \mathcal{A} nennt man Ereignisse. Für endliches oder abzählbares Ω ist \mathcal{A} immer die ganze Potenzmenge.

Definition: Sei \mathcal{A} eine σ -Algebra. Eine Abbildung $P : \mathcal{A} \rightarrow [0, 1]$ heißt Wahrscheinlichkeit, wenn folgende Eigenschaften erfüllt sind

- (a) $P(\emptyset) = 0$ und $P(\Omega) = 1$
 (b) sind $A_1, A_2, \dots \in \mathcal{A}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

Man wählt den Additionssatz in der Allgemeinheit, in der man ihn braucht, nämlich für abzählbar viele Ereignisse, einfach als Axiom. Gilt er für abzählbar viele Ereignisse A_1, A_2, \dots , dann folgt er auch für endlich viele, indem man $A_j = \emptyset$ setzt für $j > n$. Alle bisher bewiesenen Sätze gelten auch für die axiomatische Wahrscheinlichkeit, da sie nur auf dem Additionssatz und der Gleichung $P(\Omega) = 1$ aufbauen, wobei die bedingte Wahrscheinlichkeit durch die Formel $P(A|B) = \frac{P(A \cap B)}{P(B)}$ definiert wird.

Als Beispiel, das diese Axiome erfüllt, sei der diskrete Wahrscheinlichkeitsraum angeführt. Von einem solchen spricht man, wenn Ω endlich oder abzählbar und \mathcal{A} die ganze Potenzmenge von Ω ist. Eine Wahrscheinlichkeit kann man dann durch die Angabe von nichtnegativen Zahlen p_k für alle $k \in \Omega$ festlegen, wobei $\sum_{k \in \Omega} p_k = 1$ ist (p_k ist die Wahrscheinlichkeit, mit der der Ausfall k auftritt). Für $A \subset \Omega$ definieren wir $P(A) = \sum_{k \in A} p_k$. Klarerweise gilt dann $0 \leq P(A) \leq 1$ für alle $A \subset \Omega$ und $P(\Omega) = 1$. Wir interpretieren die leere Summe als 0 und haben dann auch $P(\emptyset) = 0$. Sind $A_1, A_2, \dots \subset \Omega$ paarweise disjunkt und $A = \bigcup_{j=1}^{\infty} A_j$, dann gilt $P(A) = \sum_{k \in A} p_k = \sum_{j=1}^{\infty} \sum_{k \in A_j} p_k = \sum_{j=1}^{\infty} P(A_j)$, wobei man die Summe umordnen darf, da alle $p_k \geq 0$ sind. Damit haben wir alle Axiome nachgeprüft.

Beispiele für Wahrscheinlichkeiten mit überabzählbarer Ausfallsmenge Ω werden mit Methoden aus der Maßtheorie konstruiert und können daher hier nicht gegeben werden. Für später beweisen wir noch Folgerungen aus den Axiomen.

Satz 7 (Stetigkeitssatz) Seien A_1, A_2, A_3, \dots Ereignisse.

- (a) Wenn $A_n \subset A_{n+1}$ für $n \geq 1$ und $A = \bigcup_{n=1}^{\infty} A_n$ gilt, dann ist $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.
 (b) Wenn $A_n \supset A_{n+1}$ für $n \geq 1$ und $A = \bigcap_{n=1}^{\infty} A_n$ gilt, dann ist $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

Beweis: Um (a) zu zeigen, sei $C_1 = A_1$ und $C_n = A_n \setminus A_{n-1}$ für $n \geq 2$. Für $i < j$ gilt $C_i \subset A_i \subset A_{j-1}$ und $C_j \cap A_{j-1} = \emptyset$, also auch $C_i \cap C_j = \emptyset$. Die Mengen C_1, C_2, \dots sind paarweise disjunkt. Weiters gilt $A = \bigcup_{k=1}^{\infty} C_k$, sodass $P(A) = \sum_{k=1}^{\infty} P(C_k)$ aus den Axiomen folgt. Ebenso gilt $A_n = \bigcup_{k=1}^n C_k$ und daher auch $P(A_n) = \sum_{k=1}^n P(C_k)$. Somit haben wir $P(A) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(C_k) = \lim_{n \rightarrow \infty} P(A_n)$ und (a) ist gezeigt.

Um (b) zu zeigen, sei $B_n = A_n^c$ für $n \geq 1$ und $B = A^c$. Wegen $A_n \supset A_{n+1}$ für $n \geq 1$ und $A = \bigcap_{n=1}^{\infty} A_n$ folgt $B_n \subset B_{n+1}$ für $n \geq 1$ und $B = \bigcup_{n=1}^{\infty} B_n$. Aus (a) erhalten wir dann $\lim_{n \rightarrow \infty} P(B_n) = P(B)$. Wegen Satz 2(d) gilt $P(B_n) = 1 - P(A_n)$ für $n \geq 1$ und $P(B) = 1 - P(A)$, woraus dann $\lim_{n \rightarrow \infty} P(A_n) = P(A)$ folgt. Damit ist auch (b) gezeigt. \square

Satz 8: Für beliebige Ereignisse A_1, A_2, A_3, \dots gilt $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

Beweis: Für $n \geq 1$ sei $B_n = A_n \setminus \bigcup_{j=1}^{n-1} A_j$. Die Mengen B_1, B_2, \dots sind paarweise disjunkt. Aus den Axiomen folgt $P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$. Wegen $B_n \subset A_n$ folgt $P(B_n) \leq P(A_n)$ aus Satz 2(c). Da auch $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ gilt, haben wir $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$. \square

In den folgenden Kapiteln arbeiten wir mit einer Wahrscheinlichkeit, für die wir annehmen, dass alle bisher bewiesenen Formeln gelten. Wir werden uns jedoch die Freiheit nehmen, Ereignisse auf eine weniger schwerfällige Art als die Mengenschreibweise darzustellen.

II. Zufallsvariable

1. Zufallsvariable und Verteilungsfunktion

Sind die Ausfälle eines Zufallsexperiments Zahlen (zum Beispiel Anzahlen oder Messwerte), dann verwendet man Zufallsvariable zur Darstellung von Ereignissen. Man bezeichnet den Ausfall des Zufallsexperiments mit einer Variable (für die man üblicherweise Großbuchstaben verwendet), und stellt Ereignisse durch Gleichungen, Ungleichungen und dergleichen dar. Will man zum Beispiel die Qualität von Glühbirnen prüfen, so zieht man eine Zufallsstichprobe. Sei X die Anzahl der defekten Glühbirnen in der Stichprobe. Das Ereignis "höchstens k defekte Glühbirnen in der Stichprobe" lässt sich dann als Ungleichung $X \leq k$ schreiben.

Mit Zufallsvariablen kann man auch rechnen. Es gilt $3X + 5 \leq 11 \Leftrightarrow X \leq 2$. Daher haben die durch diese beiden Ungleichungen dargestellten Ereignisse dieselbe Wahrscheinlichkeit

$$P(3X + 5 \leq 11) = P(X \leq 2)$$

Das Ereignis $X \leq 3$ ist das Komplementärereignis zu $X > 3$. Daher gilt wegen Satz 2 (d)

$$P(X \leq 3) = 1 - P(X > 3)$$

Es gilt $X \leq 3 \Leftrightarrow X \leq 2 \vee X \in (2, 3]$. Aus dem Additionssatz folgt

$$P(X \leq 3) = P(X \leq 2) + P(X \in (2, 3])$$

da die beiden Ereignisse $X \leq 2$ und $X \in (2, 3]$ unvereinbar sind.

Man kann die durch Zufallsvariable dargestellten Ereignisse auch wieder in die Mengendarstellung übersetzen. Dazu nimmt man an, dass im Hintergrund ein Zufallsexperiment mit Ausfallsmenge Ω abläuft. Eine Zufallsvariable X wird dann als Abbildung von Ω nach \mathbb{R} definiert. Die oben beschriebenen Ereignisse kann man dann als Teilmengen von Ω auffassen. Zum Beispiel wird das Ereignis $X \leq t$ zur Menge $\{\omega \in \Omega : X(\omega) \leq t\}$. Wir werden von dieser Mengendarstellung jedoch keinen Gebrauch machen.

Um Methoden aus der Analysis anwenden zu können, führen wir die Verteilungsfunktion einer Zufallsvariable ein.

Definition: Die Abbildung $F : \mathbb{R} \rightarrow [0, 1]$ definiert durch $F(t) = P(X \leq t)$ heißt Verteilungsfunktion der Zufallsvariable X .

Satz 9: Sei F die Verteilungsfunktion einer Zufallsvariable X . Dann gilt

- (a) $s < t \Rightarrow P(s < X \leq t) = F(t) - F(s)$ und $F(s) \leq F(t)$ (F ist monoton wachsend)
- (b) $\lim_{t \downarrow s} F(t) = F(s)$ für alle $s \in \mathbb{R}$ (F ist rechtsseitig stetig)
- (c) $\lim_{t \rightarrow \infty} F(t) = 1$ und $\lim_{t \rightarrow -\infty} F(t) = 0$

Beweis: Sei $s < t$. Das Ereignis $X \leq t$ tritt genau dann ein, wenn $X \leq s$ oder $s < X \leq t$ eintritt, wobei die Ereignisse $X \leq s$ und $s < X \leq t$ unvereinbar sind. Aus dem Additionssatz folgt $P(X \leq t) = P(X \leq s) + P(s < X \leq t)$. Wegen $F(t) = P(X \leq t)$ und $F(s) = P(X \leq s)$ ist damit $P(s < X \leq t) = F(t) - F(s)$ gezeigt. Da eine Wahrscheinlichkeit immer ≥ 0 ist, haben wir auch $F(s) \leq F(t)$ und (a) ist bewiesen.

Sei $(t_n)_{n \geq 1}$ eine monoton fallende Folge mit $\lim_{n \rightarrow \infty} t_n = s$. Die Ereignisse $X \leq t_n$ bilden dann eine monoton abnehmende Folge, die gegen das Ereignis $X \leq s$ geht (deren Durchschnitt das Ereignis $X \leq s$ ist). Aus dem Stetigkeitssatz folgt $\lim_{n \rightarrow \infty} P(X \leq t_n) = P(X \leq s)$, das heißt $\lim_{n \rightarrow \infty} F(t_n) = F(s)$. Damit ist auch $\lim_{t \downarrow s} F(t) = F(s)$ gezeigt. Das ist (b).

Sei $(t_n)_{n \geq 1}$ eine Folge in \mathbb{R} , die monoton wachsend gegen ∞ geht. Die Ereignisse $X \leq t_n$ bilden dann eine monoton aufsteigende Folge, die gegen das Ereignis $X < \infty$ geht (deren

Vereinigung das Ereignis $X < \infty$ ist). Nun folgt $\lim_{n \rightarrow \infty} P(X \leq t_n) = P(X < \infty)$ aus dem Stetigkeitssatz. Da das Ereignis $X < \infty$ immer eintritt, heißt das $\lim_{n \rightarrow \infty} F(t_n) = 1$. Damit ist auch $\lim_{t \rightarrow \infty} F(t) = 1$ gezeigt. Den Beweis von $\lim_{t \rightarrow -\infty} F(t) = 0$ erhalten wir, indem wir $s = -\infty$ im Beweis von (b) setzen und beachten, dass das Ereignis $X \leq -\infty$ nie eintritt, also Wahrscheinlichkeit 0 hat. Damit ist auch (c) bewiesen. \square

Bemerkung: Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Funktion, die die drei Eigenschaften aus Satz 9 erfüllt. Man kann zeigen, dass dann eine Zufallsvariable X existiert, die F als Verteilungsfunktion hat. In der Maßtheorie zeigt man, dass es für $\Omega = \mathbb{R}$ eine σ -Algebra \mathcal{B} gibt, die alle Intervalle enthält, und eine Wahrscheinlichkeit $P : \mathcal{B} \rightarrow [0, 1]$, sodass $P((-\infty, t]) = F(t)$ für alle $t \in \mathbb{R}$ gilt. Definiert man $X : \mathbb{R} \rightarrow \mathbb{R}$ als Identität, dann gilt $P(\{\omega : X(\omega) \leq t\}) = P(\{\omega : \omega \leq t\}) = P((-\infty, t]) = F(t)$ für $t \in \mathbb{R}$, sodass X eine Zufallsvariable mit Verteilungsfunktion F ist.

In den Anwendungen hat man es praktisch immer mit diskreten oder kontinuierlichen (stetigen) Zufallsvariablen zu tun, die wir jetzt definieren.

Definition: Eine Zufallsvariable X heißt diskret, wenn sie Werte in einer endlichen oder abzählbaren Menge R annimmt. Für $k \in R$ definieren wir $w(k) = P(X = k)$. Wir nennen R den Wertebereich und $w(k)$ mit $k \in R$ die Einzelwahrscheinlichkeiten der Zufallsvariablen X .

Kennt man die Einzelwahrscheinlichkeiten einer Zufallsvariablen oder hat sie mit Hilfe der Methoden aus den letzten Kapiteln bestimmt, so kann man damit Wahrscheinlichkeiten von Ereignissen berechnen.

Satz 10: Sei X eine diskrete Zufallsvariable mit Wertebereich R und Einzelwahrscheinlichkeiten $w(k)$ für $k \in R$.

- (a) Dann gilt $w(k) \geq 0$ für $k \in R$ und $\sum_{k \in R} w(k) = 1$.
 (b) Für $B \subset R$ gilt $P(X \in B) = \sum_{k \in B} w(k)$.

Beweis: Wir zeigen zuerst (b). Sei $B = \{k_1, k_2, k_3, \dots\}$. Das ist eine endliche oder abzählbare Menge. Es gilt

$$X \in B \Leftrightarrow X = k_1 \vee X = k_2 \vee X = k_3 \vee \dots$$

Da die rechtsstehenden Ereignisse unvereinbar sind, folgt aus dem Additionssatz

$$P(X \in B) = P(X = k_1) + P(X = k_2) + P(X = k_3) + \dots = w(k_1) + w(k_2) + \dots = \sum_{k \in B} w(k)$$

Damit ist (b) gezeigt.

Die erste Aussage von (a) ist klar, da $w(k) = P(X = k)$ eine Wahrscheinlichkeit und somit ≥ 0 ist. Setzt man $B = R$ in (b), so folgt $P(X \in R) = \sum_{k \in R} w(k)$. Da R alle möglichen Werte von X enthält, gilt $P(X \in R) = 1$. Damit ist auch (a) gezeigt. \square

Bemerkung: Sind $w(k)$ mit $k \in R$ die Einzelwahrscheinlichkeiten einer diskreten Zufallsvariablen X und F deren Verteilungsfunktion, dann gilt $F(t) = P(X \leq t) = P(X \in R_t) = \sum_{k \in R_t} w(k)$, wobei $R_t = \{k \in R : k \leq t\}$ gesetzt wurde. Daraus erkennt man, dass F in den Punkten $k \in R$ Sprungstellen mit Sprunghöhen $w(k)$ hat und zwischen diesen Punkten konstant ist.

Definition: Eine Zufallsvariable X heißt kontinuierlich, wenn eine integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ existiert, sodass $P(X \leq t) = \int_{-\infty}^t f(x) dx$ für alle $t \in \mathbb{R}$ gilt. Man nennt f Wahrscheinlichkeitsdichte der Zufallsvariablen X .

Kennt man eine Wahrscheinlichkeitsdichte einer Zufallsvariablen, so kann man damit Wahrscheinlichkeiten von Ereignissen berechnen.

Satz 11: Sei X eine kontinuierliche Zufallsvariable und $f : \mathbb{R} \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte von X . Dann gilt

(a) $\int_{-\infty}^{\infty} f(x) dx = 1$

(b) für ein Intervall $B \subset \mathbb{R}$ gilt $P(X \in B) = \int_B f(x) dx$

Der Beweis folgt später, wo er gleich für den mehrdimensionalen Fall gegeben wird.

Bemerkung: Oft existiert ein Intervall W (das auch unbeschränkt sein kann), sodass $f > 0$ im Innern von W und $f = 0$ außerhalb des Abschlusses von W gilt. Wir nennen dann W den Wertebereich von X . Es gilt ja $P(X \in W) = \int_W f(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1$, sodass X mit Wahrscheinlichkeit 1, das heißt immer, in W liegt.

Will man eine Wahrscheinlichkeitsdichte f einer Zufallsvariablen X berechnen, so berechnet man zuerst die Verteilungsfunktion $F(t) = P(X \leq t)$ für $t \in \mathbb{R}$. Man erhält dann eine Wahrscheinlichkeitsdichte f als Ableitung von F .

Beispiel 9: Unabhängig voneinander werden zwei Punkte a und b zufällig im Intervall $[0, 1]$ gewählt. Sei X das Minimum von a und b . Gesucht ist eine Dichte f von X .

Wir berechnen zuerst $F(t) = P(X \leq t)$. Die Ausfallsmenge unseres Zufallsexperiments ist $\Omega = [0, 1] \times [0, 1]$. Das Ereignis $X \leq t$ entspricht der Teilmenge $A = \{(a, b) : \min(a, b) \leq t\}$ von Ω . Für $t < 0$ ist $A = \emptyset$. Für $t \geq 1$ ist $A = \Omega$. Für $0 \leq t < 1$ ist $A = \Omega \setminus (t, 1] \times (t, 1]$. Für $t < 0$ gilt $F(t) = P(\emptyset) = 0$. Für $t \geq 1$ gilt $F(t) = P(\Omega) = 1$. Für $t \in [0, 1)$ verwenden wir die Formel $P(A) = |A|/|\Omega|$ für gleichwahrscheinliche Ausfälle, wobei wir $| \cdot |$ als Fläche interpretieren, und erhalten $F(t) = \frac{1 - (1-t)^2}{1} = 2t - t^2$. Wir haben also $F(t) = 0$ für $t < 0$, $F(t) = 2t - t^2$ für $0 \leq t < 1$ und $F(t) = 1$ für $t \geq 1$ gefunden. Wegen $f(x) = F'(x)$ ergibt sich $f(x) = 0$ für $x \notin [0, 1)$ und $f(x) = 2 - 2x$ für $x \in [0, 1)$. Der Wertebereich der Zufallsvariablen X ist also das Intervall $[0, 1]$.

Bemerkung: Wie man im letzten Beispiel den Funktionswert von f im Punkt 0 wählt, spielt keine Rolle. Das hat keinen Einfluß auf die mit Hilfe von f berechneten Wahrscheinlichkeiten, da diese ja Integrale über f sind. Wahrscheinlichkeitsdichten sind nicht eindeutig bestimmt. Ändert man eine Wahrscheinlichkeitsdichte zum Beispiel in einem Punkt, dann ist sie immer noch eine Wahrscheinlichkeitsdichte.

2. Binomialverteilung und geometrische Verteilung

In diesem und den nächsten Kapiteln werden die wichtigsten Verteilungen behandelt. Wir beginnen mit der Binomialverteilung.

Definition: Seien $n \geq 1$ und $0 < p < 1$. Eine diskrete Zufallsvariable X mit Wertebereich $R = \{0, 1, 2, \dots, n\}$ und Einzelwahrscheinlichkeiten

$$w(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k \in R$$

heißt binomialverteilt oder kurz $B(n, p)$ -verteilt.

Satz 12: Eine Münze, bei der W (=Wappen) mit Wahrscheinlichkeit p und Z (=Zahl) mit Wahrscheinlichkeit $1-p$ fällt, wird n Mal geworfen. Sei X die Anzahl mit der W unter diesen n Würfeln auftritt. Dann hat X die $B(n, p)$ -Verteilung.

Beweis: Die möglichen Werte für die Anzahl, mit der W unter diesen n Würfeln auftritt, sind $0, 1, 2, \dots, n$. Der Wertebereich von X ist also $R = \{0, 1, 2, \dots, n\}$.

Wir berechnen $w(k) = P(X = k)$ für $k \in R$. Wir schreiben alle Folgen der Länge n auf, die k Mal W und $n - k$ Mal Z enthalten. Diese Folgen stellen die Teilereignisse dar, in die das Ereignis $X = k$ zerfällt. Nach dem Multiplikationssatz ist die Wahrscheinlichkeit jedes dieser Teilereignisse ein Produkt aus n Faktoren, das k Mal p und $n - k$ Mal $1 - p$ enthält, das heißt $p^k(1 - p)^{n-k}$. Nach dem Additionssatz ist die Wahrscheinlichkeit des Ereignisses $X = k$ die Summe der Wahrscheinlichkeiten der Teilereignisse, also $\binom{n}{k}p^k(1 - p)^{n-k}$, da $\binom{n}{k}$ die Anzahl der die Teilereignisse darstellenden Folgen ist. Damit ist $w(k) = P(X = k) = \binom{n}{k}p^k(1 - p)^{n-k}$ für $k \in R$ gezeigt. Die Zufallsvariable X ist $B(n, p)$ -verteilt. \square

Beispiel 10: Bei einem Glücksspiel ist 0.2 die Gewinnwahrscheinlichkeit. Wie oft muss man spielen, damit man mit Wahrscheinlichkeit 0.95 mindestens 2 Gewinne erzielt?

Sei n die zu bestimmende Anzahl der Spiele. Sei X die Anzahl der Gewinne bei diesen n Spielen. Nach Satz 12 hat X die $B(n, 0.2)$ -Verteilung. Es soll n so bestimmt werden, dass $P(X \geq 2) \geq 0.95$ gilt. Das ist äquivalent zu $P(X \leq 1) < 0.05$. Es gilt $P(X \leq 1) = \binom{n}{0}0.2^0 0.8^n + \binom{n}{1}0.2^1 0.8^{n-1}$. Gesucht ist n mit $0.8^n + n \cdot 0.2 \cdot 0.8^{n-1} < 0.05$. Die linke Seite dieser Ungleichung ist gleich 0.057 für $n = 21$, gleich 0.048 für $n = 22$, gleich 0.040 für $n = 23$ und gleich 0.033 für $n = 24$, wie man durch Probieren herausfindet. Somit genügen 22 Spiele, um mit Wahrscheinlichkeit 0.95 mindestens 2 Spiele zu gewinnen.

Wir führen noch eine weitere Verteilung ein, die mit Münzenwerfen zu tun hat. Jetzt werfen wir die Münze solange, bis zum ersten Mal Z fällt.

Definition: Sei $0 < p < 1$. Eine Zufallsvariable X mit Wertebereich $R = \{1, 2, 3, \dots\}$ und Einzelwahrscheinlichkeiten

$$w(k) = p^{k-1}(1 - p) \quad \text{für } k \in R$$

heißt geometrisch verteilt.

Satz 13: Eine Münze, bei der W mit Wahrscheinlichkeit p und Z mit Wahrscheinlichkeit $1 - p$ auftritt, wird solange geworfen, bis Z fällt. Sei X die Anzahl der dafür notwendigen Würfe. Dann ist X geometrisch verteilt mit Parameter p .

Beweis: Die Anzahl der notwendigen Würfe kann $1, 2, 3, \dots$ sein. Der Wertebereich von X ist daher $R = \{1, 2, 3, \dots\}$. Das Ereignis $X = k$ tritt genau dann ein, wenn die ersten $k - 1$ Würfe W und der k -te Wurf Z ergibt. Nach dem Multiplikationssatz hat dieses Ereignis die Wahrscheinlichkeit $p^{k-1}(1 - p)$. Somit gilt $w(k) = P(X = k) = p^{k-1}(1 - p)$ für $k \in R$ und X ist geometrisch verteilt mit Parameter p . \square

Beispiel 11: Man würfelt so lange, bis 6 kommt. Wie groß ist die Wahrscheinlichkeit, dass man höchstens 10 Mal würfeln muss?

Sei X die Anzahl der Würfe bis zum ersten Mal 6 auftritt. Gesucht ist $P(X \leq 10)$. Nach Satz 13 hat X die geometrische Verteilung mit $p = \frac{5}{6}$. Für die gesuchte Wahrscheinlichkeit erhalten wir $P(X \leq 10) = \sum_{k=1}^{10} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = 1 - \left(\frac{5}{6}\right)^{10}$.

3. Poissonverteilung

In diesem Kapitel geht es um Ereignisse, die zu zufälligen Zeitpunkten eintreten.

Definition: Sei $\lambda > 0$. Eine Zufallsvariable X mit Wertebereich $R = \{0, 1, 2, \dots\}$ und Einzelwahrscheinlichkeiten

$$w(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{für } k \in R$$

heißt Poissonverteilt oder kurz $P(\lambda)$ -verteilt.

Bei der Telefonauskunft treffen zu zufälligen Zeitpunkten Telefonanrufe ein. Wir legen einen Zeitpunkt als Nullpunkt fest und bezeichnen die Anzahl der Anrufe im Zeitintervall $[0, t]$ mit X . Da X eine Anzahl ist, ist $R = \{0, 1, 2, \dots\}$ der Wertebereich von X .

Um die Einzelwahrscheinlichkeiten von X zu berechnen, müssen wir einige Annahmen machen. Sei $s > t$ und n die Anzahl der Anrufe im Zeitintervall $[0, s]$. Wir nehmen an, dass jeder Anruf unabhängig von den anderen rein zufällig in einem Zeitpunkt im Intervall $[0, s]$ ankommt. Die Wahrscheinlichkeit, dass ein Anruf im Zeitintervall $[0, t]$ eintrifft, ist dann $\frac{t}{s}$ und die Wahrscheinlichkeit, dass er im Zeitintervall $(t, s]$ eintrifft, ist $\frac{s-t}{s} = 1 - \frac{t}{s}$ (entspricht einem Münzwurf, bei dem W mit Wahrscheinlichkeit $\frac{t}{s}$ auftritt). Da X die Anzahl der Anrufe im Zeitintervall $[0, t]$ ist, hat X die $B(n, \frac{t}{s})$ -Verteilung nach Satz 12. Wir halten $\mu = \frac{t}{s}$, die durchschnittliche Anzahl der Anrufe pro Zeiteinheit, fest und lassen n und damit auch s gegen ∞ gehen.

Satz 14: Die Zufallsvariable X sei $B(n, \frac{t}{s})$ -verteilt. Wenn n und s gegen ∞ gehen, sodass $\frac{t}{s} = \mu$ fest bleibt, dann hat X im Grenzwert eine $P(\mu t)$ -Verteilung.

Beweis: Wir berechnen $w(k) = P(X = k)$ für $k \geq 0$.

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{t}{s}\right)^k \left(1 - \frac{t}{s}\right)^{n-k} = \binom{n}{k} \left(\frac{\mu t}{n}\right)^k \left(1 - \frac{\mu t}{n}\right)^{n-k} \\ &= \frac{(\mu t)^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\mu t}{n}\right)^n \left(1 - \frac{\mu t}{n}\right)^{-k} \end{aligned}$$

Wegen $\lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-k+1)}{n^k} = 1$ und $\lim_{n \rightarrow \infty} \left(1 - \frac{\mu t}{n}\right)^n = e^{-\mu t}$ erhalten wir aus obiger Rechnung, dass $\lim_{n \rightarrow \infty} P(X = k) = \frac{(\mu t)^k}{k!} e^{-\mu t}$ gilt. Das aber besagt, dass X im Grenzwert die $P(\mu t)$ -Verteilung hat. \square

Beispiel 12: Die Anzahl der Anrufe in einem Zeitintervall von t Stunden sei $P(\mu t)$ -verteilt mit $\mu = 12$ Anrufen pro Stunde. Wie groß ist die Wahrscheinlichkeit, dass zwischen 16.00 und 16.15 höchstens ein Anruf kommt? Wie groß ist die Wahrscheinlichkeit, dass zwischen 16.00 und 16.15 mindestens drei Anrufe kommen?

Sei X die Anzahl der Anrufe zwischen 16.00 und 16.15, das heißt X ist $P(\lambda)$ -verteilt mit $\lambda = 12 \cdot \frac{1}{4} = 3$. Daher gilt

$$P(X \leq 1) = w(0) + w(1) = \frac{3^0}{0!} e^{-3} + \frac{3^1}{1!} e^{-3} = 4e^{-3} = 0.199$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - w(0) - w(1) - w(2) = 1 - \left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right) e^{-3} = 0.5768$$

Mit Wahrscheinlichkeit 0.199 kommt höchstens ein Anruf. Mit Wahrscheinlichkeit 0.577 kommen mindestens drei Anrufe.

Die Poissonverteilung wird für Ereignisse verwendet, die zu zufälligen Zeitpunkten eintreten, zum Beispiel für die Kunden, die ein Geschäft betreten, für die Defekte eines Gerätes, oder für die Schadensmeldungen, die bei einer Versicherung eintreffen.

4. Exponentialverteilung und Gammaverteilung

Jetzt kommen wir zu den kontinuierlichen Zufallsvariablen. In diesem Kapitel geht es um Wartezeiten.

Definition: Sei $\lambda > 0$. Eine kontinuierliche Zufallsvariable X mit Wertebereich \mathbb{R}^+ und Wahrscheinlichkeitsdichte

$$f(x) = \lambda e^{-\lambda x} \quad \text{für } x \in \mathbb{R}^+$$

heißt exponentialverteilt oder kurz $E(\lambda)$ -verteilt.

Satz 15: Die Anzahl der Anrufe im Zeitintervall $[0, t]$ sei $P(\lambda t)$ -verteilt, wobei λ die durchschnittliche Anzahl der Anrufe pro Zeiteinheit ist (siehe letztes Kapitel). Sei Y die Wartezeit vom Zeitpunkt 0 bis zum ersten Anruf. Dann hat Y die $E(\lambda)$ -Verteilung.

Beweis: Sei $t > 0$ beliebig und X die Anzahl der Anrufe im Zeitintervall $[0, t]$. Die Wartezeit auf den ersten Anruf ist genau dann $\leq t$, wenn im Zeitintervall $[0, t]$ mindestens ein Anruf kommt. Das heißt das Ereignis $Y \leq t$ tritt genau dann ein, wenn das Ereignis $X \geq 1$ eintritt. Daher gilt $F(t) = P(Y \leq t) = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda t}$. Eine Dichte f von Y erhält man als Ableitung von F , nämlich $f(x) = F'(x) = \lambda e^{-\lambda x}$ für $x > 0$. Die Wartezeit Y kann nicht negativ sein. Der Wertebereich der Zufallsvariablen Y ist daher \mathbb{R}^+ und $f(x) = 0$ für $x \in \mathbb{R}^-$. Die Wartezeit Y auf den ersten Anruf ist somit $E(\lambda)$ -verteilt. \square

Wir untersuchen auch noch die Wartezeit auf den n -ten Anruf. Dazu definieren wir

Definition: Seien $\lambda > 0$ und $r > 0$. Eine kontinuierliche Zufallsvariable X mit Wertebereich \mathbb{R}^+ und Wahrscheinlichkeitsdichte

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \quad \text{für } x \in \mathbb{R}^+$$

heißt gammaverteilt oder kurz $G(r, \lambda)$ -verteilt. Dabei wird $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy$ so definiert, dass $\int_0^\infty f(x) dx = 1$ gilt.

Für spezielle Werte von r kann man $\Gamma(r)$ explizit berechnen. Für $n \in \{1, 2, 3, \dots\}$ gilt $\Gamma(n) = (n-1)!$ und für $n \in \{0, 1, 2, \dots\}$ gilt $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{n! 2^{2n}} \sqrt{\pi}$.

Die $G(1, \lambda)$ -Verteilung ist die $E(\lambda)$ -Verteilung. Die $G(\frac{m}{2}, \frac{1}{2})$ -Verteilung spielt in der Statistik eine wichtige Rolle und heißt dort χ^2 -Verteilung.

Satz 16: Die Anzahl der Anrufe im Zeitintervall $[0, t]$ sei $P(\lambda t)$ -verteilt. Sei Z die Wartezeit vom Zeitpunkt 0 bis zum n -ten Anruf. Dann hat Z die $G(n, \lambda)$ -Verteilung.

Beweis: Sei $t > 0$ beliebig und X die Anzahl der Anrufe im Zeitintervall $[0, t]$. Die Wartezeit Z auf den n -ten Anruf ist genau dann $\leq t$, wenn im Zeitintervall $[0, t]$ mindestens n Anrufe kommen, das heißt $Z \leq t$ tritt genau dann ein, wenn $X \geq n$ eintritt. Daher gilt

$$F(t) = P(Z \leq t) = P(X \geq n) = 1 - P(X \leq n-1) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \dots - \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}$$

Eine Dichte f von Z erhält man als Ableitung von F . Berechnet man diese, so kürzen sich alle bis auf einen Summanden weg. Man erhält $f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$ für $x \geq 0$. Da Z als Wartezeit nur positive Werte annehmen kann, ist \mathbb{R}^+ der Wertebereich von Z und $f(x) = 0$ für $x < 0$. Die Zufallsvariable Z ist somit $G(n, \lambda)$ -verteilt. \square

5. Normalverteilung

Hat man es mit Größen zu tun, die zufällig um einen festen Wert schwanken, zum Beispiel mit Messfehlern behaftete Messwerte, dann verwendet man die Normalverteilung.

Definition: Sei $\mu \in \mathbb{R}$ und $\sigma > 0$. Eine kontinuierliche Zufallsvariable X mit Wertebereich \mathbb{R} und Wahrscheinlichkeitsdichte

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

heißt normalverteilt oder kurz $N(\mu, \sigma)$ -verteilt.

Satz 17: Sei X eine Zufallsvariable. Sei $Y = \frac{X-a}{b}$, wobei $a \in \mathbb{R}$ und $b > 0$ ist. Ist F die Verteilungsfunktion von X dann hat Y die Verteilungsfunktion $G(t) = F(bt+a)$. Hat X eine Wahrscheinlichkeitsdichte f , dann hat auch Y eine, nämlich $g(x) = bf(bx+a)$.

Wenn X die $N(\mu, \sigma)$ -Verteilung hat, dann hat $Y = \frac{X-\mu}{\sigma}$ die $N(0, 1)$ -Verteilung.

Beweis: Es gilt $P(Y \leq t) = P(\frac{X-a}{b} \leq t) = P(X \leq bt+a) = F(bt+a)$, sodass $G(t) = F(bt+a)$ die Verteilungsfunktion von Y ist. Hat X eine Wahrscheinlichkeitsdichte f , dann folgt aus deren Definition $G(t) = \int_{-\infty}^{bt+a} f(y) dy = \int_{-\infty}^t f(bx+a)bdx$, sodass $g(x) = bf(bx+a)$ eine Wahrscheinlichkeitsdichte von Y ist. Damit ist der erste Teil des Satzes bewiesen.

Für eine $N(\mu, \sigma)$ -verteilte Zufallsvariable X ist $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ eine Wahrscheinlichkeitsdichte. Nach dem ersten Teil des Satzes hat $Y = \frac{X-\mu}{\sigma}$ Wahrscheinlichkeitsdichte $g(x) = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-(\sigma x + \mu - \mu)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Somit hat Y die $N(0, 1)$ -Verteilung. \square

Um die Wahrscheinlichkeitsdichte der $N(0, 1)$ -Verteilung nicht immer ausschreiben zu müssen, wurde dafür eine Bezeichnung eingeführt. Man setzt $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Man kann zeigen, dass $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ ist, sodass $\int_{-\infty}^{\infty} \varphi(x) dx = 1$ gilt, woraus $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$ folgt. Weiters bezeichnet man die Verteilungsfunktion der $N(0, 1)$ -Verteilung mit Φ , das heißt $\Phi(t) = \int_{-\infty}^t \varphi(x) dx$ für alle $t \in \mathbb{R}$. Es gilt $\varphi(-x) = \varphi(x)$ für alle $x \in \mathbb{R}$. Daraus ergibt sich $\Phi(-t) = 1 - \Phi(t)$ für alle $t \in \mathbb{R}$.

