

## 5. ERWARTUNGSWERT E UND VARIANZ V

### (5.0) Überblick

Im diskreten Fall haben wir die folgenden *wahrscheinlichkeitstheoretischen Masszahlen* der Zufallsgrösse  $X$ :

$$\text{Erwartungswert} \quad \mu = E(X) = \sum_i x_i p_i ,$$

$$\text{Varianz} \quad \sigma^2 = V(X) = \sum_i (x_i - \mu)^2 p_i .$$

$$\text{Standardabweichung (diskret und stetig)} \quad \sigma = \sqrt{V(X)}$$

Im stetigen Fall sind die wahrscheinlichkeitstheoretischen Masszahlen *Erwartungswert* und *Varianz* gegeben durch

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

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

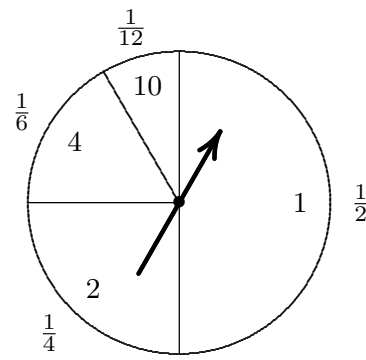
Der Erwartungswert ist ein Lagemass. Führt man das Zufallsexperiment sehr oft durch, so wird man erwarten, dass der Durchschnitt der von  $X$  angenommenen Werte ungefähr  $= \mu$  ist. Die Varianz und die Standardabweichung sind Streuungsmasse.

Zum Schluss werden noch Erweiterungen obiger Konzepte eingeführt: auch von Summen von Zufallsgrössen kann man Erwartungswerte und Varianzen berechnen:  $E[aX + bY] = E[aX] + E[bY]$  und (bei Unabhängigkeit der Summanden):  $V[aX + bY] = a^2 V[X] + b^2 V[Y]$ .

(5.1) Der Erwartungswert einer diskreten Zufallsgrösse

Beispiel 5.1.A

Wir betrachten ein Glücksrad mit Sektoren von  $180^\circ$ ,  $90^\circ$ ,  $60^\circ$  und  $30^\circ$ . Die Angaben in den Feldern stellen die jeweiligen Gewinne dar. Dadurch wird eine Zufallsgrösse  $X$  beschrieben, deren Verteilung durch die folgende Tabelle gegeben ist:



$x_i$	1	2	4	10
$p_i = P(X = x_i)$	1/2	1/4	1/6	1/12

Wir stellen uns nun die folgende Frage: Wie gross ist der durchschnittliche Gewinn bei diesem Glücksspiel, oder — anders betrachtet — wie gross muss der Spieleinsatz sein, den der Organisator des Spiels verlangen muss, damit er keinen Verlust erleidet?

Angenommen, ich spiele 12-mal. Dann kann ich ungefähr mit folgenden Gewinnen rechnen:

6 mal	1.-	:	6.-
3 mal	2.-	:	6.-
2 mal	4.-	:	8.-
1 mal	10.-	:	10.-
total		:	30.-

Der durchschnittliche Gewinn bzw. der Mindesteinsatz beträgt also  $30.- : 12 = 2.50$ .

Natürlich wird man bei nur 12-maligem Spiel kaum genau diese Gewinnverteilung erwarten. Spielt man aber sehr oft, so rechnet man aufgrund der intuitiven Interpretation der Wahrscheinlichkeit damit, dass der durchschnittliche Gewinn sehr nahe und im Grenzfall (wenn die Zahl der Spiele gegen unendlich strebt) genau bei Fr. 2.50 liegt.

☒

Diesen (hypothetischen) durchschnittlichen Gewinn nennt man den *Erwartungswert* der Zufallsgrösse  $X$ .

Wir führen nun diese Überlegung mit allgemeinen Grössen nochmals durch. Die Verteilung der diskreten Zufallsgrösse  $X$  sei gegeben durch

	$x_1$	$x_2$	$x_3$	$\dots$
	$p_1$	$p_2$	$p_3$	$\dots$

Spielen wir  $n$ -mal, so werden wir  $np_1$ -mal das Resultat  $x_1$  erwarten,  $np_2$ -mal das Resultat  $x_2$  usw. Für den Gesamtgewinn erhalten wir

$$np_1x_1 + np_2x_2 + np_3x_3 + \dots$$

Der Durchschnittsgewinn ergibt sich durch Division durch  $n$  zu

$$p_1x_1 + p_2x_2 + p_3x_3 + \dots$$

Da diese Summe eine anschauliche Bedeutung hat (z.B. Durchschnittsgewinn) ist die folgende Definition sinnvoll:

Die diskrete Zufallsgrösse  $X$  nehme die Werte  $x_i$  mit der Wahrscheinlichkeit  $p_i = P(X = x_i)$  an. Dann heisst die Zahl

$$E(X) = p_1x_1 + p_2x_2 + p_3x_3 + \dots = \sum_i p_ix_i = \sum_i P(X = x_i)x_i$$

der *Erwartungswert* von  $X$ . Statt  $E(X)$  schreibt man oft auch  $\mu$ .

### Bemerkungen

- 1) Je nachdem, ob  $X$  endlich oder abzählbar unendlich viele Werte annimmt, handelt es sich bei der Formel für  $E(X)$  um eine endliche Summe oder um eine unendliche Reihe.

Es kann vorkommen, dass diese Reihe divergiert. Dann hat  $X$  keinen Erwartungswert. Aus gewissen theoretischen Gründen fordert man sogar, dass die Reihe  $\sum_i p_i|x_i|$  konvergiert, was uns aber nicht weiter berührt.

- 2) Der Erwartungswert  $E(X)$  einer Zufallsgrösse ist eng verwandt mit dem Durchschnitt (vgl. (2.2.3.3)). Wir nehmen an, wir hätten  $n$  Messungen eines quantitativen Merkmals durchgeführt. Dabei sei  $H_1$ -mal der Wert  $x_1$ ,  $H_2$ -mal der Wert  $x_2$ , ...,  $H_k$ -mal der Wert  $x_k$  herausgekommen (mit  $H_1 + H_2 + \dots + H_k = n$ ). In (2.2.3.3.d) haben wir unter diesen Voraussetzungen die folgende Formel für den Durchschnitt  $\bar{x}$  angegeben:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k H_i x_i = \sum_{i=1}^k \frac{H_i}{n} x_i.$$

Nun ist aber  $H_i/n$  gerade die relative Häufigkeit des Auftretens von  $x_i$ . Für  $\bar{x}$  gilt also dieselbe Formel wie für  $E(X)$ , abgesehen davon, dass die Wahrscheinlichkeiten  $p_i$  durch die relativen Häufigkeiten  $H_i/n$  zu ersetzen sind.

- 3) In (2.2.3.2) und (2.2.3.3) haben wir den Durchschnitt  $\bar{x}$  als statistische Masszahl, genauer als Lagemass, interpretiert. Er beschreibt — ganz grob natürlich — die Lage der Messwerte.

Genauso können wir  $\mu = E(X)$  als *wahrscheinlichkeitstheoretische Masszahl* betrachten. Der Erwartungswert beschreibt — ebenfalls ganz grob — die “durchschnittliche Lage” der Zufallsgrösse  $X$ .

Dieser Vergleich von  $\bar{x}$  und  $\mu = E(X)$  zusammen mit der einleitend dargestellten Auffassung von  $E(X)$  als durchschnittlicher Gewinn, sollte genügen, um der abstrakt definierten Zahl  $E(X)$  einen eingängigen anschaulichen Sinn zu verleihen.

- 4) Wie bereits erwähnt, wird neben  $E(X)$  auch die Bezeichnung  $\mu$  verwendet. Diese ist etwas kürzer, doch geht der explizite Bezug zu  $X$  verloren. Wenn mehrere Zufallsgrössen, zum Beispiel  $X$  und  $Y$  involviert sind, kann man praktisch mit  $\mu_X$  und  $\mu_Y$  den Bezug herstellen.
- 5) In den Bemerkungen 6) und 7) von (4.1.3) ist erwähnt worden, dass man aus alten Zufallsgrössen neue bilden kann. Man kann zeigen (vgl. Anhang (13.2)), dass dann für den Erwartungswert von  $Y = aX + b$  eine einfache Formel gilt, nämlich

$$E(Y) = E(aX + b) = aE(X) + b .$$

Man beschreibt diese Situation in Worten dadurch, dass man sagt, der Erwartungswert sei *linear*. Dieses Resultat gilt für diskrete und stetige Zufallsgrössen.

### Zwei einfache Beispiele

- Wenn die Zufallsgrösse  $X$  die Anzahl der Exemplare einer bestimmten Pflanzenart pro Parzelle beschreibt (vgl. Beispiel 4.1.2.A), dann ist  $E(X)$  anschaulich die mittlere Anzahl von Pflanzen pro Parzelle, genommen über alle denkbaren (nicht nur die untersuchten) Parzellen.
- Entsprechend muss man sich im Beispiel 4.1.2.B ( $X =$  Körperlänge) unter  $E(X)$  die mittlere Körperlänge aller Zürcher(innen) vorstellen.  $\boxtimes$

Es folgen noch drei Beispiele, in denen es etwas mehr zu rechnen gibt.

### Beispiel 5.1.B

Wir greifen auf Beispiel 4.1.6.A zurück (Würfeln mit drei Würfeln). Die Zufallsgrösse  $X =$  “Summe der Augenzahlen” kann 16 Werte annehmen, nämlich die Zahlen 3 bis 18. Wir berechnen nun den Erwartungswert unter Verwendung der in 4.1.6.A angegebenen Wahrscheinlichkeiten und klammern dabei  $1/216$  gleich vor:

$$\begin{aligned} E(X) &= \sum_{i=1}^{16} p_i x_i = \frac{1}{216} (1 \cdot 3 + 3 \cdot 4 + 6 \cdot 5 + 10 \cdot 6 + 15 \cdot 7 + 21 \cdot 8 + 25 \cdot 9 \\ &+ 27 \cdot 10 + 27 \cdot 11 + 25 \cdot 12 + 21 \cdot 13 + 15 \cdot 14 + 10 \cdot 15 + 6 \cdot 16 + 3 \cdot 17 + 1 \cdot 18) \\ &= \frac{1}{216} \cdot 2268 = 10.5 . \end{aligned}$$

Würfelt man also sehr oft mit drei Würfeln, so wird man im Durchschnitt 10.5 Punkte erzielen.

Dies leuchtet auch direkt ein, denn 10.5 ist der Durchschnitt der minimalen (3) und der maximalen (18) Augenzahl. Ein Blick auf die Tabelle in 4.1.6.A zeigt, dass die Verteilung symmetrisch ist: 3 ist gleich wahrscheinlich wie 18, 4 wie 17 etc.  $\square$

### Beispiel 5.1.C

Ein Gerät besteht aus drei Komponenten, die in einem bestimmten Zeitraum unabhängig voneinander mit den Wahrscheinlichkeiten 0.1, 0.2 und 0.3 ausfallen. Wie gross ist die zu erwartende Anzahl der ausfallenden Komponenten?

Die Zufallsgrösse  $X = \text{“Anzahl der ausfallenden Komponenten”}$  kann die Werte 0, 1, 2, 3 annehmen. Gesucht ist ihr Erwartungswert. Wir berechnen zuerst die zugehörigen Wahrscheinlichkeiten (wer will, kann einen Baum zeichnen):

$$P(X = 0) = 0.9 \cdot 0.8 \cdot 0.7 = 0.504 \text{ (Gegenwahrscheinlichkeiten!).}$$

$$P(X = 1) = 0.1 \cdot 0.8 \cdot 0.7 + 0.9 \cdot 0.2 \cdot 0.7 + 0.9 \cdot 0.8 \cdot 0.3 = 0.398$$

(denn entweder fällt bloss die erste oder bloss die zweite oder bloss die dritte Komponente aus.) Analog

$$P(X = 2) = 0.9 \cdot 0.2 \cdot 0.3 + 0.1 \cdot 0.8 \cdot 0.3 + 0.1 \cdot 0.2 \cdot 0.7 = 0.092,$$

$$P(X = 3) = 0.1 \cdot 0.2 \cdot 0.3 = 0.006.$$

Wir erhalten

$x$	0	1	2	3
$P(X = x)$	0.504	0.398	0.092	0.006

Es folgt  $E(X) = 0 \cdot 0.504 + 1 \cdot 0.398 + 2 \cdot 0.092 + 3 \cdot 0.006 = 0.6$ .  $\square$

### Beispiel 5.1.D

Beim Schieber werden die 36 Jasskarten auf vier Personen verteilt. Wieviele Under (andernorts “Bube” genannt) erhalte ich im Durchschnitt? Gesucht ist — genauer formuliert — der Erwartungswert der Zufallsgrösse  $X = \text{“Anzahl Under”}$ . Wiederum ist die Berechnung der Wahrscheinlichkeiten die Hauptaufgabe.

