1.5.4 Quantile und Modi

Bem. 1.70. [Quantil, Modus]

Analog zu Statistik I kann man auch Quantile und Modi definieren.
Gegeben sei eine Zufallsvariable X mit Wahrscheinlichkeitsverteilung P_X und Verteilungsfunktion F_X.

a) Für $\alpha \in (0,1)$ heißt x_α zugehöriges α-*Quantil* genau dann, wenn $F(x_\alpha = \alpha)$. Insbesondere ergeben sich für α=0.5 der *Median* und für α=0.25 bzw. 0.75 das *25%* bzw. *75% Quantil*.

b) Ist X diskret, so heißt jedes x_{mod} mit

$$P(X = x_{mod}) \geq P(X = x), \text{ für alle } x \in \mathcal{X}$$

Modus von P_X.

Ist X stetig mit Dichtefunktion f_X, so heißt jedes x_{mod} mit $f_X(x_{\text{mod}}) \geq f_X(x)$, für alle $x \in \mathcal{X}$ Modus von P_X (bezüglich der Dichte f_X).

Bem. 1.71. [Modalbereiche]

Ist der Modus x_{mod} eindeutig und X diskret, so ist dies derjenige Punkt, der die höchste Wahrscheinlichkeit besitzt, einzutreten.
Dies kann man auf "zusammenhängende" Bereiche mit vorgegebener Sicherheit ausweiten:
Sei $\gamma \in (0,1)$. Ein Intervall $A_\gamma \subset \mathbb{R}$, $A_\gamma = [a,b]$, heiße *Modalintervall zum Niveau* γ, wenn gilt

$$P(X \in A_\gamma) \geq \gamma$$

und $P(X \in A_\gamma) \geq P(X \in C)$ für alle Intervalle $C \subset \mathbb{R}$, $C = [c,d]$ mit $d - c \leq b - a$.
Dann ist $A_\gamma \cap \mathcal{X}$ kleinster ("präzisester") zusammengehöriger Bereich, der mindestens Wahrscheinlichkeit γ besitzt.
Ist X diskret, so ist $A_\gamma \cap \mathcal{X}$ ein Bereich benachbarter Punkte, bei stetigem X mit Träger $\mathbb{R}$ ist A_γ ein Intervall.

1.6 Wichtige Verteilungsmodelle

Wir behandeln hier nur die Binomial-, die Poisson- und die Normalverteilung. Einige weitere Verteilungsmodelle werden direkt dort eingeführt, wo sie benötigt werden.

1.6.1 Binomialverteilung

Konstruktionsprinzip:

- Ein Zufallsexperiment wird n mal unabhängig durchgeführt.
- Wir interessieren uns jeweils nur, ob ein bestimmtes Ereignis A eintritt oder nicht.
- $X =$ „absolute Häufigkeit, mit der Ereignis A bei n unabhängigen Versuchen eintritt“.
- Träger von X: $\mathcal{X} = \{0, 1, 2, \ldots, n\}$.

Herleitung der Wahrscheinlichkeitsfunktion:

- Bezeichne $\pi = P(A)$ die Wahrscheinlichkeit für A in $einem$ Experiment.
- Das Ereignis $\{X = x\}$ tritt z.B. auf, wenn in den ersten x Versuchen A eintritt und anschließend nicht mehr. Die Wahrscheinlichkeit dafür ist

$$\begin{aligned} P(A_1 \cap \ldots \cap A_x \cap \bar{A}_{x+1} \cap \ldots \cap \bar{A}_n) &= \underbrace{\pi \cdot \ldots \cdot \pi}_{x \text{ mal}} \cdot \underbrace{(1-\pi) \cdot \ldots \cdot (1-\pi)}_{n-x \text{ mal}} \\ &= \pi^x (1-\pi)^{n-x}. \end{aligned}$$

- Diese Überlegung gilt analog für jede Auswahl von x "Plätzen". Insgesamt gibt es $\binom{n}{x}$ Möglichkeiten für die Verteilung der x Erfolge (Auftreten von A) auf n Plätze. Damit gilt:

$$P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}.$$

Definition 1.72.

Eine Zufallsvariable heißt *binomialverteilt mit dem Parameter π bei n Versuchen*, kurz $X \sim B(n, \pi)$, wenn sie die Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Die $B(1, \pi)$-Verteilung heißt auch *Bernoulliverteilung*.

