

1 Grundbegriffe der Wahrscheinlichkeitslehre

Selbst bei weitgehender Kontrolle aller Einflußgrößen sind die Ergebnisse psychologischer Datenerhebungen nicht mit Sicherheit vorherzusagen. Auch Wiederholungen unter genau gleichen Bedingungen führen in der Regel nicht zu den gleichen Ergebnissen. Die Analyse psychologischer Daten ist daher auf die Methoden der Wahrscheinlichkeitstheorie angewiesen. Dieses Kapitel gibt eine eher informelle Einführung in die grundlegenden Konzepte der Wahrscheinlichkeitstheorie. Sie stützt sich auf Brémaud (1988), Chung (1975), Gnedenko (1968), Goldberg (1969) und Hogg und Craig (1978). Das mathematische Standardwerk über Wahrscheinlichkeitstheorie ist Bauer (1991).

1.1 Mengen

Der Begriff der *Menge* wird nicht definiert, man regelt seinen Gebrauch durch Erläuterungen und Beispiele. Eine *Menge* ist eine Zusammenfassung von bestimmten, wohlunterschiedenen Objekten. Die Objekte heißen *Elemente* der Menge. Ist A eine Menge und a ein Element der Menge A , dann schreibt man $a \in A$, ist a nicht Element von A , schreibt man $a \notin A$. Manche Mengen können durch die Aufzählung ihrer Elemente angegeben werden; dies geschieht durch eine in geschweifte Klammern eingeschlossene Liste der Elemente: $A = \{a_1, \dots, a_n\}$. Allgemeiner ist die Darstellung einer Menge durch definierende Bedingungen. Soll die Menge aller a mit der Eigenschaft $E(a)$ bezeichnet werden, dann schreibt man $\{a \mid E(a)\}$. Werden mehrere Eigenschaften benötigt, dann schreibt man $\{a \mid E_1(a), \dots, E_n(a)\}$ um anzudeuten, daß alle Elemente der Menge alle Eigenschaften $E_1(a), \dots, E_n(a)$ gleichzeitig erfüllen. Einige häufig gebrauchte Mengen erhalten besondere Bezeichnungen:

1. \emptyset ist die leere Menge: $\emptyset = \{a \mid a \neq a\}$.
2. \mathbb{N} ist die Menge der natürlichen Zahlen: $\mathbb{N} = \{1, 2, 3, \dots\}$.
3. \mathbb{Z} ist die Menge der ganzen Zahlen: $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$.
4. \mathbb{R} ist die Menge der reellen Zahlen.

DEFINITION 1.1. Seien A und B Mengen.

1. Die Mengen A und B sind *identisch*, wenn sie dieselben Elemente enthalten: $A = B$ gdw. für alle a gilt $a \in A$ gdw. $a \in B$.
2. A ist eine *Teilmenge* von B , wenn jedes Element von A auch in B ist: $A \subset B$ gdw. aus $a \in A$ folgt $a \in B$. A ist eine *echte* Teilmenge von B , wenn A eine Teilmenge von B ist und es ein Element a gibt, das zwar in B , nicht aber in A enthalten ist.
3. Die *Vereinigungsmenge* $A \cup B$ von A und B ist die Menge, die alle Elemente von A und von B enthält:

$$A \cup B = \{a \mid a \in A \text{ oder } a \in B\}.$$

4. Die *Durchschnittsmenge* oder der *Durchschnitt* $A \cap B$ von A und B ist die Menge, die alle Elemente enthält, die sowohl in A als auch in B sind:

$$A \cap B = \{a \mid a \in A \text{ und } a \in B\}.$$

5. Die *Differenzmenge* $A \setminus B$ ist die Menge aller Elemente von A , die nicht in B sind:

$$A \setminus B = \{a \mid a \in A \text{ und } a \notin B\}.$$

6. Das *Komplement* von A bezüglich einer Menge Ω mit $A \subset \Omega$ ist die Menge aller Elemente von Ω , die nicht in A sind:

$$\overline{A}_\Omega = \Omega \setminus A = \{a \mid a \in \Omega \text{ und } a \notin A\}.$$

Aus einer genauen Betrachtung der oben gegebenen Definitionen lassen sich leicht folgende Rechenregeln für die definierten Mengenoperationen ableiten (die Komplementbildung bezieht sich dabei immer auf die gleiche Obermenge Ω):

1. $A \cap B = B \cap A$, $A \cup B = B \cup A$ (Kommutativität);
2. $(A \cap B) \cap C = A \cap (B \cap C)$, $(A \cup B) \cup C = A \cup (B \cup C)$ (Assoziativität);
3. $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ (Distributivität);
4. wenn $A \subset B$, dann $\overline{B} \subset \overline{A}$;
5. $\overline{(\overline{A})} = A$;
6. $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (Regeln von De Morgan).

DEFINITION 1.2. Das *kartesische Produkt* $A \times B$ von A und B ist die Menge aller geordneten Paare (a, b) , bei denen das erste Element a aus A und das zweite Element b aus B stammt.

Man beachte, daß zwei geordnete Paare (a, b) und (a', b') genau dann identisch sind, wenn sowohl a und a' , als auch b und b' identisch sind: $(a, b) = (a', b')$ gdw. $a = a'$ und $b = b'$. Statt geordneter Paare kann man allgemeiner auch geordnete n -Tupel (a_1, \dots, a_n) betrachten. Sie sind die Elemente eines n -stelligen kartesischen Produkts $A_1 \times \dots \times A_n$:

$$A_1 \times \dots \times A_n = \{(a_1, \dots, a_n) \mid a_1 \in A_1, \dots, a_n \in A_n\}.$$

Ist $A_1 = \dots = A_n$, dann schreibt man für $A_1 \times \dots \times A_n$ auch A^n .

DEFINITION 1.3. Eine Menge R ist eine *binäre Relation*, wenn es Mengen A und B gibt, so daß R eine Teilmenge von $A \times B$ ist: $R \subset A \times B$. Wir sagen dann: R ist eine *binäre Relation auf* $A \times B$. Allgemein ist eine Menge R eine *n -stellige Relation*, wenn es Mengen A_1, \dots, A_n gibt, so daß $R \subset A_1 \times \dots \times A_n$.

Ist R eine binäre Relation und (a, b) ein Element aus R , dann schreibt man statt $(a, b) \in R$ in der Regel kürzer aRb .

DEFINITION 1.4. Sei A eine Menge und R eine binäre Relation auf $A \times A$.

1. R ist *reflexiv auf* A , gdw. für alle a in A gilt aRa .
2. R ist *symmetrisch auf* A , gdw. für alle a, b in A gilt: aus aRb folgt bRa .
3. R ist *transitiv auf* A , gdw. für alle a, b, c in A gilt: aus aRb und bRc folgt aRc .

DEFINITION 1.5. Sei A eine Menge und E eine binäre Relation auf $A \times A$. Die Relation E ist eine *Äquivalenzrelation* auf A , gdw. sie auf A reflexiv, symmetrisch und transitiv ist.

Zur Bezeichnung von Äquivalenzrelationen wird häufig das Zeichen \sim verwendet. Man schreibt dann $a \sim b$ wenn $(a, b) \in \sim$.

DEFINITION 1.6. Sei A eine Menge und \sim eine Äquivalenzrelation auf A . Eine Teilmenge K von A ist eine *Äquivalenzklasse* bezüglich der Relation \sim , wenn gilt

1. $K \neq \emptyset$;
2. wenn $a \in K$ und $b \in K$, dann ist $a \sim b$;
3. wenn $a \in K$ und $a \sim b$, dann ist $b \in K$.

Ist auf einer Menge A eine Äquivalenzrelation definiert, dann ist jedes Element a von A in genau *einer* Äquivalenzklasse enthalten. Diese wird häufig mit $\llbracket a \rrbracket$ bezeichnet: $\llbracket a \rrbracket = \{b \in K \mid a \sim b\}$. Zwei Äquivalenzklassen K und K' sind entweder identisch, oder sie haben kein Element gemeinsam.

DEFINITION 1.7. Sei A eine Menge und \sim eine Äquivalenzrelation auf A . Die Menge A/\sim aller Äquivalenzklassen von A bezüglich \sim ist die von \sim *induzierte Zerlegung* von A .

Statt *Zerlegung* wird die Menge A/\sim auch *Quotientenmenge von A bezüglich \sim* genannt. Die Elemente von A/\sim sind Mengen. Ihre Vereinigung ist A , ihr paarweiser Durchschnitt ist leer.

DEFINITION 1.8. Seien A und B Mengen und f eine binäre Relation auf $A \times B$. Die Relation f heißt *Abbildung von A nach B* , wenn f die folgenden Bedingungen erfüllt:

1. f ist *linkstotal*: Für alle a in A gibt es ein b in B , so daß (a, b) in f .
2. f ist *rechtseindeutig*: Sind (a, b) und (a, c) in f , dann ist $b = c$.

Die Menge A heißt *Definitionsbereich* und die Menge B heißt *Wertebereich* der Abbildung f . Ist M eine Teilmenge von A , dann heißt

$$f(M) = \{b \in B \mid \text{es gibt ein } a \in M \text{ mit } (a, b) \in f\}$$

das *Bild* von M , und für eine Teilmenge N aus B heißt

$$f^{-1}(N) = \{a \in A \mid \text{es gibt ein } b \in N \text{ mit } (a, b) \in f\}$$

das *Urbild* von N . Die Abbildung $f|_M$ von M nach B heißt *Beschränkung* von f auf M . Sie unterscheidet sich von f nur durch den eingeschränkten Definitionsbereich.

Ist f eine Abbildung, dann schreibt man für $(a, b) \in f$ in der Regel $b = f(a)$. Statt *Abbildung* wird auch der synonyme Ausdruck *Funktion* benutzt. Um anzudeuten, daß eine Abbildung von A nach B definiert ist, verwendet man die Schreibweise $f: A \rightarrow B$. Die Zuordnung von a aus A zu $f(a)$ aus B wird durch $a \mapsto f(a)$ dargestellt.

DEFINITION 1.9. Sei $f: A \rightarrow B$ eine Abbildung.

