

StRiH. A. Gündel-vom Hofe

**Kurzskript „Statistik für Biologen“
Teil 1: Deskriptive Statistik**

1. Der Begriff Merkmal

Ein **Merkmal** X ist eine Abbildung $X: \Omega \rightarrow M$ von einer *Grundgesamtheit* Ω in eine Menge M . Die Bilder $X(\omega) \in M$ der *Untersuchungseinheiten* $\omega \in \Omega$ nennt man auch konkrete **Ausprägungen** des Merkmals X . Einen Fragebogen mit k Fragen an ein Individuum kann man als ein *k-dimensionales Merkmal* $\mathbf{X} = (X_1, \dots, X_k)$ verstehen.

Je nach Struktur von M bzw. der Wertemenge $X(\Omega) \subseteq M$ unterscheidet man zwischen

- einem **nominalen** Merkmal: $X(\Omega)$ besitzt keine weitere Struktur. Die Ausprägungen sind *Namen*;
- einem **ordinal skalierten** oder **ordinalen** Merkmal: $X(\Omega)$ ist eine *geordnete* Menge, d.h. die Ausprägungen sind Namen einer Reihenfolge;
- einem **kardinal skalierten** oder **kardinalen** Merkmal: $X(\Omega)$ trägt die Struktur des mit der euklidischen Metrik versehenen Raumes \mathbf{R}^k . In diesem Fall nennt man $\mathbf{X} = (X_1, \dots, X_k)$ auch ein *k-dimensionales kardinales Merkmal*.

Ein nominales bzw. ein ordinales Merkmal bezeichnet man auch als *qualitatives* Merkmal, ein kardinales auch als *quantitatives* Merkmal.

Besitzt X höchstens abzählbar viele mögliche Ausprägungen $X(\omega)$, so nennt man X auch ein **diskretes** Merkmal. Existiert ein Kontinuum von möglichen Ausprägungen $X(\omega) \in M$, so spricht man auch von einem **stetigen** Merkmal X .

Bemerkung:

Hat man mit x_1, \dots, x_n n Ausprägungen eines *ordinalen* oder *kardinalen* Merkmals ermittelt oder gegeben, so schreibt man für die der Reihe nach geordneten Daten im Allgemeinen $x_{(1)}, \dots, x_{(n)}$, wobei die Indizes also in Klammern gesetzt sind. In diesem Fall ist also $x_{(1)}$ das erste bzw. kleinste und $x_{(n)}$ das letzte oder größte Datum.

2. Darstellungsformen

Ergebnisse einer Befragung, eines Experiments oder einer Beobachtung werden meist in *Tabellen*, gegliedert in Zellen, Zeilen und Spalten, präsentiert. Die Tabelleninhalte werden zudem häufig *graphisch* dargestellt.

Wichtigste graphische Darstellungen sind:

- (Diskrete) **Zeitreihen** in Form einer dargestellten Folge von Wertepaaren (t, x_t) mit $t = 1, \dots, T$, welche manchmal durch *lineare Interpolation* verbunden sind. In einer korrekten Graphik muss der Nullpunkt der Ordinate angegeben oder – falls das nicht möglich oder nicht sinnvoll ist – die Ordinate deutlich unterbrochen gezeichnet werden.

- **Stab- oder Kreisdiagramme** zur Visualisierung der *Häufigkeitsverteilung eines qualitativen Merkmals* (siehe z.B. Landtags- und Bundestagswahlen). Dabei entspricht die relative Häufigkeit der Merkmalsausprägung der prozentualen Höhe der entsprechenden Striche, Stäbe oder Balken bzw. dem prozentualen Winkelanteil des Kreis-segments zur Gesamtfläche (*Tortendiagramm*).

Der Beschreibung der Häufigkeitsverteilung eines quantitativen Merkmals dient auch die

empirische Verteilungsfunktion $\hat{F}(x) = \frac{1}{n} \cdot \sum_{i=1}^n I_{(-\infty, x]}(x_i) = \frac{1}{n} \cdot \text{card}\{x_i : x_i \leq x\}$. Der

Wert $\hat{F}(x)$ stellt also die relative Häufigkeit der Elemente dar, deren Merkmalsausprägung kleiner oder gleich x ist. Dabei bezeichnet $\text{card } M$ die Anzahl der Elemente einer (endlichen) Menge M und $I_{(-\infty, x]}$ die **Indikatorfunktion** des Intervalls $(-\infty, x]$, gegeben durch

$$I_{(-\infty, x]}(a) = \begin{cases} 1, & \text{falls } a \leq x, \\ 0, & \text{falls } a > x. \end{cases} \quad \text{Ist } \hat{F} \text{ auf der Basis von } n \text{ Beobachtungswerten}$$

x_1, \dots, x_n erstellt worden, so ist \hat{F} eine monoton wachsende Treppenkurve mit $0 \leq \hat{F}(x) \leq 1$.

Zuweilen betrachtet man anstelle der *empirischen* auch eine **theoretische Verteilungsfunktion** F als ein Modell der Realität. In diesem Zusammenhang kann man zu jeder Zahl α zwischen 0 und 1 die Zahl $x_\alpha = \min\{x : F(x) \geq \alpha\}$ betrachten, das sogenannte *untere* α -**Quantil**. Entsprechend heißt die Zahl $x_{1-\alpha}$ das *obere* α -Quantil.

Hat man es mit einer ordinal bzw. kardinal der Größe nach geordneten Datenmenge $x_{(1)}, \dots, x_{(n)}$ zu tun, so lässt sich das α -Quantil auch über folgende Definition einführen:

$$x_\alpha = x_{(N)} \quad \text{mit} \quad \frac{N-1}{n} < \alpha \leq \frac{N}{n}.$$

Als besondere Quantile sind herauszuheben

- das **untere Quartil** $x_{0.25}$ für $\alpha = 0.25$,
- das **obere Quartil** $x_{0.75}$ für $1-\alpha = 0.75$ und
- der **Median** $x_{0.5}$ für $\alpha = 0.5$.

Mithilfe dieser Zahlen sowie dem *Minimum* $x_{\min} = x_{(1)}$ und dem *Maximum* $x_{\max} = x_{(n)}$ der beobachteten Werte x lässt sich ein sogenanntes **Box-and-Whiskers-Plot** oder kurz **Boxplot** zeichnen:

Dabei ist (in vertikaler Richtung gezeichnet) die Höhe der Box gegeben durch den **Quartilsabstand** $QA = x_{0.75} - x_{0.25}$. Die Gesamthöhe des Plots mit den oberhalb und unterhalb der Box gezeichneten „Whiskers“ ist gleich der **Spannweite** der Daten $x_{\max} - x_{\min} = x_{(n)} - x_{(1)}$. Sogenannte **Ausreißer** an Daten werden mit einem * gekennzeichnet. Dies sind nach einer Faustregel alle Werte $x < x_{0.25} - 1.5 \cdot QA$ sowie $x > x_{0.75} + 1.5 \cdot QA$.

Im Fall eines *gruppierten stetigen Merkmals* wählt man gern als flächentreue Darstellung der Häufigkeitsverteilung dieses Merkmals das sogenannte **Histogramm**. Dabei wird über jedem

„Gruppen“-Intervall $(g_{j-1}, g_j]$ eine Säule gezeichnet, deren Flächeninhalt der Anzahl n_j der Daten in dieser Gruppe – man spricht auch von der Gruppenbesetzung – entspricht. Insbesondere ist die Säulenhöhe h_j gegeben durch $h_j = \frac{n_j}{b_j}$ mit $b_j = g_j - g_{j-1}$.

3. Lage- und Streuungsparameter

Ein grundlegendes Bild von der Verteilung der Daten liefern sogenannte **Verteilungsparameter**, welche spezielle Aspekte der Verteilung beschreiben herausheben. Wir unterscheiden hier *Lage-* und *Streuungsparameter*.

