

Kapitel 2

**Überblick über die benötigten Grundlagen**

<b>2</b>	<b>Überblick über die benötigten Grundlagen</b>	<b>9</b>
2.1	Grundgesamtheit und Stichprobe .....	9
2.2	Zufallsvariable und Merkmal.....	10
2.3	Verteilung und Empirische Verteilung.....	12
2.4	Dichte und Häufigkeitsverteilung .....	14
2.5	Erwartungswert und Varianz .....	24
2.6	Abhängigkeit.....	31
2.7	Gängige Verteilungen und ihre Erwartungswerte und Varianzen .....	37

## 2 Überblick über die benötigten Grundlagen

### 2.1 Grundgesamtheit und Stichprobe

2.1

Mit Methoden der induktiven Statistik sollen Aussagen über Mengen von Personen oder Objekten getroffen werden. Wie bereits aus der deskriptiven Statistik bekannt, bezeichnet man solche Mengen oder Massen als Grundgesamtheiten (vergleiche auch Lehrbücher zur deskriptiven Statistik, etwa Burkschat et al. (2003), Mosler, Schmid (2003) oder in Teilen Fahrmeir et al. (2003)). Die Mehrzahl statistischer Analysen stützt sich bei ihren Aussagen jedoch nicht auf die komplette Grundgesamtheit, sondern wählt nach geeigneten Methoden Teilmengen aus Grundgesamtheiten aus. Diese so genannten Stichproben werden dann analysiert, und auf Basis der aus ihnen erhaltenen Ergebnisse werden Schlüsse auf die Grundgesamtheit gezogen.

---

#### Definition Grundgesamtheit

Eine **Grundgesamtheit** ist eine Menge von Personen oder Objekten, über die im Rahmen einer statistischen Untersuchung eine Aussage getroffen werden soll. Dabei ist die zu untersuchende Menge nach räumlichen, zeitlichen und sachlichen Kriterien genau einzugrenzen. Die Kriterien, nach denen eine Grundgesamtheit eingegrenzt wird, hängen vom Ziel der Untersuchung ab.

Die Elemente einer Grundgesamtheit heißen auch **Untersuchungseinheiten**.

---

#### Beispiel Grundgesamtheit


Zur besseren Planung von Wohnhausabrissen und -neubauten soll für die Bundesrepublik Deutschland eine nach Bundesländern gestaffelte regionale Wohnbedarfsprognose für die nächsten zehn Jahre erstellt werden. Es interessiert, wie viele Haushalte (man rechnet eine Wohnung pro Haushalt, gestaffelt nach Haushaltsgrößen) es in den einzelnen Bundesländern im Zeitraum der nächsten zehn Jahre geben wird. Die zu betrachtende Grundgesamtheit für jedes einzelne Bundesland ist daher – abgegrenzt nach den oben genannten Kriterien – die Menge aller in den nächsten zehn Jahren (zeitlich) in Haushalten zusammen lebender Personen (sachlich) in diesem Bundesland (räumlich).



**Definition Stichprobe**

Eine Teilmenge, die aus einer Grundgesamtheit zur statistischen Untersuchung einer interessierenden Fragestellung ausgewählt wird, heißt **Stichprobe**. Die Elemente einer Stichprobe werden auch **Erhebungseinheiten** genannt, die Stichprobe selbst die **Erhebungsgesamtheit**.

**B****Beispiel Stichprobe**

Im **Beispiel ►9** der Wohnbedarfsprognose ist die Grundgesamtheit eine sich in die Zukunft entwickelnde Masse. Als Stichprobe kann eine Auswahl der in einem Bundesland in Haushalten zusammen lebenden Personen an einem Stichtag der Gegenwart dienen. Anhand einer Befragung dieser Personen und zusätzlicher Information über Zu- und Abwanderung sowie die Bevölkerungsentwicklung der Vergangenheit können dann Aussagen über die zu erwartende Entwicklung getroffen werden. 

Im Rahmen dieses Buches werden wir nicht darauf eingehen, wie man zu guten Stichproben kommt. Die **Stichprobentheorie ►e** ist Inhalt eigener Veröffentlichungen (etwa Levy, Lemeshow (1999)). Gute Stichproben zeichnen sich dadurch aus, dass in ihnen die Grundgesamtheit bezüglich des interessierenden Untersuchungsziels im Kleinen abgebildet wird. Diese Eigenschaft nennt man **Repräsentativität ►e** einer Stichprobe. Wir gehen im Folgenden stets davon aus, dass die realisierten Stichproben für die interessierenden Grundgesamtheiten repräsentativ sind, so dass Schlüsse von der Stichprobe auf die Grundgesamtheit zulässig sind.

**2.2****2.2 Zufallsvariable und Merkmal**

Aus der deskriptiven Statistik ist bekannt, dass in einer statistischen Untersuchung in der Regel nicht die Untersuchungseinheiten selbst von Interesse sind, sondern sie auszeichnende Eigenschaften. Man spricht von der Erhebung so genannter **Merkmale**. Obwohl ein Merkmal bestimmte, in der Regel bekannte, Ausprägungen annehmen kann, weiß man vor der konkreten Durchführung einer Untersuchung nicht, welche Werte die einzelnen Erhebungseinheiten aufweisen. Man kann sich die Erhebung eines Merkmals an den Objekten einer Stichprobe daher auch vorstellen als die Durchführung eines (Zufalls-)Experiments, dessen Ausgang vorab nicht bekannt ist. Die hier enthaltene Zufallskomponente hat dazu geführt, dass man statt von einem Merkmal auch von einer **Zufallsvariable** spricht.

### Definition Zufallsvariable

Betrachtet wird eine Grundgesamtheit  $\Omega$ , bestehend aus Untersuchungseinheiten, an denen ein Merkmal  $X$  interessiert. Dieses Merkmal  $X$  kann aufgefasst werden als eine **Zufallsvariable**  $X : \Omega \rightarrow \mathbb{R}$ , das heißt als eine Abbildung der Grundgesamtheit auf die reellen Zahlen. Jedem Ereignis  $\omega \in \Omega$  wird durch  $X$  genau eine Zahl zugeordnet. Der Wertebereich der Zufallsvariablen  $X$  (das heißt die Menge aller möglichen Ausprägungen **►e** des Merkmals  $X$ ) sei mit  $\mathcal{X}$  bezeichnet.

Ist der Wertebereich  $\mathcal{X}$  abzählbar, so heißt  $X$  eine **diskrete** Zufallsvariable, enthält der Wertebereich  $\mathcal{X}$  ein ganzes Intervall aus den reellen Zahlen, so heißt  $X$  eine **stetige** Zufallsvariable.

Die Zufallsvariable selbst ist also eine fest definierte Funktion und daher eigentlich nicht zufällig. Dadurch, dass man bei einer statistischen Untersuchung aber vorher nicht weiß, mit welchen Elementen der Grundgesamtheit man es zu tun bekommt, sind die Werte, die  $X$  an einer Stichprobe annehmen wird, nicht vorher bekannt. Dies macht die Zufälligkeit hier aus.

So wie der Begriff der Zufallsvariable definiert ist, sind zunächst nur Merkmale  $X$  zugelassen, die reelle Zahlen als Ausprägungen liefern. Natürlich ist dies nicht immer unmittelbar gegeben, denn ein Merkmal, das beispielsweise **nominal** oder **ordinal** **►e** skaliert ist, kann als Ausprägungen auch verbale Begriffe annehmen (**männlich**, **weiblich** oder **schlecht**, **mittel**, **gut**). Um der **Definition ►11** zu genügen, wendet man bei solchen Merkmalen einen Trick an: man transformiert die verbalen Ausprägungen in Zahlen, das heißt man kodiert die Ausprägungen in Zahlenwerte um. Am ursprünglichen **Skalenniveau ►e** des Merkmals ändert sich dadurch aber nichts!

### Beispiel Zufallsvariable

In einer Untersuchung zu Fernsehgewohnheiten von Erstklässlern interessiert es, wie lange die Kinder täglich durchschnittlich fernsehen. Die betrachtete Grundgesamtheit ist die Menge aller in Deutschland lebenden Schulkinder in der ersten Klasse in einem ausgewählten Stichschuljahr. Das interessierende Merkmal  $X$  ist die durchschnittlich pro Tag vor dem Fernseher verbrachte Zeit. Die Zufallsvariable  $X$  ordnet jedem Erstklässler diese Zeit zu:

$$X : \text{Erstklässler } \omega \rightarrow \text{durchschnittliche tägliche Fernsehzeit von } \omega.$$



Liegt eine Stichprobe aus der Grundgesamtheit vor, so ist es Aufgabe der deskriptiven Statistik, die Häufigkeitsverteilung des interessierenden Merkmals zu beschreiben. Befasst man sich dagegen mit der Häufigkeitsverteilung des Merkmals in der Grundgesamtheit, so spricht man auch von der **Verteilung** oder **Wahrscheinlichkeitsverteilung** der Zufallsvariablen  $X$ .

## 2.3

## 2.3 Verteilung und Empirische Verteilung

Zur Untersuchung, mit welchen Anteilen welche Ausprägungen eines Merkmals in einer Stichprobe vorkommen, benutzt man in der deskriptiven Statistik die **empirische Verteilungsfunktion** ►e. Diese gibt zu jedem beliebigen Wert  $x$  an, wie hoch der Anteil der Erhebungseinheiten in der Stichprobe ist, deren Ausprägungen höchstens einen Wert von  $x$  besitzen. Analog definiert man die Verteilungsfunktion einer Zufallsvariablen  $X$ . Sie gibt zu jedem beliebigen Wert  $x$  an, wie hoch der Anteil der Untersuchungseinheiten in der Grundgesamtheit ist, deren Ausprägungen kleiner oder gleich  $x$  sind. Dabei setzt man die Anteile (**relativen Häufigkeiten** ►e) in der Grundgesamtheit gleich mit Wahrscheinlichkeiten. Dahinter steht die Vorstellung, dass bei zufälliger Ziehung aus einer Grundgesamtheit mit  $N$  Elementen, in der  $k$  Stück eine interessierende Eigenschaft besitzen, die Wahrscheinlichkeit, eine Untersuchungseinheit mit der interessierenden Eigenschaft zu erhalten, gerade  $\frac{k}{N}$  beträgt. Diese Umsetzung der relativen Häufigkeiten in Wahrscheinlichkeiten wird in der **Wahrscheinlichkeitsrechnung** ►e besprochen.