1. f ist *surjektiv*, wenn es zu jedem b in B ein a in A gibt, so daß $b = f(a)$. Surjektive Abbildungen sind also Relationen, die sowohl links- als auch rechtstotal (*bitotal*) und rechtseindeutig sind.
2. f ist *injektiv*, wenn mit a, a' in A aus $f(a) = f(a')$ folgt $a = a'$. Injektive Abbildungen werden auch *eineindeutig* genannt, da sie sowohl rechts- als auch linkseindeutig sind. Man spricht auch von *Einbettungen*.
3. f ist *bijektiv*, wenn f surjektiv und injektiv ist. Bijektive Abbildungen sind bitotale und eineindeutige Relationen, die vor allem bei der Betrachtung von strukturgleichen Systemen (Isomorphismen) eine Rolle spielen.

DEFINITION 1.10. Sei $f: A \rightarrow B$ eine bijektive Abbildung. Dann besteht für jedes b in B das Urbild $f^{-1}(\{b\})$ aus einem einzigen Element, das mit $f^{-1}(b)$ bezeichnet wird. Dies definiert eine Abbildung $f^{-1}: B \rightarrow A$, die jedem b in B ein $f^{-1}(b)$ in A zuordnet, sie wird *Umkehrabbildung* von f genannt.

Mit Hilfe bijektiver Abbildungen kann die Mächtigkeit von Mengen definiert werden. Wir verzichten hier allerdings auf die Einführung des Begriffs der Gleichmächtigkeit und der Kardinalzahlen, da es für uns ausreicht, wenn wir zwischen endlichen und unendlichen Mengen unterscheiden können.

DEFINITION 1.11. 1. Eine Menge A ist *endlich*, wenn es eine natürliche Zahl n und eine bijektive Abbildung von A auf $\{i \mid i \in \mathbb{N}, 1 \leq i \leq n\}$ gibt.

2. Eine nichtleere Menge A ist *unendlich*, wenn sie nicht endlich ist.

3. Eine nichtleere Menge A ist *abzählbar*, wenn es eine surjektive Abbildung von \mathbb{N} auf A gibt.

4. Eine nichtleere Menge A ist *überabzählbar*, wenn sie nicht abzählbar ist.

Unendliche Mengen A lassen sich auch durch die etwas kontraintuitive Eigenschaft charakterisieren, daß sie eine echte Teilmenge M besitzen und dabei gleichzeitig eine bijektive Abbildung von A nach M existiert.

1.2 Zufallsexperimente

Ein Experiment, dessen Ergebnis nicht mit Sicherheit vorhergesagt werden kann, wird als *Zufallsexperiment* bezeichnet, wenn die Menge aller möglichen Ergebnisse des Experiments bekannt ist. Diese Menge wird als *Ergebnis-* oder *Stichprobenraum* bezeichnet.

Beispiel. In einem Detektionsexperiment soll die Versuchsperson ein sehr schwaches Signal entdecken, das in ein Störgeräusch eingebettet ist. Es gibt zwei mögliche Ergebnisse: die Versuchsperson entdeckt das Signal (+), oder sie entdeckt es nicht (-). Der Ergebnisraum ist dann die Menge $\{+, -\}$, die diese beiden Ergebnisse enthält.

Zur wahrscheinlichkeitstheoretischen Analyse eines Experiments haben wir als erstes eine geeignete Formalisierung des möglichen Geschehens im Experiment zu geben. Diese Formalisierung besteht im wesentlichen aus einer Aufstellung aller möglichen Ereignisse und einer präzisen Formulierung der zu untersuchenden Hypothesen.

1.2.1 Der Ergebnisraum

Beispiel. Die Datenerhebung für ein Experiment bestehe darin, daß ein Ehepaar befragt wird, ob es für eine generelle Geschwindigkeitsbeschränkung auf Autobahnen sei. Wir können dann als Ergebnis notieren, wieviele „Ja“-Antworten wir bei einem Paar erhalten. Die Ergebnisse des Experiments entsprechen damit jeweils genau einem Element der Menge $\Omega_1 = \{0, 1, 2\}$.

Die Menge Ω_1 ist ein Ergebnisraum des Experiments. Sie enthält alle Ergebnisse, die möglich sind und für die Fragestellung des Experiments von Bedeutung sein könnten. Der Ergebnisraum kann auch als die Menge der möglichen Datenpunkte betrachtet werden, da seine Elemente—die Ergebnisse des Experiments—in der Regel genau diejenigen Informationen sind, die bei einem Experiment aufgezeichnet werden.

Die Definition des Ergebnisraums eines Experiments hängt nicht nur von der formalen Struktur des Experiments ab, sondern wird ganz wesentlich auch von der Art der Fragestellung bestimmt. Wir werden dies in unserem Beispiel spätestens dann bemerken, wenn wir an Hand der erhobenen Daten wissen wollen, wie Ehefrauen über das angesprochene Problem denken. Die Information, wer welche Antwort gab, ginge nämlich beim Ergebnisraum Ω_1 verloren. Um diese Frage beantworten zu können, muß der Ergebnisraum die Information enthalten, welcher Ehepartner welche Antwort gegeben hat. Wir können etwa „ JN “ notieren, wenn die Frau „Ja“ und der Mann „Nein“

sagt, „ JJ “, wenn beide „Ja“ sagen usw. Damit erhalten wir den Ergebnisraum $\Omega_2 = \{JJ, JN, NJ, NN\}$.

Wir sehen an diesen verschiedenen Ergebnisräumen, daß die Wahrscheinlichkeitslehre nicht bestimmt, wie ein Experiment zu beschreiben ist, bzw. wie die Grundbegriffe der Wahrscheinlichkeitstheorie in einem Experiment interpretiert werden müssen. Die Anwendung der Wahrscheinlichkeitslehre setzt die korrekte empirische Deutung ihrer Grundbegriffe voraus und sagt selbst nur darüber etwas, wie mit diesen Grundbegriffen umzugehen ist.

Für den ersten Grundbegriff der Wahrscheinlichkeitstheorie, den Ergebnisraum, haben wir oben zwei Beispiele kennengelernt. Auch wenn aus der Wahrscheinlichkeitslehre nicht für beliebige Experimente und Fragestellungen abgeleitet werden kann, wie der Ergebnisraum zu bestimmen ist, so lassen sich doch bestimmte Kriterien dafür angeben, ob eine Menge grundsätzlich als Ergebnisraum geeignet ist oder nicht.

DEFINITION 1.12. Ein *Ergebnisraum* Ω eines Experiments ist die Menge aller möglichen Ergebnisse des Experiments. Für jeden Ergebnisraum Ω muß gelten:

1. Jedes Element der Menge Ω bezeichnet ein mögliches Ergebnis des Experiments.
2. Jedem Ergebnis des Experiments entspricht genau ein Element von Ω .

Wir sehen an unseren Beispielen, daß für ein Experiment mehrere Ergebnisräume definiert werden können, je nach gewünschter Fragestellung. Allgemein ist es empfehlenswert, einen möglichst differenzierten Ergebnisraum zu definieren, da dann auch mehr Fragestellungen untersucht werden können, wie dies an den Beispielen deutlich wurde.

1.2.2 Ereignisse

Obwohl es aus praktischen Gründen nicht möglich ist, im Ergebnisraum die Versuchssituation vollständig zu beschreiben, muß dieser immer in Abhängigkeit von der Fragestellung des Experiments definiert werden. Der Ergebnisraum ist die Grundmenge für wahrscheinlichkeitstheoretische Überlegungen zu einem Experiment. Alle Fragen, die anhand der Daten untersucht werden sollen, bauen auf dem Ergebnisraum auf. Wir können uns etwa für das Ereignis „Beide Ehepartner geben die gleiche Antwort“ interessieren. Dieses Ereignis tritt ein, wenn „ JJ “ oder wenn „ NN “ beobachtet wird. Das Ereignis „gleiche Antwort“ ist also selbst als Teilmenge von Ω_2 darstellbar. Eine mathematische Darstellung des Ereignisses „gleiche Antwort“ ist die Menge $A = \{JJ, NN\}$, eine Teilmenge von Ω_2 .

DEFINITION 1.13. Ist Ω ein Ergebnisraum, dann ist A ein *Ereignis*, wenn A eine Teilmenge von Ω ist.

Ein Ereignis ist also eine Menge und damit ein Konzept der wahrscheinlichkeitstheoretischen Betrachtungen unseres Experiments, dessen „Eintreten“ nicht direkt empirisch beobachtbar ist. Wir können aber einfach definieren:

DEFINITION 1.14. Ist A ein Ereignis, dann sagen wir *das Ereignis A tritt ein*, wenn ein Ergebnis beobachtet wird, das Element von A ist.

In der Wahrscheinlichkeitstheorie wird also zwischen den Begriffen „Ergebnis“ und „Ereignis“ sorgfältig unterschieden. Ein Ergebnis ist beobachtbar; es ist der Ausgang des Experiments oder das aufgezeichnete Datum. „Ereignis“ ist dagegen ein abstraktes Konzept der Wahrscheinlichkeitslehre, das deshalb auch oben definiert wurde. Definiert sind Ereignisse als Mengen, und sie sind deshalb etwas Theoretisches, nicht direkt Beobachtbares. Die Elemente dieser Mengen sind Ergebnisse, die im Experiment beobachtet werden können. Das Eintreten eines Ereignisses A ist deshalb indirekt dann beobachtbar, wenn ein Ergebnis beobachtet wird, das Element von A ist.

Wir können nun auch Ereignisse definieren, die nur ein einziges Element enthalten, etwa $A = \{JJ\}$. Dieses Ereignis tritt genau dann auf, wenn das Ergebnis „ JJ “ beobachtet wird.

DEFINITION 1.15. Ein Ereignis, das nur ein einziges Ergebnis enthält, heißt *Elementarereignis*.

Auch der Ergebnisraum Ω eines Experiments ist selbst ein Ereignis. Die Besonderheit dieses Ereignisses ist, daß es bei jeder Durchführung des Experiments eintritt, da ja entsprechend der Definition jedes mögliche Ergebnis des Experiments in Ω enthalten ist.

Neben Ω gibt es noch ein weiteres—zumindest für die Wahrscheinlichkeitstheorie—wichtiges Ereignis, nämlich das Ereignis \emptyset , das überhaupt kein Ergebnis enthält und demnach niemals eintritt, da entsprechend den Eigenschaften von Ω in jedem Fall ein Element von Ω beobachtet wird.