Für mein “Blatt” im Kartenspiel gibt es  $\binom{36}{9}$  Möglichkeiten (vgl. (1.5)), denn es handelt sich um die Auswahl von 9 Karten aus 36, ohne Berücksichtigung der Reihenfolge. In wievielen dieser Möglichkeiten habe ich z.B. zwei Under? Die Überlegung ist dieselbe wie beim Zahlenlotto (Beispiel 3.4.4.A, vgl. auch (5.4)): Ich erhalte von den 4 Undern 2 Karten ( $\binom{4}{2}$  Möglichkeiten) und aus den 32 “Nicht-Undern” 7 weitere Karten ( $\binom{32}{7}$  Möglichkeiten). Somit ist (mit gerundeten Werten)

$$P(U = 2) = \frac{\binom{4}{2} \cdot \binom{32}{7}}{\binom{36}{9}} = 0.215 .$$

Analog\*:

$$P(U = 0) = \frac{\binom{4}{0} \cdot \binom{32}{9}}{\binom{36}{9}} = 0.298,$$

$$P(U = 1) = \frac{\binom{4}{1} \cdot \binom{32}{8}}{\binom{36}{9}} = 0.447,$$

$$P(U = 3) = \frac{\binom{4}{3} \cdot \binom{32}{6}}{\binom{36}{9}} = 0.039,$$

$$P(U = 4) = \frac{\binom{4}{4} \cdot \binom{32}{5}}{\binom{36}{9}} = 0.002.$$

Aus der Tabelle\*\*

$k$	0	1	2	3	4
$P(U = k)$	0.298	0.447	0.215	0.039	0.002

liest man ab:

$$E(U) = 1 \cdot 0.447 + 2 \cdot 0.215 + 3 \cdot 0.039 + 4 \cdot 0.002 = 1.002.$$

Da dieser Wert nahe bei 1 liegt, denkt man an Rundungsfehler. In der Tat liefert eine exakte Berechnung (unter Verwendung der Binomialkoeffizienten) den Wert

$$E(U) = 1,$$

der auch anschaulich sofort einleuchtet: Da die vier Under auf vier Personen verteilt werden, wird man im Mittel pro Spiel einen Under erhalten.  $\square$

(5.2) Die Varianz einer diskreten Zufallsgrösse

In der beschreibenden Statistik haben wir den Durchschnitt  $\bar{x}$  als Lagemass und die Varianz  $s^2$  als Streuungsmass kennen gelernt (2.2.3.7).

Für wahrscheinlichkeitstheoretische Verteilungen tritt, wie wir in (5.1) gesehen haben, der Erwartungswert  $\mu = E(X)$  an die Stelle des Durchschnitts  $\bar{x}$ . Ganz entsprechend tritt die nun zu definierende wahrscheinlichkeitstheoretische Varianz  $\sigma^2 = V(X)$  an die Stelle der statistischen (oder empirischen) Varianz  $s^2$ . Die Zahl  $V(X)$  dient also als Mass für die Streuung der Verteilung einer Zufallsgrösse.

Genau wie in (2.2.3.7) definiert man die Standardabweichung als positive Wurzel aus der Varianz.

$s$  : Statistische (oder empirische) Standardabweichung.

---

\* Der Wert für 4 Under wurde bereits in 3.4.4.D bestimmt.

\*\* Aus Rundungsgründen ist die Summe der Wahrscheinlichkeiten  $\neq 1$ .

$\sigma$  : Wahrscheinlichkeitstheoretische Standardabweichung.

Da  $\mu = E(X)$  das Pendant zu  $\bar{x}$  ist, wird man zur Bildung von  $\sigma^2$  vernünftigerweise die Quadrate der Abweichungen vom Erwartungswert, also

$$(x_i - \mu)^2 = (x_i - E(X))^2$$

verwenden. Diese Grössen wird man nun aber nicht einfach addieren, sondern man wird die  $x_i$ , welche mit einer grossen Wahrscheinlichkeit auftreten, stärker gewichten wollen, als jene, die nur mit einer geringen Wahrscheinlichkeit vorkommen. Etwas anders formuliert: Ein Wert  $x_i$  von  $X$ , der mit einer kleinen Wahrscheinlichkeit angenommen wird, soll nur wenig zur Varianz beitragen, auch wenn seine Abweichung von  $E(X)$  gross ist. Aus diesem Grund multipliziert man die Quadrate der Abweichungen noch mit  $p_i$ , bevor man sie addiert. Man spricht in diesem Zusammenhang auch etwa von einer "gewichteten Summe".

Die eben angestellten Überlegungen zeigen, dass die folgende Definition sinnvoll ist:

Die diskrete Zufallsgrösse  $X$  nehme die Werte  $x_i$  mit der Wahrscheinlichkeit  $p_i$  an. Ferner sei  $\mu$  der Erwartungswert von  $X$ . Dann heisst die Zahl

$$\sigma^2 = V(X) = p_1(x_1 - \mu)^2 + p_2(x_2 - \mu)^2 + \dots = \sum_i p_i(x_i - \mu)^2 = \sum_i P(X = x_i)(x_i - \mu)^2$$

die (wahrscheinlichkeitstheoretische) *Varianz* von  $X$ .

Die (positive) Wurzel  $\sigma$  aus  $V(X)$  heisst die (wahrscheinlichkeitstheoretische) *Standardabweichung* von  $X$ .

### Bemerkungen

- 1) Wenn  $X$  abzählbar unendlich viele Werte annimmt, dann steht in der Definition eine unendliche Reihe. Es ist möglich, dass diese divergiert. In diesem Fall existiert  $V(X)$  nicht. Es kann sogar noch schlimmer sein: Wenn schon  $E(X)$  nicht existieren sollte (vgl. Bemerkung 1) in (5.1)), dann existiert  $V(X)$  erst recht nicht.
- 2) In manchen Büchern findet man auch die Bezeichnung  $D^2(X)$  für die Varianz. Dies hat den Vorteil, dass man für die Standardabweichung  $D(X)$  schreiben kann.
- 3) Ähnlich wie für den Erwartungswert von  $Y = aX + b$  (vgl. Bemerkung 5) von (5.1)) gibt es eine Formel für die Varianz von  $Y$ , deren Beweis Sie ebenfalls im Anhang (13.2) finden:

$$V(Y) = V(aX + b) = a^2V(X) .$$

Die Varianz ist also *nicht* linear, was nicht verwundert, da in der Definition ein Quadrat vorkommt. Dieses Resultat gilt für diskrete und stetige Zufallsgrössen.

Beispiel 5.2.A (vgl. Beispiel 5.1.A)

Es geht hier um ein Glücksrad, das durch die folgende Verteilung der Zufallsgrösse  $X$  beschrieben wird:

	1	2	4	10
	1/2	1/4	1/6	1/12

Wir wissen schon, dass  $\mu = E(X) = 2.5$  ist. Wir erhalten daher

$$\sigma^2 = V(X) = \frac{1}{2}(1 - 2.5)^2 + \frac{1}{4}(2 - 2.5)^2 + \frac{1}{6}(4 - 2.5)^2 + \frac{1}{12}(10 - 2.5)^2 = \dots = 6.25 . \quad \square$$

Die Varianz und die Standardabweichung haben keine so direkte anschauliche Bedeutung wie der Erwartungswert. Die Situation ist ähnlich wie in der beschreibenden Statistik (2.2.3.8). In (5.3) werden wir noch etwas genauer auf diese Problematik eingehen.

Als nächstes betrachten wir noch den Spezialfall, wo die Zufallsgrösse  $X$  nur endlich viele Werte  $x_1, \dots, x_n$  annimmt, und zwar alle mit derselben Wahrscheinlichkeit  $p_i = 1/n$ . Dann ist

$$\sigma^2 = \sum_{i=1}^n p_i (x_i - \mu)^2$$

oder

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 .$$

Diese Formel wurde schon in (2.2.3.7) erwähnt. Beachten Sie, dass hier im Nenner die Zahl  $n$  steht. Bei der statistischen Varianz dagegen steht im Nenner die Zahl  $n - 1$ , siehe (2.2.3.7.b). Eine Begründung für diesen Unterschied wird in der beurteilenden Statistik gegeben (8.2.6).

Als Nächstes erwähnen wir zwei Formeln für  $V(X)$ , welche die Varianz von  $X$  mit den Erwartungswerten von gewissen andern Zufallsgrössen in Beziehung bringen. Die Herleitung finden Sie im Anhang (13.2). Es gelten die Formeln

$$V(X) = E((X - \mu)^2) = E((X - E(X))^2) ,$$

$$V(X) = E(X^2) - E(X)^2 .$$

Die zweite Formel hat den Vorteil, dass nicht für jedes  $i$  die Differenz  $x_i - E(X)$  einzeln gebildet werden muss.

Beispiel 5.2.B

Wir rechnen Beispiel 5.2.A mit der neuen Formel nochmals durch: Für die Zufallsgrösse  $X$  haben wir folgende Verteilungstabelle:

	1	2	4	10
	1/2	1/4	1/6	1/12

Den zugehörigen Erwartungswert  $\mu = E(X) = 2.5$  kennen wir bereits von früher. Nun brauchen wir noch den Erwartungswert  $E(X^2)$  der neuen Zufallsgrösse  $X^2$ . Deren Verteilung ist gegeben durch

	1	4	16	100
	1/2	1/4	1/6	1/12

(die von  $X$  angenommenen Werte werden einfach quadriert, die Wahrscheinlichkeiten bleiben unverändert). Man erhält sofort

$$E(X^2) = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 16 \cdot \frac{1}{6} + 100 \cdot \frac{1}{12} = 12.5 .$$

Die Formel  $V(X) = E(X^2) - E(X)^2$  liefert nun  $V(X) = 12.5 - 2.5^2 = 6.25$ , was wir vorher schon einmal ausgerechnet haben.  $\square$

(5.3) Näheres zur Bedeutung der Standardabweichung

Zum Schluss sei noch etwas zur konkreten Interpretation der Standardabweichung gesagt. Wir betrachten als Beispiel zwei Zufallsgrössen  $X, Y$  die beide den Erwartungswert  $\mu = E(X) = E(Y) = 100$  haben sollen. Ferner habe  $X$  die Standardabweichung  $\sigma = \sqrt{V(X)} = 10$  und  $Y$  die Standardabweichung  $\sqrt{V(Y)} = 20$ .

Die Bedeutung von  $\mu = 100$  ist nach (5.1) anschaulich leicht verständlich, es handelt sich dabei um den "mittleren Wert" der Zufallsgrösse  $X$  bzw.  $Y$ . Im Zusammenhang mit der Standardabweichung ist zunächst nur klar, dass  $Y$  breiter gestreut ist als  $X$ ,  $\sigma$  hat also vorerst eine relative, vergleichende Bedeutung. Was aber  $\sigma = 10$  "absolut" bedeutet, ist nicht so leicht einzusehen. Für ein besseres Verständnis bietet nun die folgende Formel (1) eine kleine Hilfe. Es sei  $X$  eine beliebige Zufallsgrösse mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma^2$ . Für jede positive Zahl  $r$  gilt dann

$$(1) \quad P(|X - \mu| < r\sigma) \geq 1 - \frac{1}{r^2} .$$

In Worten bedeutet dies: Die Wahrscheinlichkeit dafür, dass der Wert von  $X$  um weniger als  $r\sigma$  vom Erwartungswert abweicht, ist grösser als eine bestimmte, nur von  $r$  abhängige Zahl. Es ist übrigens offensichtlich

$$P(|X - \mu| < r\sigma) = P(\mu - r\sigma < X < \mu + r\sigma) .$$

Wir erhalten z.B.

- für  $r = 2$ :  $P(|X - \mu| < 2\sigma) \geq 3/4$ ,
- für  $r = 3$ :  $P(|X - \mu| < 3\sigma) \geq 8/9$ ,
- für  $r = 4$ :  $P(|X - \mu| < 4\sigma) \geq 15/16$ .

Wenden wir dies konkret auf unsere eingangs gegebene Zufallsgrösse  $X$  mit  $\mu = 100$ ,  $\sigma = 10$  an, so finden wir

- mit einer Wahrscheinlichkeit, die  $\geq 0.75$  ist, liegt der Wert von  $X$  im Intervall  $[80, 120]$ ,
- mit einer Wahrscheinlichkeit, die  $\geq 0.8889$  ist, liegt der Wert von  $X$  im Intervall  $[70, 130]$ ,
- mit einer Wahrscheinlichkeit, die  $\geq 0.9375$  ist, liegt der Wert von  $X$  im Intervall  $[60, 140]$ .

Damit hat der Wert der Standardabweichung  $\sigma$  in einem gewissen Sinn auch eine konkrete, absolute Bedeutung gewonnen. Der Vorteil der Formel (1) ist der, dass sie für ganz beliebige Zufallsgrössen

gilt\*. Ein Nachteil ist, dass die gelieferten Abschätzungen eher schlecht sind. Für  $r \leq 1$  besagt (1) bloss, dass  $P(|X - \mu| < r\sigma) \geq 0$  ist, was zwar stimmt, aber wenig hilfreich ist. Man wird diese Formel also nur verwenden, wenn man über die Verteilung von  $X$  nichts Näheres weiss. Wenn man konkrete Informationen hat, dann kann man meist viel bessere Abschätzungen geben. So werden wir in (5.10.7) sehen, dass im Fall der Normalverteilung  $P(|X - \mu| < 2\sigma) = 0.9545$  ist, diese Wahrscheinlichkeit ist viel grösser als der weiter oben gefundene allgemeingültige Wert 0.75.

Die Formel (1) ist eine unmittelbare Folgerung aus der so genannten *Ungleichung von Bienaymé-Tschebyscheff\*\**, die Sie im Anhang (13.5) finden.

