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 1.77. [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. 1.78.

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 1.79. [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. 1.80.

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. **??**).
- die asymptotische Grenzverteilung, also die Verteilung bei unendlich großem Stichprobenumfang, typischer statistischer Größen ist die Normalverteilung.

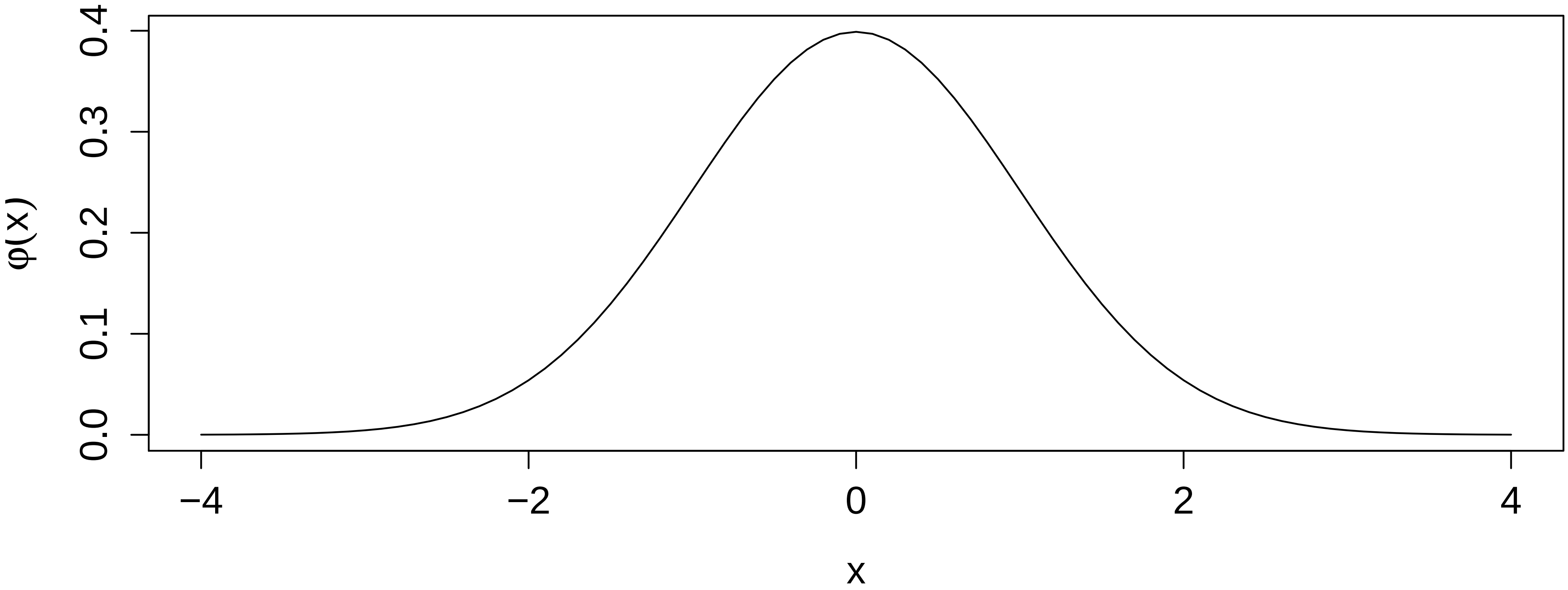
Definition 1.81.