Da Ereignisse als Mengen definiert sind, können die bekannten Mengenoperationen auf sie angewandt und dadurch aus gegebenen Ereignissen neue erzeugt werden. Sind etwa A und B Ereignisse, dann ist $A \cap B$ ebenfalls ein Ereignis. Man weiß dann natürlich auch, wann das Ereignis $A \cap B$ eintritt, nämlich genau dann, wenn ein Ergebnis beobachtet wird, das sowohl in A als auch in B enthalten ist. Wir können etwa A als das Ereignis „Beide Partner geben die gleiche Antwort“ betrachten und B als das Ereignis „Die Frau antwortet mit 'Ja'“. Dann ist $A \cap B$ das Ereignis „Bei einem Ehepaar mit gleicher Meinung sagt die Frau 'Ja'“. Abgekürzt erhalten wir $A = \{JJ, NN\}$, $B = \{JN, JJ\}$ und $A \cap B = \{JJ\}$

Auch die Operation \cup kann zur Bildung neuer Ereignisse benutzt werden: A sei das Ereignis „Die Frau sagt 'Ja'“ und B sei „Der Mann sagt 'Ja'“. Wir erhalten $A = \{JJ, JN\}$, $B = \{JJ, NJ\}$ und $A \cup B = \{JJ, JN, NJ\}$, so daß $A \cup B$ das Ereignis „Mindestens ein Partner antwortet mit 'Ja'“ darstellt.

Die dritte Operation zur Bildung von Ereignissen ist die Komplementbildung. Stellt A das Ereignis „Beide Partner sind einer Meinung“ dar, dann ist $A = \{JJ, NN\}$, und das Komplement $\bar{A} = \{JN, NJ\}$ stellt dann das Ereignis „Die Partner sind unterschiedlicher Meinung“ dar.

Für einen endlichen Ergebnisraum Ω , der n Ergebnisse enthält, können genau 2^n Ereignisse gebildet werden, da eine Menge mit n Elementen genau 2^n verschiedene Teilmengen enthält.

Wir betrachten hier vorerst nur endliche (später auch abzählbare) Ergebnisräume. In diesem Fall ist leicht zu sehen, daß den Elementarereignissen eine besondere Bedeutung zukommt. Mit Hilfe der Mengenvereinigung ist es nämlich möglich, jedes Ereignis als Vereinigung aller Elementarereignisse zu schreiben, die genau die Elemente des zu erzeugenden Ereignisses enthalten. Das Ereignis „Mindestens ein 'Ja'“ mit $A = \{JN, NJ, JJ\}$ läßt sich etwa schreiben als $A = \{JN\} \cup \{NJ\} \cup \{JJ\}$. Diese Möglichkeit wird dann besonders interessant, wenn wir den Ereignissen Wahrscheinlichkeiten zuordnen können.

Ist nämlich dann die Wahrscheinlichkeit für jedes Elementarereignis bekannt, und ist darüber hinaus bekannt, wie man die Wahrscheinlichkeit des Ereignisses $A \cup B$ errechnet, wenn die Wahrscheinlichkeiten von A und B bekannt sind, dann kann für jedes beliebige Ereignis, das im Experiment eintreten kann, die Wahrscheinlichkeit berechnet werden.

Spätestens an dieser Stelle wird auch klar, warum so großer Wert auf den Unterschied zwischen dem Begriff *Ergebnis* („ JJ “) und dem Begriff *Elementarereignis* ($\{JJ\}$) gelegt wurde. Der Grund liegt darin, daß man schreiben kann $\{JJ\} \cup \{NN\}$, aber nicht „ JJ “ \cup „ NN “, da nur $\{JJ\}$ eine Menge ist, nicht jedoch „ JJ “. Mit Elementarereignissen kann man im Sinne der Mengenlehre rechnen, mit Ergebnissen nicht.

1.2.3 Wahrscheinlichkeiten

Am Anfang dieser Einführung wurde darauf hingewiesen, daß die Wahrscheinlichkeitstheorie nicht sagt, welcher Ergebnisraum für ein bestimmtes Experiment zu definieren ist. Sie setzt jedoch gewisse Randbedingungen, die eine Menge Ω erfüllen muß, um als Ergebnisraum brauchbar zu sein. Ähnlich ist es bei der Wahrscheinlichkeit, dem zweiten wesentlichen Grundbegriff der Wahrscheinlichkeitstheorie. Die Wahrscheinlichkeitslehre sagt nichts darüber aus, wie man die Wahrscheinlichkeit eines Ereignisses in

einem Experiment erhalten kann. Sie stellt jedoch bestimmte Bedingungen, die Funktionen erfüllen müssen, um als Wahrscheinlichkeiten bestimmter Ereignisse gelten zu können. Darüber hinaus wird festgelegt, wie mit den Wahrscheinlichkeiten umzugehen ist, also wie und was mit ihnen berechnet werden kann.

Für die Definition des Begriffs „Wahrscheinlichkeit“ betrachten wir vorerst *diskrete* Ergebnisräume. Dies sind Ergebnisräume, die nur endlich oder höchstens abzählbar viele Ergebnisse enthalten.

DEFINITION 1.16. Sei Ω ein diskreter Ergebnisraum. Ferner sei P eine Funktion, die jedem Elementarereignis $\{\omega_i\}$ mit $\omega \in \Omega$ genau eine Zahl $P(\{\omega_i\})$ zuordnet. Die Funktionswerte $P(\{\omega_i\})$ heißen *Wahrscheinlichkeit des Elementarereignisses* $\{\omega_i\}$ genau dann, wenn sie folgende zwei Bedingungen erfüllen:

1. Die Wahrscheinlichkeit jedes Elementarereignisses ist nicht negativ: für alle Elementarereignisse $\{\omega_i\}$ ist $P(\{\omega_i\}) \geq 0$;
2. Die Summe aller den Elementarereignissen zugeordneten Wahrscheinlichkeiten ist 1:

$$P(\{\omega_1\}) + P(\{\omega_2\}) + \dots = 1.$$

Die Wahrscheinlichkeit beliebiger Ereignisse A wird dann in folgender Weise definiert:

DEFINITION 1.17. Die *Wahrscheinlichkeit eines Ereignisses* A ist die Summe der Wahrscheinlichkeiten aller Elementarereignisse $\{\omega\}$, die Teilmengen von A sind:

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Mit dieser Definition kann die Wahrscheinlichkeit aller möglichen Ereignisse berechnet werden. Wir mußten dazu keine Annahme über die speziellen Werte der Wahrscheinlichkeiten von Elementarereignissen machen, insbesondere brauchen diese nicht gleichwahrscheinlich sein.

1.3 Der Wahrscheinlichkeitsraum

Mit den Begriffen „Ereignis“, „Elementarereignis“ und „Wahrscheinlichkeit eines Ereignisses“ sind die drei wesentlichen Grundbegriffe der Wahrscheinlichkeitstheorie eingeführt. Wir sind dabei von diskreten Ergebnisräumen ausgegangen und haben in den Definitionen Begriffe benutzt, die vorher nicht definiert, sondern durch inhaltliche Erläuterungen eingeführt wurden, wie etwa den Begriff „Experiment“ oder „Ergebnis“. Diesen Zugang nennt man in der Mengenlehre oder der Wahrscheinlichkeitstheorie „naiv“, da er keine exakte formale Begründung der Theorie erlaubt. Deshalb soll im folgenden die axiomatische Definition der Wahrscheinlichkeit von Kolmogorov betrachtet werden. Diese ist präziser und allgemeiner als die bisher gegebenen Definitionen und stellt die wichtigste Grundlage der Wahrscheinlichkeitstheorie dar. Die Definition wird schrittweise eingeführt.

1.3.1 Die Axiomatische Definition nach Kolmogorov

DEFINITION 1.18. Sei Ω eine Menge und \mathfrak{A} eine Menge von Teilmengen von Ω , die folgende drei Bedingungen erfüllt:

1. Ω ist Element von \mathfrak{A} .
2. Ist A Element von \mathfrak{A} , dann ist auch \overline{A} , das Komplement von A bezüglich Ω , Element von \mathfrak{A} .
3. Falls jedes Element der Folge $A_1, \dots, A_n, \dots, n \geq 1$ in \mathfrak{A} ist, dann ist auch die Vereinigung $\bigcup_{i=1}^{\infty} A_i$ in \mathfrak{A} .

Die Menge \mathfrak{A} heißt dann σ -Algebra (in Ω).

Ist Ω ein Ergebnisraum, dann enthält eine σ -Algebra \mathfrak{A} in Ω alle interessierenden Ereignisse. Definition 1.18 garantiert, daß jede σ -Algebra abgeschlossen gegenüber den Mengenoperationen Komplementbildung, Vereinigung und Durchschnitt ist. Sind bestimmte Teilmengen von Ω in \mathfrak{A} enthalten, dann sind auch alle Mengen, die durch die Anwendung von Mengenoperationen aus diesen erzeugt werden können, in \mathfrak{A} enthalten (vgl. die Übungsaufgaben am Ende dieses Abschnittes).

Jedem Ereignis, also jedem Element in \mathfrak{A} , soll durch die Wahrscheinlichkeitsfunktion eine Zahl $0 \leq P \leq 1$ zugeordnet werden, die die Wahrscheinlichkeit des Ereignisses angibt. Die Zuordnung soll so geartet sein, daß sie in einem bestimmten Sinn mit den Mengenoperationen verträglich ist.

DEFINITION 1.19. Sei Ω ein Ergebnisraum und \mathfrak{A} eine zu Ω gehörige σ -Algebra. Ferner sei P eine Abbildung von \mathfrak{A} in die reellen Zahlen. Das Tripel $\langle \Omega, \mathfrak{A}, P \rangle$ heißt genau dann *Wahrscheinlichkeitsraum*, wenn die Funktion P die folgenden drei Axiome erfüllt:

1. $P(A) \geq 0$ für alle A in \mathfrak{A} .
2. $P(\Omega) = 1$.
3. Für jede Folge A_1, \dots, A_n, \dots paarweise disjunkter Ereignisse aus \mathfrak{A} gilt σ -Additivität:

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

Aufgrund dieser Definition wird eine Wahrscheinlichkeit auch ein *normiertes, nichtnegatives, vollständig additives Maß* genannt.

