

Seminar Methoden der Mustererkennung und Klassifikation
Grundlagen der Wahrscheinlichkeitsrechnung und
Statistik

Linda Briesemeister
Barbara van Schewick

27. November 1995

Inhaltsverzeichnis

1	Begriffsbildung	3
1.1	Stochastik	3
1.2	Wahrscheinlichkeitsrechnung/-theorie	3
2	Historisches	3
3	Wahrscheinlichkeit	3
3.1	Endliche Wahrscheinlichkeitsräume	3
3.1.1	Wahrscheinlichkeitsraum und -verteilung	3
3.1.2	Wahrscheinlichkeitsfunktion	4
3.1.3	Eigenschaften	5
3.1.4	Häufigkeiten	5
3.1.5	Laplace-Experiment	5
3.2	Definition und Eigenschaften der bedingten Wahrscheinlichkeiten	5
3.3	Unabhängigkeit	6
4	Zufallsgrößen	7
4.1	Zufallsgrößen	7
4.2	Wahrscheinlichkeitsverteilungen diskreter Zufallsgrößen	7
4.2.1	Wahrscheinlichkeitsfunktion	7
4.2.2	Verteilungsfunktion	7
4.3	Wahrscheinlichkeitsverteilungen stetiger Zufallsgrößen	8
4.3.1	Dichtefunktion	8
4.3.2	Verteilungsfunktion	8
4.4	Parameter von Wahrscheinlichkeitsverteilungen	9
4.4.1	Erwartungswert	9
4.4.2	Varianz	9
4.5	Wahrscheinlichkeitsverteilungen mehrdimensionaler Zufallsgrößen	9
4.5.1	Mehrdimensionale Zufallsgrößen	9
4.5.2	Wahrscheinlichkeits- und Dichtefunktion zweidimensionaler Zufallsgrößen	10
4.5.3	Parameter zweidimensionaler Verteilungen	10
5	Induktive Statistik	11
5.1	Grundbegriffe	11
5.2	Schätzverfahren	12
5.2.1	Punktschätzung	12
5.2.2	Intervallschätzung	12

1 Begriffsbildung

1.1 Stochastik

Stochastik kommt von dem griechischen Wort *stochastikos* „im Erraten geschickt“ und ist ein Sammelbegriff für die Gebiete Wahrscheinlichkeitstheorie und Mathematische Statistik.

1.2 Wahrscheinlichkeitsrechnung/-theorie

In der Wahrscheinlichkeitstheorie untersucht man zufällige Prozesse mit festen als bekannt angenommenen steuernden Wahrscheinlichkeiten [Kren91].

2 Historisches

Die ersten Anfänge der Wahrscheinlichkeitsrechnung gehen bis in die Mitte des 17. Jahrhunderts zurück und sind eng mit den Namen der Mathematiker Blaise PASCAL und Pierre de FERMAT verbunden. Ein begeisterter Spieler, der Chevalier von MERÉ, wendete sich mit einer Aufgabe, die für ihn von großer praktischer Bedeutung war und die er vergeblich zu lösen versucht hatte, an PASCAL und bat um dessen Hilfe: „Zwei Spieler wollen so viele Partien miteinander spielen, bis einer von beiden zuerst m Partien gewonnen hat. Das Spiel wird aber vorzeitig abgebrochen, und zwar nachdem der eine $n < m$ und der andere $p < m$ Partien gewonnen hat. Wie ist in diesem Fall der Spieleinsatz gerecht zu verteilen?“ PASCAL beschäftigte sich mit diesem Problem und teilte in einem Brief FERMAT seine Lösung mit. Ein Entsprechendes tat FERMAT. Schließlich gab auch Christian HYUGENS ein Ergebnis an.

Diese Gelehrten sahen schon die Rolle der Wissenschaft voraus, die die Gesetzmäßigkeiten der zufälligen Erscheinungen untersucht. Durch den geringen Entwicklungsstand der Naturwissenschaften spielten jedoch die Glücksspiele noch lange die entscheidende Rolle bei der Bildung der Begriffe und Methoden. Erst im 19. Jahrhundert machte der rasch einsetzende Aufschwung der Naturwissenschaften einen Aufbau der Wahrscheinlichkeitsrechnung über den Rahmen der Glücksspiele hinaus notwendig.

