

# Stichworte zur Statistik I

SoSe 2005

Ulrich Pötter

1. Wir setzen eine endliche Menge von Objekten als gegeben voraus, die dann mithilfe statistischer Begriffe beschrieben werden soll. Wir nehmen an, dass diese Menge  $n$  Elemente hat, und schreiben dafür:

$$\Omega = \{\omega_1, \dots, \omega_n\}$$

## Logische Symbole

2. Wir verwenden die folgenden logischen Symbole:  $\neg$  (nicht),  $\wedge$  (und),  $\vee$  (oder),  $\Rightarrow$  (wenn, dann),  $\Leftrightarrow$  (genau dann, wenn),  $\forall$  (für alle),  $\exists$  (existiert).

## Mengen

3. Zur Definition der statistischen Grundbegriffe verwenden wir einige elementare Begriffe und Schreibweisen der Mengenlehre.

Wir betrachten nur endliche Mengen. Sie werden durch Großbuchstaben bezeichnet, zum Beispiel:  $A, B, \dots, \Omega$ ; ihre Elemente werden i.d.R. durch Kleinbuchstaben bezeichnet. Die Schreibweise  $a \in A$  soll bedeuten, dass  $a$  ein Element der Menge  $A$  ist.  $a \notin A$  bedeutet, dass  $a$  nicht Element von  $A$  ist.

Ist  $A$  eine Menge, bedeutet  $|A|$  die Anzahl ihrer Elemente.

Mit dem Symbol  $\emptyset$  wird die *leere Menge* bezeichnet, und es wird festgesetzt:  $|\emptyset| = 0$ .

Die Schreibweise  $A \subseteq B$  bedeutet, dass  $A$  eine *Teilmenge* der Menge  $B$  ist, d.h. jedes Element der Menge  $A$  ist auch ein Element der Menge  $B$ :  $a \in A \Rightarrow a \in B$ . Insbesondere ist die leere Menge Teilmenge jeder Menge. Ist  $A \subseteq B$ , so heißt  $B$  *Obermenge* von  $A$ .

Zwei Mengen werden als gleich angesehen,  $A = B$ , wenn  $A \subseteq B$  und gleichzeitig  $B \subseteq A$  gilt, wenn also beide Mengen die gleichen Elemente enthalten.

Sind  $A$  und  $B$  zwei Mengen, bedeutet  $A \cup B := \{x \mid x \in A \vee x \in B\}$  die *Vereinigungsmenge* von  $A$  und  $B$ . Sie besteht aus allen Elementen, die in  $A$  oder in  $B$  enthalten sind. Die Operation  $\cup$  ist kommutativ:  $A \cup B = B \cup A$  und assoziativ:  $A \cup (B \cup C) = (A \cup B) \cup C$ .

Sind  $A$  und  $B$  zwei Mengen, bedeutet  $A \cap B := \{x \mid x \in A \wedge x \in B\}$  die *Schnittmenge* von  $A$  und  $B$ . Sie besteht aus allen Elementen, die sowohl in  $A$  als auch in  $B$  enthalten sind. Die Operation  $\cap$  ist kommutativ:  $A \cap B = B \cap A$  und assoziativ:  $A \cap (B \cap C) = (A \cap B) \cap C$ . Zwei Mengen mit  $A \cap B = \emptyset$  heißen *disjunkt*.

Vereinigungs- und Durchschnittsbildung sind distributiv:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

Ist  $A$  eine Teilmenge von  $B$ , bedeutet  $B \setminus A := \{x \mid x \in B \wedge x \notin A\}$  das *Komplement* von  $A$  in  $B$ .  $B \setminus A$  besteht aus allen Elementen, die in  $B$ , aber nicht in  $A$  enthalten sind. Ist eine Obermenge  $B$  vorgegeben, schreiben wir kurz  $A^c$  für das Komplement von  $A$  in  $B$ .

Ist  $A$  eine Menge, bedeutet  $\mathcal{P}(A) := \{B \mid B \subseteq A\}$  die *Potenzmenge* von  $A$ , die Menge aller Teilmengen von  $A$ . Insbesondere ist  $\emptyset$  ein Element der Potenzmenge jeder Menge und es gilt  $|\mathcal{P}(A)| = 2^{|A|}$ .