### Definition Verteilungsfunktion

Gegeben sei eine Zufallsvariable  $X$ . Die Funktion  $F^X$ , die die Wahrscheinlichkeit dafür beschreibt, dass  $X$  einen Wert annimmt, der kleiner oder gleich einer vorgegebenen Schranke  $x$  ist, heißt **Verteilungsfunktion** von  $X$

$$F^X(x) = P(X \leq x),$$

wobei  $F^X(x) \in [0; 1]$ ,  $x \in \mathbb{R}$  und  $\lim_{x \rightarrow -\infty} F^X(x) = 0$ ,  $\lim_{x \rightarrow \infty} F^X(x) = 1$ .



### Definition Parameter

Wird eine Verteilung eindeutig durch eine Kennzahl oder eine Gruppe (so genanntes Tupel) von Kennzahlen charakterisiert in dem Sinne, dass die gleiche Verteilung immer zu den gleichen Kennzahlen führt und dieselben Kennzahlen immer zu derselben Verteilung, so nennt man diese Kennzahlen **Parameter** der Verteilung. Zur

Verdeutlichung schreibt man für eine solche Verteilung statt  $F^X(x)$  häufig auch  $F^X(x; \vartheta)$ , wobei  $\vartheta$  für den oder die Parameter steht.

Ein Verteilungsmodell, das auf einer solchen Parametrisierung beruht, nennt man auch **parametrisches Modell**. Andernfalls spricht man von einem **nichtparametrischen Modell**.

Wir betrachten zunächst parametrische Modelle.

Häufig benutzt man die Verteilungsfunktion, um die so genannten **Quantile** anzugeben.

---

### Definition Quantil

Gegeben sei eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F^X$  und eine Zahl  $p \in (0; 1)$ .

1. Für eine diskrete Zufallsvariable  $X$  heißt eine Zahl  $x_p^*$  **(theoretisches) p-Quantil**, wenn gilt:

$$P(X < x_p^*) \leq p \quad \text{und} \quad P(X > x_p^*) \leq 1 - p.$$

Falls  $x_p^*$  aus dieser Beziehung nicht eindeutig bestimmbar ist, wählt man den kleinsten Wert, der diese Bedingung erfüllt.

2. Für eine stetige Zufallsvariable  $X$  heißt eine Zahl  $x_p^*$  **(theoretisches) p-Quantil**, wenn gilt:

$$F^X(x_p^*) = p.$$

Auch hier wählt man gegebenenfalls den kleinsten Wert  $x_p^*$ , der dies erfüllt.

Analog zur Definition der **Quantile** ►e aus der deskriptiven Statistik spricht man auch hier für  $p = 0,5$  vom **Median** und für  $p = 0,25$  bzw.  $p = 0,75$  vom **unteren** bzw. **oberen Quartil**.

## 2.4 Dichte und Häufigkeitsverteilung

In engem Zusammenhang mit der Verteilungsfunktion steht die **Dichtefunktion** (kurz: **Dichte**), die das Pendant zur relativen Häufigkeitsverteilung  $\blacktriangleright$ e darstellt. Wir unterscheiden bei der Definition der Dichte den Fall der diskreten und der stetigen Zufallsvariablen.

### Definition Dichtefunktion

1. Es sei  $X$  eine diskrete Zufallsvariable mit endlichem oder abzählbar unendlichem Wertebereich  $\mathcal{X} = \{x_1, x_2, x_3, \dots\}$ . Die **diskrete Dichte** von  $X$  ist die Funktion  $f^X$ , so dass für die Verteilungsfunktion  $F^X$  von  $X$  gilt

$$F^X(x) = \sum_{x_i \leq x} f^X(x_i).$$

Dabei kann man die Funktionswerte der diskreten Dichte angeben als

$$f^X(x_i) = P(X = x_i) \quad \text{für } i = 1, 2, \dots$$

Es gilt  $f^X(x_i) \geq 0$  für alle  $i$  und  $\sum_{x_i} f^X(x_i) = 1$ . Daraus folgt sofort, dass  $f^X(x_i) \leq 1$  ist für alle  $i$ .

Zur Berechnung der Wahrscheinlichkeit für ein Ereignis  $\{X \in A\}$  für  $A \subseteq \mathbb{R}$ , verwendet man

$$P(X \in A) = \sum_{x_i \in A} f^X(x_i) = \sum_{x_i \in A} P(X = x_i).$$

2. Es sei  $X$  eine stetige Zufallsvariable mit Wertebereich  $\mathcal{X} = \mathbb{R}$ . Die **stetige Dichte** von  $X$  ist die Funktion  $f^X$ , so dass für die Verteilungsfunktion  $F^X$  von  $X$  gilt

$$F^X(x) = \int_{-\infty}^x f^X(t) dt.$$

Dabei gilt  $f^X(x) \geq 0$  für alle  $x$  und  $\int_{-\infty}^{\infty} f^X(x) dx = 1$ . Daraus folgt **nicht**, dass immer  $f^X(x) \leq 1$  sein muss.

Die Wahrscheinlichkeit eines Ereignisses  $\{X \in A\}$  mit  $A \subseteq \mathbb{R}$  errechnet sich dann als

$$P(X \in A) = \int_A f^X(x) dx.$$

B

**Beispiel** Diskrete Dichte und Verteilungsfunktion

In manchen Fantasy-Spielen wird statt des üblichen sechsseitigen Würfels ein Würfel mit zwölf Seiten benutzt, der die Zahlen von 1 bis 12 als Ergebnis zeigen kann. Wirft man einen solchen Würfel einmal, so kann man die gewürfelte Augenzahl als Zufallsvariable  $X$  auffassen. Der Wertebereich von  $X$  ist dann  $\mathcal{X} = \{x_1, \dots, x_{12}\} = \{1, \dots, 12\}$  und  $P(X = x_i) = 1/12$  für  $i = 1, \dots, 12$ . Dabei gehen wir von einem so genannten fairen Würfel aus, der nicht zu Gunsten einer Zahl manipuliert wurde. Die diskrete Dichte von  $X$  ist damit gegeben als

$$f^X(x_i) = \frac{1}{12} \quad i = 1, \dots, 12.$$

Weiterhin lassen sich die Werte der Verteilungsfunktion bestimmen als

$x_i$	1	2	3	4	5	6	7	8	9	10	11	12
$f^X(x_i)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
$F^X(x_i)$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{10}{12}$	$\frac{11}{12}$	$\frac{12}{12}$

Damit kann man zum Beispiel die Wahrscheinlichkeit bestimmen, bei einem Wurf eine Zahl größer als 1, aber kleiner oder gleich 3 zu werfen

$$P(1 < X \leq 3) = P(X \in (1; 3]) = \sum_{x_i \in (1; 3]} = f^X(2) + f^X(3) = \frac{1}{12} + \frac{1}{12} = \frac{2}{12}$$

oder

$$P(1 < X \leq 3) = P(X \leq 3) - P(X \leq 1) = F^X(3) - F^X(1) = \frac{3}{12} - \frac{1}{12} = \frac{2}{12}.$$

◀B

**Beispiel** Stetige Dichte und Verteilungsfunktion

B

Gegeben sei eine stetige Zufallsvariable mit folgender Dichtefunktion

$$f^X(x) = \begin{cases} 1 & \text{für } 0,5 \leq x < 1 \\ 0,5 & \text{für } 0 \leq x < 0,5 \text{ oder } 1 \leq x \leq 1,5 \\ 0 & \text{sonst.} \end{cases}$$

Wollen wir überprüfen, ob es sich bei  $f$  tatsächlich um eine Dichtefunktion handelt, müssen wir dazu feststellen, ob  $f^X(x) \geq 0$  und ob  $\int_{-\infty}^{\infty} f^X(x) dx = 1$  gilt. Offensichtlich ist  $f^X(x) \geq 0$ , außerdem

$$\begin{aligned}
 \int_{-\infty}^{\infty} f^X(x) dx &= \int_0^{1,5} f^X(x) dx = \int_0^{0,5} 0,5 dx + \int_{0,5}^1 1 dx + \int_1^{1,5} 0,5 dx \\
 &= 0,5 \cdot x \Big|_0^{0,5} + 1 \cdot x \Big|_{0,5}^1 + 0,5 \cdot x \Big|_1^{1,5} \\
 &= (0,5 \cdot 0,5 - 0) + (1 \cdot 1 - 1 \cdot 0,5) + (0,5 \cdot 1,5 - 0,5 \cdot 1) \\
 &= 0,25 + 0,5 + 0,25 = 1.
 \end{aligned}$$

Damit handelt es sich um eine Dichtefunktion. Die Verteilungsfunktion  $F^X$  lässt sich damit herleiten als

$$F^X(x) = \int_{-\infty}^x f^X(t) dt = \begin{cases} 0 & \text{für } x < 0 \\ \int_0^x 0,5 dt & \text{für } 0 \leq x < 0,5 \\ \int_0^{0,5} 0,5 dt + \int_{0,5}^x 1 dt & \text{für } 0,5 \leq x < 1 \\ \int_0^{0,5} 0,5 dt + \int_{0,5}^1 1 dt + \int_1^x 0,5 dt & \text{für } 1 \leq x \leq 1,5 \\ 1 & \text{für } x > 1,5 \end{cases}$$

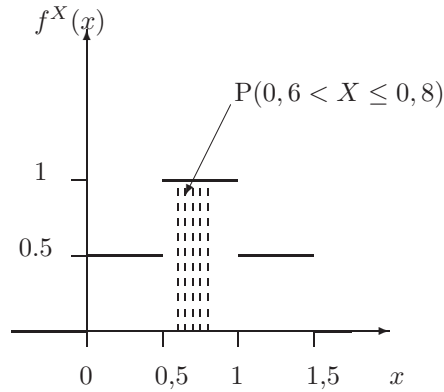
$$= \begin{cases} 0 & \text{für } x < 0 \\ \frac{x}{2} & \text{für } 0 \leq x < 0,5 \\ x - \frac{1}{4} & \text{für } 0,5 \leq x < 1 \\ \frac{1}{4} + \frac{x}{2} & \text{für } 1 \leq x \leq 1,5 \\ 1 & \text{für } x > 1,5. \end{cases}$$