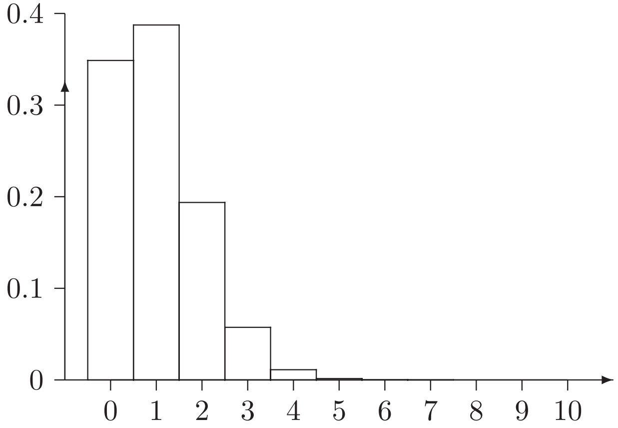
Zum Begiff *Parameter*:

Dies ist eine Größe, die die Wahrscheinlichkeitsverteilung „steuert“.

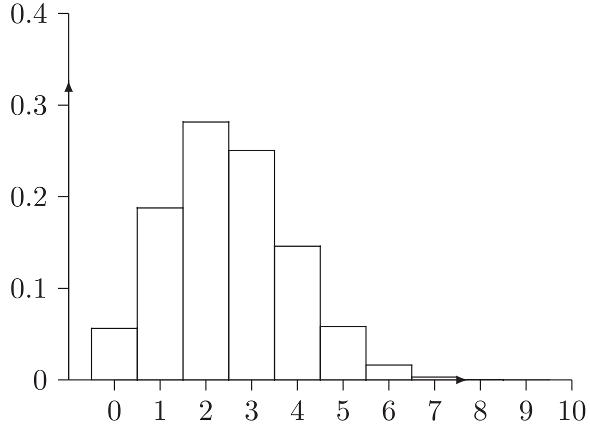
Ist der Parameter bekannt, so kennt man –wenn man die Modellklasse als bekannt voraussetzt– die Wahrscheinlichkeitsverteilung. Später, in der induktiven Statistik, spricht man dann von „parametrischen Modellen“, d.h. der Verteilungstyp wird als bekannt/gegeben („Verteilungsannahme“) vorausgesetzt (z.B. dadurch, dass man durch die Versuchsanordnung mit n unabhängigen Versuchen ein Binomialmodell erzeugt), der konkrete Wert des Parameters bzw. allgemeine Aussagen über seinen Wert sind dann anhand der Stichprobe zu bestimmen.

"Wahrscheinlichkeitshistogramme" von Binomialverteilungen mit $n = 10$: (graphische Darstellung der Wahrscheinlichkeitsfunktion)

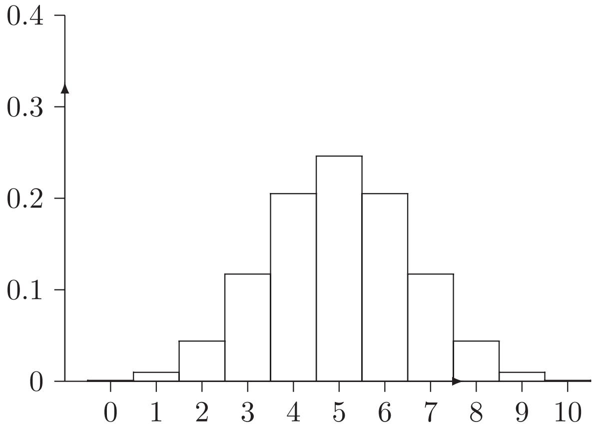
$\pi = 0.1$



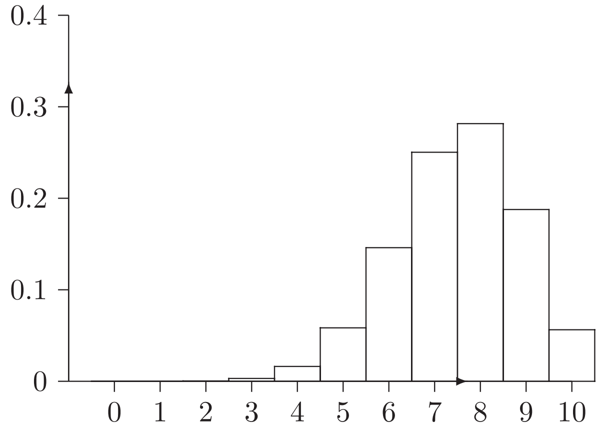
$\pi = 0.25$



$\pi = 0.5$



$\pi = 0.75$



Erwartungswert und Varianz:

- Zur Berechnung von Erwartungswert und Varianz der Binomialverteilung ist folgende Darstellung hilfreich:

$$X = X_1 + \ldots + X_n$$

  mit den binären Variablen

$$X_i = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst.} \end{cases}$$

- Die X_i sind stochastisch unabhängig mit

$$\begin{aligned} E(X_i) &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = \pi \\ \mathrm{Var}(X_i) &= E(X_i^2) - (E(X_i))^2 = 1 \cdot P(X_i = 1) - \pi^2 = \pi - \pi^2 = \pi(1 - \pi). \end{aligned}$$

- Erwartungswert der Binomialverteilung:

$$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = n\pi$$

Die direkte Berechnung über

$$E(X) = \sum_{i=1}^{n} i \cdot P(\{X = i\}) = \sum_{i=1}^{n} i \binom{n}{i} \pi^i (1-\pi)^{n-i} = \ldots = n\pi$$

ist deutlich komplizierter!

- Varianz der Binomialverteilung:

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = n\pi(1-\pi)$$

Bsp. 1.73.

Risikobereite Slolalomfahrer stürzen mit Wahrscheinlichkeit 10%, vorsichtigere mit 2%.

a) Schlagen Sie ein Modell für diese Situation vor und diskutieren Sie kurz die zugrunde gelegten Annahmen.

b) Wie groß sind jeweils die Wahrscheinlichkeiten, dass von 20 Fahrern mindestens einer stürzt?

c) Vergleichen Sie die durchschnittlich zu erwartende Anzahl von Stürzen von je 100 Rennläufern!

Exkurs: Zur Problematik der Argumentation mittels „natürlicher Häufigkeiten“ (vgl. v.a. Kap 1.2 und Kap 1.3):
Es wurde wiederholt vorgeschlagen, Wahrscheinlichkeiten anschaulich über „natürliche Häufigkeiten“ zu kommunizieren, also z. B. $P(A) = 0.3753$ darstellen als „von 10000 Personen haben 3753 die Eigenschaft A“.
Man würde demgemäß die Wahrscheinlichkeit $\pi_r = 0.1$ kommunizieren als „von 100 stürzen 10 Rennläufer“.

Diese Darstellung läuft Gefahr, die beträchtliche Variabilität zufälliger Prozesse zu verschleiern. In der Tat ist hier die Wahrscheinlichkeit, dass genau 10 von 100 Läufern stürzen,

$$P(X = 10) = \binom{100}{10} \cdot 0.1^{10} \cdot 0.9^{90} \approx 0.13,$$

also lediglich etwa 13%. „Natürliche Häufigkeiten“ müssen also unbedingt als Durchschnittswerte bzw. Erwartungswerte begriffen und kommuniziert werden.

Eigenschaften der Binomialverteilung:

- Symmetrieeigenschaft (vertausche Rolle von A und $\bar{A}$):

- Summeneigenschaft: Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$. Sind X und Y unabhängig, so gilt

$$X + Y \sim$$

Tabellierung der Binomialverteilung: Tabelliert ist oft $P(X \leq x)$

$\pi = 0.3$	$n =$11	$n =$12	...
$x \leq 0$	0.0198	0.0138	
1	0.1130	0.0850	...
2	0.3127	0.2528	
$\vdots$	$\vdots$	$\vdots$	

Daraus lassen sich die interessierenden Wahrscheinlichkeiten ablesen:

$$P(X = x) = P(X \leq x) - P(X \leq x - 1), \quad x \in \mathbb{N}_0$$

Zum Beispiel:

$$\begin{aligned} P(X = 2) &= P(X \leq 2) - P(X \leq 1) = \\ &= 0.1997 \\ \Big(&= \binom{11}{2} \cdot 0.3^2 \cdot 0.7^2. \Big) \end{aligned}$$

Wegen der Symmetrieeigenschaft gibt es meist nur Tabellen für $\pi \leq 0.5$.

Für großes n verwendet man Approximationen durch die Normalverteilung (vgl. Abschnitt 1.7.4).