Sei  $\Omega$  eine Menge und  $P = \{A_1, \dots, A_n\}$  eine Menge von Teilmengen von  $\Omega$ .  $P$  wird eine *Partition* von  $\Omega$  genannt, wenn gilt:

$$\begin{aligned} A_1 \cup \dots \cup A_n &= \Omega \\ \forall A, A' \in P, A \neq A' : A \cap A' &= \emptyset \end{aligned}$$

Sind  $A$  und  $B$  zwei Mengen, wird mit

$$A \times B := \{(a, b) \mid a \in A, b \in B\}$$

das *kartesische Produkt* von  $A$  und  $B$  bezeichnet. Eine analoge Begriffsbildung erfolgt für drei und mehr Mengen. Es ist  $\emptyset \times A = A \times \emptyset = \emptyset$ . Außerdem gilt

$$\begin{aligned} (A \cup B) \times C &= (A \times C) \cup (B \times C) \\ (A \cap B) \times C &= (A \times C) \cap (B \times C) \end{aligned}$$

## Funktionen

4. Eine *Funktion* oder *Abbildung*  $f: A \rightarrow B$  mit *Definitionsbereich*  $A$  und *Wertebereich*  $B$  ordnet jedem Element der Menge  $A$  genau ein Element der Menge  $B$  zu.  $f$  ist also Teilmenge des kartesischen Produktes  $A \times B$  und  $f$  erfüllt die beiden Bedingungen:

$$\forall a \in A \exists b \in B : (a, b) \in f$$

und

$$((a, b) \in f \wedge (a, c) \in f) \Rightarrow b = c$$

Das Element  $b = f(a)$  heißt *Funktionswert* der Funktion  $f$  zu dem *Argument*  $a$ . Die erste Bedingung besagt, dass es zu *allen* Elementen von  $A$  einen Funktionswert gibt, die zweite, dass der Funktionswert eines Arguments *eindeutig* ist.

5. Jeder Funktion  $f: A \rightarrow B$  lassen sich zwei Mengenfunktionen zuordnen. Zum einen

$$\begin{aligned} f: \mathcal{P}(A) &\longrightarrow \mathcal{P}(B) \\ f(C) &:= \{b \in B \mid \exists c \in C : f(c) = b\} = \{f(c) \mid c \in C\} \quad \forall C \in \mathcal{P}(A) \end{aligned}$$

Zum anderen die *Umkehrfunktion*

$$\begin{aligned} f^{-1}: \mathcal{P}(B) &\longrightarrow \mathcal{P}(A) \\ f^{-1}(D) &:= \{a \in A \mid f(a) \in D\} \quad \forall D \in \mathcal{P}(B) \end{aligned}$$

Ist  $C \subseteq A$ , so heißt  $f(C)$  die *Bildmenge* von  $C$ . Ist  $D \subseteq B$ , so heißt  $f^{-1}(D)$  die *Urbildmenge* von  $D$ . Wir unterscheiden die Funktion  $f$  und die zugehörige Mengenfunktion  $f$  durch die Form des Arguments: Ist das Argument eine Menge, so ist die Mengenfunktion gemeint. Diese Unterscheidung reicht im folgenden, weil wir keine Funktionen betrachten werden, deren Definitionsbereiche Mengen als Elemente enthalten. Ebenso unterscheiden

wir zwischen der Umkehrabbildung  $f^{-1}$  und der inversen Funktion  $f^{-1}: B \rightarrow A$ : Es ist die Mengenfunktion  $f^{-1}: \mathcal{P}(B) \rightarrow \mathcal{P}(A)$  gemeint, wenn das Argument von  $f^{-1}$  eine Menge ist.

Für alle  $C_1, C_2 \subseteq A$  und für alle  $D_1, D_2 \subseteq B$  gilt:

$$\begin{aligned} C_1 &\subseteq f^{-1}(f(C_1)) \\ D_1 &\supseteq f(f^{-1}(D_1)) \\ f(C_1 \cup C_2) &= f(C_1) \cup f(C_2) \\ f^{-1}(D_1 \cup D_2) &= f^{-1}(D_1) \cup f^{-1}(D_2) \\ f^{-1}(D_1 \cap D_2) &= f^{-1}(D_1) \cap f^{-1}(D_2) \end{aligned}$$

Allerdings ist i.d.R.  $f(C_1 \cap C_2) \neq f(C_1) \cap f(C_2)$ .

**6.** Sind  $f: A \rightarrow B$  und  $g: B \rightarrow C$  zwei Funktionen, so dass der Wertebereich der ersten gleich dem Definitionsbereich der zweiten Funktion ist, so kann man eine Funktion  $h: A \rightarrow C$  durch  $h(a) := g(f(a))$  definieren. Wir nennen  $h$  die *Verkettung* von  $f$  und  $g$  und schreiben  $h = g \circ f$ . Für die Umkehrabbildung  $(g \circ f)^{-1}: \mathcal{P}(C) \rightarrow \mathcal{P}(A)$  gilt  $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ .

**7.** Wenn über zwei oder mehr Funktionen mit gleichem Definitionsbereich gesprochen werden soll, etwa über  $f: A \rightarrow B$  und  $g: A \rightarrow C$ , dann ist es nützlich,  $f$  und  $g$  als *Komponenten* der Funktion  $(f, g): A \rightarrow B \times C$  aufzufassen. Dabei ist  $(f, g)(a) := (f(a), g(a)) \in B \times C$ . Die zugehörige Mengenfunktion  $(f, g): \mathcal{P}(A) \rightarrow \mathcal{P}(B \times C)$  ist durch  $(f, g)(D) := \{(f(a), g(a)) \mid a \in D\} \in \mathcal{P}(B \times C)$  für alle  $D \in \mathcal{P}(A)$  definiert. Es ist  $(f, g)(D) \subseteq f(D) \times g(D)$ . Die Umkehrfunktion  $(f, g)^{-1}: \mathcal{P}(B \times C) \rightarrow \mathcal{P}(A)$  ist durch  $(f, g)^{-1}(E) := \{a \in A \mid (f, g)(a) \in E\}$  für alle  $E \in \mathcal{P}(B \times C)$  definiert. Ist insbesondere  $E = F \times G$  mit  $F \subseteq B$  und  $G \subseteq C$ , dann ist  $(f, g)^{-1}(E) = (f, g)^{-1}(F \times G) = f^{-1}(F) \cap g^{-1}(G)$ .

## Statistische Variablen und Verteilungen

**8.** Eine *statistische Variable* ist eine Abbildung

$$X: \Omega \longrightarrow \tilde{\mathcal{X}}$$

die jedem Objekt aus  $\Omega$  einen Wert in einem *Merkmalsraum*  $\tilde{\mathcal{X}}$  zuordnet. Dem Objekt  $\omega \in \Omega$  wird also der Merkmalswert  $X(\omega) \in \tilde{\mathcal{X}}$  zugeordnet. Der Merkmalsraum  $\tilde{\mathcal{X}}$  kann beliebig gewählt werden. Die Elemente des Merkmalsraums sind strenggenommen keine Zahlen, sondern bezeichnen Merkmalswerte. Wir setzen jedoch voraus, dass sie durch Zahlen *repräsentiert* werden. Zur Repräsentation werden Teilmengen der reellen Zahlen verwendet. Da wir voraussetzen, dass  $\Omega$  als Gesamtheit der Dinge oder Situationen, über die gesprochen werden soll, immer nur endlich viele Elemente enthält, ist aber auch  $X(\Omega)$  immer endlich.

**9.** Zu jeder Variablen  $X: \Omega \rightarrow \tilde{\mathcal{X}}$  gehört eine *Häufigkeitsverteilung* oder *statistische Verteilung*. Sie ist definiert als eine Funktion

$$P[X]: \mathcal{P}(\tilde{\mathcal{X}}) \longrightarrow [0, 1]$$

die jeder Teilmenge von  $\tilde{\mathcal{X}}$  eine rationale Zahl zwischen 0 und 1 zuordnet, und zwar nach

folgender Regel:

$$P[X](A) := \frac{1}{|\Omega|} |\{\omega \mid X(\omega) \in A\}| = \frac{1}{|\Omega|} |X^{-1}(A)| \quad \text{für } A \in \mathcal{P}(\tilde{\mathcal{X}})$$

$P[X](A)$  ist also die *relative Häufigkeit* der  $\omega \in \Omega$ , für die ein in  $A$  enthaltener Merkmalswert  $X(\omega)$  auftritt.

**10.** Die Häufigkeitsverteilung ist *additiv*, das heißt: Wenn  $A, B \in \mathcal{P}(\tilde{\mathcal{X}})$  und  $A \cap B = \emptyset$ , dann gilt:

$$P[X](A \cup B) = P[X](A) + P[X](B)$$

Außerdem ist  $0 \leq P[X](A) \leq 1$  für alle  $A \subseteq \tilde{\mathcal{X}}$ ,  $P[X](\emptyset) = 0$  und  $P[X](\tilde{\mathcal{X}}) = 1$ . Ist  $A$  endlich, dann erhält man  $P[X](A) = \sum_{\tilde{x} \in A} P[X](\{\tilde{x}\})$ .