A) **Lageparameter** zur Beschreibung des *Zentrums* der Daten sind:

- Der **Modus** x_{mod} ist die am häufigsten vorkommende Ausprägung.
- Der **Median** x_{med} oder *Zentralwert* der Daten teilt die der Größe nach geordneten Daten in zwei gleich große Hälften. Es gilt, wenn n Daten gegeben sind:

$$\boxed{x_{\text{med}} = x_{\left(\frac{n+1}{2}\right)}} \quad (n = 2k + 1 \text{ ungerade}), \quad \boxed{x_{\text{med}} = \frac{1}{2} \left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right]} \quad (n = 2k \text{ gerade}).$$

- Das **arithmetische Mittel** $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ als *Schwerpunkt* der Daten.

B) **Streuungsparameter** zur Beschreibung der Streuung der Ausprägungen um das *Zentrum* sind:

- Die **Spannweite** $\boxed{x_{\text{max}} - x_{\text{min}} = x_{(n)} - x_{(1)}}$ als Differenz zwischen der größten und der kleinsten Merkmalsausprägung.
- Der **Quartilsabstand** $\boxed{QA = x_{0.75} - x_{0.25}}$ als Differenz zwischen *oberem* und *unterem* Quartil.
- Die **mittlere absolute Abweichung** als Mittelwert der Abweichungen vom Median ist

gegeben durch $\boxed{\frac{1}{n} \sum_{i=1}^n |x_i - x_{\text{med}}|}$. Insbesondere gilt für jeden beliebigen Wert

$$a \in \mathbf{R}: \quad \frac{1}{n} \sum_{i=1}^n |x_i - x_{\text{med}}| \leq \frac{1}{n} \sum_{i=1}^n |x_i - a|.$$

- Der **Median MAD der Abweichungen vom Median** $\text{MAD} = \text{Median} \left\{ |x_i - x_{\text{med}}| \right\}$
- Die **Varianz** als Mittelwert der *quadratischen* Abweichung vom Mittelwert

$$\boxed{\text{var}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$
 sowie die **Standardabweichung** $\boxed{s = \sqrt{\text{var}(\mathbf{x})}$.

Bemerkungen:

- a) Bei einer (affin) linearen Transformation $y_i = \alpha + \beta x_i$ ($i = 1, \dots, n$) der Daten transformiert sich der Mittelwert und die Varianz, wie folgt:

$$\boxed{\bar{y} = \alpha + \beta \bar{x} = \alpha + \beta \bar{x}} \quad \text{ sowie } \quad \boxed{\text{var}(\mathbf{y}) = \beta^2 \text{var}(\mathbf{x})} .$$

b) Speziell erhält man mit $\alpha = -\frac{\bar{x}}{\sqrt{\text{var}(\mathbf{x})}}$ und $\beta = \frac{1}{\sqrt{\text{var}(\mathbf{x})}}$ die sogenannte *Standardisierung der Daten*:

$\mathbf{x}_i^* = \frac{x_i - \bar{x}}{\sqrt{\text{var}(\mathbf{x})}}$. Standardisierte Daten besitzen dann den *Mittelwert*: $\bar{x}^* = 0$ und haben die *Varianz*: $\text{var}(\mathbf{x}^*) = 1$.

c) Bezüglich der Varianz gilt der sogenannte *Verschiebungssatz*:

Ist $a \in \mathbf{R}$ eine beliebige reelle Zahl, so gilt: $\boxed{\frac{1}{n} \sum_{i=1}^n (x_i - a)^2 = \text{var}(\mathbf{x}) + (\bar{x} - a)^2}$.

Speziell mit $a = 0$ erhält man daraus die alternative Berechnungsformel für die Varianz,

die da lautet: $\boxed{\text{var}(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2}$.

4. Die Kovarianz und empirische Korrelation mehrdimensionaler Daten

Untersucht man die Ausprägungen eines zweidimensionalen Merkmals $\mathbf{X} = (X, Y)$, so kann man jedes der Daten $\mathbf{X}(\omega_i) = (X(\omega_i), Y(\omega_i)) = (x_i, y_i) = \mathbf{z}_i$ graphisch als einen Punkt in der zweidimensionalen Ebene darstellen. Auf diese Weise entsteht eine *Punktwolke*.

Zeichnet man zusätzlich noch ein „Fadenkreuz“ durch den *Schwerpunkt* $S = (\bar{x}, \bar{y})$ der Punktwolke, so ist eine gewisse Steigungstendenz der Punktwolke erkennbar.

Wichtige Größen zur Beschreibung der Punktwolke und eines Zusammenhangs zwischen den beiden Komponenten X und Y des 2-dimensionalen Merkmals \mathbf{X} sind:

- Die **empirische Kovarianz** von (\mathbf{x}, \mathbf{y}) : $\boxed{\text{cov}(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}$
- Der **empirische Korrelationskoeffizient** nach Bravais und Pearson, bezogen auf die standardisierten Einzelmerkmale: $\boxed{r(\mathbf{x}, \mathbf{y}) = \text{cov}(\mathbf{x}^*, \mathbf{y}^*)}$.

Bemerkungen:

a) Für den Korrelationskoeffizienten ergibt sich immer ein Wert zwischen -1 und $+1$.

b) Insbesondere heißen zwei Datenvektoren \mathbf{x} und \mathbf{y} im Fall $r(\mathbf{x}, \mathbf{y}) = 0$ *unkorreliert*.

c) Im Fall $\mathbf{x} = \mathbf{y}$ erhält man speziell: $\boxed{\text{cov}(\mathbf{x}, \mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \text{var}(\mathbf{x})}$.

d) Betrachtet man das durch $S = (\bar{x}, \bar{y})$ laufende Fadenkreuz und bezeichnet man die dabei entstehenden Quadranten der Reihe nach entgegen dem Uhrzeigersinn mit I, II, III und IV, so deutet das *Vorzeichen* von $\text{cov}(\mathbf{x}, \mathbf{y})$ auf folgendes hin:

(i) $\text{cov}(\mathbf{x}, \mathbf{y}) > 0$: Die Punktwolke verteilt sich verstärkt auf die Quadranten I und III.

(ii) $\text{cov}(\mathbf{x}, \mathbf{y}) < 0$: Die Punktwolke verteilt sich verstärkt auf die Quadranten II und IV.

e) Analog zur Varianz kann man folgende äquivalente Formel für die Kovarianz herleiten:

$$\text{cov}(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}.$$

f) Führt man als Abkürzung $s_{xy} := \text{cov}(\mathbf{x}, \mathbf{y})$ und $s_x := \sqrt{\text{var}(\mathbf{x})}$ ein, so lässt sich der

empirische Korrelationskoeffizient auch schreiben in der Form
$$r(\mathbf{x}, \mathbf{y}) = \frac{s_{xy}}{s_x s_y}.$$

g) Umschreibt man die Punktwolke zu einem zweidimensionalen Merkmal $\mathbf{X} = (X, Y)$ mit einer freihändig gezeichneten Ellipse und betrachtet deren maximalen vertikalen Gesamtdurchmesser D sowie deren vertikalen Durchmesser d im Zentrum (\bar{x}, \bar{y}) , so be-

sagt die *Ellipsenregel*:
$$r(\mathbf{x}, \mathbf{y})^2 \approx 1 - \left(\frac{d}{D} \right)^2.$$

5. Lineare Regression und Ausgleichsgeraden

Möchte man eine Punktwolke eines zweidimensionalen Merkmals $\mathbf{X} = (X, Y)$ möglichst „gut“ durch entsprechende Bezugspunkte beschreiben, die auf einer sogenannten **Ausgleichsgeraden** in der Ebene liegen, so vollzieht man eine sogenannte **lineare Regression**.