1.6.2 Poisson Verteilung

Eine weitere wichtige diskrete Verteilung ist die Poisson-Verteilung. Sie modelliert die Anzahl (eher seltener) Ereignisse in einem Zeitintervall (Unfälle, Todesfälle; Sozialkontakte, deviante Verhaltensmuster, etc.).
Schreibweise im folgenden $\exp(a) := e^a$ mit e als Eulersche Zahl, $e = 2,718\ldots$.

Definition 1.74. [Poisson-Verteilung]

Eine Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} \exp(-\lambda), & x \in \{0, 1, \ldots\} \\ 0, & \text{sonst} \end{cases}$$

heißt *Poisson-verteilt* mit *Parameter (oder Rate)* $\lambda > 0$, kurz $X \sim Po(\lambda)$. Es gilt

$$E(X) = \lambda, \quad \mathrm{Var}(X) = \lambda$$

Bem. 1.75.

Die Poisson-Verteilung kann auch als Näherungsmodell für eine Binomialverteilung gesehen werden, wenn die Anzahl der Versuchswiederholungen n groß und die „Trefferwahrscheinlichkeit“ π sehr klein ist (seltene Ereignisse!).
Der Erwartungswert λ ist dann gleich $n \cdot \pi$.

Es gilt also abgekürzt geschrieben

$$X \sim B(n, \pi) \underset{\substack{n \text{ groß} \\ \pi \text{ klein}}}{\Longrightarrow} X \approx Po(n \cdot \pi)$$

Hat man mehrere unabhängige „Poisson-Prozesse“, also dynamische Situationen, bei denen die Ereignisanzahl Poisson-verteilt ist, also z.B. deviante Verhaltensmuster in verschiedenen Populationen, so ist die Gesamtanzahl der einzelnen Ereignisanzahlen wieder Poisson-verteilt, genauer gilt:

Satz 1.76. [Addition von Poisson-verteilten Zufallsvariablen]

Sind $X \sim Po(\lambda_X), Y \sim Po(\lambda_Y)$ voneinander unabhängig, so gilt

$$X + Y \sim Po(\lambda_X + \lambda_Y).$$

Beachte, die Unabhängigkeit ist wesentlich. Nimmt man als Extremfall zwei Ereignisse, bei denen das eine das andere voraussetzt (Scheidungen, Prozesse um das Sorgerecht für Kinder in derselben Population), so ist die Gesamtzahl nicht mehr Poisson-verteilt.

Da bei der Poisson-Verteilung Erwartungswert und Varianz identisch sind, müsste nämlich dann gelten, wenn $X + Y$ Poisson-verteilt wäre:

$$\mathrm{Var}(X + Y) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mathrm{Var}(X) + \mathrm{Var}(Y),$$

was aber bei abhängigen X und Y im Allgemeinen verletzt ist.

Bsp. 1.77.

Max geht gerne auf Open-Air Festivals. Im Durchschnitt trifft er dort 6 weibliche Bekannte und 3 männliche Bekannte.

a) Seien X und Y zwei Zufallsvariablen, die die Anzahl der getroffenen weiblichen Bekannten und die Anzahl der getroffenen männlichen Bekannten angeben. Formulieren Sie ein geeignetes Modell zur Beschreibung der Wahrscheinlichkeit von X und Y.

b) Wie groß ist die Wahrscheinlichkeit, dass Max genau 6 weibliche Bekannte trifft?

c) Wie groß ist die Wahrscheinlichkeit, dass Max mindestens einen männlichen Bekannten trifft?

d) Berechnen Sie die Wahrscheinlichkeit, dass Max weder einen männlichen Bekannten noch eine weibliche Bekannte trifft, auf 2 verschiedene Arten. Diskutieren Sie eventuell zu treffende Zusatzannahmen.