## Repräsentationen statistischer Verteilungen

**11.** Ist auch  $\tilde{\mathcal{X}} = \{\tilde{x}_1, \dots, \tilde{x}_m\}$  endlich, so dass alle Teilmengen von  $\tilde{\mathcal{X}}$  endlich sind, dann lassen sich die Werte der Häufigkeitsverteilung für alle Teilmengen  $A \subseteq \tilde{\mathcal{X}}$  in dieser Form berechnen. *Häufigkeitstabellen* repräsentieren die statistische Verteilung  $P[X]$  als Tupel  $(\tilde{x}_i, P[X](\{\tilde{x}_i\}))$ ,  $i = 1, \dots, m$ , etwa in der folgenden Form:

$$\frac{\tilde{x}_1}{P[X](\{\tilde{x}_1\})} \quad \frac{\tilde{x}_2}{P[X](\{\tilde{x}_2\})} \quad \dots \quad \frac{\tilde{x}_m}{P[X](\{\tilde{x}_m\})}$$

**12.** Ist der Merkmalsraum  $\tilde{\mathcal{X}}$  eine Menge von Zahlen, für die die  $<$ -Relation zwischen den Zahlen als Ordnung zwischen den Merkmalsausprägungen interpretiert werden kann, dann kann man eine Funktion

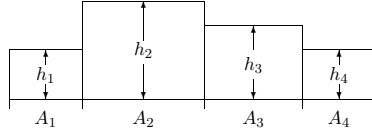
$$F[X]: \tilde{\mathcal{X}} \longrightarrow [0, 1]$$

betrachten, die jedem Merkmalswert in  $\tilde{\mathcal{X}}$  eine rationale Zahl im Intervall  $[0, 1]$  zuordnet, und zwar nach folgender Regel:

$$F[X](\tilde{x}) := P[X](\{\tilde{x}' \in \tilde{\mathcal{X}} \mid \tilde{x}' \leq \tilde{x}\})$$

für jedes  $\tilde{x} \in \tilde{\mathcal{X}}$ . Die Funktion  $F[X]$  ordnet also jedem Merkmalswert  $\tilde{x} \in \tilde{\mathcal{X}}$  die relative Häufigkeit zu, mit der in der Gesamtheit  $\Omega$  ein Merkmalswert auftritt, der kleiner oder gleich  $\tilde{x}$  ist. Die Funktion  $F[X]$  wird *Verteilungsfunktion* der Variablen  $X$  genannt. Die Verteilungsfunktion einer Variablen  $X$  kann graphisch dargestellt werden. Man verwendet dazu ein Koordinatensystem, dessen  $x$ -Achse den Merkmalsraum  $\tilde{\mathcal{X}}$  und dessen  $y$ -Achse das Zahlenintervall  $[0, 1]$  repräsentiert. Die Verteilungsfunktion kann dann als eine Treppenfunktion in diesem Koordinatensystem dargestellt werden.

**13.** Zur graphischen Darstellung der Verteilung einer Variablen werden oft *Histogramme* verwendet. Ganz allgemein definiert, handelt es sich um eine graphische Darstellung von Häufigkeitstabellen. Abbildung 1 gibt eine Illustration. In dieser Illustration wird für den Wertebereich der Variablen eine Partition in vier Klassen gebildet. Dann wird zu jeder Klasse  $A_k$  die Höhe  $h_k$  der Histogrammsäule so berechnet, dass der Flächeninhalt der Säule proportional zur Häufigkeit  $P[X](A_k)$  ist.



**Abb. 1** Schematische Illustration eines Histogramms auf der Grundlage einer Partition in vier Klassen.

Wenn der Wertebereich der Variablen  $X$  durch eine beschränkte Teilmenge der reellen Zahlen repräsentiert wird, dann bildet man eine Partition dieses Wertebereichs in eine Folge von Sub-Intervallen  $A_1, \dots, A_K$ . Wir unterscheiden vier Arten von Intervallen. (Hier sind  $a$  und  $b$  reelle Zahlen mit der Eigenschaft  $a < b$ .)

- $[a, b] = \{x \mid a \leq x \leq b\}$     einschließlich der beiden Eckpunkte
- $]a, b] = \{x \mid a < x \leq b\}$     links offen, rechts abgeschlossen
- $[a, b[ = \{x \mid a \leq x < b\}$     links abgeschlossen, rechts offen
- $]a, b[ = \{x \mid a < x < b\}$     an beiden Seiten offen

Um eine Partition des Wertebereichs in Sub-Intervalle zu erhalten, die sich nicht überschneiden, nimmt man entweder Intervalle der Form  $[a, b[$ , oder Intervalle der Form  $]a, b]$ . Wir verwenden hier Intervalle der Form  $]a, b]$ . Ist  $\mathcal{X} \subseteq ]a, b]$  und  $a = a_0 < a_1 < \dots < a_K = b$ , dann definieren wir eine Partition mit den Elementen

$$A_k := ]a_{k-1}, a_k] \quad (\text{für } k = 1, \dots, K - 1)$$

Für jedes Sub-Intervall  $A_k$  kann dann die Häufigkeit  $P[X](A_k)$  berechnet werden, mit der Beobachtungen in dieses Sub-Intervall fallen. Schließlich kann ein Histogramm gezeichnet werden. Auf der  $X$ -Achse werden die Sub-Intervalle eingetragen, und die Höhen  $h_k$  der Histogrammsäulen werden so berechnet, dass

$$h_k \cdot (a_k - a_{k-1}) = P[X](A_k)$$

gilt. Der Flächeninhalt der Histogrammsäulen zeigt dann die relativen Häufigkeiten, mit denen die entsprechenden Sub-Intervalle besetzt sind. Während man aus Häufigkeitstabellen und Verteilungsfunktionen die statistische Verteilung einer Variablen rekonstruieren kann, liefert ein Histogramm nur eine Näherung für die statistische Verteilung.