Weiterhin ist zum Beispiel

$$P(0,6 < X \leq 0,8) = \int_{0,6}^{0,8} f^X(x) dx = \int_{0,6}^{0,8} 1 dx = 1 \cdot x \Big|_{0,6}^{0,8} = 0,8 - 0,6 = 0,2$$

oder

$$P(0,6 < X \leq 0,8) = F^X(0,8) - F^X(0,6) = 0,55 - 0,35 = 0,2.$$



Man beachte außerdem, dass aus der Verteilungsfunktion auf die Dichtefunktion rückgeschlossen werden kann. Dazu wird die Ableitung von  $F^X(x)$  bestimmt

$$\frac{\partial F^X(x)}{\partial x} = \begin{cases} 0 & \text{für } x < 0 \text{ oder } x > 1,5 \\ 0,5 & \text{für } 0 < x < 0,5 \text{ oder } 1 < x < 1,5 \\ 1 & \text{für } 0,5 < x < 1. \end{cases}$$

Die Ableitung existiert nicht an den Stellen  $x = 0; 0,5; 1; 1,5$ ; an diesen Stellen unterscheiden sich die linksseitigen Ableitungen von den rechtsseitigen. Davon abgesehen, stimmen die Ableitung von  $F^X$  und die Dichte  $f^X$  überein. Es gilt also, von den genannten vier Ausnahmen abgesehen, dass

$$\frac{\partial F^X(x)}{\partial x} = f^X(x).$$

### Rechenregeln für Dichtefunktionen und Verteilungsfunktionen

1. Die Verteilungsfunktion ist das Gegenstück zur empirischen Verteilungsfunktion ►e.
2. Für eine diskrete Zufallsvariable sieht die Verteilungsfunktion wie eine Treppenfunktion aus mit Sprüngen an den Stellen  $x_i$  und Sprunghöhen  $f^X(x_i) = P(X = x_i)$ .

3. Für eine diskrete Zufallsvariable  $X$  gilt

$$P(a < X \leq b) = \sum_{a < x_i \leq b} P(X = x_i)$$

und

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F^X(b) - F^X(a).$$

4. Für eine stetige Zufallsvariable  $X$  gilt:

- Der Wert der Verteilungsfunktion  $F^X$  an einer Stelle  $x$  entspricht der Fläche unter der Kurve der stetigen Dichtefunktion  $f^X$  bis zur Stelle  $x$ .
- $P(X = x) = 0$  für  $x \in \mathbb{R}$  fest. Das heißt, für eine stetige Zufallsvariable ist die Wahrscheinlichkeit, einen bestimmten Wert anzunehmen, gleich Null.
- Weiter ist

$$P(a < X \leq b) = \int_a^b f^X(x) dx$$

und

$$\begin{aligned} P(a < X < b) &= P(a \leq X < b) = P(a < X \leq b) \\ &= P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F^X(b) - F^X(a). \end{aligned}$$

- Die stetige Dichte  $f^X$  lässt sich als Ableitung der Verteilungsfunktion  $F^X$  schreiben

$$f^X(x) = \frac{\partial F^X(x)}{\partial x},$$

vorausgesetzt, die Ableitung existiert für fast alle  $x$ . Dabei ist es zulässig, dass die Ableitung für eine endliche Menge einzelner Werte  $x$  nicht existiert (vergleiche Beispiel ►15).

Betrachtet man nicht nur ein Merkmal alleine, sondern interessiert sich dafür, wie sich zwei Merkmale gemeinsam verhalten, so geht man über zur Betrachtung gemeinsamer Dichten und gemeinsamer Verteilungen.

---

**Definition** Gemeinsame Dichte für zwei Zufallsvariablen  $X$  und  $Y$ 


---

1. Für zwei diskrete Zufallsvariablen  $X$  und  $Y$  mit Verteilungsfunktionen  $F^X$  und  $F^Y$  schreibt man die gemeinsame Dichtefunktion als

$$f^{X;Y}(x_i; y_j) = P(X = x_i; Y = y_j) \quad i, j = 1, 2, \dots,$$

wobei

$$f^{X;Y}(x_i; y_j) \geq 0 \quad \text{und} \quad \sum_{(x_i; y_j)} f^{X;Y}(x_i; y_j) = 1 \quad \text{gilt.}$$

2. Für zwei stetige Zufallsvariablen  $X$  und  $Y$  schreibt man die gemeinsame Dichtefunktion als

$$f^{X;Y}(x; y), \quad x \in \mathbb{R}, y \in \mathbb{R},$$

wobei

$$f^{X;Y}(x; y) \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^{X;Y}(x; y) dx dy = 1 \quad \text{gilt.}$$

**Rechenregeln**

Für eine Teilmenge  $R \subseteq \mathbb{R}^2$  der  $xy$ -Ebene lässt sich die Wahrscheinlichkeit für  $\{(X; Y) \in R\}$  wie folgt berechnen.

1. Falls  $X$  und  $Y$  diskrete Zufallsvariablen sind, ist

$$P((X; Y) \in R) = \sum_{(x_i; y_j) \in R} f^{X;Y}(x_i; y_j).$$

2. Falls  $X$  und  $Y$  stetige Zufallsvariablen sind, ist

$$P((X; Y) \in R) = \int \int_R f^{X;Y}(x; y) dx dy.$$

Aus der gemeinsamen Dichte von zwei Merkmalen kann auf die beiden Dichten der einzelnen Merkmale rückgeschlossen werden. Beschäftigt man sich im

Zusammenhang der gemeinsamen Betrachtung zweier Zufallsvariablen mit den Dichten der beiden einzelnen Variablen, so spricht man auch von den **Randdichten**.

### Definition Randdichten

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f^{X;Y}$ . Die **Randdichten** von  $X$  und  $Y$  sind in der folgenden Weise definiert.

- Im diskreten Fall sind die Randdichten von  $X$  bzw.  $Y$  gegeben durch

$$f^X(x_i) = P(X = x_i) = \sum_{y_j} f^{X;Y}(x_i; y_j), \quad i = 1, 2, \dots,$$

$$f^Y(y_j) = P(Y = y_j) = \sum_{x_i} f^{X;Y}(x_i; y_j), \quad j = 1, 2, \dots$$

Es wird also über diejenige Variable summiert, deren Randdichte nicht von Interesse ist, das heißt für die Randdichte von  $X$  wird über alle  $y_j$  summiert und umgekehrt.

- Im stetigen Fall sind die Randdichten von  $X$  bzw.  $Y$  gegeben durch

$$f^X(x) = \int_{-\infty}^{\infty} f^{X;Y}(x; y) dy, \quad x \in \mathbb{R},$$

$$f^Y(y) = \int_{-\infty}^{\infty} f^{X;Y}(x; y) dx, \quad y \in \mathbb{R}.$$

Für stetige Zufallsvariablen muss zur Berechnung der jeweiligen Randdichte die entsprechende andere Variable herausintegriert werden.

### B

#### Beispiel Gemeinsame Dichte und Randdichten im diskreten Fall

Seien  $X$  und  $Y$  diskrete Zufallsvariablen. Ihre gemeinsame Dichtefunktion sei gegeben als

		$y$				$f^X(x)$
		0	1	2	4	
$x$	$f^{X,Y}(x,y)$	0, 1	0, 2	0	0, 3	0, 6
		0, 1	0, 1	0, 2	0	0, 4
$f^Y(y)$		0, 2	0, 3	0, 2	0, 3	1, 0

Aus der Tabelle werden die jeweiligen Randdichten von  $X$  und  $Y$  gut sichtbar. Gesucht sei die Wahrscheinlichkeit dafür, dass die Summe  $X + Y \leq 2$  ist

$$\begin{aligned} P(X + Y \leq 2) &= \sum_{(x_i; y_j), x_i + y_j \leq 2} f^{X;Y}(x_i; y_j) \\ &= f^{X;Y}(1; 0) + f^{X;Y}(1; 1) + f^{X;Y}(2; 0) \\ &= 0, 1 + 0, 2 + 0, 1 = 0, 4. \end{aligned}$$

Seien weiter die Randdichten von  $X$  an der Stelle  $x = 1$  und von  $Y$  an der Stelle  $y = 2$  zu bestimmen

$$\begin{aligned} f^X(1) &= \sum_{y_j} f^{X;Y}(1; y_j) = f^{X;Y}(1; 0) + f^{X;Y}(1; 1) \\ &\quad + f^{X;Y}(1; 2) + f^{X;Y}(1; 4) = 0, 6 \end{aligned}$$

$$f^Y(2) = \sum_{x_i} f^{X;Y}(x_i; 2) = f^{X;Y}(1; 2) + f^{X;Y}(2; 2) = 0, 2.$$




---

### Beispiel Gemeinsame Dichte und Randdichten im stetigen Fall

B

Seien  $X$  und  $Y$  stetige Zufallsvariablen. Ihre gemeinsame Dichtefunktion sei gegeben als

$$f^{X;Y}(x; y) = \exp\{-x\} \cdot \exp\{-y\}, \quad x > 0, y > 0.$$

Berechnen wir die Wahrscheinlichkeit, dass sich  $X$  auf dem Intervall  $(-\infty; 1]$  realisiert und  $Y$  auf dem Intervall  $[1; \infty)$

$$\begin{aligned} P(X \leq 1; Y \geq 1) &= \int_1^\infty \int_{-\infty}^1 f^{X;Y}(x; y) dx dy \\ &= \int_1^\infty \exp\{-y\} \cdot \left( \int_0^1 \exp\{-x\} dx \right) dy \\ &= \int_1^\infty \exp\{-y\} \cdot \left[ -\exp\{-x\} \right]_0^1 dy \end{aligned}$$

$$\begin{aligned}
&= (1 - \exp\{-1\}) \cdot \int_1^{\infty} \exp\{-y\} dy = - (1 - \exp\{-1\}) \cdot \exp\{-y\} \Big|_1^{\infty} \\
&= (1 - \exp\{-1\}) \cdot \exp\{-1\}.
\end{aligned}$$

Ebenso erhalten wir die Randdichte von  $X$  durch Herausintegrieren von  $y$

$$f^X(x) = \int_{-\infty}^{\infty} f^{X;Y}(x; y) dy = \begin{cases} 0 & \text{für } x \leq 0 \\ \exp\{-x\} & \text{für } x > 0, \end{cases}$$

da gilt

$$\int_{-\infty}^{\infty} f^{X;Y}(x; y) dy = \int_0^{\infty} \exp\{-x\} \cdot \exp\{-y\} dy = \exp\{-x\}.$$

Analog kann die Randdichte von  $Y$  hergeleitet werden

$$f^Y(y) = \int_{-\infty}^{\infty} f^{X;Y}(x; y) dx = \begin{cases} 0 & \text{für } y \leq 0 \\ \exp\{-y\} & \text{für } y > 0. \end{cases}$$



Ist schon bekannt, dass die Zufallsvariable  $Y$  einen bestimmten Wert angenommen hat, dann kann man sich dafür interessieren, wie die Wahrscheinlichkeitsverteilung von  $X$  unter dieser **Bedingung** aussieht.