(5.4) Einige diskrete Verteilungen und ihre Masszahlen

In (4.1.9): “Anhang: Einige diskrete Verteilungen”, haben wir drei Verteilungen, denen wir schon mehrmals begegnet waren, bei ihrem in der Literatur gebräuchlichen Namen allgemein nochmals definiert. Damals standen uns die Konzepte Erwartungswert und Varianz noch nicht zur Verfügung. Deshalb repetieren wir kurz diese 3 Verteilungen nochmals.

a) Die diskrete Gleichverteilung

Man sagt, dass die diskrete Zufallsgrösse  $X$  einer *diskreten Gleichverteilung* folgt, wenn sie alle vorkommenden Werte mit derselben Wahrscheinlichkeit annimmt:

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \dots & x_n \\ \hline P(X = x_i) & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{array}$$

Der Erwartungswert bzw. die Varianz der Gleichverteilung sind gegeben durch

$$E(X) = \frac{1}{n}(x_1 + x_2 + \dots + x_n), \quad V(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

b) Die geometrische Verteilung

Allgemein heisst eine Zufallsgrösse  $X$  *geometrisch verteilt* (mit dem Parameter  $p$  ( $0 < p < 1$ )), wenn sie die (abzählbar vielen) Werte  $1, 2, 3, \dots$  annimmt und wenn gilt

$$P(X = k) = p \cdot (1 - p)^{k-1}, \quad k = 1, 2, 3, \dots$$

Man kann zeigen, dass

$$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}.$$

---

\* Im Anhang (13.5) steht der Beweis für diskrete Zufallsgrössen, welche nur endlich viele Werte annehmen. Die Formel gilt aber auch für beliebige diskrete und stetige Zufallsgrössen, für welche die Varianz existiert.

\*\* I.-J. BIENAYMÉ, 1796–1878., P.L. TSCHEBYSCHEFF, 1821–1894.

c) Die hypergeometrische Verteilung

Man hat  $N$  Objekte, wovon  $M$  "gut" sind. Aus den gesamten  $N$  Objekten werden nun  $n$  ausgewählt. Wie gross ist die Wahrscheinlichkeit  $P(X = k)$  dafür, dass unter den  $n$  ausgewählten Objekten genau  $k$  "gute" sind,  $\max(0, n + M - N) \leq k \leq \min(M, n)$ ? Es gilt:

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Mit der Abkürzung

$$p := \frac{M}{N} \quad \text{ist} \quad E(X) = np, \quad V(X) = np(1-p) \frac{N-n}{N-1}.$$

(5.5) Erwartungswert und Varianz der Binomialverteilung

Wir wollen nun noch angeben, wie die wahrscheinlichkeitstheoretischen Masszahlen im Spezialfall der Binomialverteilung aussehen. Wir gehen dazu von den allgemeinen Formeln von (5.1) bzw. (5.2) aus:

$$\mu = E(X) = \sum_{k=0}^n p_k x_k, \quad \sigma^2 = V(X) = \sum_{k=0}^n p_k (x_k - \mu)^2.$$

Im Fall der Binomialverteilung ist  $x_k = k$  und  $p_k = P(X = k) = \binom{n}{k} p^k q^{n-k}$ . Setzt man dies ein, findet man

$$\mu = E(X) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k$$

$$\sigma^2 = V(X) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (k - \mu)^2.$$

Die Berechnung von  $\mu$  und  $\sigma^2$  anhand dieser Formeln ist etwas umständlich und wurde in den Anhang (13.3) verwiesen. Für die Praxis sind vor allem die Ergebnisse dieser Rechnungen wichtig. Man findet

$$\mu = E(X) = np, \quad \sigma^2 = V(X) = npq.$$

Das Resultat für den Erwartungswert ( $\mu = np$ ) leuchtet auch ohne Rechnung ein: Führt man nämlich  $n$  Versuche durch, wobei die Erfolgswahrscheinlichkeit im Einzelversuch

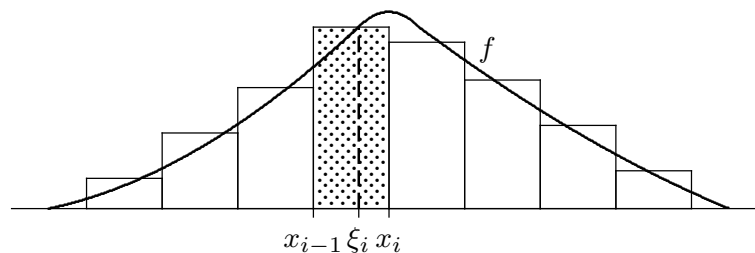
$= p$  ist, so wird man im Mittel — d.h., wenn man sehr oft eine solche Serie von  $n$  Versuchen durchführt — etwa  $np$  Erfolge erwarten.

Würfelt man z.B. mit einem unverfälschten Würfel zwölfmal, so wird man erwarten, dass wegen  $12 \cdot \frac{1}{6} = 2$  zweimal eine Sechs herauskommt, denn die Wahrscheinlichkeit für eine Sechs im Einzelversuch ist  $\frac{1}{6}$ . Diese Zahl 2 ist gerade der Erwartungswert  $np$ .

Natürlich wird man nicht mit Sicherheit sagen können, dass in einer Serie von zwölf Würfeln *genau* zwei Sechsen auftreten. Spielt man aber sehr viele solcher Serien durch und notiert sich jedes Mal die Anzahl der Sechsen, so erwartet man, dass im Durchschnitt aller Serien ziemlich genau zwei Sechsen pro Serie geworfen werden und zwar umso genauer, je mehr Serien gespielt werden. Dieser Gedankengang gibt einfach nochmals die Idee des Erwartungswerts (5.1) wieder.

(5.6) Der Erwartungswert einer stetigen Zufallsgrösse

Wir suchen eine vernünftige Definition des Erwartungswerts einer stetigen Zufallsgrösse. Es geht im Folgenden darum, die am Schluss des Abschnitts angeführte Definition von  $E(X)$  plausibel zu machen. Dazu gehen wir die in (4.3.2) gemachten Schritte rückwärts und “diskretisieren” die stetige Zufallsgrösse  $X$  mit Dichtefunktion  $f(x)$ , indem wir die  $x$ -Achse unterteilen:



Die Wahrscheinlichkeit  $p_i$  dafür, dass ein Wert in das Intervall  $[x_{i-1}, x_i]$  der Länge  $\Delta x_i = x_i - x_{i-1}$  fällt, ist gleich dem Inhalt des hervorgehobenen Rechtecks. Für eine geeignete Zahl  $\xi_i$  aus  $[x_{i-1}, x_i]$  ist dann dieser Inhalt gleich

$$f(\xi_i) \Delta x_i ,$$

und es folgt

$$p_i = f(\xi_i) \Delta x_i .$$

Diskretisieren des Merkmals bedeutet, dass wir alle Werte im Intervall  $[x_{i-1}, x_i]$  durch einen einzigen, z.B. durch  $\xi_i$  ersetzen.

Es liegt nunmehr eine diskrete Zufallsvariable vor, welche die Werte  $\xi_i$  mit der Wahrscheinlichkeit  $p_i = f(\xi_i)\Delta x_i$  annimmt. Nach der Definition von  $\mu$  (5.1) hat diese Zufallsgrösse den Erwartungswert

$$\mu = \sum_i \xi_i f(\xi_i) \Delta x_i .$$

Dies ist aber gerade eine Riemannsche Summe für die Funktion  $xf(x)$  (vgl. (10.2) im ersten Band). Gehen wir nun zur stetigen Verteilung zurück, so geht  $\Delta x_i \rightarrow 0$ , und die Riemannsche Summe strebt gegen ein bestimmtes Integral:

$$\mu = \sum_i \xi_i f(\xi_i) \Delta x_i \rightarrow \int_a^b x f(x) dx .$$

Ersetzen wir nun noch  $a$  und  $b$  durch  $-\infty$  und  $\infty$ , so haben wir die nachstehende *Definition* motiviert.

Es sei  $X$  eine stetige Zufallsgrösse mit der Dichtefunktion  $f(x)$ . Dann ist der *Erwartungswert* von  $X$  definiert durch

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

### Bemerkung

In der Definition steht ein uneigentliches Integral, das nicht zu existieren braucht (siehe Beispiel 5.8.B). Wenn dem so ist, dann hat  $X$  eben keinen Erwartungswert. Aus gewissen theoretischen Gründen fordert man sogar die Existenz von  $\int_{-\infty}^{\infty} |x|f(x) dx$ , was uns weiter nicht berührt.

### (5.7) Die Varianz einer stetigen Zufallsgrösse

Beim Erwartungswert haben wir den folgenden Übergang vom diskreten zum stetigen Fall gemacht:

$$\mu = E(X) = \sum x_i p_i \rightsquigarrow E(X) = \int_{-\infty}^{\infty} x f(x) dx .$$

In Analogie dazu steht der entsprechende Übergang für die Varianz

$$\sigma^2 = V(X) = \sum (x_i - \mu)^2 p_i \rightsquigarrow V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

Dies motiviert die folgenden Definitionen:

Es sei  $X$  eine stetige Zufallsgrösse mit der Dichtefunktion  $f(x)$ . Dann ist die *Varianz* von  $X$  definiert durch

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad \text{mit} \quad \mu = E(X)$$

(sofern sowohl  $\mu$  als auch dieses uneigentliche Integral existiert).

Die *Standardabweichung* von  $X$  ist die positive Wurzel aus der Varianz:

$$\sigma = \sqrt{V(X)} \geq 0.$$

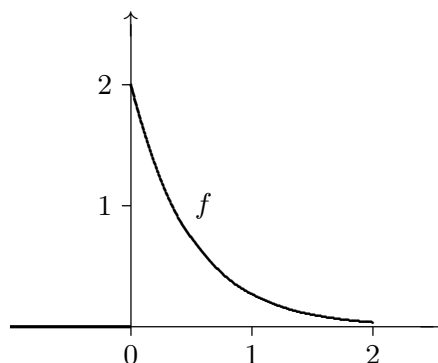
### (5.8) Beispiele zu Dichtefunktionen und Erwartungswerten

#### Beispiel 5.8.A

Es sei  $\lambda$  eine positive Zahl. Wir setzen

$$f(x) = \begin{cases} Ke^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0. \end{cases}$$

Dabei soll die Konstante  $K$  so bestimmt werden, dass  $f$  eine Dichtefunktion wird.



Wir kontrollieren die Bedingungen 1) bis 3) von (4.3.3) nach. Die Funktion  $f$  ist zwar nicht stetig (Unstetigkeit im Punkt  $x = 0$ ), aber immerhin stückweise stetig: 1) ist erfüllt und ebenso 2) (offensichtlicherweise). Die dritte Bedingung

$$(\star) \quad \int_{-\infty}^{\infty} f(t) dt = 1$$

kann erfüllt werden, wenn wir  $K$  richtig wählen. Um dieses uneigentliche Integral zu berechnen, müssen wir gemäss (20.3), siehe erster Band, zwei uneigentliche Integrale, nämlich

$$(1) \int_{-\infty}^c f(t) dt \quad \text{und} \quad (2) \int_c^{\infty} f(t) dt$$

für eine passende Zahl  $c$  getrennt betrachten. Es sollte klar sein, dass wir  $c = 0$  wählen. Das Integral (1) hat dann den Wert 0. Daher liefert nur das Integral (2) einen Beitrag zum gesamten uneigentlichen Integral aus  $(\star)$ . Wir berechnen (2), wie im ersten Band (20.2) gelernt, als Grenzwert:

$$\begin{aligned} \int_0^{\infty} f(t) dt &= \lim_{x \rightarrow \infty} \int_0^x f(t) dt = \lim_{x \rightarrow \infty} \int_0^x K e^{-\lambda t} dt \\ &= \lim_{x \rightarrow \infty} K \left( -\frac{1}{\lambda} \right) e^{-\lambda t} \Big|_0^x = \lim_{x \rightarrow \infty} \left( -\frac{K}{\lambda} (e^{-\lambda x} - 1) \right) = \frac{K}{\lambda}. \end{aligned}$$

Dabei wurde benützt, dass  $e^{-\lambda x}$  für  $x \rightarrow \infty$  gegen Null strebt. Die Bedingung  $(\star)$  ist also erfüllt, wenn  $\frac{K}{\lambda} = 1$ , also  $K = \lambda$  ist.

Wir bestimmen noch die Verteilungsfunktion  $F(x)$ . Nach Definition ist für  $x \geq 0$

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x \lambda e^{-\lambda t} dt = -(e^{-\lambda x} - 1) = 1 - e^{-\lambda x}.$$

Das Integral ist weiter oben schon ausgerechnet worden; es wurde noch der bereits bestimmte Wert  $K = \lambda$  eingesetzt. Für den (uninteressanten) Fall  $x < 0$  ist  $F(x) = 0$ .

Damit können wir nun gemäss (4.3.4) Wahrscheinlichkeiten berechnen. Wir setzen dabei  $\lambda = 2$  (dieser Fall ist in der Figur dargestellt). Es sei also  $X$  eine Zufallsgrösse, welche die entsprechende Dichtefunktion  $f$  hat. Wir erhalten beispielsweise:

- a)  $P(X \leq 1) = F(1) = 1 - e^{-2 \cdot 1} = 0.8647$ ,
- b)  $P(X > 3) = 1 - F(3) = 1 - (1 - e^{-2 \cdot 3}) = e^{-2 \cdot 3} = 0.0025$ ,
- c)  $P(2 \leq X \leq 4) = F(4) - F(2) = (1 - e^{-8}) - (1 - e^{-4}) = 0.0180$ .