### Charakterisierungen statistischer Verteilungen

**14.** Zur Charakterisierung einer Verteilungsfunktion können *Quantile* verwendet werden. Die Idee ist: Sei  $p \in (0, 1]$  eine beliebige Zahl zwischen 0 und 1. Gibt es dann eine Merkmalsausprägung  $\tilde{x}$  mit der Eigenschaft  $F[X](\tilde{x}) = p$ , dann heißt  $\tilde{x}$  das *p-Quantil* der Verteilung von  $X$ . Es wird mit  $Q_p(X)$  bezeichnet. Allerdings existiert für einen beliebigen Wert  $p$  meist kein (eindeutiges)  $\tilde{x} \in \mathcal{X}$  mit der Eigenschaft  $F[X](\tilde{x}) = p$ . Um dennoch eine Charakterisierung von Verteilungen durch eine einzige Zahl zu erreichen, die sich an

dieser Idee orientiert, gibt es viele verschiedene Vorschläge. Die wohl einfachste ist, als *p-Quantil* den kleinsten Wert der  $\tilde{x}$  zu wählen, so dass  $F[X](\tilde{x})$  größer oder gleich  $p$  ist:

$$Q_p(X) := \min\{\tilde{x} \mid F[X](\tilde{x}) \geq p\}$$

Das 0.5-Quantil wird auch *Median* genannt. Für den Median hat sich eine andere Konvention durchgesetzt:

$$Q_{0.5}(X) := \begin{cases} x_{((n+1)/2)} & \text{falls } n \text{ ungerade ist} \\ (x_{(n/2)} + x_{(n/2+1)})/2 & \text{falls } n \text{ gerade ist} \end{cases}$$

Dabei ist  $n = |\Omega|$  und  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  sind die der Größe nach geordneten Werte  $X(\omega_1), \dots, X(\omega_n)$ .

Zur praktischen Berechnung geht man in zwei Schritten vor:

a) Gegeben seien die Werte einer Variablen  $X$  für die  $n$  Objekte in  $\Omega$ . Diese Merkmalswerte werden zunächst in aufsteigender Reihenfolge geordnet. Man erhält eine Liste  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Man beachte, dass es sich hier um die tatsächlich realisierten, nicht um die im Merkmalsraum möglichen Merkmalswerte handelt. Wir müssen das ' $\leq$ '-Symbol verwenden, da es durchaus möglich ist, dass einige Merkmalswerte mehrfach auftreten.

b) In einem zweiten Schritt kann dann der Median berechnet werden: Falls  $n$  ungerade ist, wählt man den  $(n+1)/2$ -ten Wert  $x_{((n+1)/2)}$  aus der Liste. Falls  $n$  gerade ist, dann nimmt man den Mittelwert der beiden Werte  $x_{(n/2)}$  und  $x_{(n/2+1)}$ , also  $(x_{(n/2)} + x_{(n/2+1)})/2$ .

**15.** Der *Mittelwert* oder das *arithmetische Mittel* ist durch

$$M(X) := \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega) = \sum_{\tilde{x} \in \mathcal{X}} \tilde{x} P[X](\{\tilde{x}\})$$

definiert. Der Mittelwert lässt sich also ebenso wie die Quantile allein auf Grund der Kenntnis von  $X$  bzw.  $P[X]$  berechnen. Wegen der einfachen arithmetischen Form des Mittelwerts kann man einfache Rechenregeln formulieren. Insbesondere gilt für beliebige reelle Zahlen  $a$  und  $b$  und zwei Variablen  $X$  und  $Y$  mit dem selben Definitionsbereich  $\Omega$ :

$$\begin{aligned} M(aX + b) &= a M(X) + b \\ M(X + Y) &= M(X) + M(Y) \end{aligned}$$

**16.** Die *Varianz* ist die durchschnittliche quadratische Abweichung der Werte einer Variablen von ihrem Mittelwert, also:

$$V(X) := M((X - M(X))^2) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} (X(\omega) - M(X))^2$$

Die Quadratwurzel der Varianz wird *Standardabweichung* genannt. (Man beachte, dass bei der Berechnung der Varianz gelegentlich nicht durch  $|\Omega|$ , sondern durch  $|\Omega| - 1$  dividiert wird.)

