

**Herzlich willkommen zur
Vorlesung Statistik für Informatiker, SS 2021**

Matthias Birkner

Zufallsvariablen mit Dichten.

Definition 1.24. Sei X eine Zufallsvariable mit Werten in einem Intervall $S = [a, b] \subset \mathbb{R}$ mit $-\infty \leq a < b \leq \infty$ (im Fall $a > -\infty, b = \infty$ meinen wir $S = [a, \infty)$, etc.) und sei $f : S \rightarrow [0, \infty)$ eine integrierbare Funktion mit

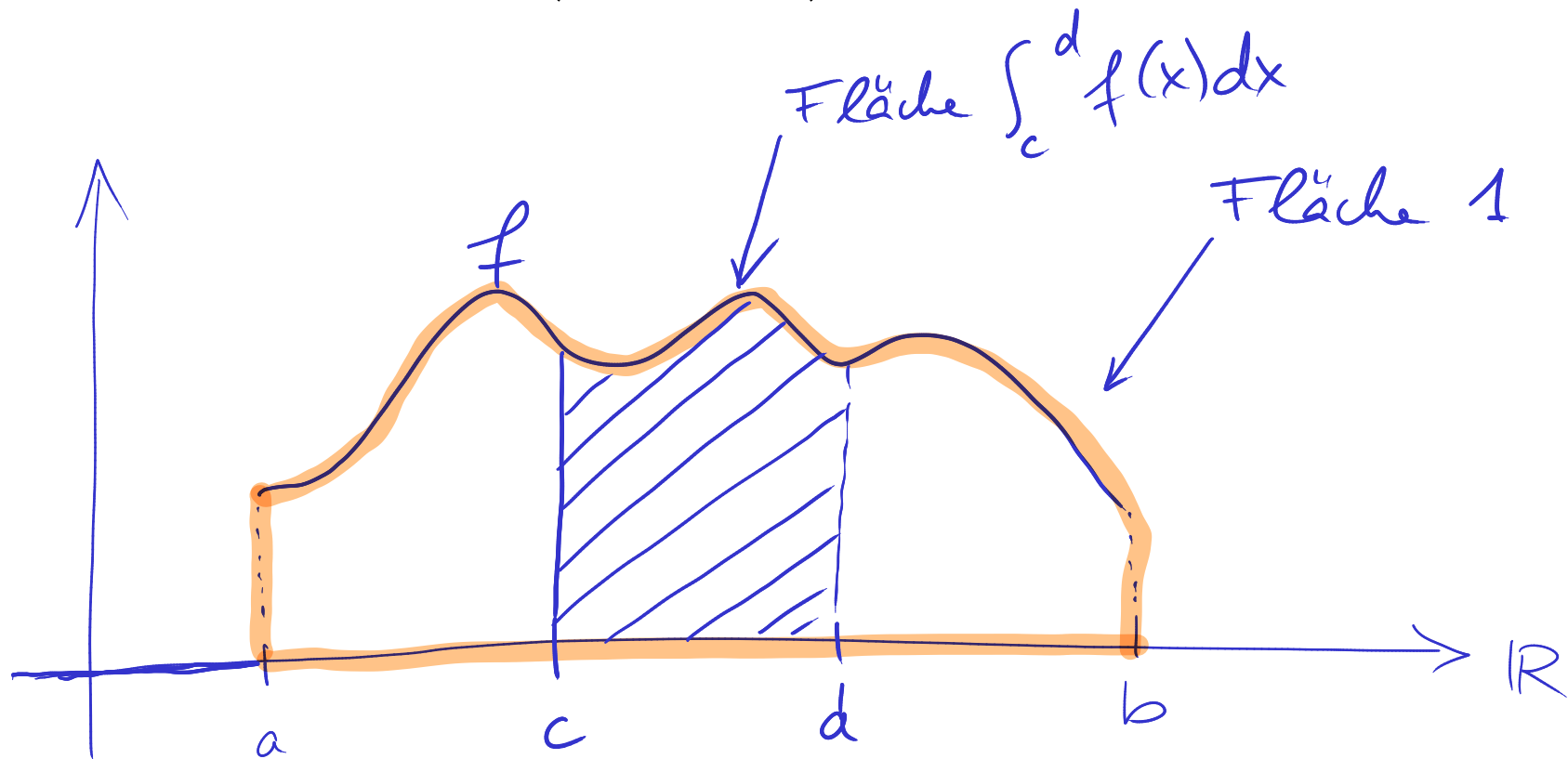
$$\int_a^b f(x) dx = 1.$$

X besitzt die *Dichte* (auch: *Wahrscheinlichkeitsdichte*) f , wenn gilt

$$P(X \in [c, d]) = \int_c^d f(x) dx \quad \text{für jedes Teilintervall } [c, d] \subset S.$$

Wir notieren oft auch f_X für die Dichte einer ZV X (um den Bezug zu X zu betonen, speziell wenn wir mehrere ZVn zugleich ins Auge fassen).

X hat Dichte f , so ist $P(X \in [c, d]) = \int_c^d f(x) dx$



Interpretation der Dichte:

X ZV mit Dichte f_X , für $x \in \mathbb{R}$ und kleines $\delta > 0$ ist

$$P(X \in [x, x + \delta]) = \int_x^{x+\delta} f_X(a) da \approx \delta f_X(x)$$

(wörtlich zumindest für Stetigkeitsstellen x von f_X), **also**

$$f_X(x) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P(X \in [x, x + \delta])$$

Man formuliert dies gelegentlich auch mit „infinitesimalen Größen“ als

$$P(X \in dx) = f_X(x) dx$$

(Dieser suggestive Ausdruck erhält einen Sinn im Sinne der „Standard-Analysis“, wenn man auf beiden Seiten x über ein Intervall $[c, d]$ integriert, dann erhält man Def. 1.24).

Bemerkung.

Für eine ZV X mit Dichte f_X ist es – im Gegensatz zum Fall mit Gewichten – nicht besonders sinnvoll, nach der Wahrscheinlichkeit von Ereignissen $\{X = x\}$ für feste Punkte $x \in \mathbb{R}$ zu fragen.

$$P(X = x) = P(X \in [x, x]) = \int_x^x f_X(a) da = 0$$

für jedes $x \in \mathbb{R}$.

aber: $P(X \in \mathbb{R}) = 1$

beachte: $\mathbb{R}$ sind nicht abzählbar

Beispiel 1.25 (Einige „klassische“ eindimensionale Verteilungen mit Dichte).

1. (uniforme Verteilung) $a, b \in \mathbb{R}, a < b$.

$\text{Unif}_{[a,b]}$ mit Dichte $\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$

ist W' dichte:

$$\int_a^b \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx$$

"

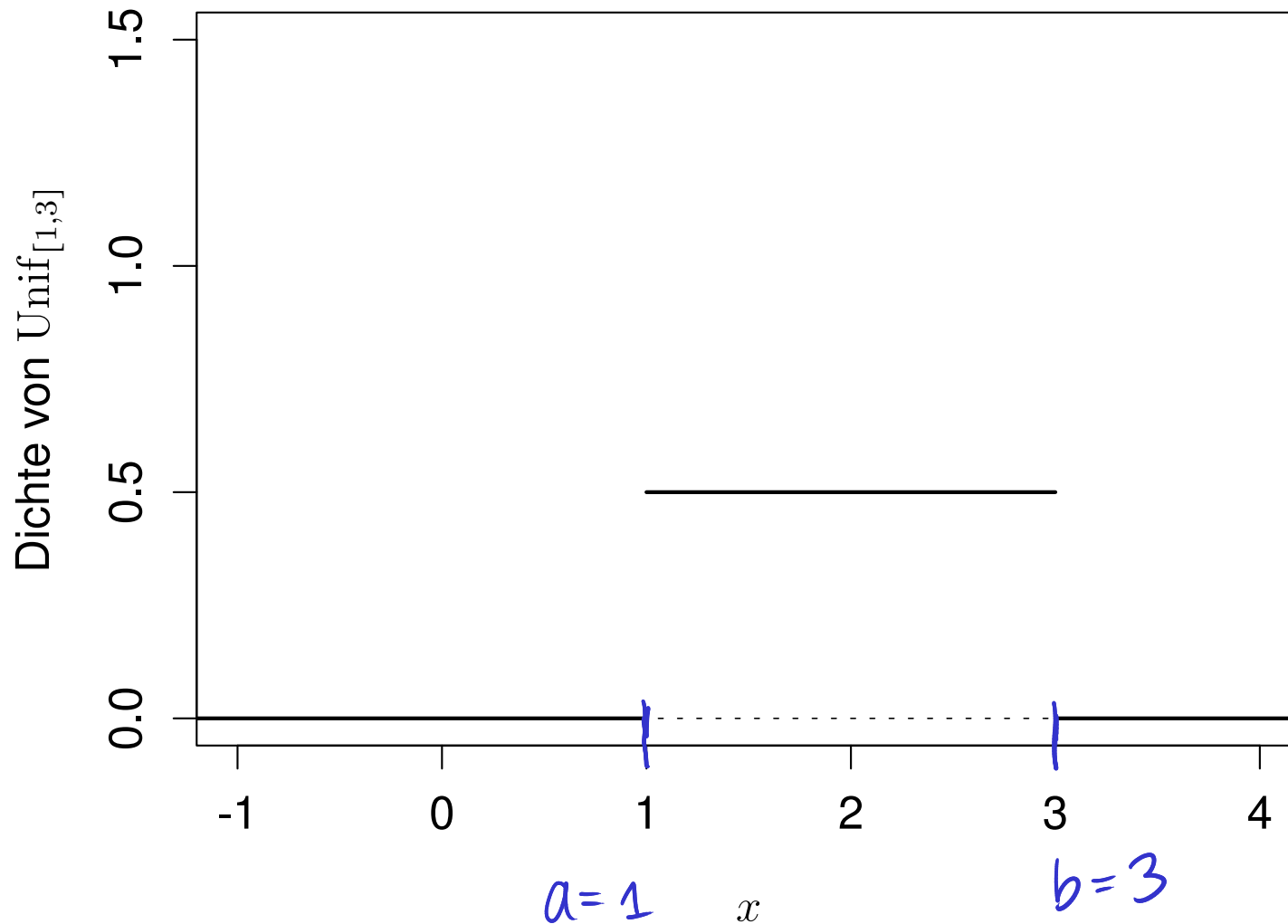
$$\int_a^b \frac{1}{b-a} dx$$

"