3 Wahrscheinlichkeit

Zentral ist der Begriff des **Zufallexperiments**, bei dem der Versuchsausgang von unabhängig angestellten Wiederholungen nicht notwendig stets der gleiche ist, also nicht determiniert ist. Um diese Zufallsexperimente mathematisch untersuchen zu können, müssen geeignete Modelle gefunden werden, innerhalb derer man durch mathematische Methoden Aussagen machen kann. Wie gut ein Modell die Realität widerspiegelt, hängt oft von Erfahrung ab und läßt sich später nur empirisch prüfen.

3.1 Endliche Wahrscheinlichkeitsräume

3.1.1 Wahrscheinlichkeitsraum und -verteilung

Zunächst betrachten wir Zufallsexperimente mit endlich vielen möglichen Versuchsausgängen ω , die zusammengefaßt die endliche **Ergebnismenge** Ω bilden. Wir

identifizieren $A \subseteq \Omega$ mit dem **Ereignis**, daß ein $\omega \in A$ das beobachtete **Ergebnis** ist. Damit läßt sich mengentheoretische Notation einsetzen: So bedeutet $A \cap B$ das Ereignis, daß sich A und B ereignen, sowie $A \cup B$, daß sich A oder B ereignet. Das Komplement A^C von A in Ω bezeichnet das Ereignis, daß A nicht geschieht. Zwei Ereignisse A und B heißen **unvereinbar**, wenn die Mengen A und B disjunkt sind. Die leere Menge \emptyset heißt auch das **unmögliche** Ereignis; Ω das **sichere** Ereignis. Die Menge aller Ereignisse ist die Potenzmenge $\mathcal{P}(\Omega)$.

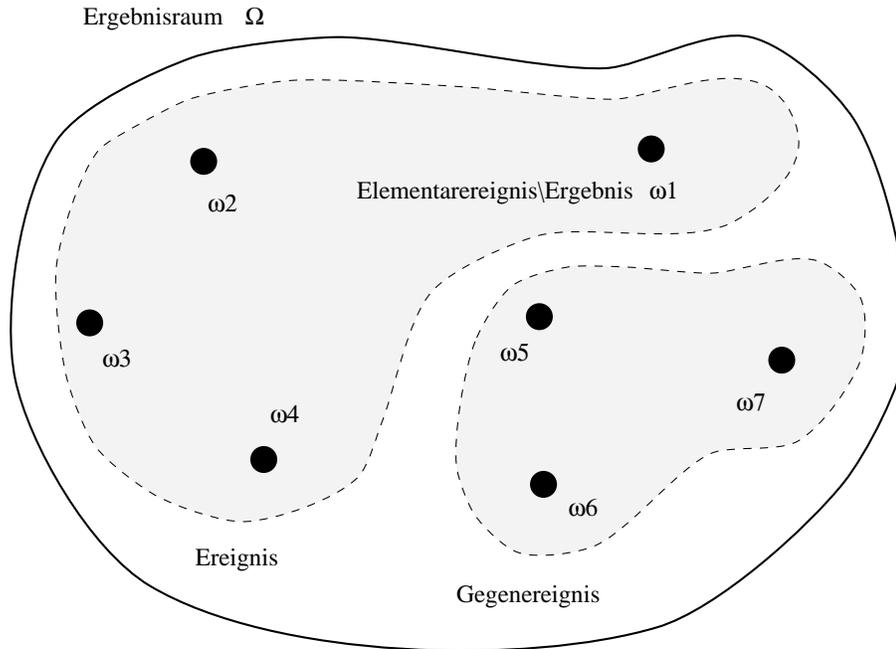


Abbildung 1: Endliche Wahrscheinlichkeitsräume

Eine Abbildung $P: \mathcal{P}(\Omega) \rightarrow [0..1]$ heißt **Wahrscheinlichkeitsverteilung**, wenn sie folgende Eigenschaften hat:

- (1) $P(\Omega) = 1$
- (2) $P(A) \geq 0$ für alle $A \in \mathcal{P}(\Omega)$
- (3) $P(A \cup B) = P(A) + P(B)$ für alle disjunkten $A, B \in \mathcal{P}(\Omega)$ (**Additivität**)

Dieses Axiomensystem stammt von KOLMOGOROW und ist äquivalent zu folgenden Eigenschaften:

$$\sum_{\omega \in \Omega} P(\{\omega\}) = 1 \quad \text{und}$$

$$P(\{\omega\}) \geq 0 \quad \text{für alle } \omega \in \Omega$$

$P(A)$ heißt **Wahrscheinlichkeit** von A ; das Paar (Ω, P) heißt der dem Experiment zugeordnete **Wahrscheinlichkeitsraum**.