6. Zufallsvektoren

Besteht der Ausfall eines Zufallsexperiments nicht aus einer, sondern aus mehreren Zahlen, dann verwendet man Zufallsvektoren. Wir geben ein Beispiel. Das Zufallsexperiment ist die Auswahl einer Person. Sei X_1 deren Körpergröße und X_2 deren Körpergewicht. Der Zufallsvektor $X = (X_1, X_2)$ beschreibt den uns interessierenden Ausfall des Zufallsexperiments.

Definition: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor. Die Funktion $F : \mathbb{R}^n \rightarrow [0, 1]$ definiert durch $F(t_1, t_2, \dots, t_n) = P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n)$ heißt Verteilungsfunktion des Zufallsvektors X .

Bemerkung: Die Beistriche in $P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n)$ bedeuten ein logisches und. Man sollte eigentlich \wedge schreiben. Jedoch ist diese Schreibweise mit Beistrichen üblich.

Die Eigenschaften der Verteilungsfunktion behandeln wir nur für $n = 2$. Analoge Resultate gelten auch für den Fall $n > 2$.

Satz 18: Sei $F : \mathbb{R}^2 \rightarrow [0, 1]$ die Verteilungsfunktion eines zweidimensionalen Zufallsvektors $X = (X_1, X_2)$. Sei $B = (r_1, s_1] \times (r_2, s_2]$ mit $r_1 < s_1$ und $r_2 < s_2$ ein Rechteck. Dann gilt $P(X \in B) = F(s_1, s_2) - F(s_1, r_2) - F(r_1, s_2) + F(r_1, r_2)$.

Beweis: Das Ereignis $\{X_1 \leq s_1, X_2 \leq s_2\}$ lässt sich in die beiden unvereinbaren Ereignisse $\{r_1 < X_1 \leq s_1, X_2 \leq s_2\}$ und $\{X_1 \leq r_1, X_2 \leq s_2\}$ zerlegen. Aus dem Additionssatz folgt $P(X_1 \leq s_1, X_2 \leq s_2) = P(r_1 < X_1 \leq s_1, X_2 \leq s_2) + P(X_1 \leq r_1, X_2 \leq s_2)$. Die Definition der Verteilungsfunktion ergibt dann $P(r_1 < X_1 \leq s_1, X_2 \leq s_2) = F(s_1, s_2) - F(r_1, s_2)$. Indem man s_2 durch r_2 ersetzt, hat man auch $P(r_1 < X_1 \leq s_1, X_2 \leq r_2) = F(s_1, r_2) - F(r_1, r_2)$.

Das Ereignis $\{r_1 < X_1 \leq s_1, X_2 \leq s_2\}$ zerlegt man in die beiden unvereinbaren Ereignisse $\{r_1 < X_1 \leq s_1, r_2 < X_2 \leq s_2\}$ und $\{r_1 < X_1 \leq s_1, X_2 \leq r_2\}$. Der Additionssatz ergibt

$P(r_1 < X_1 \leq s_1, X_2 \leq s_2) = P(r_1 < X_1 \leq s_1, r_2 < X_2 \leq s_2) + P(r_1 < X_1 \leq s_1, X_2 \leq r_2)$.
Daraus folgt $P(r_1 < X_1 \leq s_1, r_2 < X_2 \leq s_2) = F(s_1, s_2) - F(r_1, s_2) - F(s_1, r_2) + F(r_1, r_2)$
durch Einsetzen obiger Ergebnisse. Das ist das gewünschte Resultat. \square

Satz 19: Sei $F : \mathbb{R}^2 \rightarrow [0, 1]$ die Verteilungsfunktion eines Zufallsvektors $X = (X_1, X_2)$.

- (a) Für $r_1 < s_1$ und $r_2 < s_2$ gilt $F(s_1, s_2) - F(s_1, r_2) - F(r_1, s_2) + F(r_1, r_2) \geq 0$.
- (b) $\lim_{n \rightarrow \infty} F(t_1^{(n)}, t_2^{(n)}) = F(t_1, t_2)$ für alle Folgen $t_1^{(n)}$ und $t_2^{(n)}$ mit $t_1^{(n)} \downarrow t_1$ und $t_2^{(n)} \downarrow t_2$.
- (c) $\lim_{t_1 \rightarrow -\infty} F(t_1, t_2) = 0$, $\lim_{t_2 \rightarrow -\infty} F(t_1, t_2) = 0$ und $\lim_{t \rightarrow \infty} F(t, t) = 1$.

Beweis: Da eine Wahrscheinlichkeit immer nichtnegativ ist, erhalten wir (a) aus Satz 18.

Sind $t_1^{(n)}$ und $t_2^{(n)}$ monoton fallende Folgen in \mathbb{R} mit $t_1^{(n)} \downarrow t_1$ und $t_2^{(n)} \downarrow t_2$, dann bilden die Ereignisse $\{X_1 \leq t_1^{(n)}, X_2 \leq t_2^{(n)}\}$ eine monoton abnehmende Folge, die gegen das Ereignis $\{X_1 \leq t_1, X_2 \leq t_2\}$ geht. Es folgt $\lim_{n \rightarrow \infty} P(X_1 \leq t_1^{(n)}, X_2 \leq t_2^{(n)}) = P(X_1 \leq t_1, X_2 \leq t_2)$ aus dem Stetigkeitssatz. Damit ist (b) gezeigt. Setzt man $t_2^{(n)} = t_2$ für alle n und $t_1 = -\infty$, dann hat man $\lim_{n \rightarrow \infty} P(X_1 \leq t_1^{(n)}, X_2 \leq t_2) = P(X_1 \leq -\infty, X_2 \leq t_2) = 0$. Damit ist $\lim_{t_1 \rightarrow -\infty} P(X_1 \leq t_1, X_2 \leq t_2) = 0$ gezeigt. Setzt man $t_1^{(n)} = t_1$ für alle n und $t_2 = -\infty$, dann hat man $\lim_{n \rightarrow \infty} P(X_1 \leq t_1, X_2 \leq t_2^{(n)}) = P(X_1 \leq t_1, X_2 \leq -\infty) = 0$. Damit ist $\lim_{t_2 \rightarrow -\infty} P(X_1 \leq t_1, X_2 \leq t_2) = 0$ gezeigt. Das sind die ersten beiden Aussagen von (c).

Schließlich sei $t^{(n)}$ eine monoton wachsende Folge in \mathbb{R} , die gegen ∞ geht. Dann bilden die Ereignisse $\{X_1 \leq t^{(n)}, X_2 \leq t^{(n)}\}$ eine monoton aufsteigende Folge, die gegen das Ereignis $\{X_1 < \infty, X_2 < \infty\}$ geht, das immer eintritt. Es folgt $\lim_{n \rightarrow \infty} P(X_1 \leq t^{(n)}, X_2 \leq t^{(n)}) = 1$ aus dem Stetigkeitssatz. Damit ist $\lim_{t \rightarrow \infty} P(X_1 \leq t, X_2 \leq t) = 1$ gezeigt. Das ist die dritte Aussage von (c). \square

Definition: Ein Zufallsvektor $X = (X_1, X_2, \dots, X_n)$ heißt diskret, wenn die Zufallsvariablen X_1, X_2, \dots, X_n diskret sind, das heißt sie haben endliche oder abzählbare Wertebereiche R_1, R_2, \dots, R_n . Sei $S = R_1 \times R_2 \times \dots \times R_n$. Wir definieren die Einzelwahrscheinlichkeiten des Zufallsvektors X durch

$$u(k) = P(X = k) \quad \text{für } k \in S$$

In Koordinatenschreibweise heißt das $u(k_1, k_2, \dots, k_n) = P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n)$ für $k_1 \in R_1, k_2 \in R_2, \dots, k_n \in R_n$. Wir nennen S den Wertebereich des Zufallsvariables X . Es ist jedoch möglich, dass $u(k) = 0$ für manche $k \in S$ gilt.

Satz 20: Sei X ein diskreter Zufallsvektor, dessen Einzelwahrscheinlichkeiten durch $u(k)$ mit $k \in S$ gegeben sind.

- (a) Dann gilt $u(k) \geq 0$ für alle $k \in S$ und $\sum_{k \in S} u(k) = 1$.
- (b) Für $B \subset S$ gilt $P(X \in B) = \sum_{k \in B} u(k)$.

Beweis: Der Beweis ist derselbe wie für Satz 10. \square

Beispiel 13 (Multinomialverteilung) Seien $1, 2, \dots, n$ die Ausfälle eines Zufallsexperiments, die mit Wahrscheinlichkeiten p_1, p_2, \dots, p_n auftreten, wobei $p_j > 0$ für $1 \leq j \leq n$ und $p_1 + p_2 + \dots + p_n = 1$ gilt. Wir wiederholen das Zufallsexperiment m Mal. Für $1 \leq j \leq n$ sei X_j die Anzahl der Wiederholungen, bei denen der Ausfall j eintritt. Gesucht sind die Einzelwahrscheinlichkeiten des Zufallsvektors $X = (X_1, X_2, \dots, X_n)$.

Der Wertebereich der Zufallsvariablen X_j ist $R_j = R = \{0, 1, \dots, m\}$, das ist die Menge der Anzahlen, mit denen der Ausfall j bei den m Wiederholungen auftreten kann. Der Wertebereich des Zufallsvektors X ist dann $S = R^n$. Nun ist $X_1 + X_2 + \dots + X_n$ die Anzahl m aller durchgeführten Wiederholungen. Es muss daher $X_1 + X_2 + \dots + X_n = m$ gelten,

und wir erhalten $P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = 0$, wenn $k_1 + k_2 + \dots + k_n \neq m$ ist. Durch eine analoge Vorgangsweise wie im Beweis von Satz 12 folgt

$$P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{m!}{k_1!k_2!\dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n} \quad \text{wenn } k_1 + k_2 + \dots + k_n = m$$

Dabei ist $\frac{m!}{k_1!k_2!\dots k_n!}$ die Anzahl der Folgen der Länge m , die k_1 Mal 1, k_2 Mal 2, \dots , k_n Mal n enthalten, und aus dem Multiplikationssatz folgt, dass $p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$ die Wahrscheinlichkeit ist, mit der jede dieser Folgen auftritt. Eine Anwendung des Additionssatzes liefert dann das angegebene Resultat.

Nun kommen wir zu den Zufallsvektoren, die eine Wahrscheinlichkeitsdichte haben. Sie werden kontinuierliche Zufallsvektoren genannt.

Definition: Ein Zufallsvektor $X = (X_1, X_2, \dots, X_n)$ heißt kontinuierlich, wenn eine integrierbare Funktion $g : \mathbb{R}^n \rightarrow [0, \infty)$ existiert, sodass

$$P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_n} g(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1$$

für alle $t_1, t_2, \dots, t_n \in \mathbb{R}$ gilt. Die Funktion g nennt man Wahrscheinlichkeitsdichte des Zufallsvektors X .

Satz 21: Sei $X = (X_1, X_2, \dots, X_n)$ ein kontinuierlicher Zufallsvektor mit Wahrscheinlichkeitsdichte $g : \mathbb{R}^n \rightarrow [0, \infty)$.

(a) Dann gilt $\int_{\mathbb{R}^n} g(x) dx = 1$.

(b) Für integrierbare Teilmengen $B \subset \mathbb{R}^n$ gilt $P(X \in B) = \int_B g(x) dx$.

Beweis: Wir führen den Beweis nur für $n = 2$. Für $n > 2$ funktioniert der Beweis analog. Wir beginnen mit (b). Sei B zuerst ein Rechteck $C = (r_1, s_1] \times (r_2, s_2]$. Aus Satz 18 folgt $P(X \in C) = F(s_1, s_2) - F(s_1, r_2) - F(r_1, s_2) + F(r_1, r_2)$. Aus obiger Definition folgt nun $F(s_1, s_2) - F(s_1, r_2) = \int_{-\infty}^{s_1} \int_{-\infty}^{s_2} g(x_1, x_2) dx_2 dx_1 - \int_{-\infty}^{s_1} \int_{-\infty}^{r_2} g(x_1, x_2) dx_2 dx_1 = \int_{-\infty}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$. Ersetzt man in dieser Gleichung r_1 durch s_1 ein, so hat man auch $F(r_1, s_2) - F(r_1, r_2) = \int_{-\infty}^{r_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$. Setzt man in obige Gleichungen ein, so folgt $P(X \in C) = \int_{-\infty}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1 - \int_{-\infty}^{r_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1 = \int_{r_1}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$. Damit ist $P(X \in C) = \int_C g(x) dx$ gezeigt.

Sei $B = C_1 \cup C_2 \cup \dots \cup C_j$ eine disjunkte Vereinigung von endlich vielen solchen Rechtecken. Das Ereignis $X \in B$ tritt genau dann ein, wenn eines der Ereignisse $X \in C_1, \dots, X \in C_j$ eintritt. Wegen der Disjunktheit der Rechtecke C_i sind diese Ereignisse unvereinbar. Aus dem Additionssatz erhalten wir dann $P(X \in B) = P(X \in C_1) + P(X \in C_2) + \dots + P(X \in C_j)$. Da $P(X \in C_i) = \int_{C_i} g(x) dx$ für $1 \leq i \leq j$ bereits im letzten Absatz bewiesen wurde, haben wir auch $P(X \in B) = \int_{C_1} g(x) dx + \int_{C_2} g(x) dx + \dots + \int_{C_j} g(x) dx = \int_B g(x) dx$, wobei die letzte Gleichheit wegen der Disjunktheit der Rechtecke C_i folgt.

Sei B eine beschränkte integrierbare Menge. Sei $\varepsilon > 0$. Dann existieren Mengen B_1 und B_2 , die disjunkte Vereinigungen von endlich vielen Rechtecken wie oben sind, sodass $B_1 \subset B \subset B_2$ und $\int_{B_2} g(x) dx - \int_{B_1} g(x) dx < \varepsilon$ gilt. Wegen $B_1 \subset B \subset B_2$ und wegen Satz 2(c) erhalten wir $P(X \in B_1) \leq P(X \in B) \leq P(X \in B_2)$. Ebenso folgt $\int_{B_1} g(x) dx \leq \int_B g(x) dx \leq \int_{B_2} g(x) dx$, da ja $g \geq 0$ gilt. Wir haben $P(X \in B_1) = \int_{B_1} g(x) dx$ und $P(X \in B_2) = \int_{B_2} g(x) dx$ bereits oben bewiesen. Daher liegen die beiden Werte $P(X \in B)$ und $\int_B g(x) dx$ in einem Intervall, das Länge ε hat. Damit ist $|P(X \in B) - \int_B g(x) dx| < \varepsilon$ gezeigt. Da $\varepsilon > 0$ beliebig klein gewählt werden kann, muss auch $P(X \in B) = \int_B g(x) dx$ gelten.

Sei B schließlich eine unbeschränkte integrierbare Teilmenge des \mathbb{R}^2 . Für jedes $m \in \mathbb{N}$ sei $B_m = B \cap \{x \in \mathbb{R}^2 : \|x\| \leq m\}$. Da B_m eine beschränkte integrierbare Menge ist, wurde $P(X \in B_m) = \int_{B_m} g(x) dx$ bereits gezeigt. Wegen $B_1 \subset B_2 \subset \dots$ und $\bigcup_{m=1}^{\infty} B_m = B$ bilden die Ereignisse $X \in B_m$ eine aufsteigende Folge, die gegen das Ereignis $X \in B$ geht. Aus dem Stetigkeitssatz folgt $P(X \in B) = \lim_{m \rightarrow \infty} P(X \in B_m) = \lim_{m \rightarrow \infty} \int_{B_m} g(x) dx = \int_B g(x) dx$. Damit ist (b) vollständig bewiesen.

Da das Ereignis $X \in \mathbb{R}^2$ immer eintritt, haben wir $P(X \in \mathbb{R}^2) = 1$. Setzt man $B = \mathbb{R}^2$ in die Formel aus (b) ein, so erhält man (a). \square

Wir fragen nach dem Zusammenhang zwischen einem Zufallsvektor $X = (X_1, X_2, \dots, X_n)$ und den Zufallsvariablen X_j , aus denen er besteht.

Satz 22: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor.

(a) Sei $F : \mathbb{R}^n \rightarrow [0, 1]$ die Verteilungsfunktion von X . Die Verteilungsfunktion der Zufallsvariable X_j ist dann durch $F_j(t) = \lim_{m \rightarrow \infty} F(m, \dots, m, t, m, \dots, m)$ gegeben (wobei t in der j -ten Koordinate steht).

(b) Seien R_1, R_2, \dots, R_n die Wertebereiche der diskreten Zufallsvariablen X_1, X_2, \dots, X_n und $u(k_1, k_2, \dots, k_n)$ mit $k_1 \in R_1, k_2 \in R_2, \dots, k_n \in R_n$ die Einzelwahrscheinlichkeiten des Zufallsvektors X . Die Einzelwahrscheinlichkeiten der Zufallsvariable X_j sind dann durch $w_j(k) = \sum_{k_1 \in R_1} \dots \sum_{k_{j-1} \in R_{j-1}} \sum_{k_{j+1} \in R_{j+1}} \dots \sum_{k_n \in R_n} u(k_1, \dots, k_{j-1}, k, k_{j+1}, \dots, k_n)$ mit $k \in R_j$ gegeben.

(c) Sei $g : \mathbb{R}^n \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte des Zufallsvektors X . Dann ist durch $f_j(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) dx_n \dots dx_{j+1} dx_{j-1} \dots dx_1$ eine Wahrscheinlichkeitsdichte der Zufallsvariable X_j gegeben.

Beweis: Es gilt $P(X_j \leq t) = P(X_1 \in \mathbb{R}, \dots, X_{j-1} \in \mathbb{R}, X_j \leq t, X_{j+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R})$. Wir erhalten $P(X_j \leq t) = \lim_{m \rightarrow \infty} P(X_1 \leq m, \dots, X_{j-1} \leq m, X_j \leq t, X_{j+1} \leq m, \dots, X_n \leq m)$ aus dem Stetigkeitssatz. Das heißt $P(X_j \leq t) = \lim_{m \rightarrow \infty} F(m, \dots, m, t, m, \dots, m)$ und (a) ist gezeigt.

Es gilt $P(X_j = k) = P(X \in R_1 \times \dots \times R_{j-1} \times \{k\} \times R_{j+1} \times \dots \times R_n)$. Aus Satz 20(b) folgt $P(X_j = k) = \sum_{k_1 \in R_1} \dots \sum_{k_{j-1} \in R_{j-1}} \sum_{k_{j+1} \in R_{j+1}} \dots \sum_{k_n \in R_n} u(k_1, \dots, k_{j-1}, k, k_{j+1}, \dots, k_n)$. Damit ist (b) gezeigt.

Es gilt $P(X_j \leq t) = P(X \in \mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, t] \times \mathbb{R} \times \dots \times \mathbb{R})$. Aus Satz 21(b) folgt, dass $\int_{-\infty}^t \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) dx_n \dots dx_{j+1} dx_{j-1} \dots dx_1 dx$ gleich $P(X_j \leq t)$ ist, wobei auch die Integrationsreihenfolge vertauscht wurde. Die Funktion, die im Integral $\int_{-\infty}^t \dots dx$ steht, ist eine Wahrscheinlichkeitsdichte von X_j . Damit haben wir auch (c) gezeigt. \square

7. Unabhängigkeit von Zufallsvariablen

Nachdem wir früher schon die Unabhängigkeit von Ereignissen definiert haben, führen wir nun auch einen Unabhängigkeitsbegriff für Zufallsvariablen ein.

Definition: Die Zufallsvariablen X_1, X_2, \dots, X_n heißen unabhängig, wenn

$$P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) = P(X_1 \leq t_1)P(X_2 \leq t_2) \dots P(X_n \leq t_n)$$

für alle $t_1, t_2, \dots, t_n \in \mathbb{R}$ gilt. Ist F_j die Verteilungsfunktion der Zufallsvariablen X_j für $1 \leq j \leq n$, und F die Verteilungsfunktion des Zufallsvektors $X = (X_1, X_2, \dots, X_n)$ dann ist das gleichbedeutend mit $F(t_1, t_2, \dots, t_n) = F_1(t_1)F_2(t_2) \dots F_n(t_n)$.

Bemerkung: Setzt man in dieser Definition $t_1 = m \in \mathbb{N}$ und lässt m gegen ∞ gehen, dann geht das Ereignis $X_1 \leq m$ monoton aufsteigend gegen das Ereignis $X_1 < \infty$ und der Stetigkeitssatz liefert $P(X_2 \leq t_2, \dots, X_n \leq t_n) = P(X_2 \leq t_2) \dots P(X_n \leq t_n)$. Wendet man das wiederholt an, so folgt aus dieser Definition, dass

$$P(X_{j_1} \leq t_{j_1}, X_{j_2} \leq t_{j_2}, \dots, X_{j_k} \leq t_{j_k}) = P(X_{j_1} \leq t_{j_1})P(X_{j_2} \leq t_{j_2}) \dots P(X_{j_k} \leq t_{j_k})$$

für alle Teilmengen $\{j_1, j_2, \dots, j_k\}$ von $\{1, 2, \dots, n\}$ mit $k \geq 2$ gilt. Daraus erkennt man, dass obige Definition äquivalent dazu ist, dass die Ereignisse $X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n$ unabhängig sind für alle $t_1, t_2, \dots, t_n \in \mathbb{R}$.

Diese Bemerkung zeigt auch: Sind die Zufallsvariablen X_1, X_2, \dots, X_n unabhängig und ist $\{j_1, j_2, \dots, j_k\}$ eine Teilmenge von $\{1, 2, \dots, n\}$ mit $k \geq 2$, dann sind auch die Zufallsvariablen $X_{j_1}, X_{j_2}, \dots, X_{j_k}$ unabhängig.

Der nächste Satz zeigt, wie sich Einzelwahrscheinlichkeiten und Wahrscheinlichkeitsdichte des Zufallsvektors $X = (X_1, X_2, \dots, X_n)$ bei Vorliegen von Unabhängigkeit berechnen lassen.

Satz 23: Seien X_1, X_2, \dots, X_n unabhängige Zufallsvariable und $X = (X_1, X_2, \dots, X_n)$.

(a) Sind diese Zufallsvariablen diskret und R_j der Wertebereich und $w_j(k)$ mit $k \in R_j$ die Einzelwahrscheinlichkeiten von X_j für $1 \leq j \leq n$, dann sind durch

$$u(k_1, k_2, \dots, k_n) = w_1(k_1)w_2(k_2) \dots w_n(k_n) \quad \text{mit } k_1 \in R_1, k_2 \in R_2, \dots, k_n \in R_n$$

die Einzelwahrscheinlichkeiten des Zufallsvektors X gegeben.

(b) Sind die Zufallsvariablen X_1, X_2, \dots, X_n kontinuierlich mit Wahrscheinlichkeitsdichten f_1, f_2, \dots, f_n , dann ist durch

$$g(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \dots f_n(x_n) \quad \text{für } x_1 \in \mathbb{R}, x_2 \in \mathbb{R}, \dots, x_n \in \mathbb{R}$$

eine Wahrscheinlichkeitsdichte des Zufallsvektors X gegeben.

Beweis: Sei F die Verteilungsfunktion des Zufallsvektors $X = (X_1, X_2, \dots, X_n)$ und F_j die Verteilungsfunktion der Zufallsvariablen X_j für $1 \leq j \leq n$. Nach Voraussetzung gilt dann $F(t_1, t_2, \dots, t_n) = F_1(t_1)F_2(t_2) \dots F_n(t_n)$ für $t_1, t_2, \dots, t_n \in \mathbb{R}$.

Wir beweisen (a) nur für $n = 2$. Mit Hilfe des Stetigkeitssatzes und Satz 18 folgt

$$\begin{aligned} u(k_1, k_2) &= P(X_1 = k_1, X_2 = k_2) = \lim_{m \rightarrow \infty} P(k_1 - \frac{1}{m} < X_1 \leq k_1, k_2 - \frac{1}{m} < X_2 \leq k_2) \\ &= \lim_{m \rightarrow \infty} (F(k_1, k_2) - F(k_1 - \frac{1}{m}, k_2) - F(k_1, k_2 - \frac{1}{m}) + F(k_1 - \frac{1}{m}, k_2 - \frac{1}{m})) \\ &= \lim_{m \rightarrow \infty} (F_1(k_1)F_2(k_2) - F_1(k_1 - \frac{1}{m})F_2(k_2) - F_1(k_1)F_2(k_2 - \frac{1}{m}) + F_1(k_1 - \frac{1}{m})F_2(k_2 - \frac{1}{m})) \\ &= \lim_{m \rightarrow \infty} (F_1(k_1) - F_1(k_1 - \frac{1}{m}))(F_2(k_2) - F_2(k_2 - \frac{1}{m})) \\ &= \lim_{m \rightarrow \infty} P(k_1 - \frac{1}{m} < X_1 \leq k_1)P(k_2 - \frac{1}{m} < X_2 \leq k_2) \\ &= P(X_1 = k_1)P(X_2 = k_2) \\ &= w_1(k_1)w_2(k_2) \end{aligned}$$

Damit ist (a) für $n = 2$ bewiesen. Ein analoger Beweis gilt auch für größere n .

Wir erhalten (b) mit Hilfe der Definition einer Wahrscheinlichkeitsdichte.

$$\begin{aligned} F(t_1, t_2, \dots, t_n) &= F_1(t_1)F_2(t_2) \dots F_n(t_n) = \int_{-\infty}^{t_1} f_1(x_1) dx_1 \int_{-\infty}^{t_2} f_2(x_2) dx_2 \dots \int_{-\infty}^{t_n} f_n(x_n) dx_n \\ &= \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_n} f_1(x_1)f_2(x_2) \dots f_n(x_n) dx_n \dots dx_2 dx_1 \end{aligned}$$

Die Funktion in diesem n -fachen Integral ist nach Definition eine Wahrscheinlichkeitsdichte des Zufallsvektors X . Damit ist (b) gezeigt. \square

Beispiel 14: Jemand hat zwei voneinander unabhängige Telefone. Bei Telefon A ist die Wartezeit auf den nächsten Anruf $E(\lambda)$ -verteilt, bei Telefon B ist sie $E(\mu)$ -verteilt. Wie groß ist die Wahrscheinlichkeit, dass Telefon A vor Telefon B läutet?

Sie X_1 die Wartezeit auf den nächsten Anruf bei Telefon A und X_2 die Wartezeit auf den nächsten Anruf bei Telefon B. Nach Satz 23 (b) ist $g(x, y) = \lambda e^{-\lambda x} \mu e^{-\mu y}$ für $(x, y) \in \mathbb{R}_+^2$ und $= 0$ für $(x, y) \notin \mathbb{R}_+^2$ eine Dichte des Zufallsvektors (X_1, X_2) , da X_1 und X_2 unabhängig sind. Sei $M = \{(x, y) \in \mathbb{R}_+^2 : x < y\}$. Mit Satz 21 (b) folgt $P(X_1 < X_2) = P((X_1, X_2) \in M) = \int_M g(x, y) d(x, y) = \int_0^\infty \int_x^\infty \lambda e^{-\lambda x} \mu e^{-\mu y} dy dx = \int_0^\infty \lambda e^{-\lambda x} e^{-\mu x} dx = \frac{\lambda}{\lambda + \mu}$.

Es gilt auch eine Umkehrung des letzten Satzes.

Satz 24: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor.

(a) Sei $F : \mathbb{R}^n \rightarrow [0, 1]$ die Verteilungsfunktion des Zufallsvektors X . Wenn $F(t_1, t_2, \dots, t_n)$ in ein Produkt $F_1(t_1)F_2(t_2) \dots F_n(t_n)$ zerfällt, sodass $\lim_{s \rightarrow \infty} F_j(s) = 1$ für $1 \leq j \leq n$ gilt, dann ist F_j die Verteilungsfunktion der Zufallsvariablen X_j für $1 \leq j \leq n$ und die Zufallsvariablen X_1, X_2, \dots, X_n sind unabhängig.

(b) Seien $u(k_1, k_2, \dots, k_n)$ mit $k_1 \in R_1, k_2 \in R_2, \dots, k_n \in R_n$ die Einzelwahrscheinlichkeiten des Zufallsvektors X . Wenn $u(k_1, k_2, \dots, k_n)$ in ein Produkt $w_1(k_1)w_2(k_2) \dots w_n(k_n)$ zerfällt mit $\sum_{i \in R_j} w_j(i) = 1$ für $1 \leq j \leq n$, dann sind $w_j(k)$ mit $k \in R_j$ die Einzelwahrscheinlichkeiten der Zufallsvariable X_j für $1 \leq j \leq n$ und die Zufallsvariablen X_1, X_2, \dots, X_n sind unabhängig.

(c) Sei $g : \mathbb{R}^n \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte von X . Wenn $g(x_1, x_2, \dots, x_n)$ in ein Produkt $f_1(x_1)f_2(x_2) \dots f_n(x_n)$ zerfällt mit $\int_{-\infty}^\infty f_j(y) dy = 1$ für $1 \leq j \leq n$, dann ist f_j eine Wahrscheinlichkeitsdichte der Zufallsvariablen X_j für $1 \leq j \leq n$ und die Zufallsvariablen X_1, X_2, \dots, X_n sind unabhängig.

Beweis: Wir zeigen (a). Nach Satz 22(a) ist $\lim_{m \rightarrow \infty} F(m, \dots, m, t, m, \dots, m)$ die Verteilungsfunktion der Zufallsvariablen X_j , wobei t in der j -ten Koordinate steht. Nach Voraussetzung gilt $F(m, \dots, m, t, m, \dots, m) = F_1(m) \dots F_{j-1}(m)F_j(t)F_{j+1}(m) \dots F_n(m)$ und $\lim_{m \rightarrow \infty} F_i(m) = 1$ für $1 \leq i \leq n$. Es folgt $\lim_{m \rightarrow \infty} F(m, \dots, m, t, m, \dots, m) = F_j(t)$, womit gezeigt ist, dass F_j die Verteilungsfunktion von X_j ist. Die Unabhängigkeit der Zufallsvariablen X_1, X_2, \dots, X_n folgt dann direkt aus der Definition.

Ähnlich beweisen wir (b). Aus den Voraussetzungen folgt für $1 \leq j \leq n$ und $k \in R_j$, dass

$$\sum_{k_1 \in R_1} \dots \sum_{k_{j-1} \in R_{j-1}} \sum_{k_{j+1} \in R_{j+1}} \dots \sum_{k_n \in R_n} u(k_1, \dots, k_{j-1}, k, k_{j+1}, \dots, k_n) = w_j(k)$$

gilt. Wegen Satz 22(b) sind dann durch $w_j(k)$ mit $k \in R_j$ die Einzelwahrscheinlichkeiten der Zufallsvariable X_j gegeben. Mit Hilfe von Satz 20(b) erhalten wir nun

$$\begin{aligned} P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) &= \sum_{k_1 \leq t_1} \dots \sum_{k_n \leq t_n} u(k_1, \dots, k_n) \\ &= \sum_{k_1 \leq t_1} w_1(k_1) \dots \sum_{k_n \leq t_n} w_n(k_n) = P(X_1 \leq t_1)P(X_2 \leq t_2) \dots P(X_n \leq t_n) \end{aligned}$$

womit die Unabhängigkeit der Zufallsvariablen X_1, X_2, \dots, X_n gezeigt ist.

Schließlich kommen wir zu (c). Aus den Voraussetzungen folgt für $1 \leq j \leq n$ und $x \in \mathbb{R}$, dass $\int_{-\infty}^\infty \dots \int_{-\infty}^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty g(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) dx_n \dots dx_{j+1} dx_{j-1} \dots dx_1 = f_j(x)$ gilt. Wegen Satz 22(c) ist dann f_j eine Wahrscheinlichkeitsdichte der Zufallsvariable X_j . Mit Hilfe der Definition einer Wahrscheinlichkeitsdichte folgt nun

$$\begin{aligned} P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) &= \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} g(x_1, \dots, x_n) dx_n \dots dx_1 \\ &= \int_{-\infty}^{t_1} f_1(x_1) dx_1 \dots \int_{-\infty}^{t_n} f_n(x_n) dx_n = P(X_1 \leq t_1)P(X_2 \leq t_2) \dots P(X_n \leq t_n) \end{aligned}$$

womit die Unabhängigkeit der Zufallsvariablen X_1, X_2, \dots, X_n gezeigt ist. \square

Aus Zufallsvariablen kann man neue Zufallsvariable berechnen. Sind X_1 und X_2 Zufallsvariable, dann ist auch $X_2 - X_1$ oder $X_1^2 + X_2^2$ eine. Ist allgemeiner $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine Funktion (die wir üblicherweise als stetig annehmen), dann ist auch $\psi(X_1, X_2)$ eine Zufallsvariable. Der nächste Satz behandelt die Unabhängigkeit von solchen Funktionen von Zufallsvariablen.

Satz 25: Die Zufallsvariablen X_1, X_2, \dots, X_n seien unabhängig. Sei $\psi : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$ stetig und $Y = \psi(X_{k+1}, X_{k+2}, \dots, X_n)$. Dann sind auch die Zufallsvariablen X_1, X_2, \dots, X_k, Y unabhängig. (Dieser Satz gilt für viel allgemeinere Funktionen ψ als nur für stetige. Er gilt für messbare Funktionen, die man in der Maßtheorie kennenlernt.)