$$\frac{1}{b-a} \cdot \int_a^b 1 dx$$

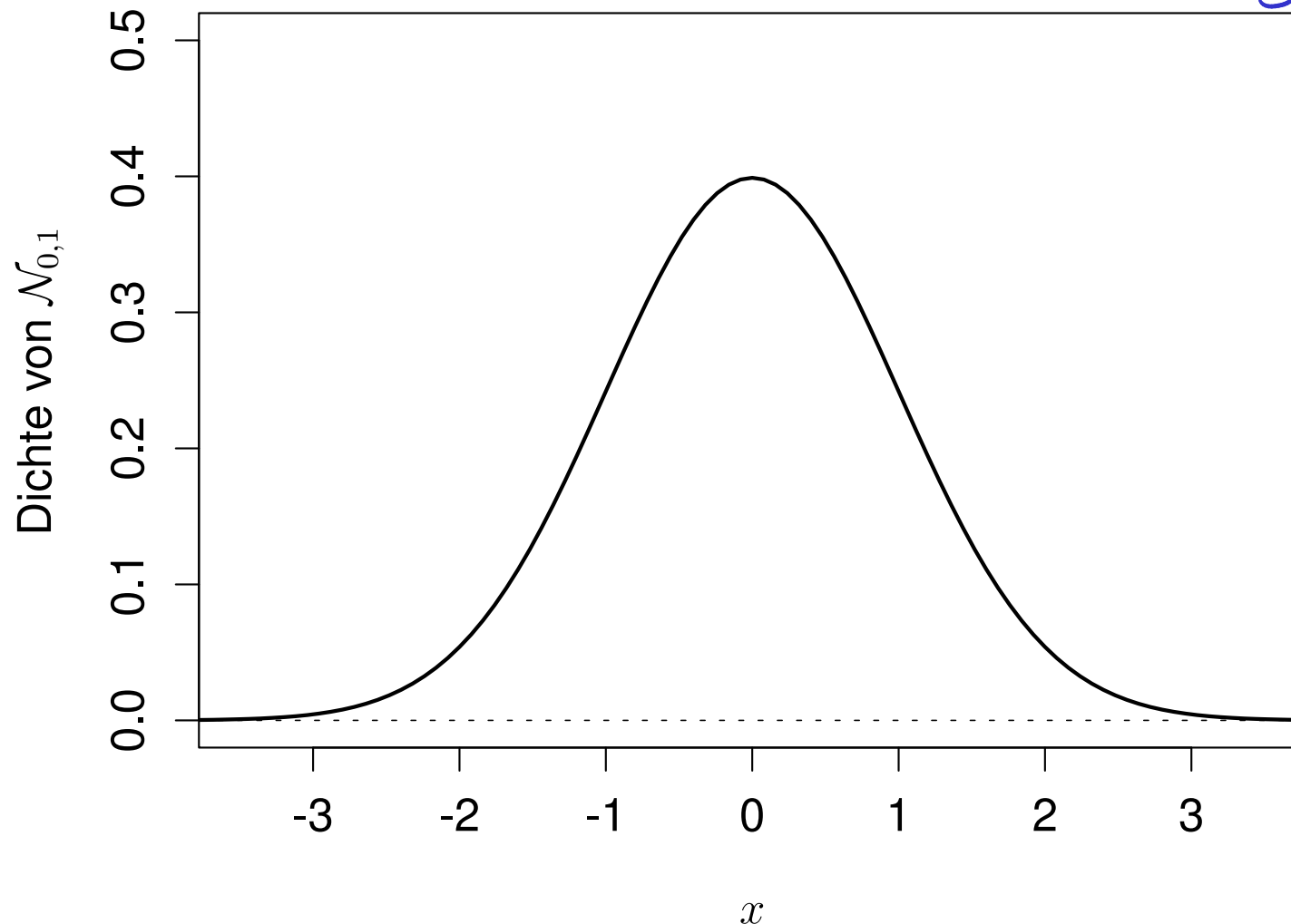
"

$$\frac{b-a}{b-a} = 1$$



Beispiel 1.25. 2. $\mu \in \mathbb{R}$, $\sigma > 0$. $\mathcal{N}_{\mu, \sigma^2}$ mit Dichte $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ heißt Normalverteilung mit Mittelwert μ und Varianz σ^2 .
 $\mathcal{N}_{0,1}$ heißt die *Standardnormalverteilung*.

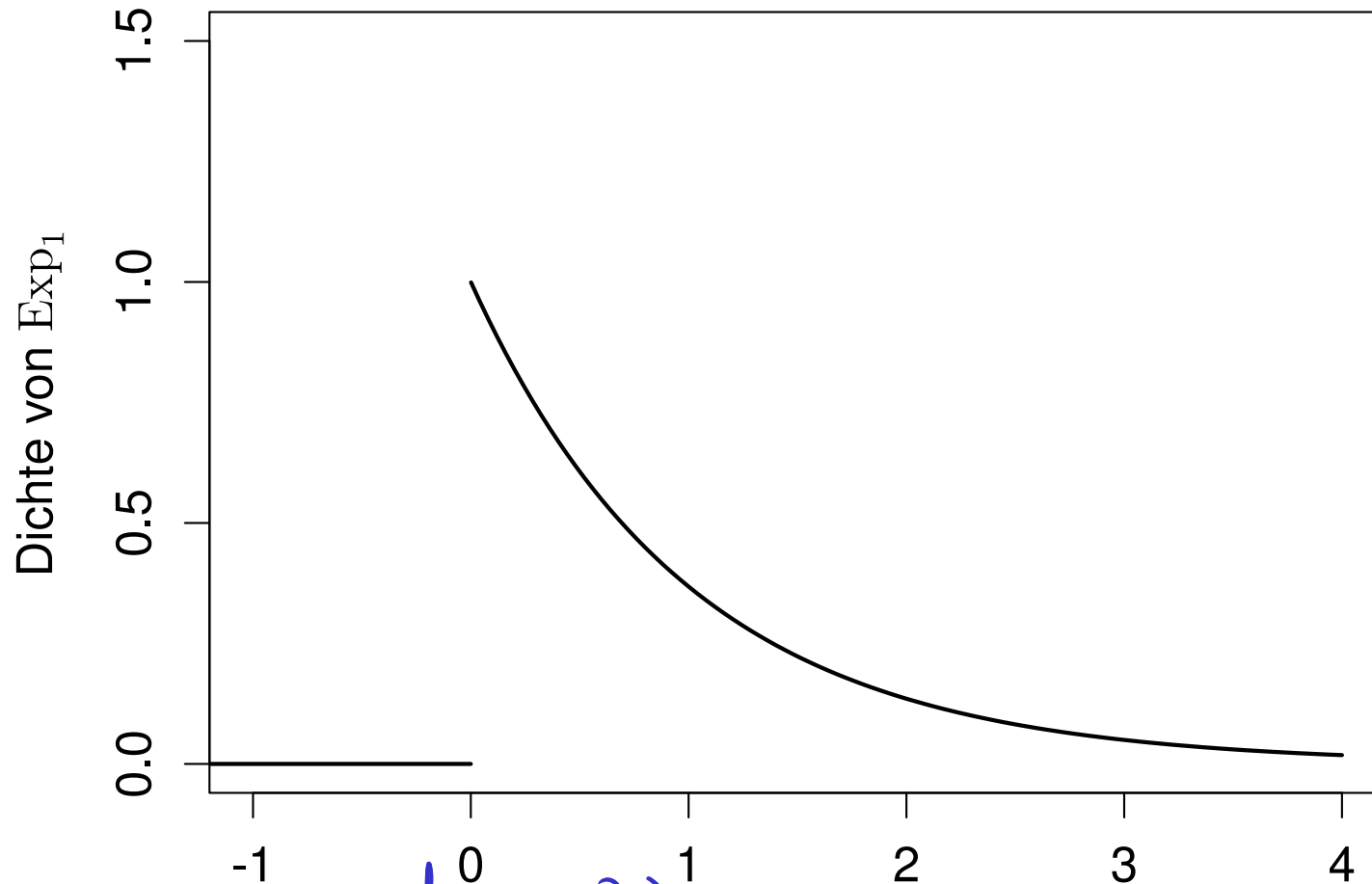
Wertebereich
 $S = (-\infty, \infty)$
" $\mathbb{R}$



Beispiel 1.25. 3. (Exponentialverteilung[en]) $\theta > 0$,

Exp_θ hat Dichte $\theta e^{-\theta x} \mathbf{1}_{[0, \infty)}(x)$

Wertebereich
 $S = [0, \infty)$



bem.: $\int_0^\infty \theta e^{-\theta x} dx = \frac{d}{dx} (-e^{-\theta x}) \Big|_0^\infty = [-e^{-\theta x}]_0^\infty = -0 - (-1) = 1$