3.1.2 Wahrscheinlichkeitsfunktion

Die Wahrscheinlichkeit eines Ereignisses A ist die Summe der Wahrscheinlichkeiten der Ergebnisse, bei denen A eintritt (folgt aus der Additivität). P ist also durch

die Werte aller $P(\{\omega\})$ mit $\omega \in \Omega$ bestimmt und die Abbildung $\omega \rightarrow P(\{\omega\})$ heißt **Wahrscheinlichkeitsfunktion**. Die Wahrscheinlichkeitsverteilung kann somit durch die zugehörige Wahrscheinlichkeitsfunktion angegeben werden und umgekehrt.

3.1.3 Eigenschaften

Eine Wahrscheinlichkeitsverteilung besitzt immer folgende Eigenschaften, die leicht mittels der Axiome (1) bis (3) beweisbar sind.

Für beliebige Ereignisse $A \in \mathcal{P}(\Omega)$ gilt

$$(4) \quad P(\bar{A}) = 1 - P(A), \text{ speziell ist} \\ P(\emptyset) = 0.$$

Das Axiom (3) läßt sich auch auf beliebige $A, B \in \mathcal{P}(\Omega)$ erweitern

$$(5) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (\text{Additionssatz}).$$

3.1.4 Häufigkeiten

Wird ein Zufallsexperiment tatsächlich n -mal ausgeführt, so unterscheidet man zwischen der **absoluten Häufigkeit** $k_n(A)$, also der Zahl der Versuchsausgänge in A , und der **relativen Häufigkeit** $h_n(A) = \frac{k_n(A)}{n}$. Mit wachsender Versuchszahl n stabilisiert sich $h_n(A)$ und nähert sich der Wahrscheinlichkeit $P(A)$ an. Dies nennt man auch das **empirische Gesetz der großen Zahl**.

3.1.5 Laplace-Experiment

Wenn alle Versuchsausgänge $\omega \in \Omega$ gleichwahrscheinlich sind, heißt P **Gleichverteilung** auf Ω und (Ω, P) **Laplacescher Wahrscheinlichkeitsraum**. In diesem Fall ist

$$P(\omega) = \frac{1}{|\Omega|} \quad \text{für alle } \omega \in \Omega$$

und

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{für alle } A \in \mathcal{P}(\Omega)$$

die **Laplace-Wahrscheinlichkeit** von A .

3.2 Definition und Eigenschaften der bedingten Wahrscheinlichkeiten

Häufig steht, bevor das Ergebnis eines Zufallsexperiments bekannt ist, schon die Information zur Verfügung, daß das Ergebnis zu einer bestimmten Teilmenge des Stichprobenraumes gehört. Dies führt zur Betrachtung von bedingten Wahrscheinlichkeiten.

Man fragt also nach der Wahrscheinlichkeit von A unter der Bedingung, daß B schon eingetreten ist.

Die **bedingte Wahrscheinlichkeit** $P(A|B)$ von A bei gegebenem B für beliebige Ereignisse B mit $P(B) > 0$ ist

$$(6) \quad P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Oft ist in der Praxis die bedingte Wahrscheinlichkeit gegeben und die Schnittwahrscheinlichkeit gesucht. Dazu formt man die Gleichung für die bedingte Wahrscheinlichkeit um und erhält so den **Multiplikationssatz**

$$(7) \quad P(A \cap B) = P(B) \cdot P(A|B), \text{ wenn } P(B) > 0.$$

Mit dem Multiplikationssatz kann man dann den **Satz der totalen Wahrscheinlichkeit** formulieren:

$$(8) \quad P(A) = P(B) \cdot P(A|B) + P(\bar{B}) \cdot P(A|\bar{B}), \text{ wenn } P(B) > 0 \text{ und } P(\bar{B}) > 0.$$

Der Satz der totalen Wahrscheinlichkeit läßt sich anhand des Mengendiagramms erklären. Für Mengen gilt

$$(9) \quad A = (B \cap A) \cup (\bar{B} \cap A)$$

Jetzt kann man dies einsetzen und erhält

$$\begin{aligned} P(A) &= P[(B \cap A) \cup (\bar{B} \cap A)] && \text{mit (9)} \\ &= P(B \cap A) + P(\bar{B} \cap A) && \text{mit (3)} \\ &= P(B) \cdot P(A|B) + P(\bar{B}) \cdot P(A|\bar{B}) && \text{mit (7)} \end{aligned}$$