Beweis: Wir führen den Beweis nur für den Fall, dass die Zufallsvariablen X_1, X_2, \dots, X_n Wahrscheinlichkeitsdichten f_1, f_2, \dots, f_n haben. Der allgemeine Beweis geht genauso, wobei man das Riemann–Stieltjes–Integral verwendet. Wegen Satz 23(b) hat der Zufallsvektor (X_1, X_2, \dots, X_n) Wahrscheinlichkeitsdichte $g(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n)$. Da nach einer früheren Bemerkung auch die Zufallsvariablen $X_{k+1}, X_{k+2}, \dots, X_n$ unabhängig sind, hat der Zufallsvektor $(X_{k+1}, X_{k+2}, \dots, X_n)$ wieder nach Satz 23(b) Wahrscheinlichkeitsdichte $h(x_{k+1}, x_{k+2}, \dots, x_n) = f_{k+1}(x_{k+1})f_{k+2}(x_{k+2})\dots f_n(x_n)$. Seien t_1, t_2, \dots, t_k, t in \mathbb{R} beliebig. Sei $C_t = \psi^{-1}((-\infty, t]) \subset \mathbb{R}^{n-k}$ und $B = (-\infty, t_1] \times (-\infty, t_2] \times \dots \times (-\infty, t_k] \times C_t$. Wir rechnen die Definition der Unabhängigkeit für die Zufallsvariablen X_1, \dots, X_k, Y nach.

$$\begin{aligned} & P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k, Y \leq t) \\ &= P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_k \leq t_k, (X_{k+1}, \dots, X_n) \in \psi^{-1}(-\infty, t]) \\ &= P((X_1, X_2, \dots, X_n) \in B) \\ &= \int_B f_1(x_1)f_2(x_2)\dots f_n(x_n) d(x_1, x_2, \dots, x_n) \quad \text{nach Satz 21(b)} \\ &= \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_k} \int_{C_t} f_1(x_1)f_2(x_2)\dots f_n(x_n) d(x_{k+1}, \dots, x_n) dx_k \dots dx_1 \\ &= \int_{-\infty}^{t_1} f_1(x_1) dx_1 \dots \int_{-\infty}^{t_k} f_k(x_k) dx_k \int_{C_t} f_{k+1}(x_{k+1}) \dots f_n(x_n) d(x_{k+1}, \dots, x_n) \\ &= P(X_1 \leq t_1)P(X_2 \leq t_2) \dots P(X_k \leq t_k)P((X_{k+1}, \dots, X_n) \in C_t) \quad \text{nach Satz 21(b)} \\ &= P(X_1 \leq t_1)P(X_2 \leq t_2) \dots P(X_k \leq t_k)P(Y \leq t) \end{aligned}$$

Damit ist die Unabhängigkeit der Zufallsvariablen X_1, X_2, \dots, X_k, Y gezeigt. \square

Bemerkung: Durch wiederholtes Anwenden von Satz 25 erhält man folgende Resultate:

- Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariablen und $\psi_1, \psi_2, \dots, \psi_n$ stetige Funktionen von \mathbb{R} nach \mathbb{R} , dann sind die Zufallsvariablen $\psi_1(X_1), \psi_2(X_2), \dots, \psi_n(X_n)$ ebenfalls unabhängig.
- Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariablen und $\psi_1 : \mathbb{R}^k \rightarrow \mathbb{R}$ und $\psi_2 : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$ stetige Funktionen, dann sind auch $\psi_1(X_1, X_2, \dots, X_k)$ und $\psi_2(X_{k+1}, X_{k+2}, \dots, X_n)$ unabhängig.
- Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariablen, ist $0 = k_0 < k_1 < k_2 < \dots < k_r = n$ und sind $\psi_m : \mathbb{R}^{k_m - k_{m-1}} \rightarrow \mathbb{R}$ für $1 \leq m \leq r$ stetige Funktionen, dann sind die Zufallsvariablen $\psi_1(X_1, \dots, X_{k_1}), \psi_2(X_{k_1+1}, \dots, X_{k_2}), \dots, \psi_r(X_{k_{r-1}+1}, \dots, X_{k_r})$ ebenfalls unabhängig.

8. Funktionen von Zufallsvariablen

Bereits am Ende des letzten Kapitels sind Funktionen von Zufallsvariablen vorgekommen, die wir noch verallgemeinern. Sei $X = (X_1, X_2, \dots, X_n)$ ein n -dimensionaler Zufallsvektor und $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine geeignete (stetig, differenzierbar) Abbildung. Dann ist $\psi(X)$ ein m -dimensionaler Zufallsvektor. Ist zum Beispiel $X = (X_1, X_2, X_3, X_4)$ der Ausfall bei viermaligem Würfeln und $\psi(x_1, x_2, x_3, x_4) = (\max(x_1, x_2, x_3, x_4), \min(x_1, x_2, x_3, x_4))$, dann gibt

$\psi(X)$ die größte und die kleinste erzielte Augenzahl an. Wir stellen uns die Aufgabe, aus der Verteilung von X die Verteilung von $\psi(X)$ zu berechnen.

Satz 26: Sei X eine kontinuierliche Zufallsvariable und $f : \mathbb{R} \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte von X . Dann hat die Zufallsvariable X^2 Wahrscheinlichkeitsdichte h , die definiert ist durch $h(x) = 0$ für $x \leq 0$ und $h(x) = \frac{1}{2\sqrt{x}}(f(\sqrt{x}) + f(-\sqrt{x}))$ für $x > 0$.

Beweis: Sei H die Verteilungsfunktion der Zufallsvariablen X^2 . Da X^2 nicht negativ sein kann, erhalten wir $H(t) = P(X^2 \leq t) = 0$ für $t < 0$. Für $t > 0$ gilt

$$H(t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} f(y) dy = \int_0^{\sqrt{t}} f(y) dy + \int_{-\sqrt{t}}^0 f(y) dy$$

Variablentransformationen $y = \sqrt{x}$ und $y = -\sqrt{x}$ in den letzten beiden Integralen liefern

$$H(t) = \int_0^t f(\sqrt{x}) \frac{1}{2\sqrt{x}} dx + \int_t^0 f(-\sqrt{x}) \frac{-1}{2\sqrt{x}} dx = \int_0^t (f(\sqrt{x}) + f(-\sqrt{x})) \frac{1}{2\sqrt{x}} dx$$

Wir haben somit gezeigt, dass $H(t) = \int_{-\infty}^t h(x) dx$ gilt, wenn man $h(x) = 0$ für $x \leq 0$ und $h(x) = (f(\sqrt{x}) + f(-\sqrt{x})) \frac{1}{2\sqrt{x}}$ für $x > 0$ setzt. Daraus folgt, dass diese Funktion h eine Wahrscheinlichkeitsdichte von X^2 ist. \square

Beispiel 15: Sei X eine $N(0, 1)$ -verteilte Zufallsvariable. Gesucht ist die Verteilung von X^2 .

Nach Voraussetzung hat die Zufallsvariable X Wahrscheinlichkeitsdichte $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Nach Satz 26 hat dann X^2 Wahrscheinlichkeitsdichte h mit $h(x) = 0$ für $x < 0$ und

$$h(x) = \frac{1}{2\sqrt{x}} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \right) = \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} = \frac{(\frac{1}{2})^{1/2}}{\sqrt{\pi}} x^{\frac{1}{2}-1} e^{-\frac{1}{2}x}$$

für $x > 0$. Man erkennt, dass X^2 die $G(\frac{1}{2}, \frac{1}{2})$ -Verteilung hat.

Nun beweisen wir ein allgemeines Resultat für n -dimensionale Zufallsvektoren und invertierbare Funktionen ψ . Für eine Menge A bezeichne 1_A die Indikatorfunktion, das heißt $1_A(x) = 1$ für $x \in A$ und $1_A(x) = 0$ für $x \notin A$.

Satz 27: Sei $U \subset \mathbb{R}^n$ offen und $\psi : U \rightarrow \mathbb{R}^n$ eine injektive stetig differenzierbare Abbildung mit $\det D\psi(x) \neq 0$ für alle $x \in U$. Sei X ein n -dimensionaler Zufallsvektor, dessen Werte in U liegen und der Wahrscheinlichkeitsdichte $g : \mathbb{R}^n \rightarrow [0, \infty)$ hat. Dann ist

$$h(x) = 1_{\psi(U)}(x) g(\psi^{-1}(x)) |\det D\psi^{-1}(x)|$$

eine Wahrscheinlichkeitsdichte des n -dimensionalen Zufallsvektors Y .

Beweis: Für eine integrierbare Teilmenge C von \mathbb{R}^n gilt $\int_{C \setminus U} g(x) dx = P(X \in C \setminus U) = 0$, da X nur Werte in U annimmt, woraus $\int_C g(x) dx = \int_{C \cap U} g(x) dx$ folgt.

Sei $B = (-\infty, t_1] \times (-\infty, t_2] \times \dots \times (-\infty, t_n]$. Mit Hilfe von Satz 21 (b) und da $g(y) = 0$ für $y \notin U$ vorausgesetzt wird, erhalten wir

$$P(Y \in B) = P(\psi(X) \in B) = P(X \in \psi^{-1}(B)) = \int_{\psi^{-1}(B)} g(y) dy = \int_{\psi^{-1}(B) \cap U} g(y) dy$$

Wir verwenden die mehrdimensionale Transformationsformel für Integrale und führen die neue Integrationsvariable $x = \psi(y)$ ein. Da y durch die Menge $\psi^{-1}(B) \cap U$ läuft, läuft x durch die Menge $\psi(\psi^{-1}(B) \cap U) = B \cap \psi(U)$. Wegen $y = \psi^{-1}(x)$ gilt $dy = |\det D\psi^{-1}(x)| dx$. Da wir $\det D\psi(y) \neq 0$ für alle $y \in U$ voraussetzen, existiert ja $\det D\psi^{-1}(x) = \frac{1}{\det D\psi(\psi^{-1}(x))}$ für alle $x \in \psi(U)$. Wir erhalten also

$$P(Y \in B) = \int_{B \cap \psi(U)} g(\psi^{-1}(x)) |\det D\psi^{-1}(x)| dx = \int_B 1_{\psi(U)}(x) g(\psi^{-1}(x)) |\det D\psi^{-1}(x)| dx$$

Es folgt, dass die Funktion im Integral $\int_B \dots dx$ eine Wahrscheinlichkeitsdichte des Zufallsvektors Y ist, und der Satz ist bewiesen. \square

Folgerungen: Wir wenden Satz 27 für verschiedene Funktionen ψ an.

(a) Hat die Zufallsvariable X den Wertebereich \mathbb{R}^+ und Wahrscheinlichkeitsdichte f , dann hat \sqrt{X} Dichte h definiert durch $h(x) = 2xf(x^2)$ für $x \geq 0$ und $h(x) = 0$ für $x < 0$.

Das folgt mit $\psi(x) = \sqrt{x}$, $U = \psi(U) = \mathbb{R}^+$, $\psi^{-1}(x) = x^2$ und $D\psi^{-1}(x) = 2x$.

(b) Seien a und b in \mathbb{R} , wobei $a \neq 0$ ist. Wenn die Zufallsvariable X Wahrscheinlichkeitsdichte f hat, dann hat $aX + b$ Wahrscheinlichkeitsdichte $h(x) = \frac{1}{|a|}f\left(\frac{x-b}{a}\right)$ für $x \in \mathbb{R}$.

Das folgt mit $\psi(x) = ax + b$, $U = \psi(U) = \mathbb{R}$, $\psi^{-1}(x) = \frac{x-b}{a}$ und $D\psi^{-1}(x) = \frac{1}{a}$.

(c) Ist X eine Zufallsvariable, die den Wert 0 nicht annimmt und Wahrscheinlichkeitsdichte f hat, dann hat $\frac{1}{X}$ Wahrscheinlichkeitsdichte $h(x) = \frac{1}{x^2}f\left(\frac{1}{x}\right)$ für $x \in \mathbb{R} \setminus \{0\}$.

Das folgt mit $\psi(x) = \frac{1}{x}$, $U = \psi(U) = \mathbb{R} \setminus \{0\}$, $\psi^{-1}(x) = \frac{1}{x}$ und $D\psi^{-1}(x) = -\frac{1}{x^2}$.

(d) Sei A eine invertierbare $n \times n$ -Matrix. Hat der n -dimensionale Zufallsvektor X Wahrscheinlichkeitsdichte $f : \mathbb{R}^n \rightarrow [0, \infty)$, dann hat der Zufallsvektor AX Wahrscheinlichkeitsdichte $h(x) = f(A^{-1}x)|\det A^{-1}|$ für $x \in \mathbb{R}^n$.

Das folgt mit $\psi(x) = Ax$, $U = \psi(U) = \mathbb{R}^n$, $\psi^{-1}(x) = A^{-1}x$ und $D\psi^{-1}(x) = A^{-1}$.

(e) Hat $X = (X_1, X_2)$ Wahrscheinlichkeitsdichte $f : \mathbb{R}^2 \rightarrow [0, \infty)$, dann hat die Zufallsvariable $X_1 + X_2$ Wahrscheinlichkeitsdichte $g(x) = \int_{-\infty}^{\infty} f(x-y, y) dy$ für $x \in \mathbb{R}$.

Sei $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Es folgt $A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ und AX hat Dichte $h(x, y) = f(x-y, y)$ nach (d) wegen $A^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x-y \\ y \end{pmatrix}$ und $\det A^{-1} = 1$. Da $X_1 + X_2$ die erste Komponente von AX ist, hat $X_1 + X_2$ Wahrscheinlichkeitsdichte $g(x) = \int_{-\infty}^{\infty} h(x, y) dy$ nach Satz 22 (c).

(f) Hat $X = (X_1, X_2)$ Wahrscheinlichkeitsdichte $f : \mathbb{R}^2 \rightarrow [0, \infty)$, dann hat die Zufallsvariable $\frac{X_1}{X_2}$ Wahrscheinlichkeitsdichte $g(x) = \int_{-\infty}^{\infty} f(xy, y) \cdot |y| dy$ für $x \in \mathbb{R}$.

Sei $U = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R} \setminus \{0\}\}$ und $\psi(x, y) = \left(\frac{x}{y}, y\right)$. Es folgt $\psi^{-1}(x, y) = (xy, y)$, $D\psi^{-1}(x, y) = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}$ und $\det D\psi^{-1}(x, y) = y$. Eine Wahrscheinlichkeitsdichte von $\left(\frac{X_1}{X_2}, X_2\right)$ ist daher $h(x, y) = f(xy, y) \cdot |y|$. Nach Satz 22 (c) ist $g(x) = \int_{-\infty}^{\infty} h(x, y) dy$ eine Wahrscheinlichkeitsdichte von $\frac{X_1}{X_2}$.

Das zu Satz 27 analoge Resultat für diskrete Zufallsvektoren ist sehr einfach.

Satz 28: Der Zufallsvektor X habe Wertebereich $S = R_1 \times R_2 \times \cdots \times R_n$ und Einzelwahrscheinlichkeiten $u(k)$ mit $k \in S$. Sei $\psi : S \rightarrow \mathbb{R}^m$ eine Abbildung und $Y = \psi(X)$. Dann hat Y den Wertebereich $\psi(S)$ und Einzelwahrscheinlichkeiten $v(i) = \sum_{k \in \psi^{-1}(i)} u(k)$ mit $i \in \psi(S)$.

Beweis: Es gilt $P(Y = i) = P(\psi(X) = i) = P(X \in \psi^{-1}(i)) = \sum_{k \in \psi^{-1}(i)} u(k)$ für alle $i \in \psi(S)$, wobei Satz 20(b) verwendet wurde. \square

9. Mehrdimensionale Normalverteilung

Wir definieren die mehrdimensionale Normalverteilung. Als Anwendungen der Sätze aus den letzten Kapiteln beweisen wir einige Eigenschaften.

Definition: Sei M eine symmetrische positiv definite $n \times n$ -Matrix. Ein n -dimensionaler Zufallsvektor mit Wahrscheinlichkeitsdichte

$$g(x) = \frac{1}{\sqrt{(2\pi)^n \det M}} e^{-\frac{1}{2}x^t M^{-1}x} \quad \text{für } x \in \mathbb{R}^n$$

heißt zentriert normalverteilt mit Kovarianzmatrix M oder kurz $ZN(M)$ -verteilt.

Satz 29: Sei A eine invertierbare $n \times n$ -Matrix.

(a) Sei X ein n -dimensionaler $ZN(M)$ -verteilter Zufallsvektor. Sei $Y = AX$. Dann ist Y

ein $ZN(L)$ -verteilter Zufallsvektor mit $L = AMA^t$.

(b) Seien X_1, X_2, \dots, X_n unabhängig und alle $N(0, 1)$ -verteilt und sei $X = (X_1, X_2, \dots, X_n)$. Sei A orthogonal, das heißt $A^{-1} = A^t$, und sei $Y = AX$. Die Komponenten Y_1, Y_2, \dots, Y_n von Y sind dann ebenfalls unabhängig und alle $N(0, 1)$ -verteilt.

Beweis: Nach Definition ist $g(x) = \frac{1}{\sqrt{(2\pi)^n \det M}} e^{-\frac{1}{2}x^t M^{-1}x}$ für $x \in \mathbb{R}^n$ eine Dichte des Zufallsvektors X in (a). Weiters hat wegen Satz 27 der Zufallsvektor $Y = AX$ Dichte $h(x) = \frac{1}{\sqrt{(2\pi)^n \det M}} e^{-\frac{1}{2}(A^{-1}x)^t M^{-1}A^{-1}x} |\det A^{-1}| = \sqrt{\frac{(\det A^{-1})^2}{(2\pi)^n \det M}} e^{-\frac{1}{2}x^t (A^{-1})^t M^{-1}A^{-1}x}$. Für die Matrix $L = AMA^t$ gilt $\frac{\det M}{(\det A^{-1})^2} = \det M (\det A)^2 = \det A \det M \det A^t = \det L$ und $L^{-1} = (A^t)^{-1} M^{-1} A^{-1} = (A^{-1})^t M^{-1} A^{-1}$. Es folgt $h(x) = \frac{1}{\sqrt{(2\pi)^n \det L}} e^{-\frac{1}{2}x^t L^{-1}x}$ und wir haben gezeigt, dass Y die $ZN(L)$ -Verteilung hat.

Mit Hilfe von Satz 23(b) folgt, dass der Zufallsvektor X in (b) Wahrscheinlichkeitsdichte $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2} \dots \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_n^2} = \frac{1}{\sqrt{2\pi^n}} e^{-\frac{1}{2}\|x\|^2} = \frac{1}{\sqrt{2\pi^n}} e^{-\frac{1}{2}x^t I x}$ hat, das heißt X hat die $ZN(I)$ -Verteilung, wobei I die Einheitsmatrix ist. Aus (a) folgt jetzt, dass $Y = AX$ die $ZN(L)$ -Verteilung hat mit $L = AIA^t = AA^t = AA^{-1} = I$. Somit hat auch der Zufallsvektor Y Wahrscheinlichkeitsdichte g . Da $g(x)$ in ein Produkt zerfällt, sind nach Satz 24(c) die Zufallsvariablen Y_1, Y_2, \dots, Y_n unabhängig und alle $N(0, 1)$ -verteilt. \square

10. Verteilungen in der Statistik

In der Statistik beschreibt man eine Stichprobe durch Zufallsvariable X_1, X_2, \dots, X_n , die meistens als unabhängig und oft auch als normalverteilt angenommen werden. In verschiedenen statistischen Verfahren spielt die Summe der Quadrate eine wichtige Rolle. Daher ist es notwendig deren Verteilung zu kennen.

Satz 30: Die Zufallsvariablen Y und Z seien unabhängig, Y sei $G(r, \lambda)$ -verteilt und Z sei $G(s, \lambda)$ -verteilt. Dann hat $Y + Z$ die $G(r + s, \lambda)$ -Verteilung.

Beweis: Die Dichte der $G(r, \lambda)$ -Verteilung ist $f_{\lambda,r}(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$ für $x \in \mathbb{R}^+$ und $= 0$ für $x \in \mathbb{R}^-$. Da Y und Z unabhängig sind, ist $(x, y) \mapsto f_{\lambda,r}(x) f_{\lambda,s}(y)$ eine Dichte des Zufallsvektors (Y, Z) . Als Dichte g für $Y + Z$ ergibt sich $g(x) = \int_0^x f_{\lambda,r}(x-y) f_{\lambda,s}(y) dy$ für $x \in \mathbb{R}^+$ und $g(x) = 0$ für $x \in \mathbb{R}^-$ nach Folgerung (e) von Satz 27. Man erhält

$$g(x) = \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} \int_0^x (x-y)^{r-1} y^{s-1} dy$$

indem man für $f_{\lambda,r}$ und $f_{\lambda,s}$ einsetzt. Durch die Substitution $y = xz$ ergibt sich daraus

$$g(x) = \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} x^{r+s-1} \int_0^1 (1-z)^{r-1} z^{s-1} dz$$

Wir berechnen $\Gamma(r)\Gamma(s)$. Mit Hilfe der Substitution $v = u(1-t)$ und $w = tu$, die die Eigenschaft hat, dass (v, w) die Menge $\mathbb{R}^+ \times \mathbb{R}^+$ durchläuft, wenn t durch $(0, 1)$ und u durch \mathbb{R}^+ läuft, und für die $|\frac{\partial(v,w)}{\partial(t,u)}| = u$ gilt, erhalten wir

$$\begin{aligned} \Gamma(r)\Gamma(s) &= \int_0^\infty \int_0^\infty e^{-(w+v)} v^{r-1} w^{s-1} dw dv = \int_0^\infty \int_0^1 e^{-u} (u-ut)^{r-1} (tu)^{s-1} u dt du \\ &= \int_0^\infty e^{-u} u^{r+s-1} du \int_0^1 (1-t)^{r-1} t^{s-1} dt = \Gamma(r+s) \int_0^1 (1-t)^{r-1} t^{s-1} dt \end{aligned}$$

Setzt man das oben ein, so ergibt sich

$$g(x) = \frac{\lambda^{r+s}}{\Gamma(r+s)} e^{-\lambda x} x^{r+s-1}$$

Das aber ist die Wahrscheinlichkeitsdichte der $G(r + s, \lambda)$ -Verteilung. \square

Satz 31: Seien X_1, X_2, \dots, X_n unabhängige Zufallsvariable, die alle $N(0, 1)$ -verteilt sind. Dann hat $X_1^2 + X_2^2 + \dots + X_n^2$ die $G(\frac{n}{2}, \frac{1}{2})$ -Verteilung.

Beweis: Für $n = 1$ wurde dieses Resultat in Beispiel 15 gezeigt. Wir nehmen an, dass für $Y = X_1^2 + X_2^2 + \dots + X_{n-1}^2$ die $G(\frac{n-1}{2}, \frac{1}{2})$ -Verteilung schon nachgewiesen wurde. Da $Z = X_n^2$ die $G(\frac{1}{2}, \frac{1}{2})$ -Verteilung hat und wegen Satz 25 die Zufallsvariablen Y und Z unabhängig sind, hat $X_1^2 + X_2^2 + \dots + X_n^2 = Y + Z$ nach Satz 30 die $G(\frac{n}{2}, \frac{1}{2})$ -Verteilung. Damit ist der Satz durch Induktion bewiesen. \square

Die in Satz 31 auftretende $G(\frac{n}{2}, \frac{1}{2})$ -Verteilung wird in der Statistik chi-quadrat-Verteilung genannt. Wir behandeln zwei weitere Verteilungen, die in der Statistik auftreten.

Definition: Sei $n \in \{1, 2, 3, \dots\}$. Eine Zufallsvariable X heißt T -verteilt mit Parameter n , wenn sie Wertebereich \mathbb{R} hat und Wahrscheinlichkeitsdichte

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \quad \text{für } x \in \mathbb{R}$$

Wir nennen sie kurz $T(n)$ -verteilt.

Definition: Seien m und n in $\{1, 2, 3, \dots\}$. Eine Zufallsvariable X heißt F -verteilt mit Parametern m und n , wenn sie Wertebereich \mathbb{R}^+ hat und Wahrscheinlichkeitsdichte

$$f(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}} \quad \text{für } x \in \mathbb{R}^+$$

Wir nennen sie kurz $F(m, n)$ -verteilt.

Satz 32: Die Zufallsvariablen X und Y seien unabhängig, X sei $N(0, 1)$ -verteilt und Y sei $G(\frac{n}{2}, \frac{1}{2})$ -verteilt. Dann hat $\frac{X}{\sqrt{Y/n}}$ die $T(n)$ -Verteilung.

Beweis: Eine Wahrscheinlichkeitsdichte von Y ist $f(x) = \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$ für $x > 0$ und $= 0$ für $x \leq 0$. Eine Wahrscheinlichkeitsdichte von Y/n ist $nf(nx)$ und die von $\sqrt{Y/n}$ ist 0 für $x \leq 0$ und $2xn f(nx^2)$ für $x > 0$, beides nach Satz 27. Wegen Satz 25 sind X und $\sqrt{Y/n}$ unabhängig, sodass der Zufallsvektor $(X, \sqrt{Y/n})$ wegen Satz 23(b) Dichte

$$g(x, y) = \begin{cases} \varphi(x) 2yn f(ny^2) & \text{für } x \in \mathbb{R}, y > 0 \\ 0 & \text{für } x \in \mathbb{R}, y \leq 0 \end{cases}$$

hat, wobei $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ eine Wahrscheinlichkeitsdichte von X ist. Nach Folgerung (f) von Satz 27 ergibt sich eine Wahrscheinlichkeitsdichte $h(x)$ von $\frac{X}{\sqrt{Y/n}}$ durch

$$h(x) = \int_{-\infty}^{\infty} g(xy, y) |y| dy = \int_0^{\infty} \varphi(xy) 2yn f(ny^2) |y| dy$$

Setzt man φ und f ein, so ergibt sich

$$h(x) = \frac{2^{-\frac{n}{2}+1} n^{\frac{n}{2}}}{\sqrt{2\pi}\Gamma(\frac{n}{2})} \int_0^{\infty} e^{-\frac{x^2 y^2}{2}} y^{n-1} e^{-\frac{ny^2}{2}} y dy = \frac{n^{\frac{n}{2}}}{\sqrt{\pi}\Gamma(\frac{n}{2})} \int_0^{\infty} e^{-\frac{y^2}{2}(x^2+n)} (\frac{y^2}{2})^{\frac{n-1}{2}} y dy$$

Führt man die neue Integrationsvariable $u = \frac{y^2}{2}(n+x^2)$ ein, so erhält man

$$h(x) = \frac{n^{\frac{n}{2}}}{\sqrt{\pi}\Gamma(\frac{n}{2})} (n+x^2)^{-\frac{n-1}{2}} \int_0^{\infty} e^{-u} u^{\frac{n-1}{2}} \frac{1}{n+x^2} du = \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \int_0^{\infty} e^{-u} u^{\frac{n-1}{2}} du$$

Da $\Gamma(\frac{n+1}{2}) = \int_0^{\infty} e^{-u} u^{\frac{n+1}{2}-1} du$ gilt, ist das bereits die Dichte der $T(n)$ -Verteilung. \square

Satz 33: Die Zufallsvariablen X und Y seien unabhängig, X sei $G(\frac{k}{2}, \frac{1}{2})$ -verteilt und Y sei $G(\frac{m}{2}, \frac{1}{2})$ -verteilt. Dann hat $\frac{\frac{1}{k}X}{\frac{1}{m}Y}$ die $F(k, m)$ -Verteilung.

Beweis: Sei f_n die Wahrscheinlichkeitsdichte der $G(\frac{n}{2}, \frac{1}{2})$ -Verteilung, das heißt

$$f_n(x) = \begin{cases} \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Dann haben $\frac{1}{k}X$ und $\frac{1}{m}Y$ Wahrscheinlichkeitsdichten $x \mapsto kf_k(kx)$ und $x \mapsto mf_m(mx)$ nach Satz 27. Da $\frac{1}{k}X$ und $\frac{1}{m}Y$ unabhängig sind, hat der Zufallsvektor $(\frac{1}{k}X, \frac{1}{m}Y)$ Dichte $(x, y) \mapsto kf_k(kx)mf_m(my)$. Eine Dichte g von $\frac{\frac{1}{k}X}{\frac{1}{m}Y}$ ist dann nach Folgerung (f) von Satz 27

$$g(x) = \int_{-\infty}^{\infty} kf_k(kxy)mf_m(my)|y|dy = \int_0^{\infty} kf_k(kxy)mf_m(my)ydy$$

da ja $f_m(my) = 0$ für $y \leq 0$ gilt. Für $x \leq 0$ hat man $g(x) = 0$, da in diesem Fall wegen $y \in [0, \infty)$ ja $f_k(kxy) = 0$ gilt. Indem wir für f_k und f_m einsetzen, erhalten wir für $x > 0$

$$g(x) = (\frac{1}{2})^{\frac{k}{2}} (\frac{1}{2})^{\frac{m}{2}} \frac{1}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})} k^{\frac{k}{2}} m^{\frac{m}{2}} x^{\frac{k}{2}-1} \int_0^{\infty} y^{\frac{k+m}{2}-1} e^{-\frac{kx+m}{2}y} dy$$

Führt man die neue Integrationsvariable $t = \frac{kx+m}{2}y$ ein, so erhält man

$$\int_0^{\infty} y^{\frac{k+m}{2}-1} e^{-\frac{kx+m}{2}y} dy = (\frac{2}{kx+m})^{\frac{k+m}{2}} \int_0^{\infty} t^{\frac{k+m}{2}-1} e^{-t} dt = (\frac{2}{kx+m})^{\frac{k+m}{2}} \Gamma(\frac{k+m}{2})$$

Setzt man das oben ein, so hat man

$$g(x) = \frac{\Gamma(\frac{k+m}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})} k^{\frac{k}{2}} m^{\frac{m}{2}} \frac{x^{\frac{k}{2}-1}}{(kx+m)^{\frac{k+m}{2}}}$$

die Wahrscheinlichkeitsdichte der $F(k, m)$ -Verteilung. □

11. Erwartungswert und Varianz

Sei R die endliche Menge der möglichen Ausfälle eines Zufallsexperiments, wobei R eine Teilmenge der reellen Zahlen ist. Seien $w(i)$ mit $i \in R$ die Wahrscheinlichkeiten, mit denen diese Ausfälle auftreten. So ein Zufallsexperiment wurde durch eine diskrete Zufallsvariable X mit Wertebereich R und Einzelwahrscheinlichkeiten $w(i)$ beschrieben.

Wir führen das Zufallsexperiment k Mal durch und bilden den Mittelwert der dabei auftretenden Ausfälle. Ist $N_k(i)$ die Anzahl der Wiederholungen, bei denen der Ausfall i eintritt, dann ist $\frac{1}{k} \sum_{i \in R} iN_k(i)$ der Mittelwert der Ausfälle der k Wiederholungen. Wegen $w(i) = \lim_{k \rightarrow \infty} \frac{1}{k} N_k(i)$ erhalten wir $\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i \in R} iN_k(i) = \sum_{i \in R} iw(i)$ als Mittelwert der Ausfälle bei vielen Wiederholungen.

Definition: Sei X eine diskrete Zufallsvariable mit Wertebereich R und Einzelwahrscheinlichkeiten $w(i)$ mit $i \in R$. Dann nennt man $E(X) = \sum_{i \in R} iw(i)$ den Erwartungswert der Zufallsvariablen X . Dabei nimmt man an, dass $\sum_{i \in R} |i|w(i) < \infty$ gilt. Ansonsten sagt man, dass $E(X)$ nicht existiert.

Beispiel 16: Für eine $B(n, p)$ -verteilte Zufallsvariable X soll der Erwartungswert $E(X)$ berechnet werden.

Der Wertebereich ist $R = \{0, 1, \dots, n\}$ und $w(i) = \binom{n}{i} p^i (1-p)^{n-i}$ mit $i \in R$ sind die Einzelwahrscheinlichkeiten. Wir setzen in die Formel ein

$$\begin{aligned} E(X) &= \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n n \binom{n-1}{i-1} p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np \end{aligned}$$

Für eine $B(n, p)$ -verteilte Zufallsvariable X gilt also $E(X) = np$.

Satz 34: Sei X ein Zufallsvektor mit Wertebereich $S = R_1 \times R_2 \times \dots \times R_n$ und Einzelwahrscheinlichkeiten $u(k)$ mit $k \in S$. Sei $\psi : S \rightarrow \mathbb{R}$ eine Abbildung und $Y = \psi(X)$. Dann gilt $E(Y) = \sum_{k \in S} \psi(k)u(k)$, wenn $\sum_{k \in S} |\psi(k)|u(k) < \infty$ gilt.

Beweis: Nach Satz 28 hat die Zufallsvariable Y den Wertebereich $\psi(S) \subset \mathbb{R}$ und die Einzelwahrscheinlichkeiten $v(i) = \sum_{k \in \psi^{-1}(i)} u(k)$ mit $i \in \psi(S)$. Unter Verwendung von $k \in \psi^{-1}(i) \Leftrightarrow i = \psi(k)$ und da man Reihen mit nichtnegativen Summanden beliebig umordnen kann, erhalten wir

$$\sum_{i \in \psi(S)} |i|v(i) = \sum_{i \in \psi(S)} |i| \sum_{k \in \psi^{-1}(i)} u(k) = \sum_{i \in \psi(S)} \sum_{k \in \psi^{-1}(i)} |\psi(k)|u(k) = \sum_{k \in S} |\psi(k)|u(k)$$

Wenn $\sum_{k \in S} |\psi(k)|u(k) < \infty$ ist, dann gilt auch $\sum_{i \in \psi(S)} |i|v(i) < \infty$, das heißt $E(Y)$ existiert. Obige Rechnung gilt dann aber auch ohne Absolutbeträge, da man absolut konvergente Reihen ebenfalls beliebig umordnen kann, sodass wir $\sum_{i \in \psi(S)} iv(i) = \sum_{k \in S} \psi(k)u(k)$ erhalten. Da $E(Y) = \sum_{i \in \psi(S)} iv(i)$ nach Definition gilt, ist der Satz bewiesen. \square

Für kontinuierliche Zufallsvariable definieren wir den Erwartungswert analog zu den diskreten Zufallsvariablen.

Definition: Sei X eine Zufallsvariable mit Wahrscheinlichkeitsdichte $f : \mathbb{R} \rightarrow [0, \infty)$. Dann nennt man $E(X) = \int_{-\infty}^{\infty} xf(x) dx$ den Erwartungswert der Zufallsvariablen X . Dabei nimmt man an, dass $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$ gilt. Ansonsten sagt man, dass $E(X)$ nicht existiert.

Bemerkung: Das uneigentliche Integral $\int_{-\infty}^{\infty} xf(x) dx$ in der Definition des Erwartungswerts wird definiert als $\lim_{n \rightarrow \infty} \int_{-a_n}^{b_n} xf(x) dx$, wobei $(a_n)_{n \geq 1}$ und $(b_n)_{n \geq 1}$ monoton wachsende Folgen sind, die gegen ∞ gehen. Ist die Bedingung $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$ erfüllt, dann hängt $\lim_{n \rightarrow \infty} \int_{-a_n}^{b_n} xf(x) dx$ nicht von der Wahl der Folgen $(a_n)_{n \geq 1}$ und $(b_n)_{n \geq 1}$ ab. Sonst erhält man durch verschiedene Wahl der Folgen $(a_n)_{n \geq 1}$ und $(b_n)_{n \geq 1}$ auch verschiedene Grenzwerte.

Beispiel 17: Sei X eine $E(\lambda)$ -verteilte Zufallsvariable. Zu berechnen ist $E(X)$.

Setzt man die Wahrscheinlichkeitsdichte der Exponentialverteilung in die Formel für den Erwartungswert ein, so erhält man durch partielle Integration

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = -\frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = \frac{1}{\lambda}.$$

In den folgenden Beweisen nehmen wir an, dass die Zufallsvariablen Wahrscheinlichkeitsdichten besitzen. Mit Hilfe des Riemann-Stieltjes-Integrals, auf das wir später zurückkommen, lassen sich Definition des Erwartungswerts und diese Beweise in einer allgemeineren Form schreiben, sodass sie dann auch ohne die Annahme einer Wahrscheinlichkeitsdichte gelten.