Definition 1.26 (Verteilungsfunktion). Für eine Zufallsvariable X mit Werten in $\mathbb{R}$ (bzw. in einer Teilmenge $S \subset \mathbb{R}$) heißt die Funktion

$$F_X(x) := P(X \leq x), \quad x \in \mathbb{R} \quad (1.9)$$

die *Verteilungsfunktion* von X . $= P(X \in (-\infty, x])$

Wenn X mit Wertebereich $S \subset \mathbb{R}$ die Dichte f_X besitzt, so gilt offenbar

$$F_X(x) = \int_{-\infty}^x f_X(a) da \quad (1.10)$$

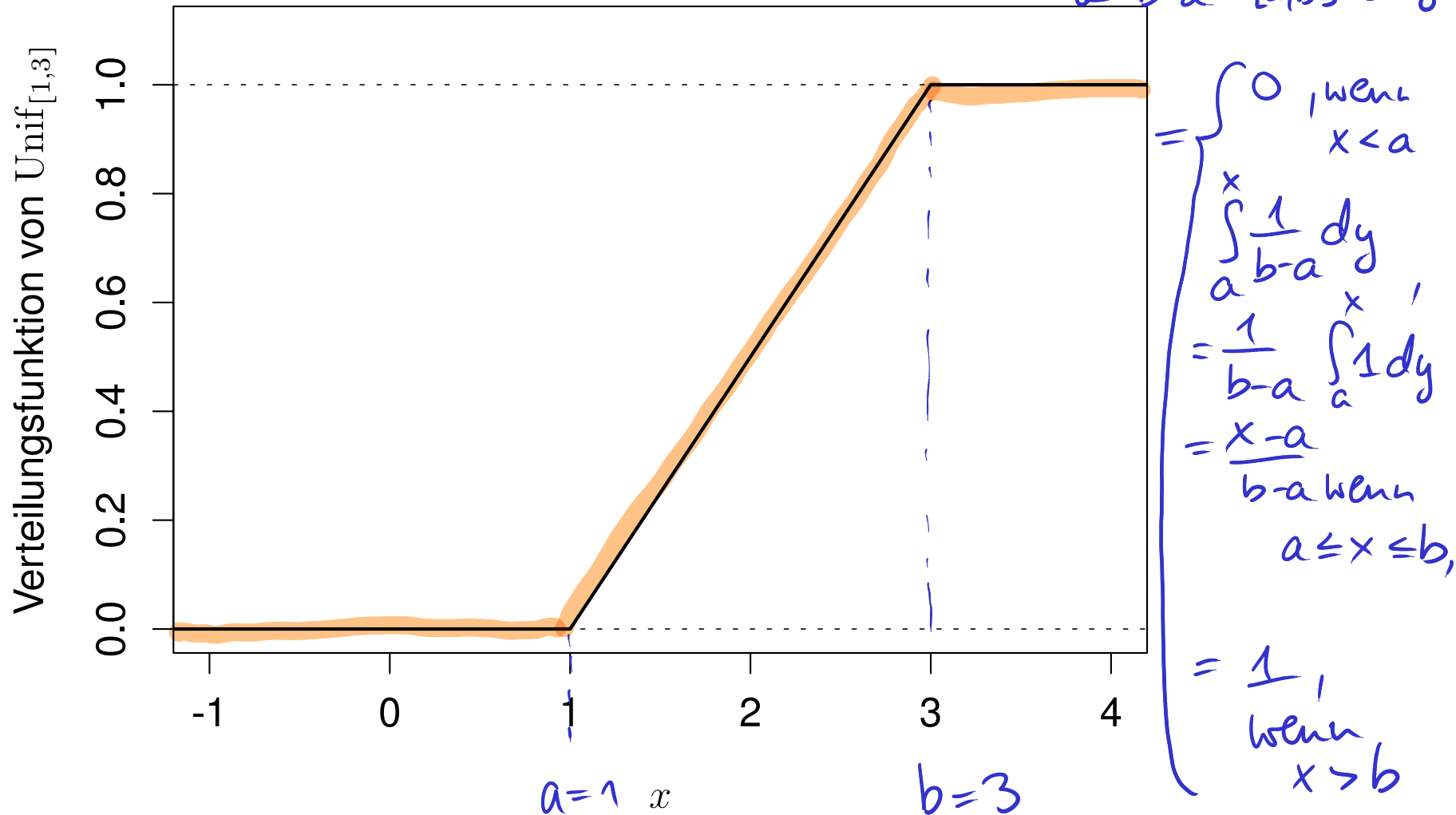
(mit Setzung $f_X(a) = 0$ für $a \notin S$, dem Wertebereich von X).

Beispiel 1.25 (Fortsetzung). 1. (uniforme Verteilung) $a, b \in \mathbb{R}, a < b$.

$\text{Unif}_{[a,b]}$ mit Dichte $\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$,

Verteilungsfunktion $\max \left\{ \min \left\{ \frac{x-a}{b-a}, 1 \right\}, 0 \right\}$

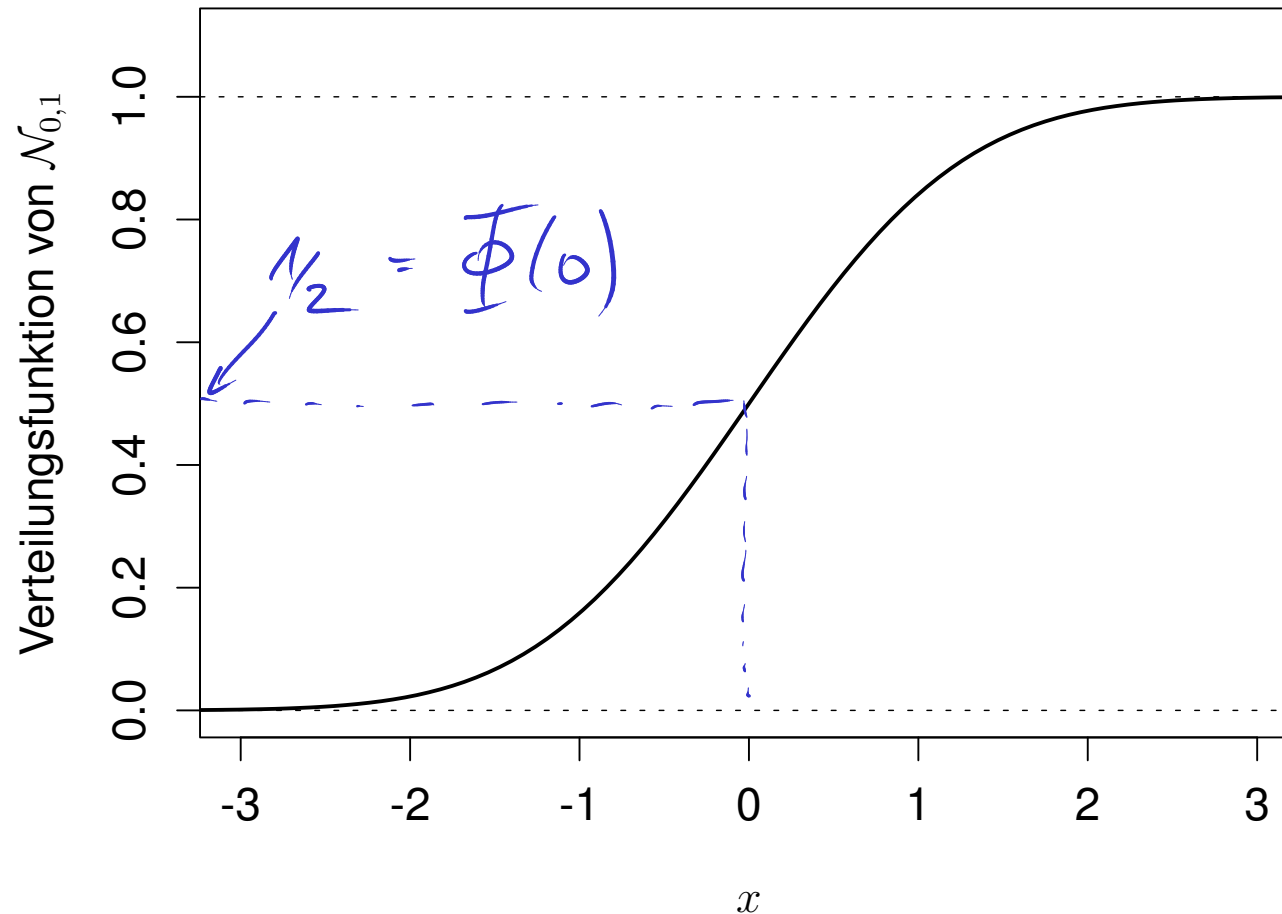
$$F_{\text{Unif}_{[a,b]}}(x) = \int_{-\infty}^x \frac{1}{b-a} \mathbf{1}_{[a,b]}(y) dy$$



Beispiel 1.25 (Fortsetzung). 2. Die Verteilungsfunktion der *Standardnormalverteilung* $\mathcal{N}_{0,1}$,

$$\Phi(x) := F_{\mathcal{N}_{0,1}}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