Der Satz der totalen Wahrscheinlichkeit ist verallgemeinerbar. Wenn disjunkte Teilmengen $B_i \subseteq \Omega$ vereinigt Ω ergeben und für alle Wahrscheinlichkeiten $P(B_i) > 0$, dann gilt

$$(10) \quad P(A) = \sum_i P(B_i) \cdot P(A|B_i).$$

Um den Zusammenhang zwischen $P(A|B)$ und $P(B|A)$ herzustellen, wird die **Formel von BAYES** benutzt:

$$(11) \quad P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}$$

3.3 Unabhängigkeit

Mit der Formel für bedingte Wahrscheinlichkeiten ist es plausibel, ein Ereignis A von einem Ereignis B unabhängig zu nennen, wenn $P(A) = P(A|B)$ gilt. Da aber $P(A|B)$ nur für Ereignisse B mit $P(B) > 0$ definiert ist, formuliert man allgemeiner. Zwei Ereignisse A und B heißen **stochastisch unabhängig**, wenn

$$(12) \quad P(A|B) = P(A) \text{ bzw. } P(B|A) = P(B).$$

Die Definition von Unabhängigkeit ist auf n Ereignisse erweiterbar, allerdings etwas komplizierter. Auf die Darstellung soll hier verzichtet werden, da diese Formel für unsere Zwecke nicht vonnöten ist.

4 Zufallsgrößen

4.1 Zufallsgrößen

Die Ergebnisse von Zufallsexperimenten können qualitative oder quantitative Größen sein. Um qualitative Ergebnisse weiterverwerten und mit ihnen rechnen zu können, ist es sinnvoll, sie durch quantitative Werte zu beschreiben. Man benötigt also eine Vorschrift, die jedem Ausgang eines Zufallsexperimentes eine reelle Zahl zuordnet.

Eine Zuordnung $X : \Omega \rightarrow \mathbb{R}$, die jedem Ergebnis eines Zufallsversuches $\omega \in \Omega$ eine reelle Zahl $x \in \mathbb{R}$ zuordnet, heißt **Zufallsgröße**. Zufallsgrößen werden häufig auch als **Zufallsvariablen** bezeichnet.

Von der Zufallsgröße selbst ist ihre konkrete Realisierung zu unterscheiden. Die **Zufallsgröße** ist die Funktion $X : \Omega \rightarrow \mathbb{R}$; sie wird mit großen lateinischen Buchstaben, meist vom Ende des Alphabets, bezeichnet. Die **konkrete Realisierung** der Zufallsgröße ist der Wert, den sie bei Ausführung des Zufallsexperimentes annehmen kann; für sie werden kleine lateinische Buchstaben verwendet.

Eine Zufallsgröße heißt **diskret**, wenn sie abzählbar viele Werte annehmen kann (Würfeln). Eine Zufallsgröße, die überabzählbar viele Werte annehmen kann, heißt **stetig** (Lebensdauer eines Glühwürmchens).

4.2 Wahrscheinlichkeitsverteilungen diskreter Zufallsgrößen

4.2.1 Wahrscheinlichkeitsfunktion

Sei X eine diskrete Zufallsgröße mit den Werten x_1, \dots, x_n . Dann lassen sich alle Ergebnisse des Zufallsversuches $\omega \in \Omega$ mit $X(\omega) = x_i$ zu einem Ereignis zusammenfassen, das man mit der Gleichung $X = x_i$ beschreiben kann.

Ordnet man jedem möglichen Wert x_i die Wahrscheinlichkeit $P(X=x_i)$ zu, mit der X diesen Wert annehmen kann, erhält man die **Wahrscheinlichkeitsfunktion** f_X . Es gilt

$$f_X(x_i) = P(X = x_i)$$

Aus der Definition der Wahrscheinlichkeit folgt, daß die Wahrscheinlichkeitsfunktion folgende **Eigenschaften** besitzt:

$$0 \leq f_X(x_i) \leq 1 \quad \text{für alle } x_i$$
$$\sum_i f_X(x_i) = 1$$

4.2.2 Verteilungsfunktion

Die **Verteilungsfunktion** F_X gibt die Wahrscheinlichkeit dafür an, daß die Zufallsgröße X Werte kleiner oder gleich dem Wert x annimmt. Man erhält sie aus der Wahrscheinlichkeitsfunktion durch Aufsummieren der Wahrscheinlichkeiten. Es gilt:

$$F_X(x_i) = P(X \leq x_i)$$
$$F_X(x_i) = \sum_{x \leq x_i} f_X(x)$$

Die Verteilungsfunktion F_X einer Zufallsvariablen hat folgende **Eigenschaften**:

F_X ist monoton steigend; F_X ist rechtsseitig stetig und es gilt:

$$\begin{aligned}\lim_{x \rightarrow -\infty} F_X &= 0 \\ \lim_{x \rightarrow +\infty} F_X &= 1\end{aligned}$$

4.3 Wahrscheinlichkeitsverteilungen stetiger Zufallsgrößen

4.3.1 Dichtefunktion

Bei stetigen Zufallsgrößen tritt an die Stelle der Wahrscheinlichkeitsfunktion die Dichtefunktion.