Als Kriterium für eine möglichst optimale Gerade dient die von Gauß stammende **Methode der kleinsten Quadrate**: Ersetzt man die Punkte $\mathbf{z}_i = (x_i, y_i)$ durch entsprechende Punkte $\hat{\mathbf{z}}_i = (\hat{x}_i, \hat{y}_i)$ auf der gedachten Ausgleichsgeraden, so soll die sogenannte *Fehlerquadratsumme* SSE^1 , definiert durch

$$SSE = \sum_{i=1}^n \|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2 = \sum_{i=1}^n [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

möglichst klein werden.

Es bieten sich vor allem folgende zwei Ausgleichsgeraden an, die jeweils durch den Schwerpunkt $S = (\bar{x}, \bar{y})$ der Punktwolke verlaufen:

(i) Die **Ausgleichsgerade von y nach x**:

Setze $\hat{x}_i = x_i$ und $\hat{y}_i = \alpha + \beta x_i$ ($i = 1, \dots, n$). Dann ist α, β so zu bestimmen, dass

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\alpha + \beta x_i)]^2$$
 minimal wird. Die gesuchten Geraden-

parameter berechnen sich dann nach den Formeln:
$$\beta = \frac{\text{cov}(\mathbf{x}, \mathbf{y})}{\text{var}(\mathbf{x})}$$
 sowie
$$\alpha = \bar{y} - \beta \bar{x}.$$

(ii) Die **Ausgleichsgerade von x nach y**:

Setze $\hat{y}_i = y_i$ und $\hat{x}_i = \gamma + \delta y_i$ ($i = 1, \dots, n$). Dann ist γ, δ so zu bestimmen, dass

$$SSE = \sum_{i=1}^n (x_i - \hat{x}_i)^2 = \sum_{i=1}^n [x_i - (\gamma + \delta y_i)]^2$$
 minimal wird. Es ist:
$$\delta = \frac{\text{cov}(\mathbf{x}, \mathbf{y})}{\text{var}(\mathbf{y})}$$
 und

$$\gamma = \bar{x} - \delta \bar{y}.$$
 Für die Geradengleichung gilt dann:
$$x = \gamma + \delta y$$
 bzw.
$$y = -\frac{\gamma}{\delta} + \frac{1}{\delta} x.$$

¹ SSE steht für Sum of Squares of Errors

Teil 2: Grundlagen der Wahrscheinlichkeitstheorie

1. Naive Mengenlehre

Da wir inzwischen öfter den Begriff der Menge und Schreibweisen der Mengenlehre verwendet haben, soll an dieser Stelle ein kurzer Exkurs zur (naiven) Mengenlehre erfolgen. Die Mengenlehre liefert sozusagen die universelle Basis für die Sprache der Mathematik.

Definition der Menge nach Georg Cantor (1895):

„Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten, wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens, welche die Elemente von M genannt werden, zu einem Ganzen.“

Dabei gelten folgende Regeln:

- Von der „Definition“ Cantors ausgehend soll von einem beliebigen Objekt a eindeutig aussagbar sein, ob es zu einer gegebenen Menge M gehört bzw. *Element der Menge* ist oder nicht. In Zeichen:

$a \in M$: „ a ist Element der Menge M “ bzw. „ a gehört zu M “ bzw. „ a ist aus M “

$a \notin M$: „ a ist nicht Element der Menge M “ bzw. „ a gehört nicht zu M “ bzw. „ a ist nicht aus M “.

- Zur Kennzeichnung der Menge als *Zusammenfassung* von Objekten verwendet man die *geschweiften Mengenklammern* „{“ und „}“. Dabei unterscheidet man zur Beschreibung der Menge das

a) das *aufzählende Verfahren*:

Die Menge wird durch explizite Aufzählung ihrer Elemente beschrieben, wobei es hierbei auf die Reihenfolge der Elemente *nicht* ankommt und es gleichgültig ist, ob ein Element *mehrfach* aufgezählt wird oder nur *einfach*. Damit gilt z.B.:

$$A = \{2, 3, 5, 7\} = \{5, 3, 7, 2\} = \{2, 3, 3, 5, 2, 7, 5, 5\}$$

Im Falle *unendlicher* Mengen oder größerer *endlicher* Mengen verwendet man das Symbol „...“, um Platz zu sparen, wenn für die in der Menge zusammengefaßten Objekte ein eindeutiges Bildungsgesetz erkennbar ist.

b) das *beschreibende Verfahren*:

Ausgehend von einer *Grundmenge* G werden die Elemente, welche zur betrachteten Menge gehören, durch eine (oder mehrere) *Eigenschaften* in Form einer Aussageform $p(x)$ beschrieben, welche für einige Objekte x *wahr* ist, für andere *falsch*. In Zeichen:

$A = \{x \in G \mid p(x)\}$: „ A ist die Menge aller x aus G , für die $p(x)$ wahr ist“.

Insbesondere gilt dann: $a \in A \Leftrightarrow p(a)$ ist wahr,
 $a \notin A \Leftrightarrow p(a)$ ist falsch.

Auch spielt es keine Rolle, wie die *Individuenvariable* genannt wird, und es kann anstelle von „/“ zuweilen auch „:“ bzw. „;“ verwendet werden. Also gilt z.B.

$$A = \{x \in G \mid p(x)\} = \{t \in G \mid p(t)\} = \{y \in G : p(y)\} = \{m \in G ; p(m)\}$$

- Zugelassen ist auch die Menge, die *kein* Element enthält, die sogenannte *leere Menge*. In Zeichen: $\emptyset := \{\}$ bzw. $\emptyset := \{x \in G \mid x \neq x\}$.

- Ist x ein beliebiges Objekt, so nennt man die Menge $A = \{x\}$ die zu x gehörige *Elementarmenge* oder *Einermenge*. Zu unterscheiden ist hierbei deutlich zwischen der *Menge* $\{x\}$ und dem *Element* x .

Wir kommen im folgenden nun zu zwei wesentlichen Mengenbeziehungen.

Definition:

Seien A und B zwei Mengen mit Elementen aus einer Grundmenge G .

- 1) A heißt *Teilmenge* (oder *Untermenge*) von B bzw. B eine *Erweiterungsmenge* (oder *Obermenge*) von A (in Zeichen: $A \subseteq B$), falls für alle $x \in G$ gilt:
 $x \in A \Rightarrow x \in B$ (d.h. wenn x in A , dann auch x in B).
- 2) A und B heißen *gleich* (in Zeichen: $A = B$), falls für alle $x \in G$ gilt:
 $x \in A \Leftrightarrow x \in B$ (d.h. x in A genau dann, wenn x in B).

Für die Teilmengenrelation „ \subseteq “ gelten folgende wichtigen Eigenschaften:

- $\emptyset \subseteq A$ für alle Mengen A . Damit ist die Sonderstellung der leeren Menge \emptyset erklärt.
- $A \subseteq A$ für alle Mengen A . Damit ist die Teilmengenbeziehung *reflexiv*.
- $A \subseteq B$ und $B \subseteq C \Rightarrow A \subseteq C$ für alle Mengen A, B und C . Damit ist die Teilmengenbeziehung *transitiv*.
- $A = B \Leftrightarrow A \subseteq B$ und $B \subseteq A$ für alle Mengen A und B . Damit ist die *Gleichheit* von Mengen *mittels* der *Teilmengenbeziehung* beschrieben. Tatsächlich wird die Gleichheit zweier Mengen A und B konkret oft durch Nachweis der doppelten Teilmengenbeziehung hergeleitet.
- Man kann auch zu einer gegebenen Menge A alle ihre Teilmengen zu einer neuen Menge von Mengen zusammenfassen, der sogenannten *Potenzmenge* von A . In Zeichen:

$$\wp(A) := \{ M \mid M \subseteq A \} .$$

Man beachte: Die Elemente von $\wp(A)$ sind selbst wieder Mengen!!! Speziell gilt hierbei immer: $\wp(A) \neq \emptyset$, denn es ist $\emptyset \in \wp(A)$ wegen $\emptyset \subseteq A$ sowie $A \in \wp(A)$ wegen $A \subseteq A$.