1.6.3 Normalverteilung

Die Normalverteilung ist wohl das wichtigste Verteilungsmodell der Statistik, denn

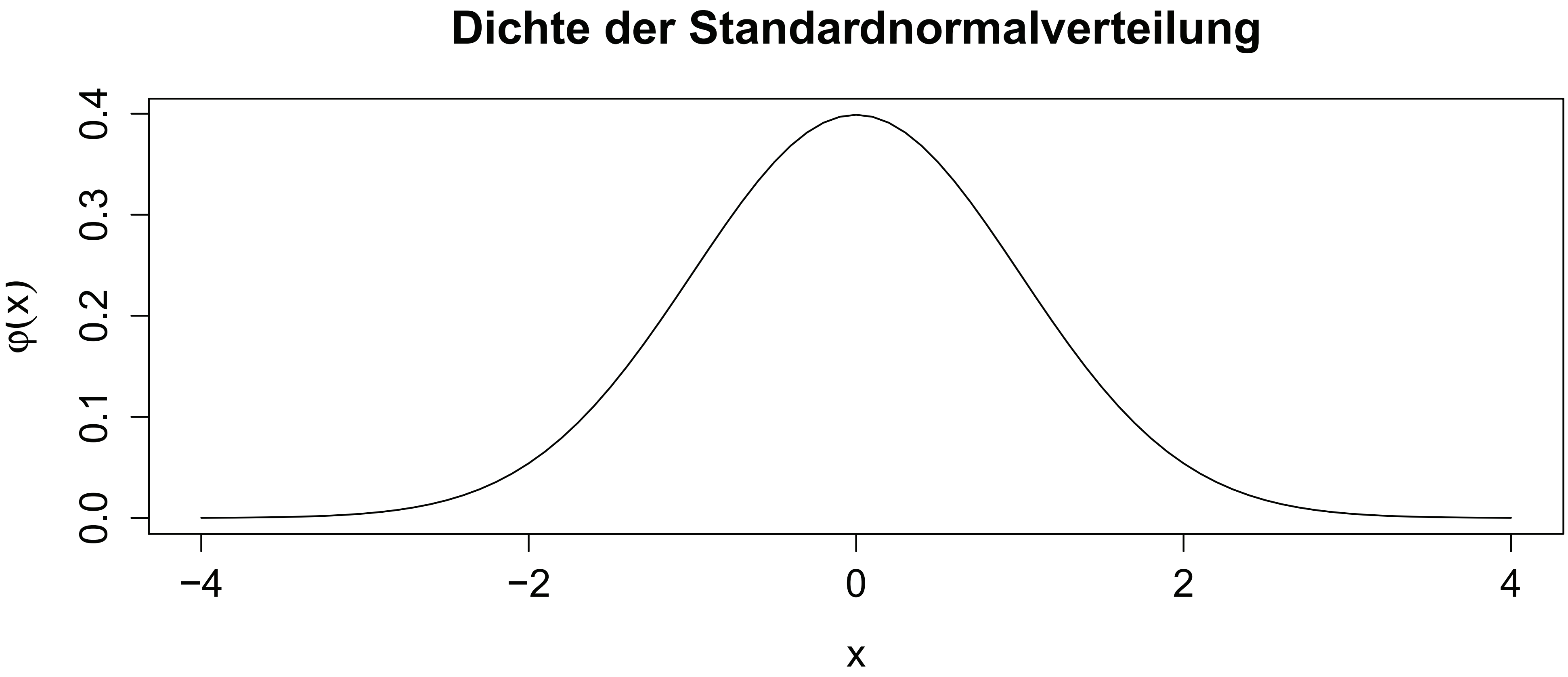
- viele Zufallsvariablen sind (nach einer geeigneten Transformation) (ungefähr) normalverteilt.
- beim Zusammenwirken vieler zufälliger Einflüsse ist der geeignet aggregierte Gesamteffekt oft approximativ normalverteilt (Zentraler Grenzwertsatz, Kap. 1.7).
- die asymptotische Grenzverteilung, also die Verteilung bei unendlich großem Stichprobenumfang, typischer statistischer Größen ist die Normalverteilung.