Der Erwartungswert wurde durch $E(X) = \int_{-\infty}^{\infty} xf(x) dx$ definiert, wobei f eine Dichte der Zufallsvariable X ist. Die Dichte ist jedoch nicht eindeutig. Man erhält aber immer denselben Wert für $E(X)$, welche Dichte von X man auch wählt. Das zeigt folgender Satz.

Satz 35: Für eine Zufallsvariable X gilt $E(X) = \int_0^{\infty} P(X > t) dt - \int_{-\infty}^0 P(X \leq t) dt$, wobei $E(X)$ genau dann existiert, wenn beide Integrale $< \infty$ sind.

Bemerkung: Diese Formel ist äquivalent zu $E(X) = \int_0^{\infty} 1 - F(t) dt - \int_{-\infty}^0 F(t) dt$, wobei F die Verteilungsfunktion von X ist. Wenn $X \geq 0$ gilt (das heißt Wertebereich $\subset \mathbb{R}^+$), dann ist $P(X \leq t) = 0$ für $t < 0$ und es folgt $E(X) = \int_0^{\infty} P(X > t) dt = \int_0^{\infty} 1 - F(t) dt$. Insbesondere gilt $E(X) \geq 0$, wenn $X \geq 0$ ist.

Beweis: Sei f eine Dichte von X . Wir berechnen die beiden Integrale auf der rechten Seite $\int_{-\infty}^0 P(X \leq t) dt = \int_{-\infty}^0 \int_{-\infty}^t f(x) dx dt = \int_{-\infty}^0 \int_x^0 f(x) dt dx = \int_{-\infty}^0 f(x) \int_x^0 dt dx = \int_{-\infty}^0 -xf(x) dx$
 $\int_0^{\infty} P(X > t) dt = \int_0^{\infty} \int_t^{\infty} f(x) dx dt = \int_0^{\infty} \int_0^x f(x) dt dx = \int_0^{\infty} f(x) \int_0^x dt dx = \int_0^{\infty} xf(x) dx$

Nun existiert $E(X)$ genau dann, wenn $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ ist, das heißt genau dann, wenn $\int_{-\infty}^0 P(X \leq t)dt = \int_{-\infty}^0 |x|f(x)dx < \infty$ und $\int_0^{\infty} P(X > t)dt = \int_0^{\infty} |x|f(x)dx < \infty$ gilt. Es folgt dann $E(X) = \int_0^{\infty} xf(x)dx + \int_{-\infty}^0 xf(x)dx = \int_0^{\infty} P(X > t)dt - \int_{-\infty}^0 P(X \leq t)dt$. \square

Satz 36: Sei X ein kontinuierlicher Zufallsvektor und $f : \mathbb{R}^n \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte von X . Sei $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig (oder stückweise stetig) und $Y = \psi(X)$. Dann gilt $E(Y) = \int_{\mathbb{R}^n} \psi(x)f(x)dx$, wobei $E(Y)$ genau dann existiert, wenn $\int_{\mathbb{R}^n} |\psi(x)|f(x)dx < \infty$ ist.

Beweis: Wir verwenden die Formel aus Satz 35. Sei $I_t = (t, \infty)$. Dann gilt

$$\begin{aligned} \int_0^{\infty} P(Y > t)dt &= \int_0^{\infty} P(\psi(X) > t)dt = \int_0^{\infty} P(X \in \psi^{-1}(I_t))dt = \int_0^{\infty} \int_{\psi^{-1}(I_t)} f(x)dx dt \\ &= \int_0^{\infty} \int_{\mathbb{R}^n} 1_{\psi^{-1}(I_t)}(x)f(x)dx dt \end{aligned}$$

Wegen $1_{\psi^{-1}(I_t)}(x) = 1 \Leftrightarrow x \in \psi^{-1}(I_t) \Leftrightarrow \psi(x) \in I_t \Leftrightarrow t < \psi(x) \Leftrightarrow 1_{(-\infty, \psi(x))}(t) = 1$ erhalten wir $1_{\psi^{-1}(I_t)}(x) = 1_{(-\infty, \psi(x))}(t)$. Setzt man das oben ein so folgt

$$\int_0^{\infty} P(Y > t)dt = \int_{\mathbb{R}^n} \int_0^{\infty} 1_{(-\infty, \psi(x))}(t) dt f(x)dx = \int_{\mathbb{R}^n} \max(0, \psi(x))f(x)dx$$

Genauso erhalten wir mit $J_t = (-\infty, t]$

$$\begin{aligned} \int_{-\infty}^0 P(Y \leq t)dt &= \int_{-\infty}^0 P(\psi(X) \leq t)dt = \int_{-\infty}^0 P(X \in \psi^{-1}(J_t))dt = \int_{-\infty}^0 \int_{\psi^{-1}(J_t)} f(x)dx dt \\ &= \int_{-\infty}^0 \int_{\mathbb{R}^n} 1_{\psi^{-1}(J_t)}(x)f(x)dx dt = \int_{\mathbb{R}^n} \int_{-\infty}^0 1_{[\psi(x), \infty)}(t) dt f(x)dx \\ &= \int_{\mathbb{R}^n} \max(0, -\psi(x))f(x)dx \end{aligned}$$

Nach Satz 35 existiert $E(Y)$ genau dann, wenn sowohl $\int_0^{\infty} P(Y > t)dt < \infty$ als auch $\int_{-\infty}^0 P(Y \leq t)dt < \infty$ gilt, das heißt wenn sowohl $\int_{\mathbb{R}^n} \max(0, \psi(x))f(x)dx < \infty$ als auch $\int_{\mathbb{R}^n} \max(0, -\psi(x))f(x)dx < \infty$ erfüllt ist. Das ist äquivalent zu $\int_{\mathbb{R}^n} |\psi(x)|f(x)dx < \infty$, da einerseits $|u| = \max(0, u) + \max(0, -u)$ und andererseits $0 \leq \max(0, u) \leq |u|$ und $0 \leq \max(0, -u) \leq |u|$ für alle $u \in \mathbb{R}$ gilt. Damit ist die Bedingung für die Existenz von $E(Y)$ gefunden. Gilt sie, dann folgt $E(Y) = \int_0^{\infty} P(Y > t)dt - \int_{-\infty}^0 P(Y \leq t)dt$ aus Satz 35, das heißt $E(Y) = \int_{\mathbb{R}^n} \psi(x)f(x)dx$, da $u = \max(0, u) - \max(0, -u)$ für alle $u \in \mathbb{R}$ gilt. \square

Nun ist es einfach, Eigenschaften des Erwartungswerts zu beweisen.

Satz 37: Seien X und Y Zufallsvariable, für die der Erwartungswert existiert. Dann gilt

- (a) für a und b in \mathbb{R} existiert $E(aX + b)$ und es gilt $E(aX + b) = aE(X) + b$
- (b) $E(X + Y)$ existiert und es gilt $E(X + Y) = E(X) + E(Y)$
- (c) X und Y unabhängig $\Rightarrow E(XY)$ existiert und es gilt $E(XY) = E(X)E(Y)$

Beweis: Sei $g : \mathbb{R}^2 \rightarrow [0, \infty)$ eine Wahrscheinlichkeitsdichte des Zufallsvektors (X, Y) . Nach Satz 22(c) sind dann $f_1(x) = \int_{-\infty}^{\infty} g(x, y)dy$ und $f_2(y) = \int_{-\infty}^{\infty} g(x, y)dx$ Dichten der Zufallsvariablen X und Y . Nach Voraussetzung gilt $\int_{-\infty}^{\infty} |x|f_1(x)dx < \infty$ und $\int_{-\infty}^{\infty} |y|f_2(y)dy < \infty$.

Um (a) zu zeigen, wenden wir Satz 36 mit $\psi(x) = ax + b$ an und erhalten

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b)f_1(x)dx = a \int_{-\infty}^{\infty} xf_1(x)dx + b \int_{-\infty}^{\infty} f_1(x)dx = aE(X) + b$$

wobei $E(aX + b)$ existiert, da $\int_{-\infty}^{\infty} |ax + b|f_1(x)dx \leq |a| \int_{-\infty}^{\infty} |x|f_1(x)dx + |b| < \infty$ gilt.

Um (b) zu zeigen, wenden wir Satz 36 mit $\psi(x, y) = x + y$ an und erhalten

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)g(x, y)dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xg(x, y)dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yg(x, y)dx dy \\ &= \int_{-\infty}^{\infty} xf_1(x)dx + \int_{-\infty}^{\infty} yf_2(y)dy = E(X) + E(Y) \end{aligned}$$

Genauso zeigt man $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x + y|g(x, y)dx dy \leq \int_{-\infty}^{\infty} |x|f_1(x)dx + \int_{-\infty}^{\infty} |y|f_2(y)dy < \infty$, womit die Existenz von $E(X + Y)$ nachträglich gezeigt ist.

Um (c) zu zeigen, wenden wir Satz 36 mit $\psi(x, y) = xy$ an. Wegen Satz 23(b) können wir $g(x, y) = f_1(x)f_2(y)$ annehmen. Wir erhalten

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyg(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_1(x)f_2(y) dx dy \\ &= \int_{-\infty}^{\infty} xf_1(x) dx \int_{-\infty}^{\infty} yf_2(y) dy = E(X)E(Y) \end{aligned}$$

Genauso zeigt man $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |xy|g(x, y) dx dy = \int_{-\infty}^{\infty} |x|f_1(x) dx \int_{-\infty}^{\infty} |y|f_2(y) dy < \infty$, womit die Existenz von $E(XY)$ nachträglich gezeigt ist. \square

Folgerungen: Durch wiederholtes Anwenden von Satz 37 erhält man: Sind X_1, X_2, \dots, X_n Zufallsvariable, für die der Erwartungswert existiert, und sind a_1, a_2, \dots, a_n, b in \mathbb{R} , dann existiert $E(a_1X_1 + a_2X_2 + \dots + a_nX_n + b)$ und ist gleich $a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n) + b$. Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariable, für die der Erwartungswert existiert, dann existiert auch $E(X_1X_2 \dots X_n)$ und ist gleich $E(X_1)E(X_2) \dots E(X_n)$. Das ergibt sich mit Induktion aus Satz 37 (c), wobei man verwendet, dass für $2 \leq k \leq n$ die beiden Zufallsvariablen $Y = X_1X_2 \dots X_{k-1}$ und X_k wegen Satz 25 unabhängig sind.

Sind X und Y Zufallsvariable mit $X \leq Y$, dann gilt $Y - X \geq 0$, und daher folgt dann auch $E(Y) - E(X) = E(Y - X) \geq 0$, das heißt $E(X) \leq E(Y)$. Ist X eine Zufallsvariable und $a \in \mathbb{R}$ mit $X \leq a$, dann folgt $a - X \geq 0$, also auch $a - E(X) = E(a - X) \geq 0$, das heißt $E(X) \leq a$.

Nun kommen wir zur Varianz einer Zufallsvariablen

Definition: Sei X eine Zufallsvariable. Dann heißt $V(X) = E((X - E(X))^2)$ die Varianz der Zufallsvariablen X , vorausgesetzt die vorkommenden Erwartungswerte existieren.

Bemerkung: Wegen $(X - E(X))^2 \geq 0$ gilt $V(X) = E((X - E(X))^2) \geq 0$. Die Varianz misst die durchschnittliche quadratische Abweichung der Zufallsvariablen von ihrem Erwartungswert. Hat X eine Dichte f , dann folgt $V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$ aus Satz 36.

Beispiel 18: Sei X eine $E(\lambda)$ -verteilte Zufallsvariable. Zu berechnen ist $V(X)$.

In einem früheren Beispiel wurde bereits $E(X) = \frac{1}{\lambda}$ gezeigt. Durch Einsetzen in die Formel $V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$ und zweimalige partielle Integration erhält man

$$V(X) = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = -(x - \frac{1}{\lambda})^2 e^{-\lambda x} \Big|_0^{\infty} - 2(x - \frac{1}{\lambda}) \frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} + \int_0^{\infty} 2 \frac{e^{-\lambda x}}{\lambda} dx = \frac{1}{\lambda^2}.$$

Die Rechenregeln für die Varianz kommen später, nachdem wir die Kovarianz eingeführt haben. Zuerst zeigen wir noch zwei Ungleichungen.

Satz 38 (Markov-Ungleichung) Sei $u : \mathbb{R} \rightarrow [0, \infty)$ eine monoton wachsende Abbildung und X eine Zufallsvariable, für die $E(u(X))$ existiert. Dann gilt $P(X > t) \leq \frac{1}{u(t)} E(u(X))$ für alle $t \in \mathbb{R}$ mit $u(t) > 0$.

Beweis: Sei f eine Dichte von X und $t \in \mathbb{R}$ so, dass $u(t) > 0$. Dann gilt $\frac{u(x)}{u(t)} \geq 1$ für $x \geq t$, da u monoton wachsend ist. Es folgt $\int_t^{\infty} f(x) dx \leq \int_t^{\infty} \frac{u(x)}{u(t)} f(x) dx = \frac{1}{u(t)} \int_t^{\infty} u(x) f(x) dx$. Da $u(x) \geq 0$ für alle $x \in \mathbb{R}$ gilt, erhalten wir daraus $\int_t^{\infty} f(x) dx \leq \frac{1}{u(t)} \int_{-\infty}^{\infty} u(x) f(x) dx$. Wegen Satz 11(b) und Satz 36 heißt das $P(X > t) \leq \frac{1}{u(t)} E(u(X))$. \square

Satz 39 (Tschebyscheff-Ungleichung) Sei X eine Zufallsvariable, für die $E(X)$ und $V(X)$ existieren. Dann gilt $P(|X - E(X)| > t) \leq \frac{1}{t^2} V(X)$ für alle $t > 0$.

Beweis: Sei $u(x) = x^2$ für $x \geq 0$ und $u(x) = 0$ für $x < 0$. Verwendet man diese Funktion und $|X - E(X)|$ anstelle von X in Satz 38, dann erhält man die gewünschte Ungleichung. \square

Um die lineare Abhängigkeit von Zufallsvariablen zu studieren, führt man die Kovarianz zweier Zufallsvariable ein.

Definition: Seien X und Y Zufallsvariable. Die Kovarianz der Zufallsvariablen X und Y wird definiert durch $C(X, Y) = E((X - E(X))(Y - E(Y)))$, wobei vorausgesetzt wird, dass die vorkommenden Erwartungswerte existieren.

Bemerkung: Es gilt $C(X, X) = V(X)$ und $C(X, Y) = C(Y, X)$, wie sich sofort aus den Definitionen ergibt.

Satz 40: Sei $k \geq 2$. Seien U_1, U_2, \dots, U_n Zufallsvariable, sodass $E(U_j^k)$ für $1 \leq j \leq n$ existiert. Sind $m_1, m_2, \dots, m_n \geq 0$ mit $m_1 + m_2 + \dots + m_n \leq k$, dann existiert auch $E(U_1^{m_1} U_2^{m_2} \dots U_n^{m_n})$. Ist insbesondere X eine Zufallsvariable, sodass $E(X^k)$ existiert, dann existiert auch $E(X^m)$ für $1 \leq m \leq k - 1$.

Beweis: Unter den angegebenen Voraussetzungen gilt $|x_1^{m_1} x_2^{m_2} \dots x_n^{m_n}| \leq 1 + \sum_{j=1}^n |x_j^k|$ für alle $x_1, x_2, \dots, x_n \in \mathbb{R}$. Das sieht man, indem man $a = \max\{1, |x_1|, |x_2|, \dots, |x_n|\}$ setzt, womit $|x_1^{m_1} x_2^{m_2} \dots x_n^{m_n}| \leq a^{m_1 + \dots + m_n} \leq a^k = \max\{1, |x_1|^k, \dots, |x_n|^k\} \leq 1 + \sum_{j=1}^n |x_j^k|$ folgt.

Sei $g: \mathbb{R}^n \rightarrow [0, \infty)$ eine Dichte des Zufallsvektors (U_1, U_2, \dots, U_n) . Aus der Voraussetzung und aus Satz 36 mit $\psi(x_1, \dots, x_n) = x_j^k$ folgt $\int_{\mathbb{R}^n} |x_j^k| g(x) dx < \infty$ für $1 \leq j \leq n$. Mit obiger Ungleichung hat man $\int_{\mathbb{R}^n} |x_1^{m_1} x_2^{m_2} \dots x_n^{m_n}| g(x) dx \leq \int_{\mathbb{R}^n} g(x) dx + \sum_{j=1}^n \int_{\mathbb{R}^n} |x_j^k| g(x) dx < \infty$. Aus Satz 36 mit $\psi(x_1, \dots, x_n) = x_1^{m_1} x_2^{m_2} \dots x_n^{m_n}$ folgt, dass $E(U_1^{m_1} U_2^{m_2} \dots U_n^{m_n})$ existiert.

Die letzte Aussage ist der Spezialfall $n = 1$. \square

Satz 41: Seien X und Y Zufallsvariable, sodass $E(X^2)$ und $E(Y^2)$ existieren. Dann existiert auch $C(X, Y)$ und es gilt $C(X, Y) = E(XY) - E(X)E(Y)$. Sind X und Y unabhängig, dann gilt $C(X, Y) = 0$.

Beweis: Nach Voraussetzung und Satz 40 existieren $k = E(X)$, $m = E(Y)$ und $E(XY)$. Nach Satz 37 existiert dann der Erwartungswert von $(X - k)(Y - m) = XY - mX - kY + km$, das ist aber $C(X, Y)$, sodass die Existenz von $C(X, Y)$ gezeigt ist. Aus Satz 37 folgt dann auch $C(X, Y) = E((X - k)(Y - m)) = E(XY) - mE(X) - kE(Y) + km = E(XY) - km$. Damit ist $C(X, Y) = E(XY) - E(X)E(Y)$ gezeigt. Sind X und Y unabhängig, dann haben wir $E(XY) = E(X)E(Y)$ nach Satz 37(c), sodass $C(X, Y) = 0$ gilt. \square

Satz 42: Seien X_1, X_2, \dots, X_m und Y_1, Y_2, \dots, Y_n Zufallsvariable, deren Quadrate einen Erwartungswert besitzen. Weiters seien a_1, a_2, \dots, a_m, b und c_1, c_2, \dots, c_n, d in \mathbb{R} . Dann existiert $C(\sum_{i=1}^m a_i X_i + b, \sum_{j=1}^n c_j Y_j + d)$ und ist gleich $\sum_{i=1}^m \sum_{j=1}^n a_i c_j C(X_i, Y_j)$.

Beweis: Sei $U = \sum_{i=1}^m a_i X_i + b$ und $V = \sum_{j=1}^n c_j Y_j + d$. Wegen Satz 40 existiert $E(X_i)$ für $1 \leq i \leq m$. Aus Satz 37 folgt, dass $E(U)$ existiert und $E(U) = \sum_{i=1}^m a_i E(X_i) + b$ gilt, also $U - E(U) = \sum_{i=1}^m a_i (X_i - E(X_i))$. Ebenso folgt $V - E(V) = \sum_{j=1}^n c_j (Y_j - E(Y_j))$. Wir haben daher $(U - E(U))(V - E(V)) = \sum_{i=1}^m \sum_{j=1}^n a_i c_j (X_i - E(X_i))(Y_j - E(Y_j))$. Nach Voraussetzung und Satz 41 existiert $C(X_i, Y_j) = E((X_i - E(X_i))(Y_j - E(Y_j)))$ für $1 \leq i \leq m$ und $1 \leq j \leq n$. Aus Satz 37 folgt dann, dass $C(U, V) = E((U - E(U))(V - E(V)))$ existiert und $C(U, V) = \sum_{i=1}^m \sum_{j=1}^n a_i c_j C(X_i, Y_j)$ gilt. \square

Mit Hilfe der letzten Sätze erhalten wir auch die Rechenregeln für die Varianz.

Satz 43: Sei X eine Zufallsvariable, für die $E(X^2)$ existiert. Dann existieren auch $E(X)$ und $V(X)$ und es gilt $V(X) = E(X^2) - E(X)^2$. Für a und b in \mathbb{R} existiert auch $V(aX + b)$ und es gilt $V(aX + b) = a^2 V(X)$.

Beweis: Wegen $V(X) = C(X, X)$ ist die erste Aussage ein Spezialfall von Satz 41. Die zweite Aussage ist ein Spezialfall von Satz 42 wegen $V(aX + b) = C(aX + b, aX + b)$. \square

Beispiel 19: Sei X eine $N(\mu, \sigma)$ -verteilte Zufallsvariable. Gesucht sind $E(X)$ und $V(X)$.

Sei $Y = \frac{X-\mu}{\sigma}$. Dann hat Y wegen Satz 17 die $N(0, 1)$ -Verteilung und somit die Wahrscheinlichkeitsdichte $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Wegen $\varphi(-x) = \varphi(x)$ folgt $E(Y) = \int_{-\infty}^{\infty} x\varphi(x) dx = 0$. Wegen $\varphi'(x) = -x\varphi(x)$ erhalten wir $1 = \int_{-\infty}^{\infty} 1 \cdot \varphi(x) dx = x\varphi(x) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x^2\varphi(x) dx$. Wegen $\lim_{x \rightarrow \infty} x\varphi(x) = 0$ und $\lim_{x \rightarrow -\infty} x\varphi(x) = 0$ ist damit $E(Y^2) = \int_{-\infty}^{\infty} x^2\varphi(x) dx = 1$ gezeigt. Aus Satz 43 folgt $V(Y) = 1$.

Nun erhalten wir $E(X) = E(\sigma Y + \mu) = \sigma E(Y) + \mu = \mu$ mit Hilfe von Satz 37 (a) und $V(X) = V(\sigma Y + \mu) = \sigma^2 V(Y) = \sigma^2$ folgt aus Satz 43.

Satz 44: Seien X_1, X_2, \dots, X_n Zufallsvariable, sodass $E(X_j^2)$ für alle j existiert. Dann gilt

- (a) $V(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \sum_{j=1}^n C(X_i, X_j) = \sum_{j=1}^n V(X_j) + 2 \sum_{1 \leq i < j \leq n} C(X_i, X_j)$
 (b) X_1, X_2, \dots, X_n unabhängig $\Rightarrow V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$

Beweis: Es gilt $V(X_1 + X_2 + \dots + X_n) = C(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j)$. Mit Hilfe von Satz 42 erhält man $V(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \sum_{j=1}^n C(X_i, X_j)$. Wegen $C(X_j, X_j) = V(X_j)$ und $C(X_i, X_j) = C(X_j, X_i)$ für $i \neq j$ ist damit (a) gezeigt.

Sind die Zufallsvariablen X_1, X_2, \dots, X_n unabhängig, dann gilt $C(X_i, X_j) = 0$ für $i \neq j$ nach Satz 41, sodass $V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$ aus (a) folgt. \square

Erwartungswert und Kovarianz kann man auch für Zufallsvektoren einführen.

Definition: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor. Wir definieren

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} \quad \text{und} \quad D(X) = \begin{pmatrix} C(X_1, X_1) & \dots & C(X_1, X_n) \\ \vdots & \ddots & \vdots \\ C(X_n, X_1) & \dots & C(X_n, X_n) \end{pmatrix}$$

und nennen $E(X)$ den Erwartungswertvektor und $D(X)$ die Dispersionsmatrix oder Kovarianzmatrix des Zufallsvektors X .

Satz 45: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor, sodass $E(X_j^2)$ für $1 \leq j \leq n$ existiert. Sei A eine $m \times n$ -Matrix, sei $b \in \mathbb{R}^m$ und $Y = AX + b$. Dann gilt $E(Y) = AE(X) + b$ und $D(Y) = AD(X)A^t$.

Beweis: Sei $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ und $b = (b_i)_{1 \leq i \leq m}$. Mit Satz 40 und Satz 37 folgt

$$E(Y_i) = E(\sum_{j=1}^n a_{ij} X_j + b_i) = \sum_{j=1}^n a_{ij} E(X_j) + b_i \quad \text{für } 1 \leq i \leq m$$

Fasst man diese Gleichungen zusammen, so hat man $E(Y) = AE(X) + b$.

Für $1 \leq i \leq m$ und $1 \leq k \leq m$ folgt mit Hilfe von von Satz 42

$$C(Y_i, Y_k) = C(\sum_{j=1}^n a_{ij} X_j + b_i, \sum_{l=1}^n a_{kl} X_l + b_k) = \sum_{j=1}^n a_{ij} \sum_{l=1}^n a_{kl} C(X_j, X_l)$$

Fasst man diese Gleichungen zusammen, so hat man $D(Y) = AD(X)A^t$. \square

Beispiel 20: Sei X ein $ZN(M)$ -verteilter n -dimensionaler Zufallsvektor. Gesucht ist der Erwartungswertvektor $E(X)$ und die Kovarianzmatrix $D(X)$.

Da M eine positiv definite $n \times n$ -Matrix ist, gilt $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ für die Eigenwerte von M . Sei U die Diagonalmatrix mit $\lambda_1, \lambda_2, \dots, \lambda_n$ in der Diagonale. Die symmetrische Matrix M ist diagonalisierbar. Es existiert eine orthogonale $n \times n$ -Matrix A mit $AMA^t = U$,

sodass dann wegen Satz 29 (a) der Zufallsvektor $Y = AX$ die $ZN(U)$ -Verteilung hat. Er hat Dichte $g(x) = \frac{1}{\sqrt{2\pi^n} \sqrt{\det U}} e^{-\frac{1}{2}x^t U^{-1}x} = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{1}{2\lambda_j}x_j^2}$. Wegen Satz 24 (c) sind die Komponenten Y_1, Y_2, \dots, Y_n des Zufallsvektors Y unabhängig und Y_j ist $N(0, \sqrt{\lambda_j})$ -verteilt. Nach Beispiel 19 gilt $E(Y_j) = 0$ für $1 \leq j \leq n$, das heißt der Erwartungswertvektor $E(Y)$ ist gleich 0. Aus Satz 45 folgt $E(X) = E(A^{-1}Y) = A^{-1}E(Y) = 0$. Damit ist $E(X)$ berechnet. Ebenso gilt $C(Y_i, Y_j) = 0$ für $i \neq j$ wegen Satz 41 und $C(Y_i, Y_i) = V(Y_i) = \lambda_i$ für $1 \leq i \leq n$ nach Beispiel 19. Wir haben also $D(Y) = U$. Mit Hilfe von Satz 45 erhalten wir $D(X) = D(A^{-1}Y) = A^{-1}U(A^{-1})^t = A^{-1}AM A^t(A^{-1})^t = A^{-1}AM(A^{-1}A)^t = M$. Damit ist auch $D(X)$ berechnet.

Eine symmetrische $n \times n$ -Matrix M heißt positiv semidefinit, wenn $a^t M a \geq 0$ für alle $a \in \mathbb{R}^n$ gilt, und positiv definit, wenn $a^t M a > 0$ für alle $a \in \mathbb{R}^n \setminus \{0\}$ gilt. Eine positiv semidefinite Matrix M ist genau dann positiv definit, wenn $\det M \neq 0$ gilt. Wir beweisen weitere Eigenschaften der Kovarianzmatrix.

Satz 46: Für eine Zufallsvariable Y ist $V(Y) = 0$ äquivalent zu $P(Y = E(Y)) = 1$.

Beweis: Sei $U = (Y - E(Y))^2$. Gilt $P(Y = E(Y)) = 1$, dann auch $P(U = 0) = 1$. Aus der Definition des Erwartungswerts folgt $E(U) = 0$, das heißt $V(Y) = E((Y - E(Y))^2) = 0$.

Ist $V(Y) = 0$, dann folgt $P(|Y - E(Y)| > \frac{1}{k}) \leq k^2 V(Y) = 0$ für alle $k \in \mathbb{N}$ aus Satz 39. Die Ereignisse $|Y - E(Y)| > \frac{1}{k}$ bilden eine monoton aufsteigende Folge, die für $k \rightarrow \infty$ gegen das Ereignis $|Y - E(Y)| > 0$ geht. Aus dem Stetigkeitssatz folgt $P(|Y - E(Y)| > 0) = \lim_{k \rightarrow \infty} P(|Y - E(Y)| > \frac{1}{k}) = 0$. Somit gilt $P(Y = E(Y)) = 1 - P(|Y - E(Y)| > 0) = 1$. \square

Satz 47: Sei $X = (X_1, X_2, \dots, X_n)$ ein Zufallsvektor, sodass $E(X_j^2)$ für $1 \leq j \leq n$ existiert. Für alle $a \in \mathbb{R}^n$ gilt dann $V(\langle a, X \rangle) = a^t D(X)a$ und $D(X)$ ist positiv semidefinit. Weiters ist $\det D(X)$ genau dann gleich 0, wenn $a_1, a_2, \dots, a_n \in \mathbb{R}$ existieren, die nicht alle 0 sind, sodass $P(\sum_{j=1}^n a_j(X_j - E(X_j)) = 0) = 1$ gilt, das heißt mit Wahrscheinlichkeit 1 sind die Zufallsvariablen $X_1 - E(X_1), X_2 - E(X_2), \dots, X_n - E(X_n)$ linear abhängig.

Beweis: Seien $a_1, a_2, \dots, a_n \in \mathbb{R}$ die Komponenten von a . Aus Satz 42 folgt

$$V(\sum_{j=1}^n a_j X_j) = C(\sum_{j=1}^n a_j X_j, \sum_{j=1}^n a_j X_j) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j C(X_i, X_j) = a^t D(X)a$$

Da eine Varianz immer ≥ 0 ist, folgt auch $a^t D(X)a \geq 0$ für alle $a \in \mathbb{R}^n$, womit die ersten beiden Aussagen bereits bewiesen sind.

Für die positiv semidefinite Matrix $D(X)$ gilt $\det D(X) = 0$ genau dann, wenn ein $a \in \mathbb{R}^n \setminus \{0\}$ existiert mit $a^t D(X)a = 0$. Nach obiger Rechnung ist das äquivalent zur Existenz von $a_1, a_2, \dots, a_n \in \mathbb{R}$, die nicht alle 0 sind, sodass $V(\sum_{j=1}^n a_j X_j) = 0$ gilt. Wegen Satz 46 und da $E(\sum_{j=1}^n a_j X_j) = \sum_{j=1}^n a_j E(X_j)$ nach Satz 37 gilt, ist das wieder äquivalent zur Existenz von $a_1, a_2, \dots, a_n \in \mathbb{R}$, die nicht alle 0 sind, sodass $P(\sum_{j=1}^n a_j(X_j - E(X_j)) = 0) = 1$ gilt. \square

Bemerkung: Die Varianz $V(Y)$ einer Zufallsvariable Y gibt an, wie stark die Zufallsvariable um ihren Erwartungswert streut. Sei X ein n -dimensionaler Zufallsvektor und $a \in \mathbb{R}^n$ mit $\|a\| = 1$. Dann ist $Y = \langle a, X \rangle$ die Länge der orthogonalen Projektion des Vektors X auf eine Gerade mit Richtung a . Nach Satz 47 gilt $V(Y) = a^t D(X)a$. Diese Zahl gibt an, wie stark X in Richtung a streut. Man sieht, dass die Dispersionsmatrix $D(X)$ die Information über das Streuverhalten des Zufallsvektors X enthält.

Die im letzten Satz gemachte Aussage über die lineare Abhängigkeit von Zufallsvariablen wird auf zwei Zufallsvariable spezialisiert.

Definition: Seien X_1 und X_2 Zufallsvariable, für die $E(X_1^2)$ und $E(X_2^2)$ existieren und für die $V(X_1) > 0$ und $V(X_2) > 0$ gilt. Dann heißt $K(X_1, X_2) = C(X_1, X_2)/\sqrt{V(X_1)V(X_2)}$ der Korrelationskoeffizient der Zufallsvariablen X_1 und X_2 .

Satz 48: Seien X_1 und X_2 Zufallsvariable, für die $E(X_1^2)$ und $E(X_2^2)$ existieren. Seien $V(X_1) > 0$ und $V(X_2) > 0$ und sei $X = (X_1, X_2)$. Dann gilt $K(X_1, X_2)^2 = 1 - \frac{\det D(X)}{V(X_1)V(X_2)}$ und $K(X_1, X_2)^2 \leq 1$. Es gilt $K(X_1, X_2)^2 = 1$ genau dann, wenn $P(X_1 = cX_2 + d) = 1$ für ein $d \in \mathbb{R}$ und ein $c \neq 0$ mit $\text{sign } c = K(X_1, X_2)$ gilt.

Beweis: Aus $D(X) = \begin{pmatrix} V(X_1) & C(X_1, X_2) \\ C(X_1, X_2) & V(X_2) \end{pmatrix}$ folgt $\det D(X) = V(X_1)V(X_2) - C(X_1, X_2)^2$. Dividiert man durch $V(X_1)V(X_2)$, so hat man bereits die Formel für $K(X_1, X_2)^2$. Da $D(X)$ nach Satz 47 positiv semidefinit ist, haben wir $\det D(X) \geq 0$. Es folgt $K(X_1, X_2)^2 \leq 1$.

Wegen der soeben bewiesenen Formel ist $K(X_1, X_2)^2 = 1$ äquivalent zu $\det D(X) = 0$. Das ist äquivalent zur Existenz eines $c \neq 0$ mit $P(X_1 - E(X_1) - c(X_2 - E(X_2))) = 0 = 1$ wegen Satz 47. Das wieder ist äquivalent zur Existenz eines $c \neq 0$ und eines $d \in \mathbb{R}$ mit $P(X_1 - cX_2 - d = 0) = 1$, da dann $E(X_1 - cX_2 - d) = 0$ und somit d gleich $E(X_1) - cE(X_2)$ sein muss.