Aus der Definition eines Wahrscheinlichkeitsraums lassen sich einige Eigenschaften ableiten, deren Beweise gute Übungsaufgaben bilden:

1. $\emptyset \in \mathfrak{A}$. Da Ω in \mathfrak{A} ist, muß auch dessen Komplement, die leere Menge in \mathfrak{A} sein.
2. Die Menge aller Teilmengen von Ω ist eine σ -Algebra in Ω .
3. $P(\emptyset) = 0$.
4. $P(A) \leq 1$.
5. $P(\bar{A}) = 1 - P(A)$.
6. Wenn $A \subset B$, dann $P(A) \leq P(B)$.
7. Für beliebige Ereignisse A, B gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Die Definition 1.19 bestimmt den Rahmen für die Formalisierung eines Zufallsexperiments. Wie bereits früher bei der Darstellung des naiven Konzepts von Wahrscheinlichkeiten, wird auch durch diese Definition nicht bestimmt, welche konkreten Werte die Wahrscheinlichkeiten bestimmter Ereignisse haben. Es werden nur strukturelle Forderungen an Wahrscheinlichkeitsfunktionen und Ergebnisräume gestellt. Die empirische Interpretation des Tripels $\langle \Omega, \mathfrak{A}, P \rangle$ ist weiterhin frei, solange die strukturellen Erfordernisse erfüllt sind. Für diese Interpretation sind mehrere Möglichkeiten bekannt. Wir stellen hier nur zwei kurz vor, da sie in der Literatur auch häufig als „Definitionen“ der Wahrscheinlichkeit bezeichnet werden. Aus der Sicht von Definition 1.19 stellen sie aber nur spezielle Interpretationen des Konzepts „Wahrscheinlichkeit“ dar.

Beispiel. Wir betrachten ein Zufallsexperiment, das aus dem einmaligen Werfen von zwei Würfeln besteht. Der Ergebnisraum Ω ist die Menge aller Paare der Form $\omega = (i, j)$, wobei i und j die Ziffern 1 bis 6 sind: $\Omega = \{1, 2, 3, 4, 5, 6\}^2$. Als σ -Algebra wird die Menge aller Teilmengen von Ω definiert. Die Wahrscheinlichkeitsfunktion sei bestimmt durch $P(\{\omega\}) = (\frac{1}{6})^2$ für jedes ω in Ω . Es ist leicht zu sehen, daß dann für alle Ereignisse A in \mathfrak{A} gilt: $P(A) = (\frac{1}{6})^2 |A|$, wobei $|A|$ die Anzahl von Elementen in A ist. Das Ereignis „eine Summe von 3“ ist $A = \{(1, 2), (2, 1)\}$, seine Wahrscheinlichkeit ist $P(A) = \frac{1}{18}$.

1.3.2 Die Interpretation nach Laplace

Die klassische Interpretation der Wahrscheinlichkeit nach Laplace geht von den Elementarereignissen aus. Dies sind, wie bereits früher definiert, Mengen, die nur ein einziges Ergebnis enthalten und paarweise disjunkt sind. Von jedem Elementarereignis wird angenommen, daß es die gleiche Wahrscheinlichkeit wie jedes andere Elementarereignis hat. Um nun die Wahrscheinlichkeit für ein beliebiges Ereignis A zu finden, wird die Menge der Elementarereignisse in zwei disjunkte Teilmengen zerlegt: die erste Teilmenge enthält alle Ergebnisse, deren Realisierung gleichbedeutend ist mit dem Eintreten von A , die zweite Teilmenge enthält die Ergebnisse, deren Realisierung das Eintreten von \bar{A} bedeutet. Die erste Teilmenge wird als die Menge der für A günstigen Fälle bezeichnet, die Anzahl ihrer Elemente sei m . Die Gesamtzahl aller möglichen Elementarereignisse sei n . Als Wahrscheinlichkeit des Ereignisses A wird dann definiert

$$P(A) = \frac{m}{n},$$

also die Anzahl der dem Ereignis A günstigen Fälle dividiert durch die Gesamtzahl aller möglichen Elementarereignisse. Es kann gezeigt werden, daß diese Definition die Bedingungen von Definition 1.19 erfüllt.

Die Nachteile des Konzepts von Laplace sind aber offensichtlich:

- Es kann nur auf endliche Mengen von Elementarereignissen angewandt werden.
- Es setzt voraus, daß alle Elementarereignisse gleichwahrscheinlich sind.

Vor allem letztere Einschränkung macht die klassische Definition nach Laplace für empirische Anwendungen unbrauchbar. Sie kann bestenfalls zur Analyse von Glücksspielen benutzt werden, wobei selbst dort „ideale“ Spielgeräte, wie etwa Würfel oder Glücksräder angenommen werden müssen. In der Regel sind empirische Elementarereignisse nicht gleichwahrscheinlich, sondern ihre Wahrscheinlichkeit hängt von den empirischen Interpretationen oder Bedingungen ab.

1.3.3 Grenzwerte relativer Häufigkeiten nach von Mises

Das Konzept der statistischen Wahrscheinlichkeitstheorie nach von Mises geht nicht von der Struktur des Ergebnisraums aus, sondern versucht eine rein empirische Begründung zu geben. In diesem System werden Wahrscheinlichkeiten als Grenzwerte relativer Häufigkeiten definiert, wenn die Anzahl der unabhängigen Beobachtungen unendlich groß wird:

$$P(A) = \lim_{n \rightarrow \infty} \frac{m}{n},$$

wobei n die Gesamtzahl der Experimente ist und m die Anzahl der Experimente, in denen das zufällige Ereignis A beobachtet wird.

Es ist klar, daß der Anspruch dieser Theorie, eine rein empirische Begründung des Wahrscheinlichkeitskonzepts zu geben, nicht eingelöst wird, da weder aus theoretischen Gründen folgt, daß der Grenzwert des Quotienten $\frac{m}{n}$ tatsächlich gegen einen festen Wert konvergiert, noch diese Annahme empirisch begründet werden kann, da sie eine unendliche Anzahl von Beobachtungen voraussetzt.

Übungsaufgaben

1. Zeigen Sie folgende Konsequenzen von Definition 1.18:

- (a) $\emptyset \in \mathfrak{A}$.
 (b) Falls jedes Element der Folge A_1, \dots, A_n in \mathfrak{A} ist, dann ist auch der Durchschnitt $\bigcap_{i=1}^n A_i$ in \mathfrak{A} . Hinweis: Benutzen Sie die Formel von de Morgan:

$$\bigcap_{i=1}^n A_i = \overline{\left(\bigcup_{i=1}^n \overline{A_i} \right)}.$$

2. Zeigen Sie, daß aus den Axiomen der Definition 1.19 folgt: $P(\overline{A}) = 1 - P(A)$. Hinweis: $\Omega = A \cup \overline{A}$.
 3. Zeigen Sie, daß aus $A \subset B$ folgt, $P(A) \leq P(B)$. Hinweis: Aus $A \subset B$ folgt, daß $A \cup (B - A) = B$.
 4. Zeigen Sie für beliebige Ereignisse A, B : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
 5. Aus $A = \emptyset$ folgt $P(A) = 0$. Gilt dies auch umgekehrt?
 6. Zeigen Sie, daß der klassische Wahrscheinlichkeitsbegriff von der Wahrscheinlichkeit als Quotient von Anzahl der günstigen durch Anzahl der möglichen Fälle die Bedingungen 1 - 3 aus Definition 1.19 erfüllt.
 7. Nehmen Sie an, die Grenzwerte

$$\lim_{n \rightarrow \infty} \frac{m}{n}$$

für die Quotienten der Anzahl m von Fällen, in denen ein Ereignis A eintritt, und der Gesamtzahl der Experimente n existieren. Zeigen Sie, daß diese Definition die Axiome 1 bis 3 von Definition 1.19 erfüllt.

1.4 Bedingte Wahrscheinlichkeiten

Beispiel. Eine Versuchsperson bekommt eine Liste von Vokabeln, beispielsweise englisch-deutsche Wortpaare, für eine Minute zum Lernen. Danach wird sie abgefragt, wobei ihr jeweils das deutsche Wort vom Versuchsleiter vorgelegt wird, und sie muß das zugehörige englische Wort sagen. Der gesamte Zyklus wird dreimal durchgeführt. Bei jedem Durchgang wird notiert, ob die Versuchsperson beim Test einen Fehler gemacht hat (F) oder nicht (R). Ein Ergebnis des Experiments könnte dann sein: „FFR“. Als Ergebnisraum kann die Menge $\Omega = \{FFF, FFR, FRF, FRR, RFF, RFR, RRF, RRR\}$ definiert werden. Ein Ereignis ist dann etwa „nach dem ersten Versuch keinen Fehler mehr“: $A = \{FRR, RRR\}$

Nun soll angenommen werden, daß es für jedes Elementarereignis $\{\omega\}$ dieses Experiments eine Wahrscheinlichkeit $P(\{\omega\})$ gibt. Eine mögliche Frage ist dann etwa die nach der Wahrscheinlichkeit eines Fehlers in einem bestimmten Durchgang, da man erwartet, daß sich diese Wahrscheinlichkeit im Laufe des Experiments ändert. Darüber hinaus kann die Wahrscheinlichkeit eines Fehlers aber auch davon abhängen, was im vorhergehenden Durchgang geschah, denn wenn die Versuchsperson im zweiten Durchgang alle Vokabeln gelernt hat, dann wird im dritten ihre Fehlerwahrscheinlichkeit anders sein, als wenn sie im zweiten Durchgang auch noch Fehler hatte. Wenn nach dem zweiten Durchgang die Wahrscheinlichkeit eines Fehlers im dritten Durchgang angegeben werden soll, dann kann die Kenntnis der Antworten aus dem zweiten Durchgang diese Wahrscheinlichkeiten erheblich verändern. Das Konzept der „bedingten Wahrscheinlichkeit“ dient dazu, den Einfluß von Vorinformation auf die Bestimmung von Wahrscheinlichkeiten noch ausstehender Ereignisse zu untersuchen. Als Vorinformation dient dabei die Kenntnis von bereits eingetretenen Ereignissen.

DEFINITION 1.20. Seien A und B Ereignisse, wobei $P(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit* $P(A|B)$ von A unter der Bedingung B definiert durch

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

1.4.1 Stochastische Unabhängigkeit

Die bedingte Wahrscheinlichkeit $P(A|B)$ gibt somit die Wahrscheinlichkeit für das Ereignis A in den Fällen an, in denen B bereits eingetreten ist. Nicht immer ist jedoch die Kenntnis bereits eingetretener Ereignisse hilfreich für die Vorhersage zukünftiger Ereignisse:

DEFINITION 1.21. Zwei Ereignisse A und B heißen genau dann *stochastisch unabhängig*, wenn $P(A|B) = P(A)$. Ist der Zusammenhang klar, dann werden die beiden Ereignisse statt *stochastisch unabhängig* häufig einfach *unabhängig* genannt.