### Definition Bedingte Dichte

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f^{X;Y}(x; y)$  und zugehörigen Randdichten  $f^X(x)$  und  $f^Y(y)$ . Die **bedingte Dichte** von  $X$  für gegebenes  $Y = y$  ist definiert als

$$f^{X|Y}(x|y) = \frac{f^{X;Y}(x; y)}{f^Y(y)} \quad \text{für } f^Y(y) \neq 0.$$

Für  $f^Y(y) = 0$  ist  $f^{X|Y}(x|y)$  nicht definiert.

Umgekehrt ist die **bedingte Dichte** von  $Y$  gegeben  $X = x$  definiert als

$$f^{Y|X}(y|x) = \frac{f^{X;Y}(x; y)}{f^X(x)} \quad \text{für } f^X(x) \neq 0.$$

Die obige Definition kann sowohl für diskrete als auch stetige Zufallsvariablen angewendet werden. Sind  $X$  und  $Y$  diskret, so entspricht die bedingte Dichte von  $X$  gegeben  $Y = y$  der Wahrscheinlichkeit, dass  $X$  den Wert  $x$  annimmt, wenn sich  $Y$  als  $y$  realisiert hat, also  $f^{X|Y}(x|y) = P(X = x | Y = y)$ .

### Beispiel Bedingte Dichte

B

Seien  $X$  und  $Y$  zwei stetige Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f^{X;Y}(x; y) = \begin{cases} 2 & \text{für } x > 0, y > 0, x + y < 1 \\ 0 & \text{sonst.} \end{cases}$$

Zur Bestimmung der bedingten Dichte  $f^{Y|X}(y|x)$  benötigen wir zunächst die Randdichte von  $X$ . Diese erhält man durch Herausintegrieren der Variable  $Y$  aus der gemeinsamen Dichtefunktion von  $X$  und  $Y$

$$f^X(x) = \int_{-\infty}^{\infty} f^{X;Y}(x; y) dy = \begin{cases} \int_0^{1-x} 2 dy = 2 \cdot (1 - x) & \text{für } 0 < x < 1 \\ 0 & \text{sonst.} \end{cases}$$

Für gegebenes  $0 < x < 1$  berechnet sich die bedingte Dichte nun zu

$$f^{Y|X}(y|x) = \frac{f^{X;Y}(x; y)}{f^X(x)} = \begin{cases} \frac{2}{2 \cdot (1-x)} = \frac{1}{1-x} & \text{für } y > 0, y < 1 - x \\ 0 & \text{sonst.} \end{cases}$$

Interessant ist, dass für festes  $x$  die bedingte Verteilung von  $Y$  eine Rechteckverteilung ►42 auf dem Intervall  $(0; 1 - x)$  ist. ◀B

### Definition Bedingte Verteilung

◀

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f^{X;Y}(x; y)$  und zugehörigen Randdichten  $f^X(x)$  und  $f^Y(y)$ . Die bedingte Verteilung von  $X$  für gegebenes  $Y = y$  ist,

- wenn  $X$  und  $Y$  diskret sind, definiert als

$$F^{X|Y}(x|y) = \sum_{x_i \leq x} f^{X|Y}(x_i|y).$$

- wenn  $X$  und  $Y$  stetig sind, definiert als

$$F^{X|Y}(x|y) = \int_{-\infty}^x f^{X|Y}(t|y) dt.$$

Die in der deskriptiven Statistik benutzten Kenngrößen für die Häufigkeitsverteilungen von Merkmalen finden ihre Gegenstücke in den entsprechenden Größen für Zufallsvariablen.

## 2.5 Erwartungswert und Varianz

Zur zusammenfassenden Beschreibung von Datensätzen werden in der deskriptiven Statistik unter Anderem Maße für die **Lage** und die **Streuung** von Daten berechnet. Üblich sind das **arithmetische Mittel** zur Charakterisierung der Lage und die **empirische Varianz** (Stichprobenvarianz) und **Standardabweichung** (Stichprobenstandardabweichung) zur Charakterisierung der Variabilität. Als Lage- und Streuungsmaße für Wahrscheinlichkeitsverteilungen dienen die entsprechenden theoretischen Konstrukte **Erwartungswert** und **Varianz** bzw. **Standardabweichung**.

### Definition Erwartungswert

Betrachtet wird eine Zufallsvariable  $X$  mit Dichtefunktion  $f^X$ .

1. Ist  $X$  diskrete Zufallsvariable, so ist der **Erwartungswert**  $E[X]$  von  $X$  das gewichtete Mittel

$$E[X] = \sum_{x_i} x_i \cdot f^X(x_i) = x_1 \cdot f^X(x_1) + x_2 \cdot f^X(x_2) + \dots$$

2. Ist  $X$  stetige Zufallsvariable, so ist der **Erwartungswert**  $E[X]$  von  $X$  definiert als

$$E[X] = \int_{-\infty}^{\infty} x \cdot f^X(x) dx.$$

### Beispiel (Fortsetzung) Diskrete Dichte

Für die diskrete Zufallsvariable aus Beispiel errechnet sich der Erwartungswert wie folgt

$$E[X] = \sum_{i=1}^{12} x_i \cdot f^X(x_i) = \frac{1}{12} \cdot (1 + 2 + \dots + 12) = \frac{78}{6} = 6,5.$$

**Beispiel** (Fortsetzung ►15) Stetige Dichte**B**

Für die stetige Zufallsvariable aus **Beispiel ►15** errechnet sich der Erwartungswert wie folgt

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{\infty} x \cdot f^X(x) dx = \int_0^{0,5} x \cdot 0,5 dx + \int_{0,5}^1 x \cdot 1 dx + \int_1^{1,5} x \cdot 0,5 dx \\
 &= \left. \frac{x^2}{2} \cdot 0,5 \right|_0^{0,5} + \left. \frac{x^2}{2} \cdot 1 \right|_{0,5}^1 + \left. \frac{x^2}{2} \cdot 0,5 \right|_1^{1,5} \\
 &= 0,0625 + 0,375 + 0,3125 = 0,75.
 \end{aligned}$$

 **B****Eigenschaften und Rechenregeln zum Erwartungswert**

- Der Erwartungswert existiert nicht immer. Es kann Dichten geben, so dass die Summe bzw. das Integral von  $x \cdot f^X(x)$  nicht endlich ist. In diesem Fall sagt man, dass  $E[X]$  nicht existiert.
- Der Erwartungswert ist das theoretische Gegenstück zum **arithmetischen Mittel ►46 ►e**. Man kann  $E[X]$  interpretieren als den „Schwerpunkt“ der Dichte, das heißt als die Stelle, an der man die Dichtefunktion unterstützen müsste, um sie im Gleichgewicht zu halten.
- Ist die Dichtefunktion  $f^X$  von  $X$  symmetrisch um eine Stelle  $a$ , das heißt  $f^X(a+x) = f^X(a-x)$  für alle  $x$ , dann ist  $E[X] = a$ .
- Transformiert man die Zufallsvariable  $X$  linear, das heißt man betrachtet  $Y = a \cdot X + b$  für Konstanten  $a, b$ , so gilt

$$E[Y] = E[a \cdot X + b] = a \cdot E[X] + b.$$

Dies ist die so genannte **Linearität des Erwartungswerts**.

- Transformiert man die Zufallsvariable  $X$  mit einer beliebigen Funktion  $g$ , das heißt man betrachtet  $Y = g(X)$ , so gilt

$$E[Y] = E[g(X)] = \sum_{x_i} g(x_i) \cdot f^X(x_i),$$

falls  $X$  eine diskrete Zufallsvariable, bzw.

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f^X(x) dx,$$

falls  $X$  eine stetige Zufallsvariable ist.

### Definition Varianz und Standardabweichung

Sei  $X$  eine Zufallsvariable mit Dichtefunktion  $f^X$ , und der Erwartungswert  $E[X]$  existiere. Die **Varianz** von  $X$  ist definiert durch

$$\text{Var}[X] = E[(X - E[X])^2].$$

Die Größe  $\text{Std}[X] = \sqrt{\text{Var}[X]}$  heißt **Standardabweichung** von  $X$ .

1. Ist  $X$  diskret, so rechnet man

$$\text{Var}[X] = \sum_{x_i} (x_i - E[X])^2 \cdot f^X(x_i).$$

2. Ist  $X$  stetig, so rechnet man

$$\text{Var}[X] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f^X(x) dx.$$

### Eigenschaften und Rechenregeln zur Varianz

- Die Varianz ist das theoretische Gegenstück zur **Stichprobenvarianz** ►e.
- Die Varianz kann alternativ über den **Verschiebungssatz** berechnet werden

$$\text{Var}[X] = \text{E}[X^2] - (\text{E}[X])^2,$$

wobei im diskreten Fall  $\text{E}[X^2] = \sum x_i^2 \cdot f^X(x_i)$ , im stetigen Fall  $\text{E}[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f^X(x) dx$  ist.