Seien im folgenden A und B beliebige Teilmengen einer Grundmenge G . Dann lassen sich folgende spezielle *Mengenoperationen* einführen, um aus A und B neue Mengen zu erzeugen:

Mengenoperationen:	
Schnittmenge / Durchschnitt:	$A \cap B := \{ x \in G \mid x \in A \text{ und } x \in B \}$ <p><u>Spezialfall:</u> $A \cap B := \emptyset$; dann heißen A und B <i>disjunkt</i>.</p>
Vereinigungsmenge / Vereinigung:	$A \cup B := \{ x \in G \mid x \in A \text{ oder } x \in B \}$

Differenzmenge / Differenz:	$A \setminus B := \{ x \in G \mid x \in A \text{ und } x \notin B \}$
Komplementärmenge / Komplement von G	$A^C := G \setminus A = \{x \in G \mid x \notin A\}$
Cartesisches Produkt von A und B	$A \times B := \{ (x, y) \mid x \in A \text{ und } y \in B \}$ Die Elemente $(x, y) \in A \times B$ mit $x \in A, y \in B$ heißen <i>geordnete Paare</i>
Charakterisierende Eigenschaft der geordneten Paare:	$(a, b) = (c, d) \Leftrightarrow a = c \text{ und } b = d$ Also <i>komponentenweise Gleichheit</i> gefordert

Bemerkungen:

- Das in der Definition der *Vereinigung* auftretende (logische) „oder“ ist ein *einschließendes* „oder“, umfaßt also auch das (logische) „und“.
- Zur Beschreibung der *Komplementärmenge* von A findet man in Büchern teilweise auch die Schreibweisen A' bzw. \bar{A} bzw. $-A$.
- Zwischen der Teilmengenbeziehung zweier Mengen A und B und der Teilmengenbeziehung ihrer *Komplementärmen*gen A^C und B^C besteht folgender Zusammenhang:

$$(A \subseteq B) \Leftrightarrow (B^C \subseteq A^C)$$

- Man beachte, daß es sich bei den *geordneten Paaren* (a, b) mit $a \in A, b \in B$ um *Elemente* einer Menge (nämlich des Cartesischen Produkts von A und B) handelt, bei $\{a, b\}$ hingegen um eine *Menge*.
- Bei den *geordneten Paaren* kommt es auf die Reihenfolge der Komponenten an, nicht so bei den *Mengen*, d.h. es gilt:

$$(a, b) \neq (b, a) \text{ für } a \neq b, \text{ aber stets } \{a, b\} = \{b, a\}$$

- In Verallgemeinerung zu den geordneten Paaren (a, b) mit $a \in A, b \in B$ lassen sich auch sogenannte *n-Tupel* mit n Komponenten einführen. Dabei gilt:

$$(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n) \Leftrightarrow a_1 = b_1 \text{ und } a_2 = b_2 \text{ und } \dots \text{ und } a_n = b_n$$

Auf diese Weise erhält man das verallgemeinerte *Cartesische Produkt* von n (nicht leeren) Mengen A_1, A_2, \dots, A_n :

$$A_1 \times A_2 \times \dots \times A_n := \{ (a_1, a_2, \dots, a_n) \mid a_1 \in A_1 \text{ und } a_2 \in A_2 \text{ und } \dots \text{ und } a_n \in A_n \}$$

- Ist $A_1 = A_2 = \dots = A_n = A$ ($n \in \mathbf{N}$), so schreibt man abkürzend:

$$A^n := \underbrace{A \times A \times \dots \times A}_{n\text{-mal}} = \{ (a_1, a_2, \dots, a_n) \mid a_1, a_2, \dots, a_n \in A \}$$

Im Fall $A = \mathbf{R}$ bietet das Cartesische Produkt die Möglichkeit zur Beschreibung von Punkten in der *Ebene* / dem *Raum* durch $P = (x, y) \in \mathbf{R}^2$ bzw. $P = (x, y, z) \in \mathbf{R}^3$.

Für die Mengenoperationen „ \cap “, „ \cup “ und Komplementbildung gelten - analog zu den Zahlen - bestimmte „Rechenregeln“, die in gewisser Weise an die Rechenregeln für Zahlen bezüglich der Operationen „+“, „ \cdot “ und die Vorzeichenbildung „-“ erinnern. Zudem spielen die leere und die Grundmenge bzgl. der Mengenoperationen eine ähnliche Rolle wie 0 und 1 im Reich der Zahlen.

Gesetze der Mengenoperationen		
	Durchschnitt „ \cap “	Vereinigung „ \cup “
Assoziativgesetze	$(A \cap B) \cap C = A \cap (B \cap C)$	$(A \cup B) \cup C = A \cup (B \cup C)$
Kommutativgesetze	$A \cap B = B \cap A$	$A \cup B = B \cup A$
Distributivgesetze	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
Absorptionsgesetze	$A \cap (A \cup B) = A$	$A \cup (A \cap B) = A$
Idempotenzgesetze	$A \cap A = A$	$A \cup A = A$
Gesetz für das Komplement	$A \cap A^c = \emptyset$	$A \cup A^c = G$
Gesetz des doppelten Komplements	$(A^c)^c = A$	
De Morgans Gesetze	$(A \cap B)^c = A^c \cup B^c$	$(A \cup B)^c = A^c \cap B^c$
Operationen mit G und \emptyset	$G \cap A = A$ $\emptyset \cap A = \emptyset$	$\emptyset \cup A = A$ $G \cup A = G$
	$\emptyset^c = G, G^c = \emptyset$	

Man beachte:

Im Gegensatz zu den Rechengesetzen gilt in der Mengenalgebra ein strenges *Dualitätsprinzip*, wie der Vergleich zweier jeweils in einer Zeile stehenden Gesetze zeigt:

- Vertauscht man konsequent „ \cap “ und „ \cup “ sowie „ \emptyset “ und „ G “ in einem der Mengengesetze, erhält man das entsprechende dazu „duale“ Mengengesetz.

2. Wahrscheinlichkeitsaxiomatik

Die Grundlage der Wahrscheinlichkeitstheorie bildet der auf der Mengenlehre und Abbildungstheorie basierende Begriff des **Wahrscheinlichkeitsraumes**. Dazu gehört:

1. der **Grundraum** oder **Ergebnisraum** $\Omega \neq \emptyset$ als Menge aller möglichen Ergebnisse eines Zufallsexperiments,
2. eine (endliche) σ -**Algebra** S , bestehend aus Teilmengen von Ω , für welche gilt:

$$\text{a) } \boxed{\Omega \in S}, \text{ b) } \boxed{A \in S \Rightarrow A^c = \Omega - A \in S}, \text{ c) } \boxed{A_i \in S (i = 1, \dots, n) \Rightarrow \bigcup_{i=1}^n A_i \in S}.$$

Die Elemente $A \in S$ dieser σ -Algebra S werden *Ereignisse* genannt. Insbesondere heißt $\Omega \in S$ selbst das *sichere* und $\emptyset \in S$ das *unmögliche Ereignis*.

3. eine Abbildung $P: S \rightarrow \mathbf{R}$, definiert auf einer σ -Algebra S und **Wahrscheinlichkeitsmaß** oder **Wahrscheinlichkeit** genannt, für welche gilt:

a) $0 \leq P(A) \leq 1$ für alle $A \in S$, b) $P(\Omega) = 1$,

c) $P\left(\bigcup_{i=1}^n A_i\right) = P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) = P(A_1) + \dots + P(A_n)$ für $A_i \cap A_k = \emptyset, i \neq k$.

Die Eigenschaft (c) heißt auch die σ -**Additivität** des Wahrscheinlichkeitsmaßes P bezüglich paarweise disjunkter Ereignisse. Alle drei Eigenschaften zusammen nennt man auch die **Axiome von Kolmogorov**.