Ebenso wie für den Mittelwert ergeben sich einfache Rechenregeln, insbesondere ist:

$$\begin{aligned} V(X) &= M(X^2) - M(X)^2 \\ V(aX + b) &= a^2 V(X) \end{aligned}$$

Die Varianz erlaubt eine Abschätzung der Häufigkeit einer Abweichung vom Mittelwert. Für alle  $a > 0$  gilt die *Tschebyscheffsche Ungleichung*:

$$\begin{aligned} P[X](\{\tilde{x} \mid a < |\tilde{x} - M(X)|\}) &\leq \frac{V(X)}{a^2} \\ P[X](\{\tilde{x} \mid a \geq |\tilde{x} - M(X)|\}) &\geq 1 - \frac{V(X)}{a^2} \end{aligned}$$

## Mehrdimensionale Verteilungen

**17.** Bisher haben wir nur eine einzelne Variable,  $X$ , betrachtet. Meistens benötigt man jedoch mehrere Variablen, um die Objekte in der Gesamtheit  $\Omega$  sinnvoll zu charakterisieren. Praktisch gesprochen erhebt man dann für jedes Objekt mehrere Merkmale, wobei jedes Merkmal durch eine Variable repräsentiert wird. Die Gesamtheit der Beobachtungen wird dann durch eine *mehrdimensionale Variable* dargestellt. In formaler Notation:

$$(X, Y, Z, \dots): \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \dots$$

Zu jedem  $\omega \in \Omega$  gibt es dann eine komplexe Beobachtung

$$(X(\omega), Y(\omega), Z(\omega), \dots)$$

also gleichzeitig einen Merkmalswert in  $\tilde{\mathcal{X}}$ , in  $\tilde{\mathcal{Y}}$ , in  $\tilde{\mathcal{Z}}$ , usw. Wenn die mehrdimensionale Variable  $K$  Komponenten hat, spricht man auch von einer  *$K$ -dimensionalen Variablen*.

**18.** Viele der Begriffsbildungen, die wir für einzelne Variablen eingeführt haben, lassen sich auf mehrdimensionale Variablen übertragen. Die *mehrdimensionale Häufigkeitsverteilung* ist definiert als eine Funktion

$$P[X, Y, Z, \dots]: \mathcal{P}(\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \dots) \longrightarrow [0, 1]$$

wobei für eine Teilmenge  $A \subseteq \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \dots$

$$P[X, Y, Z, \dots](A) := \frac{1}{|\Omega|} |(X, Y, Z, \dots)^{-1}(A)| = \frac{1}{|\Omega|} |\{\omega \in \Omega \mid (X, Y, Z, \dots)(\omega) \in A\}|$$

die relative Häufigkeit ist, mit der Objekte in  $\Omega$  Merkmalswerte  $(X(\omega), Y(\omega), Z(\omega), \dots)$  in  $A$  haben.

**19.** Hat die mehrdimensionale Variable nur wenige Komponenten, kann man die gemeinsame Verteilung in Form einer *mehrdimensionalen Häufigkeitstabelle* darstellen. Dies bietet sich insbesondere an, wenn es sich um eine zwei-dimensionale Variable handelt. Ihre Häufigkeitsverteilung kann dann übersichtlich durch eine zwei-dimensionale Tabelle angegeben werden.

**20.** Die *mehrdimensionale Verteilungsfunktion* ist definiert durch

$$\begin{aligned} F[X, Y, Z, \dots]: \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \times \dots &\longrightarrow [0, 1] \\ F[X, Y, Z, \dots](\tilde{x}, \tilde{y}, \tilde{z}, \dots) &:= P[X, Y, Z, \dots](\{\tilde{x}' \leq \tilde{x}\} \times \{\tilde{y}' \leq \tilde{y}\} \times \{\tilde{z}' \leq \tilde{z}\} \times \dots) \\ &= \frac{1}{|\Omega|} |\{\omega \in \Omega \mid X(\omega) \leq \tilde{x} \wedge Y(\omega) \leq \tilde{y} \wedge Z(\omega) \leq \tilde{z} \wedge \dots\}| \end{aligned}$$

**21.** Ist eine mehrdimensionale Variable  $(X, Y, Z, \dots)$  gegeben, kann man jede ihrer Komponenten einzeln betrachten. Man spricht dann von den *Randverteilungen* der einzelnen Variablen. Ist die *gemeinsame Verteilung* der Variablen  $P[X, Y, Z, \dots]$  gegeben, dann erhält man die Randverteilung etwa der Variablen  $X$  durch

$$P[X](A) = P[X, Y, Z, \dots](A \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \times \dots)$$

wobei  $A \subseteq \tilde{\mathcal{X}}$  ist. Es ist klar, dass dann ein wesentlicher Aspekt verloren geht, nämlich der Zusammenhang der Variablen. Kennt man nur die Randverteilungen der einzelnen Variablen, kann man daraus im allgemeinen die gemeinsame Verteilung nicht rekonstruieren.