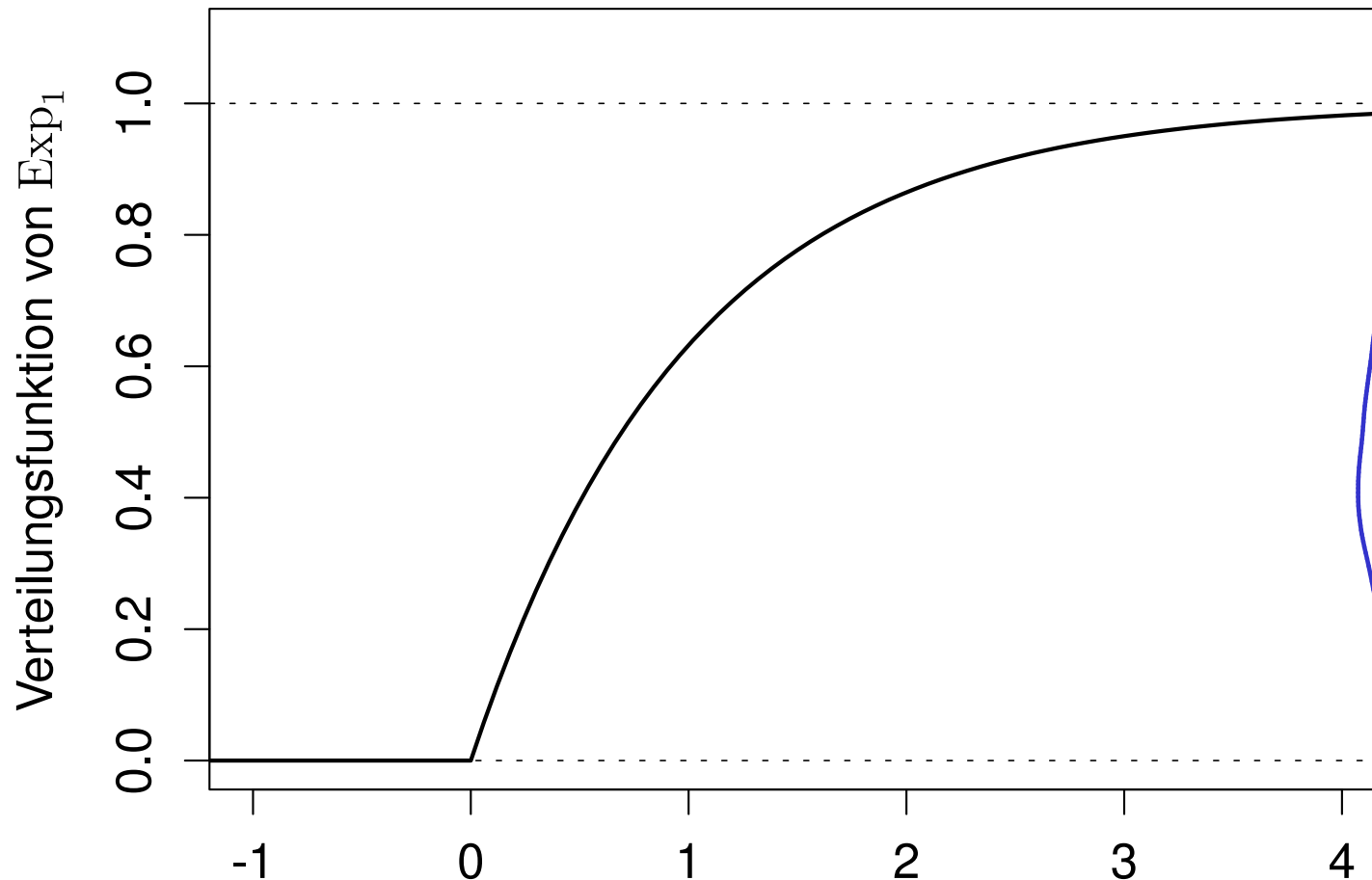
ist tabelliert bzw. in vielen Computerprogrammen implementiert (z.B. `pnorm` in R)



Beispiel 1.25 (Fortsetzung). 3. (Exponentialverteilung[en]) $\theta > 0$,

Exp_θ hat Dichte $\theta e^{-\theta x} \mathbf{1}_{[0, \infty)}(x)$,

Verteilungsfunktion $F_{\text{Exp}_\theta}(x) = (1 - e^{-\theta x}) \mathbf{1}_{[0, \infty)}(x)$



$$\begin{aligned}
 &= \int_{-\infty}^x \theta e^{-\theta y} \mathbf{1}_{[0, \infty)}(y) dy \\
 &= \begin{cases} 0, & x < 0 \\ \int_0^x \theta e^{-\theta y} dy, & x \geq 0 \end{cases} \\
 &= [-e^{-\theta y}]_0^x \\
 &= 1 - e^{-\theta x}, \\
 &\text{falls } x \geq 0
 \end{aligned}$$

x

Bemerkung 1.27. 1. Die Dichte / Verteilungsfunktion von X hängt nur von der Verteilung von X ab:

wenn $Y \stackrel{d}{=} X$ („Gleichheit in Verteilung“), also $P(X \in B) = P(Y \in B)$ für alle B gilt, so hat Y dieselbe (offenbar).

Wir sprechen daher oft auch kurz von der Dichte bzw. Verteilungsfunktion einer Wahrscheinlichkeitsverteilung auf $\mathbb{R}$, ohne die zugehörige ZV explizit zu machen.

2. Wenn X Dichte f_X und Verteilungsfunktion F_X besitzt, so ist

$$\frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^x f_X(a) da = f_X(x)$$

(zumindest an Stetigkeitspunkten von f_X)

$$F_X(x) = P(X \in (-\infty, x]) = \int_{-\infty}^x f_X(a) da$$

Bemerkung 1.27.3. X ZV mit Werten in $S \subset \mathbb{R}$ mit Verteilungsfunktion F_X , $c < d$, so ist

$$= P(X \in (-\infty, d] \setminus (-\infty, c])$$

$$P(X \in (c, d]) = P(X \leq d) - P(X \leq c) = F(d) - F(c)$$

(und falls $P(X = c) = 0$, z.B. weil X eine Dichte besitzt, so ist natürlich auch $P(X \in [c, d]) = P(X = c) + P(X \in (c, d]) = F(d) - F(c)$).

Für $B = \bigcup_{i=1}^n (c_i, d_i]$ mit $c_1 < d_1 < c_2 < d_2 < \dots < c_n < d_n$ ist (mit Eigenschaft (A) aus Def. 1.7)

$$P(X \in B) = \sum_{i=1}^n P(X \in (c_i, d_i]) = \sum_{i=1}^n (F_X(d_i) - F_X(c_i))$$

(und „allgemeine“ Mengen $B \subset \mathbb{R}$ können auf diese Weise approximiert werden).

In diesem Sinne „weiß F_X alles“ über die Verteilung von X .



Bemerkung 1.27. 4. (Bezug zum diskreten Fall). Sei X ZV mit (höchstens) abzählbarem Wertebereich $S = \{x_1, x_2, \dots\} \subset \mathbb{R}$ und Gewichten $\rho_X(x_n)$ wie in Def. 1.11, d.h.

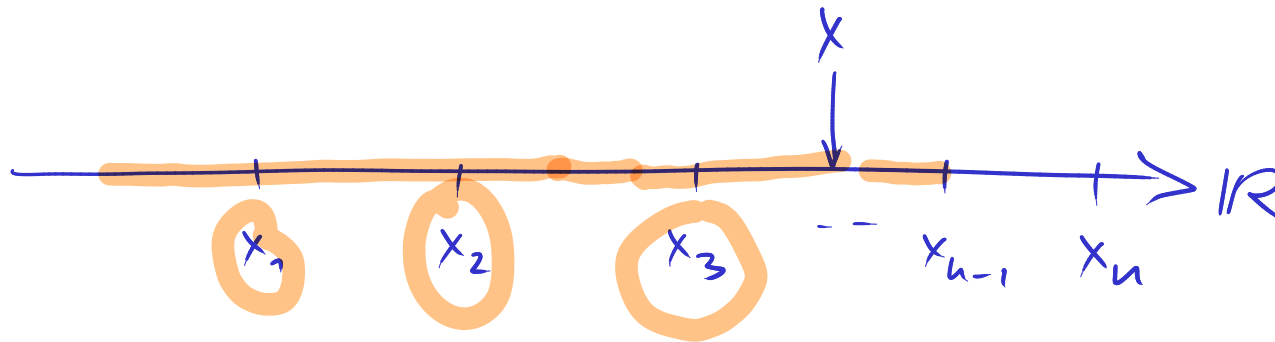
$$P(X \in B) = \sum_{n: x_n \in B} \rho_X(x_n), \quad B \subseteq \mathbb{R}$$