Bei unabhängigen Ereignissen A und B führt das Wissen, daß B eingetreten ist, nicht zu einer Veränderung der Wahrscheinlichkeit von A . Die Wahrscheinlichkeit, daß A und B gleichzeitig eintreten, ist dann einfach $P(A \cap B) = P(A)P(B)$ oder äquivalent dazu $P(A|B) = P(A|\overline{B})$. Die Definition der Unabhängigkeit kann auf Familien von Ereignissen ausgedehnt werden.

DEFINITION 1.22. Sei \mathfrak{C} eine Familie von Ereignissen (also eine Menge von Ereignissen). Die Familie \mathfrak{C} ist eine *Familie unabhängiger Ereignisse*, wenn für alle endlichen Teilfamilien $\{A_1, \dots, A_n\}$ von \mathfrak{C} gilt

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i).$$

Paarweise Unabhängigkeit (Def. 1.21) in einer Familie von Ereignissen hat nicht automatisch auch die Unabhängigkeit der gesamten Familie zur Folge. Dies zeigt Übungsaufgabe 1 am Ende dieses Abschnitts.

Ein weit verbreitetes Mißverständnis des Konzepts der stochastischen Unabhängigkeit betrifft den Zusammenhang zwischen „disjunkt“ und „unabhängig“. Disjunkte Ereignisse sind *nicht* unabhängig. Im Gegenteil, disjunkte Ereignisse können nicht unabhängig sein, denn aus der Disjunktheit folgt, daß das zweite Ereignis nicht eintreten kann, wenn das erste eingetreten ist. Wären diese Ereignisse unabhängig, dann müßte gelten

$$P(A \cap B) = P(\emptyset) = 0 = P(A)P(B),$$

eines der Ereignisse müßte daher die Wahrscheinlichkeit 0 haben. „Disjunkt“ ist also gleichbedeutend mit „unverträglich“.

1.4.2 Die Formel von Bayes

Will man von der bedingten Wahrscheinlichkeit eines Ereignisses zur unbedingten Wahrscheinlichkeit übergehen, dann kann dies durch die Betrachtung aller möglichen Bedingungen erreicht werden, unter denen das entsprechende Ereignis eintreten kann. Seien B_i Elemente einer Zerlegung von Ω , also eine Reihe paarweise disjunkter Ereignisse: $B_i \cap B_j = \emptyset$, falls $i \neq j$, mit $i, j = 1, \dots, n$, wobei die Vereinigung aller B_i gleich Ω sei:

$$\Omega = \bigcup_{i=1}^n B_i.$$

Dann gilt

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

und wegen dem Distributivgesetz auch

$$A = \bigcup_{i=1}^n A \cap B_i.$$

Wegen der Additivität der Wahrscheinlichkeit, und da alle B_i disjunkt sind, gilt dann für die Wahrscheinlichkeit von A

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(A | B_i)P(B_i). \end{aligned}$$

Diese Gleichung wird *Formel der totalen Wahrscheinlichkeit* genannt. Sie dient zusammen mit der Definition der bedingten Wahrscheinlichkeit zur Herleitung der „Formel von Bayes“. Aufgrund der Definition der bedingten Wahrscheinlichkeit gilt nämlich:

$$\begin{aligned} P(A | B)P(B) &= P(A \cap B) \\ &= P(B | A)P(A). \end{aligned}$$

Durch Umformen erhält man daraus

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

Wird hier das Ereignis B als „Ursache“ des beobachtbaren Ereignisses A betrachtet und kennt man die Wahrscheinlichkeit $P(A | B)$, dann kann damit die Wahrscheinlichkeit der „Ursache“ B bei gegebener Beobachtung A berechnet werden.

Setzt man in Gleichung (1.1) für $P(A)$ die Formel der totalen Wahrscheinlichkeit ein, dann erhält man die Formel von Bayes:

$$P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^n P(A | B_i)P(B_i)}. \quad (1.1)$$

Die Formel von Bayes kann zur adaptiven Schätzung der Wahrscheinlichkeit von Hypothesen benutzt werden. Sei A ein Ereignis, das unter verschiedenen Bedingungen eintritt, über die es die Hypothesen B_1, \dots, B_n gibt und deren a priori Wahrscheinlichkeiten $P(B_1), \dots, P(B_n)$ bekannt seien. Ferner seien die Wahrscheinlichkeiten des Auftretens von A bei Gültigkeit der Hypothesen B_i bekannt, also die bedingten Wahrscheinlichkeiten $P(A | B_i)$. Tritt nun bei einem Zufallsexperiment das Ereignis A ein, dann lassen sich mit Hilfe der Formel von Bayes die Wahrscheinlichkeiten der Hypothesen B_i neu berechnen.

1.4.3 Ereignisfolgen

Mit Hilfe eines Induktionsbeweises läßt sich folgende Beziehung zeigen:

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

In dieser Formel steht $P(A_1, \dots, A_n)$ für $P(\bigcap_{i=1}^n A_i)$. Betrachtet man die A_i als eine in der Zeit auftretende Folge von Ereignissen, dann können mit dieser Formel spezielle stochastische Prozesse untersucht werden. So spricht man etwa von einem *Markov-Prozess*, wenn $P(A_i | A_1, \dots, A_{i-1}) = P(A_i | A_{i-1})$. Jedes Ereignis A_i hängt in einem Markov-Prozess nur vom direkt vorausgehenden, nicht aber von früheren Ereignissen ab. Markov-Prozesse bilden die Grundlage mehrerer stochastischer Lernmodelle im Rahmen der Reiz-Stichproben-Theorie (Wickens, 1982).

1.4.4 Bedingte Unabhängigkeit

DEFINITION 1.23. Sei C ein Ereignis aus dem Wahrscheinlichkeitsraum $\langle \Omega, \mathfrak{A}, P \rangle$ mit positiver Wahrscheinlichkeit. Wir definieren eine Abbildung von der σ -Algebra \mathfrak{A} in \mathbb{R} durch

$$P_C(A) = P(A|C).$$

Dann ist $\langle \Omega, \mathfrak{A}, P_C \rangle$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse A und B , die bezüglich der so definierten Wahrscheinlichkeit P_C unabhängig sind, heißen *bedingt unabhängig bezüglich C* .

Für Ereignisse A und B , die bezüglich C bedingt unabhängig sind, gilt nach dieser Definition

$$P(A \cap B | C) = P(A | C)P(B | C).$$

Das Konzept der bedingten Unabhängigkeit ist eine wesentliche Grundlage der psychologischen Testtheorie. Dort gibt es Modelle, die annehmen, daß die richtigen Lösungen zweier Testaufgaben durch eine Person bezüglich dieser einen Person bedingt unabhängige Ereignisse sind. Diese Eigenschaft wird dort „lokal stochastische Unabhängigkeit“ genannt.

Übungsaufgaben

1. [Aus Brémaud (1988, S. 14)] Sei $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ ein Ergebnisraum und \mathfrak{A} die Menge aller Teilmengen von Ω . Auf \mathfrak{A} sei die Wahrscheinlichkeitsfunktion P definiert durch $P(\{\omega_i\}) = \frac{1}{4}$ für $i = 1, \dots, 4$. Die Ereignisse A , B und C seien folgendermaßen definiert: $A = \{\omega_1, \omega_2\}$, $B = \{\omega_2, \omega_3\}$ und $C = \{\omega_1, \omega_3\}$. Zeigen Sie daß $\mathfrak{C} = \{A, B, C\}$ keine Familie unabhängiger Ereignisse ist, obwohl A , B und C paarweise unabhängig sind.

1.5 Zufallsvariablen

Die Analyse von Zufallsexperimenten mit Hilfe von Mengen und Wahrscheinlichkeiten ist verhältnismäßig umständlich, da für allgemeine Mengen nur sehr einfache Operationen zur Verfügung stehen, die das Theoretisieren erschweren. Man versucht daher Regeln zu finden, nach denen die Elemente des Ergebnisraums durch Zahlen dargestellt werden können. Diese Zuordnung von Zahlen zu Ergebnissen soll es ermöglichen, die strukturellen Eigenschaften des Wahrscheinlichkeitsraums numerisch zu analysieren. Die wesentliche Restriktion bei der Zuordnung von Zahlen zu Ergebnissen ist, daß den Mengen, die im Bereich der Zahlen gebildet werden können, Mengen von Ergebnissen—Ereignisse—entsprechen. Einer Menge von Zahlen entspricht dann das Ereignis, das alle Ergebnisse enthält, deren Zahlenwert in der Zahlenmenge enthalten ist. Aus technischen Gründen wird hier als Bildmenge $\overline{\mathbb{R}}$ üblicherweise die um $\{-\infty, +\infty\}$ erweiterte Menge der reellen Zahlen \mathbb{R} benutzt: $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

1.5.1 Reelle Zufallsvariablen

DEFINITION 1.24. Sei $\langle \Omega, \mathfrak{A}, P \rangle$ ein Wahrscheinlichkeitsraum. Eine Abbildung X von Ω in $\overline{\mathbb{R}}$ ist eine *Zufallsvariable* genau dann, wenn für jede reelle Zahl x die Menge $\{\omega | X(\omega) \leq x\}$ in \mathfrak{A} ist.

Die Bedingung, daß unter einer Zufallsvariablen X das Urbild eines jeden Intervalls der Form $(-\infty, x] = \{y | y \leq x\}$ ein Ereignis sein muß, nennt man auch *\mathfrak{A} -Meßbarkeit* der Funktion X . Sie ermöglicht es, die Wahrscheinlichkeit P von der σ -Algebra \mathfrak{A} auf den Wertebereich der Zufallsvariablen X zu übertragen.

DEFINITION 1.25. Sei X eine Zufallsvariable. Die Funktion

$$F(x) = P(\{\omega \mid X(\omega) \leq x\})$$

heißt *Verteilungsfunktion* der Zufallsvariablen X .