Die **Dichtefunktion** f_X einer stetigen Zufallsgröße X ist eine intervallweise stetige Funktion, für die gilt:

$$\begin{aligned}\int_{-\infty}^{+\infty} f_X(x) dx &= 1 \\ f_X(x) &\geq 0\end{aligned}$$

Im Gegensatz zur Wahrscheinlichkeitsfunktion einer diskreten Zufallsgröße gibt die Dichtefunktion einer stetigen Zufallsgröße im Punkt x nicht die Wahrscheinlichkeit dafür an, daß die Zufallsgröße den Wert x annimmt. Bestimmbar ist nur die Wahrscheinlichkeit, daß der Wert von X innerhalb eines bestimmten Intervalls $x_1 < x \leq x_2$ liegt. Sie entspricht der Fläche unterhalb der Dichtefunktion über diesem Intervall. Es gilt:

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx$$

Wie daraus folgt, hat ein einzelner Wert einer stetigen Zufallsvariablen stets die Wahrscheinlichkeit Null: Er bildet nämlich ein "Intervall der Länge Null", das mit der Dichtefunktion eine Fläche der Größe Null einschließt. Daraus folgt jedoch nicht, daß dieses Ereignis unmöglich ist; sonst könnte die Zufallsgröße überhaupt keinen Wert annehmen.

4.3.2 Verteilungsfunktion

Wie bei diskreten Zufallsgrößen gibt die **Verteilungsfunktion** F_X einer stetigen Zufallsgröße X die Wahrscheinlichkeit dafür an, daß die Zufallsgröße Werte kleiner oder gleich x annimmt. Während man die Verteilungsfunktion einer diskreten Zufallsgröße jedoch aus der Wahrscheinlichkeitsfunktion durch Summieren berechnen kann, erhält man die Verteilungsfunktion der stetigen Zufallsgröße durch Integrieren der Dichtefunktion in einem bestimmten Intervall. Es gilt:

$$\begin{aligned}F_X(x) &= P(X \leq x) \\ F_X(x) &= \int_{-\infty}^x f_X(\xi) d\xi\end{aligned}$$

Die Verteilungsfunktion ist stetig; ihre Eigenschaften sind dieselben wie die der Verteilungsfunktion einer diskreten Zufallsvariablen.

Wie man leicht sieht, erhält man zu einer gegebenen Verteilungsfunktion F_X die zugehörige Dichtefunktion f_X durch Differenzieren von F_X :

$$f_X(x) = \frac{dF_X(x)}{dx}$$

4.4 Parameter von Wahrscheinlichkeitsverteilungen

4.4.1 Erwartungswert

Oft reicht die Kenntnis der Wahrscheinlichkeitsverteilung für die Einschätzung einer Situation nicht aus. Man benötigt einen zusätzlichen Wert, der die Wahrscheinlichkeitsverteilung der Zufallsgröße charakterisiert und einen Vergleich mit anderen Wahrscheinlichkeitsverteilungen ermöglicht.

Diese Funktion erfüllt der **Erwartungswert $E(X)$** , auch μ genannt. Der Erwartungswert ist der Durchschnitt aller von X angenommenen Werte unter Berücksichtigung der Häufigkeit ihres Auftretens. Es gilt:

$$E(X) = \sum_i x_i \cdot f_X(x_i) \quad \text{für } X \text{ diskret}$$

$$E(X) = \int_{-\infty}^{+\infty} f_X(x) dx \quad \text{für } X \text{ stetig}$$

4.4.2 Varianz

Die Werte, die eine Zufallsgröße annimmt, streuen im allgemeinen mehr oder weniger um den Erwartungswert $E(X)$. Ein Maß für die Streuung der Werte einer Zufallsvariablen X um den Erwartungswert $E(X)$ ist die **Varianz $VAR(X)$** . Es gilt:

$$VAR(X) = \sum_i (x_i - E(X))^2 \cdot f_X(x_i) \quad \text{für } X \text{ diskret}$$

$$VAR(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f_X(x) dx \quad \text{für } X \text{ stetig}$$

Durch die Benutzung der quadratischen Abweichungen werden größere Abstände zum Erwartungswert stärker berücksichtigt.