Für den Wahrscheinlichkeitsraum schreibt man zusammenfassend das Tripel: $(\Omega; S; P)$.

Bemerkungen:

• Aus den Axiomen der σ -Algebra folgt u.a.:

(i) $\emptyset = \Omega^C \in S$ sowie wegen der de Morganschen Gesetze:

(ii) $A_i \in S (i = 1, \dots, n) \Rightarrow \bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^C\right)^C \in S$.

• Aus den Axiomen von Kolmogorov folgen u.a. auch die Formeln:

(i) $P(A^C) = 1 - P(A)$, (ii) $P(\emptyset) = 0$,

(iii) $P(A - B) = P(A) - P(A \cap B)$ für $A, B \in S$ beliebig sowie

(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für $A, B \in S$ beliebig.

• Es gibt auch sogenannte *unendliche* σ -Algebren, in denen man z.B. Ereignisse der Form A : „Es wird solange gewürfelt, bis eine Sechs erscheint“ beschrieben werden können. Wir werden diese jedoch in unserer Vorlesung außer Acht lassen.

• Jedes System $\{A_i | i = 1, \dots, n\} \subseteq S$ von Ereignissen, welche eine Zerlegung des Ereignisraums Ω bilden – also mit den Eigenschaften (i) $\Omega = \bigcup_{i=1}^n A_i$ und (ii) $A_i \cap A_k = \emptyset$ für $i \neq k$ – heißt auch ein **vollständiges Ereignisfeld**.

3. Laplace-Experimente

Einen Sonderfall für einen Wahrscheinlichkeitsraum stellt der *Laplacesche* Wahrscheinlichkeitsraum dar. Dabei gilt:

• $\Omega = \{\omega_1, \dots, \omega_n\}$ ist eine endliche Menge

• $S = \wp(\Omega)$ ist die Potenzmenge, enthält also alle *Elementarereignisse* $A_i = \{\omega_i\} \in S$.

• $P(A_i) = P(\{\omega_i\}) = \frac{1}{n}$ für $i = 1, \dots, n$; d.h. alle Elementarereignisse sind *gleichwahrscheinlich*.

Bezeichnet man mit $|A|$ für ein Ereignis $A \in S$ die Anzahl der in A enthaltenen Ergebnisse $\omega \in A$ – auch *Kardinalzahl* von A genannt –, dann erhält man für das Wahr-

scheinlichkeitsmaß P die Darstellung:
$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

Experimente, welche sich mithilfe des Modells eines Laplaceschen Wahrscheinlichkeitsraumes beschreiben lassen, heißen auch *Laplace-Experimente*. Beispiele dafür sind:

- Das Werfen einer idealen Münze mit „Kopf“ und „Zahl“,
- Das Würfeln mit einem idealen Würfel,
- Das Ziehen einer Karte aus einem 32-teiligen gut gemischten Kartenspiel
- Das zufällige Ziehen von k Kugeln aus einer Urne mit n verschiedenen Kugeln usw.

Man kann auch mehrere solcher Laplace-Experimente zu einem *mehrstufigen* Laplace-Experiment oder *stochastischen Prozess* kombinieren. Dessen Veranschaulichung gelingt in durchsichtiger Weise mittels eines sogenannten *Baumdiagramms*, dessen einzelne *Pfade* die möglichen verschiedenen Ausgänge dieses Mehrstufenexperiments symbolisieren. Dabei gilt:

- Die Wahrscheinlichkeit eines möglichen Gesamtergebnisses – also eines Pfades – ist gleich dem Produkt aller Wahrscheinlichkeiten entlang eines Pfades.

4. Grundlagen der Kombinatorik und grundlegende Urnenmodelle

Das *Grundmodell* der endlichen Kombinatorik lautet:

In einer Urne befinden sich n verschiedene Kugeln mit den Nummern $1, 2, \dots, n$. Aus dieser Urne wird eine *Stichprobe vom Umfang s* (d.h. s Kugeln) entnommen.

Die dazu *duale Version dieses Modells* lautet:

Es werden s Kugeln auf n verschiedene Urnen verteilt.

Es gibt hierbei im wesentlichen vier Arten der Ziehung bzw. Verteilung, die sich kombinatorisch aus den Fallunterscheidungen (i) *mit / ohne Zurücklegen* und (ii) *mit / ohne Berücksichtigung der Reihenfolge* ergeben:

- Ziehen *mit Zurücklegen mit Berücksichtigung der Reihenfolge*; die Ergebnisse heißen Variationen mit Wiederholung
- Ziehen *ohne Zurücklegen mit Berücksichtigung der Reihenfolge*; die Ergebnisse heißen Variationen ohne Wiederholung bzw. im Fall $s = n$ Permutationen
- Ziehen *mit Zurücklegen ohne Berücksichtigung der Reihenfolge*; die Ergebnisse heißen Kombinationen mit Wiederholung
- Ziehen *ohne Zurücklegen ohne Berücksichtigung der Reihenfolge*; die Ergebnisse heißen Kombinationen ohne Wiederholung

Um die Anzahlen der entsprechenden Möglichkeiten zu beschreiben, benötigt man zwei wichtige Zahlenkonstrukte:

- die **Fakultät** einer natürlichen Zahl $n \in \mathbf{N}_0$,
- den **Binomialkoeffizienten** n über k für zwei natürliche Zahlen $n, k \in \mathbf{N}_0$.

<p>Definition der Fakultät mittels Produkt:</p>	<p>Für $n \in \mathbf{N}_0$ ist die <i>Fakultät</i> $n!$ (lies: „n Fakultät“) definiert durch:</p> $n! := \prod_{k=1}^n k = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 \quad \text{mit}$ $0! := \prod_{k=1}^0 k = 1 \quad (\text{leeres Produkt}).$
<p>rekursive Definition der Fakultät:</p>	<p>Startwert: $0! := 1$, Rekursionsgleichung: $(n+1)! := n!(n+1) \quad (n \in \mathbf{N}_0)$</p>
<p>Definition des (allgemeinen) Binomialkoeffizienten mittels Produkt:</p>	<p>Für $\alpha \in \mathbf{R}$ beliebig und $k \in \mathbf{N}_0$ ist der <i>Binomialkoeffizient</i> $\binom{\alpha}{k}$ (lies: „α über k“) definiert durch:</p> $\binom{\alpha}{k} := \prod_{j=1}^k \frac{\alpha+1-j}{j} = \frac{\alpha \cdot (\alpha-1) \cdot \dots \cdot (\alpha-k+1)}{k!} \quad \text{mit}$ $\binom{\alpha}{0} = \prod_{j=1}^0 \frac{\alpha+1-j}{j} = 1 \quad (\text{leeres Produkt}).$
<p>rekursive Definition der Binomialkoeffizienten:</p>	<p>Für $\alpha \in \mathbf{R}$ beliebig gilt:</p> <p>Startwert: $\binom{\alpha}{0} := 1$,</p> <p>Rekursionsgleichung: $\binom{\alpha}{k+1} := \frac{\alpha-k}{k+1} \cdot \binom{\alpha}{k} \quad (k \in \mathbf{N}_0)$.</p>
<p>Erste Rechenregeln für die Binomialkoeffizienten:</p>	<p>Ist $\alpha = n \in \mathbf{N}_0$, $k \in \mathbf{N}_0$ und $k \leq n$, so gilt:</p> <p>(1) $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, (2) $\binom{n}{k} = \binom{n}{n-k}$,</p> <p>(3) $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$,</p> <p>(4) $\sum_{j=k}^n \binom{j}{k} = \binom{k}{k} + \binom{k+1}{k} + \dots + \binom{n}{k} = \binom{n+1}{k+1}$,</p> <p>(5) $\sum_{k=0}^n \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n$.</p>

Weitere Rechenregeln für die Binomialkoeffizienten:	(6) $\sum_{k=0}^n (-1)^k \binom{n}{k} = \binom{n}{0} - \binom{n}{1} + \dots + (-1)^n \binom{n}{n} = 0$ (7) für $n = k_1 + k_2 + \dots + k_s$ mit $k_i \in \mathbf{N}_0$ ($1 \leq i \leq s$) gilt: $\binom{n}{k_1} \binom{n-k_1}{k_2} \dots \binom{n-k_1-\dots-k_{s-1}}{k_s} = \frac{n!}{k_1! k_2! \dots k_s!}$
---	--

Bemerkungen:

- Die Formel (3) gilt sogar allgemein für jedes beliebige $\alpha \in \mathbf{R}$ und $k \in \mathbf{N}$. Mittels vollständiger Induktion kann unter Rückgriff auf diese sogar bewiesen werden:

$$\boxed{\binom{n}{k} \in \mathbf{N}_0} \text{ für } n \in \mathbf{N}_0 \text{ und } k = 0, \dots, n.$$

- Für $n, k \in \mathbf{N}_0$ folgt aus der Definition des Binomialkoeffizienten: $\boxed{\binom{n}{k} = 0}$ für $k > n$.