dann ergibt sich als Verteilungsfunktion

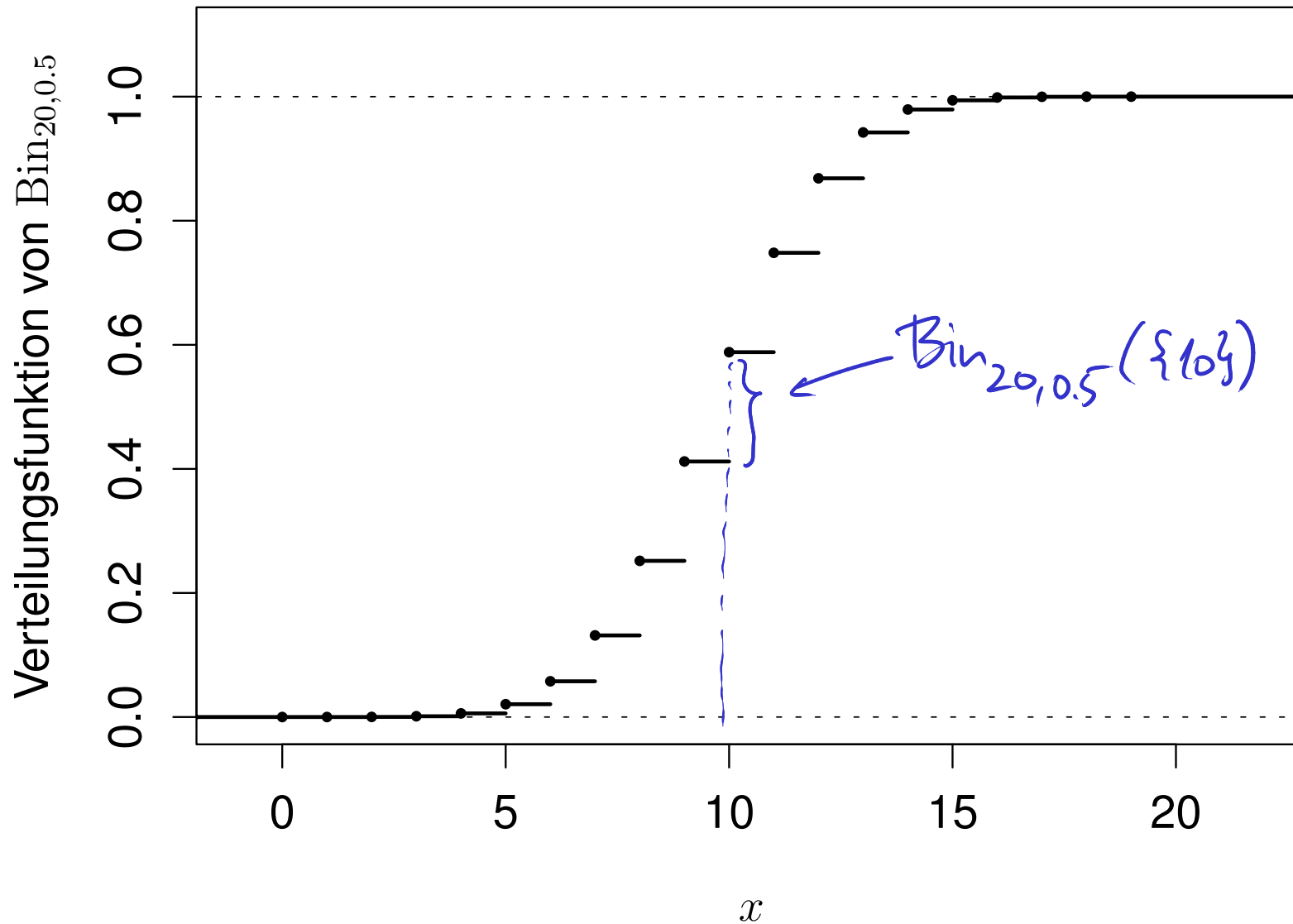
$$F_X(x) = \sum_{n: x_n \leq x} \rho_X(x_n).$$

speziell:
für $B = (-\infty, x]$
mit $x \in \mathbb{R}$

(Diese ist stückweise konstant mit [höchstens] abzählbar vielen Sprüngen.)



Beispiel: Verteilungsfunktion einer Binomialverteilung



Bemerkung 1.27. 5. Stets ist F_X nicht-fallend und rechtsstetig (wenn X eine Dichte besitzt, so ist F_X stetig) mit

$$\lim_{x \rightarrow \infty} F_X(x) = 1, \quad \lim_{x \rightarrow -\infty} F_X(x) = 0.$$

damit 1) für $x_1 < x_2$ ist $\{X \leq x_1\} \subset \{X \leq x_2\}$

$$\Rightarrow F_X(x_1) = P(X \leq x_1) \leq P(X \leq x_2) = F_X(x_2)$$

2) Sei $x_n \downarrow x$, $\{X \leq x_n\} \supset \{X \leq x_{n+1}\} \supset \dots \supset \{X \leq x\}$
 $\quad \quad \quad n \rightarrow \infty$

$$\Rightarrow \underbrace{P(X \leq x_n)}_{F(x_n)} \xrightarrow{n \rightarrow \infty} P(X \leq x) = F(x)$$

6. Umgekehrt gibt es zu jeder Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit den Eigenschaften aus Def. 1.26, 5. eine ZV X mit $F_X = F$.

(Wir kommen darauf zurück, siehe Beob. 1.29 unten.)

Definition 1.28. Die (verallgemeinerte) inverse Funktion von F_X ,

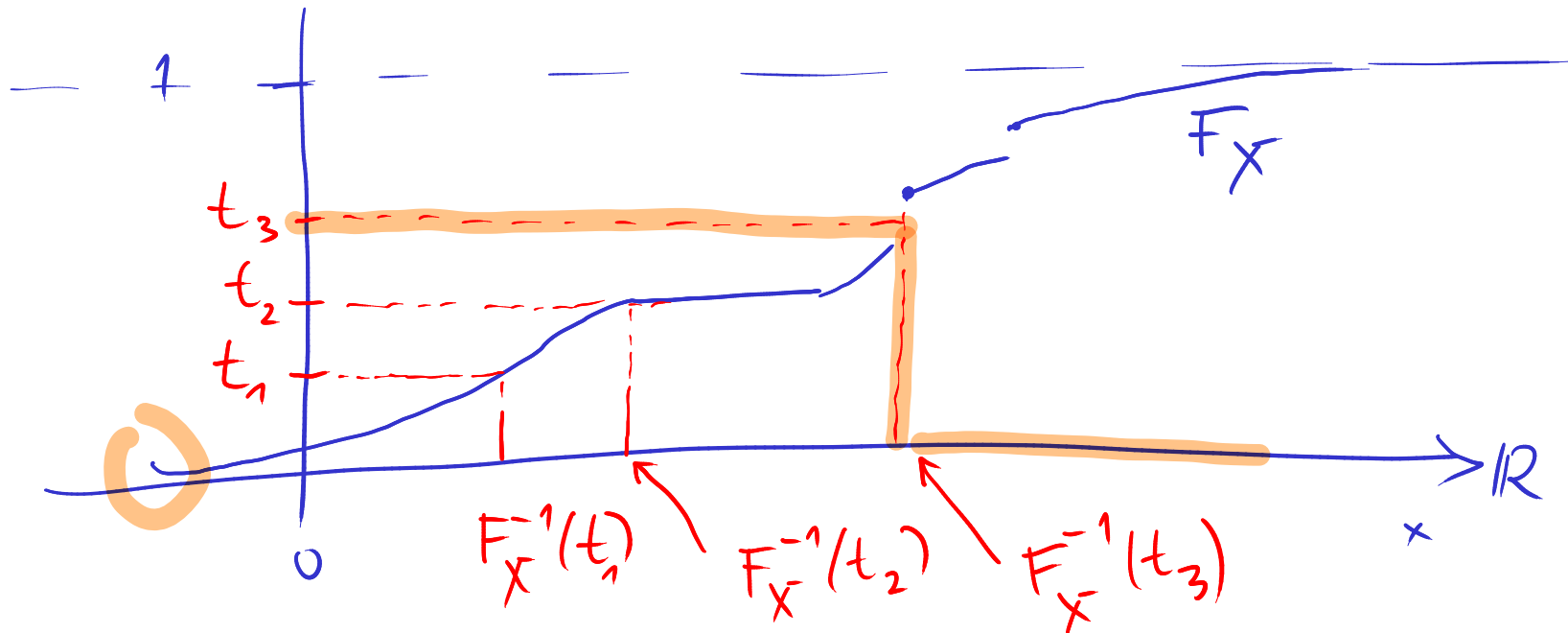
$$F_X^{-1}(t) := \inf \{x \in \mathbb{R} : F_X(x) \geq t\}, \quad t \in [0, 1]$$

(mit Setzung $\inf \emptyset = +\infty$) heißt auch die *Quantilfunktion* von X .

(Beachte, dass die so definierte Funktion F_X^{-1} linksstetig ist. In der Literatur gibt es leicht verschiedene Definitionen der „Quantilfunktion“, man prüfe ggfs. jeweils die verwendete Konvention.)