Alle diese Wahrscheinlichkeiten können als Flächeninhalte unter dem Graphen von  $f$  aufgefasst werden, vgl. ähnliche Skizzen in (4.3.4).

Die hier besprochene Verteilung heisst *Exponentialverteilung*.

Nun bestimmen wir noch den Erwartungswert von  $X$ . Nach (4.3.6) ist dieser durch das folgende uneigentliche Integral gegeben:

$$\lambda = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx.$$

Mit partieller Integration, (13.5) im ersten Band, finden wir

$$\int \lambda x e^{-\lambda x} dx = \lambda \left( -\frac{1}{\lambda} x e^{-\lambda x} + \frac{1}{\lambda} \int e^{-\lambda x} dx \right) = -x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x}.$$

Einsetzen der Grenzen ergibt

$$\int_0^u \lambda x e^{-\lambda x} dx = -u e^{-\lambda u} - \frac{1}{\lambda} (e^{-\lambda u} - 1).$$

Jetzt lassen wir  $u \rightarrow \infty$  gehen. Dabei streben  $e^{-\lambda u}$  und  $u e^{-\lambda u}$  beide gegen Null. Die zweite Behauptung ergibt sich aus der in (19.8.a), erster Band, erwähnten Beziehung  $\lim_{x \rightarrow \infty} e^x / x^n = \infty$ . Setzen wir nämlich  $n = 1$  und gehen zum Kehrwert über, so wird  $\lim_{x \rightarrow \infty} x / e^x = \lim_{x \rightarrow \infty} x e^{-x} = 0$ , woraus die Behauptung folgt, wenn man noch  $x$  durch  $\lambda u$  ersetzt. (Vgl. dazu eine analoge Rechnung in (20.2), erster Band, Beispiel 4.) Aus alledem ergibt sich schliesslich  $\lambda = E(X) = \frac{1}{\lambda}$ .

Eine ähnliche, aber noch etwas kompliziertere Rechnung liefert für die Varianz den Wert  $V(X) = 1/\lambda^2$ .

□

Beispiel 5.8.B

Wir betrachten die Funktion

$$f(x) = \frac{c}{1+x^2}$$

und wollen zunächst  $c$  so bestimmen, dass  $f$  eine Dichtefunktion wird. Sicher sind die Bedingungen 1) und 2) von (4.3.3) erfüllt. Die Bedingung 3) besagt

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{c}{1+x^2} dx = c \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = 1 .$$

Nun haben wir aber den Wert des letzten Integrals in Beispiel 1 von (20.3) im ersten Band bestimmt; er ist gleich  $\pi$ . Somit ist  $c = \frac{1}{\pi}$  zu setzen.

Wir berechnen damit  $P(-1 \leq X \leq 1)$  für die so verteilte Zufallsgrösse  $X$ . Diesmal verwenden wir nicht wie in Beispiel 4.3.8.A die Verteilungsfunktion, sondern operieren direkt mit dem Integral (siehe den Schluss von (4.3.4.b)).

$$P(-1 \leq X \leq 1) = \frac{1}{\pi} \int_{-1}^1 \frac{1}{1+x^2} dx = \frac{1}{\pi} \arctan x \Big|_{-1}^1 = \frac{1}{\pi} \left( \frac{\pi}{4} - \left(-\frac{\pi}{4}\right) \right) = \frac{1}{2} .$$

Diese Verteilung wird *Cauchy-Verteilung*\* genannt. Diese Verteilung ist identisch mit der  $t_1$ -Verteilung (siehe 6.2.6).

Wir weisen noch nach, dass diese Verteilung keinen Erwartungswert besitzt. Dazu genügt es zu zeigen, dass das uneigentliche Integral

$$\int_0^{\infty} x f(x) dx = \int_0^{\infty} \frac{x}{1+x^2} dx$$

nicht existiert, denn dann existiert auch das entsprechende Integral mit den Grenzen  $-\infty$  und  $\infty$  nicht. Nun ist aber (man verwendet die Substitution  $v = 1+x^2$ ,  $dv = 2x dx$ ; vgl. (13.2) und (13.4) im ersten Band)

$$\int_0^u \frac{x}{1+x^2} dx = \frac{1}{2} \ln(1+u^2) ,$$

und dieser Ausdruck strebt mit wachsendem  $u$  gegen unendlich. Das uneigentliche Integral existiert daher nicht;  $X$  hat keinen Erwartungswert. ☒

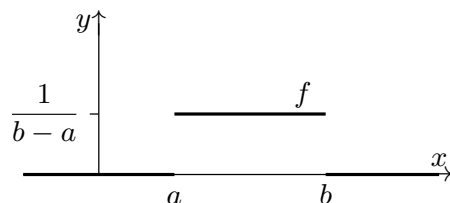
(5.9) Die stetige Gleichverteilung

Die für uns wichtigste stetige Verteilung, die Normalverteilung, wird in Teil 5.10 näher behandelt. In der beurteilenden Statistik werden wir sodann die  $t$ -Verteilung und die  $\chi^2$ -Verteilung antreffen. Als ein sehr einfaches Beispiel einer stetigen Verteilung erwähnen wir hier die *stetige Gleichverteilung*, auch Rechteckverteilung genannt.

---

\* A.L. CAUCHY, 1789–1857.

Ihre Dichtefunktion  $f(x)$  ist in einem vorgegebenen Intervall  $[a, b]$  konstant, und zwar hat sie dort den Wert  $1/(b-a)$ ; ausserhalb dieses Intervalls ist  $f(x) = 0$ . Ihr Graph sieht so aus:



Sind die Bedingungen, die wir in (4.3.3) an eine Dichtefunktion stellen, erfüllt?

- 1)  $f$  ist stückweise stetig (aber nicht stetig, denn  $f$  hat Sprungstellen in  $a$  und  $b$ ).
- 2)  $f(x) \geq 0$  für alle  $x$  nach Definition.
- 3) Nur das Rechteck liefert einen Beitrag zur Integration von  $-\infty$  bis  $\infty$ . Geometrisch ist klar, dass sein Flächeninhalt = 1 ist (denn  $f(x) = 1/(b-a)$  in  $[a, b]$ ), somit ist  $\int_{-\infty}^{\infty} f(t) dt = 1$ .

Alle drei Bedingungen gelten:  $f(x)$  ist tatsächlich eine Dichtefunktion.

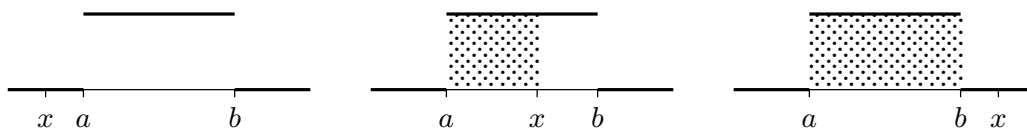
Die zugehörige Verteilungsfunktion ist gegeben durch

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

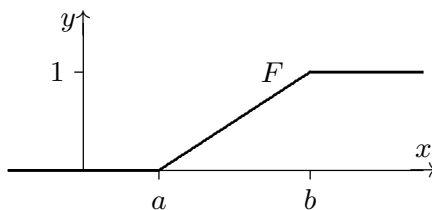
Man könnte dieses Integral analytisch berechnen (mit Fallunterscheidung  $x < a$ ,  $a \leq x < b$ ,  $b \leq x$ ). Aufgrund geometrischer Überlegungen (Flächeninhalt; vgl. die drei unten stehenden Figuren) erkennt man aber sofort, dass gilt:

$$(*) \quad F(x) = \begin{cases} 0 & \text{für } x < a \\ (x - a)/(b - a) & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

(Das Rechteck hat ja die Höhe  $1/(b-a)$ .)



Der Graph der Verteilungsfunktion  $F(x)$  hat die folgende Form:



Man erkennt auch sofort, dass  $F'(x) = f(x)$  ist (abgesehen von den Stellen  $a$  und  $b$ , wo  $F$  nicht differenzierbar ist, vgl. Bemerkung c) von (4.3.3)). Im nächsten Beispiel kommt eine solche Gleichverteilung vor.

### Beispiel 5.9.A

Wir betrachten ein Glücksrad, wie z.B. in (4.1.8). Der Winkel  $X$  (mit  $0 \leq X < 2\pi$ ), bei welchem der Zeiger stehenbleibt, sei unsere Zufallsgrösse. Sie kann jeden Wert in  $[0, 2\pi)$  annehmen und sollte deshalb aufgrund unserer Anschauung eine stetige Zufallsgrösse sein. Um auch der theoretischen Definition zu genügen, müssen wir eine Dichtefunktion gemäss (4.3.3) angeben.

Unser Gefühl sagt uns, dass bei einem "ehrlichen" Glücksrad jeder Ausgang dieselbe Wahrscheinlichkeit hat. Da  $X$  aber stetig ist, ist  $P(X = x) = 0$  für alle  $x \in [0, 2\pi)$ . Dieser scheinbare Widerspruch löst sich, wenn wir die "Ehrlichkeit" des Rades so formulieren: Im Intervall  $[0, 2\pi]$  ist die *Dichtefunktion* konstant. Dies ist nun gerade der Fall einer stetigen Gleichverteilung, mit  $a = 0$ ,  $b = 2\pi$ .

Die Graphen von  $f(x)$  und  $F(x)$  entsprechen den oben stehenden Figuren, wenn man  $a = 0$ ,  $b = 2\pi$  setzt.

Mit der weiter oben erwähnten Formel (\*) für  $F(x)$  erhalten wir nun zum Beispiel:

$$P\left(X \leq \frac{\pi}{2}\right) = F\left(\frac{\pi}{2}\right) = \frac{\frac{\pi}{2} - 0}{2\pi - 0} = \frac{1}{4},$$

$$P(X \leq \pi) = F(\pi) = \frac{\pi - 0}{2\pi - 0} = \frac{1}{2},$$

in Übereinstimmung mit der Anschauung und den Überlegungen in (4.1.8). □

Schliesslich berechnen wir noch als Anwendung der Formeln aus (5.6) und (5.7) Erwartungswert und Varianz der stetigen Gleichverteilung. In den entsprechenden Integralen können die Grenzen  $-\infty$  und  $\infty$  durch  $a$  und  $b$  ersetzt werden, da die Dichtefunktion  $f(x)$  ausserhalb des Intervalls  $[a, b]$  den Wert 0 annimmt. Wir erhalten also zunächst

$$\mu = E(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Der Erwartungswert ist also gleich dem Durchschnitt der Intervallgrenzen, d.h., gleich der Mitte des Intervalls  $[a, b]$ .

Für die Varianz lassen wir die Details der (etwas langweiligen, aber problemlosen) Rechnung aus.

$$\sigma^2 = V(X) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \dots = \frac{(b-a)^2}{12}.$$

Beispiel 5.9.B

Auf der S-Bahn-Linie S9 verkehrt tagsüber alle 30 Minuten ein Zug. Ein Bahnbenützer pflegt zufällig auf dem Bahnhof einzutreffen. Wie lange muss er im Durchschnitt warten?

Der gesunde Menschenverstand suggeriert, dass die mittlere Wartezeit 15 Minuten beträgt. Dies wollen wir nun rechnerisch bestätigen. Die Wartezeit  $X$  (in Minuten) ist eine Zufallsvariable, die Werte zwischen 0 und 30 annimmt (Pünktlichkeit der Züge einmal vorausgesetzt!). Die Voraussetzung, dass unser Freund des öffentlichen Verkehrs zufällig eintrifft, dass also kein Zeitintervall bevorzugt ausgewählt wird, bedeutet gerade, dass  $X$  einer Gleichverteilung folgt. Dies ist — vom Standpunkt der Wahrscheinlichkeitsrechnung aus — die Hauptüberlegung. Die mittlere Wartezeit ist dann der Erwartungswert  $\mu = E(X)$ , der nach der oben hergeleiteten Formel  $= 15$  ist.

## (5.10) Die Normalverteilung

## (5.10.1) Überblick

Die stetige Zufallsgrösse  $X$  folgt der *Normalverteilung* mit den *Parametern*  $\mu, \sigma^2$  ( $\mu \in \mathbb{R}$ ), abgekürzt “der Verteilung  $N(\mu; \sigma^2)$ ”, wenn sie die Dichtefunktion

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (5.10.2)$$

hat.

Im Spezialfall, wo  $\mu = 0$  und  $\sigma^2 = 1$  ist, spricht man von der *Standard-Normalverteilung* und bezeichnet die Dichtefunktion mit  $\varphi(x)$ . (5.10.3)

Der *Graph* der Dichtefunktion der Normalverteilung  $N(\mu; \sigma^2)$  ist glockenförmig, symmetrisch bezüglich der Geraden  $x = \mu$ , und er hat Wendepunkte an den Stellen  $\mu - \sigma$  und  $\mu + \sigma$ . (5.10.2)

Der *Erwartungswert* der Normalverteilung  $N(\mu; \sigma^2)$  ist gleich  $\mu$ ; ihre *Standardabweichung* ist gleich  $\sigma$  und demnach die Varianz gleich  $\sigma^2$ . (5.10.6)

Die zugehörige *Verteilungsfunktion* bezeichnet man mit  $\Phi_{\mu, \sigma^2}(x)$ , im Falle der Standard-Normalverteilung kurz mit  $\Phi(x)$ . Die Werte von  $\Phi(x)$  sind tabelliert; zur Berechnung von  $\Phi_{\mu, \sigma^2}(x)$  für allgemeine Werte von  $\mu$  und  $\sigma^2$  muss man die folgende Formel benützen: (5.10.3)

$$\Phi_{\mu, \sigma^2}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Für jedes  $\alpha \in [0, 1]$  ist der *kritische Wert*  $z_\alpha$  dadurch definiert, dass (5.10.5)

$$P(|X| \geq z_\alpha) = \alpha$$

ist, wobei  $X$  der Standard-Normalverteilung folgt.