- Man kann zeigen, dass für $\alpha \in \mathbf{R}$ beliebig und $k \in \mathbf{N}_0$ gilt: $\boxed{\binom{\alpha+1}{k} := \frac{\alpha+1}{\alpha-k+1} \cdot \binom{\alpha}{k}}$.

- Zur Abkürzung der rechten Seite in (7) führt man in einigen Büchern den sogenannten *Multinomialkoeffizienten* ein: $\boxed{\binom{n}{k_1 k_2 \dots k_s} := \frac{n!}{k_1! k_2! \dots k_s!}}$.

Binomialkoeffizienten wurden zuerst in *China* verwendet. Ihre zugehörige symbolische Schreibweise stammt allerdings vom Mathematiker *Leonhard Euler*. Auf den Mathematiker (und Philosophen) *Blaise Pascal* geht folgende schematische Anordnung der Binomialkoeffizienten in Form eines *Dreiecks* zurück unter Rückgriff auf (3) und den Startwert:

$$\begin{array}{cccccccc}
 & & & & \binom{0}{0} & & & & \\
 & & & & = 1 & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & \\
 & & & & = 1 & & = 1 & & \\
 & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \\
 & & & & = 1 & & = 2 & & = 1 & \\
 & & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
 & & & & = 1 & & = 3 & & = 3 & & = 1 & \\
 & & & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
 & & & & = 1 & & = 4 & & = 6 & & = 4 & & = 1 & \\
 & & & & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & \\
 & & & & = 1 & & = 5 & & = 10 & & = 10 & & = 5 & & = 1 & \\
 \dots & & & & \dots & & \dots & & \dots & & \dots & & \dots & & \dots & \dots
 \end{array}$$

Mittels der *Fakultät* und des *Binomialkoeffizienten* ergeben sich nun die folgenden *Formeln* für die einzelnen *Anzahlen* an möglichen Ausgängen in den beschriebenen 4 Grundmodellen der Urnenziehung (n Kugeln, s Ziehungen) :

Ziehen:	ohne Zurücklegen	mit Zurücklegen
mit Berücksichtigung der Reihenfolge: (Variationen / Permutationen)	$(n)_s := n \cdot (n-1) \cdot \dots \cdot (n-s+1)$ $= \frac{n!}{(n-s)!}$ Speziell für $n = s$: $(n)_n := n!$	n^s
ohne Berücksichtigung der Reihenfolge: (Kombinationen)	$\binom{n}{s} = \frac{n!}{s!(n-s)!}$	$\binom{n+s-1}{s} = \frac{(n+s-1)!}{s!(n-1)!}$

Bemerkung:

Man kann auch sogenannte Variationen mit festgelegter Auswahl bilden. In diesem Fall ist die Anzahl der verschiedenen Möglichkeiten gegeben durch:

$$\left[\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \dots \cdot \binom{n-n_1-\dots-n_{s-1}}{n_s} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_s!} \right].$$

Dabei verteilt man n verschiedene Kugeln auf s leere Urnen so, daß in Urne Nr. 1 genau n_1 Kugeln, in Urne Nr. 2 genau n_2 Kugeln, ... in Urne Nr. s genau n_s Kugeln zu liegen kommen. Man beachte, dass gilt: $n = n_1 + n_2 + \dots + n_s$.

5. Bedingte Wahrscheinlichkeiten

Seien $A, B \subseteq \Omega$ zwei Ereignisse in einem Wahrscheinlichkeitsraum (Ω, S, P) mit $P(B) \neq 0$, so kann man nach der sogenannten **bedingten Wahrscheinlichkeit von A unter der Bedingung (Hypothese) B** fragen:

$$\left[P_B(A) = P(A | B) := \frac{P(A \cap B)}{P(B)} \right].$$

Entsprechend ergibt sich dann für die bedingte Wahrscheinlichkeit von B unter der Bedingung (Hypothese) A :

$$\left[P_A(B) = P(B | A) := \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \right].$$

Im Fall eines *Laplaceschen* Wahrscheinlichkeitsraumes Ω erhält man für die beiden Ereignisse $A, B \subseteq \Omega$ speziell die bedingten Wahrscheinlichkeiten

$$\left[P_B(A) := P(A | B) = \frac{|A \cap B|}{|B|} \right] \text{ und } \left[P_A(B) := P(B | A) = \frac{|A \cap B|}{|A|} \right]$$

als ein Maß für den „Anteil“ des Ereignisses A im neu gewählten Grundraum $B \subseteq \Omega$.

Bemerkung:

Objektivistisch gesehen misst $P_B(A) = P(A | B)$ die relative Häufigkeit, mit der das Ereignis A eingetreten ist, in Bezug auf die Gesamtheit der Male, in denen B eingetreten ist.

Subjektivistisch interpretiert stellt $P(A | B)$ unsere Einschätzung dar, dass das Ereignis A eintritt, wenn wir wissen, dass B eingetreten ist.

Eine wichtige Folgerung aus der Definition ist die **Bayes-Formel**, welche den Zusammenhang zwischen den beiden verschiedenen bedingten Wahrscheinlichkeiten in Bezug auf zwei gegebene Ereignisse $A, B \subseteq \Omega$ aufzeigt:

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

Sie resultiert aus der Umstellung der definierenden Formel für die bedingte Wahrscheinlichkeit nach $P(A \cap B)$, wie folgt:

$$P(A | B) \cdot P(B) = P(A \cap B) = P(B \cap A) = P(B | A) \cdot P(A)$$

und lässt sich in der Sicht der sogenannten *subjektiven Wahrscheinlichkeit* so interpretieren:

Stelle A eine unbekannte potentielle Ursache für eine Beobachtung oder ein Symptom B dar, und trete A mit der Wahrscheinlichkeit $P(A)$ auf. Liegt A vor, so trete das Symptom B mit der Wahrscheinlichkeit $P(B | A)$ auf. Nun wird B beobachtet: Mit welcher Wahrscheinlichkeit liegt A vor? Die Antwort liefert die Bayes-Formel.

In diesem Zusammenhang heißt $P(A)$ auch die *A-priori*-Wahrscheinlichkeit von A , welche unser Wissen **vor** der Beobachtung repräsentiert, und $P(A | B)$ auch die *A-posteriori*-Wahrscheinlichkeit von A , welche unser Wissen **nach** der Beobachtung repräsentiert.