Die Quadratwurzel aus der Varianz heißt **Standardabweichung**, genannt $\sigma(X)$. Während die Varianz als Dimension das Quadrat der Dimension des Merkmals hat, für das sie berechnet wurde, hat die Standardabweichung dieselbe Dimension wie das Merkmal. Es gilt:

$$\sigma(X) = \sqrt{VAR(X)}$$

4.5 Wahrscheinlichkeitsverteilungen mehrdimensionaler Zufallsgrößen

4.5.1 Mehrdimensionale Zufallsgrößen

Bis jetzt wurden nur eindimensionale Zufallsgrößen betrachtet. Manchmal interessiert man sich jedoch für die gemeinsame Verteilung mehrerer Zufallsgrößen. Eine **mehrdimensionale Zufallsgröße X** mit $X : \Omega \rightarrow \mathbb{R}^n$ ist die Zusammenfassung von n Zufallsgrößen zu einer n -dimensionalen Zufallsgröße.

4.5.2 Wahrscheinlichkeits- und Dichtefunktion zweidimensionaler Zufallsgrößen

Zum besseren Verständnis werden im folgenden nur zweidimensionale Zufallsgrößen betrachtet. Sei Z eine **diskrete zweidimensionale Zufallsgröße**, die aus den eindimensionalen Zufallsgrößen X und Y besteht. X habe die Werte x_1, \dots, x_n , Y die Werte y_1, \dots, y_m . Dann ist die **Wahrscheinlichkeitsfunktion** f_Z wie folgt definiert:

$$f_Z(z_{ij}) = f_{XY}(x_i, y_j) = P(X = x_i; Y = y_j) \text{ für } i = 1, \dots, n; j = 1, \dots, m$$

Die Wahrscheinlichkeitsfunktion hat folgende **Eigenschaften**:

$$\begin{aligned} f_Z(z_{ij}) &= f_{XY}(x_i, y_j) \geq 0 \\ \sum_{ij} f_Z(z_{ij}) &= \sum_i \sum_j f_{XY}(x_i, y_j) = 1 \end{aligned}$$

Sei nun Z eine **stetige zweidimensionale Zufallsvariable**, bestehend aus den eindimensionalen stetigen Zufallsvariablen X und Y . Dann ist die **Dichtefunktion** f_{XY} eine integrierbare Funktion mit

$$\begin{aligned} f_{XY}(x_i, y_j) &\geq 0 \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy &= 1. \end{aligned}$$

Während im eindimensionalen Fall die Wahrscheinlichkeit dafür, daß eine Zufallsvariable einen Wert in einem bestimmten Intervall annimmt, durch die Fläche zwischen der zugehörigen Dichtefunktion und der x-Achse über dem Intervall dargestellt wurde, kann man im zweidimensionalen Fall nur die **Wahrscheinlichkeit** dafür angeben, daß die Ausprägungen von X und Y in einem Rechteck liegen. Diese Wahrscheinlichkeit entspricht nicht mehr einer Fläche, sondern dem Raum zwischen der Dichtefunktion und der x-y-Ebene über dem Rechteck. Es gilt also:

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dx dy$$

4.5.3 Parameter zweidimensionaler Verteilungen

Bei zweidimensionalen Zufallsvariablen werden zwei Ereignisse oder Merkmale gleichzeitig erfaßt. Häufig möchte man dann wissen, ob zwischen diesen Merkmalen ein Zusammenhang besteht. Sei nun wieder Z eine aus den eindimensionalen Zufallsgrößen X und Y bestehende Zufallsgröße.

Ein unnormiertes Maß für die gemeinsame Variabilität von X und Y ist die **Kovarianz** $\text{COV}(\mathbf{X}, \mathbf{Y})$. Es gilt:

$$\text{COV}(X, Y) = \sigma_{XY} = E[(X-E(X)) \cdot (Y-E(Y))]$$

Gibt es keine Abhängigkeit bzw. keinen Zusammenhang zwischen beiden Merkmalen, dann ist $\text{COV}(X,Y) = 0$. Die Kovarianz ist umso größer, je häufiger zu einer Beobachtung x_i und y_i die Differenzen $x_i - E(X)$ und $y_i - E(Y)$ dasselbe Vorzeichen haben, wenn also X bei den Ereignissen größere Werte annimmt, bei denen auch Y die größeren Werte annimmt. Negative Kovarianz deutet auf eine umgekehrte Tendenz hin.