## (5.10.2) Definition der Normalverteilung

Die für uns wichtigste stetige Verteilung ist die Normalverteilung (auch Gauss-Verteilung\* genannt). Sie hängt von zwei “Parametern”  $\mu$  und  $\sigma^2$  ( $\mu$  beliebig) ab; es gibt also unendlich viele Normalverteilungen. Man benützt oft das Zeichen  $N(\mu; \sigma^2)$  um anzugeben, dass eine Normalverteilung mit den Parametern  $\mu$  und  $\sigma^2$  vorliegt.

\* Nach C.F. GAUSS, 1777–1855.

Hinweis

In manchen Büchern, auch in R, verwendet man die Bezeichnung  $N(\mu; \sigma)$  statt  $N(\mu; \sigma^2)$ . In diesem Fall bedeutet  $N(0; 2)$  also eine Normalverteilung mit  $\mu = 0$ ,  $\sigma = 2$ . Wir werden aber bei der Form  $N(\mu; \sigma^2)$  bleiben, weil dies in den weiterführenden Vorlesungen sinnvoll ist. Weil die Varianz  $\sigma^2$  ein quadratisches Mass ist, argumentiert man aber oft mit  $\sigma$  (Verwechslungsgefahr!).

Wie jede stetige Verteilung kann die Normalverteilung durch ihre Dichtefunktion gegeben werden (4.3.3). Wir definieren:

Die stetige Zufallsgrösse  $X$  folgt der *Normalverteilung*  $N(\mu; \sigma^2)$  ( $\mu, \sigma^2 \in \mathbb{R}$ ,  $\sigma > 0$ ), wenn ihre Dichtefunktion die Gestalt

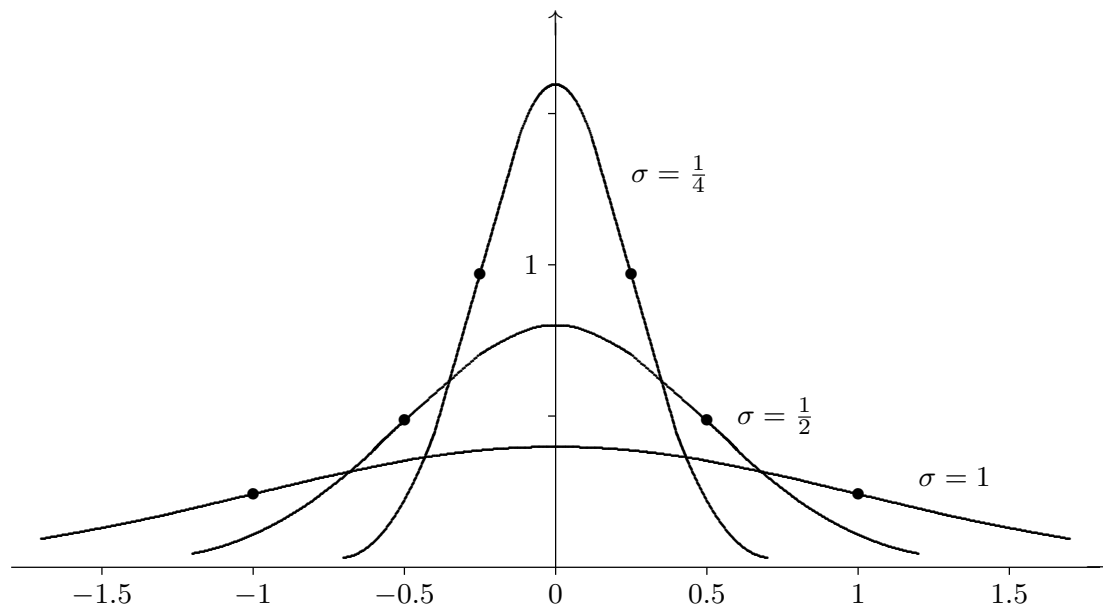
$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

hat. Im Fall  $N(0; 1)$ , wo also  $\mu = 0$  und  $\sigma^2 = 1$  ist, spricht man von der *Standard-Normalverteilung*. Ihre Dichtefunktion wird einfach mit  $\varphi$  bezeichnet; sie ist gegeben durch

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Wir haben bereits in (4.3.3) erwähnt, dass diese Funktion (die Indizes  $\mu, \sigma^2$  hatten wir dort weggelassen) die Bedingungen erfüllt, die man an eine Dichtefunktion stellt.

Eine Diskussion des Graphen dieser Funktion steht im Anhang (13.6). Wir fassen die wichtigsten Punkte zusammen. Zunächst stellen wir den Graphen der Dichtefunktion  $\varphi_{\mu, \sigma^2}$  für  $\mu = 0$ ,  $\sigma = \frac{1}{4}, \frac{1}{2}, 1$  vor:



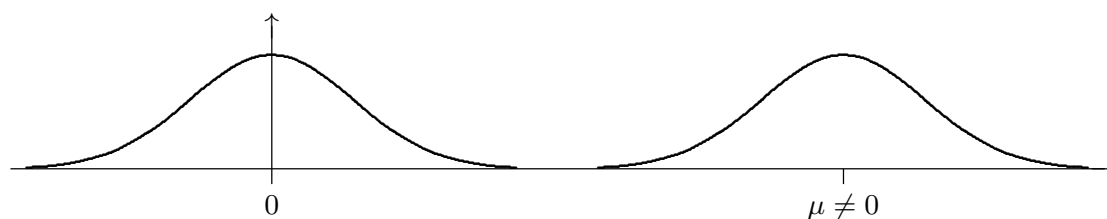
Der Graph hat die folgenden Eigenschaften:

- 1) Er ist symmetrisch in Bezug auf die Gerade  $x = \mu$ .
- 2) Er hat Wendepunkte in  $\mu - \sigma$ ,  $\mu + \sigma$ .
- 3) Es ist  $\lim_{x \rightarrow \infty} \varphi_{\mu, \sigma^2}(x) = \lim_{x \rightarrow -\infty} \varphi_{\mu, \sigma^2}(x) = 0$ .
- 4) Spezielle Funktionswerte:

$$\varphi_{\mu, \sigma^2}(\mu) = \frac{1}{\sqrt{2\pi}\sigma}, \quad \varphi_{\mu, \sigma^2}(\mu - \sigma) = \varphi_{\mu, \sigma^2}(\mu + \sigma) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{e}}.$$

Der Graph wird ‘‘Glockenkurve’’ oder ‘‘Gauss-Kurve’’ genannt. Je grosser  $\sigma$  ist, desto breiter und niedriger wird die Glockenkurve, was auch an den drei Fallen der Figur illustriert wird.

Ein von 0 verschiedener Wert von  $\mu$  bewirkt eine Parallelverschiebung des Graphen:



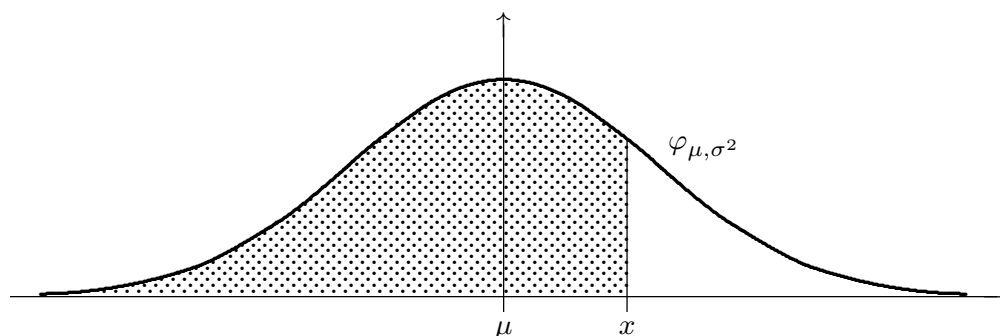
Die Verteilungsfunktion von  $N(\mu; \sigma^2)$  wird mit  $\Phi_{\mu, \sigma^2}(x)$  bezeichnet. Nach (4.3.3) ist sie gegeben durch

$$\Phi_{\mu, \sigma^2}(x) = \int_{-\infty}^x \varphi_{\mu, \sigma^2}(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

Wir erinnern daran, dass fur die Verteilungsfunktion definitionsgemass gilt

$$\Phi_{\mu, \sigma^2}(x) = P(X \leq x).$$

Geometrisch gesehen ist diese Zahl gleich dem Flacheninhalt des in der unten stehenden Figur mit Punkten markierten Stucks unter der Glockenkurve.



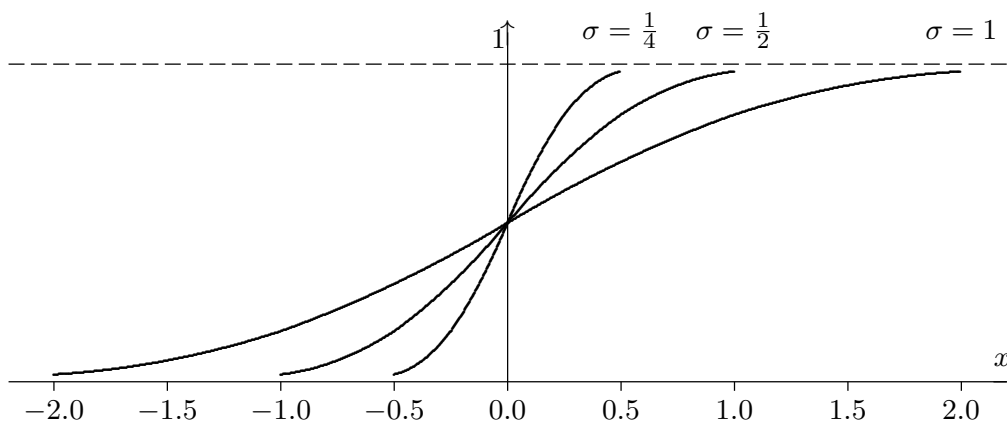
Wir haben schon im ersten Band in (12.3.e) erwähnt, dass die Funktion  $e^{-x^2}$  keine elementare Stammfunktion hat. Daher kann auch  $\Phi_{\mu,\sigma^2}$  als Integral der eng mit  $e^{-x^2}$  verwandten Funktion  $\varphi_{\mu,\sigma^2}(x)$  nicht auf elementare Weise berechnet werden. Zur Berechnung der Werte von  $\Phi_{\mu,\sigma^2}(x)$  muss man vielmehr Reihenentwicklungen heranziehen (vgl. Beispiel 3. von (19.6) im ersten Band).

Immerhin ist aus Symmetriegründen klar, dass  $\Phi_{\mu,\sigma^2}(\mu) = \frac{1}{2}$  ist. Benützt man noch, dass gilt (vgl. (4.3.3))

$$\Phi'_{\mu,\sigma^2}(x) = \varphi_{\mu,\sigma^2}(x),$$

so kann man den Graphen von  $\Phi_{\mu,\sigma^2}$  bereits einigermaßen zeichnen. Genauere Skizzen erhält man, wenn man die Wertetabellen der Verteilungsfunktion verwendet (vgl. (5.10.3)):

*Graph der Verteilungsfunktion  $\Phi_{\mu,\sigma^2}$  für  $\mu = 0$ ,  $\sigma = \frac{1}{4}, \frac{1}{2}, 1$*



Dabei strebt  $\Phi_{\mu,\sigma^2}(x) \rightarrow 1$  für  $x \rightarrow \infty$  und  $\Phi_{\mu,\sigma^2}(x) \rightarrow 0$  für  $x \rightarrow -\infty$ .

(5.10.3) Berechnung der Werte der Verteilungsfunktion,  $Z$ -Transformierte

Wie eben erwähnt wurde, können die Werte der Verteilungsfunktion nicht auf elementare Weise bestimmt werden. Jene der Dichtefunktion sind dagegen leicht mit dem Taschenrechner zu ermitteln.

Die Werte der Verteilungsfunktion der Normalverteilung sind in Tabellen niedergelegt (vgl. Tabelle in (6.2.3)), doch beziehen sich diese Tabellen stets auf die Standard-Normalverteilung  $N(0;1)$ , also die Normalverteilung mit den Parametern  $\mu = 0$  und  $\sigma^2 = 1$ . Für die Dichte- bzw. die Verteilungsfunktion der Standard-Normalverteilung schreiben wir kurz  $\varphi$  bzw.  $\Phi$  (ohne Indizes).

Die Werte der Verteilungsfunktion  $\Phi(x)$  der Standard-Normalverteilung können also Tabellen entnommen werden. Um nun  $\Phi_{\mu,\sigma^2}(x)$  für beliebiges  $\mu, \sigma^2$  berechnen zu können, benützt man die folgende wichtige Formel:

$$(*) \quad \boxed{\Phi_{\mu,\sigma^2}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)} .$$

Etwas abstrakter formuliert: Man geht von der ursprünglichen Zufallsgrösse  $X$  zur neuen Zufallsgrösse  $Z := \frac{X-\mu}{\sigma}$  über. Dieser Vorgang heisst "Standardisierung". Man nennt die neue Zufallsgrösse auch die  $Z$ -Transformierte von  $X$ .