Es folgen einige Sätze und Begriffe im Zusammenhang mit der bedingten Wahrscheinlichkeit:

a) **Satz von der totalen Wahrscheinlichkeit:**

Bildet $\{A_i | i = 1, \dots, n\}$ ein vollständiges Ereignisfeld – also eine Zerlegung des Ereignisraums Ω mit den Eigenschaften (i) $\Omega = \bigcup_{i=1}^n A_i$ und (ii) $A_i \cap A_k = \emptyset$ für $i \neq k$ –, so

gilt für jedes Ereignis $B \subseteq \Omega$: **(**)**
$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B | A_i).$$

Bemerkung: (**) folgt aus der disjunkten Zerlegung $B = B \cap \left(\bigcup_{i=1}^n A_i \right) = \bigcup_{i=1}^n (B \cap A_i)$.

b) **Satz von Bayes**

Aus dem Einsetzen der Formel (**) für die totale Wahrscheinlichkeit in die Bayes-Formel (*) erhält man für ein Ereignis $B \subseteq \Omega$ und ein vollständiges Ereignisfeld $\{A_i | i = 1, \dots, n\}$ die folgende Formel:

$$(***) \quad P(A_k | B) = \frac{P(B | A_k)}{\sum_{i=1}^n P(A_i) \cdot P(B | A_i)} \cdot P(A_k) \quad \text{für } k = 1, \dots, n.$$

Bemerkung: Im häufig angewendeten Fall einer Zerlegung $\Omega = A \cup A^C$ ergibt sich dann z.B. speziell:

$$P(A | B) = \frac{P(B | A)}{P(A) \cdot P(B | A) + P(A^C) \cdot P(B | A^C)} \cdot P(A)$$

mit $P(A^C) = 1 - P(A)$.

c) Die (stochastische) Unabhängigkeit von Ereignissen

Zwei Ereignisse $A, B \subseteq \Omega$ heißen **stochastisch unabhängig**, falls gilt:

$$P(A \cap B) = P(A) \cdot P(B).$$

Dann folgt: $P_B(A) := P(A | B) = P(A)$ bzw. $P_A(B) := P(B | A) = P(B)$. Also gilt:

Für das Eintreten des Ereignisses A ist das Eintreten des Ereignisses B irrelevant und umgekehrt.

In Verallgemeinerung nennt man n Ereignisse A_1, \dots, A_n **(total) unabhängig**, falls für jede beliebige Teilmenge A_{i_1}, \dots, A_{i_k} , $k \in \{1, \dots, n\}$ aus diesen Ereignissen gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k}).$$

Damit sind im Fall von drei Ereignissen $A, B, C \subseteq \Omega$ folgende 4 Gleichungen zu prüfen:
 $P(A \cap B) = P(A) \cdot P(B)$, $P(A \cap C) = P(A) \cdot P(C)$, $P(B \cap C) = P(B) \cdot P(C)$,
 $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$.

Bemerkungen:

- Sind A und B stochastisch unabhängig, dann sind es auch jeweils die Ereignisse (i) A^C und B , (ii) A und B^C sowie (iii) A^C und B^C .
- Allgemeiner: Ersetzt man im Fall von n stochastisch unabhängigen Ereignissen A_1, \dots, A_n einen beliebigen Teil A_{i_1}, \dots, A_{i_k} , $k \in \{1, \dots, n\}$ dieser Ereignisse durch ihre Gegenereignisse $A_{i_1}^C, \dots, A_{i_k}^C$, so bleibt die stochastische Unabhängigkeit für die neue Ereignisfamilie bewahrt.

6. Zufallsgrößen, Zufallsvariablen als Abbildungen

Der Begriff der *Abbildung* ist wie auch der Begriff der *Menge* von fundamentaler Bedeutung für die Mathematik.

Definition:

Seien A und B zwei Mengen. Unter einer *Abbildung von A in B* (oder: *nach B*) versteht man eine Zuordnungsvorschrift f , durch welche *jedem* $x \in A$ *genau ein* $y \in B$ zuordnet wird. Gilt für die (Ziel-)Menge B : $B \subseteq \mathbf{R}$, so spricht man auch gern von einer (reellen) *Funktion*.

Einige Bemerkungen bzw. Begriffe:

- Man nennt A den *Definitions- oder Urbildbereich* von f (in Zeichen: $D_f = A$), B den *Werte-, Ziel- oder Bildbereich* von f .
- Das durch f abgebildete Element $x \in A$ heißt auch *Argument* oder *Variable* von f , das dem Element $x \in A$ eindeutig zugeordnete Element $y \in B$ das *Bild* von x unter f oder *Funktionswert* von f an der Stelle x . Das Element x selbst nennt man auch (ein) *Urbild* von y unter f . Man schreibt: $y = f(x)$.
- Zur Beschreibung der Abbildung f von $D_f = A$ in B findet man die *symbolischen Schreibweisen* $f: D_f \rightarrow B, x \mapsto f(x)$ oder $f: x \mapsto f(x)$ mit $x \in D_f$ oder auch $y = f(x)$ mit $x \in D_f$. Manchmal benutzt man in Kurzform auch nur $x \mapsto f(x)$ o.ä.
- Ist $M \subseteq A$ eine Teilmenge von A , so heißt die Menge $f(M) := \{f(x) \mid x \in M\} \subseteq B$ das *Bild* von M unter f . Speziell heißt $f(A) \subseteq B$ auch das *Bild* von f . Oft findet man als Definition für den *Wertebereich* von f in diesem Zusammenhang: $W_f = f(A)$.
- Ist $N \subseteq B$ eine Teilmenge von B , so heißt die Menge $f^{-1}(N) := \{x \in A \mid f(x) \in N\} \subseteq A$ das *Urbild* von N unter f . Speziell gilt: $f^{-1}(B) = D_f = A$. Als *Urbild eines Elements* $y \in B$ unter f erhält man also speziell: $f^{-1}(\{y\}) := \{x \in A \mid f(x) = y\}$.

Den allgemeinen Begriff der Abbildung bzw. reellen Funktion spezifizieren wir nun im Zusammenhang mit einem gegebenen Wahrscheinlichkeitsraum (Ω, \mathcal{S}, P) :

Definition:

Unter einer **Zufallsvariablen** oder **Zufallsgröße** auf Ω versteht man eine Abbildung $X: \Omega \rightarrow \mathbf{R}$, $\omega \mapsto X(\omega)$, durch welche jedem möglichen Ergebnis $\omega \in \Omega$ eine eindeutige reelle Zahl zugeordnet wird. Wichtig ist hierbei, dass für jedes reelle Intervall $[a, b] \subseteq \mathbf{R}$ das *Urbild* $X^{-1}([a, b]) := \{\omega \in \Omega \mid a \leq X(\omega) \leq b\} \subseteq \Omega$ ein Element der σ -Algebra \mathcal{S} darstellt.

Bemerkungen:

- Man beachte, dass aus eher historischen Gründen die Zufallsvariablen als Abbildungen in der Wahrscheinlichkeitstheorie gern mit Großbuchstaben aus dem hinteren Alphabet bezeichnet werden. Also: X, Y, Z u.ä.
- Ist das Bild $X(\Omega) \subseteq \mathbf{R}$ speziell eine *endliche* Menge, dann reicht zum Nachweis, dass es sich bei der Abbildung X um eine Zufallsgröße auf Ω handelt, die Tatsache, dass für jede reelle Zahl $x \in \mathbf{R}$ das Urbild $X^{-1}(x) = \{\omega \in \Omega \mid X(\omega) = x\}$ ein Element der Ereignisalgebra \mathcal{S} darstellt.
- Man schreibt häufig $\{X = x\} := X^{-1}(x) \subseteq \Omega$ für das Urbild von $x \in \mathbf{R}$ unter der Zufallsgröße X sowie, davon ausgehend, $P(X = x) := P(\{X = x\}) = P(\{\omega \in \Omega \mid X(\omega) = x\})$ für die zugehörige Wahrscheinlichkeit dieses Ereignisses.