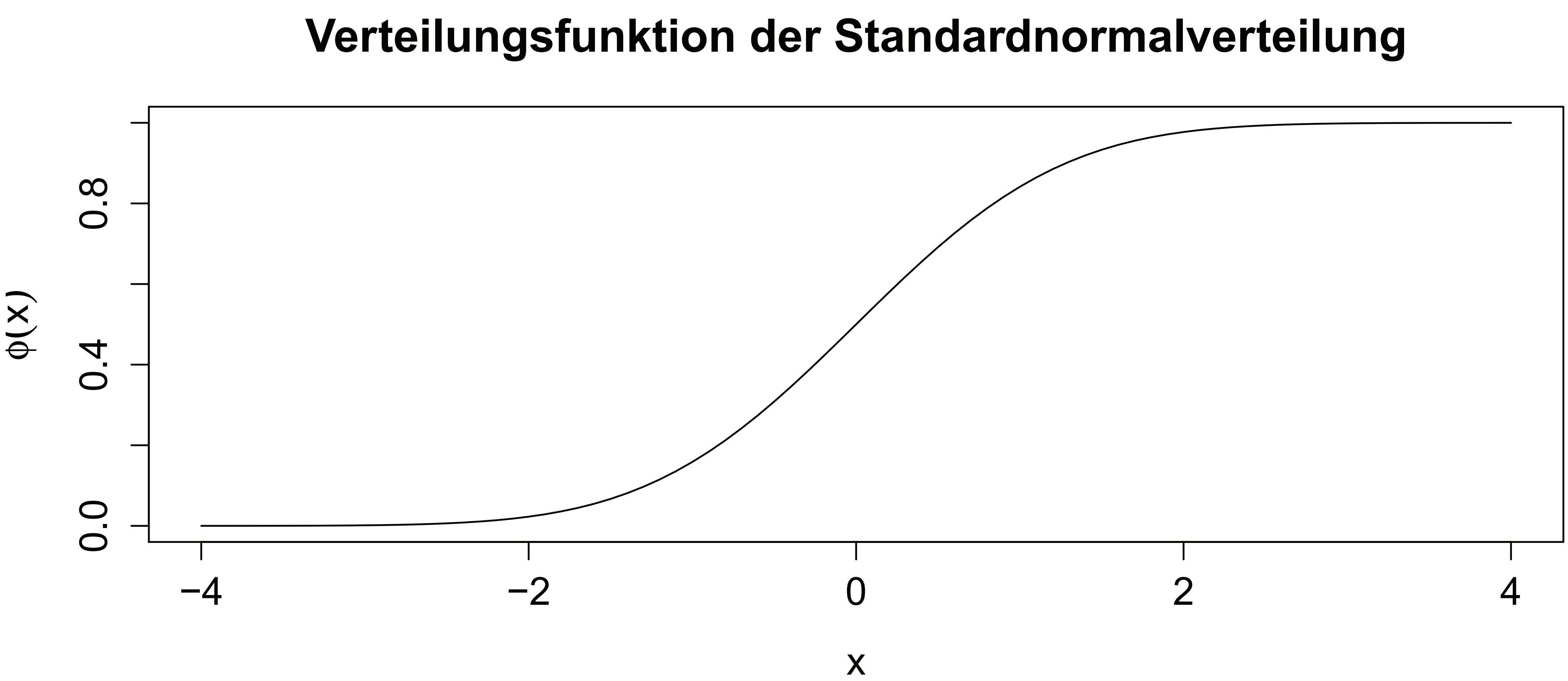
Definition 1.78.

Eine stetige Zufallsvariable X heißt *normalverteilt* mit den Parametern μ und σ^2, in Zeichen $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \ x \in \mathbb{R} \tag{1.9}$$

und *standardnormalverteilt*, in Zeichen $X \sim \mathcal{N}(0; 1)$, falls $\mu = 0$ und $\sigma^2 = 1$ gilt (π ist hier die Kreiszahl $\pi = 3.14\ldots$, $\exp$ ist wieder die „e-Funktion“: $\exp(a) := e^a$).


Dichte der Standardnormalverteilung
φ(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x


Verteilungsfunktion der Standardnormalverteilung
φ(x)
0.0
0.4
0.8
−4
−2
0
2
4
x

Grundlegende Eigenschaften:

a) Die Dichte ist symmetrisch um μ, d.h.

$$f(\mu - x) \quad = \quad f(\mu + x)\,.$$

Man kann zeigen: Je größer σ^2, desto flacher verläuft die Kurve.

b) Die *Dichte der Standardnormalverteilung* wird oft mit $\varphi(x)$ bezeichnet, also

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

und die zugehörige Verteilungsfunktion mit

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u) du.$$

c) $\Phi(x)$ lässt sich nicht in geschlossener Form durch elementare Funktionen beschreiben $\Longrightarrow$ numerische Berechnung, Tabellierung.

d) μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim \mathcal{N}\ \mu, \sigma^2)$, dann

$$\mathrm{E}(X) = \mu \quad \text{und} \quad \mathrm{Var}(X) = \sigma^2.$$

Grundlegendes zum Rechnen mit Normalverteilungen:

- Es gilt:

$$\Phi(-x) = 1 - \Phi(x)$$

  (folgt aus der Symmetrie der Dichte). Für die Tabellierung reicht es also $x \geq 0$ zu betrachten.

- Gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die zugehörige standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma}$$

  standardnormalverteilt.

- Entscheidende Eigenschaft für die Tabellierung: Es reicht, die Standardnormalverteilung zu tabellieren. Normalverteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 muss man, wie unten erläutert, zuerst standardisieren, dann kann man auch die Standardnormalverteilungstabelle verwenden.

- Tabelliert sind die Werte der Verteilungsfunktion $\Phi(x) = P(X \leq x)$ für $x \geq 0$.

  Ablesebeispiel: $\Phi(1.75) =$

- Funktionswerte für negative Argumente: $\Phi(-x) = 1 - \Phi(x)$

	0.00	0.01	$\cdots$	0.05	$\cdots$	0.09
$\vdots$						
1.5	0.9332	0.9345	$\cdot$	0.9394	$\cdot$	0.9441
1.6	0.9452	0.9463	$\cdot$	0.9505	$\cdot$	0.9545
1.7	0.9554	0.9564	$\cdot$	0.9599	$\cdot$	0.9633
1.8	0.9641	0.9649	$\cdot$	0.9678	$\cdot$	0.9706
1.9	0.9713	0.9719	$\cdot$	0.9744	$\cdot$	0.9767
2.0	0.9772	0.9778	$\cdot$	0.9798	$\cdot$	0.9817
$\vdots$						

Berechnung bei „allgemeiner Normalverteilung“: Wie bestimmt man bei $X \sim \mathcal{N}(\mu, \sigma^2)$ die Wahrscheinlichkeiten $P(X \leq a)$ aus der Tabelle der Standardnormalverteilung?

Bem. 1.79. [Abgeschlossenheit gegenüber Linearkombinationen]

Seien X_1 und X_2 unabhängig und $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2),\ i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt:

$$
\begin{aligned}
Y_1 &:= a_1 X_1 + b \sim \mathcal{N}(a_1\mu_1 + b; a_1^2\sigma_1^2) \\
Y_2 &:= a_1 X_1 + a_2 X_2 \sim \mathcal{N}(a_1\mu_1 + a_2\mu_2; a_1^2\sigma_1^2 + a_2^2\sigma_2^2).
\end{aligned}
$$

Das Ergebnis lässt sich auf mehrere Summanden verallgemeinern.

Bsp. 1.80. [aus Fahrmeir et al.]

- Schultischhöhe: $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)\,, \quad \mu_Y = 113\,, \quad \sigma_Y^2 = 16$
  Stuhlhöhe: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)\,, \quad \mu_X = 83\,, \quad \sigma_X^2 = 25$

- X und Y unabhängig

- optimale Sitzposition: Tisch zwischen 27 und 29 cm höher als Stuhl.

- Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Paar zueinander gut passt?

Differenz: $Y - X$ soll zwischen $[27, 29]$ sein.
Definiere also

$$V := Y - X = Y + (-X)$$

Wegen $-X \sim \mathcal{N}(-83, 25)$ gilt dann

$$V \sim \mathcal{N}(113 - 83, 16 + 25) = \mathcal{N}(30, 41).$$

Außerdem ergibt sich durch Standardisieren:

$$
\begin{aligned}
27 \leq V \leq 29 \quad &\iff \quad 27 - 30 \leq V - 30 \leq 29 - 30 \\
&\iff \quad \frac{27 - 30}{\sqrt{41}} \leq \frac{V - 30}{\sqrt{41}} \leq \frac{29 - 30}{\sqrt{41}}
\end{aligned}
$$

Damit lässt sich die gesuchte Wahrscheinlichkeit bestimmen:

$$
\begin{aligned}
P(27 \leq V \leq 29) \quad &= \quad P\left(-0.469 \leq \frac{V - 30}{\sqrt{41}} \leq -0.156\right) = \\
&= \quad \Phi(-0.156) - \Phi(-0.469) = \\
&= \quad (1 - \Phi(0.156)) - (1 - \Phi(0.469)) = \\
&= \quad -0.5636 + 0.6808 = 0.1172
\end{aligned}
$$