Wir leiten nun die Beziehung (\*) her. Zuerst schreiben wir die Formel für die Dichtefunktion der Standard-Normalverteilung auf, indem wir in der Definition von  $\varphi_{\mu,\sigma^2}$  die Parameter  $\mu = 0$  und  $\sigma^2 = 1$  wählen. Wir erhalten wie in (5.10.2)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} .$$

Ersetzen wir hier  $x$  durch  $\frac{x-\mu}{\sigma}$  und dividieren noch durch  $\sigma$ , so finden wir

$$\frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \varphi_{\mu,\sigma^2}(x) ,$$

also

$$(**) \quad \varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) .$$

Für die Formel (\*) benützen wir aus dem ersten Band die Definition des uneigentlichen Integrals (20.2) sowie die Substitutionsregel (13.2):

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^x \varphi_{\mu,\sigma^2}(t) dt = \lim_{a \rightarrow -\infty} \int_a^x \varphi_{\mu,\sigma^2}(t) dt = \lim_{a \rightarrow -\infty} \int_a^x \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt$$

(wegen Formel (\*\*)). Nun substituieren wir

$$u = \frac{t-\mu}{\sigma}, \quad du = \frac{1}{\sigma} dt .$$

Die untere Grenze  $a$  wird zu  $\frac{a-\mu}{\sigma} = \tilde{a}$  (als Abkürzung), die obere Grenze  $x$  wird zu  $\frac{x-\mu}{\sigma}$ . Wichtig ist, dass mit  $a \rightarrow -\infty$  auch  $\tilde{a} \rightarrow -\infty$  strebt. Wir finden so

$$\begin{aligned} \Phi_{\mu,\sigma^2}(x) &= \lim_{a \rightarrow -\infty} \int_a^x \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt = \lim_{\tilde{a} \rightarrow -\infty} \int_{\tilde{a}}^{\frac{x-\mu}{\sigma}} \varphi(u) du \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(u) du = \Phi\left(\frac{x-\mu}{\sigma}\right) , \end{aligned}$$

wobei am Schluss die Definition von  $\Phi = \Phi_{0,1}$  benützt wurde. Damit ist die Formel (\*) bewiesen.

## (5.10.4) Berechnung von Wahrscheinlichkeiten

Wir geben nun einige Beispiele zur Anwendung der Formel von (5.10.3):

- 1) Gegeben sei die Normalverteilung  $N(-1; 4)$ , das heisst  $\sigma = 2$  (Vorsicht: Fehlerquelle!).

Gesucht sei  $\Phi_{-1,4}(3)$ .

Wir setzen  $x = 3$ ,  $\mu = -1$  und  $\sigma = 2$  in Formel (\*) ein und finden unter Verwendung der Tabelle in (6.2.3)

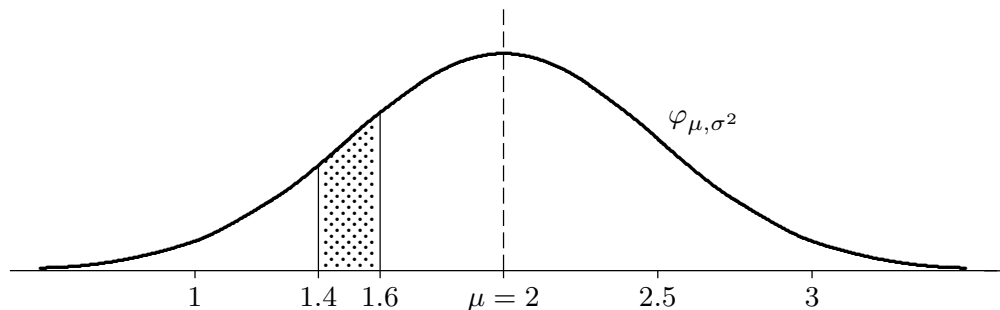
$$\Phi_{-1,4}(3) = \Phi\left(\frac{3 - (-1)}{2}\right) = \Phi(2) = 0.9772 . \quad \boxtimes$$

- 2) Eine Zufallsgrösse  $X$  sei gemäss  $N(2; \frac{1}{4})$  normal verteilt. Mit welcher Wahrscheinlichkeit liegt  $X$  im Intervall  $[1.4, 1.6]$ ?

Wegen Formel b) (und Bemerkung c) von (4.3.4)) ist

$$P(a \leq X \leq b) = \Phi_{\mu, \sigma^2}(b) - \Phi_{\mu, \sigma^2}(a) .$$

Geometrisch entspricht diese Wahrscheinlichkeit dem Inhalt des markierten Flächenstücks.



In der obigen Formel setzen wir  $\mu = 2$ ,  $\sigma = \frac{1}{2}$ ,  $a = 1.4$ ,  $b = 1.6$  und erhalten

$$P(1.4 \leq X \leq 1.6) = \Phi_{2, \frac{1}{4}}(1.6) - \Phi_{2, \frac{1}{4}}(1.4) .$$

Die beiden Funktionswerte berechnen sich zu

$$\Phi_{2, \frac{1}{4}}(1.6) = \Phi\left(\frac{1.6 - 2}{0.5}\right) = \Phi(-0.8) = 0.2119 ,$$

$$\Phi_{2, \frac{1}{4}}(1.4) = \Phi\left(\frac{1.4 - 2}{0.5}\right) = \Phi(-1.2) = 0.1151 .$$

Die gesuchte Wahrscheinlichkeit ist also  $= 0.0968$ .

Für dieselbe Zufallsgrösse bestimmen wir noch  $P(X \geq \pi)$ . Wegen Formel a) und Bemerkung d) von (4.3.4) ist  $P(X \geq \pi) = 1 - P(X \leq \pi) = 1 - \Phi_{2, \frac{1}{4}}(\pi)$ . Standardisierung liefert

$$\Phi_{2, \frac{1}{4}}(\pi) = \Phi\left(\frac{\pi - 2}{0.5}\right) \approx \Phi(2.28) .$$

Nun ist dieser Wert in der Tabelle in (6.2.3) leider nicht zu finden. Wir müssen deshalb interpolieren (für diesbezügliche Einzelheiten siehe (6.2.3)). Es ist  $\Phi(2.25) = 0.9878$  und  $\Phi(2.30) = 0.9893$ . Wir finden so  $\Phi(2.28) \approx 0.9888$ . Es folgt

$$P(X \geq \pi) = 1 - 0.9888 = 0.0112 . \quad \square$$

- 3) Für die Standard-Normalverteilung  $N(0; 1)$  können wir ohne Umrechnung die Tabelle in (6.2.3) verwenden. Hierzu ein weiteres Beispiel: Mit welcher Wahrscheinlichkeit gilt für die standard-normal verteilte Zufallsvariable  $X$ , dass

$$|X| \geq 2$$

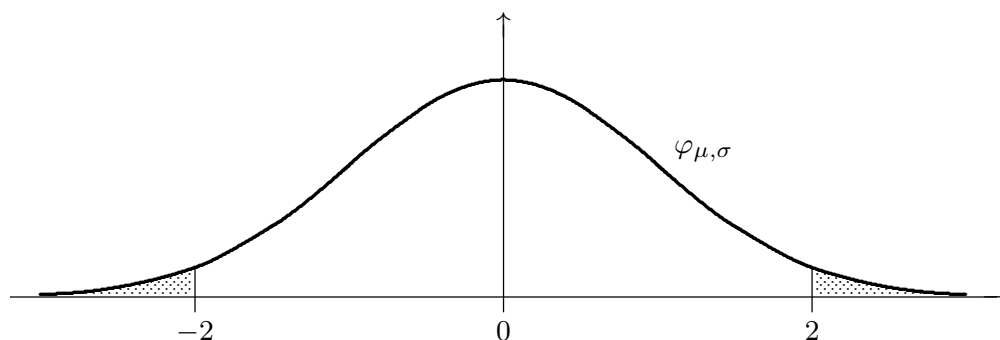
ist? Gesucht ist also  $P(|X| \geq 2)$ .

Da “ $|X| \geq 2$ ” gleichbedeutend mit “ $X \leq -2$  oder  $X \geq 2$ ” ist, gilt

$$P(|X| \geq 2) = P(X \leq -2) + P(X \geq 2) ,$$

und aus Symmetriegründen ist hier

$$P(X \leq -2) = P(X \geq 2) .$$



Nun ist  $P(X \leq -2) = \Phi(-2)$  direkt der Tabelle zu entnehmen:  $\Phi(-2) = 0.0228$ . Wie wir eben gesehen haben, ist  $P(|X| \geq 2)$  das Doppelte dieser Zahl:

$$P(|X| \geq 2) = 0.0456$$

oder ungefähr 5%. \square

## (5.10.5) Kritische Werte

Wir knüpfen unmittelbar an das Beispiel 3) an und betrachten weiterhin eine standard-normal verteilte Zufallsgrösse  $X$ . Wir haben gesehen, dass gilt

$$P(|X| \geq 2) \approx 0.05 .$$

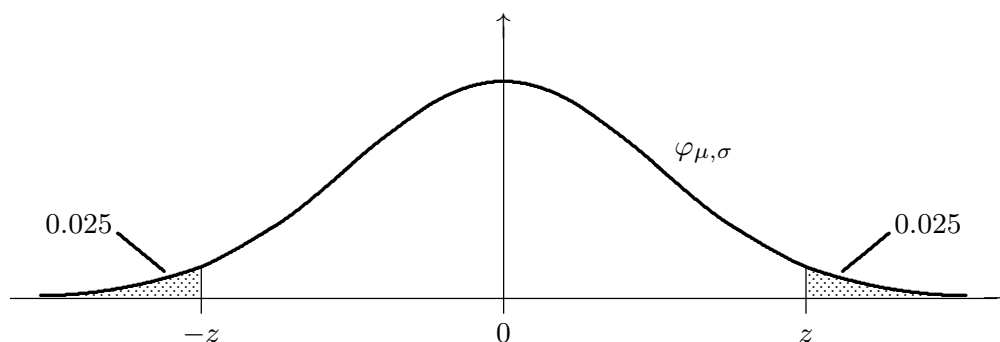
Die Wahrscheinlichkeit, dass  $X$  um mehr als 2 vom Mittelwert 0 abweicht, ist also recht klein, nämlich eben etwa 5%.

Wir fragen uns nun: Durch welche Zahl  $z$  muss die Zahl 2 im obigen Beispiel ersetzt werden, damit *exakt* gilt:

$$P(|X| \geq z) = 0.05 ?$$

Mit der üblichen Interpretation der Wahrscheinlichkeit als Flächeninhalt können wir die Frage wie folgt umformulieren:

Gesucht ist eine Zahl  $z$  so, dass der Inhalt der hervorgehobenen Flächenstücke zusammen = 0.05 ist.



Aus Symmetriegründen haben die beiden Stücke denselben Inhalt, nämlich je 0.025. Wir müssen daher  $z$  so bestimmen, dass

$$\Phi(-z) = P(X \leq -z) = 0.025$$

ist. Nun ist nach der Tabelle in (6.2.3)

$$\Phi(-2) = 0.0228 < 0.025 ,$$

$$\Phi(-1.9) = 0.0287 > 0.025 .$$

Die Zahl  $-z$  liegt also irgendwo zwischen  $-2.0$  und  $-1.9$ . Damit gilt für  $z$ :

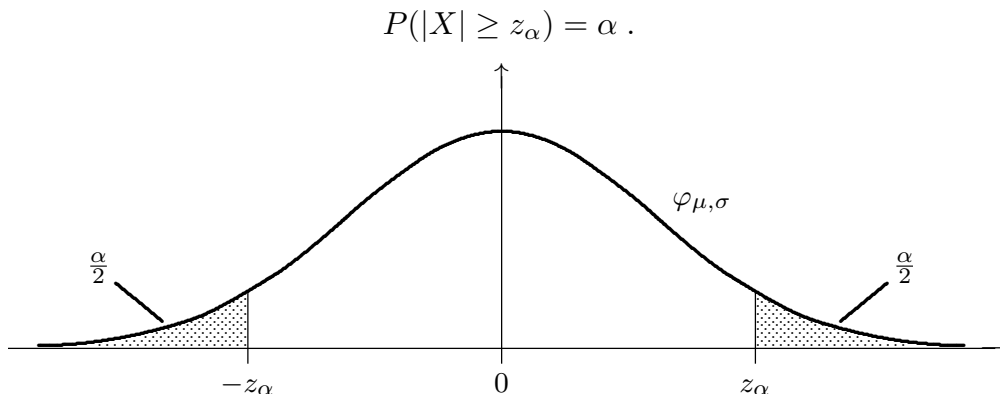
$$1.9 < z < 2.0 .$$

Genauere Berechnungen ergeben  $z = 1.959964$ . In der Praxis rundet man normalerweise auf 1.96.