- Da Zufallsgrößen X, Y, Z reelle Funktionen sind, kann man sie mittels verschiedener Rechenoperationen, wie *Addieren*, *Subtrahieren*, *Vervielfachen*, *Multiplizieren* und *Dividieren*, zu einer neuen Zufallsgröße zusammenfassen, indem man diese Operationen auf den Bildern der entsprechenden Zufallsgrößen durchführt. Also z.B.:

$$(X + Y)(\omega) := X(\omega) + Y(\omega) \text{ und } (X \cdot Y)(\omega) := X(\omega) \cdot Y(\omega) \quad (\omega \in \Omega).$$

In den meisten uns interessierenden Fällen haben wir es mit einer sogenannten *diskreten* Zufallsvariable zu tun, d.h. einer reellen Funktion, welche auf Ω nur endlich viele verschiedene Werte annimmt. Also gilt in diesem Fall: $X(\Omega) = \{x_1, x_2, \dots, x_n\}$.

In diesem Fall definieren wir die **Wahrscheinlichkeitsverteilung** und die **Verteilungsfunktion**:

Definition:

Sei $X: \Omega \rightarrow \mathbf{R}$ eine diskrete Zufallsvariable mit $X(\Omega) = \{x_1, x_2, \dots, x_n\}$.

- a) Unter der (*Wahrscheinlichkeits-*)*Verteilung* der Variablen X versteht man eine Liste aller möglichen Werte $x_i \in X(\Omega)$ mit zugehörigen Wahrscheinlichkeiten $p_i = P(X = x_i) > 0$,

$$i = 1, \dots, n. \text{ Insbesondere gilt immer: } \sum_{i=1}^n p_i = p_1 + \dots + p_n = \sum_{i=1}^n P(X = x_i) = 1.$$

- b) Die reelle Funktion $F_X: \mathbf{R} \rightarrow [0, 1]$, gegeben durch die Vorschrift: $F_X(x) := P(X \leq x)$ ($x \in \mathbf{R}$), nennt man die *Verteilungsfunktion* der Zufallsvariablen X .

Bemerkung:

In Analogie zur empirischen Verteilungsfunktion in der deskriptiven Statistik besteht ein enger Zusammenhang zwischen der Wahrscheinlichkeitsverteilung und der Verteilungsfunktion, gegeben durch die Gleichung: $F_X(x) = \sum_{x_i \leq x} P(X = x_i)$.

In anderen Worten: $F_X(x)$ ergibt sich durch Aufsummieren aller Einzelwahrscheinlichkeiten $p_i = P(X = x_i)$ über die Werte $x_i \in X(\Omega)$ mit $x_i \leq x$ für gegebenes $x \in \mathbf{R}$.

7. Einige spezielle diskrete Wahrscheinlichkeitsverteilungen

Ausgehend von den verschiedenen Wahrscheinlichkeitsmodellen eines Experiments erhält man verschiedene Wahrscheinlichkeitsverteilungen für entsprechend Zufallsvariablen. Wir listen die wichtigsten im Folgenden auf.

- A) Die diskrete Gleichverteilung im Fall eines *Laplace-Experiments*:

Für den Wahrscheinlichkeitsraum $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ gelte $P(\{\omega_i\}) = \frac{1}{n}$, $i = 1, \dots, n$ (Laplace-Experiment). Weiterhin sei $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable mit $X(\omega_i) = x_i$ für $i = 1, \dots, n$ und $x_i \neq x_k$ für $i \neq k$. Die Verteilung, gegeben durch $p_i = P(X = x_i) = \frac{1}{n}$ für $i = 1, \dots, n$, heißt die **diskrete Gleichverteilung** von X über den Zahlen x_1, \dots, x_n .

B) Die Zweipunktverteilung im Fall eines *einstufigen Bernoulli-Experiments*:

Zugrunde liegt ein sogenanntes *Bernoulli-Experiment* mit nur zwei möglichen Ausgängen („Erfolg“ und „Misserfolg“). Ist $\Omega = \{\omega_1, \omega_2\}$ der entsprechende zweielementige Ergebnisraum, so nennt man die Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$ mit $X(\omega_1) = 0$, $X(\omega_2) = 1$ auch *Bernoullivariablen*. Die zu X gehörende Verteilung, welche gegeben ist durch $p = P(X = 1)$ und $q = P(X = 0) = 1 - p$, heißt dann **Zweipunkt-Verteilung** von X .

Achtung: Im Fall $p = q = \frac{1}{2}$ hat man es wieder mit einem Laplace-Experiment zu tun.

C) Die Binomialverteilung im Fall eines *n-stufigen Bernoulli-Experiments*:

Führt man ein *n-stufiges Bernoulli-Experiment* durch, wobei die Einzelergebnisse in den einzelnen Stufen voneinander unabhängig sind und jeweils mit „0“ und „1“ bezeichnet sind, dann betrachtet man auf $\Omega = \{0,1\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \{0,1\}\}$ gern die Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$, welche die Anzahl der „Erfolge“ in $\omega \in \Omega$ zählt. Insbesondere gilt:

$$X(\omega) = \omega_1 + \omega_2 + \dots + \omega_n = \sum_{i=1}^n \omega_i. \text{ Ist } p \text{ wieder die Einzelwahrscheinlichkeit für Erfolg}$$

(also: „1“) und $q = 1 - p$ die Einzelwahrscheinlichkeit für Misserfolg (also: „0“), so bezeichnet man X als **binomialverteilt** mit den Parametern n und $p > 0$. Insbesondere

gilt dann für $k = 1, \dots, n$:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

Achtung:

Bezüglich des zugrunde liegenden *Urnenmodells* handelt es sich beim mehrstufigen Bernoulli-Experiment um ein *n*-maliges *Ziehen mit Zurücklegen ohne Berücksichtigung der Reihenfolge* aus einer Urne mit 2 Sorten an Kugeln („1“ und „0“), wobei p und q den jeweiligen prozentualen Anteil beider Sorten im Gesamtsortiment angibt.

D) Die hypergeometrische Verteilung im Fall *Stichprobenwahl* in einer in zwei disjunkte Klassen zerfallenden Gesamtheit:

Wählt man aus einer Gesamtpopulation $G \neq \emptyset$ der Größe $|G| = N$, in welcher r Individuen ein bestimmtes Merkmal tragen und die restlichen $N - r$ Individuen nicht, eine zufällige *Stichprobe* im Umfang von n Elementen, so gelangt man zur **hypergeometrischen Verteilung**. Dazu definieren wir

$$\Omega = \left\{ \omega = A_n \mid A_n \subseteq G, |A_n| = n \right\}$$

als Ergebnisraum und betrachten die Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$, welche die Anzahl x der Individuen in der Stichprobe $\omega = A_n$ vom Umfang n zählt, welche das betrachtete Merkmal besitzen. Dann folgt für die Wahrscheinlichkeitsverteilung von X :

$$P(X = x) = \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}} \quad \text{für } x = 1, \dots, r.$$

Achtung:

Bezüglich des zugrunde liegenden *Urnenmodells* handelt es sich um ein *n*-maliges *Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge* aus einer Urne mit 2

Sorten an Kugeln („1“ und „0“), wobei $p = \frac{r}{N}$ und $q = \frac{N-r}{N}$ den jeweiligen prozentualen Anteilen beider Sorten in der Gesamtpopulation entsprechen.

Bemerkung:

Im Fall der Ziehung einer Stichprobe von relativ kleinem Umfang n im Verhältnis zur Gesamtpopulationsgröße N gleichen sich die *hypergeometrische* und die *Binomialverteilung* ziemlich dicht aneinander an, sofern x kleiner als r gewählt ist. Denn:

Die Wahrscheinlichkeit, dass eine Kugel aus einer großen Gesamtmenge bei einer relativ kleinen Zahl an Ziehungen trotz Zurücklegen ein zweites Mal gezogen wird, ist äußerst klein.

Also ersetzt man im Fall kleiner Stichproben aus einer großen Population in der angewandten Stochastik öfter das eigentlich dem Experiment korrekt entsprechende hypergeometrische Verteilungsmodell durch das leichter handzuhabende, da leichter zu berechnende, Binomialverteilungsmodell.