- Transformiert man die Zufallsvariable  $X$  linear, das heißt man betrachtet  $Y = a \cdot X + b$  für Konstanten  $a, b$ , so gilt

$$\text{Var}[Y] = \text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X]$$

und für die Standardabweichung

$$\text{Std}[Y] = |a| \cdot \text{Std}[X].$$

### Beispiel Varianz einer diskreten Zufallsvariable

B

Sei  $X$  eine diskrete Zufallsvariable mit Dichtefunktion

$$f^X(x) = \begin{cases} p & \text{für } x = 2 \\ \frac{1-p}{2} & \text{für } x \in \{1; 3\} \end{cases} \quad \text{für } p \in (0; 1).$$

Zu berechnen sei die Varianz. Dazu berechnen wir zunächst den Erwartungswert von  $X$

$$\begin{aligned} \text{E}[X] &= \sum_{x_i} x_i \cdot f^X(x_i) = 1 \cdot f^X(1) + 2 \cdot f^X(2) + 3 \cdot f^X(3) \\ &= 1 \cdot \frac{1-p}{2} + 2 \cdot p + 3 \cdot \frac{1-p}{2} = 2. \end{aligned}$$

Nun lässt sich die Varianz wie folgt berechnen

$$\begin{aligned} \text{Var}[X] &= \sum_{x_i} (x_i - \text{E}[X])^2 \cdot f^X(x_i) \\ &= (1-2)^2 \cdot f^X(1) + (2-2)^2 \cdot f^X(2) + (3-2)^2 \cdot f^X(3) \end{aligned}$$

$$= \frac{1-p}{2} + \frac{1-p}{2} = 1-p.$$

Die Berechnung der Varianz mit Hilfe des **Verschiebungssatzes** ►27 führt zum gleichen Ergebnis: Dazu berechnen wir zunächst  $E[X^2]$

$$E[X^2] = \sum_{x_i} x_i^2 \cdot f^X(x_i) = 1^2 \cdot f^X(1) + 2^2 \cdot f^X(2) + 3^2 \cdot f^X(3) = 5 - p.$$

Die Anwendung des Verschiebungssatzes ergibt dann

$$\text{Var}[X] = E[X^2] - (E[X])^2 = 5 - p - 4 = 1 - p.$$



Für zwei Merkmale  $X$  und  $Y$  gemeinsam können ebenfalls Erwartungswerte bestimmt werden.

### Rechenregeln für den Erwartungswert diskreter Zufallsvariablen

- Der Erwartungswert einer beliebigen Funktion  $g(X; Y)$  ist definiert als

$$E[g(X; Y)] = \sum_{(x_i; y_j)} g(x_i; y_j) \cdot f^{X;Y}(x_i; y_j).$$

- Insbesondere gilt, wenn  $g(x; y) = x \cdot y$

$$E[X \cdot Y] = \sum_{(x_i; y_j)} x_i \cdot y_j \cdot f^{X;Y}(x_i; y_j).$$

### Rechenregeln für den Erwartungswert stetiger Zufallsvariablen

- Für eine beliebige Funktion  $g(X; Y)$  von  $X$  und  $Y$  ist der Erwartungswert definiert als

$$E[g(X; Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x; y) \cdot f^{X;Y}(x; y) dx dy.$$

- Insbesondere gilt, wenn  $g(x; y) = x \cdot y$

$$E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f^{X;Y}(x; y) dx dy.$$

### Beispiel Erwartungswert von $X \cdot Y$ im diskreten Fall

B

Seien  $X$  und  $Y$  die diskreten Zufallsvariablen aus dem Beispiel ►20. Der Erwartungswert von  $(X \cdot Y)$  berechnet sich zu

$$\begin{aligned} E[X \cdot Y] &= \sum_{(x_i, y_j)} x_i \cdot y_j \cdot f^{X;Y}(x_i; y_j) \\ &= (1 \cdot 0) \cdot f^{X;Y}(1; 0) + (1 \cdot 1) \cdot f^{X;Y}(1; 1) + \dots \\ &\quad + (2 \cdot 4) \cdot f^{X;Y}(2; 4) = 2, 4. \end{aligned}$$

◀B

Die Definition der **bedingten Dichte** ►22 einer Zufallsvariablen  $X$  für gegebenes  $Y = y$  führt zum Konzept der so genannten **bedingten Erwartungswerte**. So wie der einfache Erwartungswert auf Basis der Dichte einer einzelnen Zufallsvariable definiert wird, basiert die Definition des bedingten Erwartungswerts auf der bedingten Dichte.

### Definition Bedingte Erwartungswerte

◀

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichte  $f^{X;Y}(x; y)$  und zugehörigen Randdichten  $f^X(x)$  und  $f^Y(y)$ . Für eine beliebige Funktion  $g$  ist der **bedingte Erwartungswert** von  $g(X; Y)$  gegeben  $Y = y$

- für zwei diskrete Zufallsvariablen  $X$  und  $Y$  definiert als

$$E[g(X; Y)|Y = y] = \sum_{x_i} g(x_i; y) \cdot f^{X|Y}(x_i|y),$$

- für zwei stetige Zufallsvariablen  $X$  und  $Y$  definiert als

$$E[g(X, Y)|Y = y] = \int_{-\infty}^{\infty} g(x; y) \cdot f^{X|Y}(x|y) dx.$$

Entsprechend sind die bedingten Erwartungswerte von  $Y$  gegeben  $X = x$  definiert über die bedingte Dichte von  $Y$  gegeben  $X = x$ .

Zum Verständnis der bedingten Erwartungswerte ist es hilfreich, nicht nur feste Realisationen  $y$  von  $Y$  als Bedingung anzunehmen, sondern die Bedingung selbst wieder als zufällig aufzufassen. Damit betrachtet man den bedingten Erwartungswert  $E[g(X; Y)|Y]$ , als Funktion von  $Y$ , selbst wieder als Zufallsvariable.

### Eigenschaften bedingter Erwartungswerte

- Für die speziellen Funktionen  $g_1(x; y) = x$  und  $g_2(x; y) = y$  sind  $E[X|Y = y]$  und  $E[Y|X = x]$  die so genannten **bedingten Erwartungswerte** von  $X$  für gegebenes  $Y = y$  bzw. von  $Y$  für gegebenes  $X = x$ .
- Der bedingte Erwartungswert  $E[g(X; Y)|Y]$  kann als Funktion in Abhängigkeit von  $Y$  aufgefasst werden.
- Es lässt sich zeigen, dass die Zufallsvariable  $E[X|Y]$  den Erwartungswert  $E[X]$  besitzt, das heißt es gilt

$$E[E[X|Y]] = E[X].$$

Entsprechend gilt  $E[E[Y|X]] = E[Y]$ .


## B

### Beispiel (Fortsetzung ►23) Bedingter Erwartungswert

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f^{X;Y}(x; y)$  und bedingter Dichte aus **Beispiel ►23**. Der bedingte Erwartungswert  $E[Y|X = x]$  für festes  $X = x$  und  $0 < x < 1$  errechnet sich dann wie folgt

$$E[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f^{Y|X}(y|x) dy$$

$$= \int_0^{1-x} \frac{y}{1-x} dy = \frac{y^2}{2 \cdot (1-x)} \Big|_0^{1-x} = \frac{1-x}{2}.$$

Fasst man nun den bedingten Erwartungswert  $E[Y|X]$  als Funktion von  $X$  auf, erhält man  $E[Y|X] = \frac{1-X}{2}$ , also wieder eine zufällige Größe. 

## 2.6 Abhängigkeit

2.6

Bei der gemeinsamen Betrachtung zweier Merkmale interessiert man sich häufig dafür, ob und gegebenenfalls wie stark die beiden Merkmale miteinander zusammenhängen. Dazu berechnet man in der deskriptiven Statistik **Zusammenhangsmaße** ►e wie **Kontingenz-** und **Korrelationskoeffizienten** ►e. Als zugrunde liegende theoretische Konzepte betrachten wir die **stochastische Unabhängigkeit** ►31, die **Kovarianz** und die **Korrelation** ►32.

Gilt, dass für festes  $y$  die bedingte Dichte von  $X$  der Randdichte von  $X$  entspricht, also  $f^{X|Y}(x|y) = f^X(x)$ , so sind  $X$  und  $Y$  voneinander **stochastisch unabhängig**. Das heißt, die Realisierung von  $Y$  hat keinen Einfluss auf die Realisierung von  $X$ . Dies ist äquivalent zur folgenden Definition der Unabhängigkeit.

### Definition Unabhängigkeit von Zufallsvariablen

Seien  $X$  und  $Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f^{X,Y}(x,y)$  und zugehörigen Randdichten  $f^X(x)$  und  $f^Y(y)$ . Dann sind  $X$  und  $Y$  **(stochastisch) unabhängig**, wenn

$$f^{X,Y}(x,y) = f^X(x) \cdot f^Y(y)$$

für alle  $x$  und  $y$  aus den Wertebereichen von  $X$  und  $Y$  gilt. Man beachte, dass hier die beiden Fälle diskreter und stetiger Zufallsvariablen abgedeckt sind.

### Rechenregeln für unabhängige Zufallsvariablen

Sind die Zufallsvariablen  $X$  und  $Y$  unabhängig, dann gilt für beliebige Funktionen  $g(X), h(Y)$

$$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)].$$

Da die Funktionen  $g$  und  $h$  auch der Identität entsprechen können, gilt insbesondere

$$E[X \cdot Y] = E[X] \cdot E[Y],$$

wenn  $X$  und  $Y$  unabhängig sind.