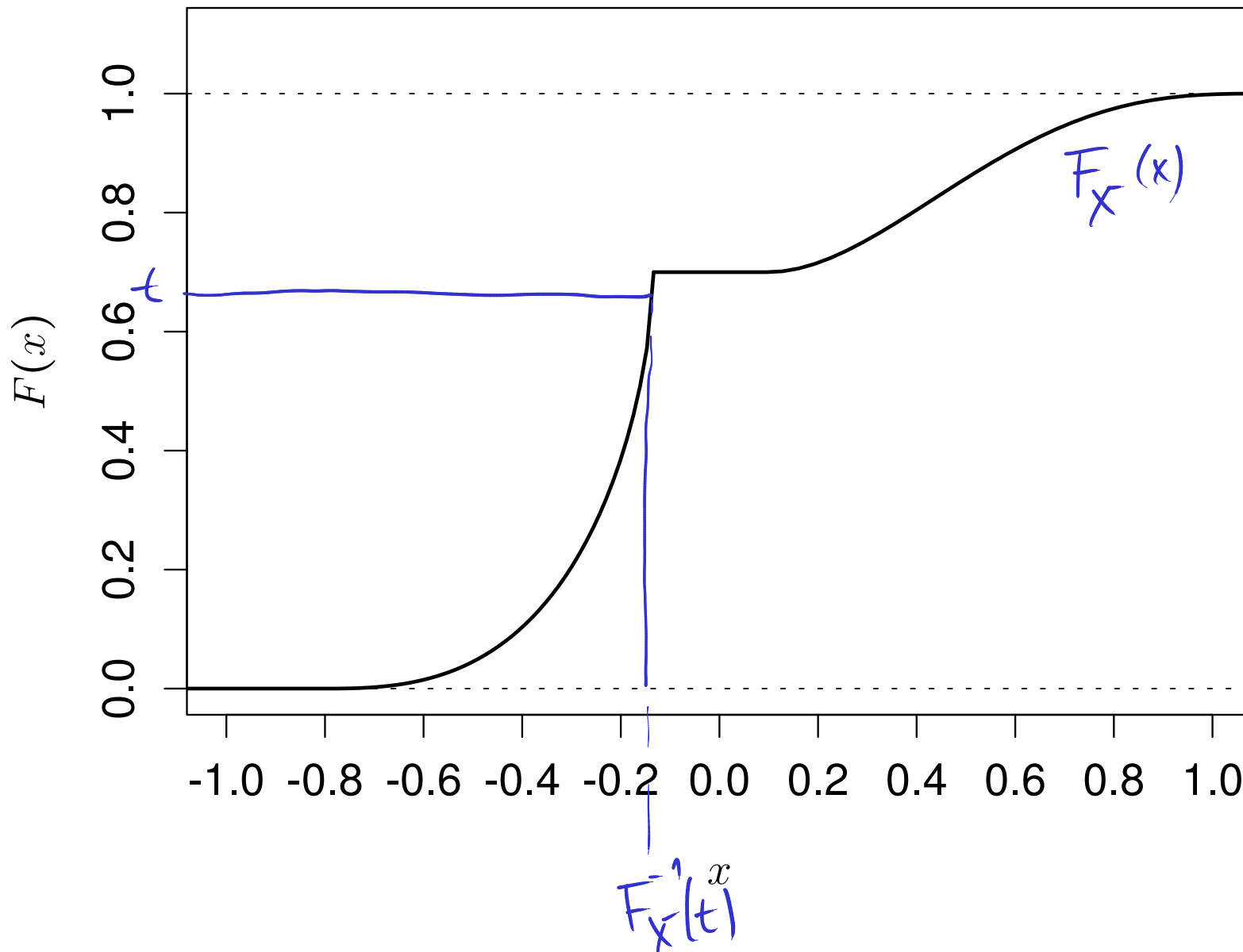
Mit dieser Definition ergibt sich für $x \in \mathbb{R}$, $t \in [0, 1]$ die Beziehung

$$F_X^{-1}(t) \leq x \iff t \leq F_X(x).$$

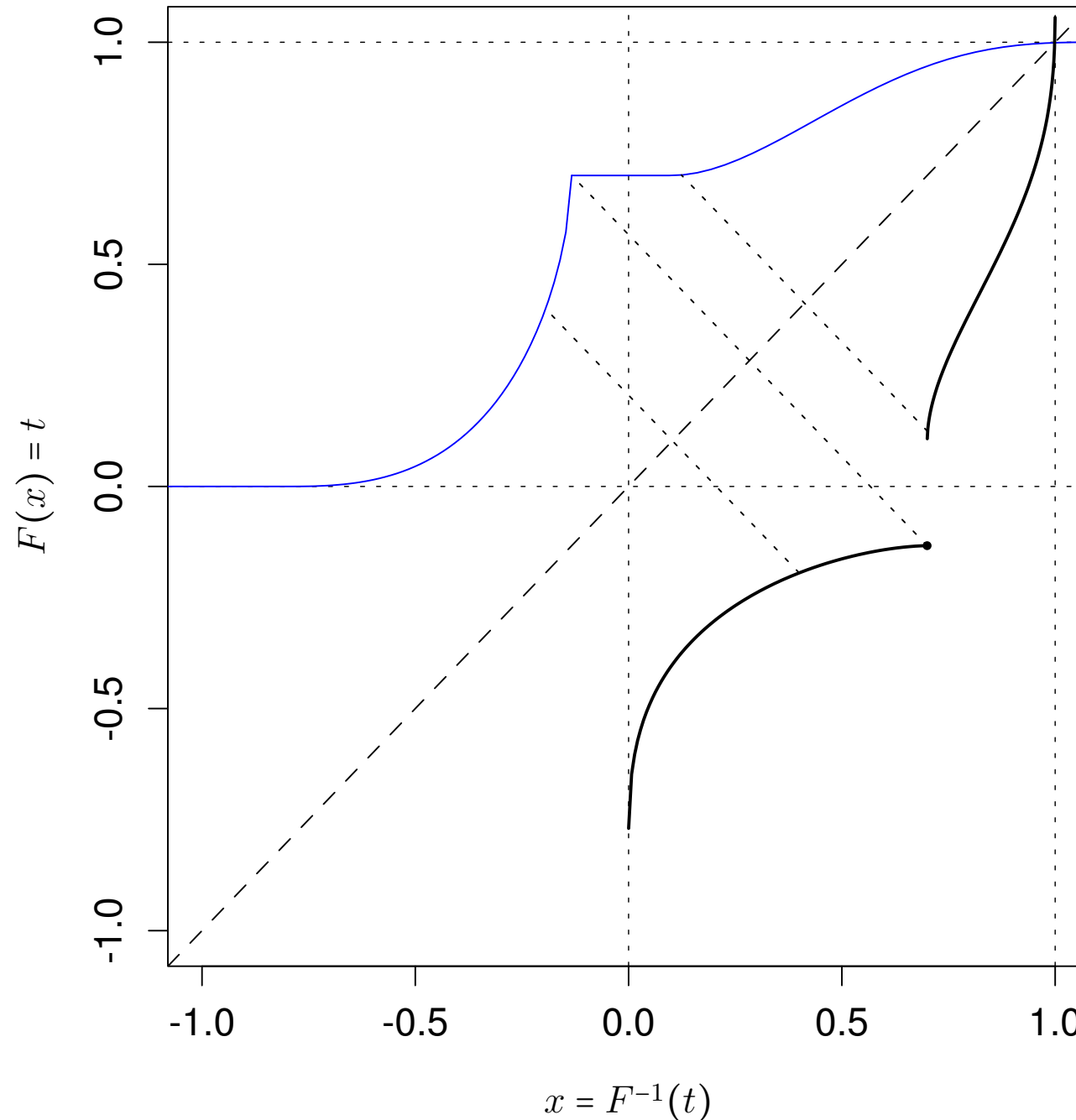


Beispiel: Ablesen von F^{-1}

$$F_X^{-1}(t) := \inf\{x \in \mathbb{R} : F_X(x) \geq t\}, \quad t \in [0, 1]$$



Erinnerung: Inverse via Spiegelung an der Diagonalen



Beobachtung 1.29 (Erzeugung reeller ZVn mit vorgegebener Verteilung). Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion ,

$$F^{-1}(t) := \inf \{x \in \mathbb{R} : F(x) \geq t\}, \quad t \in [0, 1]$$

die inverse Verteilungsfunktion oder Quantilfunktion (aus Def. 1.28),
 U reelle ZV, $U \sim \text{Unif}_{[0,1]}$, dann hat

$$X := F^{-1}(U)$$

(mit R :
 $r_{\text{unif}}(1)$)

die Verteilungsfunktion $F_X = F$.

Denn: Es gilt $F^{-1}(t) \leq x \iff t \leq F(x)$
(für $t \in [0, 1]$, $x \in \mathbb{R}$).

Sei $x \in \mathbb{R}$

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) \\ &= P(0 \leq U \leq F(x)) = \frac{F(x) - 0}{1 - 0} = F(x) \end{aligned}$$

↙ $\in [0, 1]$

┘

Beispiel 1.30. 1. Exp_θ hat Verteilungsfunktion

$$F_{\text{Exp}_\theta}(x) = (1 - e^{-\theta x}) \mathbf{1}_{[0, \infty)}(x) \text{ mit}$$

inverser Funktion $F_{\text{Exp}_\theta}^{-1}(t) = -\frac{1}{\theta} \log(1-t)$, also ist $-\frac{1}{\theta} \log(1-U) \sim \text{Exp}_\theta$

(und natürlich ebenso $-\frac{1}{\theta} \log(U)$)

$x > 0$

$$(1 - e^{-\theta x}) = t \iff e^{-\theta x} = 1 - t$$

$$\iff -\theta x = \log(1-t)$$

2. $p(k), k \in \mathbb{N}_0$ Wahrscheinlichkeitsgewichte, $F(x) = \sum_{k: k \leq x} p(k)$ zugehörige Verteilungsfunktion (vgl. Bem. 1.27, 4.), so hat

$$X := \min \left\{ n \in \mathbb{N}_0 : \sum_{k=0}^n p(k) \geq U \right\}$$

die Gewichte $(p(k))_{k \in \mathbb{N}_0}$.

(Dies ist etwa eine Möglichkeit, eine Poisson-verteilte ZV zu simulieren.)

Bericht: Verteilungen in R.