Es bleibt daher nur noch, $\text{sign } c$ unter diesen äquivalenten Bedingungen zu bestimmen. Wegen $P(X_1 - E(X_1) - c(X_2 - E(X_2))) = 0 = 1$ folgt für alle Zufallsvariablen U , für die $E(U^2)$ existiert, dass $P((X_1 - E(X_1) - cX_2 - cE(X_2))(U - E(U)) = 0) = 1$ gilt. Daraus wieder folgt $E((X_1 - E(X_1) - cX_2 - cE(X_2))(U - E(U))) = 0$, das heißt $C(X_1 - cX_2, U) = 0$ oder $C(X_1, U) = cC(X_2, U)$. Für $U = X_2$ ergibt sich $C(X_1, X_2) = cV(X_2)$ und für $U = X_1$ folgt $V(X_1) = cC(X_1, X_2) = c^2V(X_2)$. Damit erhalten wir $K(X_1, X_2) = cV(X_2)/\sqrt{c^2V(X_2)^2} = c/\sqrt{c^2} = \text{sign } c$. \square

III. Transformationen und Grenzwertsätze

1. Integrale

Um diskrete und kontinuierliche Zufallsvariable nicht immer getrennt behandeln zu müssen, führt man eine Verallgemeinerung des Riemannintegrals, das sogenannte Riemann-Stieltjes-Integral einer Funktion $g : (a, b] \rightarrow \mathbb{R}$ bezüglich einer Verteilungsfunktion F ein. Es wird genauso wie das Riemann-Integral definiert, nur wird die Länge des Intervalls $(u, v]$ durch $F(v) - F(u)$ gemessen. Das ist die Wahrscheinlichkeit, dass eine Zufallsvariable mit Verteilungsfunktion F ins Intervall $(u, v]$ fällt. Zu jeder Zerlegung $a = t_0 < t_1 < \dots < t_m = b$ des Intervalls $(a, b]$ definieren wir die Obersumme $\sum_{j=1}^m (F(t_j) - F(t_{j-1})) \sup_{(t_{j-1}, t_j]} g$ und die Untersumme $\sum_{j=1}^m (F(t_j) - F(t_{j-1})) \inf_{(t_{j-1}, t_j]} g$. Ist das Infimum der Obersummen gleich dem Supremum der Untersummen, dann heißt die Funktion g integrierbar, und dieser gemeinsame Wert wird mit $\int_a^b g(x) dF(x)$ bezeichnet. Wir wollen gar nicht versuchen zu bestimmen, welche Funktionen integrierbar sind. Ist $g : [a, b] \rightarrow \mathbb{R}$ stetig, dann kann man jedoch die Existenz von $\int_a^b g(x) dF(x)$ genauso wie für das Riemannintegral zeigen. Durch Grenzübergänge $b \rightarrow \infty$ und $a \rightarrow -\infty$ kann man Integrale über unbeschränkte Intervalle definieren. Hat F eine Dichte f , dann kann man zeigen, dass $\int_a^b g(x) dF(x) = \int_a^b g(x) f(x) dx$ gilt. Ist F die Verteilungsfunktion einer diskreten Zufallsvariable mit Wertebereich R und Einzelwahrscheinlichkeiten $w(k)$, dann gilt $\int_a^b g(x) dF(x) = \sum_{k \in R \cap (a, b]} g(k) w(k)$.

Mit Hilfe des Riemann-Stieltjes-Integrals kann man den Erwartungswert einer Zufallsvariablen X mit Verteilungsfunktion F durch $E(X) = \int_{-\infty}^{\infty} x dF(x)$ definieren, wobei vorausgesetzt wird, dass $\int_{-\infty}^{\infty} |x| dF(x) < \infty$ gilt.

Ist $F : \mathbb{R}^n \rightarrow [0, 1]$ eine n -dimensionale Verteilungsfunktion, dann kann man analog ein Integral $\int_B g(x_1, \dots, x_n) dF(x_1, \dots, x_n)$ definieren, wobei B eine integrierbare Teilmenge des \mathbb{R}^n ist. Anstelle der Fläche eines Rechtecks oder des Volumens eines Quaders nimmt man wieder die entsprechende Wahrscheinlichkeit. Für $n = 2$ und das Rechteck $(r_1, s_1] \times (r_2, s_2]$ ist das $F(s_1, s_2) - F(s_1, r_2) - F(r_1, s_2) + F(r_1, r_2)$. Wenn $F(t_1, t_2, \dots, t_n) = F_1(t_1) F_2(t_2) \dots F_n(t_n)$ für alle $t_1, t_2, \dots, t_n \in \mathbb{R}$ gilt, dann lässt sich $\int_{I_1 \times I_2 \times \dots \times I_n} g(x_1, \dots, x_n) dF(x_1, \dots, x_n)$ als iteriertes Integral $\int_{I_1} \int_{I_2} \dots \int_{I_n} g(x_1, \dots, x_n) dF_n(x_n) \dots dF_2(x_2) dF_1(x_1)$ schreiben.

Mit diesem mehrdimensionalen Riemann-Stieltjes-Integral lassen sich dann Satz 25 und die Sätze über den Erwartungswert allgemeiner beweisen, insbesondere hat man die Formel $E(\psi(Y)) = \int_{\mathbb{R}^n} \psi(x_1, \dots, x_n) dF(x_1, \dots, x_n)$, wobei $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig ist und F die Verteilungsfunktion des n -dimensionalen Zufallsvektors Y ist (der Erwartungswert existiert, wenn $\int_{\mathbb{R}^n} |\psi(x_1, \dots, x_n)| dF(x_1, \dots, x_n) < \infty$ gilt). Man erhält die allgemeinen Beweise, indem man überall $f(x) dx$ durch $dF(x)$, $f_n(x) dx$ durch $dF_n(x)$, $g(x, y) dx dy$ durch $dG(x, y)$ und so weiter ersetzt. Auch alle folgenden Beweise werden nur für den Fall, dass eine Wahrscheinlichkeitsdichte existiert, geführt werden.

Es folgen einige Hilfssätze über Integrale, die wir später brauchen werden. Für eine komplexwertige Funktion f definieren wir $f' = f'_1 + i f'_2$ und $\int f(x) dx = \int f_1(x) dx + i \int f_2(x) dx$, wobei f_1 der Real- und f_2 der Imaginärteil von f ist.

Hilfssatz A: Sei $f : \mathbb{R} \rightarrow \mathbb{R}^+$ eine Wahrscheinlichkeitsdichte und I ein offenes Intervall in \mathbb{R} . Sei $u : I \times \mathbb{R} \rightarrow \mathbb{C}$ eine stetig differenzierbare Funktion und $v(t) = \int_{\mathbb{R}} u(t, x) f(x) dx$ existiere für alle $t \in I$. Wenn für jeden Punkt $t \in I$ eine Umgebung $U_t \subset I$ und eine integrierbare Funktion $\gamma : \mathbb{R} \rightarrow \mathbb{R}^+$ existieren mit $\int_{\mathbb{R}} \gamma(x) f(x) dx < \infty$ und $|\frac{\partial u}{\partial s}(s, x)| \leq \gamma(x)$ für alle $s \in U_t$ und $x \in \mathbb{R}$, dann gilt $v'(t) = \int_{\mathbb{R}} \frac{\partial u}{\partial t}(t, x) f(x) dx$ für alle $t \in I$.

Beweis: Wir führen den Beweis zuerst für reellwertiges u , das heißt $u : I \times \mathbb{R} \rightarrow \mathbb{R}$. Sei $\alpha > 0$ so gewählt, dass $[t - \alpha, t + \alpha] \subset U_t$ ist. Für $|h| \leq \alpha$ gilt dann $\frac{v(t+h)-v(t)}{h} = \int_{\mathbb{R}} w_h(x) f(x) dx$ mit $w_h(x) = \frac{u(t+h,x)-u(t,x)}{h} = \frac{\partial u}{\partial t}(t + \xi_h, x)$ für ein ξ_h zwischen 0 und h . Wegen $|h| \leq \alpha$ gilt auch $|\xi_h| \leq \alpha$ und somit $t + \xi_h \in U_t$, sodass $|w_h(x)| \leq \gamma(x)$ für alle $x \in \mathbb{R}$ und $|h| \leq \alpha$ nach Voraussetzung folgt. Ebenso folgt $|\frac{\partial u}{\partial t}(t, x)| \leq \gamma(x)$ für alle $x \in \mathbb{R}$.

Sei $\varepsilon > 0$ vorgegeben. Wegen $\int_{\mathbb{R}} \gamma(x) f(x) dx < \infty$ existiert ein kompaktes Intervall $K \subset \mathbb{R}$ mit $\int_{\mathbb{R} \setminus K} \gamma(x) f(x) dx < \frac{\varepsilon}{3}$. Daher gilt für $|h| \leq \alpha$

$$\begin{aligned} \left| \int_{\mathbb{R} \setminus K} w_h(x) f(x) dx \right| &\leq \int_{\mathbb{R} \setminus K} |w_h(x)| f(x) dx \leq \int_{\mathbb{R} \setminus K} \gamma(x) f(x) dx < \frac{\varepsilon}{3} \quad \text{und} \\ \left| \int_{\mathbb{R} \setminus K} \frac{\partial u}{\partial t}(t, x) f(x) dx \right| &\leq \int_{\mathbb{R} \setminus K} \left| \frac{\partial u}{\partial t}(t, x) \right| f(x) dx \leq \int_{\mathbb{R} \setminus K} \gamma(x) f(x) dx < \frac{\varepsilon}{3} \end{aligned}$$

Auf der Menge $[t - \alpha, t + \alpha] \times K$ ist die Funktion $\frac{\partial u}{\partial t}$ gleichmäßig stetig. Es existiert ein $\delta \in (0, \alpha]$, sodass $|\frac{\partial u}{\partial t}(r, x) - \frac{\partial u}{\partial t}(s, y)| < \frac{\varepsilon}{3}$ gilt, wenn die beiden Punkte (r, x) und (s, y) in $[t - \alpha, t + \alpha] \times K$ liegen und Abstand $< \delta$ haben. Für $x \in K$ und $|h| < \delta$ liegen die Punkte (t, x) und $(t + \xi_h, x)$ in $[t - \alpha, t + \alpha] \times K$ und haben Abstand $|\xi_h|$, der $< \delta$ ist. Daher gilt $|\frac{\partial u}{\partial t}(t + \xi_h, x) - \frac{\partial u}{\partial t}(t, x)| < \frac{\varepsilon}{3}$ und somit auch

$$\left| \int_K w_h(x) f(x) dx - \int_K \frac{\partial u}{\partial t}(t, x) f(x) dx \right| \leq \int_K |w_h(x) - \frac{\partial u}{\partial t}(t, x)| f(x) dx \leq \frac{\varepsilon}{3} \int_K f(x) dx \leq \frac{\varepsilon}{3}$$

Setzt man das zusammen, so erhält man für $|h| < \delta$

$$\begin{aligned} \left| \int_{\mathbb{R}} w_h(x) f(x) dx - \int_{\mathbb{R}} \frac{\partial u}{\partial t}(t, x) f(x) dx \right| &\leq \left| \int_K w_h(x) f(x) dx - \int_K \frac{\partial u}{\partial t}(t, x) f(x) dx \right| \\ &\quad + \left| \int_{\mathbb{R} \setminus K} w_h(x) f(x) dx \right| + \left| \int_{\mathbb{R} \setminus K} \frac{\partial u}{\partial t}(t, x) f(x) dx \right| < \varepsilon \end{aligned}$$

Wir haben zu jedem $\varepsilon > 0$ ein $\delta > 0$ gefunden mit $|\frac{v(t+h)-v(t)}{h} - \int_{\mathbb{R}} \frac{\partial u}{\partial t}(t, x) f(x) dx| < \varepsilon$ für alle $h \in (-\delta, \delta)$. Damit ist $v'(t) = \lim_{h \rightarrow 0} \frac{v(t+h)-v(t)}{h} = \int_{\mathbb{R}} \frac{\partial u}{\partial t}(t, x) f(x) dx$ gezeigt.

Sei jetzt u komplexwertig, das heißt $u : I \times \mathbb{R} \rightarrow \mathbb{C}$, und u_1 und u_2 seien Real- und Imaginärteil von u . Es gilt dann $\frac{\partial u}{\partial s}(s, x) = \frac{\partial u_1}{\partial s}(s, x) + i \frac{\partial u_2}{\partial s}(s, x)$, woraus $|\frac{\partial u_1}{\partial s}(s, x)| \leq |\frac{\partial u}{\partial s}(s, x)| \leq \gamma(x)$ und $|\frac{\partial u_2}{\partial s}(s, x)| \leq |\frac{\partial u}{\partial s}(s, x)| \leq \gamma(x)$ für alle $s \in U_t$ und alle $x \in \mathbb{R}$ folgt. Somit sind die Voraussetzungen des Hilfssatzes für die beiden reellwertigen Funktionen u_1 und u_2 erfüllt. Seien $v_1(t) = \int_{\mathbb{R}} u_1(t, x) f(x) dx$ und $v_2(t) = \int_{\mathbb{R}} u_2(t, x) f(x) dx$. Dann folgt aus obigem Beweis, dass $v_1'(t) = \int_{\mathbb{R}} \frac{\partial u_1}{\partial t}(t, x) f(x) dx$ und $v_2'(t) = \int_{\mathbb{R}} \frac{\partial u_2}{\partial t}(t, x) f(x) dx$ gilt. Multipliziert man die zweite Gleichung mit i und addiert sie zur ersten, so hat man $v'(t) = \int_{\mathbb{R}} \frac{\partial u}{\partial t}(t, x) f(x) dx$. \square

Sei $a < b$. In den nächsten beiden Hilfssätzen behandeln wir die Funktion $\ell_{a,b} : \mathbb{R} \rightarrow [0, 1]$ definiert durch $\ell_{a,b}(x) = 1$ für $x < a$, durch $\ell_{a,b}(x) = \frac{b-x}{b-a}$ für $a \leq x \leq b$ und durch $\ell_{a,b}(x) = 0$ für $x > b$. Sie ist also stetig, gleich 1 links von a , gleich 0 rechts von b und linear auf $[a, b]$.

Hilfssatz B: Sei $a < b$ und $\ell_{a,b}$ wie soeben definiert. Für jedes $\varepsilon > 0$ und jedes $r > 0$ existiert dann ein trigonometrisches Polynom $\varrho(x) = \sum_{j=1}^m c_j \sin \alpha_j x + \sum_{j=0}^m d_j \cos \alpha_j x$, sodass $|\int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx - \int_{-\infty}^{\infty} \varrho(x) f(x) dx| \leq \int_{-\infty}^{\infty} |\ell_{a,b}(x) - \varrho(x)| f(x) dx \leq \int_{\mathbb{R} \setminus (-r, r]} f(x) dx + \varepsilon$ für alle Wahrscheinlichkeitsdichten f gilt.

Beweis: Wir wählen ein $\alpha > 0$, sodass $\frac{\pi}{\alpha} > r$ gilt. Wir definieren $\tilde{\ell} : \mathbb{R} \rightarrow [0, 1]$ folgendermaßen. Für $x \in [-r, r]$ sei $\tilde{\ell}(x) = \ell_{a,b}(x)$. Auf $[-\frac{\pi}{\alpha}, -r]$ sei $\tilde{\ell}$ linear mit $\tilde{\ell}(-r) = \ell_{a,b}(-r)$ und $\tilde{\ell}(-\frac{\pi}{\alpha}) = 1$. Auf $[r, \frac{\pi}{\alpha}]$ sei $\tilde{\ell}$ ebenfalls linear mit $\tilde{\ell}(r) = \ell_{a,b}(r)$ und $\tilde{\ell}(\frac{\pi}{\alpha}) = 1$. Damit ist $\tilde{\ell}$ auf $[-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}]$ definiert und stetig und hat in den Endpunkten jeweils den Wert 1. Wir können $\tilde{\ell}$ daher zu einer stetigen Funktion mit Periode $\frac{2\pi}{\alpha}$ auf ganz \mathbb{R} fortsetzen.

Man kann dann zeigen, dass ein trigonometrisches Polynom ϱ , wie oben angegeben, existiert mit $|\tilde{\ell}(x) - \varrho(x)| < \varepsilon$ für alle $x \in \mathbb{R}$. Da die Funktionen $\tilde{\ell}$ und $\ell_{a,b}$ beide Wertebereich

$[0, 1]$ haben und auf $(-r, r]$ übereinstimmen, erhalten wir $|\ell_{a,b}(x) - \tilde{\ell}(x)| \leq 1_{\mathbb{R} \setminus (-r, r]}(x)$ für alle $x \in \mathbb{R}$. Mit Hilfe der Dreiecksungleichung folgt $|\ell_{a,b}(x) - \varrho(x)| \leq 1_{\mathbb{R} \setminus (-r, r]}(x) + \varepsilon$ für alle $x \in \mathbb{R}$. Sei f irgendeine Wahrscheinlichkeitsdichte. Durch Multiplikation mit $f(x)$ und Integration erhalten wir $\int_{-\infty}^{\infty} |\ell_{a,b}(x) - \varrho(x)| f(x) dx \leq \int_{\mathbb{R} \setminus (-r, r]} f(x) dx + \varepsilon$. \square

Hilfssatz C: Sei $a < b$ und $\ell_{a,b}$ wie oben definiert. Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion, die eine Wahrscheinlichkeitsdichte f besitzt. Dann gilt $F(a) \leq \int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx \leq F(b)$ und $\lim_{b \downarrow a} \int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx = F(a)$.

Beweis: Aus der Definition von $\ell_{a,b}$ folgt $1_{(-\infty, a]}(x) \leq \ell_{a,b}(x) \leq 1_{(-\infty, b]}(x)$ für alle $x \in \mathbb{R}$ und daher $F(a) = \int_{-\infty}^{\infty} 1_{(-\infty, a]}(x) f(x) dx \leq \int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx \leq \int_{-\infty}^{\infty} 1_{(-\infty, b]}(x) f(x) dx = F(b)$. Das ist die erste Aussage. Aus dieser folgt $0 \leq \int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx - F(a) \leq F(b) - F(a)$. Da F rechtsseitig stetig ist und daher $\lim_{b \downarrow a} F(b) = F(a)$ gilt, erhalten wir auch die zweite Aussage. \square

2. Momenterzeugende Funktion (Laplace-Transformation)

Für eine Zufallsvariable X sei D_X die Menge aller $t \in \mathbb{R}$, für die $E(e^{tX})$ existiert. Die Funktion $m_X : D_X \rightarrow \mathbb{R}$ definiert durch $m_X(t) = E(e^{tX})$ heißt die momenterzeugende Funktion der Zufallsvariablen X . Hat X eine Dichte f , dann gilt $m_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$ für $t \in D_X$ nach Satz 36. Man nennt m_X auch die Laplace-Transformierte der Funktion f .

Satz 49: Sei X eine Zufallsvariable. Dann ist D_X ein Intervall und $0 \in D_X$ (auch $D_X = \{0\}$ ist möglich). Weiters ist m_X im Innern von D_X unendlich oft differenzierbar. Für $t \in \text{int } D_X$ und $k \geq 1$ existiert $E(X^k e^{tX})$ und es gilt $m_X^{(k)}(t) = E(X^k e^{tX})$.

Beweis: Für $t = 0$ ist e^{tX} die Konstante 1, deren Erwartungswert existiert und gleich 1 ist. Somit gilt $0 \in D_X$ und $m_X(0) = 1$. Seien a und b in D_X mit $a \leq b$. Ist $t \in [a, b]$, dann gilt $e^{tx} \leq e^{ax} + e^{bx}$ für alle $x \in \mathbb{R}$. Ist f eine Wahrscheinlichkeitsdichte von X , so folgt $\int_{-\infty}^{\infty} e^{tx} f(x) dx \leq \int_{-\infty}^{\infty} e^{ax} f(x) dx + \int_{-\infty}^{\infty} e^{bx} f(x) dx < \infty$, da $E(e^{aX})$ und $E(e^{bX})$ existieren, sodass auch $E(e^{tX})$ existiert und t in D_X liegt. Mit zwei Punkten enthält D_X auch alle Punkte, die dazwischen liegen. Das beweist, dass D_X ein Intervall ist.

Für $t \in \text{int } D_X$ zeigen wir, dass $E(X^k e^{tX})$ existiert und dass $m_X^{(k)}(t) = E(X^k e^{tX}) = \int_0^{\infty} x^k e^{tx} f(x) dx$ gilt. Für $k = 0$ ist das die Definition von D_X und m_X . Wir nehmen an, dass der Beweis für $k = n-1$ schon erbracht ist. Wir wählen $\alpha, a, b, \beta \in D_X$ mit $\alpha < a < t < b < \beta$ und setzen $c = (\frac{n}{a-\alpha})^n$ und $d = (\frac{n}{\beta-b})^n$. Man kann zeigen, dass $|x|^n e^{sx} \leq ce^{\alpha x} + de^{\beta x}$ für alle $x \in \mathbb{R}$ und alle $s \in [a, b]$ gilt. Sei $\gamma(x) = ce^{\alpha x} + de^{\beta x}$. Wegen $\alpha \in D_X$ und $\beta \in D_X$ gilt $\int_{-\infty}^{\infty} \gamma(x) f(x) dx < \infty$. Es folgt $\int_{-\infty}^{\infty} |x|^n e^{tx} f(x) dx < \infty$, sodass $E(X^n e^{tX})$ existiert. Wir wenden Hilfssatz A auf $u(s, x) = x^{n-1} e^{sx}$ mit $s \in I = \text{int } D_X$ und $x \in \mathbb{R}$ an. Wählt man $U_t = (a, b)$ und γ wie oben, dann sind die Voraussetzungen von Hilfssatz A erfüllt. Nach Induktionsvoraussetzung gilt $m_X^{(n-1)}(t) = \int_{-\infty}^{\infty} x^{n-1} e^{tx} f(x) dx$. Nach Hilfssatz A können wir unter dem Integral differenzieren und erhalten $m_X^{(n)}(t) = \int_{-\infty}^{\infty} x^n e^{tx} f(x) dx = E(X^n e^{tX})$, die Formel für $k = n$, womit die Induktion beendet ist. \square

Wir definieren die Momente einer Zufallsvariable und verwenden die momenterzeugende Funktion, um diese zu berechnen. Untenstehender Satz folgt unmittelbar aus Satz 49.

Definition: Für eine Zufallsvariable X und $k \geq 1$ nennt man $E(X^k)$ das k -te Moment der Zufallsvariablen X , falls dieser Erwartungswert existiert.

Satz 50: Sei X eine Zufallsvariable, für die 0 im Innern von D_X liegt. Dann existieren alle Momente und es gilt $E(X^k) = m_X^{(k)}(0)$ für $k \geq 1$.

Wir berechnen die momenterzeugende Funktion für konkrete Verteilungen und wenden darauf Satz 50 an.

Beispiel 21: Für eine $B(n, p)$ -verteilte Zufallsvariable X sind m_X , $E(X)$ und $V(X)$ gesucht.

Mit Hilfe des binomischen Lehrsatzes erhalten wir

$$m_X(t) = \sum_{j=0}^n e^{tj} \binom{n}{j} p^j (1-p)^{n-j} = (e^t p + 1 - p)^n$$

sodass m_X auf ganz \mathbb{R} definiert ist. Berechnet man die ersten beiden Ableitungen, so erhält man $m_X'(0) = np$ und $m_X''(0) = n^2 p^2 - np^2 + np$. Aus Satz 50 folgt $E(X) = np$ und $E(X^2) = n^2 p^2 - np^2 + np$. Damit ist auch $V(X) = E(X^2) - E(X)^2 = np(1-p)$ berechnet.

Beispiel 22: Für eine $G(r, \lambda)$ -verteilte Zufallsvariable X sind m_X , $E(X)$ und $V(X)$ gesucht.

Durch Einführen der neuen Variable $y = (\lambda - t)x$ und wegen $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy$ folgt

$$\begin{aligned} m_X(t) &= \int_0^\infty e^{tx} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty \left(\frac{y}{\lambda-t}\right)^{r-1} e^{-y} \frac{1}{\lambda-t} dy = \left(\frac{\lambda}{\lambda-t}\right)^r \frac{1}{\Gamma(r)} \int_0^\infty y^{r-1} e^{-y} dy = \left(\frac{\lambda}{\lambda-t}\right)^r \end{aligned}$$

In diesem Fall ist $m_X(t)$ für $t \in (-\infty, \lambda)$ definiert. Satz 50 ergibt $E(X) = m_X'(0) = \frac{r}{\lambda}$ und $E(X^2) = m_X''(0) = \frac{r^2 + r}{\lambda^2}$, woraus dann $V(X) = E(X^2) - E(X)^2 = \frac{r}{\lambda^2}$ folgt.

3. Charakteristische Funktion (Fouriertransformation)

Für eine Zufallsvariable X wird die charakteristische Funktion $h_X : \mathbb{R} \rightarrow \mathbb{C}$ definiert durch $h_X(t) = E(e^{itX}) = E(\cos tX) + iE(\sin tX)$. Hat X eine Wahrscheinlichkeitsdichte f , dann gilt $h_X(t) = \int_{-\infty}^\infty f(x) \cos tx dx + i \int_{-\infty}^\infty f(x) \sin tx dx$. Man nennt h_X auch die Fouriertransformierte der Funktion f . Da \sin und \cos beschränkte Funktionen sind, existiert h_X für alle Zufallsvariablen X . Es gilt $h_X(0) = \int_{-\infty}^\infty f(x) dx = 1$.

Beispiel 23: Zu berechnen ist h_X für eine $N(0, 1)$ -verteilte Zufallsvariable X .

Eine Dichte von X ist $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Da $x \mapsto \varphi(x) \sin tx$ eine ungerade Funktion ist, erhalten wir $\int_0^\infty \varphi(x) \sin tx dx = -\int_{-\infty}^0 \varphi(x) \sin tx dx$, woraus $\int_{-\infty}^\infty \varphi(x) \sin tx dx = 0$ folgt, sodass $h_X(t) = E(\cos tX) = \int_{-\infty}^\infty \varphi(x) \cos tx dx$ gilt. Wir schreiben dafür $h(t)$.

Wir wenden Hilfssatz A mit $I = \mathbb{R}$ und $\gamma(x) = |x|$ an. Es gilt ja $\int_{-\infty}^\infty \gamma(x) \varphi(x) dx < \infty$ und $|\frac{\partial}{\partial t} \cos tx| = |x \sin tx| \leq \gamma(x)$ für alle $t \in \mathbb{R}$. Wir erhalten $h'(t) = -\int_{-\infty}^\infty \varphi(x) x \sin tx dx$. Man rechnet $-x\varphi(x) = \varphi'(x)$ nach. Setzt man das ein und integriert partiell, so ergibt sich $h'(t) = -t \int_{-\infty}^\infty \varphi(x) \cos tx dx = -th(t)$. Sei $g(t) = e^{\frac{1}{2}t^2} h(t)$. Wegen $h'(t) = -th(t)$ folgt $g'(t) = 0$ für alle $t \in \mathbb{R}$. Daher ist g eine konstante Funktion. Weiters gilt $g(0) = h(0) = \int_{-\infty}^\infty \varphi(x) dx = 1$, sodass wir $g(t) = 1$ für alle t erhalten. Damit ist $h_X(t) = e^{-\frac{1}{2}t^2}$ berechnet.

Charakteristische Funktionen sind gleichmäßig stetig. Das besagt der folgende Hilfssatz, wenn man die Funktion g identisch 1 wählt.

Hilfssatz D: Sei f eine Wahrscheinlichkeitsdichte und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine integrierbare Funktion mit $\int_{-\infty}^\infty |g(x)| f(x) dx < \infty$. Dann ist die Funktion $t \mapsto \int_{-\infty}^\infty e^{itx} g(x) f(x) dx$ gleichmäßig stetig.

Beweis: Sei $u(t) = \int_{-\infty}^{\infty} \sin tx g(x)f(x) dx$ und $\varepsilon > 0$. Wegen $\int_{-\infty}^{\infty} |g(x)|f(x) dx = a < \infty$ existiert ein $c > 0$ mit $\int_V |g(x)|f(x) dx < \frac{\varepsilon}{4}$, wobei V für $\mathbb{R} \setminus [-c, c]$ steht. Sei $\delta = \frac{\varepsilon}{2ca}$. Da die Ableitung von \sin betragsmäßig ≤ 1 ist, erhalten wir $|\sin tx - \sin sx| \leq |tx - sx| \leq c|t - s|$, wenn x im Intervall $[-c, c]$ liegt. Für $s, t \in \mathbb{R}$ mit $|t - s| < \delta$ gilt dann

$$|u(t) - u(s)| \leq \int_V |g(x) \sin tx| f(x) dx + \int_V |g(x) \sin sx| f(x) dx + \int_{-c}^c |\sin tx - \sin sx| |g(x)| f(x) dx \\ \leq 2 \int_V |g(x)| f(x) dx + c|t - s| \int_{-c}^c |g(x)| f(x) dx \leq 2 \frac{\varepsilon}{4} + c\delta a = \varepsilon$$

Somit ist u gleichmäßig stetig. Genauso zeigt man, dass $v(t) = \int_{-\infty}^{\infty} \cos tx g(x)f(x) dx$ gleichmäßig stetig ist. Also ist auch $t \mapsto v(t) + iu(t) = \int_{-\infty}^{\infty} e^{itx} g(x)f(x) dx$ gleichmäßig stetig. \square

Der folgende Satz liefert wieder eine Methode zum Berechnen der Momente und wird auch zum Beweis des zentralen Grenzwertsatzes benötigt.

Satz 51: Sei X eine Zufallsvariable und $k \geq 1$. Wenn $E(X^k)$ existiert, dann ist h_X eine k Mal stetig differenzierbare Funktion und es gilt $E(X^m) = \frac{1}{i^m} h_X^{(m)}(0)$ für $m \leq k$.

Beweis: Sei f eine Dichte von X . Wir zeigen $h_X^{(m)}(t) = \int_{-\infty}^{\infty} i^m x^m e^{itx} f(x) dx$ für alle $t \in \mathbb{R}$ und $0 \leq m \leq k$ durch Induktion. Für $m = 0$ ist die zu beweisende Gleichung die Definition von h_X . Sei $0 \leq m < k$ und $h_X^{(m)}(t) = \int_{-\infty}^{\infty} (ix)^m e^{itx} f(x) dx$ schon gezeigt. Wir wenden Hilfssatz A für $u(s, x) = (ix)^m e^{isx}$ mit $s \in I = \mathbb{R}$ und $x \in \mathbb{R}$ an. Ist $t \in I$ und wählt man $U_t = \mathbb{R}$ und $\gamma(x) = |x^{m+1}|$, dann sind die Voraussetzungen von Hilfssatz A erfüllt, da die Existenz von $E(X^{m+1})$ aus Satz 40 folgt, das heißt $\int_{-\infty}^{\infty} |x^{m+1}| f(x) dx < \infty$, und wegen $\frac{\partial}{\partial s} e^{isx} = \frac{\partial}{\partial s} \cos sx + i \frac{\partial}{\partial s} \sin sx = -x \sin sx + ix \cos sx = ix e^{isx}$ auch $\frac{\partial}{\partial s} u(s, x) = (ix)^{m+1} e^{isx}$ gilt. Wir können unter dem Integral differenzieren. Es folgt $h_X^{(m+1)}(t) = \int_{-\infty}^{\infty} (ix)^{m+1} e^{itx} f(x) dx$ und die Induktion ist beendet.

Wir haben $h_X^{(m)}(t) = i^m \int_{-\infty}^{\infty} x^m e^{itx} f(x) dx$ für $0 \leq m \leq k$ gezeigt. Da $\int_{-\infty}^{\infty} |x^k| f(x) dx < \infty$ vorausgesetzt wird, ist $h_X^{(k)}$ wegen Hilfssatz D stetig, sodass h_X eine k Mal stetig differenzierbare Funktion ist. Man erhält $h_X^{(m)}(0) = i^m \int_{-\infty}^{\infty} x^m f(x) dx = i^m E(X^m)$ für $0 \leq m \leq k$, indem man $t = 0$ setzt in obenstehender Gleichung. \square

Beispiel 24: Sei X eine $N(0, 1)$ -verteilte Zufallsvariable. Gesucht sind die Momente von X .

Es gilt $h_X(t) = e^{-\frac{1}{2}t^2} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} t^{2k} = \sum_{m=0}^{\infty} c_m t^m$ mit $c_m = 0$ für ungerades m und $c_m = \frac{(-1)^k}{2^k k!}$ für $m = 2k$. Es folgt $h_X^{(m)}(0) = m! c_m$ für $m \geq 0$. Da $\int_{-\infty}^{\infty} |x|^m \varphi(x) dx < \infty$ für alle m gilt und daher $E(X^m)$ für alle m existiert, erhalten wir $E(X^m) = 0$ für ungerades m und $E(X^m) = \frac{(2k)!}{2^k k!}$ für $m = 2k$ mit Hilfe von Satz 51. Insbesondere folgt $E(X) = 0$, $V(X) = E(X^2) = 1$, $E(X^3) = 0$ und $E(X^4) = 3$.

Es folgen Rechenregeln für die charakteristische Funktion und ein Eindeutigkeitssatz.

Satz 52: Es gilt $h_{X_1+X_2+\dots+X_n}(t) = h_{X_1}(t)h_{X_2}(t)\dots h_{X_n}(t)$ für unabhängige Zufallsvariable X_1, X_2, \dots, X_n . Für eine Zufallsvariable X und $a, b \in \mathbb{R}$ gilt $h_{aX+b}(t) = e^{itb} h_X(at)$.

Beweis: Sind X und Y unabhängig, dann sind wegen Satz 25 auch $f(X)$ und $g(Y)$ unabhängig, wenn f und g stetige Funktionen sind. Es folgt wegen Satz 37

$$h_{X+Y}(t) = E(\cos(tX + tY)) + iE(\sin(tX + tY)) \\ = E(\cos tX \cos tY - \sin tX \sin tY) + iE(\sin tX \cos tY + \cos tX \sin tY) \\ = E(\cos tX)E(\cos tY) - E(\sin tX)E(\sin tY) + iE(\sin tX)E(\cos tY) + iE(\cos tX)E(\sin tY) \\ = (E(\cos tX) + iE(\sin tX))(E(\cos tY) + iE(\sin tY)) = h_X(t)h_Y(t)$$

Das ist die erste Aussage für zwei Zufallsvariable. Wir beweisen sie für n Zufallsvariable mit Induktion. Es gelte $h_{X_1+\dots+X_{m-1}}(t) = h_{X_1}(t) \dots h_{X_{m-1}}(t)$. Da nach Voraussetzung und Satz 25 die beiden Zufallsvariablen $X = X_1 + \dots + X_{m-1}$ und $Y = X_m$ unabhängig sind, gilt auch $h_{X+Y}(t) = h_X(t)h_Y(t)$, sodass $h_{X_1+\dots+X_m}(t) = h_{X_1}(t) \dots h_{X_{m-1}}(t)h_{X_m}(t)$ gezeigt ist.

Es gilt $h_{aX+b}(t) = E(\cos(atX + bt)) + iE(\sin(atX + bt))$. Eine analoge Rechnung wie oben führt zu $h_{aX+b}(t) = (\cos bt + i \sin bt)h_X(at)$. Das ist die zweite Aussage. \square

Beispiel 25: Gesucht ist $h_X(t)$ für eine $N(\mu, \sigma)$ -verteilte Zufallsvariable X .

Sei $U = \frac{X-\mu}{\sigma}$. Dann hat U die $N(0, 1)$ -Verteilung, sodass $h_U(t) = e^{-\frac{1}{2}t^2}$ nach Beispiel 24 gilt. Aus Satz 52 erhalten wir $h_X(t) = h_{\sigma U + \mu}(t) = e^{it\mu} e^{-\frac{1}{2}\sigma^2 t^2}$.

Satz 53 (Eindeutigkeitssatz) Seien X und Y Zufallsvariable. Wenn $h_X = h_Y$ gilt, dann haben X und Y die gleiche Verteilung.