Der **Korrelationskoeffizient** ρ_{XY} ist ein normiertes Maß für die Ausprägtheit eines linearen Zusammenhanges zwischen X und Y. Es gilt:

$$\rho_{XY} = \frac{\text{COV}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Die Werte von ρ_{XY} liegen zwischen -1 und +1. Liegen alle Werte (x_i, y_i) auf einer Geraden, dann gilt $|\rho_{XY}| = 1$. Bei einer steigenden Geraden gilt $\rho_{XY} = 1$, bei einer fallenden Gerade $\rho_{XY} = -1$. Der Wert von ρ_{XY} ist von nahe bei Null, wenn keine aufwärts oder abwärts gerichtete Tendenz erkennbar ist. Sind zwei die Merkmale X und Y unabhängig, gilt $\rho_{XY} = 0$; die Umkehrung der Aussage gilt jedoch nicht.

5 Induktive Statistik

5.1 Grundbegriffe

Die **deskriptive Statistik** befaßt sich mit Datenmengen, die im Rahmen einer Untersuchung vollständig erfaßt und analysiert werden sollen. Alle mit Hilfe der deskriptiven Statistik gewonnenen Aussagen beziehen sich nur auf die untersuchte Datenmenge und sind im allgemeinen nicht verallgemeinerbar.

Häufig ist aber eine vollständige Untersuchung der Menge, über die man eine Aussage machen will, nicht möglich, zu teuer oder nicht zweckmäßig. Untersucht wird daher nur ein Teil der gesamten Masse. Die **induktive Statistik** befaßt sich mit der Frage, wie man mit Hilfe der Wahrscheinlichkeitsrechnung aus den Ergebnissen der statistischen Untersuchung einer Teilmasse auf die übergeordnete Gesamtmasse schließen kann.

Als **Grundgesamtheit** bezeichnet man die statistische Masse, über die man eine Aussage machen möchte.

Die **Stichprobe** ist der Teil einer statistischen Masse, der analysiert wird, um Aufschluß über die Grundgesamtheit zu erhalten.

Entscheidende Voraussetzung für die Anwendung der Wahrscheinlichkeitsrechnung ist die **zufällige Auswahl der Stichprobenelemente**. In diesem Fall sind die an den Stichprobenelementen feststellbaren Werte des untersuchten Merkmals konkrete Realisationen von Zufallsgrößen. Über Zufallsgrößen, ihre Wahrscheinlichkeitsverteilungen und ihre Parameter lassen sich aber mit Hilfe der Wahrscheinlichkeitsrechnung Aussagen machen.

Mathematisch schlägt sich dies im Begriff der **Stichprobenfunktion** nieder. Sei X das Merkmal einer Grundgesamtheit, dessen Eigenschaften untersucht werden sollen. Eine Stichprobe vom Umfang n liefert dann die Stichprobenwerte x_1, \dots, x_n . Werden die Elemente zufällig entnommen, kann das Auftreten einer möglichen Merkmalsausprägung an einem Element durch die Zufallsgröße X beschrieben werden. Die Merkmalsausprägung jedes der n Stichprobenelemente i läßt sich daher

durch die Zufallsgröße X_i beschreiben. Eine Funktion h , die den aus der Stichprobe gewonnenen Zufallsgrößen X_1, \dots, X_n einen Wert $H(X_1, \dots, X_n)$ zuordnet, heißt Stichprobenfunktion. Als Funktion von Zufallsgrößen ist auch H wieder eine Zufallsgröße, so daß die Aussagen zu Wahrscheinlichkeitsverteilungen anwendbar bleiben.

5.2 Schätzverfahren

Schätzverfahren haben das Ziel, mittels einer Stichprobe unbekannte Parameter der Grundgesamtheit wie zum Beispiel den Erwartungswert oder die unbekannte Verteilung der Grundgesamtheit zu schätzen. Im folgenden werden nur Parameterschätzungen behandelt. Man unterscheidet **Punkt-** und **Intervallschätzungen**.

5.2.1 Punktschätzung

Bei einer **Punktschätzung** wird für den zu schätzenden Parameter ein einzelner Wert bestimmt.