R verwendet folgende Namenskonvention: Wenn name für eine Verteilung steht, so ist

- `dname` die Dichte- bzw. Gewichtsfunktion von `name`
- `pname` die Verteilungsfunktion von `name` (“p” steht für “probability distribution function”)
- `qname` die Quantilfunktion von `name`
- `rname` simuliert gemäß `name` (“r” steht für “random”)

Beispiele:

- Uniforme Verteilung $\text{Unif}_{[a,b]}$: `[d|p|q|r]unif(...,min=a,max=b)`
- Normalverteilung $\mathcal{N}_{\mu,\sigma^2}$: `[d|p|q|r]norm(...,mean= $\mu$ ,sd= $\sigma$ )`
(beachte: R parametrisiert mit σ , nicht mit σ^2)
- Exponentialverteilung Exp_{λ} : `[d|p|q|r]exp(...,rate= $\lambda$ )`
- Poissonverteilung Poi_{λ} : `[d|p|q|r]pois(...,lambda= $\lambda$ )`
- Binomialverteilung $\text{Bin}_{n,p}$: `[d|p|q|r]binom(...,size=n,prob=p)`

(Siehe auch die Online-Hilfe in R.)

Transformation von Dichten

Beobachtung 1.31. X habe Dichte f_X , sei $a > 0$, $b \in \mathbb{R}$, so hat $Y := aX + b$ die Dichte

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

$$\left(y = ax + b \Leftrightarrow x = \frac{y-b}{a} \right)$$

beachte:

Nebenbem.:

$$Y \in [y, y+\delta]$$



$$X \in \left[\frac{y-b}{a}, \frac{y-b}{a} + \frac{\delta}{a} \right]$$

$$P(Y \leq y) = P\left(X \leq \frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} f_X(x) dx$$

Substituiere $x = \frac{z-b}{a} \Leftrightarrow z = ax + b$, $\frac{dz}{dx} = a$, $\frac{dx}{dz} = \frac{1}{a}$

$$= \int_{-\infty}^y f_X\left(\frac{z-b}{a}\right) \frac{1}{a} dz$$

Beispiel 1.32. 1. $X \sim \mathcal{N}_{0,1}$, $\mu \in \mathbb{R}$, $\sigma > 0$, so ist $Y := \sigma X + \mu \sim \mathcal{N}_{\mu, \sigma^2}$

Insbesondere gilt $\mathcal{N}_{\mu, \sigma^2}((-\infty, x]) = \Phi((x - \mu)/\sigma)$

(mit
 $\Phi(x) = F_{\mathcal{N}_{0,1}}(x)$)

↙ Dichte $e^{-x} \mathbf{1}_{[0, \infty)}(x)$

2. $X \sim \text{Exp}_1$, $a > 0$, so hat $Y := aX$ die Dichte $\frac{1}{a}e^{-x/a} \mathbf{1}_{[0, \infty)}(x)$
(d.h. $Y \sim \text{Exp}_{1/a}$)

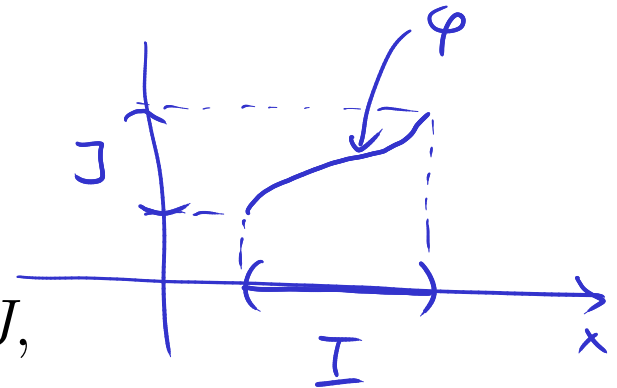
Transformation von Dichten, 2

Proposition 1.33 (Allgemeine Dichtetransformation im Fall $\mathbb{R}^1$).

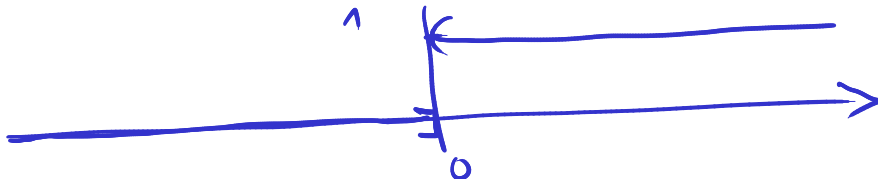
X reelle ZV mit Dichte f_X , d.h. $F_X(x) = \int_{-\infty}^x f_X(z) dz$, $I \subset \mathbb{R}$
(möglicherweise unbeschränktes) offenes Intervall mit $P(X \in I) = 1$,
 $J \subset \mathbb{R}$, $\varphi : I \rightarrow J$ stetig differenzierbar, bijektiv.

Dann hat $Y := \varphi(X)$ die Dichte

$$f_Y(y) = \begin{cases} \frac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}, & y \in J, \\ 0, & y \notin J. \end{cases}$$



[Bem.: Wenn φ nicht bijektiv ist, so wird $\varphi(X)$ i.A. keine Dichte besitzen, z.B. $X \sim \mathcal{N}_{0,1}$,
 $\varphi(x) = \mathbf{1}_{(0,\infty)}(x)$, so ist $\varphi(X) \sim \text{Ber}_{1/2}$.]



$Y = \varphi(X)$ hat Dichte $f_X(\varphi^{-1}(y))/\varphi'(\varphi^{-1}(y))$

„Dichtetransformel“
(1-dimensionaler Fall)

(wir nehmen an, dass φ strikt monoton wachsend ist.)

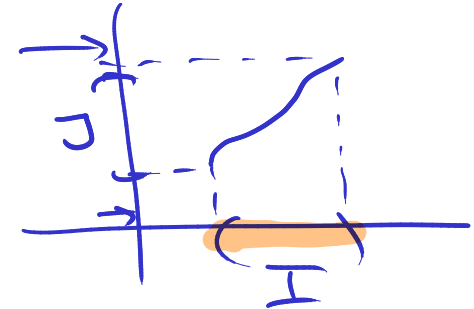
$z \leq \inf J \Rightarrow P(Y \leq z) = 0$,
also $f_Y(z) = 0$ für alle $z < \inf J$

$z \geq \sup J \Rightarrow P(Y \leq z) = 1$
also $f_Y(z) = 0$ für alle $z > \sup J$

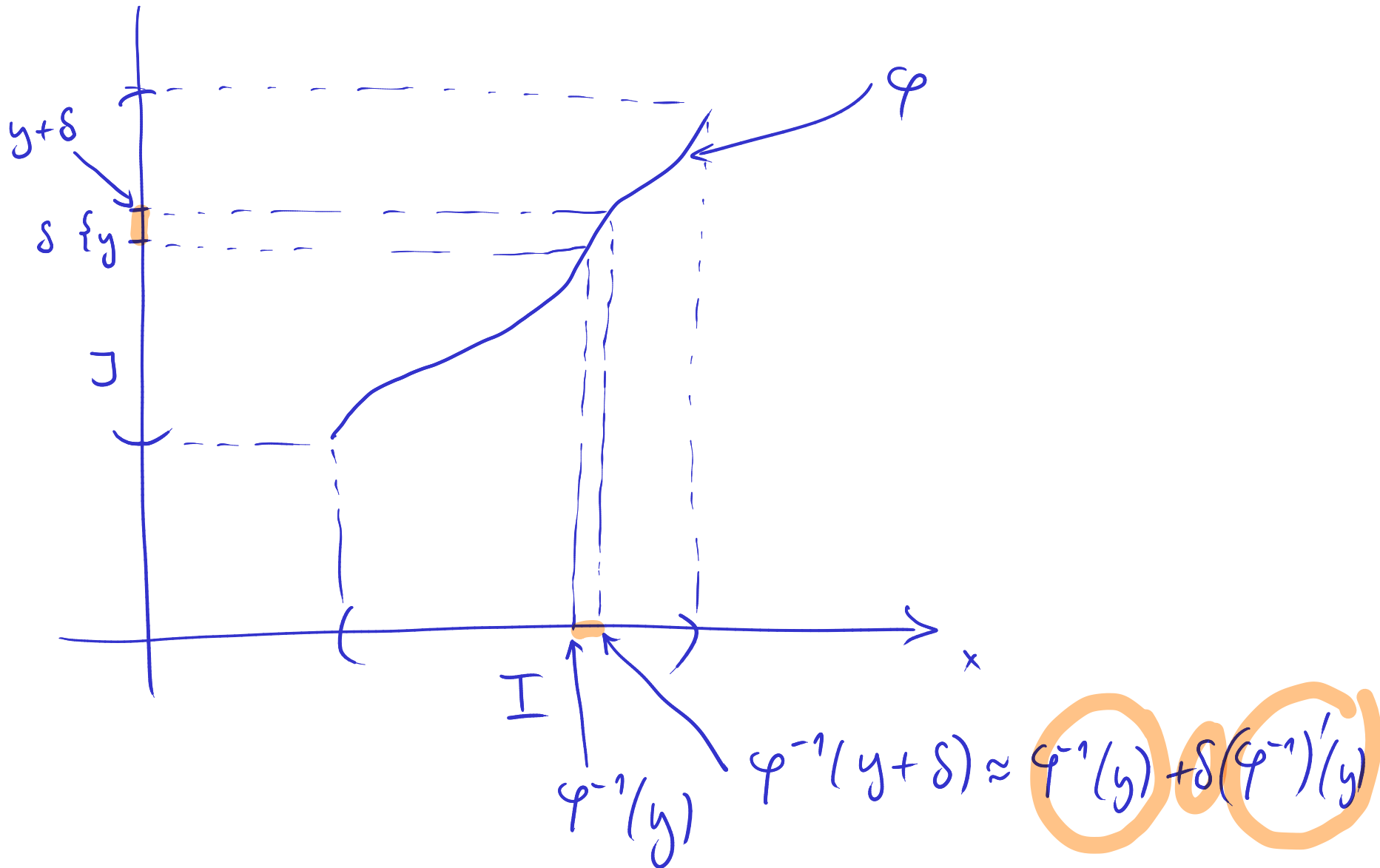
$z \in J$: $P(Y \leq z) = P(\varphi(X) \leq z) = P(X \leq \varphi^{-1}(z))$

$$\inf I \rightarrow \int_{\inf J}^{\varphi^{-1}(z)} f_X(x) dx = \int_{\inf J}^z f_X(\varphi^{-1}(y)) \frac{1}{\varphi'(\varphi^{-1}(y))} dy$$

substituiere $x = \varphi^{-1}(y) \Leftrightarrow \varphi(x) = y$, $\frac{dx}{dy} = (\varphi^{-1})'(y) = \frac{1}{\varphi'(\varphi^{-1}(y))}$