**22.** Einen Aspekt der gemeinsamen Verteilung charakterisiert die *Kovarianz* zweier Variabler  $(X, Y)$ . Sie ist definiert als:

$$\begin{aligned} \text{Cov}(X, Y) &:= M((X - M(X))(Y - M(Y))) \\ &= \frac{1}{|\Omega|} \sum_{\omega \in \Omega} (X(\omega) - M(X))(Y(\omega) - M(Y)) \\ &= \sum_{(\tilde{x}, \tilde{y}) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}} (\tilde{x} - M(X))(\tilde{y} - M(Y)) P[X, Y](\{\tilde{x}, \tilde{y}\}) \\ &= M(X \cdot Y) - M(X)M(Y) \end{aligned}$$

Für die Summe zweier Variabler ist

$$\begin{aligned} V(X + Y) &= M(((X + Y) - M(X + Y))^2) \\ &= M(((X - M(X)) + (Y - M(Y)))^2) \\ &= M((X - M(X))^2 + (Y - M(Y))^2 + 2(X - M(X))(Y - M(Y))) \\ &= V(X) + V(Y) + 2M((X - M(X))(Y - M(Y))) \\ &= V(X) + V(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

Analog findet man die Gleichung

$$V(X - Y) = V(X) + V(Y) - 2\text{Cov}(X, Y)$$

Für die Kovarianz gilt die Abschätzung

$$-\sqrt{V(X)V(Y)} \leq \text{Cov}(X, Y) \leq \sqrt{V(X)V(Y)}$$

so dass man an Stelle der Kovarianz oft die *Korrelation* betrachtet, die durch

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

definiert ist. Die Korrelation nimmt also nur Werte zwischen -1 und 1 an. Ist  $M(X) = M(Y) = 0$  und  $V(X) = V(Y) = 1$ , dann erhält man aus der Tschebyscheffschen Ungleichung für die Variable  $X - Y$

$$P[X, Y](\{a < |X - Y|\}) \leq \frac{V(X - Y)}{a^2} = \frac{2 - 2\text{Corr}(X, Y)}{a^2}$$

Je mehr sich die Korrelation zwischen  $X$  und  $Y$  dem Wert 1 nähert, desto seltener treten Paare von Merkmalswerten auf, die sich um mehr als  $a$  unterscheiden.

## Bedingte Verteilungen

**23.** Sei  $X$  eine ein- oder mehrdimensionale Variable und sei  $B \subseteq \tilde{\mathcal{X}}$ . Ist  $P[X](B) > 0$ , dann heißt

$$\begin{aligned} P[X|X \in B]: \mathcal{P}(\tilde{\mathcal{X}}) &\longrightarrow [0, 1] \\ P[X|X \in B](A) &:= \frac{P[X](A \cap B)}{P[X](B)} \end{aligned}$$

*bedingte Häufigkeitsverteilung unter der Bedingung  $X \in B$ .* Die bedingte Häufigkeitsverteilung hat alle Eigenschaften einer gewöhnlichen Häufigkeitsverteilung, ist also eine additive Mengenfunktion. Sie kann als die Häufigkeitsverteilung aufgefasst werden, die durch die Beschränkung von  $P[X]$  auf die Teilmenge  $X^{-1}(B)$  von  $\Omega$  entsteht.

**24.** Durch Umstellen der Definitionsgleichung erhält man

$$P[X](A \cap B) = P[X|X \in B](A) P[X](B)$$

Ist nun  $\{B_1, B_2, \dots, B_m\}$  eine Partition von  $\tilde{\mathcal{X}}$ , dann ist

$$P[X](A) = \sum_{j=1}^m P[X|X \in B_j](A) P[X](B_j)$$

Insbesondere ergibt sich

$$P[X|X \in A](B_j) = \frac{P[X](A \cap B_j)}{P[X](A)} = \frac{P[X|X \in B_j](A) P[X](B_j)}{\sum_{k=1}^m P[X|X \in B_k](A) P[X](B_k)}$$

was manchmal *Satz von Bayes* genannt wird.