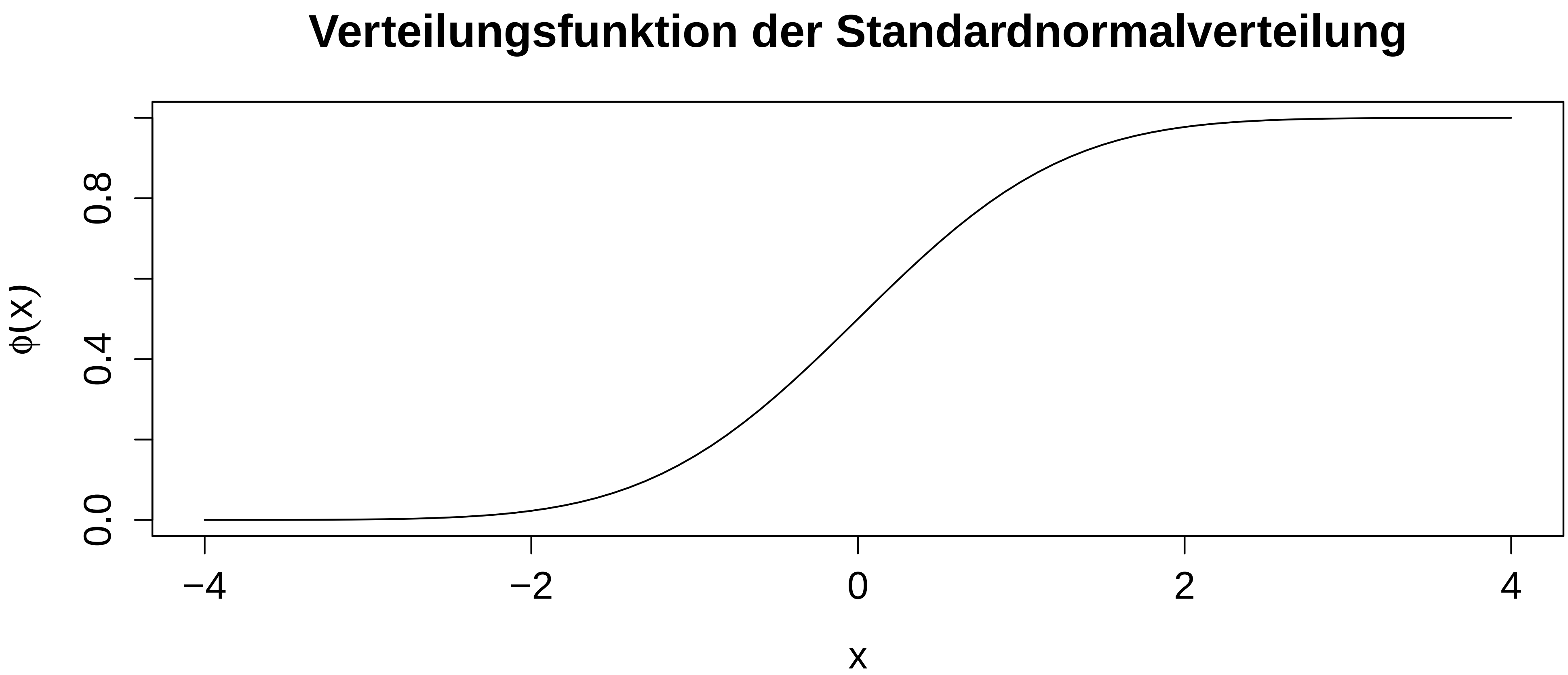
Eine stetige Zufallsvariable X heißt *normalverteilt* mit den Parametern μ und σ^2, in Zeichen $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \ x \in \mathbb{R} \tag{1.12}$$

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




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	···	0.05	···	0.09
⋮						
1.5	0.9332	0.9345	·	0.9394	·	0.9441
1.6	0.9452	0.9463	·	0.9505	·	0.9545
1.7	0.9554	0.9564	·	0.9599	·	0.9633
1.8	0.9641	0.9649	·	0.9678	·	0.9706
1.9	0.9713	0.9719	·	0.9744	·	0.9767
2.0	0.9772	0.9778	·	0.9798	·	0.9817
⋮						

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. 1.82. [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:

$$\begin{aligned} Y_1 &:= a_1X_1 + b \sim \mathcal{N}(a_1\mu_1 + b; a_1^2\sigma_1^2) \\ Y_2 &:= a_1X_1 + a_2X_2 \sim \mathcal{N}(a_1\mu_1 + a_2\mu_2; a_1^2\sigma_1^2 + a_2^2\sigma_2^2). \end{aligned}$$

Das Ergebnis lässt sich auf mehrere Summanden verallgemeinern.

Bsp. 1.83. [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:

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

Damit lässt sich die gesuchte Wahrscheinlichkeit bestimmen:

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