$Y = \varphi(X)$ hat Dichte $f_X(\varphi^{-1}(y))/\varphi'(\varphi^{-1}(y))$



Beispiel 1.34. 1. $U \sim \text{Unif}_{[0,1]}$, so hat

$X := -\log(U)$ Dichte e^{-x} für $x \geq 0$ (wie wir in Bsp. 1.30 gesehen haben)

2. $U \sim \text{Unif}_{[0,1]}$, $n \in \mathbb{N}$, so hat $Y := U^n$ Dichte $f_Y(y) = n^{-1}y^{1/n-1}$ (für $0 \leq y \leq 1$)

1.1.3 Zum mehrdimensionalen Fall

Dichten im mehrdimensionalen Fall

Definition 1.35. Sei $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ eine (geeignet) integrierbare* Funktion mit

$$\int_{\mathbb{R}^d} f(x) dx = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 \cdots dx_d = 1.$$

Eine Zufallsvariable X mit Werten in $\mathbb{R}^d$ besitzt die *Dichte* f , wenn für (geeignete) Teilmengen $A \subset \mathbb{R}^d$ gilt

$$\begin{aligned} P(X \in A) &= \int_A f(x) dx \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_A(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \cdots dx_d. \end{aligned}$$

*Wir denken hier z.B. an ein d -fach iteriertes Riemann-Integral.

Dichten im mehrdimensionalen Fall

Analog zum 1-dimensionalen Fall besitzt die Dichte f_X einer d -dimensionalen ZV X die Interpretation

$$\begin{aligned} P(X \in [x_1, x_1 + \delta_1] \times [x_2, x_2 + \delta_2] \times \cdots \times [x_d, x_d + \delta_d]) \\ \approx \delta_1 \delta_2 \cdots \delta_d \cdot f_X((x_1, \dots, x_d)) \end{aligned}$$

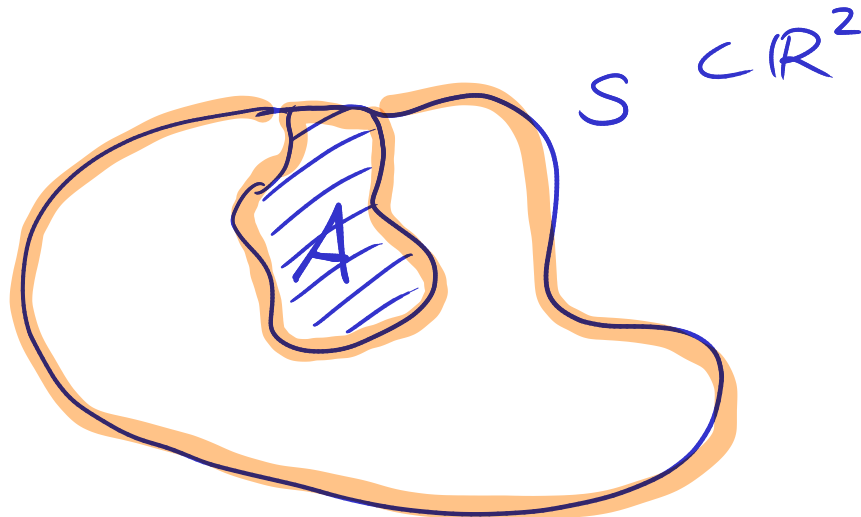
$[x_d, x_d + \delta_d]$

für $(x_1, \dots, x_d) \in \mathbb{R}^d$ und $0 < \delta_1, \delta_2, \dots, \delta_d \ll 1$.

Beispiel 1.36. 1. Uniforme oder Laplace-Verteilung auf einem beschränkten Gebiet $S \subset \mathbb{R}^d$: X mit Dichte $f_X(x) = \frac{1}{\text{vol}(S)} \mathbf{1}_S(x)$ erfüllt für $A \subset S$ (geeignet*)

$$P(X \in A) = \int_A \frac{1}{\text{vol}(S)} \mathbf{1}_S(x) dx = \frac{\int_A 1 dx}{\int_S 1 dx} = \frac{\text{vol}(A)}{\text{vol}(S)}.$$

(Z.B. der zufällig gewählte Punkt Z aus Kapitel 0 kann als uniform auf $S = [0, 1]^2$ verteilt modelliert werden.)



*in dem Sinne, dass ein d -dimensionales „Volumen“ $\text{vol}(A)$ definierbar ist

Beispiel 1.36 (Forts.).

2. Die 2-dimensionale Standard-Normalverteilung hat Dichte $f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$

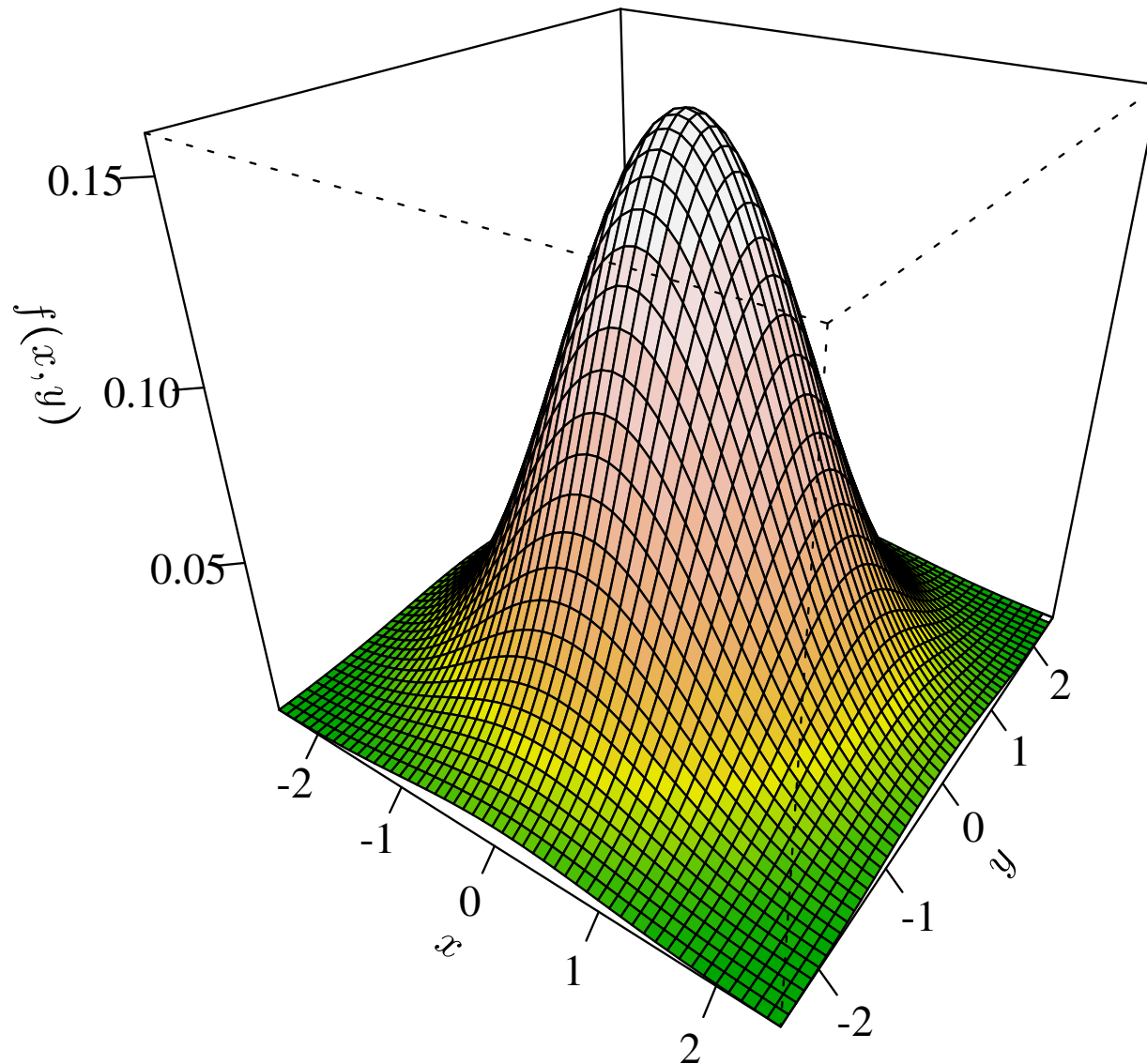
3. Allgemeiner: Die d -dimensionale Standard-Normalverteilung hat Dichte

$$f(x_1, \dots, x_d) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_d^2)\right)$$

$= \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|x\|_2^2\right)$

Die 2-dimensionale Standard-Normalverteilung hat

Dichte $f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$



Die 2-dimensionale Standard-Normalverteilung hat

Dichte $f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$