Statt $P(\{\omega \mid X(\omega) \leq x\})$ wird auch häufig $P(X \leq x)$ geschrieben. Verteilungsfunktionen haben folgende allgemeine Eigenschaften:

1. Sie sind monoton steigend: wenn $x \leq x'$, dann ist $F(x) \leq F(x')$.
2. Für kleine Werte von x nähert sich $F(x)$ an den Wert 0 an:

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$

3. Für große Werte von x nähert sich $F(x)$ an den Wert 1 an:

$$\lim_{x \rightarrow \infty} F(x) = 1.$$

Wegen der letzten beiden Bedingungen kann der Definitionsbereich einer Verteilungsfunktion von \mathbb{R} auf $\overline{\mathbb{R}}$ ausgedehnt werden. Man setzt $F(-\infty) = 0$ und $F(\infty) = 1$.

DEFINITION 1.26. Zufallsvariablen, die nur Werte aus \mathbb{R} annehmen, werden *reelle Zufallsvariablen* genannt. Falls eine reelle Zufallsvariable eine Verteilungsfunktion F besitzt, für die eine nichtnegative Funktion f existiert, so daß gilt

$$F(x) = \int_{-\infty}^x f(y) dy,$$

dann sagt man, daß X eine *Wahrscheinlichkeitsdichte besitzt*. Die Funktion f wird *Wahrscheinlichkeitsdichte* genannt.

Es läßt sich leicht zeigen, daß aus der Definition 1.26 einer Dichtefunktion folgt:

$$P(a \leq X \leq b) = \int_a^b f(y) dy$$

und

$$\int_{-\infty}^{+\infty} f(y) dy = 1.$$

1.5.2 Diskrete Zufallselemente

DEFINITION 1.27. Sei E eine abzählbare Menge und $\langle \Omega, \mathfrak{A}, P \rangle$ ein Wahrscheinlichkeitsraum. Eine Abbildung X von Ω in die Menge E , so daß für alle x in E die Menge $\{\omega \mid X(\omega) = x\}$ in \mathfrak{A} ist, heißt *diskretes Zufallselement* von E . Ist E eine Teilmenge der reellen Zahlen \mathbb{R} , dann wird X auch *diskrete Zufallsvariable* genannt.

Auf dem Wertebereich E eines diskreten Zufallselements kann die Funktion

$$\begin{aligned} p(x) &= P(\{\omega \mid X(\omega) = x\}) \\ &= P(X = x) \end{aligned}$$

definiert werden, sie wird *Wahrscheinlichkeitsfunktion* des diskreten Zufallselements X genannt. Für eine Teilmenge A von E gilt damit

$$\begin{aligned} P(X \in A) &= P(\{\omega \mid X(\omega) \in A\}) \\ &= \sum_{x \in A} p(x). \end{aligned}$$

Für die Verteilungsfunktion $F(x)$ einer diskreten Zufallsvariablen X gilt

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{y \leq x} p(y). \end{aligned}$$

In der Regel werden wir nur solche diskreten Zufallselemente betrachten, deren Wertebereich eine Teilmenge von \mathbb{R} ist, die also diskrete Zufallsvariablen sind. Diskrete Zufallselemente mit anderen Wertebereichen werden in der psychologischen Testtheorie benutzt, um die zufällige Auswahl einer Testperson aus einer Population zu beschreiben.

1.5.3 Indikatorfunktionen

Sei $\langle \Omega, \mathfrak{A}, P \rangle$ ein Wahrscheinlichkeitsraum und A ein Ereignis daraus. Eine Indikatorfunktion 1_A ist eine Funktion von Ω in $\{0, 1\}$, definiert durch

$$1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases} \quad (1.2)$$

Damit ist $X = 1_A$ eine diskrete Zufallsvariable mit dem Wertebereich $E = \{0, 1\}$, die Ereignisse $\{\omega \mid X(\omega) = 1\}$ und A sind identisch. Es ist also $P(X = 1) = P(A)$ und $P(X = 0) = 1 - P(A)$.

1.5.4 Unabhängige Zufallsvariablen

DEFINITION 1.28. Reelle Zufallsvariablen X_1, \dots, X_n werden als (*stochastisch*) *unabhängig* bezeichnet, wenn für alle x_1, \dots, x_n aus \mathbb{R} gilt

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n).$$

Die Schreibweise $P(X_1 \leq x_1, \dots, X_n \leq x_n)$ ist eine Abkürzung für

$$P\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right).$$

Falls die Zufallsvariablen X_i Dichten f_i besitzen, dann gilt

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(y_1) \cdots f_n(y_n) dy_1 \cdots dy_n.$$

Die Dichte von (X_1, \dots, X_n) ist also das Produkt der Dichten aller X_i :

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i).$$

Diskrete Zufallselemente X und Y sind genau dann unabhängig, wenn für alle x aus dem Wertebereich von X und alle y aus dem Wertebereich von Y gilt

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Das Konzept der Unabhängigkeit kann auch auf Familien von Zufallsvariablen ausgedehnt werden: Man nennt eine Familie \mathfrak{H} von Zufallsvariablen unabhängig, wenn für jede endliche Teilfamilie $\{X_1, \dots, X_n\} \subset \mathfrak{H}$ die Bedingung

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

für alle reellen Zahlen x_i , $1 \leq i \leq n$ gilt.

1.5.5 Zufallsstichproben

Eine Zufallsstichprobe erhält man, wenn in einer Population eine Folge von Zufallsexperimenten durchgeführt wird. Eine einzelne Beobachtung besteht im Registrieren eines Ergebnisses und des Wertes der damit verbundenen Zufallsvariablen. Wesentlich dabei ist, daß jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit hat, beobachtet zu werden. Die einzelnen Beobachtungen müssen stochastisch unabhängige Ereignisse darstellen. Die Wiederholungen stellen also unabhängige Wiederholungen eines einzigen Zufallsexperiments dar.

DEFINITION 1.29. Eine *Zufallsstichprobe* vom Umfang n ist eine Folge X_1, \dots, X_n stochastisch unabhängiger und identisch verteilter Zufallsvariablen.

Für alle Zufallsvariablen X_i einer Zufallsstichprobe $X_1, \dots, X_i, \dots, X_n$ gibt es daher eine bestimmte Verteilungsfunktion F , so daß

$$P(X_i \leq x) = F(x).$$

Außerdem sind alle Zufallsvariablen X_i und X_j der Zufallsstichprobe mit $i \neq j$ stochastisch unabhängig.

1.6 Verteilungsparameter

1.6.1 Modus, Median, Quantile einer Zufallsvariablen

DEFINITION 1.30. Ist X eine reelle Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$ und der Verteilungsfunktion $F(x)$, dann ist

1. jeder Wert x_{Mod} , an dem $f(x)$ maximal ist, ein *Modus* oder *Modalwert* von X ,
2. jeder Wert x_π mit $F(x_\pi) = \pi$ ein π -*Quantil* von X .

Ist X ein Zufallselement mit einem abzählbaren Wertebereich E und einer auf E definierten Wahrscheinlichkeitsfunktion $p(x)$, dann ist

1. jeder Wert x_{Mod} , an dem $p(x)$ maximal ist, ein *Modus* von X ,
2. jeder Wert x_π mit

$$p(X \leq x_\pi) = \pi \text{ und } p(X \geq x_\pi) = 1 - \pi$$

ein π -*Quantil* von X .

Das 0.5-Quantil von X heißt *Median* von X .

1.6.2 Erwartungswert und Varianz

DEFINITION 1.31. Ist X eine reelle Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$, dann ist, falls der Ausdruck existiert,

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

der *Erwartungswert* von X und

$$\begin{aligned} \sigma^2(X) &= \mathcal{E}[(X - \mathcal{E}(X))^2] \\ &= \int_{-\infty}^{\infty} [x - \mathcal{E}(X)]^2 f(x) dx \end{aligned}$$

die *Varianz* von X . Die *Standardabweichung* von X ist $\sigma(X)$, die positive Wurzel aus der Varianz.

DEFINITION 1.32. Sei X ein Zufallselement mit einem abzählbaren Wertebereich E und einer auf E definierten Wahrscheinlichkeitsfunktion $p(x)$. Ferner sei h eine ebenfalls auf E definierte, reellwertige Funktion, so daß

$$\sum_{x \in E} |h(x)|p(x) < \infty. \quad (1.3)$$

Dann ist

$$\mathcal{E}[h(X)] = \sum_{x \in E} h(x)p(x)$$

der Erwartungswert von $h(X)$. *Varianz* und *Standardabweichung* von X sind definiert wie in Definition 1.31.

Bei diskreten Zufallsvariablen, also diskreten Zufallselementen, deren Wertebereich eine Teilmenge von \mathbb{R} ist, betrachtet man häufig den Erwartungswert von $h(x) = x$, der identischen Abbildung.

Beispiel. Wir betrachten einen Wurf mit einer idealen Münze. Dem Ergebnis $\omega_1 = \text{„Kopf“}$ wird die Zahl $X(\omega_1) = 1$, dem Ergebnis $\omega_2 = \text{„Zahl“}$ die Zahl $X(\omega_2) = 0$ zugeordnet. Die Wahrscheinlichkeiten seien $P(\{\omega_i\}) = 0.5$ für $i = 1, 2$. Die Wahrscheinlichkeitsfunktion von X ist dann durch $p(0) = 0.5$ und $p(1) = 0.5$ definiert. Der Erwartungswert von X ist

$$\mathcal{E}(X) = p(0)0 + p(1)1 = p(1) = 0.5.$$

Die Varianz von X ist

$$\sigma^2(X) = p(0)(0 - 0.5)^2 + p(1)(1 - 0.5)^2 = 0.25.$$

Hier einige Regeln für das Rechnen mit Erwartungswerten. Dabei werden a und b als konstante Zahlen betrachtet und X und Y als Zufallsvariablen.