Damit ist

$$P(|X| \geq 1.959964) = 0.05 .$$

Anstelle von 0.05 hätte man natürlich eine beliebige andere Wahrscheinlichkeit  $\alpha$  nehmen können. Statt  $z$  schreibt man dann  $z_\alpha$ . Diese Zahl ist also gegeben durch die Beziehung

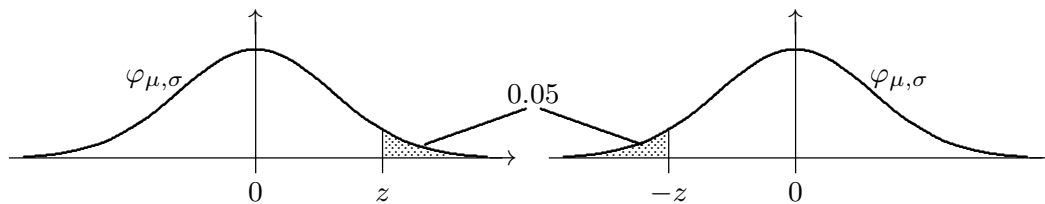


Die so definierten Zahlen  $z_\alpha$  nennen wir *kritische Werte*\*. Wir führen einige an (vgl. auch (6.2.3) für eine etwas ausführlichere Tabelle, vgl. auch in R: `qnorm(0.975)=1.959964`):

$\alpha$	0.001	0.01	0.02	0.05	0.1	0.2
$z_\alpha$	3.291	2.576	2.326	1.960	1.645	1.282

Die eingangs behandelte Fragestellung wird aus nahe liegenden Gründen “zweiseitig” genannt. Man kann das Problem auch “einseitig” betrachten und kommt dann auf folgende Fragen: Wie gross muss  $z$  gewählt werden, damit gilt

$$P(X \geq z) = 0.05 \quad \text{bzw.} \quad P(X \leq -z) = 0.05 ?$$



Aus Symmetriegründen führen beide Fragen auf denselben Wert von  $z$ . Das markierte Flächenstück hat in beiden Fällen den Inhalt 0.05. Wären wir im zweiseitigen Fall, so würde dem fraglichen  $z$  der kritische Wert  $z_{0.1}$  für  $\alpha = 0.1$  entsprechen. Nach der obigen Tabelle ist also  $z = 1.645$ ; wird auf lediglich 2 Stellen nach dem Komma gerundet (häufig), verwendet man die Zahl 1.64:

$$P(X \leq -1.645) = P(X \geq 1.645) = 0.05 .$$

\* Die Wahl des Namens erklärt sich bei der Besprechung der statistischen Tests in Kapitel 9. Für unser  $z_\alpha$  sind auch andere Bezeichnungen wie  $z_{\frac{\alpha}{2}}$  oder  $u_{1-\frac{\alpha}{2}}$  üblich.

Allgemein gilt: Der “einseitige” kritische Wert zur Wahrscheinlichkeit  $\alpha$  ist gleich dem “zweiseitigen” kritischen Wert zur Wahrscheinlichkeit  $2\alpha$ . Beachten Sie, dass sich die Tabellen gewöhnlich auf den zweiseitigen Fall beziehen.

#### Beispiel 5.10.5.A

Der Intelligenzquotient (IQ) ist so normiert, dass er einer Normalverteilung mit  $\mu = 100$  und  $\sigma = 15$  folgt (damit ist  $\sigma^2 = 225$ ). Wir entschliessen uns, die obersten 5% als “superklug” zu bezeichnen. Von welchem IQ an ist eine Person superklug? Dazu sei  $X$  die Zufallsgrösse “Intelligenzquotient”. Gesucht ist eine Zahl  $x$  mit

$$P(X > x) = 0.05 .$$

Dies ist gleichbedeutend mit

$$P(X \leq x) = \Phi_{100,225}(x) = 0.95 .$$

Standardisieren ergibt

$$\Phi_{100,225}(x) = \Phi\left(\frac{x - 100}{15}\right) = 0.95 .$$

Die Tabelle in (6.2.3), in der wir jetzt für ein gegebenes  $\Phi(z)$ , nämlich  $\Phi(z) = 0.95$  das zugehörige  $z$  suchen müssen, zeigt, dass dieses  $z$  zwischen 1.60 und 1.65 liegen muss. Einen genaueren Wert erhalten wir, wenn wir beachten, dass unser gesuchtes  $z$  gerade der kritische Wert für  $\alpha = 0.1$  ist (Vorsicht wegen einseitige versus zweiseitige Fragestellung!). Somit ist  $z = 1.645$  (siehe die kleine Aufstellung weiter oben), also

$$\frac{x - 100}{15} = 1.645 ,$$

woraus  $x = 124.675$  folgt.

⊠

#### (5.10.6) Erwartungswert, Varianz und Standardabweichung

In (5.6) bzw. (5.7) haben wir für beliebige stetige Zufallsgrössen definiert:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx, \quad V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

Im Falle der Normalverteilung  $N(\mu; \sigma^2)$  ist anstelle von  $f$  die Dichtefunktion  $\varphi_{\mu, \sigma^2}$  einzusetzen. Diese Integrale lassen sich mit einem gewissen Aufwand berechnen (siehe unten); man erhält folgende Resultate:

1. Der Erwartungswert der Normalverteilung  $N(\mu; \sigma^2)$  ist  $E(X) = \mu$ .
2. Die Varianz der Normalverteilung  $N(\mu; \sigma^2)$  ist  $V(X) = \sigma^2$ .
3. Die Standardabweichung der Normalverteilung  $N(\mu; \sigma^2)$  ist  $\sigma$ .

Die Bezeichnungen in der Formel für die Dichtefunktion sind also “geschickt” gewählt worden: Das  $\mu$  aus der Formel ist gerade der Erwartungswert, das  $\sigma$  die Standardabweichung der Zufallsgrösse  $X$ .

Wir skizzieren noch kurz, wie  $E(X)$  im Einzelnen berechnet wird; die entsprechende Rechnung für  $V(X)$  ist etwas komplizierter und soll weggelassen werden.  $E(X)$  ist gegeben durch

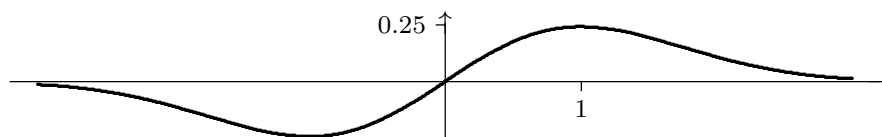
$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \varphi_{\mu, \sigma^2}(x) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx . \end{aligned}$$

(Hier wurde Formel (\*\*)) von (5.10.3) verwendet.)

Substituieren wir  $t = \frac{x - \mu}{\sigma}$ , so dass  $dt = \frac{dx}{\sigma}$  und  $x = \sigma t + \mu$  ist, so erhalten wir (wir wenden hier die Substitutionsregel auf uneigentliche Integrale an, eine detailliertere Überlegung dazu wurde im Beweis von Formel (\*) in (5.10.3) vorgeführt):

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} (\sigma t + \mu) \varphi(t) dt \\ &= \sigma \int_{-\infty}^{\infty} t \varphi(t) dt + \mu \int_{-\infty}^{\infty} \varphi(t) dt . \end{aligned}$$

Das erste Integral ist aus Symmetriegründen = 0, denn  $\varphi(t)$  ist symmetrisch zur  $y$ -Achse, woraus folgt, dass  $t\varphi(t)$  punktsymmetrisch bezüglich des Nullpunktes ist:



Diese Symmetrieüberlegung ist streng genommen nur dann richtig, wenn wir wissen, dass das uneigentliche Integral  $\int_{-\infty}^{\infty} \varphi(t) dt$  tatsächlich existiert. Dazu muss man nach (20.3), erster Band, zeigen, dass  $\int_{-\infty}^0 \varphi(t) dt$  und  $\int_0^{\infty} \varphi(t) dt$  einzeln existieren. Es genügt, das zweite Integral zu betrachten. Der konstante Faktor  $1/\sqrt{2\pi}$  ist offensichtlich irrelevant, so dass wir die Existenz von  $\int_0^{\infty} te^{-t^2/2} dt$  zeigen müssen. Mit der weitem Substitution  $u = t^2/2$ ,  $du = t dt$  folgt aber mit einfacher Rechnung

$$\int_0^{\infty} te^{-t^2/2} dt = \int_0^{\infty} e^{-u} du = 1 ,$$

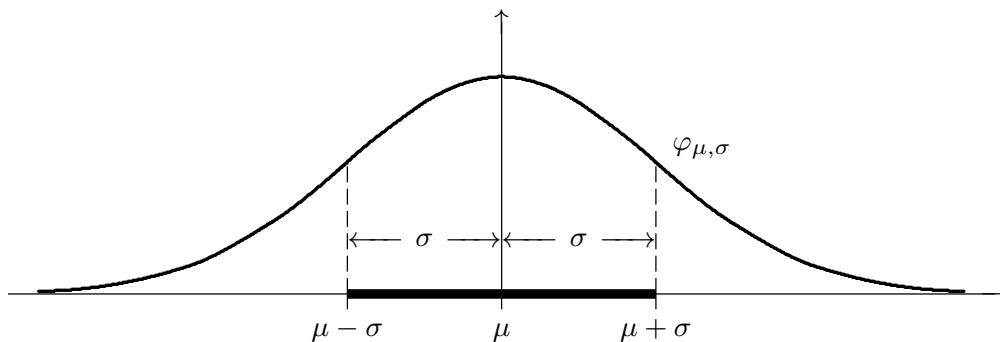
somit existieren die fraglichen uneigentlichen Integrale. (Siehe (20.3), im ersten Band, für ein ganz ähnliches Beispiel.)

Das zweite Integral hat den Wert 1, da  $\varphi$  eine Dichtefunktion ist (Bedingung 3) von (4.3.3)). Zusammen ergibt sich  $E(X) = \mu$ , was behauptet wurde.

## (5.10.7) Streuungsintervalle

Die Standardabweichung  $\sigma$  hat ja bekanntlich im Allgemeinen keine sehr anschauliche Bedeutung. Sie ist insofern ein Mass für die Streuung, als ein grosser Wert von  $\sigma$  eine grosse Streuung beschreibt; was die Zahl  $\sigma$  aber absolut gesehen bedeutet, ist nicht so unmittelbar einsichtig (vgl. jedoch (4.1.11)).

Für die Normalverteilung aber hat  $\sigma$  einen sehr konkreten Sinn, denn  $\sigma$  ist ja der Abstand der Wendepunkte von der Symmetrieachse der Glockenkurve (5.10.2) und beschreibt damit ganz direkt die “Breite” der Kurve.



Das Intervall  $[\mu - \sigma, \mu + \sigma]$  heisst das “*einfache Streuungsintervall*” (vgl. (2.2.3.8)). Wir berechnen nun die Wahrscheinlichkeit dafür, dass die Zufallsgrösse  $X$  in diesem Intervall liegt: Es ist (vgl. die eingerahmte Formel in (5.10.4) sowie die Tabelle in (6.2.3))

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= \Phi_{\mu, \sigma^2}(\mu + \sigma) - \Phi_{\mu, \sigma^2}(\mu - \sigma) \\ &= \Phi\left(\frac{(\mu + \sigma) - \mu}{\sigma}\right) - \Phi\left(\frac{(\mu - \sigma) - \mu}{\sigma}\right) \\ &= \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826 . \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass ein Wert der Zufallsgrösse im einfachen Streuungsintervall liegt, ist also  $= 0.6826$  oder  $68.26\%$  (vgl. die Bemerkung in (2.2.3.8)).

Ganz entsprechend behandelt man das doppelte und das dreifache Streuungsintervall. Man findet

$$\begin{aligned} P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 0.9545 , \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 0.9974 . \end{aligned}$$

## (5.11) Erwartungswert, Varianz und Standardabweichung von Summen

Wir haben in (5.1) die Formel  $E(aX+b) = aE(X)+b$  kennengelernt und in (5.2) die Formel  $V(aX+b) = a^2V(X)$ . Dies sind Spezialfälle der folgenden Formeln für Summen von Zufallsgrössen. In Kapitel 7 brauchen wir diese Summen und deren Kennzahlen, in

Kapitel 8 werden Summen von Zufallsgrößen detailliert eingeführt. Hier begnügen wir uns mit zwei anschaulichen Beispielen zur Motivation der allgemeinen Formeln:

1. Ein fairer Würfel hat Erwartungswert  $E(X) = 3.5$ . Wenn wir gleichzeitig 2 faire Würfel werfen, vermuten wir (zu Recht), dass der Erwartungswert der Summe der Augenzahlen gleich  $E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$  ist. Wenn wir uns weiter fragen, was der Erwartungswert vom Dreifachen der Augenzahl des ersten Würfels plus des Zweifachen der Augenzahl des zweiten Würfels ist, so vermuten wir (wieder zu Recht), dass dies gleich  $E(3X + 2Y) = 3E(X) + 2E(Y) = 3 \cdot 3.5 + 2 \cdot 3.5 = 17.5$  ist. Das folgt aus der allgemeinen Formel (auf den Beweis wird verzichtet)

$$E(aX + bY) = aE(X) + bE(Y). \quad (\text{Erwartungswert ist linear})$$

Den Spezialfall  $E(aX + b) = aE(X) + b$  von oben bekommt man, indem man für  $Y$  die konstante Zufallsgröße 1 wählt.