Die bedingte Dichtefunktion von  $X$  für gegebenes  $Y = y$  war definiert als

$$f^{X|Y}(x|y) = \frac{f^{X;Y}(x; y)}{f^Y(y)} \quad \text{für } f^Y(y) \neq 0.$$

Mit der Unabhängigkeit gilt dann

$$\begin{aligned} f^{X|Y}(x|y) &= \frac{f^{X;Y}(x; y)}{f^Y(y)} \quad \text{für } f^Y(y) \neq 0 \\ &= \frac{f^X(x) \cdot f^Y(y)}{f^Y(y)} = f^X(x). \end{aligned}$$

Daher sind die Formulierungen der Unabhängigkeit über die bedingten Dichten und über die gemeinsame Dichte äquivalent.

### Definition Kovarianz und Korrelation

Für zwei Zufallsvariablen  $X$  und  $Y$  ist die **Kovarianz** zwischen  $X$  und  $Y$  definiert als

$$\text{Cov}[X, Y] = E[(X - E[X]) \cdot (Y - E[Y])].$$

Der **Korrelationskoeffizient** (kurz: die **Korrelation**) zwischen  $X$  und  $Y$  ist gegeben als

$$\text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}}.$$

- Sind  $X$  und  $Y$  diskret, so lässt sich die Formel für die Kovarianz darstellen durch

$$\text{Cov}[X, Y] = \sum_{(x_i, y_j)} (x_i - E[X]) \cdot (y_j - E[Y]) \cdot f^{X;Y}(x_i; y_j).$$

- Für zwei stetige Zufallsvariablen  $X, Y$  ergibt sich

$$\text{Cov}[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X]) \cdot (y - E[Y]) \cdot f^{X;Y}(x; y) dx dy.$$

### Rechenregeln und Eigenschaften zu Kovarianz und Korrelation

- Die Korrelation ist das theoretische Gegenstück zum Korrelationskoeffizienten nach Bravais und Pearson ►e.
- Zur vereinfachten Berechnung der Kovarianz verwendet man den **Verschiebungssatz für die Kovarianz**

$$\text{Cov}[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y].$$

- Transformiert man  $X$  und  $Y$  linear in  $a \cdot X + b$  und  $c \cdot Y + d$  für konstante Werte  $a, b, c, d$ , so gilt

$$\text{Cov}[a \cdot X + b, c \cdot Y + d] = a \cdot c \cdot \text{Cov}[X, Y].$$

- Für zwei Zufallsvariablen  $X$  und  $Y$  gilt außerdem

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}[X, Y].$$

- Wenn  $X$  und  $Y$  stochastisch unabhängig sind, so gilt  $\text{Cov}[X, Y] = 0$ . Dies ist leicht einzusehen, denn

$$\begin{aligned} \text{Cov}[X, Y] &= E[X \cdot Y] - E[X] \cdot E[Y] \\ &= E[X] \cdot E[Y] - E[X] \cdot E[Y] = 0, \end{aligned}$$

da  $E[X \cdot Y] = E[X] \cdot E[Y]$  aus der Unabhängigkeit von  $X$  und  $Y$  gefolgert werden kann. Der Umkehrschluss ist **nicht** zulässig. Das heißt, aus  $\text{Cov}[X, Y] = 0$  folgt im Allgemeinen nicht die Unabhängigkeit der beiden Zufallsvariablen.

## Ergänzungen

Betrachtet man nicht nur zwei, sondern eventuell auch mehr als zwei Zufallsvariablen  $X_1, \dots, X_n$  gemeinsam, so gelten außerdem noch die folgenden Rechenregeln.

### Rechenregeln für mehr als zwei Zufallsvariablen

- $X_1, \dots, X_n$  sind stochastisch unabhängig, falls

$$f^{X_1; \dots; X_n}(x_1; \dots; x_n) = f^{X_1}(x_1) \cdot \dots \cdot f^{X_n}(x_n).$$

Dabei bezeichnet  $f^{X_1, \dots, X_n}$  die gemeinsame Dichte von  $X_1, \dots, X_n$  und  $f^{X_i}$  die Randdichte von  $X_i$ ,  $i = 1, \dots, n$ .

- Für Konstanten  $a_1, \dots, a_n$  gilt

$$\mathbb{E} \left[ \sum_{i=1}^n a_i \cdot X_i \right] = \sum_{i=1}^n a_i \cdot \mathbb{E}[X_i].$$

- Für Konstanten  $a_1, \dots, a_n, b_1, \dots, b_m$  gilt

$$\text{Cov} \left[ \sum_{i=1}^n a_i \cdot X_i, \sum_{j=1}^m b_j \cdot Y_j \right] = \sum_{i=1}^n \sum_{j=1}^m a_i \cdot b_j \cdot \text{Cov}[X_i, Y_j].$$

- Falls  $X_1, \dots, X_n$  stochastisch unabhängig, gilt für die Varianz

$$\text{Var} \left[ \sum_{i=1}^n a_i \cdot X_i \right] = \sum_{i=1}^n a_i^2 \cdot \text{Var}(X_i).$$

Über die **Verteilungsfunktion** ►12 wird ein Merkmal charakterisiert. Zur statistischen Beschreibung einer Stichprobe verwendet man die folgende modellhafte Idee. Man geht davon aus, dass jeder beobachtete Wert des Merkmals in der Stichprobe (der Merkmalswert jeder Erhebungseinheit) eine Realisation eines Grundmerkmals  $X$  ist. Um die Werte für die einzelnen Erhebungseinheiten voneinander zu unterscheiden, stellt man sich weiter vor, dass die  $i$ -te Untersuchungseinheit selbst das Merkmal  $X_i$  besitzt, das dieselben Charakteristika aufweist wie das Grundmerkmal  $X$ .

**Definition** Stichprobenvariablen

Ein interessierendes Merkmal lasse sich beschreiben durch eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F^X(x; \vartheta)$ . Eine Stichprobe  $x_1, \dots, x_n$  lässt sich dann auffassen als eine Realisierung von Zufallsvariablen  $X_1, \dots, X_n$ , die stochastisch unabhängig sind und alle dieselbe Verteilung wie  $X$  besitzen. Die Zufallsvariablen  $X_1, \dots, X_n$  nennt man **Stichprobenvariablen**.

Durch die Modellvorstellung, dass die Stichprobenvariablen unabhängig und identisch wie die Ausgangsvariable  $X$  verteilt sind, sichert man, dass die realisierte Stichprobe  $x_1, \dots, x_n$  für das interessierende Merkmal  $X$  in der Grundgesamtheit repräsentativ ist.

**Rechenregeln** für Stichprobenvariablen

Seien  $X_1, \dots, X_n$  unabhängige Stichprobenvariablen, die identisch verteilt sind wie eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F^X(x)$  und Dichtefunktion  $f^X(x)$ .

- Die gemeinsame Dichtefunktion von  $X_1, \dots, X_n$  ist

$$f^{X_1; \dots; X_n}(x_1; \dots; x_n) = \prod_{i=1}^n f^X(x_i).$$

- $E[X_i] = E[X]$ ,  $\text{Var}[X_i] = \text{Var}[X]$ ,  $i = 1, \dots, n$ , wenn Erwartungswert und Varianz von  $X$  existieren.

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

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \cdot \sum_{i=1}^n X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^n E[X_i] = E[X], \\ \text{Var}[\bar{X}] &= \text{Var}\left[\frac{1}{n} \cdot \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n} \cdot \text{Var}[X]. \end{aligned}$$

Größen, die häufig im Zusammenhang mit Stichprobenvariablen betrachtet werden, sind die so genannten **Ordnungsstatistiken**. Ordnungsstatistiken sind relevant beispielsweise bei der Bestimmung der Verteilung des **Minimums** und des **Maximums**.

### Definition Ordnungsstatistiken

Betrachten wir ein mindestens ordinal skaliertes Merkmal, das durch eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F^X$  und zugehöriger Dichtefunktion  $f^X$  beschrieben wird. Die Stichprobenvariablen  $X_1, \dots, X_n$  seien unabhängig und identisch wie  $X$  verteilt, wobei  $x_1, \dots, x_n$  eine realisierte Stichprobe vom Umfang  $n$  ist. Die Beobachtungen werden der Größe nach geordnet, beginnend mit der kleinsten

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}.$$

Dann können  $x_{(1)}, \dots, x_{(n)}$  als Realisationen von  $X_{(1)}, \dots, X_{(n)}$  aufgefasst werden. Diese Zufallsvariablen  $X_{(1)}, \dots, X_{(n)}$  heißen **Ordnungsstatistiken**.

### Regel Verteilung von Ordnungsstatistiken

Sei  $X$  eine stetige Zufallsvariable mit Verteilungsfunktion  $F^X$ . Seien weiter  $X_1, \dots, X_n$  unabhängige und wie  $X$  verteilte Stichprobenvariablen und  $X_{(1)}, \dots, X_{(n)}$  die entsprechenden Ordnungsstatistiken. Dann ist die Randverteilung der  $i$ -ten Ordnungsstatistik,  $i = 1, \dots, n$ , gegeben durch

$$F^{X_{(i)}}(x) = \sum_{j=i}^n \binom{n}{j} \cdot (F^X(x))^j \cdot (1 - F^X(x))^{n-j}, \quad x \in \mathbb{R}.$$

Setzen wir  $i = 1$ , so erhalten wir die Verteilung des Minimums, das der Ordnungsstatistik  $X_{(1)}$  entspricht.

### Regel Verteilung des Minimums

Die Verteilung des Minimums ist für  $x \in \mathbb{R}$  gegeben als

$$F^{X_{(1)}}(x) = \sum_{j=1}^n \binom{n}{j} \cdot (F^X(x))^j \cdot (1 - F^X(x))^{n-j} = 1 - (1 - F^X(x))^n.$$

Die Dichtefunktion des Minimums erhalten wir durch Ableiten der Verteilungsfunktion

$$f^{X_{(1)}}(x) = n \cdot (1 - F^X(x))^{n-1} \cdot f^X(x), \quad x \in \mathbb{R}.$$

Analog ergibt sich für  $i = n$  die Verteilung des Maximums  $X_{(n)}$ .

### Regel Verteilung des Maximums

Die Verteilung des Maximums ist für  $x \in \mathbb{R}$  gegeben als

$$F^{X_{(n)}}(x) = \sum_{j=n}^n \binom{n}{j} \cdot (F^X(x))^j \cdot (1 - F^X(x))^{n-j} = (F^X(x))^n.$$

Die Dichtefunktion

$$f^{X_{(n)}}(x) = n \cdot (F^X(x))^{n-1} \cdot f^X(x), \quad x \in \mathbb{R},$$

erhält man wieder durch Ableiten der Verteilungsfunktion.