**25.** Dem Begriff der bedingten Häufigkeitsverteilung kommt bei der Behandlung mehrdimensionaler Variablen ein zentraler Stellenwert zu. Wir betrachten insbesondere bedingte Verteilungen, deren Bedingungen durch Komponenten der gemeinsamen Verteilung formuliert sind. Wir geben hier die Definition für den Fall einer zwei-dimensionalen Variablen  $(X, Y)$  an. Es seien  $B \subseteq \tilde{\mathcal{X}}$  und  $A \subseteq \tilde{\mathcal{Y}}$  zwei beliebige Mengen aus den jeweiligen Merkmalsräumen. Dann ist

$$P[Y|X \in B](A) = \frac{P[X, Y](A \times B)}{P[X](B)}$$

Dies ist die Häufigkeit, mit der  $Y$  einen Wert in  $A$  annimmt, unter der Bedingung, dass  $X$  einen Wert in  $B$  annimmt. Dabei wird natürlich vorausgesetzt, dass  $P[X](B) > 0$  ist. Diese bedingte Häufigkeit ist ein spezieller Fall der allgemeinen Definition bedingter Häufigkeiten, weil nur Teilmengen der Form  $A \times B$  aus  $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$  betrachtet werden.

**26.** Bedingte Verteilungen lassen sich ebenso wie unbedingte Verteilungen repräsentieren und charakterisieren. Wir betrachten hier nur bedingte Verteilungsfunktionen, bedingte Mittelwerte und bedingte Varianzen. Die *bedingte Verteilungsfunktion* ist definiert durch

$$F[Y|X \in B]: \tilde{\mathcal{Y}} \longrightarrow [0, 1]$$

$$\begin{aligned} F[Y|X \in B](\tilde{y}) &:= P[Y|X \in B](\{\tilde{y}' \leq \tilde{y}\}) \\ &= \frac{1}{|X^{-1}(B)|} |\{\omega \in \Omega | X(\omega) \in B \wedge Y(\omega) \leq \tilde{y}\}| \end{aligned}$$

*Bedingte Mittelwerte* und *bedingte Varianzen* werden entsprechend definiert:

$$\begin{aligned} M(Y|X \in B) &:= \frac{1}{|X^{-1}(B)|} \sum_{\omega \in X^{-1}(B)} Y(\omega) \\ &= \sum_{\tilde{y} \in \tilde{\mathcal{Y}}} \tilde{y} P[Y|X \in B](\{\tilde{y}\}) \\ V(Y|X \in B) &:= M((Y - M(Y|X \in B))^2 | X \in B) \\ &= M(Y^2 | X \in B) - M(Y | X \in B)^2 \end{aligned}$$

**27.** Es ist oft sehr nützlich, die bedingten Repräsentationen und Charakterisierungen als neue statistische Variable aufzufassen. Um das zu erreichen, muss die bedingte Repräsentation oder Charakterisierung als Funktion aller  $\omega \in \Omega$  aufgefasst werden. Dazu betrachten wir für jedes  $\omega$  die Menge

$$X^{-1}(X(\omega)) = \{\omega' | X(\omega') = X(\omega)\}$$

d.h. die Menge aller Elemente von  $\Omega$ , die in der Variablen  $X$  den gleichen Wert aufweisen wie  $\omega$ . Nun können wir bedingte Repräsentationen und Charakterisierungen für alle  $\omega \in \Omega$  bilden, indem wir diese Menge als Bedingung betrachten. Für die bedingte Verteilungsfunktion erhalten wir:

$$\begin{aligned} F[Y|X]: \tilde{\mathcal{Y}} \times \Omega &\longrightarrow [0, 1] \\ F[Y|X](\tilde{y})(\omega) &:= F[Y|X \in X(\omega)](\tilde{y}) = \frac{1}{|X^{-1}(X(\omega))|} |\{\omega' | Y(\omega') \leq \tilde{y} \wedge X(\omega') = X(\omega)\}| \end{aligned}$$

wobei wir  $F[Y|X](\tilde{y})$  als statistische Variable auffassen. Für die bedingten Mittelwerte und Varianzen erhalten wir:

$$\begin{aligned} M[Y|X](\omega) &:= M(Y|X = X(\omega)) = \frac{1}{|X^{-1}(X(\omega))|} \sum_{\omega' \in X^{-1}(X(\omega))} Y(\omega') \\ V[Y|X](\omega) &:= V(Y|X = X(\omega)) = M[Y^2|X](\omega) - (M[Y|X](\omega))^2 \end{aligned}$$

Der Zusammenhang zwischen unbedingten und bedingten Größen lässt nun sich einfach schreiben:

$$\begin{aligned} F[Y](\tilde{y}) &= M(F[Y|X](\tilde{y})) \\ M(Y) &:= M(M[Y|X]) \\ V(Y) &:= M(V[Y|X]) + V(M[Y|X]) \end{aligned}$$