1. Der Erwartungswert einer Konstanten ist die Konstante selbst:

$$\mathcal{E}(a) = a.$$

2. Der Erwartungswert ist ein linearer Operator:

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y). \quad (1.4)$$

Für die Varianz gelten folgende Regeln, wobei a wiederum eine konstante Zahl sein soll:

- 1.

$$\sigma^2(a) = 0; \quad (1.5)$$

$$\sigma^2(X + a) = \sigma^2(X); \quad (1.6)$$

$$\sigma^2(aX) = a^2\sigma^2(X). \quad (1.7)$$

2. Zum Berechnen der Varianz einer Zufallsvariablen benutzt man die Formel

$$\sigma^2(X) = \mathcal{E}(X^2) - [\mathcal{E}(X)]^2. \quad (1.8)$$

Sie läßt sich folgendermaßen ableiten:

$$\begin{aligned} \sigma^2(X) &= \mathcal{E}[(X - \mathcal{E}(X))^2] \\ &= \mathcal{E}[X^2 - 2X\mathcal{E}(X) + (\mathcal{E}(X))^2] \\ &= \mathcal{E}(X^2) - \mathcal{E}[2X\mathcal{E}(X)] + \mathcal{E}[(\mathcal{E}(X))^2] \\ &= \mathcal{E}(X^2) - 2\mathcal{E}(X)\mathcal{E}(X) + [\mathcal{E}(X)]^2 \\ &= \mathcal{E}(X^2) - [\mathcal{E}(X)]^2. \end{aligned}$$

Der Schritt von $\mathcal{E}[2X\mathcal{E}(X)]$ nach $2\mathcal{E}(X)\mathcal{E}(X)$ ist berechtigt, weil im ersten Ausdruck $\mathcal{E}(X)$ eine Konstante ist und daher vor den Erwartungswertoperator gezogen werden darf.

1.6.3 Kovarianz und Korrelationskoeffizient

DEFINITION 1.33. Seien X und Y zwei Zufallsvariablen. Der Erwartungswert

$$\sigma(X, Y) = \mathcal{E}[(X - \mathcal{E}(X))(Y - \mathcal{E}(Y))]$$

heißt *Kovarianz von X und Y* .

Es läßt sich zeigen, daß

$$\begin{aligned}\sigma(X, Y) &= \mathcal{E}[(X - \mathcal{E}(X))(Y - \mathcal{E}(Y))] \\ &= \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y).\end{aligned}$$

DEFINITION 1.34. Sind X und Y Zufallsvariablen, dann heißt

$$\rho(X, Y) = \frac{\sigma(X, Y)}{\sigma(X)\sigma(Y)}$$

der *Korrelationskoeffizient* von X und Y .

Für 3 Zufallsvariablen X, Y, Z und reelle Konstanten a, b gelten folgende Regeln:

1.
$$\sigma^2(aX + bY) = a^2\sigma^2(X) + b^2\sigma^2(Y) + 2ab\sigma(X, Y). \quad (1.9)$$

Diese Beziehung ergibt sich mit Gl. (1.8):

$$\begin{aligned}\sigma^2(aX + bY) &= \mathcal{E}[(aX + bY)^2] - [\mathcal{E}(aX + bY)]^2 \\ &= \mathcal{E}[(aX)^2 + 2aXbY + (bY)^2] - [\mathcal{E}(aX) + \mathcal{E}(bY)]^2 \\ &= \mathcal{E}[(aX)^2] - [\mathcal{E}(aX)]^2 + \mathcal{E}[(bY)^2] - [\mathcal{E}(bY)]^2 \\ &\quad + 2\mathcal{E}[aXbY] - 2\mathcal{E}(aX)\mathcal{E}(bY) \\ &= \sigma^2(aX) + \sigma^2(bY) + 2\sigma(aX, bY) \\ &= a^2\sigma^2(X) + b^2\sigma^2(Y) + 2ab\sigma(X, Y).\end{aligned}$$

Analog hierzu läßt sich zeigen, daß

$$\sigma^2(X - Y) = \sigma^2(X) + \sigma^2(Y) - 2ab\sigma(X, Y). \quad (1.10)$$

2. Sind X und Y stochastisch unabhängig, dann gilt

$$\mathcal{E}(XY) = \mathcal{E}(X)\mathcal{E}(Y).$$

3. Sind X und Y stochastisch unabhängig, dann ist die Varianz $\sigma^2(X + Y)$ der Summe gleich der Summe der Varianzen: $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$. Dies gilt nicht im allgemeinen Fall, wenn $\sigma(X, Y) \neq 0$.

4. $\sigma(X, Y) = \sigma(Y, X)$.

5. Sind X_1, \dots, X_n beliebige Zufallsvariablen, dann gilt für die Varianz der Summe:

$$\sigma^2\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sum_{j=1}^n \sigma(X_i, X_j). \quad (1.11)$$

6. Für die Kovarianz der Summe $X + Y$ und der Variablen Z gilt

$$\begin{aligned}\sigma(X + Y, Z) &= \mathcal{E}[(X + Y)Z] - \mathcal{E}(X + Y)\mathcal{E}(Z) \\ &= \mathcal{E}(XZ + YZ) - [\mathcal{E}(X)\mathcal{E}(Z) + \mathcal{E}(Y)\mathcal{E}(Z)] \\ &= \sigma(X, Z) + \sigma(Y, Z)\end{aligned} \quad (1.12)$$

1.7 Das Gesetz der großen Zahlen

Gnedenko (1968) schreibt:

Die umfangreichen von der Menschheit angesammelten Erfahrungen zeigen, daß Erscheinungen, die eine Wahrscheinlichkeit nahe Eins besitzen, fast immer eintreten. [...] Dieser Tatbestand spielt für viele praktische Schlußfolgerungen aus der Wahrscheinlichkeitsrechnung eine grundlegende Rolle, da die erwähnte *Erfahrungstatsache* es in der Praxis gestattet, wenig wahrscheinliche Ereignisse für *praktisch unmöglich* und Ereignisse mit Wahrscheinlichkeit nahe Eins für *praktisch sicher* anzunehmen. Doch kann man auf die ganz natürliche Frage, wie groß eine Wahrscheinlichkeit sein muß, damit man ein Ereignis für praktisch unmöglich halten kann, keine eindeutige Antwort geben. Das ist auch verständlich, da man im praktischen Leben die Wichtigkeit der Ereignisse berücksichtigen muß, mit denen man es zu tun bekommt (Gnedenko, 1968, S 185).

Das Gesetz der großen Zahlen besteht aus Aussagen über Ereignisse, die mit einer Wahrscheinlichkeit von 1, also „fast sicher“, eintreten. Ein Beispiel für eine solche Aussage ist folgende: Seien X_i , $i = 1, \dots, n$ Zufallsvariablen und

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

deren arithmetisches Mittel. Wenn es eine reelle Konstante $a > 0$ gibt, so daß für beliebige $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X} - a| \geq \varepsilon) = 0,$$

dann sagt man, *die Zufallsgröße \bar{X} konvergiert in Wahrscheinlichkeit gegen a .*

1.7.1 Die Tschebyschewsche Ungleichung

Sei X ein diskretes Zufallselement mit dem Wertebereich E und h eine reelle Funktion auf E , die die Gleichung (1.3) erfüllt. Dann gilt für $a > 0$ die *Markovsche Ungleichung*:

$$P(|h(X)| \geq a) \leq \frac{\mathcal{E}[|h(x)|]}{a} \quad (1.13)$$

Begründung. Sei $C = \{x \mid |h(x)| \geq a\}$ dann ist

$$\begin{aligned} |h(X)| &= 1_C(x)|h(x)| + 1_{\bar{C}}(x)|h(x)| \\ &\geq 1_C(x)|h(x)| \\ &\geq 1_C(x)a, \end{aligned}$$

da falls $x \in C$ wegen der Definition von C gilt $|h(x)| \geq a$. Es ergibt sich

$$\begin{aligned} \mathcal{E}[|h(X)|] &\geq \mathcal{E}[a 1_C(X)] = a \mathcal{E}[1_C(X)] \\ &= a P(X \in C) \\ &= a P(|h(X)| \geq a), \end{aligned}$$

da $X \in C$ gdw. $|h(X)| \geq a$. \square

Setzt man in der Markovschen Ungleichung (1.13) $h(X) = [X - \mathcal{E}(X)]^2$ und $a = \varepsilon^2$ mit $\varepsilon > 0$, so erhält man die in der psychologischen Testtheorie häufig benutzte Tschebyschewsche Ungleichung:

$$P(|X - \mathcal{E}(X)| \geq \varepsilon) \leq \frac{\sigma^2(X)}{\varepsilon^2}. \quad (1.14)$$

Es läßt sich leicht zeigen, daß sowohl die Markovsche als auch die Tschebyschewsche Ungleichung auch für stetige Zufallsvariablen gelten.

1.7.2 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz ist einer der wichtigsten Sätze der Wahrscheinlichkeitstheorie. Er macht eine Aussage über die Summe beliebig verteilter, unabhängiger Zufallsvariablen: Sind X_i , $i = 1, \dots, n$ unabhängig verteilte Zufallsvariablen, und ist

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

deren Summe, dann ist die Zufallsvariable S_n approximativ normalverteilt mit dem Erwartungswert $\mathcal{E}(S_n)$ und der Varianz $\sigma^2(S_n)$.

Übungsaufgaben

1. Was ist der Erwartungswert einer Indikatorfunktion?
2. Zeigen Sie folgende Eigenschaften von Indikatorfunktionen:
 - (a) $1_{A \cap B} = \min(1_A, 1_B) = 1_A \cdot 1_B$;
 - (b) $1_{A \cup B} = \max(1_A, 1_B)$;
 - (c) $1_{\overline{A}} = 1 - 1_A$, da $1_\Omega = 1$;

1.8 Mehrdimensionale Zufallsvariablen

1.8.1 Gemeinsame Wahrscheinlichkeitsfunktionen und Randwahrscheinlichkeiten

Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und X und Y Zufallselemente auf Ω mit den Wertebereichen E_X und E_Y . Die auf $E_X \times E_Y$ definierte Funktion

$$\begin{aligned} p(x, y) &= P(X = x, Y = y) \\ &= P(\{\omega \mid X(\omega) = x, Y(\omega) = y\}) \end{aligned}$$

ist die *gemeinsame Wahrscheinlichkeitsfunktion* des Zufallsvektors (X, Y) . Um die Wahrscheinlichkeitsfunktion von X alleine zu erhalten, braucht man nur die gemeinsame Wahrscheinlichkeitsfunktion $p(x, y)$ über alle möglichen Werte von Y zu summieren:

$$P(X = x) = \sum_{y \in E_Y} p(x, y) = p(x, *).$$

Man nennt dann $p(x, *)$ die *Randwahrscheinlichkeitsfunktion* von X .