2. In (5.5) haben wir bereits den Erwartungswert und die Varianz der Binomialverteilung berechnet. Wir erhielten  $E(X) = np$  und  $V(X) = np(1 - p) = npq$ . Dies kann man auch folgendermassen erhalten: eine binomialverteilte Zufallsgröße  $X$  zählt die Anzahl Erfolge in  $n$  unabhängigen Versuchen mit Erfolgswahrscheinlichkeit  $p$ . Wenn wir die  $n$  Versuche als Bernoulli-Experiment, siehe (4.2), auffassen und  $X_i$  den Erfolg in Versuch  $i$  misst, also  $X_i(\omega) = 1$  bei Erfolg und  $X_i(\omega) = 0$  bei Misserfolg, so ist  $X$  die Summe der  $X_1, \dots, X_n$ :

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

Wegen der Linearität des Erwartungswerts gilt dann auch hier:

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np.$$

Allgemein gilt also

$$E(aX + bY) = aE(X) + bE(Y); \quad (\text{Erwartungswert ist linear})$$

für die Varianz gilt bei Unabhängigkeit der Zufallsgrößen (siehe Kapitel 8 für eine exakte Definition der Unabhängigkeit von Zufallsgrößen):

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$

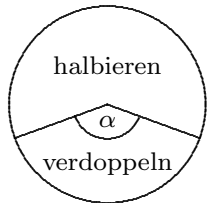
Den Spezialfall  $V(aX + b) = a^2V(X)$  von oben bekommt man, indem man für  $Y$  die konstante Zufallsgröße 1 wählt - weil dann  $V(Y) = 0!$

(5.∞) Aufgaben

5–1 Die Zufallsgrösse  $X$  ist durch die nachstehende Tabelle definiert:

$x_i$	-2	-1	1	3	4
$p_i$	0.1	0.1	0.2	0.3	0.3

- a) Geben Sie die Tabellen für die Zufallsgrössen  $Y = 2X + 1$  und  $Z = X^2$  an.
  - b) Berechnen Sie  $E(Y)$ ,  $E(Z)$ ,  $V(Y)$ ,  $V(Z)$ .
  - c) Verifizieren Sie die Gültigkeit der Formel  $V(X) = E(X^2) - (E(X))^2$  von (5.2).
- 5–2 Bei einem Würfelspiel darf man einmal würfeln, es sei denn, es sei eine Sechs gefallen. In diesem Fall hat man noch einen zweiten (und in jedem Falle letzten) Wurf zugute. Man gewinnt die totale Augenzahl in Franken. Wie gross ist der zu erwartende Gewinn? a) Bei einem ehrlichen Würfel. b) Bei einem manipulierten Würfel, wo die Eins die Wahrscheinlichkeit  $2/9$ , die Sechs eine solche von  $1/9$  hat (übrige Wahrscheinlichkeiten unverändert).
- 5–3 Für eine Prüfung müssen zwanzig Stoffgebiete vorbereitet werden, von denen dann vier zufällig ausgewählte an die Reihe kommen. Eine betroffene Person hat leider nur zehn dieser Gebiete gelernt. Die Zufallsgrösse  $G$  (für gelernt [oder für Glück]) gibt an, wieviele der vier Fragen aus einem gelernten Gebiet stammen. a) Geben Sie die Verteilung von  $G$  in Tabellenform an. b) Wie gross ist der Erwartungswert  $E(G)$ ? Was bedeutet er anschaulich?
- 5–4 Ein (nicht sehr faires) Glücksspiel hat eine Gewinnwahrscheinlichkeit von 25% und offeriert im Gewinnfall eine Auszahlung von Fr. 5.–, wobei der Einsatz Fr. 2.50 beträgt. Ich spiele so oft, bis ich zum ersten Mal gewonnen oder bis ich zehn Franken verloren habe. Wie gross ist der zu erwartende Reingewinn?
- 5–5 Ein Glücksspiel: Durch Drehen des Glücksrades wird entschieden, ob Ihr Kapital verdoppelt oder halbiert wird. Dies wird zweimal durchgeführt, wobei Sie mit einem Einsatz von Fr. 8.– beginnen. Wie gross muss der Winkel  $\alpha$  sein, damit das Spiel fair ist, d.h., damit der Erwartungswert für das Kapital am Schluss gleich dem Einsatz ist?



5–6 Ein rot lackierter Holzwürfel mit der Kantenlänge  $n$  cm ( $n = 2, 3, 4, \dots$ ) wird in  $n^3$  Würfelchen mit der Kantenlänge 1 cm zersägt. Unter diesen Würfelchen wird eines zufällig ausgewählt. Die Zufallsgrösse  $X_n$  bezeichnet die Anzahl roter Seitenflächen dieses Würfelchens. Wie gross ist der Erwartungswert von  $X_n$ ?

5–7 Ein Konditor stellt üppige Crème-Torten her, die er aber gleichentags verkaufen muss. Die Herstellung kostet ihn Fr. 8.–, der Verkaufspreis beträgt Fr. 20.–. Mehr als vier Torten pro Tag sind noch nie verlangt worden. Aus Erfahrung kennt er die folgenden Wahrscheinlichkeiten:

$k$	0	1	2	3	4
Wahrscheinlichkeit dafür, dass $k$ Torten verlangt werden	0.05	0.25	0.40	0.20	0.10

Wieviel Torten muss er pro Tag produzieren, damit sein Gewinn (auf das Crème-Torten-Geschäft bezogen) möglichst gross wird?

5–8 Eine Person hat drei (zumindest im Dunkeln, und in dieser Situation spielt unsere Geschichte) nicht unterscheidbare Schlüssel an ihrem Schlüsselring. Sie probiert die Schlüssel, bis der passende gefunden ist. Die Zufallsgrösse  $X$  stellt die Anzahl der dazu nötigen Versuche dar. Berechnen Sie den Erwartungswert von  $X$  in den folgenden Fällen: a) Ein Schlüssel, der erfolglos ausprobiert wurde, wird nicht wieder verwendet. b) Die Person ist nicht (mehr) in der Lage, sich zu merken, welche Schlüssel schon getestet wurden und probiert völlig zufällig (aber

- hartnäckig). Tipp: Zur Bestimmung der Summe der auftretenden unendlichen Reihe können Sie sich in (19.6) im ersten Band umsehen.
- 5–9 Eine Münze wird so lange geworfen, bis “Kopf” oben liegt. Die Zufallsgrösse  $X$  beschreibt die Anzahl der notwendigen Würfe. So bedeutet  $X = 1$ , dass gleich beim ersten Wurf “Kopf” gefallen ist etc. Wie gross ist der Erwartungswert von  $X$ ? (Wie in Aufgabe 5–8 kann auch hier ein Blick in (19.6) im ersten Band hilfreich sein.)
- 5–10 Die Zufallsgrösse  $X$  nehme die abzählbar unendlich vielen Werte  $k = 1, 2, 3, \dots$  mit den Wahrscheinlichkeiten  $P(X = k) = 1/k(k + 1)$  an. Zeigen Sie, dass der Erwartungswert von  $X$  nicht existiert.
- 5–11 a) Berechnen Sie Erwartungswert, Varianz und Standardabweichung für die Binomialverteilung mit  $n = 50$ ,  $p = 0.2$ . b) Eine Binomialverteilung hat den Erwartungswert 80 und die Standardabweichung 8. Wie gross sind  $n, p$  und  $q$ ?
- 5–12 Untersuchen Sie in der Situation von Aufgabe 4–23, ob  $X$  einen Erwartungswert hat. Wenn ja, wie gross ist er?
- 5–13 Untersuchen Sie in der Situation von Aufgabe 4–24, ob  $X$  einen Erwartungswert hat. Wenn ja, wie gross ist er?
- 5–14 Berechnen Sie in der Situation von Aufgabe 4–28  $E(X)$ .
- 5–15 Zeichnen Sie im selben Koordinatensystem die Graphen (inkl. Extrema und Wendepunkte) der Dichtefunktion der folgenden Normalverteilungen: a)  $N(-1; 0.09)$ , b)  $N(-1; 0.36)$ , c)  $N(1; 0.36)$ .
- 5–16 a) Skizzieren Sie den Graphen der Dichtefunktion von  $N(50; 25)$  (wählen Sie die Einheiten auf den beiden Achsen vernünftig), und stellen Sie die Wahrscheinlichkeit des Ereignisses  $48 \leq X \leq 53$  geometrisch dar. b) Berechnen Sie diese Wahrscheinlichkeit.
- 5–17 Berechnen Sie (Interpolation wo nötig):
- $P(X \geq 999)$  für  $N(1000; 100)$ .
  - $P(1.3 \leq X \leq 1.4)$  für  $N(1; 0.16)$ .
  - $P(|X| < 2)$  für  $N(0.5; 16)$ .
  - Eine normal verteilte Zufallsgrösse  $X$  hat Erwartungswert 50 und Varianz 4. Berechnen Sie  $P(X > 52.5)$ .
- 5–18 Die Zufallsgrösse  $X$  sei normal verteilt mit  $\mu = 100$ ,  $\sigma = 4$ . Gesucht ist die Zahl  $z$  mit  $P(97 \leq X \leq z) = 0.6$ .
- Skizze anhand des Graphen.
  - Bestimmen Sie dieses  $z$  so gut wie möglich mit der Tabelle.
- 5–19 Wir nehmen an, die Gewicht (in g) von Eiern einer bestimmten Singvogelart sei normal verteilt mit  $\mu = 2.1$ ,  $\sigma = 0.2$ . Es werden 100 Eier untersucht. Wieviele davon erwarten Sie a) mit einem Gewicht von  $\leq 2.15$  g, b) mit einer Gewicht zwischen 1.9 und 2.3 g?
- 5–20 Das Körpergewicht einer gewissen Sorte von Menschen sei normal verteilt mit  $\mu = 80$  kg,  $\sigma = 10$  kg. Die schwersten 10% müssen ein Sondertraining absolvieren. Bei welchem Gewicht ist die Grenze festzusetzen?
- 5–21 Von einer Population von Menschen kennt man die durchschnittliche Körpergrösse; sie beträgt 170 cm. Wir nehmen an, die Körpergrösse sei normalverteilt. Ferner weiss man, dass 80% dieser Menschen zwischen 160 cm und 180 cm gross sind. a) Bestimmen Sie die Parameter  $\mu$  und  $\sigma$  dieser Normalverteilung. b) Wie gross ist die Wahrscheinlichkeit dafür, dass eine zufällig aus dieser Gruppe ausgewählte Person zwischen 175 cm und 185 cm gross ist? c) Mit welcher Wahrscheinlichkeit ist eine Person aus dieser Population kleiner als 180 cm?
- 5–22 Eine Maschine stellt Tonkugelchen her, deren Durchmesser normal verteilt ist. Diese Kugelchen werden zuerst mit einem Sieb der Maschenweite 8 mm gesiebt; dabei fallen 15% der Produktion

- durch (wörtlich!). Der Rest wird noch mit einem Sieb von 11 mm Maschenweite geprüft, wobei 40% des Restes im Sieb zurückbleiben. Bestimmen Sie den mittleren Durchmesser aller Kügelchen.
- 5–23 Bei der Untersuchung einer gewissen Population hat man festgestellt, dass 33% der Personen ein Gewicht von  $\leq 55$  kg und 5% der Personen ein solches von  $> 70$  kg haben. Wir nehmen an, die Zufallsgrösse  $X =$  Körpergewicht sei normal verteilt. a) Berechnen Sie Erwartungswert  $\mu$  und Standardabweichung  $\sigma$  von  $X$  mit einer Genauigkeit von zwei Stellen nach dem Dezimalpunkt. b) Mit welcher Wahrscheinlichkeit liegt das Gewicht einer zufällig aus dieser Gruppe herausgegriffenen Person zwischen 57 kg und 64 kg? c) Bestimmen Sie die Zahl  $z$  so, dass genau 25% aller Personen aus der Population mehr als  $z$  kg wiegen.
- 5–24 Der Intelligenzquotient ist so normiert, dass er einer Normalverteilung mit  $\mu = 100$ ,  $\sigma = 15$  folgt. Wie gross ist die Wahrscheinlichkeit dafür, dass unter 5 Personen a) genau zwei, b) mindestens zwei einen IQ  $> 130$  haben?
- 5–25 Eine Anlage produziert Nägel, bei denen die Abweichung  $X$  vom Sollmass normal verteilt ist mit  $E(X) = 0$  mm und  $V(X) = 0.04$  mm<sup>2</sup>. Ich kaufe eine Schachtel mit 100 Nägeln. Wie gross ist die Wahrscheinlichkeit dafür, dass höchstens 3 davon um mehr als 0.3 mm zu lang sind?
- 5–26 Zwei Bäckereien A und B backen Brote, deren Gewicht normal verteilt ist, mit  $\mu = 250$  g. Die Standardabweichung des Gewichts beträgt bei A 15 g, bei B aber 25 g. a) Eine Person kauft bei A, eine andere, unabhängig davon, bei B ein Brot. Mit welcher Wahrscheinlichkeit wiegt jedes der beiden Brote mehr als 265 g? b) In einer Verkaufsstelle stammen 40% der Brote von A, der Rest von B. Die Brote sind nicht nach ihrer Herkunft unterscheidbar. Jemand kauft zufällig ein Brot. Mit welcher Wahrscheinlichkeit ist es leichter als 238 g? c) In der Situation von b) habe ich ein Brot gekauft, das leichter als 238 g ist. Mit welcher Wahrscheinlichkeit wurde es in der Bäckerei A gebacken?
- 5–27 Ein Taschenrechnermodell verfügt über drei Tasten, die mit  $P(t)$ ,  $Q(t)$ ,  $R(t)$  beschriftet sind. Ich habe leider die Betriebsanleitung verlegt, weiss aber noch, dass diese Tasten etwas mit der Standardnormalverteilung zu tun haben. Wenn ich jeweils 0.1 eingebe, erhalte ich mit der Taste  $P(t)$  den Wert 0.5398, mit  $Q(t)$  den Wert 0.4602 und mit  $R(t)$  schliesslich den Wert 0.0398. Können Sie die Bedeutung der drei Tasten erläutern? Was erhalten Sie, wenn Sie jeweils 0.2 eingeben?