Beweis: Seien f und g Dichten von X und Y und F und G deren Verteilungsfunktionen. Seien a und b in \mathbb{R} mit $a < b$. Sei $\varepsilon > 0$ vorgegeben und r so gewählt, dass $\int_{\mathbb{R} \setminus (-r, r]} f(x) dx < \varepsilon$ und $\int_{\mathbb{R} \setminus (-r, r]} g(x) dx < \varepsilon$ gilt. Nach Hilfssatz B existiert ein trigonometrisches Polynom $\varrho(x) = \sum_{j=1}^m c_j \sin \alpha_j x + \sum_{j=0}^m d_j \cos \alpha_j x$ mit $|\int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx - \int_{-\infty}^{\infty} \varrho(x) f(x) dx| < 2\varepsilon$ und $|\int_{-\infty}^{\infty} \ell_{a,b}(x) g(x) dx - \int_{-\infty}^{\infty} \varrho(x) g(x) dx| < 2\varepsilon$. Aus der Voraussetzung $h_X = h_Y$ erhalten wir $\int_{-\infty}^{\infty} \cos \alpha_j x f(x) dx = \int_{-\infty}^{\infty} \cos \alpha_j x g(x) dx$ und $\int_{-\infty}^{\infty} \sin \alpha_j x f(x) dx = \int_{-\infty}^{\infty} \sin \alpha_j x g(x) dx$ für alle j . Daraus folgt $\int_{-\infty}^{\infty} \varrho(x) f(x) dx = \int_{-\infty}^{\infty} \varrho(x) g(x) dx$. Mit Hilfe der Dreiecksungleichung ergibt sich $|\int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx - \int_{-\infty}^{\infty} \ell_{a,b}(x) g(x) dx| < 4\varepsilon$. Da $\varepsilon > 0$ beliebig gewählt werden kann, ist $\int_0^{\infty} \ell_{a,b}(x) f(x) dx = \int_0^{\infty} \ell_{a,b}(x) g(x) dx$ gezeigt. Lässt man jetzt b von rechts gegen a gehen, dann erhält man $F(a) = G(a)$ aus Hilfssatz C. \square

Satz 54: Seien X_1, X_2, \dots, X_n unabhängige Zufallsvariable, und für $1 \leq j \leq n$ habe X_j die $N(\mu_j, \sigma_j)$ -Verteilung. Sei $Y = X_1 + X_2 + \dots + X_n$. Dann hat Y die $N(\mu, \sigma)$ -Verteilung mit $\mu = \mu_1 + \mu_2 + \dots + \mu_n$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$.

Beweis: Es gilt $h_{X_j}(t) = e^{it\mu_j} e^{-\frac{1}{2}t^2\sigma_j^2}$ für $1 \leq j \leq n$. Satz 52 liefert $h_Y(t) = e^{it\mu} e^{-\frac{1}{2}t^2\sigma^2}$ mit $\mu = \mu_1 + \mu_2 + \dots + \mu_n$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$. Das ist die charakteristische Funktion einer $N(\mu, \sigma)$ -verteilten Zufallsvariablen. Nach Satz 53 hat Y auch die $N(\mu, \sigma)$ -Verteilung. \square

4. Konvergenz in Verteilung

Im folgenden Kapitel soll der zentrale Grenzwertsatz behandelt werden. Er besagt, dass die in geeigneter Weise gebildeten Mittel einer Folge von Zufallsvariablen in Verteilung konvergieren. Zuerst müssen wir präzisieren, was Konvergenz in Verteilung heißt.

Definition: Seien X_n für $n \geq 1$ und X Zufallsvariablen mit Verteilungsfunktionen F_n und F . Man sagt, dass die Folge X_n in Verteilung gegen X konvergiert, wenn $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ für alle Punkte $t \in \mathbb{R}$ gilt, in denen F stetig ist.

Beispiel 26: Für $n \geq 1$ sei X_n eine Zufallsvariable mit Verteilungsfunktion $F_n = 1_{[\frac{1}{n}, \infty)}$ und X eine Zufallsvariable mit Verteilungsfunktion $F = 1_{[0, \infty)}$. Hat man Verteilungskonvergenz?

Ist $t \leq 0$, dann gilt $F_n(t) = 0$ für $n \geq 1$. Ist $t > 0$, dann gilt $F_n(t) = 1$ für $n \geq \frac{1}{t}$. Es folgt $\lim_{n \rightarrow \infty} F_n(t) = 1_{(0, \infty)}(t)$ für alle $t \in \mathbb{R}$. Wir haben also $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ für $t \in \mathbb{R} \setminus \{0\}$ und $\lim_{n \rightarrow \infty} F_n(0) \neq F(0)$. Da F im Punkt 0 unstetig ist, liegt Verteilungskonvergenz vor. Das wäre nicht der Fall, würde man in der Definition der Verteilungskonvergenz auch Konvergenz in den Unstetigkeitspunkten von F verlangen.

Wir geben verschiedene Charakterisierungen der Konvergenz in Verteilung. Dazu brauchen wir zwei Hilfssätze.

Hilfssatz E: Sei $c > 0$ und X eine Zufallsvariable mit Wahrscheinlichkeitsdichte f . Dann gilt $\int_{\mathbb{R} \setminus (-\frac{2}{c}, \frac{2}{c}]} f(x) dx \leq \frac{2}{c} \int_0^c 1 - E(\cos tX) dt$.

Beweis: Für $x \in (-\frac{2}{c}, \frac{2}{c}]$ gilt $1 - \frac{\sin cx}{cx} \geq 0$ wegen $|\sin y| \leq |y|$. Für $x \in \mathbb{R} \setminus (-\frac{2}{c}, \frac{2}{c}]$ gilt $1 - \frac{\sin cx}{cx} \geq \frac{1}{2}$ wegen $|cx| \geq 2$ und $|\sin cx| \leq 1$. Somit erhalten wir

$$\int_{-\infty}^{\infty} (1 - \frac{\sin cx}{cx}) f(x) dx \geq 0 + \frac{1}{2} \int_{\mathbb{R} \setminus (-\frac{2}{c}, \frac{2}{c}]} f(x) dx$$

Andererseits folgt mit Hilfe des Satzes von Fubini

$$\begin{aligned} \frac{2}{c} \int_0^c 1 - E(\cos tX) dt &= \frac{2}{c} \int_0^c \int_{-\infty}^{\infty} (1 - \cos tx) f(x) dx dt = \frac{2}{c} \int_{-\infty}^{\infty} \int_0^c (1 - \cos tx) dt f(x) dx \\ &= \frac{2}{c} \int_{-\infty}^{\infty} (c - \frac{\sin cx}{x}) f(x) dx = 2 \int_{-\infty}^{\infty} (1 - \frac{\sin cx}{cx}) f(x) dx \end{aligned}$$

Beides zusammen ergibt die gewünschte Abschätzung. \square

Der zweite Hilfssatz folgt aus dem Satz über dominierte Konvergenz (siehe Maßtheorie). Er wird hier nicht bewiesen.

Hilfssatz F: Sei $J \subset \mathbb{R}$ ein beschränktes Intervall. Seien g_n für $n \geq 1$ und g integrierbare Funktionen auf J , sodass $\lim_{n \rightarrow \infty} g_n(x) = g(x)$ für alle $x \in J$ gilt. Wenn ein $\gamma > 0$ existiert mit $|g_n(x)| \leq \gamma$ für alle $x \in J$ und alle $n \geq 1$, dann gilt $\lim_{n \rightarrow \infty} \int_J g_n(x) dx = \int_J g(x) dx$.

Satz 55: Für Zufallsvariable X_n mit $n \geq 1$ und X sind äquivalent

- (a) X_n konvergiert für $n \rightarrow \infty$ in Verteilung gegen X
- (b) $\lim_{n \rightarrow \infty} E(\psi(X_n)) = E(\psi(X))$ für alle stetigen beschränkten Funktionen $\psi : \mathbb{R} \rightarrow \mathbb{R}$
- (c) $\lim_{n \rightarrow \infty} h_{X_n}(t) = h_X(t)$ für alle $t \in \mathbb{R}$
- (d) $\lim_{n \rightarrow \infty} E(\ell_{a,b}(X_n)) = E(\ell_{a,b}(X))$ für alle a und b in \mathbb{R} mit $a < b$

Beweis: Wir zeigen nur (c) \Rightarrow (d) \Rightarrow (a), da wir nur diesen Teil des Satzes zum Beweis des zentralen Grenzwertsatzes brauchen. (Die Implikation (b) \Rightarrow (c) ist trivial, da $x \mapsto \cos tx$ und $x \mapsto \sin tx$ für jedes $t \in \mathbb{R}$ stetige beschränkte Funktionen sind.)

Für $n \geq 1$ sei f_n eine Wahrscheinlichkeitsdichte von X_n und F_n deren Verteilungsfunktion. Ebenso sei f eine Wahrscheinlichkeitsdichte von X und F deren Verteilungsfunktion.

Um zu zeigen, dass (d) aus (c) folgt, seien a und b in \mathbb{R} mit $a < b$ und $\varepsilon > 0$ vorgegeben. Sei $s > 0$ so gewählt, dass $\int_{\mathbb{R} \setminus (-s, s]} f(x) dx < \frac{\varepsilon}{9}$ gilt. Wir wählen $c \in (0, \frac{2}{s}]$ so, dass $\cos tx \geq 1 - \frac{\varepsilon}{9}$ für alle $x \in (-s, s]$ gilt, wenn $|t| \leq c$ ist. Für $|t| \leq c$ erhalten wir dann

$$1 - E(\cos tX) = \int_{-\infty}^{\infty} (1 - \cos tx) f(x) dx \leq \int_{-s}^s \frac{\varepsilon}{9} f(x) dx + \int_{\mathbb{R} \setminus (-s, s]} 2f(x) dx < \frac{\varepsilon}{9} + 2\frac{\varepsilon}{9} = \frac{\varepsilon}{3}$$

Da wir (c), das heißt $\lim_{n \rightarrow \infty} h_{X_n}(t) = h_X(t)$ für alle $t \in \mathbb{R}$ voraussetzen, haben wir auch $\lim_{n \rightarrow \infty} E(\cos tX_n) = E(\cos tX)$. Für alle $t \in \mathbb{R}$ und alle $n \geq 1$ gilt $-1 \leq \cos tX_n \leq 1$ und daher auch $-1 \leq E(\cos tX_n) \leq 1$ und $0 \leq 1 - E(\cos tX_n) \leq 2$. Aus Hilfssatz F mit $\gamma = 2$ folgt somit $\lim_{n \rightarrow \infty} \int_0^c 1 - E(\cos tX_n) dt = \int_0^c 1 - E(\cos tX) dt$. Wegen $\int_0^c 1 - E(\cos tX) dt \leq \frac{\varepsilon}{3}$ existiert ein n_1 mit $\int_0^c 1 - E(\cos tX_n) dt < \frac{\varepsilon}{2}$ für $n \geq n_1$. Es folgt $\int_{\mathbb{R} \setminus (-\frac{2}{c}, \frac{2}{c}]} f_n(x) dx < \varepsilon$ für alle $n \geq n_1$ aus Hilfssatz E. Oben wurde $c \leq \frac{2}{s}$ gewählt, woraus $s \leq \frac{2}{c}$ folgt und daher auch $\int_{\mathbb{R} \setminus (-\frac{2}{c}, \frac{2}{c}]} f(x) dx \leq \int_{\mathbb{R} \setminus (-s, s]} f(x) dx < \frac{\varepsilon}{9} < \varepsilon$. Wir wenden Hilfssatz B mit $r = \frac{2}{c}$ an.

Das liefert uns ein trigonometrisches Polynom $\varrho(x) = \sum_{j=1}^m c_j \sin \alpha_j x + \sum_{j=0}^m d_j \cos \alpha_j x$ mit $|\int_{-\infty}^{\infty} \ell_{a,b}(x) f(x) dx - \int_{-\infty}^{\infty} \varrho(x) f(x) dx| < 2\varepsilon$ und $|\int_{-\infty}^{\infty} \ell_{a,b}(x) f_n(x) dx - \int_{-\infty}^{\infty} \varrho(x) f_n(x) dx| < 2\varepsilon$ für alle $n \geq n_1$. Da wir $\lim_{n \rightarrow \infty} h_{X_n}(t) = h_X(t)$ für alle $t \in \mathbb{R}$ voraussetzen, gilt auch $\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varrho(x) f_n(x) dx = \int_{-\infty}^{\infty} \varrho(x) f(x) dx$. Somit existiert dann auch noch ein n_2 , sodass

$|\int_{-\infty}^{\infty} \varrho(x)f_n(x)dx - \int_{-\infty}^{\infty} \varrho(x)f(x)dx| < \varepsilon$ für alle $n \geq n_2$ gilt. Wir setzen $n_0 = \max(n_1, n_2)$. Für $n \geq n_0$ haben wir dann $|\int_{-\infty}^{\infty} \ell_{a,b}(x)f_n(x)dx - \int_{-\infty}^{\infty} \ell_{a,b}(x)f(x)dx| < 5\varepsilon$. Damit ist $\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \ell_{a,b}(x)f_n(x)dx = \int_{-\infty}^{\infty} \ell_{a,b}(x)f(x)dx$ gezeigt. Das ist (d).

Um zu zeigen, dass (a) aus (d) folgt, sei $\varepsilon > 0$ und $t \in \mathbb{R}$ ein Punkt, in dem F stetig ist. Es existieren $a < t$ und $b > t$ mit $F(t) - \varepsilon < F(a) \leq F(t) \leq F(b) < F(t) + \varepsilon$. Da (d) vorausgesetzt wird, existiert ein n_0 , sodass $|\int_{-\infty}^{\infty} \ell(x)f_n(x)dx - \int_{-\infty}^{\infty} \ell(x)f(x)dx| < \varepsilon$ für alle $n \geq n_0$ sowohl für $\ell = \ell_{t,b}$ als auch für $\ell = \ell_{a,t}$ gilt. Aus Hilfssatz C folgt für $n \geq n_0$

$$F_n(t) \leq \int_{-\infty}^{\infty} \ell_{t,b}(x)f_n(x)dx < \int_{-\infty}^{\infty} \ell_{t,b}(x)f(x)dx + \varepsilon \leq F(b) + \varepsilon < F(t) + 2\varepsilon$$

$$F_n(t) \geq \int_{-\infty}^{\infty} \ell_{a,t}(x)f_n(x)dx > \int_{-\infty}^{\infty} \ell_{a,t}(x)f(x)dx - \varepsilon \geq F(a) - \varepsilon > F(t) - 2\varepsilon$$

Somit gilt $|F_n(t) - F(t)| < 2\varepsilon$ für $n \geq n_0$. Damit ist $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ für alle $t \in \mathbb{R}$, in denen F stetig ist, gezeigt. Das ist (a). \square

5. Grenzwertsätze

Spielt man jeden Tag dasselbe Glücksspiel und sind X_1, X_2, \dots die dabei erzielten Gewinne, dann sind das unabhängige Zufallsvariable, die alle dieselbe Verteilung F haben. Der Durchschnittsgewinn der ersten n Tage ist $M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Wie verhält sich M_n , wenn man n gegen ∞ gehen lässt?

Da die Zufallsvariablen X_j alle dieselbe Verteilung haben, haben sie auch denselben Erwartungswert $m = \int_{-\infty}^{\infty} x dF(x)$ und dieselbe Varianz $v = \int_{-\infty}^{\infty} (x - m)^2 dF(x)$, falls diese existieren. Wir nennen die Zufallsvariablen X_1, X_2, \dots unabhängig, wenn die Zufallsvariablen X_1, X_2, \dots, X_n für jedes n unabhängig sind. Der folgende Satz ist das sogenannte schwache Gesetz der großen Zahl.

Satz 56: Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben und m ihr gemeinsamer Erwartungswert. Ihre gemeinsame Varianz v existiere. Sei $M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Für alle $\varepsilon > 0$ gilt dann $\lim_{n \rightarrow \infty} P(|M_n - m| > \varepsilon) = 0$.

Beweis: Aus Satz 37 folgt $E(M_n) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n}nm = m$ und ebenso $V(M_n) = \frac{1}{n^2}(V(X_1) + V(X_2) + \dots + V(X_n)) = \frac{1}{n^2}nv = \frac{v}{n}$ aus Satz 43 und Satz 44(b). Für jedes $\varepsilon > 0$ erhalten wir dann $P(|M_n - m| > \varepsilon) \leq \frac{V(M_n)}{\varepsilon^2} = \frac{v}{n\varepsilon^2}$ mit Hilfe von Satz 39. Daraus folgt $\lim_{n \rightarrow \infty} P(|M_n - m| > \varepsilon) = 0$. \square

Unter den selben Voraussetzungen wie in Satz 56 kann man auch $P(\lim_{n \rightarrow \infty} M_n = m) = 1$ zeigen. Diese Aussage, dass M_n für $n \rightarrow \infty$ mit Wahrscheinlichkeit 1 gegen m konvergiert, nennt man das starke Gesetz der großen Zahl.

Die Gesetze der großen Zahl besagen, dass die Folge $M_n - m$ für $n \rightarrow \infty$ in einem gewissen Sinn gegen 0 geht. Der zentrale Grenzwertsatz gibt genauere Auskunft über diese Konvergenz, indem die Konvergenz der Folge $\sqrt{n}(M_n - m)$ für $n \rightarrow \infty$ untersucht wird. Diese Folge konvergiert nicht mehr gegen 0, sondern im Grenzwert stellt sich eine Normalverteilung ein. Zum Beweis des zentralen Grenzwertsatzes brauchen wir einen Hilfssatz aus der Analysis.

Hilfssatz G: Sei $\alpha_1, \alpha_2, \dots$ eine Folge in \mathbb{C} mit $\lim_{n \rightarrow \infty} n\alpha_n = 0$. Für $x \in \mathbb{R}$ gilt dann $\lim_{n \rightarrow \infty} (1 + \frac{x}{n} + \alpha_n)^n = e^x$.

Beweis: Für $n \geq 1$ sei $w_n = (1 + \frac{x}{n} + \alpha_n)^n - (1 + \frac{x}{n})^n$. Wir zeigen $\lim_{n \rightarrow \infty} w_n = 0$. Dann ist auch $\lim_{n \rightarrow \infty} (1 + \frac{x}{n} + \alpha_n)^n = \lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n = e^x$ gezeigt. Es gilt

$$|w_n| = |\sum_{j=1}^n \binom{n}{j} \alpha_n^j (1 + \frac{x}{n})^{n-j}| \leq \sum_{j=1}^n \binom{n}{j} |\alpha_n|^j (1 + \frac{|x|}{n})^{n-j}$$

Da $\lim_{n \rightarrow \infty} n\alpha_n = 0$ vorausgesetzt wird, existiert ein $c > 0$ mit $|\alpha_n| \leq \frac{c}{n}$ für alle n . Weiters gilt $\binom{n}{j} = \frac{n}{j} \binom{n-1}{j-1} \leq n \binom{n-1}{j-1}$. Damit folgt

$$|w_n| \leq \left(1 + \frac{|x|}{n}\right)^n n |\alpha_n| \sum_{j=1}^n \binom{n-1}{j-1} \left(\frac{c}{n}\right)^{j-1} = \left(1 + \frac{|x|}{n}\right)^n n |\alpha_n| \left(1 + \frac{c}{n}\right)^{n-1}$$

Wegen $\lim_{n \rightarrow \infty} \left(1 + \frac{|x|}{n}\right)^n = e^{|x|}$, wegen $\lim_{n \rightarrow \infty} \left(1 + \frac{c}{n}\right)^{n-1} = e^c$ und wegen $\lim_{n \rightarrow \infty} n |\alpha_n| = 0$ erhalten wir $\lim_{n \rightarrow \infty} w_n = 0$. \square

Nun können wir den zentralen Grenzwertsatz beweisen. Ist $g : \mathbb{R} \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion, dann gilt $g(t) = g(0) + g'(0)t + g''(0)\frac{t^2}{2} + r(t)\frac{t^2}{2}$ mit $\lim_{t \rightarrow 0} r(t) = 0$ nach der Taylorformel. Ist $g : \mathbb{R} \rightarrow \mathbb{C}$ eine zweimal stetig differenzierbare Funktion, dann haben wir $g_1(t) = g_1(0) + g_1'(0)t + g_1''(0)\frac{t^2}{2} + r_1(t)\frac{t^2}{2}$ für den Realteil g_1 von g und ebenso $g_2(t) = g_2(0) + g_2'(0)t + g_2''(0)\frac{t^2}{2} + r_2(t)\frac{t^2}{2}$ für den Imaginärteil g_2 von g , wobei $\lim_{t \rightarrow 0} r_1(t) = 0$ und $\lim_{t \rightarrow 0} r_2(t) = 0$ gilt. Multipliziert man die zweite Gleichung mit i und addiert sie zur ersten, so erhält man $g(t) = g(0) + g'(0)t + g''(0)\frac{t^2}{2} + r(t)\frac{t^2}{2}$ mit $r(t) = r_1(t) + ir_2(t)$, das heißt $\lim_{t \rightarrow 0} r(t) = 0$. Somit gilt die Taylorformel auch für komplexwertige Funktionen.

Satz 57 (Zentraler Grenzwertsatz) Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben, und m ihr gemeinsamer Erwartungswert. Ihre gemeinsame Varianz v sei > 0 . Sei $M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Dann konvergiert die Folge $Z_n = \sqrt{n} \frac{M_n - m}{\sqrt{v}}$ für $n \rightarrow \infty$ in Verteilung gegen eine $N(0, 1)$ -verteilte Zufallsvariable.

Beweis: Sei $Y_k = \frac{X_k - m}{\sqrt{v}}$ für $k \geq 1$. Es folgt $E(Y_k) = 0$ aus Satz 37 und $E(Y_k^2) = V(Y_k) = 1$ aus Satz 43. Die Zufallsvariablen Y_k sind unabhängig wegen Satz 25 und haben alle dieselbe Verteilung wegen Satz 17. Es gilt $Z_n = \sqrt{n} \frac{M_n - m}{\sqrt{v}} = \frac{1}{\sqrt{nv}} \sum_{k=1}^n (X_k - m) = \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k$.

Wir berechnen die charakteristische Funktion von Z_n . Wegen Satz 51 ist h_{Y_k} zweimal stetig differenzierbar und es gilt $h'_{Y_k}(0) = iE(Y_k) = 0$ und $h''_{Y_k}(0) = -E(Y_k^2) = -1$. Aus der Taylorformel für die komplexwertige Funktion h_{Y_k} erhalten wir

$$h_{Y_k}(t) = h_{Y_k}(0) + h'_{Y_k}(0)t + h''_{Y_k}(0)\frac{t^2}{2} + r(t)\frac{t^2}{2} = 1 - \frac{1}{2}t^2 + r(t)\frac{t^2}{2}$$

mit $\lim_{t \rightarrow 0} r(t) = 0$. Weiters gilt $h_{Y_1 + Y_2 + \dots + Y_n}(t) = h_{Y_1}(t) \dots h_{Y_n}(t) = \left(1 - \frac{1}{2}t^2 + r(t)\frac{t^2}{2}\right)^n$ nach Satz 52. Wegen $Z_n = \frac{1}{\sqrt{n}}(Y_1 + Y_2 + \dots + Y_n)$ erhalten wir wieder aus Satz 52

$$h_{Z_n}(t) = h_{Y_1 + Y_2 + \dots + Y_n}\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{1}{2}\frac{t^2}{n} + \frac{t^2}{2n}r\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Für festes $t \in \mathbb{R}$ sei $\alpha_n = \frac{t^2}{2n}r\left(\frac{t}{\sqrt{n}}\right)$ für $n \geq 1$. Es folgt $\lim_{n \rightarrow \infty} n\alpha_n = \frac{t^2}{2} \lim_{n \rightarrow \infty} r\left(\frac{t}{\sqrt{n}}\right) = 0$ wegen $\lim_{s \rightarrow 0} r(s) = 0$. Mit Hilfe von Hilfssatz G erhalten wir nun

$$\lim_{n \rightarrow \infty} h_{Z_n}(t) = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{2}\frac{t^2}{n} + \alpha_n\right)^n = e^{-\frac{1}{2}t^2}$$

Damit ist gezeigt, dass die charakteristische Funktion von $Z_n = \sqrt{n} \frac{M_n - m}{\sqrt{v}}$ für $n \rightarrow \infty$ gegen die charakteristische Funktion einer $N(0, 1)$ -verteilten Zufallsvariable konvergiert. Wegen Satz 55 konvergiert dann die Folge Z_n für $n \rightarrow \infty$ in Verteilung gegen eine $N(0, 1)$ -verteilte Zufallsvariable. \square

Bemerkung: Nimmt man zusätzlich an, dass $\gamma = E(|X_k - m|^3)$ existiert, dann kann man $\sup_{t \in \mathbb{R}} |F_n(t) - \Phi(t)| \leq \frac{4}{5} \frac{\gamma}{\sqrt{n}\sqrt{v}^3}$ für alle n zeigen, wobei F_n die Verteilungsfunktion von $Z_n = \sqrt{n} \frac{M_n - m}{\sqrt{v}}$ und Φ die Verteilungsfunktion der $N(0, 1)$ -Verteilung ist. Man hat also sogar eine Fehlerabschätzung.

IV. Statistik

Zuerst werden die grundlegenden statistischen Methoden eingeführt, das sind Parameterschätzer, Konfidenzintervalle und statistische Tests. Dann werden die wohl wichtigsten statistischen Verfahren behandelt, die Varianzanalyse und die lineare Regression. Es gibt natürlich noch viele andere statistische Verfahren.

1. Parameterschätzer

Aus einer festgelegten Grundgesamtheit (population) wird eine Stichprobe gezogen. Mit Hilfe dieser Stichprobe sollen Aussagen über die Grundgesamtheit gemacht werden. Im einfachsten Fall ist das die Schätzung eines Parameters.

Zum Beispiel können wir die Menge aller erwachsenen Österreicher als Grundgesamtheit nehmen. Sei p der unbekannte Anteil der Raucher in der Grundgesamtheit. Das ist der Parameter, der geschätzt werden soll. Wir wählen zufällig n Personen aus der Grundgesamtheit, sodass jede Person in der Grundgesamtheit die gleiche Chance hat, gewählt zu werden. Sei $X_j = 0$, wenn die j -te Person kein Raucher ist, und $X_j = 1$, wenn die j -te Person Raucher ist. Die Stichprobe (sample) wird durch die Zufallsvariablen X_1, X_2, \dots, X_n beschrieben. Diese Zufallsvariablen sind unabhängig (Ziehen mit Zurücklegen). Als Schätzwert für den Anteil p der Raucher in der Grundgesamtheit verwenden wir den Anteil der Raucher in der Stichprobe. Da $X_1 + X_2 + \dots + X_n$ die Anzahl der Raucher in der Stichprobe ist, ist das $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$.

Als weiteres Beispiel wählen wir die Körpergröße. Die Grundgesamtheit sei wieder die Menge aller erwachsenen Österreicher (oder die aller Frauen). Sei μ der Mittelwert und σ^2 die Varianz der Körpergröße in der Grundgesamtheit. Diese beiden Parameter sollen geschätzt werden. Wir wählen zufällig n Personen aus der Grundgesamtheit. Die Stichprobe X_1, X_2, \dots, X_n gibt die Körpergrößen dieser n Personen an. Diese Zufallsvariablen sind unabhängig (zufälliges Ziehen). Als Schätzwert für den Mittelwert μ in der Grundgesamtheit (population mean) verwenden wir das Stichprobenmittel (sample mean) $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Als Schätzwert für die Varianz σ^2 in der Grundgesamtheit bietet sich die Stichprobenvarianz $\frac{1}{n} \sum_{j=1}^n (X_j - M)^2$ an.

Ein Schätzer (estimator) für einen Parameter θ ist eine Funktion $u(X_1, X_2, \dots, X_n)$, die aus der Stichprobe X_1, X_2, \dots, X_n berechnet wird. Er heißt erwartungstreu (unbiased), wenn $E(u(X_1, \dots, X_n)) = \theta$ gilt. In obigem Beispiel gilt $E(X_j) = \mu$ für $1 \leq j \leq n$, da ja jede Person in der Grundgesamtheit die gleiche Chance hat, gewählt zu werden. Es folgt $E(M) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n}n\mu = \mu$ aus den Rechenregeln für den Erwartungswert. Somit ist M ein erwartungstreuer Schätzer für den Mittelwert μ der Körpergröße in der Grundgesamtheit. Im Beispiel mit den Rauchern gilt $E(X_j) = p$ für $1 \leq j \leq n$, woraus wie oben $E(M) = p$ folgt. In diesem Fall ist M ein erwartungstreuer Schätzer für den Raucheranteil p in der Grundgesamtheit.

Die Wurzel der Varianz eines erwartungstreuen Schätzers heißt Standardfehler des Schätzers. Der Standardfehler ist ein Maß für die Abweichung des Schätzers vom zu schätzenden Parameter, also für die Genauigkeit der Schätzung. Ist σ^2 die Varianz in der Grundgesamtheit, dann gilt $V(M) = \frac{1}{n^2}(V(X_1) + V(X_2) + \dots + V(X_n)) = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2$, da die Stichprobe X_1, X_2, \dots, X_n aus unabhängigen Zufallsvariablen besteht und da $V(X_j) = \sigma^2$ für $1 \leq j \leq n$ gilt. Somit ist $\frac{\sigma}{\sqrt{n}}$ der Standardfehler des Schätzers M für den Mittelwert μ . Er geht mit wachsendem Stichprobenumfang n gegen 0.

Etwas komplizierter ist die Situation, wenn der zu schätzende Parameter die Varianz σ^2 in der Grundgesamtheit ist. Um zu überprüfen, ob die Stichprobenvarianz ein erwartungstreuer Schätzer für σ^2 ist, berechnen wir den Erwartungswert von $\sum_{k=1}^n (X_k - M)^2$.

$$\begin{aligned} \mathbb{E}(\sum_{j=1}^n (X_j - M)^2) &= \mathbb{E}(\sum_{j=1}^n ((X_j - \mu) - (M - \mu))^2) \\ &= \mathbb{E}(\sum_{j=1}^n (X_j - \mu)^2 - 2 \sum_{j=1}^n (X_j - \mu)(M - \mu) + \sum_{j=1}^n (M - \mu)^2) \\ &= \mathbb{E}(\sum_{j=1}^n (X_j - \mu)^2 - 2(nM - n\mu)(M - \mu) + n(M - \mu)^2) \\ &= \sum_{j=1}^n \mathbb{E}((X_j - \mu)^2) - \mathbb{E}(n(M - \mu)^2) \\ &= \sum_{j=1}^n \mathbb{V}(X_j) - n\mathbb{V}(M) = \sum_{j=1}^n \sigma^2 - n \frac{\sigma^2}{n} \\ &= (n - 1)\sigma^2 \end{aligned}$$

Um einen erwartungstreuen Schätzer zu erhalten, muss man $\frac{1}{n-1}$ vor die Summe schreiben und nicht $\frac{1}{n}$, was eigentlich naheliegender wäre. Das ist der Grund, warum man üblicherweise $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - M)^2$ als Schätzer für die Varianz σ^2 verwendet.

2. Konfidenzintervalle

Aus der Stichprobe X_1, X_2, \dots, X_n soll ein Intervall berechnet werden, das einen Parameter θ mit großer Wahrscheinlichkeit einschließt. Dazu gibt man ein $\gamma \in (0, 1)$ vor (übliche Werte für γ sind 0.95 oder 0.99) und berechnet aus X_1, X_2, \dots, X_n ein möglichst kleines Intervall I für das $\mathbb{P}(\theta \in I) \geq \gamma$ gilt. So ein Intervall I nennt man γ -Konfidenzintervall für θ . Dabei ist zu beachten, dass der unbekannte Parameter θ eine feste Zahl ist, aber I von der zufällig gewählten Stichprobe abhängt. Für γ gibt es verschiedene Namen, zum Beispiel Sicherheitswahrscheinlichkeit oder Konfidenzniveau.

Um Konfidenzintervalle berechnen zu können, muss man Annahmen über die Verteilung in der Grundgesamtheit treffen. Will man ein Konfidenzintervall für die durchschnittliche Körpergröße μ bestimmen, so könnte man die Körpergröße als $N(\mu, \sigma)$ -verteilt annehmen, wobei μ und σ unbekannte Parameter sind. Ein anderes Beispiel ist das Messen einer physikalischen Größe μ . Wird insgesamt n Mal gemessen und nimmt man an, dass die Messfehler einer Normalverteilung gehorchen, dann hat man n voneinander unabhängige $N(\mu, \sigma)$ -verteilte Messwerte für μ , wobei σ die Genauigkeit des Messvorgangs angibt.

Wir nehmen an, dass in der Grundgesamtheit eine $N(\mu, \sigma)$ -Verteilung vorliegt. Da zufällig aus dieser Grundgesamtheit gezogen wird, besteht die Stichprobe X_1, X_2, \dots, X_n dann aus unabhängigen Zufallsvariablen, die alle $N(\mu, \sigma)$ -verteilt sind. Wir nehmen vorläufig an, dass σ bekannt ist und berechnen ein γ -Konfidenzintervall für den Parameter μ . Dazu sei Φ die Verteilungsfunktion der $N(0, 1)$ -Verteilung. Bekanntlich gilt $\Phi(-x) = 1 - \Phi(x)$ für $x \in \mathbb{R}$.

Definition: Für $\delta \in (0, 1)$ sei t_δ die eindeutige Lösung von $\Phi(t_\delta) = 1 - \delta$. Diese Lösung ist eindeutig, da $\Phi : \mathbb{R} \rightarrow (0, 1)$ eine invertierbare Abbildung ist. Es gilt auch $\Phi(-t_\delta) = \delta$.

Satz 58: Sei $0 < \gamma < 1$ gegeben. Seien X_1, X_2, \dots, X_n unabhängige $N(\mu, \sigma)$ -verteilte Zufallsvariable, wobei σ bekannt ist. Sei $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ das Stichprobenmittel und $\delta = \frac{1-\gamma}{2}$. Für $I = [M - \frac{\sigma t_\delta}{\sqrt{n}}, M + \frac{\sigma t_\delta}{\sqrt{n}}]$ gilt dann $\mathbb{P}(\mu \in I) = \gamma$, das heißt I ist ein γ -Konfidenzintervall für μ .

Beweis: Aus Satz 54 folgt, dass $X_1 + X_2 + \dots + X_n$ die $N(n\mu, \sqrt{n}\sigma)$ -Verteilung hat. Wegen Satz 17 hat $\sqrt{n} \frac{M - \mu}{\sigma} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma}$ dann die $N(0, 1)$ -Verteilung. Es folgt

$$\mathbb{P}(-t_\delta \leq \sqrt{n} \frac{M - \mu}{\sigma} \leq t_\delta) = \Phi(t_\delta) - \Phi(-t_\delta) = 1 - \delta - \delta = \gamma$$

Wir formen dieses Ereignis um

$$-t_\delta \leq \sqrt{n} \frac{M - \mu}{\sigma} \leq t_\delta \Leftrightarrow -\frac{\sigma t_\delta}{\sqrt{n}} \leq M - \mu \leq \frac{\sigma t_\delta}{\sqrt{n}} \Leftrightarrow M - \frac{\sigma t_\delta}{\sqrt{n}} \leq \mu \leq M + \frac{\sigma t_\delta}{\sqrt{n}} \Leftrightarrow \mu \in I$$

Setzt man dieses umgeformte Ereignis oben ein, so erhält man $P(\mu \in I) = \gamma$, das gewünschte Resultat. \square

Beispiel 27: Die Füllmengen von Waschmittelpackungen einer bestimmten Sorte seien $N(\mu, \sigma)$ -verteilt. Sei $\sigma = 0.02$ kg die Genauigkeit der Abfüllmaschine. Man wiegt 10 zufällig gewählte Packungen und erhält 1.05, 0.99, 1.03, 1.03, 1.01, 1.02, 1.01, 0.97, 1.01, 0.98 in kg. Gesucht ist ein 99%-Konfidenzintervall für den durchschnittlichen Packungsinhalt μ .