Sei q der unbekannte Parameter, der geschätzt werden soll. Dann wird für eine Punktschätzung der Grundgesamtheit eine Zufallsstichprobe x_1, \dots, x_n entnommen, deren Elemente die Realisation der Zufallsgrößen X_1, \dots, X_n darstellen. Aus x_1, \dots, x_n wird dann der Wert einer für diesen Parameter passenden Stichprobenfunktion \hat{Q} berechnet.

Eine **Punktschätzung** ist also der sich für bestimmte Stichprobenwerte x_1, \dots, x_n ergebende Wert \hat{q} der Stichprobenfunktion \hat{Q} :

$$\hat{q} = Q(x_1, \dots, x_n)$$

Die für die einzelnen Parameter verwendbaren Schätzfunktionen und ihre Eigenschaften sind zum Beispiel in [Bron87] beschrieben. Eine häufig verwendete Schätzfunktion für den Erwartungswert μ der Grundgesamtheit ist etwa der Stichprobenmittelwert \bar{X} mit:

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

Im Normalfall wird der durch die Punktschätzung erhaltene Wert nicht dem wirklichen Wert des zu schätzenden Parameters entsprechen. Führt man jedoch über diesem Ansatz sehr viele Schätzungen aus, gleichen sich die möglichen Fehler ungefähr aus.

5.2.2 Intervallschätzung

Eine **Intervallschätzung** liefert ein Intervall, in dem der zu schätzende Parameter mit einer bestimmten Wahrscheinlichkeit erwartet wird.

Dabei wird aus den zur Stichprobe gehörenden Zufallsgrößen X_1, \dots, X_n ein **Zufallsintervall** $[Q_1; Q_2]$ berechnet, das den unbekannt Parameter q mit einer vorgegebenen Wahrscheinlichkeit überdeckt. Die Grenzen des Intervalls, Q_1 und Q_2 , sind Zufallsgrößen, die aus X_1, \dots, X_n berechnet werden.

Im Regelfall wird durch eine Intervallschätzung ein **Konfidenzintervall** berechnet. Ein Konfidenzintervall ist ein Zufallsintervall $[Q_1; Q_2]$, das den unbekanntem Parameter q mit der Wahrscheinlichkeit $1 - \alpha$ überdeckt. Es gilt also:

$$P(Q_1 \leq q \leq Q_2) = 1 - \alpha$$

α heißt auch **Irrtumswahrscheinlichkeit**.

Aus einer konkreten Stichprobe kann man also eine konkrete Realisation des Konfidenzintervalles berechnen. Der wirkliche Wert des unbekanntem Parameters q liegt dann entweder innerhalb oder außerhalb des Intervalls (oder auf den Grenzen). Zieht man sehr viele Stichproben, so kann man erwarten, daß bei einem $(1 - \alpha) \cdot 100\%$ -Konfidenzintervall $(1 - \alpha) \cdot 100\%$ der Intervalle den Parameter q überdecken.

Die konkreten Berechnungsvorschriften für ein Konfidenzintervall sind von der Wahrscheinlichkeitsverteilung der Grundgesamtheit abhängig; sie sind zum Beispiel [Bron87] zu entnehmen.

Literaturhinweise

Bei unserer Arbeit fiel uns auf, daß uns der Einstieg am besten mit Schulbüchern gelang, die für die gymnasiale Oberstufe konzipiert waren. Besonders können wir [Big/Kö92] jedem ans Herz legen, der verschüttete Kenntnisse wieder auffrischen möchte und dies eher am praktischen Verständnis orientiert tun will. Viele, interessante Beispielaufgaben und die didaktische Aufbereitung des Stoffes lassen die Motivation bis zuletzt nicht verschwinden.

Wer lieber den axiomatischen Zugang wählt und Wert darauf legt, daß sich alles im wohldefinierten Rahmen abspielt, dem sei [Kren91] empfohlen. Dieses Buch ist sowohl zum Selbststudium als auch als Referenz geeignet.

Werden die Grundbegriffe schon beherrscht und läßt alles oben gesagte nur Langeweile aufkommen, ist als Nachschlagewerk die Bibel der Anwender der Mathematik [Bron87] geeignet.

Literatur

[Big/Kö92] Bigalke, Anton und Köhler, Norbert: *Mathematik 13.2, Stochastik*. Cornelsen Berlin 1992.

[Bron87] Bronstein, I.N. und Semendjajew, K.A.: *Taschenbuch der Mathematik*. BSB B. G. Teubner Verlagsgesellschaft, Leipzig 1987²³.

[Kren91] Kregel, Ulrich: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Braunschweig 1991.