## 2.7 Gängige Verteilungen und ihre Erwartungswerte und Varianzen

---

2.7

Einige Standardsituationen kommen bei statistischen Analysen immer wieder vor. Mit diesen Situationen verbunden sind Merkmale, die bestimmte Typen von Verteilungen besitzen. Im Folgenden stellen wir die gängigsten dieser Verteilungen vor, jeweils zusammen mit Dichtefunktion, Erwartungswert und Varianz der entsprechend verteilten Zufallsvariablen, sowie einigen grundlegenden Eigenschaften. Die hier vorgestellten Verteilungen werden in den folgenden Kapiteln benötigt. Darüber hinaus gibt es viele weitere Verteilungen, die hier nicht besprochen werden, wie zum Beispiel die **Negativ-Binomialverteilung**, die **Beta-Verteilung**, die **Cauchy-Verteilung**, die **logistische Verteilung** und andere ►e. Übersichten findet man beispielsweise in Evans et al. (2000).

### Diskrete Verteilungen

Eine faire Münze mit den beiden Seiten **Kopf** und **Zahl** wird  $n$ -mal voneinander unabhängig geworfen. Es wird jeweils notiert, welche Seite oben liegt. Das erhobene Merkmal  $X$  sei die Anzahl der Würfe, in denen **Kopf** oben gelegen hat. Dann ist für den einzelnen Wurf die Wahrscheinlichkeit, dass **Kopf** oben liegt, gleich  $1/2$  bei einer fairen Münze. Jeder einzelne Wurf stellt ein so genanntes **Bernoulli-Experiment** dar.

**Definition Bernoulli-Experiment**

Betrachtet wird ein einzelnes Zufallsexperiment mit den zwei möglichen Ausgängen Erfolg und Misserfolg. Dabei tritt mit Wahrscheinlichkeit  $p \in [0; 1]$  ein Erfolg ein,  $p$  heißt dementsprechend **Erfolgswahrscheinlichkeit**. Ein solches Zufallsexperiment heißt **Bernoulli-Experiment**.

**Definition Bernoulli-Verteilung**

Eine Zufallsvariable  $X$ , die den Wert 1 annimmt, falls ein interessierendes Ereignis eintritt, und den Wert 0, falls es nicht eintritt, und die eine Dichtefunktion  $f^X$  der Form

$$f^X(x) = p^x \cdot (1 - p)^{1-x} \quad \text{für } x = 0, 1$$

besitzt, heißt **bernoulliverteilt** mit **Parameter**  $p$ .

Schreibweise:  $X \sim \text{Bin}(1; p)$ .

**Eigenschaften**

- Der Parameter  $p$  ist definiert auf dem Intervall  $[0; 1]$ .
- Erwartungswert und Varianz einer bernoulliverteilten Zufallsvariablen sind

$$\mathbb{E}[X] = p, \quad \text{Var}[X] = p \cdot (1 - p).$$

Zur Darstellung der Binomialverteilung benötigen wir den Binomialkoeffizienten.

**Definition Binomialkoeffizient**

Der **Binomialkoeffizient** aus zwei natürlichen Zahlen  $m$  und  $k$  ist definiert als

$$\binom{m}{k} = \frac{m!}{k! \cdot (m - k)!}, \quad \text{falls } m \geq k.$$

Falls  $m < k$ , wird festgelegt, dass  $\binom{m}{k} = 0$  gilt.

Dabei ist die **Fakultät**  $k!$  einer natürlichen Zahl  $k$  definiert als

$$k! = 1 \cdot 2 \cdot \dots \cdot (k - 1) \cdot k,$$

wobei per Definition  $1! = 1$  und  $0! = 1$  gesetzt wird.

Sprechweisen:  $k! = k$  Fakultät,  $\binom{m}{k} = m$  über  $k$ .

### Definition Binomialverteilung

Eine diskrete Zufallsvariable  $X$ , die die Werte  $0, 1, \dots, n$  annehmen kann, mit Dichtefunktion

$$f^X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad \text{für } x = 0, 1, \dots, n$$

heißt **binomialverteilt** mit **Parametern**  $n$  und  $p$ .

Schreibweise:  $X \sim \text{Bin}(n; p)$ .

Die Binomialverteilung wird verwendet, wenn die Anzahl der Erfolge in  $n$  voneinander unabhängigen Bernoulli-Versuchen von Interesse ist. Dabei wird angenommen, dass die Erfolgswahrscheinlichkeit  $p$  in jedem der  $n$  Versuche gleich ist. Ein Beispiel ist eine klinische Studie, in der bei 100 Patienten der Heilungserfolg durch die Behandlung mit einem Medikament beobachtet wird. Erfolg tritt dabei ein, wenn ein Patient geheilt wird. Die Zufallsvariable  $X$  beschreibt die Anzahl der geheilten Patienten.

### Eigenschaften

- Der Parameter  $p$  ist definiert auf dem Intervall  $[0; 1]$ .
- Nimmt der Parameter  $p$  die Werte Null oder Eins an, also die Grenzen seines Definitionsbereiches, so degeneriert die Binomialverteilung zu einer so genannten **Einpunktverteilung** ►e, die einen Spezialfall der Binomialverteilung darstellt.
- Die Bernoulliverteilung ist ein Spezialfall der Binomialverteilung mit  $n = 1$ .
- Sind  $X_1, \dots, X_n$  stochastisch unabhängig und identisch bernoulliverteilt mit Parameter  $p$ , dann ist ihre Summe  $\sum_{i=1}^n X_i$  binomialverteilt mit Parametern  $n$  und  $p$ .
- Ist  $X \sim \text{Bin}(n; p)$ , dann ist

$$\mathbb{E}[X] = n \cdot p, \quad \text{Var}[X] = n \cdot p \cdot (1 - p).$$

### Definition Geometrische Verteilung

Eine diskrete Zufallsvariable  $X$ , die die Werte  $1, 2, \dots$  annehmen kann, mit Dichtefunktion

$$f^X(x) = p \cdot (1 - p)^{x-1} \quad \text{für } x \in \mathbb{N} = \{1, 2, \dots\}$$

heißt **geometrisch verteilt** mit **Parameter**  $p$ .

Schreibweise:  $X \sim \text{Geo}(p)$ .

Die geometrische Verteilung wird benutzt, wenn die Anzahl der Versuche bis zum Eintreten des ersten Erfolgs in einem Bernoulli-Experiment von Interesse ist. Ein Beispiel ist die Anzahl der Freiwürfe eines Spielers in einem Basketballspiel bis zum ersten Treffer. Wir nehmen dabei an, dass die Würfe voneinander unabhängig sind mit gleicher Trefferwahrscheinlichkeit  $p$ .

#### Eigenschaften

- Der Parameter  $p$  ist definiert auf dem Intervall  $(0; 1)$ .
- Ist  $X \sim \text{Geo}(p)$ , so gilt

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}[X] = \frac{1-p}{p^2}.$$

### Definition Hypergeometrische Verteilung

Eine diskrete Zufallsvariable  $X$ , die die Werte  $0, 1, \dots, n$  annehmen kann, mit Dichtefunktion

$$f^X(k) = \frac{\binom{r}{k} \cdot \binom{s-r}{n-k}}{\binom{s}{n}} \quad \text{für } k \in \{0, \dots, n\}$$

heißt **hypergeometrisch verteilt** mit **Parametern**  $s, r, n$ .

Schreibweise:  $X \sim \text{Hyp}(s, r, n)$ .

#### Eigenschaften

- Die Parameter  $s, r, n$  sind definiert auf  $\mathbb{N}$ , wobei  $r \leq s, n \leq s$  gelten muss. Die Werte der Dichtefunktion sind nur dann echt größer als Null, wenn  $k \in \{\max\{0, n + r - s\}, \dots, \min\{r, n\}\}$ .

- Ist  $X \sim \text{Hyp}(s, r, n)$ , so gilt

$$E[X] = n \cdot \frac{r}{s}, \quad \text{Var}[X] = \frac{n \cdot r \cdot (s - r) \cdot (s - n)}{s^2 \cdot (s - 1)}.$$

### Definition Poissonverteilung

Eine diskrete Zufallsvariable  $X$ , die Werte  $0, 1, 2, \dots$  annehmen kann, mit Dichtefunktion

$$f^X(x) = \frac{\lambda^x}{x!} \cdot \exp\{-\lambda\} \quad \text{für } x \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$$

heißt **poissonverteilt** mit **Parameter**  $\lambda$ .

Schreibweise:  $X \sim \text{Poi}(\lambda)$ .

Die Poissonverteilung ist bekannt als Verteilung der seltenen Ereignisse. Sie wird oft eingesetzt, wenn die Anzahl der innerhalb eines kleinen Zeitraums eintretenden Ereignisse gezählt wird. Dabei ist die Wahrscheinlichkeit, dass in einem kleinen Zeitraum ein solches Ereignis eintritt, typischerweise klein. Sei beispielsweise  $X$  die durchschnittliche Anzahl der Verkehrsunfälle pro Stunde an einer bestimmten Kreuzung. Die Wahrscheinlichkeit, dass innerhalb einer Stunde dort ein Unfall passiert, ist relativ gering. Die Anzahl der Verkehrsunfälle kann als poissonverteilt angenommen werden.

### Eigenschaften

- Der Parameter  $\lambda$  ist definiert auf dem Intervall  $(0; \infty)$ .
- Ist  $X \sim \text{Poi}(\lambda)$ , so gilt

$$E[X] = \lambda, \quad \text{Var}[X] = \lambda.$$

## Stetige Verteilungen

### Definition Rechteckverteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

$$f^X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

heißt **rechteckverteilt (gleichverteilt) auf dem Intervall**  $[a; b]$ .

Schreibweise:  $X \sim \mathcal{R}[a; b]$ .