Aus $\gamma = 0.99$ folgt $\delta = \frac{1-\gamma}{2} = 0.005$ und $t_\delta = 2.58$. Es gilt $M = \frac{1.05 + \dots + 0.98}{10} = 1.01$ und $\frac{\sigma t_\delta}{\sqrt{n}} = \frac{0.02 \cdot 2.58}{\sqrt{10}} = 0.019$. Daraus folgt $I = [1.01 - 0.019, 1.01 + 0.019] = [0.991, 1.029]$. Mit einer Wahrscheinlichkeit von 0.99 ist μ in diesem Intervall enthalten.

Bevor man ein Konfidenzintervall ermittelt, muss man zwei Entscheidungen treffen. Die eine ist die Wahl des Konfidenzniveaus γ , das angibt, mit welcher Wahrscheinlichkeit der unbekannte Parameter μ vom Konfidenzintervall überdeckt wird. Die andere ist die Wahl des Stichprobenumfangs. Hat man sich für ein Konfidenzniveau γ entschieden, dann ist auch t_δ festgelegt und man kann die Länge $|I|$ des Konfidenzintervalls I durch den Stichprobenumfang n steuern. Aus der Formel in Satz 58 folgt $|I| = \frac{2\sigma t_\delta}{\sqrt{n}}$. Vergrößert man den Stichprobenumfang n , dann wird das Konfidenzintervall kleiner, das heißt die Genauigkeit wird größer.

Beispiel 28: Wieviele Packungen muss man in obigem Beispiel überprüfen, damit die Länge des 99%-Konfidenzintervalls höchstens 0.01 kg beträgt?

Für $\gamma = 0.99$ haben wir $t_\delta = 2.58$ gefunden. Es wird $|I| = \frac{2\sigma t_\delta}{\sqrt{n}} \leq 0.01$ verlangt. Daraus folgt $\sqrt{n} \geq \frac{2\sigma t_\delta}{0.01} = \frac{2 \cdot 0.02 \cdot 2.58}{0.01} = 10.32$, das heißt $n \geq 106.5$. Man muss mindestens 107 Packungen überprüfen.

3. Statistische Tests

Es wird eine Hypothese H_0 formuliert, die eine Aussage über die Grundgesamtheit macht. Aufgrund einer Stichprobe aus dieser Grundgesamtheit wird entschieden, ob man diese Hypothese verwirft oder beibehält. Dabei soll die Wahrscheinlichkeit, eine falsche Entscheidung zu treffen, klein gehalten werden.

Als Beispiel wählen wir eine Maschine, die Waschmittelpackungen mit $\mu_0 = 1$ kg Inhalt abfüllt. Wir nehmen an, dass die Füllmengen $N(\mu, \sigma)$ -verteilt sind, wobei die Füllgenauigkeit σ bekannt sein soll. Die durchschnittliche Füllmenge μ lässt sich einstellen. Um zu kontrollieren, ob die Maschine richtig eingestellt ist, soll die Hypothese $H_0 : \mu = \mu_0$ getestet werden. Wir wählen zufällig n der gefüllten Packungen aus. Die Inhalte dieser Packungen bilden die Stichprobe X_1, X_2, \dots, X_n mit deren Hilfe für oder gegen H_0 zu entscheiden ist.

Diese Entscheidung trifft man mit Hilfe der sogenannten Teststatistik. Das ist eine Zufallsvariable, die sich als Funktion der Stichprobe schreiben lässt und deren Wert etwas über die Gültigkeit der Hypothese H_0 aussagt. Um eine Entscheidungsregel anzugeben, muss man außerdem die Verteilung der Teststatistik bei Gültigkeit von H_0 kennen.

In obigem Beispiel wählt man $U = \frac{\sqrt{n}}{\sigma}(M - \mu_0)$ mit $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ als Teststatistik. Diese Teststatistik U nimmt Werte in \mathbb{R} an. Da M ein Schätzer für μ ist, spricht ein Wert von U , der nahe bei 0 liegt, für die Hypothese. Je weiter der Wert von U von 0 entfernt ist, umso mehr spricht er gegen die Hypothese. Eine Entscheidungsregel für diesen Test wird durch ein $c > 0$ festgelegt: Man verwirft die Hypothese, wenn $|U| > c$ gilt,

und behält sie bei, wenn $|U| \leq c$ ist. Die Menge $\mathbb{R} \setminus [-c, c]$ nennt man den Verwerfungsbereich des Tests, da die Hypothese genau dann verworfen wird, wenn die Teststatistik U in diese Menge fällt.

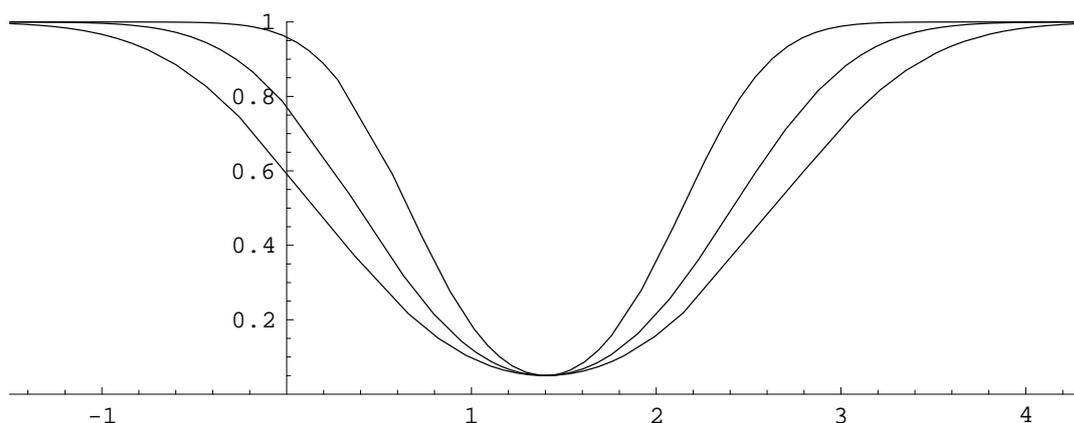
Um c zu bestimmen, gibt man die sogenannte Irrtumswahrscheinlichkeit $\alpha \in (0, 1)$ vor (typische Werte für α sind 0.05 und 0.01). Man wählt den Verwerfungsbereich so, dass bei Gültigkeit der Hypothese, die Wahrscheinlichkeit, diese zu verwerfen, kleiner oder gleich α ist. Unter dieser Bedingung soll der Verwerfungsbereich möglichst groß sein.

Bei Gültigkeit von H_0 hat U die $N(0, 1)$ -Verteilung, wie wir im Beweis von Satz 58 gesehen haben. Die Wahrscheinlichkeit, H_0 zu verwerfen, ist $P(|U| > c)$. Bei Gültigkeit von H_0 ist $P(|U| > c) = 2\Phi(-c)$. Somit ist c minimal zu wählen, sodass $2\Phi(-c) \leq \alpha$ gilt. Da Φ monoton und stetig ist, ist c die Lösung der Gleichung $2\Phi(-c) = \alpha$.

Die Wahl der Entscheidungsregel stellt sicher, dass eine richtige Hypothese höchstens mit Wahrscheinlichkeit α verworfen wird. Das gilt für jeden Stichprobenumfang n . Will man sich gegen die Beibehaltung einer falschen Hypothese absichern, dann muss man den Stichprobenumfang groß genug wählen. Man gibt eine Distanz $d > 0$ vor. Dann kann man n so groß wählen, dass mit Wahrscheinlichkeit $\geq 1 - \alpha$ verworfen wird, wenn die Abweichung von der Hypothese $\geq d$ ist.

Sei $G(\mu)$ die Wahrscheinlichkeit, die Hypothese zu verwerfen, in Abhängigkeit vom unbekanntem Parameter μ . Man nennt G Gütefunktion oder Mächtigkeit des Tests. Im Beweis von Satz 58 wurde gezeigt, dass $V = \frac{\sqrt{n}}{\sigma}(M - \mu)$ die $N(0, 1)$ -Verteilung hat. Wegen $U = V + \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)$ folgt $U \notin [-c, c] \Leftrightarrow V \notin [-c - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}, c - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}]$ und daraus $G(\mu) = P(V \notin [-c - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}, c - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}]) = \Phi(-c - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}) + \Phi(-c + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma})$. Setzt man $\mu = \mu_0$, so erhält man $G(\mu_0) = 2\Phi(-c) = \alpha$, wie es wegen obiger Wahl des Verwerfungsbereichs ja sein muss. Weiters gilt $G(\mu_0 + x) = G(\mu_0 - x)$ für alle $x \in \mathbb{R}$. Auf dem Intervall $(-\infty, \mu_0]$ ist G monoton fallend und auf dem Intervall $[\mu_0, \infty)$ ist G monoton wachsend. Wählt man n so groß, dass $G(\mu_0 + d) \geq 1 - \alpha$ ist, dann gilt auch $G(\mu_0 + x) \geq 1 - \alpha$ für alle x mit $|x| \geq d$. So ein n ist ein geeigneter Stichprobenumfang.

Die folgende Zeichnung zeigt G für $\mu_0 = 1.4$, $\sigma = 2$, $\alpha = 0.05$ und für die Stichprobenumfänge $n = 10$, $n = 15$ und $n = 28$. Man sieht, dass $G(1.4) = 0.05$ gilt. Je größer n ist, umso



schneller wächst G , wenn μ von μ_0 abweicht. Wenn $\mu = 0$ ist (Abweichung $d = 1.4$ von μ_0), dann wird die Hypothese $\mu = \mu_0$ bei einem Stichprobenumfang von $n = 28$ bereits mit Wahrscheinlichkeit ≥ 0.95 verworfen, während bei einem Stichprobenumfang von $n = 15$ die Wahrscheinlichkeit, die Hypothese zu verwerfen, unter 0.8 liegt, wie man an der y -Achse ablesen kann.

Beispiel 29: Die Füllmengen seien $N(\mu, \sigma)$ -verteilt mit $\sigma = 0.02$ kg. Sei $\alpha = 0.05$ und $d = 0.01$ kg. Gesucht ist der Verwerfungsbereich zur Irrtumswahrscheinlichkeit α und ein Stichprobenumfang, sodass mit Wahrscheinlichkeit $\geq 1 - \alpha$ verworfen wird, wenn die Abweichung von der Hypothese $\geq d$ ist.

Es gilt $c = -\Phi^{-1}(\frac{\alpha}{2}) = 1.96$. Der Verwerfungsbereich ist $(-\infty, -1.96) \cup (1.96, \infty)$. Für den Stichprobenumfang n muss dann $\Phi(-1.96 - \frac{\sqrt{n}}{\sigma}0.01) + \Phi(-1.96 + \frac{\sqrt{n}}{\sigma}0.01) \geq 0.95$ gelten. Der erste Summand links ist sehr klein. Man kann ihn vernachlässigen und erhält $\frac{\sqrt{n}}{\sigma}0.01 \geq 1.96 + \Phi^{-1}(0.95) = 3.61$, woraus $\sqrt{n} \geq 361\sigma = 7.22$ und $n \geq 52.13$ folgt.

Plant man einen Test, dann formuliert man die Hypothese, wählt eine Irrtumswahrscheinlichkeit α und legt eine Abweichung d von der Hypothese fest, die man als wesentlich ansieht. Vorausgesetzt, es lässt sich eine geeignete Teststatistik finden, kann man dann mit obigen Methoden den Verwerfungsbereich und den Stichprobenumfang berechnen. Fällt der Wert der Teststatistik in den Verwerfungsbereich, dann verwirft man die Hypothese, wobei die Wahrscheinlichkeit für einen Irrtum durch α beschränkt ist. Fällt der Wert der Teststatistik nicht in den Verwerfungsbereich, dann behält man die Hypothese bei, wobei bei einer Abweichung $> d$ von der Hypothese die Wahrscheinlichkeit für einen Irrtum wieder durch α beschränkt ist.

Interessiert man sich nur für Abweichungen von μ_0 in einer Richtung, so wählt man zum Beispiel eine Hypothese $H_0 : \mu \geq \mu_0$. So einen Test nennt man einseitig im Gegensatz zu obigem zweiseitigen Test. Die Teststatistik ist wieder $U = \frac{\sqrt{n}}{\sigma}(M - \mu_0)$. Da jetzt aber ein großer Wert von U für und ein kleiner Wert von U gegen die Hypothese spricht, hat der Verwerfungsbereich die Form $(-\infty, b)$ und b soll maximal sein, sodass $P(U < b) \leq \alpha$ bei Richtigkeit der Hypothese gilt. Sei $V = \frac{\sqrt{n}}{\sigma}(M - \mu)$. Wenn H_0 gilt, dann hat man $V \leq U$ und damit $P(U < b) \leq P(V < b) = \Phi(b)$, sodass für b die Lösung von $\Phi(b) = \alpha$ zu wählen ist. Die Gütefunktion des einseitigen Tests ist $G(\mu) = \Phi(b - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma})$. Das ist eine monoton fallende Funktion. Es gilt $G(\mu) \leq \alpha$ für $\mu \geq \mu_0$, wie es wegen obiger Wahl des Verwerfungsbereichs ja sein muss. Den Stichprobenumfang n kann man aus der Gleichung $G(\mu_0 - d) \geq 1 - \alpha$ bestimmen.

Es gibt auch kompliziertere Tests als den hier behandelten, wie wir noch sehen werden. Für diese gelten im Prinzip dieselben Überlegungen. Allerdings ist es dann nicht mehr so einfach, Abweichungen von der Hypothese zu beschreiben, und der Stichprobenumfang kann oft nicht mehr explizit berechnet werden, sondern muss geschätzt werden.

Bei der Wahl der Teststatistik hat man gewisse Freiheiten. Zumindest kann man eine streng monotone Funktion auf sie anwenden und erhält wieder eine geeignete Teststatistik. Man braucht ja nur den Verwerfungsbereich mit derselben monotonen Funktion abzubilden. Man versucht nun die Teststatistik so zu transformieren, dass sie Wertebereich $(0, 1]$ hat, dass ein Wert nahe 0 gegen und ein Wert nahe 1 für die Hypothese spricht, und dass sie bei Gültigkeit der Hypothese gleichmäßig auf $(0, 1]$ verteilt ist, das heißt Wahrscheinlichkeitsdichte $1_{(0,1]}$ hat. Man nennt so eine Teststatistik die Signifikanz (P-Wert) des Tests. Der Verwerfungsbereich ist dann einfach das Intervall $(0, \alpha)$. Es enthält die Werte, die am meisten gegen die Hypothese sprechen, und die Wahrscheinlichkeit, dass bei Gültigkeit der Hypothese die Signifikanz K in dieses Intervall fällt, ist wegen der gleichmäßigen Verteilung gleich $P(K \in (0, \alpha)) = \int_0^\alpha 1_{(0,1]}(x) dx = \alpha$. Zum Berechnen der Signifikanz verwenden wir folgenden Satz.

Satz 59: Sei Z eine Zufallsvariable und F eine Verteilungsfunktion. Hat Z Wertebereich \mathbb{R}^+ und ist $F : [0, \infty) \rightarrow [0, 1)$ bijektiv, dann hat die Zufallsvariable $K = 1 - F(Z)$ Wertebereich

$(0, 1]$, wobei der Wert 0 von Z dem Wert 1 von K entspricht und ein großer Wert von Z einem Wert von K entspricht, der nahe 0 liegt. Hat Z die Verteilung F , dann hat K die Wahrscheinlichkeitsdichte $1_{(0,1]}$.

Beweis: Da $F : [0, \infty) \rightarrow [0, 1)$ eine streng monoton wachsende und bijektive Funktion ist, bildet $x \mapsto 1 - F(x)$ das Intervall $[0, \infty)$ streng monoton fallend und bijektiv auf $(0, 1]$ ab. Daraus folgt die erste Aussage. Da Z die Verteilung F hat, folgt für $t \in (0, 1]$, dass $P(K \leq t) = P(Z \geq F^{-1}(1 - t)) = 1 - F(F^{-1}(1 - t)) = t$ gilt. Damit ist gezeigt, dass K die Wahrscheinlichkeitsdichte $1_{(0,1]}$ hat. \square

Für den Test der Hypothese $H_0 : \mu = \mu_0$ kann man auch $Z = U^2$ als Teststatistik wählen. Der Wertebereich von Z ist \mathbb{R}^+ und der Verwerfungsbereich für die Teststatistik Z ist dann $[c^2, \infty)$. Bei Gültigkeit der Hypothese hat Z die $C(1)$ -Verteilung (Beispiel 15). Wir nennen die $G(\frac{n}{2}, \frac{1}{2})$ -Verteilung hier $C(n)$ -Verteilung (chi-quadrat-Verteilung). Die Verteilungsfunktion der $C(1)$ -Verteilung bezeichnen wir mit D . Sie bildet \mathbb{R}^+ streng monoton wachsend auf $[0, 1)$ ab. Somit ist $K = 1 - D(Z)$ ebenfalls eine geeignete Teststatistik. Wegen Satz 59 hat sie Wertebereich $(0, 1]$, wobei ein Wert nahe 0 gegen und ein Wert nahe 1 für H_0 spricht, und bei Gültigkeit von H_0 ist K gleichmäßig auf $(0, 1]$ verteilt. Daher ist K die Signifikanz des Tests.

4. Das lineare Modell

Um auch normalverteilte Grundgesamtheiten mit unbekanntem σ behandeln zu können, benötigen wir Resultate aus der Geometrie. Sei V ein linearer Teilraum des \mathbb{R}^n der Dimension $d < n$. Die orthogonale Projektion P auf V ist eine Abbildung von \mathbb{R}^n nach \mathbb{R}^n , die jedem Punkt $x \in \mathbb{R}^n$ den Punkt in V zuordnet, der von x den kleinsten Abstand hat. Im folgenden Satz wird die orthogonale Projektion berechnet.

Satz 60: Sei v_1, v_2, \dots, v_d eine Basis des linearen Teilraumes V . Sei A die $n \times d$ -Matrix, deren Spalten v_1, v_2, \dots, v_d sind. Sei P die $n \times n$ -Matrix $A(A^t A)^{-1} A^t$. Für jedes $x \in \mathbb{R}^n$ ist dann Px die orthogonale Projektion von x auf V .

Beweis: Ist $y \in \mathbb{R}^d$ ein Spaltenvektor, dann bezeichnen wir mit y_1, y_2, \dots, y_d seine Koordinaten. Wir zeigen zuerst, dass die $d \times d$ -Matrix $A^t A$ invertierbar ist. Sei $y \in \mathbb{R}^d \setminus \{0\}$. Dann ist $Ay = \sum_{j=1}^d y_j v_j \neq 0$, da v_1, v_2, \dots, v_d linear unabhängige Vektoren sind. Es folgt $\langle y, A^t Ay \rangle = \langle Ay, Ay \rangle = \|Ay\|^2 > 0$. Wäre $A^t A$ nicht invertierbar, dann würde ein Vektor $y \in \mathbb{R}^d \setminus \{0\}$ existieren mit $A^t Ay = 0$, also auch $\langle y, A^t Ay \rangle = 0$, ein Widerspruch.

Es gilt $V = \{Ay : y \in \mathbb{R}^d\}$, da V die Menge der Linearkombinationen der Spalten von A ist. Für alle $z \in \mathbb{R}^d$ gilt $\langle x - Px, Az \rangle = \langle A^t x - A^t Px, z \rangle = \langle A^t x - A^t A(A^t A)^{-1} A^t x, z \rangle = \langle A^t x - A^t x, z \rangle = 0$. Setzt man $z = (A^t A)^{-1} A^t x - y$, so hat man $\langle x - Px, Px - Ay \rangle = 0$, sodass $\|x - Ay\|^2 = \|x - Px + Px - Ay\|^2 = \|x - Px\|^2 + \|Px - Ay\|^2$ für alle $y \in \mathbb{R}^d$ gilt. Somit ist $\|x - Ay\|$ genau dann minimal, wenn $y \in \mathbb{R}^d$ so gewählt wird, dass $Ay = Px$ gilt, zum Beispiel $y = (A^t A)^{-1} A^t x$. Das zeigt, dass Px der Punkt in V ist, der von x den kleinsten Abstand hat. Somit ist Px die orthogonale Projektion von x auf V . \square

Beispiel 30: Sei $e \in \mathbb{R}^n$ der Spaltenvektor, dessen Koordinaten alle 1 sind. Sei V der von e aufgespannte eindimensionale lineare Teilraum des \mathbb{R}^n . Zu berechnen ist die orthogonale Projektion P auf V .

Sei E die $n \times n$ -Matrix, deren Eintragungen alle 1 sind. Die $n \times 1$ -Matrix A besteht aus der einen Spalte e . Somit ist $A^t A$ die 1×1 -Matrix mit Eintragung n und $(A^t A)^{-1}$ die 1×1 -Matrix mit Eintragung $\frac{1}{n}$. Wegen $AA^t = E$ folgt $P = A(A^t A)^{-1} A^t = \frac{1}{n} AA^t = \frac{1}{n} E$.

Wir haben eine Folge von ineinandergeschachtelten linearen Teilräumen des \mathbb{R}^n vorliegen und eine Stichprobe, die aus unabhängigen normalverteilten Zufallsvariablen besteht. Wir untersuchen die Projektionen der Stichprobe auf die Teilräume.

Satz 61: Seien $V_0 = \{0\} \subset V_1 \subset \dots \subset V_{r-1} \subset V_r = \mathbb{R}^n$ Teilräume mit Dimensionen $d_0 = 0 < d_1 < \dots < d_{r-1} < d_r = n$. Sei P_j die orthogonale Projektion auf V_j , sodass insbesondere $P_0x = 0$ und $P_r x = x$ für alle $x \in \mathbb{R}^n$ gilt. Seien X_1, X_2, \dots, X_n unabhängige Zufallsvariablen, wobei X_j die $N(\rho_j, \sigma)$ -Verteilung hat. Seien $X = (X_1, X_2, \dots, X_n)$ und $\rho = (\rho_1, \rho_2, \dots, \rho_n)$ Spaltenvektoren. Dann gilt

(a) Für stetige Funktionen $u_j : \mathbb{R}^n \rightarrow \mathbb{R}$ sind die Zufallsvariablen $u_j(P_j X - P_{j-1} X)$ mit $1 \leq j \leq r$ unabhängig.

(b) Wenn $P_j \rho = P_{j-1} \rho$ gilt, dann hat $\frac{1}{\sigma^2} \|P_j X - P_{j-1} X\|^2$ die $C(d_j - d_{j-1})$ -Verteilung.

Beweis: Wir führen eine orthogonale Basis v_1, v_2, \dots, v_n im \mathbb{R}^n ein, sodass v_1, v_2, \dots, v_{d_j} den Teilraum V_j aufspannen. Sei K die orthogonale Matrix, deren Zeilen v_1, v_2, \dots, v_n sind, sodass K alte Koordinaten in Koordinaten bezüglich der neuen Basis v_1, v_2, \dots, v_n überführt. Es gilt $K^{-1} = K^t$ und $\|K^t x\| = \|x\|$ für alle $x \in \mathbb{R}^n$.

Sei I_m die $n \times n$ -Diagonalmatrix, die zuerst m Einser und dann $n-m$ Nullen in der Diagonale hat. Da v_1, v_2, \dots, v_{d_j} den Teilraum V_j aufspannen, ist I_{d_j} die orthogonale Projektion auf V_j in neuen Koordinaten. Deshalb gilt $I_{d_j} = K P_j K^t$, also $K^t I_{d_j} = P_j K^t$.

Sei $Y_j = \frac{X_j - \rho_j}{\sigma}$ für $1 \leq j \leq n$. Das sind unabhängige $N(0, 1)$ -verteilte Zufallsvariablen. Für $Y = (Y_1, Y_2, \dots, Y_n)$ gilt $Y = \frac{1}{\sigma}(X - \rho)$. Sei $Z = KY$. Da K eine orthogonale Matrix ist, sind auch die Komponenten Z_1, Z_2, \dots, Z_n von Z unabhängig und alle $N(0, 1)$ -verteilt. Es folgt $P_j X = P_j(\rho + \sigma Y) = P_j \rho + \sigma P_j Y = P_j \rho + \sigma P_j K^t Z = P_j \rho + \sigma K^t I_{d_j} Z$. Daraus erhalten wir $P_j X - P_{j-1} X = P_j \rho - P_{j-1} \rho + \sigma K^t (I_{d_j} - I_{d_{j-1}}) Z$.

Nun ist $I_{d_j} - I_{d_{j-1}}$ die $n \times n$ -Diagonalmatrix, die zuerst d_{j-1} Nullen, dann $d_j - d_{j-1}$ Einser und schließlich $n - d_j$ Nullen in der Diagonale hat. Somit ist $(I_{d_j} - I_{d_{j-1}}) Z$ ein Vektor, dessen erste d_{j-1} Koordinaten null sind, der in den Koordinaten $d_{j-1} + 1, d_{j-1} + 2, \dots, d_j$ die Eintragungen $Z_{d_{j-1}+1}, Z_{d_{j-1}+2}, \dots, Z_{d_j}$ hat, und dessen $n - d_j$ letzte Koordinaten wieder null sind. Daraus und aus obiger Darstellung von $P_j X - P_{j-1} X$ erkennt man, dass $u_j(P_j X - P_{j-1} X)$ eine stetige Funktion von $Z_{d_{j-1}+1}, Z_{d_{j-1}+2}, \dots, Z_{d_j}$ ist, aber von den anderen Z_k nicht abhängt. Da die Zufallsvariablen Z_1, Z_2, \dots, Z_n unabhängig sind, erhalten wir aus einer Folgerung von Satz 25, dass auch $u_1(P_1 X - P_0 X), u_2(P_2 X - P_1 X), \dots, u_r(P_r X - P_{r-1} X)$ unabhängige Zufallsvariable sind. Das ist (a).

Gilt $P_j \rho = P_{j-1} \rho$, dann folgt aus obiger Gleichung und wegen $\|K^t x\| = \|x\|$ für $x \in \mathbb{R}^n$, dass $\frac{1}{\sigma^2} \|P_j X - P_{j-1} X\|^2 = \|K^t (I_{d_j} - I_{d_{j-1}}) Z\|^2 = \|(I_{d_j} - I_{d_{j-1}}) Z\|^2 = \sum_{k=d_{j-1}+1}^{d_j} Z_k^2$ gilt. Das ist eine Summe von Quadraten von $d_j - d_{j-1}$ unabhängigen $N(0, 1)$ -verteilten Zufallsvariablen und hat daher nach Satz 31 die $C(d_j - d_{j-1})$ -Verteilung. \square

Bemerkung: $P_j - P_{j-1}$ ist die Projektion auf das orthogonale Komplement von V_j in V_{j-1} . Das ist ein Teilraum der Dimension $d_j - d_{j-1}$. Ein anderes Wort für Dimension ist Freiheitsgrade (degrees of freedom). Diese Bezeichnung findet man oft in der Statistik.

Jetzt können wir normalverteilte Grundgesamtheiten mit unbekanntem σ behandeln.

Satz 62: Die Zufallsvariablen X_1, X_2, \dots, X_n seien unabhängig und alle $N(\mu, \sigma)$ -verteilt. Sei $M = \frac{1}{n} \sum_{j=1}^n X_j$, $Y = \sqrt{n} \frac{M - \mu}{\sigma}$ und $Z = \frac{1}{\sigma^2} \sum_{j=1}^n (X_j - M)^2 = \frac{n-1}{\sigma^2} S^2$. Dann gilt

(a) Y und Z sind unabhängig, Y ist $N(0, 1)$ -verteilt und Z ist $C(n-1)$ -verteilt

(b) $\sqrt{n} \frac{M - \mu}{S} = \frac{\sqrt{n-1} Y}{\sqrt{Z}}$ ist $T(n-1)$ -verteilt.

Beweis: Seien $e \in \mathbb{R}^n$ und die $n \times n$ -Matrix E wie in obigem Beispiel. Sei V_1 der von e aufgespannte eindimensionale Teilraum des \mathbb{R}^n . Es gilt $V_0 = \{0\} \subset V_1 \subset V_2 = \mathbb{R}^n$. Seien P_0, P_1 und P_2 die zugehörigen orthogonalen Projektionen und $X = (X_1, X_2, \dots, X_n)$. Dann gilt $P_1X - P_0X = P_1X = \frac{1}{n}EX = Me$ und $\frac{1}{\sigma^2}\|P_2X - P_1X\|^2 = \frac{1}{\sigma^2}\|X - Me\|^2 = Z$. Definiert man $u(x) = \sqrt{n} \frac{x_1 - \mu}{\sigma}$, so folgt $Y = u(P_1X - P_0X)$, sodass wir die Unabhängigkeit von Y und Z aus Satz 61(a) erhalten. Wie im Beweis von Satz 58 folgt, dass Y die $N(0, 1)$ -Verteilung hat. Für das in Satz 61 definierte ρ gilt hier $\rho = \mu e \in V_1$. Somit folgt $P_2\rho = P_1\rho$ und nach Satz 61(b) hat Z die $C(n-1)$ -Verteilung. Damit ist (a) bewiesen. Mit Hilfe von Satz 32 folgt aus (a) auch, dass $\frac{\sqrt{n-1}Y}{\sqrt{Z}}$ die $T(n-1)$ -Verteilung hat. Das ist (b). \square

Definition: Sei T_m die Verteilungsfunktion der $T(m)$ -Verteilung. Ihre Dichte ist symmetrisch zur y -Achse, daher gilt $T_m(-x) = 1 - T_m(x)$. Für $\delta \in (0, 1)$ sei $u_{\delta, m}$ die eindeutige Lösung x von $1 - T_m(x) = \delta$.

Satz 63: Sei $0 < \gamma < 1$ gegeben. Seien X_1, X_2, \dots, X_n unabhängig und $N(\mu, \sigma)$ -verteilt. Sei $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$, $S = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - M)^2}$ und $\delta = \frac{1-\gamma}{2}$. Für das Intervall $K = [M - \frac{Su_{\delta, n-1}}{\sqrt{n}}, M + \frac{Su_{\delta, n-1}}{\sqrt{n}}]$ gilt dann $P(\mu \in K) = \gamma$.

Beweis: Genauso wie für Satz 58 unter Verwendung von Satz 62(b). \square

Bemerkung: Genauso einfach kann man auch die statistischen Tests für die Hypothesen $H_0 : \mu = \mu_0$ und $H_0 : \mu \geq \mu_0$ übertragen. Man braucht nur σ durch S ersetzen und statt der Verteilungsfunktion Φ die Verteilungsfunktion T_{n-1} verwenden.

Wir können auch ein Konfidenzintervall für σ bestimmen.

Definition: Sei C_m die Verteilungsfunktion der $C(m)$ -Verteilung. Für $\delta \in (0, 1)$ sei $v_{\delta, m}$ die eindeutige Lösung x von $1 - C_m(x) = \delta$.

Satz 64: Sei $0 < \gamma < 1$ gegeben. Seien X_1, X_2, \dots, X_n unabhängig und $N(\mu, \sigma)$ -verteilt. Sei $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$, $S = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - M)^2}$ und $\delta = \frac{1-\gamma}{2}$. Für das Intervall $L = [S\sqrt{\frac{n-1}{v_{\delta, n-1}}}, S\sqrt{\frac{n-1}{v_{1-\delta, n-1}}}]$ gilt dann $P(\sigma \in L) = \gamma$.

Beweis: Nach Satz 62(a) hat $\frac{n-1}{\sigma^2}S^2$ die $C(n-1)$ -Verteilung. Daher gilt

$$P(v_{1-\delta, n-1} \leq \frac{n-1}{\sigma^2}S^2 \leq v_{\delta, n-1}) = C_{n-1}(v_{\delta, n-1}) - C_{n-1}(v_{1-\delta, n-1}) = 1 - 2\delta = \gamma$$

Formt man diese Ungleichungen entsprechend um, so erhält man $P(\sigma \in L) = \gamma$. \square

Liegt in der Grundgesamtheit eine andere Verteilung als die Normalverteilung vor, dann besteht die Stichprobe aus unabhängigen Zufallsvariablen, die alle diese andere Verteilung haben. Sei μ deren Erwartungswert. Man kann zeigen, dass die Verteilung von $\sqrt{n} \frac{M - \mu}{S}$ gegen die $N(0, 1)$ -Verteilung konvergiert, wenn der Stichprobenumfang n gegen ∞ geht (zentraler Grenzwertsatz). Für einen großen Stichprobenumfang n ist dann $I = [M - t_\delta \frac{S}{\sqrt{n}}, M + t_\delta \frac{S}{\sqrt{n}}]$ mit $\delta = \frac{1-\gamma}{2}$ und $t_\delta = \Phi^{-1}(1 - \delta)$ ein näherungsweise γ -Konfidenzintervall für den Parameter μ . Man spricht von einem asymptotischen Konfidenzintervall.

Wir berechnen ein asymptotisches Konfidenzintervall für den Anteil p der Raucher in der Grundgesamtheit. In diesem Fall haben wir $X_j = 1$, wenn die j -te befragte Person ein Raucher ist, und $X_j = 0$, wenn sie kein Raucher ist. Die Stichprobe besteht aus unabhängigen $B(1, p)$ -verteilten Zufallsvariablen X_1, X_2, \dots, X_n . Der relative Anteil der Raucher in der

Stichprobe ist dann $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Wegen $X_j^2 = X_j$ für $1 \leq j \leq n$ folgt $\sum_{j=1}^n (X_j - M)^2 = \sum_{j=1}^n X_j^2 - 2 \sum_{j=1}^n X_j M + nM^2 = nM - 2nM^2 + nM^2 = nM(1 - M)$ und daher $S^2 = \frac{n}{n-1}M(1 - M)$. Für große n haben wir daher $S \approx \sqrt{M(1 - M)}$. Somit ist $I = \left[M - \frac{t_\delta}{\sqrt{n}} \sqrt{M(1 - M)}, M + \frac{t_\delta}{\sqrt{n}} \sqrt{M(1 - M)} \right]$ ein asymptotisches Konfidenzintervall für den Anteil p der Raucher in der Grundgesamtheit.

Beispiel 31: Wie groß muss man den Stichprobenumfang n wählen, wenn die Länge des 95%-Konfidenzintervalls I höchstens 0.04 sein soll?

Die Länge von I ist $\frac{2t_\delta}{\sqrt{n}} \sqrt{M(1 - M)}$. Wegen $\sqrt{M(1 - M)} \leq \frac{1}{2}$ genügt es n so zu wählen, dass $\frac{t_\delta}{\sqrt{n}} \leq 0.04$ gilt. Für $\gamma = 0.95$ ist $t_\delta = 1.96$, also $\sqrt{n} \geq \frac{1.96}{0.04} = 49$, das heißt $n \geq 2401$.