Im Fall stetiger Zufallsvariablen X und Y sagt man, daß diese eine *gemeinsame* Wahrscheinlichkeitsdichte f besitzen, wenn es eine Funktion f gibt, so daß

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, du \, dv$$

für alle reellen Zahlen x und y . Für die *Randwahrscheinlichkeiten* $P(X \leq x)$ gilt dann

$$P(X \leq x) = \int_{-\infty}^x f(u, *) \, du$$

mit

$$f(u, *) = \int_{-\infty}^{\infty} f(u, v) \, dv,$$

der *Randwahrscheinlichkeitsdichte* von X . Analoge Vereinbarungen können natürlich für Y getroffen werden.

1.8.2 Bedingte Wahrscheinlichkeitsfunktionen und bedingte Erwartungswerte

Seien X und Y diskrete Zufallselemente mit den Wertebereichen E_X und E_Y und der gemeinsamen Wahrscheinlichkeitsfunktion $f(x, y)$. $f(x)$ und $f(y)$ seien die Randwahrscheinlichkeitsfunktionen von X und Y . Sei

$$A = \{(x, y) \mid x = x', -\infty < y < \infty\}$$

und

$$B = \{(x, y) \mid -\infty < x < \infty, y = y'\},$$

wobei x' und y' so gewählt seien, daß

$$P(A) = P(X = x') = f(x') > 0.$$

Dann ist

$$\begin{aligned} P(B \mid A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(X = x', Y = y')}{P(X = x')} \\ &= \frac{f(x', y')}{f(x')}. \end{aligned}$$

Die Wahrscheinlichkeit, daß $Y = y'$ unter der Bedingung, daß $X = x'$, ist also gleich dem Quotienten $f(x', y')/f(x')$.

DEFINITION 1.35. Ist $f(x) > 0$, dann ist

$$f(y \mid x) = \frac{f(x, y)}{f(x)}$$

die *bedingte Wahrscheinlichkeitsfunktion des diskreten Zufallselements Y* , gegeben für das diskrete Zufallselement X gilt $X = x$. Entsprechend ist

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

für $f(y) > 0$ die *bedingte Wahrscheinlichkeit des diskreten Zufallselements X* gegeben $Y = y$.

Für stetige Zufallsvariablen X und Y , die eine gemeinsame Wahrscheinlichkeitsdichte $f(x, y)$ und Randwahrscheinlichkeitsdichten $f(x)$ und $f(y)$ besitzen, werden analog zum diskreten Fall die bedingten Wahrscheinlichkeitsdichten

$$f(y \mid x) = \frac{f(x, y)}{f(x)}$$

für $f(x) > 0$ und

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

für $f(y) > 0$ definiert. Die Funktionen $f(y \mid x)$ und $f(x \mid y)$ sind nichtnegativ, und es gilt

$$\begin{aligned} \int_{-\infty}^{\infty} f(y \mid x) dy &= \int_{-\infty}^{\infty} \frac{f(x, y)}{f(x)} dy \\ &= \frac{1}{f(x)} \int_{-\infty}^{\infty} f(x, y) dy \\ &= \frac{1}{f(x)} f(x) \\ &= 1. \end{aligned}$$

Damit ist gezeigt, daß $f(y \mid x)$ die Eigenschaften einer Wahrscheinlichkeitsdichte besitzt. Es können daher auch Wahrscheinlichkeiten und Erwartungswerte mit Hilfe von $f(y \mid x)$ berechnet werden. Wir erhalten etwa

$$P(Y < y \mid X = x) = \int_{-\infty}^y f(u \mid x) du.$$

Der Erwartungswert

$$\mathcal{E}(Y | x) = \int_{-\infty}^{\infty} y f(y | x) dy$$

wird als *bedingter Erwartungswert von Y, gegeben X = x* bezeichnet.

Im Fall diskreter Zufallsvariablen werden statt Integrationen einfach Summationen benutzt: Sind X und Y diskrete Zufallselemente mit den Wertebereichen E_X und E_Y , und ist h eine reellwertige Funktion auf E_Y , dann ist

$$\mathcal{E}(h(Y) | X = x) = \sum_{y \in E_Y} h(y) P(Y = y | X = x)$$

der bedingte Erwartungswert von $h(Y)$, gegeben $X = x$. Wegen der Formel der totalen Wahrscheinlichkeit gilt

$$P(Y = y) = \sum_{x \in E_X} P(X = x) P(Y = y | X = x).$$

Multipliziert man hier beide Seiten mit $h(y)$ und summiert über E_Y , erhält man

$$\sum_{y \in E_Y} h(y) P(Y = y) = \sum_{y \in E_Y} h(y) \sum_{x \in E_X} P(X = x) P(Y = y | X = x),$$

was äquivalent ist zu

$$\mathcal{E}[h(Y)] = \sum_{x \in E_X} P(X = x) \mathcal{E}[h(Y) | X = x].$$

Es gilt also

$$\mathcal{E}[h(Y)] = \mathcal{E}[\mathcal{E}(h(Y) | X)]. \quad (1.15)$$

Übungsaufgaben

1. Beweisen Sie, daß Verteilungsfunktionen monoton steigend sind.
2. X und Y seien zwei unabhängige, diskrete Zufallselemente. Zeigen Sie daß $\mathcal{E}[h(X)g(Y)] = \mathcal{E}[h(X)] \mathcal{E}[g(Y)]$ gilt.

1.9 Bedingte Erwartungen

Wir betrachten hier einen Wahrscheinlichkeitsraum $\langle \Omega, \mathfrak{A}, P \rangle$ mit einer Zufallsvariablen $X: \Omega \rightarrow \mathbb{R}$ (Def. 1.24) und einem diskreten Zufallselement $H: \Omega \rightarrow E_H$ (Def. 1.27). Die bedingten Erwartungswerte $\mathcal{E}(X | H(\omega) = h)$ sind Parameter der bedingten Verteilung von X, gegeben $H = h$. Da H ein Zufallselement, also eine Funktion auf Ω ist, kann auch der bedingte Erwartungswert als Funktion auf Ω betrachtet werden. Diese Funktion wird als *bedingte Erwartung* bezeichnet. Sie ist, da sie eine Funktion auf Ω ist, eine Zufallsvariable.

DEFINITION 1.36. Die *bedingte Erwartung* einer Zufallsvariablen X, gegeben das diskrete Zufallselement H, ist die Zufallsvariable $T_X: \Omega \rightarrow \mathbb{R}$, die auf folgende Weise definiert ist: für jedes $\omega \in \Omega$ gilt

$$T_X(\omega) = t_X[H(\omega)],$$

wobei $t_X: E_H \rightarrow \mathbb{R}$ für alle h in E_H definiert ist durch

$$t_X(h) = \mathcal{E}(X | H = h).$$

Die bedingte Erwartung ist also eine Zufallsvariable, d. h. eine Abbildung von Ω in \mathbb{R} . Die Funktionswerte der Funktion T_X sind die bedingten Erwartungswerte $\mathcal{E}(X | H(\omega) = h)$.

Sei \sim_H eine auf Ω definierte Äquivalenzrelation mit $\omega_1 \sim_H \omega_2$ gdw. $H(\omega_1) = H(\omega_2)$. Zwei Ergebnisse ω_1 und ω_2 des Zufallsexperiments sind

äquivalent im Sinne der Äquivalenzrelation \sim_H , wenn sie den gleichen Funktionswert H haben. Für die Äquivalenzklasse, in der sich ein Element ω befindet, schreibt man $[\omega]$. Man kann dann die Zerlegung Ω/\sim_H von Ω bezüglich \sim_H betrachten, also die Menge aller \sim_H -Äquivalenzklassen. Auf ihr ist die durch H induzierte σ -Algebra \mathfrak{B} , eine Teilmenge von \mathfrak{A} aufgebaut. Die bedingte Erwartung T_X ist dann eine \mathfrak{B} -meßbare Zufallsvariable. T_X ist auf jeder Äquivalenzklasse $[\omega] \subset \Omega$ konstant und gleich dem bedingten Erwartungswert von X , gegeben $[\omega]$:

$$T_X(\omega) = \mathcal{E}(X | [\omega]).$$

Da \mathfrak{B} eine σ -Algebra über der Quotientenmenge Ω/\sim_H ist, heißt \mathfrak{B} -Meßbarkeit, daß T_X auf jeder \sim_H -Äquivalenzklasse $[\omega]$ konstant ist. Wir werden im folgenden zur Vereinfachung der Schreibweise das Argument ω der Funktion $T_X(\omega)$ weglassen, wie das bei Zufallsvariablen üblich ist, falls der Definitionsbereich klar aus dem Kontext hervorgeht.

1.9.1 Rechenregeln für bedingte Erwartungen

1. Die bedingte Erwartung ist ein linearer Operator. Seien X und Y beliebige Zufallsvariablen, deren Erwartungswerte existieren, und a sei eine beliebige reelle Zahl. Dann gilt

$$\begin{aligned} T_{X+Y}(\omega) &= \mathcal{E}(X + Y | [\omega]) \\ &= \mathcal{E}(X | [\omega]) + \mathcal{E}(Y | [\omega]) \\ &= T_X(\omega) + T_Y(\omega) \end{aligned} \quad (1.16)$$

$$\begin{aligned} T_{aX}(\omega) &= \mathcal{E}(aX | [\omega]) \\ &= a \mathcal{E}(X | [\omega]) \\ &= a T_X(\omega). \end{aligned} \quad (1.17)$$

2. Seien X und Y Zufallsvariablen, so daß $\mathcal{E}(X)$ und $\mathcal{E}(XY)$ existieren und Y auf jedem Element von Ω/\sim_H konstant ist. Dann gilt

$$\begin{aligned} T_{XY}(\omega) &= \mathcal{E}(XY | [\omega]) \\ &= Y(\omega) \mathcal{E}(X | [\omega]) \\ &= Y T_X(\omega). \end{aligned} \quad \begin{array}{l} \text{da } Y \text{ auf } [\omega] \text{ konstant} \\ (1.18) \end{array}$$

Als Spezialfälle hiervon ergeben sich

$$T_Y = Y \quad (1.19)$$

$$T_{T_X} = T_X. \quad (1.20)$$

3. Wegen $\mathcal{E}(T_X | \Omega) = \mathcal{E}(X)$ gilt

$$\mathcal{E}(T_X) = \mathcal{E}(X). \quad (1.21)$$

4. Aus (1.16) und (1.21) folgt

$$T_{X - \mathcal{E}(X)} = T_X - \mathcal{E}(T_X). \quad (1.22)$$