### Eigenschaften

- Für die Parameter gilt  $a, b \in \mathbb{R}$  mit  $a < b$ .
- Ist  $X \sim \mathcal{R}[a; b]$ , dann gilt

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}[X] = \frac{(b-a)^2}{12}.$$

### Definition Normalverteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

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

heißt **normalverteilt** mit **Parametern**  $\mu$  und  $\sigma^2$ .

Schreibweise:  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

Die spezielle Normalverteilung  $\mathcal{N}(0, 1)$  mit Parametern  $\mu = 0$  und  $\sigma^2 = 1$  heißt **Standardnormalverteilung**. Ihre Verteilungsfunktion wird mit  $\Phi$  bezeichnet.

Die Normalverteilung ist eine der wichtigsten statistischen Verteilungen. Viele Verteilungen konvergieren in gewissem Sinn gegen die Normalverteilung, so dass bei großen Stichprobenumfängen häufig die Analyse so betrieben werden kann, als ob die Beobachtungen Realisationen normalverteilter Stichprobenvariablen wären.

### Eigenschaften

- Für die Parameter gelten folgende Definitionsbereiche:  $\mu \in \mathbb{R}$  und  $\sigma^2 \in \mathbb{R}^+$ .

- Ist  $X \sim \mathcal{N}(\mu, \sigma^2)$ , dann gilt

$$\mathbb{E}[X] = \mu, \quad \text{Var}[X] = \sigma^2.$$

- Eine normalverteilte Zufallsvariable  $X$  kann immer so **standardisiert** werden, dass ihre Transformation  $Z$  standardnormalverteilt ( $Z \sim \mathcal{N}(0, 1)$ ) ist. Ist  $X \sim \mathcal{N}(\mu, \sigma^2)$ , dann gilt

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1),$$

das heißt  $P(Z \leq z) = \Phi(z)$ .

- Ist  $X \sim \mathcal{N}(\mu, \sigma^2)$ , dann ist eine lineare Transformation  $Y$  von  $X$  wieder normalverteilt, und es gilt

$$Y = a \cdot X + b \sim \mathcal{N}(a \cdot \mu + b, a^2 \cdot \sigma^2).$$

- Sind  $X_1, \dots, X_n$  stochastisch unabhängig mit  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ , dann ist

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Im Spezialfall  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  für alle  $i$  ist dann

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Zur Darstellung der so genannten  **$\chi^2$ -Verteilung** wird die Gammafunktion benötigt.

### Definition Gammafunktion

Für beliebige Werte  $\alpha > 0$  ist die Gammafunktion an der Stelle  $\alpha$  definiert als

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \cdot \exp\{-x\} dx.$$

#### Eigenschaften

- $\Gamma(1) = 1$ .
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ .
- $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$  für  $\alpha > 0$ .
- $\Gamma(\alpha + 1) = \alpha!$  für  $\alpha \in \mathbb{N}$ .

### Definition $\chi^2$ -Verteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

$$f^X(x) = \frac{1}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot x^{(n/2)-1} \cdot \exp\{-x/2\} \quad \text{für } x > 0$$

heißt  **$\chi^2$ -verteilt** mit  $n$  **Freiheitsgraden**, sprich *chiquadrat*-verteilt.  
Schreibweise:  $X \sim \chi_n^2$ .

#### Eigenschaften

- Der Definitionsbereich von  $n$  ist die Menge der natürlichen Zahlen, also  $n \in \mathbb{N}$ .
- Für  $x \leq 0$  gilt  $f^X(x) = 0$ .
- Die  $\chi^2$ -Verteilung ist nicht symmetrisch.

- Ist  $X \sim \chi_n^2$ , so ist

$$\mathbb{E}[X] = n, \quad \text{Var}[X] = 2 \cdot n.$$

- Sind  $Z_1, \dots, Z_n$  stochastisch unabhängig mit  $Z_i \sim \mathcal{N}(0, 1)$ , dann ist

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

### Definition $t$ -Verteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

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

heißt **t-verteilt** mit  $n$  **Freiheitsgraden**.

Schreibweise:  $X \sim t_n$ .

### Eigenschaften

- Die  $t$ -Verteilung wird auch **Student-t-Verteilung** genannt.
- Der Definitionsbereich von  $n$  ist die Menge der natürlichen Zahlen, also  $n \in \mathbb{N}$ .
- Die Verteilung ist symmetrisch um Null.
- Für ein beliebiges **p-Quantil** ►13 von  $t_n$  gilt aufgrund der Symmetrie

$$t_{n;p} = -t_{n;1-p}.$$

- Ist  $X \sim t_n$ , dann gilt

$$\mathbb{E}[X] = 0 \quad \text{für } n > 1, \quad \text{Var}[X] = \frac{n}{n-2} \quad \text{für } n > 2.$$

- Für große Werte von  $n$  nähert sich die  $t_n$ -Verteilung der  $\mathcal{N}(0, 1)$ -Verteilung. Als Faustregel für eine gute Approximation gilt  $n \geq 30$ .

- Ist  $Z \sim \mathcal{N}(0, 1)$ ,  $V \sim \chi_n^2$ , und sind  $Z$  und  $V$  stochastisch unabhängig, dann ist

$$\frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n.$$

- Sind  $X_1, \dots, X_n$  unabhängig und identisch  $\mathcal{N}(\mu, \sigma^2)$ -verteilt, so ist

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \sim t_{n-1},$$

wobei  $\bar{X}$  das arithmetische Mittel und  $S$  die Stichprobenstandardabweichung von  $X_1, \dots, X_n$  ist. Beide Größen werden hier als Zufallsvariablen aufgefasst, definiert als

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i \quad \text{und} \quad S = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2}.$$

### Definition *F*-Verteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

$$f^X(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \cdot \Gamma\left(\frac{m}{2}\right)} \cdot \frac{n^{n/2} \cdot m^{m/2} \cdot x^{(n/2)-1}}{(m + n \cdot x)^{(n+m)/2}} \quad \text{für } x > 0$$

heißt ***F*-verteilt mit  $n$  und  $m$  Freiheitsgraden**.

Schreibweise:  $X \sim F_{n,m}$ .

### Eigenschaften

- Der Definitionsbereich der Freiheitsgrade  $n$  und  $m$  ist die Menge der natürlichen Zahlen,  $n, m \in \mathbb{N}$  mit  $m > 2$ .
- Für  $x \leq 0$  gilt  $f^X(x) = 0$ .
- Die *F*-Verteilung ist nicht symmetrisch.
- Ist  $X \sim F_{n,m}$ , so ist

$$\mathbb{E}[X] = \frac{m}{m-2}, \quad m > 2, \quad \text{Var}[X] = \frac{2 \cdot m^2 \cdot (n+m-2)}{n \cdot (m-2)^2 \cdot (m-4)}, \quad m > 4.$$

- Ist  $X \sim F_{n,m}$ , so ist  $\frac{1}{X} \sim F_{m,n}$ .
- Ist  $V_1 \sim \chi_n^2$ ,  $V_2 \sim \chi_m^2$ , und sind  $V_1$  und  $V_2$  stochastisch unabhängig, dann ist

$$\frac{V_1/n}{V_2/m} \sim F_{n,m}.$$

---

### Definition Exponentialverteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

$$f^X(x) = \lambda \cdot \exp\{-\lambda \cdot x\} \quad \text{für } x > 0$$

heißt **exponentialverteilt** mit **Parameter**  $\lambda$ .

Schreibweise:  $X \sim \text{Exp}(\lambda)$ .

### Eigenschaften

- Für den Parameter  $\lambda$  gilt  $\lambda > 0$ .
- Für  $x \leq 0$  gilt  $f^X(x) = 0$ .
- Die Exponentialverteilung ist nicht symmetrisch.
- Ist  $X$  exponentialverteilt mit Parameter  $\lambda$ , so ist

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}.$$

---

### Definition Gammaverteilung

Eine stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  und Dichtefunktion

$$f^X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot \exp\{-\lambda \cdot x\} \quad \text{für } x > 0$$

heißt **gammaverteilt** mit **Parametern**  $\lambda$  und  $\alpha$ .

Schreibweise:  $X \sim \Gamma(\lambda, \alpha)$ .

**Eigenschaften**

- Für die Parameter  $\lambda$  und  $\alpha$  gilt  $\lambda, \alpha > 0$ .
- Für  $x \leq 0$  gilt  $f^X(x) = 0$ .
- Die Gammaverteilung ist nicht symmetrisch.
- Ist  $X$  gammaverteilt mit Parametern  $\lambda$  und  $\alpha$ , so ist

$$E[X] = \frac{\alpha}{\lambda}, \quad \text{Var}[X] = \frac{\alpha}{\lambda^2}.$$

- Sind  $X_1, \dots, X_n$  unabhängig und identisch gammaverteilt mit Parametern  $\lambda$  und  $\alpha$ , so ist die Summe der  $X_i, i = 1, \dots, n$ , ebenfalls gammaverteilt, und zwar mit Parametern  $\lambda$  und  $\alpha \cdot n$

$$\sum_{i=1}^n X_i \sim \Gamma(\lambda, \alpha \cdot n).$$

- Die  $\chi^2$ -Verteilung ist ein Spezialfall der Gammaverteilung. Ist  $X \sim \chi_n^2$ , so ist  $X$  zugleich gammaverteilt mit Parametern  $\lambda = 1/2$  und  $\alpha = n/2$ .
- Die Exponentialverteilung ist ebenfalls ein Spezialfall der Gammaverteilung. Ist  $X \sim \text{Exp}(\lambda)$ , so ist  $X$  zugleich gammaverteilt mit Parametern  $\lambda$  und  $\alpha = 1$ .
- Sind  $X_1, \dots, X_n$  unabhängig und identisch exponentialverteilt mit Parameter  $\lambda$ , so ist die Summe der  $X_i, i = 1, \dots, n$ , gammaverteilt mit Parametern  $\lambda$  und  $n$

$$\sum_{i=1}^n X_i \sim \Gamma(\lambda, n).$$

Schließende Statistik

Grundlegende Methoden

Genschel, U.; Becker, C.

2005, XI, 352 S., Softcover

ISBN: 978-3-540-21838-8