5. Varianzanalyse

Die Grundgesamtheit zerfällt in k Untergruppen. Es wird angenommen, dass in der j -ten Untergruppe eine $N(\mu_j, \sigma)$ -Verteilung vorliegt. In den Untergruppen haben wir also gleiche Varianzen aber möglicherweise verschiedene Mittelwerte. Ein Beispiel ist die Körpergröße in verschiedenen Altersgruppen. Es soll die Hypothese $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ getestet werden. Zu entscheiden ist, ob die Mittelwerte in den Untergruppen gleich oder nicht alle gleich sind. Aus der j -ten Untergruppe wird eine Stichprobe vom Umfang n_j gezogen. Diese Stichproben werden zu einer Gesamtstichprobe X_1, X_2, \dots, X_n vom Umfang $n = n_1 + \dots + n_k$ zusammengefasst, sodass X_1, X_2, \dots, X_{n_1} die Stichprobe aus der ersten Untergruppe darstellt, $X_{n_1+1}, X_{n_1+2}, \dots, X_{n_1+n_2}$ die aus der zweiten Untergruppe und so weiter. Wir führen die Indexmengen $I_j = \{n_1 + \dots + n_{j-1} + 1, n_1 + \dots + n_{j-1} + 2, \dots, n_1 + \dots + n_{j-1} + n_j\}$ für $1 \leq j \leq k$ ein, sodass die Zufallsvariablen X_i mit $i \in I_j$ die Stichprobe aus der j -ten Untergruppe darstellen. Aus den oben gemachten Annahmen folgt dann

- (A) Die Zufallsvariablen X_1, X_2, \dots, X_n in der Gesamtstichprobe sind unabhängig.
 (B) Die Zufallsvariablen X_i für $i \in I_j$ sind $N(\mu_j, \sigma)$ -verteilt, wobei $1 \leq j \leq k$ gilt.

Ein weiteres Beispiel ist der Vergleich der Erträge von k verschiedenen Kartoffelsorten. Die k Kartoffelsorten bilden die Untergruppen der Grundgesamtheit. Die Zufallsvariablen X_i für $i \in I_j$ stellen die Erträge der j -ten Sorte bei einem Anbauversuch dar. Dabei wird ein Feld in n gleich große Parzellen geteilt, von denen n_1 mit Sorte 1, n_2 mit Sorte 2 und so weiter bepflanzt werden. Die Ergebnisse eines solchen Versuchs sehen so aus, wobei $k = 3$ ist

Ertrag	Sorte	v_1	v_2	v_3
57	1	1	0	0
50	1	1	0	0
53	1	1	0	0
58	1	1	0	0
52	1	1	0	0
42	2	0	1	0
49	2	0	1	0
42	2	0	1	0
47	2	0	1	0
56	3	0	0	1
62	3	0	0	1
56	3	0	0	1

Die Unterteilung in die Mengen I_1 , I_2 und I_3 ist in der zweiten Spalte angegeben.

Um eine Teststatistik für die Hypothese $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ zu finden, gehen wir von Schätzern M_j und S_j^2 für μ_j und $\sigma_j^2 = \sigma^2$ aus

$$M_j = \frac{1}{n_j} \sum_{i \in I_j} X_i \quad \text{und} \quad S_j^2 = \frac{1}{n_j - 1} \sum_{i \in I_j} (X_i - M_j)^2$$

Weiters bilden wir den Mittelwert der Gesamtstichprobe

$$M = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{j=1}^k n_j M_j$$

Damit definieren wir jetzt verschiedene Quadratsummen. Die Summe der Quadrate der Abweichungen zwischen den Stichproben wird definiert durch

$$L = \sum_{j=1}^k n_j (M_j - M)^2$$

Da M_j ein Schätzer für μ_j und M ein Mittelwert von M_1, M_2, \dots, M_k ist, wird L umso kleiner sein, je näher $\mu_1, \mu_2, \dots, \mu_k$ beieinander liegen. Daher spricht ein kleiner Wert von L für die Hypothese und ein großer Wert von L gegen die Hypothese. Die Summe der Quadrate der Abweichungen innerhalb der Stichproben wird definiert durch

$$F = \sum_{j=1}^k (n_j - 1) S_j^2 = \sum_{j=1}^k \sum_{i \in I_j} (X_i - M_j)^2$$

Da $\frac{1}{n-k} F$ ein Mittelwert der Schätzer S_j^2 von $\sigma_j^2 = \sigma^2$ ist, kann man $\frac{1}{n-k} F$ als Schätzer für σ^2 auffassen. Schließlich wird auch noch die totale Summe der Quadrate der Abweichungen definiert durch $T = \sum_{i=1}^n (X_i - M)^2$.

Aus den Quadratsummen L und F können wir eine geeignete Teststatistik Q bilden.

$$Q = \frac{L/(k-1)}{F/(n-k)}$$

Der Wertebereich von Q ist \mathbb{R}^+ . Man kann Q als Verhältnis der Abweichungen zwischen den Stichproben aus den einzelnen Untergruppen der Grundgesamtheit zu den Abweichungen innerhalb dieser Stichproben auffassen. Sind die Abweichungen zwischen den Stichproben wesentlich größer als die Abweichungen innerhalb der Stichproben, was gleichbedeutend mit einem großen Wert von Q ist, dann spricht das gegen die Hypothese. Liegt der Wert von Q hingegen nahe 0, dann spricht das für die Hypothese.

Um den Test durchführen zu können, müssen wir die Verteilung von Q bei Gültigkeit der Hypothese berechnen. Zu diesem Zweck drücken wir die Quadratsummen mit Hilfe geeigneter Projektionen aus. Sei $e \in \mathbb{R}^n$ der Vektor, dessen Eintragungen alle 1 sind. Für $1 \leq j \leq k$ sei $v_j \in \mathbb{R}^n$ der Vektor, der in den Koordinaten I_j Eintragungen 1 und sonst Eintragungen 0 hat. Sei V_1 der von e aufgespannte eindimensionale Teilraum des \mathbb{R}^n und V_2 der von den Vektoren v_1, v_2, \dots, v_k aufgespannte k -dimensionale Teilraum des \mathbb{R}^n . Es gilt $V_0 = \{0\} \subset V_1 \subset V_2 \subset V_3 = \mathbb{R}^n$, da $e = v_1 + v_2 + \dots + v_k \in V_2$ ist. Die Dimensionen sind $d_1 = 1$, $d_2 = k$ und $d_3 = n$. Seien P_0, P_1, P_2 und P_3 die zugehörigen orthogonalen Projektionen. Dann gilt $P_3 X = X$ und $P_1 X = M e$ wie im Beweis von Satz 62 berechnet. Um $P_2 X$ gemäß Satz 60 zu berechnen, sei A die $n \times k$ -Matrix mit den Spalten v_1, v_2, \dots, v_k . Dann ist $A^t A$ die $k \times k$ -Diagonalmatrix mit den Diagonaleintragungen n_1, n_2, \dots, n_k und $(A^t A)^{-1} A^t X$ ist der Spaltenvektor (M_1, M_2, \dots, M_k) . Multipliziert man diesen Vektor von links mit A , so folgt $P_2 X = A(A^t A)^{-1} A^t X = M_1 v_1 + M_2 v_2 + \dots + M_k v_k$. Somit ist

$$\|P_2 X - P_1 X\|^2 = \sum_{j=1}^k n_j (M_j - M)^2 = L$$

$$\|X - P_2 X\|^2 = \sum_{j=1}^k \sum_{i \in I_j} (X_i - M_j)^2 = F$$

$$\|X - P_1 X\|^2 = \sum_{i=1}^n (X_i - M)^2 = T$$

Mit Hilfe dieser Darstellung der Quadratsummen lässt sich folgender Satz beweisen.

Satz 65: Es seien (A) und (B) vorausgesetzt.

- (a) Wenn $\mu_1 = \mu_2 = \dots = \mu_k$ gilt, dann hat Q die $F(k-1, n-k)$ -Verteilung
 (b) Es gilt $F + L = T$

Beweis: Wegen $P_1X \in V_1 \subset V_2$ und $P_2X \in V_2$ folgt $P_2X - P_1X \in V_2$. Da P_2 die orthogonale Projektion auf V_2 ist, erhalten wir $\langle X - P_2X, P_2X - P_1X \rangle = 0$ und daher auch $\|X - P_1X\|^2 = \|X - P_2X + P_2X - P_1X\|^2 = \|X - P_2X\|^2 + \|P_2X - P_1X\|^2$. Damit ist $T = F + L$ gezeigt. Für den Vektor ϱ aus Satz 61 gilt $\varrho = \mu_1v_1 + \mu_2v_2 + \dots + \mu_kv_k$, da (B) vorausgesetzt wird. Es gilt also $\varrho \in V_2$ und somit $P_2\varrho = \varrho$. Wenn nun auch $\mu_1 = \mu_2 = \dots = \mu_k$ vorausgesetzt wird, dann gilt $\varrho \in V_1$ und $\varrho = P_1\varrho$. Aus Satz 61(b) folgt, dass $\frac{1}{\sigma^2}L$ die $C(k-1)$ -Verteilung und $\frac{1}{\sigma^2}F$ die $C(n-k)$ -Verteilung hat. Wegen Satz 61(a) sind $\frac{1}{\sigma^2}L$ und $\frac{1}{\sigma^2}F$ unabhängig. Somit hat $Q = \frac{1}{k-1} \frac{1}{\sigma^2}L / \frac{1}{n-k} \frac{1}{\sigma^2}F$ nach Satz 33 die $F(k-1, n-k)$ -Verteilung. \square

Sei H die Verteilungsfunktion der $F(k-1, n-k)$ -Verteilung, eine bijektive streng monoton wachsende Funktion von $[0, \infty)$ nach $[0, 1)$. Wegen Satz 59 ist $K = 1 - H(Q)$ eine Teststatistik mit Wertebereich $(0, 1]$, wobei ein Wert von K nahe 0 gegen und ein Wert von K nahe 1 für die Hypothese spricht. Bei Gültigkeit der Hypothese hat Q die $F(k-1, n-k)$ -Verteilung, sodass K gleichverteilt auf $(0, 1]$ ist. Somit ist K die Signifikanz des Tests. Der zur Teststatistik K gehörige Verwerfungsbereich ist $(0, \alpha]$.

Beispiel 32: Für die in obiger Tabelle angegebenen Daten soll die Varianzanalyse durchgeführt werden.

Wir haben $k = 3$ Stichproben mit Stichprobenumfängen $n_1 = 5$, $n_2 = 4$ und $n_3 = 3$, sodass die Gesamtstichprobe Umfang $n = 12$ hat. Die Stichprobenmittel sind $M_1 = 54$, $M_2 = 45$ und $M_3 = 58$. Das Gesamtmittel ist dann $M = \frac{1}{n}(n_1M_1 + n_2M_2 + n_3M_3) = 52$. Daraus berechnet man $L = 5 \cdot 2^2 + 4 \cdot 7^2 + 3 \cdot 6^2 = 324$. Die Stichprobenvarianzen sind $S_1 = \frac{1}{4} \cdot 46$, $S_2 = \frac{1}{3} \cdot 38$ und $S_3 = \frac{1}{2} \cdot 24$, woraus $F = 46 + 38 + 24 = 108$ folgt. Somit ist $\frac{1}{k-1}L = 162$, $\frac{1}{n-k}F = 12$ und $Q = 13.5$. Die Signifikanz ist 0.002, sodass man unterschiedliche Durchschnittserträge annehmen muss. An den Werten von M_1 , M_2 und M_3 kann man ablesen, welche Sorten besser sind.

6. Lineare Regression

Die Länge x eines Stabes ist eine lineare Funktion der Temperatur u . Es gilt $x = \beta_1 + \beta_2u$ für gewisse Konstanten β_1 und β_2 , die vom Material abhängen, aus dem der Stab besteht. Um β_1 und β_2 zu bestimmen, bringt man den Stab auf Temperaturen u_1, u_2, \dots, u_n und misst jeweils seine Länge. Man erhält die Längen X_1, X_2, \dots, X_n , die wir als Zufallsvariable auffassen, da es möglicherweise noch andere zufällige Einflüsse auf die gemessenen Längen des Stabes gibt (Messfehler). Wir haben also

$$X_j = \beta_1 + \beta_2u_j + \sigma N_j \quad \text{für } 1 \leq j \leq n$$

wobei die Zufallsvariablen N_j die zufälligen Störungen bei den einzelnen Messungen beschreiben. Wir fassen die Zufallsvariablen X_j zum Vektor X und die Zufallsvariablen N_j zum Vektor N zusammen und gehen über zur Vektorschreibweise

$$X = A\beta + \sigma N \quad \text{mit } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad \text{und } A = \begin{pmatrix} 1 & u_1 \\ \vdots & \vdots \\ 1 & u_n \end{pmatrix}$$

Die Zufallsvariablen N_j werden als unabhängig und $N(0, 1)$ -verteilt angenommen.

Wir können diese Fragestellung gleich verallgemeinern. Es ist nicht notwendig, dass x eine lineare Funktion in u ist. Wir geben Funktionen r_1, r_2, \dots, r_k vor und behandeln eine Regressionsfunktion der Form

$$x = \beta_1 r_1(u) + \beta_2 r_2(u) + \dots + \beta_k r_k(u)$$

wobei die Parameter $\beta_1, \beta_2, \dots, \beta_k$ zu bestimmen sind. Wie oben erhalten wir

$$X = A\beta + \sigma N \quad \text{mit} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} r_1(u_1) & \cdots & r_k(u_1) \\ \vdots & & \vdots \\ r_1(u_n) & \cdots & r_k(u_n) \end{pmatrix}$$

Wählt man zum Beispiel $r_j(u) = u^{j-1}$, dann wird die Länge des Stabes durch ein Polynom der Temperatur dargestellt.

Auch kann x eine Funktion in mehreren Variablen sein, zum Beispiel

$$x = \beta_1 + \beta_2 u + \beta_3 v + \beta_4 w$$

wobei für vorgegebene $(u_1, v_1, w_1), (u_2, v_2, w_2), \dots, (u_n, v_n, w_n)$ Werte X_1, X_2, \dots, X_n gemessen werden, die aufgrund sonstiger Einflüsse wieder Zufallsvariable sind. Dann ist

$$X = A\beta + \sigma N \quad \text{mit} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_4 \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} 1 & u_1 & v_1 & w_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_n & v_n & w_n \end{pmatrix}$$

In allen Fällen erhalten wir die Gleichung $X = A\beta + \sigma N$ mit vorgegebener $n \times k$ -Matrix A , wobei die Komponenten N_1, N_2, \dots, N_n von N als unabhängig und $N(0, 1)$ -verteilt angenommen werden und die unbekannt Parameter $\beta_1, \beta_2, \dots, \beta_k$ und σ aus den gemessenen Werten X_1, X_2, \dots, X_n zu bestimmen sind.

Um einen Schätzer für den Parametervektor β zu finden, ignorieren wir die zufälligen Störungen, indem wir vorläufig $\sigma = 0$ setzen. Es werden mehr Messungen durchgeführt als unbekannte Parameter $\beta_1, \beta_2, \dots, \beta_k$ vorhanden sind. Deshalb ist A eine $n \times k$ -Matrix mit $k < n$. Wir nehmen an, dass A linear unabhängige Spalten hat, sodass $(A^t A)^{-1}$ existiert (siehe Beweis von Satz 60). Das zu lösende Gleichungssystem $A\beta = X$ ist überbestimmt. Wir bestimmen daher β so, dass die Norm $\|A\beta - X\|$ minimal wird. Da $A\beta$ in dem von den Spalten von A aufgespannten Teilraum liegt, ist $\|A\beta - X\|$ genau dann minimal, wenn $A\beta$ die orthogonale Projektion von X auf diesen Teilraum ist. Nach Satz 60 ist das genau dann der Fall, wenn $A\beta = A(A^t A)^{-1} A^t X$ gilt. Indem man diese Gleichung von links mit der Matrix $(A^t A)^{-1} A^t$ multipliziert, erhält man $\beta = (A^t A)^{-1} A^t X$ als eindeutige Lösung. Also ist $\|A\beta - X\|$ genau dann minimal, wenn $\beta = (A^t A)^{-1} A^t X$ gilt. Der Zufallsvektor $B = (A^t A)^{-1} A^t X$ sollte ein geeigneter Schätzer für den Vektor β sein.

Oben haben wir verschiedene Versionen von Regressionsfunktionen behandelt. Die Komponenten B_1, B_2, \dots, B_k von B sind Schätzer für die dort vorkommenden unbekannt Parameter $\beta_1, \beta_2, \dots, \beta_k$. Setzen wir B_1, B_2, \dots, B_k anstelle von $\beta_1, \beta_2, \dots, \beta_k$ ein, so erhalten wir eine geschätzte Regressionsfunktion.

Um Aussagen über die Güte dieser Schätzer machen zu können untersuchen wir die Verteilung der Zufallsvariablen B_1, B_2, \dots, B_k und Linearkombinationen davon.

Satz 66: Die Komponenten des Zufallsvektors N seien unabhängig und $N(0, 1)$ -verteilt. Sei $c \in \mathbb{R}^k \setminus \{0\}$. Dann hat die Zufallsvariable $\frac{1}{\sigma} \langle c, B - \beta \rangle = \frac{1}{\sigma} \sum_{j=1}^k c_j (B_j - \beta_j)$ die $N(0, \sqrt{c^t (A^t A)^{-1} c})$ -Verteilung.

Beweis: Wir hatten $B = (A^t A)^{-1} A^t X$. Setzt man $X = A\beta + \sigma N$ ein, so erhält man

$B = \beta + \sigma LN$ mit $L = (A^t A)^{-1} A^t$. Es folgt $\langle c, B \rangle = \langle c, \beta \rangle + \sigma \langle c, LN \rangle$ und daraus wieder $\frac{1}{\sigma} \langle c, B - \beta \rangle = \langle c, LN \rangle = \langle L^t c, N \rangle$. Sei $v = L^t c$. Da N_i die $N(0, 1)$ -Verteilung hat, hat $v_i N_i$ die $N(0, |v_i|)$ -Verteilung, wenn $v_i \neq 0$ ist. Weiters sind die Zufallsvariablen N_1, N_2, \dots, N_n unabhängig. Indem man die Summanden mit $v_i = 0$ weglässt, folgt mit Hilfe von Satz 54, dass $\langle L^t c, N \rangle = \sum_{i=1}^n v_i N_i$ die $N(0, \sqrt{\sum_{i=1}^n v_i^2})$ -Verteilung hat.

Weil $A^t A$ eine symmetrische Matrix ist und somit auch ihre Inverse $(A^t A)^{-1}$ symmetrisch ist, erhalten wir $LL^t = (A^t A)^{-1} A^t A (A^t A)^{-1} = (A^t A)^{-1}$. Somit ergibt sich

$$\sum_{i=1}^n v_i^2 = \|v\|^2 = \langle L^t c, L^t c \rangle = \langle c, LL^t c \rangle = c^t LL^t c = c^t (A^t A)^{-1} c$$

und $\langle c, LN \rangle$ hat die $N(0, \sqrt{c^t (A^t A)^{-1} c})$ -Verteilung. Man beachte, dass $c^t (A^t A)^{-1} c \neq 0$ gilt, da mit $A^t A$ auch $(A^t A)^{-1}$ eine positiv definite Matrix ist und $c \in \mathbb{R}^k \setminus \{0\}$ vorausgesetzt wurde. \square

Dieser Satz liefert noch nicht das, was wir wollen. Wir möchten die Verteilung von $\langle c, B - \beta \rangle$ berechnen, aber da steht $\frac{1}{\sigma}$ davor. Wir brauchen daher einen Schätzer für das unbekannte σ .

Wie es schon bei der Varianzanalyse geschehen ist, führen wir Summen von Quadraten ein. Dazu sei $M = \frac{1}{n} \sum_{i=1}^n X_i$ der Mittelwert der gemessenen Werte. Weiters sei $\hat{X} = AB$. Die Matrix A wurde gerade so gewählt, dass die i -te Komponente \hat{X}_i des Zufallsvektors \hat{X} der Wert der geschätzten Regressionsfunktion im Punkt u_i ist oder im Punkt (u_i, v_i, \dots) , wenn die Regressionsfunktion mehrere Variable hat. Das sieht man leicht, indem man AB für die zu Beginn dieses Kapitels angeführten Beispiele berechnet. Man nennt $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n$ auch die vorhergesagten Werte.

Die Summe der Quadrate der Abweichungen der gemessenen Werte X_i von den vorhergesagten Werten \hat{X}_i wird definiert durch

$$F = \|X - \hat{X}\|^2 = \sum_{i=1}^n (X_i - \hat{X}_i)^2$$

Das ist die Summe der nicht durch die Regressionsfunktion erklärten Fehleranteile in den Messwerten. Das Mittel $\frac{1}{n-k} F$ dieser Abweichungen ist ein Schätzer für die Fehlervarianz σ^2 .

Die Summe der Quadrate der Abweichungen der vorhergesagten Werte \hat{X}_i vom Mittelwert M der gemessenen Werte wird definiert durch

$$L = \sum_{i=1}^n (\hat{X}_i - M)^2$$

Man interpretiert L als die Summe der durch die Regressionsfunktion erklärten Anteile in den Messwerten. Die Summe der Quadrate der Abweichungen der gemessenen Werte X_i von ihrem Mittelwert M wird definiert durch

$$T = \sum_{i=1}^n (X_i - M)^2$$

Diese Quadratsumme hat nichts mit der Regressionsfunktion zu tun und gibt die Gesamtsumme der in den Messwerten enthaltenen Abweichungen von ihrem Mittelwert an.

Seien $e \in \mathbb{R}^n$ der Vektor, dessen Eintragungen alle gleich 1 sind. Sei V_1 der von e aufgespannte eindimensionale Teilraum des \mathbb{R}^n und V_2 der von den Spalten der Matrix A aufgespannte k -dimensionale Teilraum des \mathbb{R}^n . Wir nehmen an, dass die Regressionsfunktion die Konstante enthält, sodass der Vektor e eine Spalte der Matrix A ist. Dann gilt $V_0 = \{0\} \subset V_1 \subset V_2 \subset V_3 = \mathbb{R}^n$. Seien P_0, P_1, P_2 und P_3 die zugehörigen orthogonalen Projektionen. Dann gilt $P_3 X = X$ und $P_1 X = Me$, wie wir bereits wissen. Aus Satz 60 folgt $P_2 = A(A^t A)^{-1} A^t$, also $\hat{X} = AB = A(A^t A)^{-1} A^t X = P_2 X$. Somit erhalten wir

$$F = \|X - P_2 X\|^2, \quad L = \|P_2 X - P_1 X\|^2 \quad \text{und} \quad T = \|X - P_1 X\|^2$$

Wegen $P_2 X - P_1 X \in V_2$ sind die Vektoren $X - P_2 X$ und $P_2 X - P_1 X$ zueinander orthogonal. Es folgt $\|X - P_2 X\|^2 + \|P_2 X - P_1 X\|^2 = \|X - P_1 X\|^2$, das heißt $F + L = T$. Die Summe T

aller Abweichungen wird zerlegt in einen Anteil L , der durch die Regressionsfunktion erklärt wird, und einen Anteil F , der von sonstigen Einflüssen (Fehlern) herrührt. Man kann auch $\frac{F}{T} + \frac{L}{T} = 1$ schreiben und $\frac{L}{T}$ als den relativen Anteil der durch Regressionsfunktion erklärten Abweichungen auffassen. Er wird in Prozent angegeben, heißt Bestimmtheitsmaß, wird mit R^2 bezeichnet und ist ein Maß für die Approximationsgüte der gemessenen Werte durch die Regressionsfunktion.

Satz 67: Die Komponenten des Zufallsvektors N seien unabhängig und $N(0, 1)$ -verteilt. Für $c \in \mathbb{R}^k \setminus \{0\}$ sei $U = \frac{\langle c, B - \beta \rangle}{\sqrt{c^t(A^t A)^{-1}c} \sqrt{\frac{1}{n-k}F}}$. Dann hat U die $T(n - k)$ -Verteilung.

Beweis: Sei $\varrho = A\beta$. Dann gilt $X_i = \varrho_i + \sigma N_i$ für $1 \leq i \leq n$. Da die Zufallsvariablen $N_1, N_2, N_3, \dots, N_n$ unabhängig und $N(0, 1)$ -verteilt sind, sind auch $X_1, X_2, X_3, \dots, X_n$ unabhängig und X_i ist $N(\varrho_i, \sigma)$ -verteilt. Somit gelten die Voraussetzungen von Satz 61 mit $\varrho = A\beta$. Wegen $A\beta \in V_2$ gilt $\varrho = P_2\varrho$. Wegen Satz 61 (b) hat $\frac{1}{\sigma^2}F = \frac{1}{\sigma^2}\|X - P_2X\|^2$ die $C(n - k)$ -Verteilung. Weiters erhalten wir $B = (A^t A)^{-1}A^t X = (A^t A)^{-1}A^t P_2X$ wegen $P_2 = A(A^t A)^{-1}A^t$, das heißt B und somit auch $W = \frac{\langle c, B - \beta \rangle}{\sigma \sqrt{c^t(A^t A)^{-1}c}}$ ist eine Funktion von $P_2X = P_2X - P_0X$. Wegen Satz 61 (a) ist W unabhängig von $Z = \frac{1}{\sigma^2}F$, das eine Funktion von $X - P_2X = P_3X - P_2X$ ist. Außerdem hat W nach Satz 66 die $N(0, 1)$ -Verteilung. Somit hat $U = \frac{W}{\sqrt{Z/(n-k)}}$ die $T(n - k)$ -Verteilung nach Satz 32. \square

Mit Hilfe der Zufallsvariablen U und ihrer in Satz 67 berechneten Verteilung findet man Konfidenzintervalle. Ist $u_{\delta, m}$ wie früher definiert dann gilt $P(-u_{\delta, n-k} \leq U \leq u_{\delta, n-k}) = \gamma$ mit $\delta = \frac{1-\gamma}{2}$. Daraus berechnet man das folgende γ -Konfidenzintervall für $\langle c, \beta \rangle$.

$$\left[\langle c, B \rangle - u_{\delta, n-k} \sqrt{c^t(A^t A)^{-1}c} \sqrt{\frac{1}{n-k}F}, \langle c, B \rangle + u_{\delta, n-k} \sqrt{c^t(A^t A)^{-1}c} \sqrt{\frac{1}{n-k}F} \right]$$

Da die Regressionsfunktion linear in $\beta_1, \beta_2, \dots, \beta_k$ ist, erhält man durch Einsetzen geeigneter $c \in \mathbb{R}^k$ Konfidenzintervalle für die Funktionswerte der Regressionsfunktion. Ist zum Beispiel $x = \beta_1 + \beta_2 u$ die Regressionsfunktion, so erhält man mit $c = \begin{pmatrix} 1 \\ u \end{pmatrix}$ ein Konfidenzintervall für $\beta_1 + \beta_2 u$.

Setzt man für c den j -ten Einheitsvektor ein, dann ist $d_j = \sqrt{c^t(A^t A)^{-1}c}$ die Wurzel der j -ten Diagonaleintragung der Matrix $(A^t A)^{-1}$ und $\langle c, \beta \rangle = \beta_j$. Somit ist

$$\left[B_j - u_{\delta, n-k} d_j \sqrt{\frac{1}{n-k}F}, B_j + u_{\delta, n-k} d_j \sqrt{\frac{1}{n-k}F} \right]$$

ein γ -Konfidenzintervall für β_j .

Beispiel 33: Eine Regressionsgerade $x = \beta_1 + \beta_2 u$ kann man auch anders darstellen, indem man den Nullpunkt auf der Skala der u -Werte in den Punkt a verlegt. Man erhält dann $x = \tilde{\beta}_1 + \tilde{\beta}_2(u - a)$ mit $\tilde{\beta}_1 = \beta_1 + \beta_2 a$ und $\tilde{\beta}_2 = \beta_2$. Wie wirkt sich so eine Transformation auf die Genauigkeit der Schätzwerte, das heißt auf die Länge der Konfidenzintervalle für die Parameter aus?

Sei $e \in \mathbb{R}^n$ der Vektor, dessen Eintragungen alle 1 sind, und w der Vektor mit Komponenten u_1, u_2, \dots, u_n . Die Matrix A hat die Spalten e und w , die Matrix \tilde{A} für die transformierte Regressionsgerade hat die Spalten e und $w - ae$. Da der von e und w aufgespannte Teilraum derselbe ist wie der von e und $w - ae$ aufgespannte, ist die orthogonale Projektion P_2 auf diesen Teilraum für beide Fälle gleich. Somit ist auch $F = \|X - P_2X\|^2$ für beide Fälle gleich. Sollte es einen Unterschied in der Länge des Konfidenzintervalls geben, dann kann das nur an der Wurzel d_j der j -ten Diagonaleintragung der Matrix $(A^t A)^{-1}$ liegen. Diese

sollte möglichst klein sein. Da A die Spalten e und w hat, folgt

$$A^t A = \begin{pmatrix} n & n\bar{w} \\ n\bar{w} & \|w\|^2 \end{pmatrix} \text{ mit } \bar{w} = \frac{1}{n} \sum_{j=1}^n w_j \text{ und } \|w\|^2 = \sum_{j=1}^n w_j^2$$

Man rechnet nach, dass $s = \det A^t A = n \sum_{j=1}^n (w_j - \bar{w})^2$ gilt. Ersetzt man w durch $w - ae$, so ändert sich s nicht. Es gilt also auch $\det \tilde{A}^t \tilde{A} = s$. Die Diagonaleintragungen von $(A^t A)^{-1}$ sind daher $\frac{\|w\|^2}{s}$ und $\frac{n}{s}$ und die von $(\tilde{A}^t \tilde{A})^{-1}$ sind $\frac{\|w - ae\|^2}{s}$ und $\frac{n}{s}$. Die kürzesten Konfidenzintervalle für die Parameter erhält man, wenn man a so wählt, dass $\|w - ae\|^2$ minimal wird. Das ist der Fall, wenn $a = \bar{w}$ ist, das heißt wenn man den Nullpunkt in den Schwerpunkt der Punktmenge u_1, u_2, \dots, u_n legt.

Schließlich kann man Satz 67 auch noch dazu verwenden, um einen Test für die Hypothese $H_0 : \beta_j = 0$ anzugeben. Als Teststatistik verwendet man $U_j = \frac{B_j}{d_j \sqrt{F/(n-k)}}$. Bei Richtigkeit der Hypothese hat sie wegen Satz 67 die $T(n-k)$ -Verteilung und U_j^2 hat die $F(1, n-k)$ -Verteilung. Der Wertebereich von U_j^2 ist \mathbb{R}^+ , wobei ein großer Wert gegen die Hypothese und ein Wert nahe 0 für die Hypothese spricht. Bezeichnet man mit H die Verteilungsfunktion der $F(1, n-k)$ -Verteilung, dann ist $1 - H(U_j^2)$ die Signifikanz des Tests nach Satz 59.

Inhaltsverzeichnis

I. Rechnen mit Wahrscheinlichkeiten	1
1. Ereignisse und deren Wahrscheinlichkeit	1
2. Gleichwahrscheinliche Ausfälle	2
3. Bedingte Wahrscheinlichkeit	3
4. Axiome für die Wahrscheinlichkeit	6
II. Zufallsvariable	8
1. Zufallsvariable und Verteilungsfunktion	8
2. Binomialverteilung und geometrische Verteilung	10
3. Poissonverteilung	11
4. Exponentialverteilung und Gammaverteilung	12
5. Normalverteilung	13
6. Zufallsvektoren	14
7. Unabhängigkeit von Zufallsvariablen	17
8. Funktionen von Zufallsvariablen	20
9. Mehrdimensionale Normalverteilung	22
10. Verteilungen in der Statistik	23
11. Erwartungswert und Varianz	25
III. Transformationen und Grenzwertsätze	33
1. Integrale	33
2. Momenterzeugende Funktion (Laplacetransformation)	35
3. Charakteristische Funktion (Fouriertransformation)	36
4. Konvergenz in Verteilung	38
5. Grenzwertsätze	40
IV. Statistik	42
1. Parameterschätzer	42
2. Konfidenzintervalle	43
3. Statistische Tests	44
4. Das lineare Modell	47
5. Varianzanalyse	50
6. Lineare Regression	52

Zusammenstellung der wichtigen Verteilungen

Name	Parameter	Abkürz.	Wertebereich	Einzelwahrscheinlichkeiten Wahrscheinlichkeitsdichte	Ew.	Varianz	momentenerzeug. Funktion	charakterist. Funktion
Binomial- verteilung	$n \in \mathbb{N}$ $p \in (0, 1)$	$B(n, p)$	$\{0, 1, \dots, n\}$	$w(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(1-p + pe^t)^n$	$(1-p + pe^{it})^n$
geometrische Verteilung	$p \in (0, 1)$		$\{1, 2, \dots\}$	$w(k) = (1-p)p^{k-1}$	$\frac{1}{1-p}$	$\frac{p}{(1-p)^2}$	$\frac{(1-p)e^t}{1-pe^t}$	$\frac{(1-p)e^{it}}{1-pe^{it}}$
negative Binomialvert.	$n \in \mathbb{N}$ $p \in (0, 1)$		$\{n, n+1, \dots\}$	$w(k) = \binom{k-1}{n-1} (1-p)^n p^{k-n}$	$\frac{n}{1-p}$	$\frac{np}{(1-p)^2}$	$\frac{(1-p)^n e^{nt}}{(1-pe^t)^n}$	$\frac{(1-p)^n e^{int}}{(1-pe^{it})^n}$
Poisson- verteilung	$\lambda \in \mathbb{R}^+$	$P(\lambda)$	$\{0, 1, 2, \dots\}$	$w(k) = \frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ	$e^{-\lambda} e^{\lambda e^t}$	$e^{-\lambda} e^{\lambda e^{it}}$
Normal- verteilung	$\mu \in \mathbb{R}$ $\sigma \in \mathbb{R}^+$	$N(\mu, \sigma)$	\mathbb{R}	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{t\mu + \frac{1}{2}t^2\sigma^2}$	$e^{it\mu - \frac{1}{2}t^2\sigma^2}$
Exponential- verteilung	$\lambda \in \mathbb{R}^+$	$E(\lambda)$	\mathbb{R}^+	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$	$\frac{\lambda}{\lambda-it}$
Gamma- verteilung	$\lambda \in \mathbb{R}^+$ $r \in \mathbb{R}^+$	$G(r, \lambda)$	\mathbb{R}^+	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$(\frac{\lambda}{\lambda-t})^r$	$(\frac{\lambda}{\lambda-it})^r$
Beta- verteilung	$\alpha > 0$ $\beta > 0$		$(0, 1)$	$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta(\alpha+\beta)^{-2}}{\alpha+\beta+1}$		
T -Verteilung	$n \in \mathbb{N}$	$T(n)$	\mathbb{R}	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$	0			
F -Verteilung	$m \in \mathbb{N}$ $n \in \mathbb{N}$	$F(m, n)$	\mathbb{R}^+	$f(x) = \frac{\Gamma(\frac{m+n}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}$				