Wochen 3 und 4: Verteilungen von Zufallsvariablen

WBL 15/17, 04.05.2015

Alain Hauser <alain.hauser@bfh.ch>

Monty Hall-Problem

US-amerikanische Fernseh-Show "Let's make a deal", moderiert von Monty Hall:

- Hauptgewinn ist hinter einer von 3 Türen versteckt.
- Sie als KandidatIn wählen eine der 3 Türen.
- Der Moderator öffnet Ihnen eine andere Türe; dahinter ist der Hauptpreis *nicht*
- *Wechseln Sie die Tür oder bleiben Sie bei Ihrer Wahl?*

Diskrete Verteilungen

Lernziele

Sie können...

- ...die Verteilung einer diskreten Zufallsvariablen angeben
- ...Erwartungswert und Varianz diskreter Zufallsvariablen berechnen
- ...mit binomial- und Poisson-verteilten Zufallsvariablen rechnen
- ...eine Zählgrösse mit einer passenden Verteilung modellieren

Vorlesung basiert auf Kapitel 2.6 und 2.8 des Skripts.

Modelle für Zähldaten: diskrete Zufallsvariablen

- Die Zufallsvariablen, die wir bisher betrachtet haben, konnten bloss Werte aus einer (möglicherweise unendlich langen) Liste annehmen: $X : \Omega \to \{x_1, x_2, \dots\}$
- Solche Zufallsvariablen nennt man **diskrete Zufallsvariablen**

Wiederholung: diskrete Zufallsvariablen

- Verteilung einer diskreten Zufallsvariablen ist spezifiziert durch die **Wahrscheinlichkeitsverteilung** $P(X = x_1), P(X = x_2), P(X = x_3), \dots$, die jedem möglichen Wert von X die Wahrscheinlichkeit zuordnet, dass er angenommen wird
- Normalisierung: $\sum_k P(X = x_k) = 1$
- Wahrscheinlichkeit, dass X in einer Menge A liegt:

$$P(X \in A) = \sum_{k : x_k \in A} P(X = x_k)$$

- Zusammenhang zur kumulativen Verteilungsfunktion:

$$F_X(x) = P(X \leq x) = \sum_{k : x_k \leq x} P(X = x_k)$$

Wiederholung: Erwartungswert und Varianz

Eine diskrete Zufallsvariable X kann die Werte $x_1, x_2, x_3, \dots$ annehmen.

Definition (Erwartungswert und Varianz)

Der **Erwartungswert** von X ist definiert als

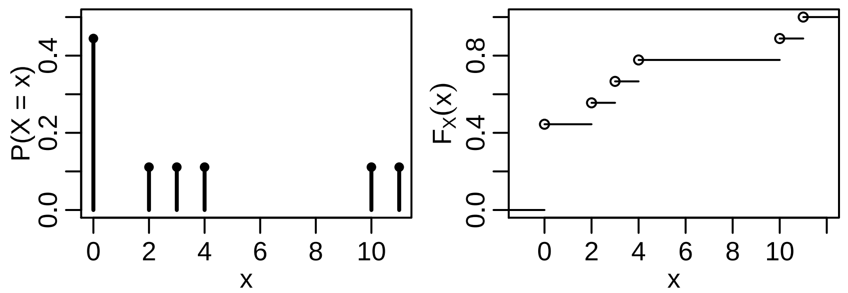
$$E[X] = \sum_k x_k P(X = x_k) \ .$$

Die **Varianz** von X ist definiert als

$$\mathrm{Var}(X) = \sum_k (x_k - E[X])^2 P(X = x_k) \ .$$

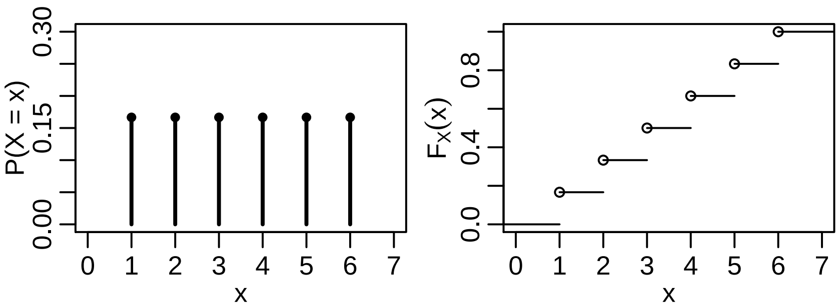
Beispiel: zufällig gezogene Jasskarte

- Zufallsexperiment: Sie ziehen aus den 9 Jasskarten der Farbe "Eichel" zufällig eine heraus; X: Wert der gezogenen Karte.
- Wahrscheinlichkeitsverteilung und kumulative Verteilungsfunktion von X:



Beispiel: fairer Würfel

- Ein Würfel kann die Werte $\{1, 2, \ldots, 6\}$ annehmen; wenn er fair ist, wird jede Zahl mit derselben Wahrscheinlichkeit angenommen.
- Wahrscheinlichkeitsverteilung und kumulative Verteilungsfunktion der gewürfelten Augenzahl X:



Vergleich der Beispiele

- Es scheint plausibel, dass wir eine Wahrscheinlichkeitsverteilung wie die des fairen Würfels öfter verwenden können als die des Werts einer zufällig gezogenen Jasskarte
- Jasskarten-Beispiel: Wahrscheinlichkeitsverteilung genau auf ein sehr spezifisches Beispiel abgestimmt
- Fairer Würfel: Beispiel eines Zufallsexperiments mit diskretem Resultat, bei dem jeder mögliche Ausgang gleich wahrscheinlich ist
- Verteilungen, die in vielen Situationen auftreten, haben spezielle Namen erhalten
- Wir werden ein paar wichtige Verteilungen betrachten:
  - Gleichverteilung
  - Binomialverteilung
  - Poisson-Verteilung

Diskrete Gleichverteilung

- Wir betrachten eine diskrete Zufallsvariable X, die die Werte $1, 2, 3, \ldots, n$ annehmen kann.
- X heisst **gleichverteilt**, wenn alle Werte mit Wahrscheinlichkeit $\frac{1}{n}$ angenommen werden
- Beispiel: fairer Würfel
- Modell für beliebiges Laplace-Experiment, bei dem die verschiedenen Ausgänge durchnummeriert werden

Bernoulli-Verteilung

- Einfachste (nicht-triviale) Zufallsvariable: binäre Zufallsvariable $X \in \{0, 1\}$
- Verwendung: Messung von Erfolg/Misserfolg (Qualitätskontrolle), Ausgang eines medizinischen Tests, Erfassung eines Merkmals in einer Population, etc.
- Verteilung einer binären Variable durch einzelnen Parameter bestimmt: brauche Wahrscheinlichkeit für $X = 1$
- Formal: $P(X = 1) = \pi \in [0, 1] \;\Rightarrow\; P(X = 0) = 1 - \pi$
- Notation: $X \sim \text{Bernoulli}(\pi)$

Anzahl Erfolge in n Versuchen

- Bernoulli-Verteilung zeigt “Erfolg” oder “Misserfolg” in einzelnem Zufallsexperiment an
- Spannendere Grösse bzw. Zufallsvariable: Anzahl “Erfolge” in n unabhängigen Wiederholungen eines Zufallsexperiments
- Beispiel: medizinischer Test stuft gesunde Person mit 10% Wahrscheinlichkeit fälschlicherweise als krank ein. 8 zufällig zusammengekommene, gesunde Personen lassen sich testen.
- *Was ist die Wahrscheinlichkeit, dass keine der 8 Personen fälschlicherweise ein positives Testergebnis erhalten? Was ist die Wahrscheinlichkeit, dass 1, 2 oder 3 der 8 Personen fälschlicherweise als krank eingestuft werden?*

Binomialverteilung I

Eine Zufallsvariable, die die Anzahl “Erfolge” in n unabhängigen Versuchen zählt, heisst **binomialverteilt**.

Definition (Binomialverteilung)

Eine diskrete Zufallsvariable $X \in \{0, 1, \dots, n\}$ hat **Binomialverteilung** mit Parametern n und π, falls

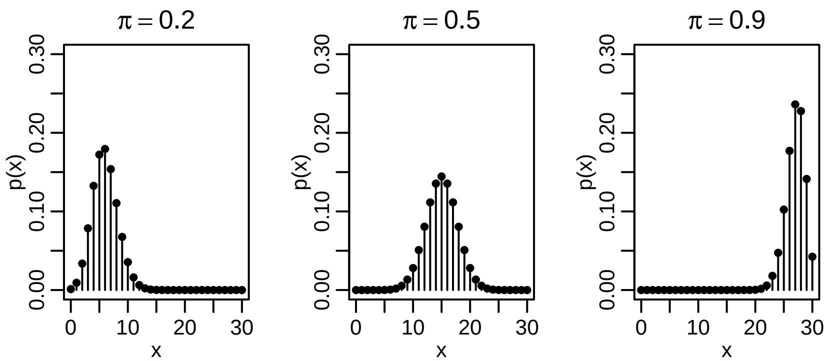
$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \ .$$

Wir schreiben dafür kurz $X \sim \text{Bin}(n, \pi)$ mit $n \in \mathbb{N}$, $\pi \in (0, 1)$.

- π bezeichnet gerade die Wahrscheinlichkeit für einen “Erfolg” in einem *einzelnen* Versuch
- Die Bernoulli-Verteilung ist demnach ein Spezialfall der Binomialverteilung: $\text{Bernoulli}(\pi) = \text{Bin}(1, \pi)$.

Binomialverteilung II

Wahrscheinlichkeitsverteilung der Binomialverteilung $\text{Bin}(30, \pi)$ for verschiedene Werte des Parameters π:



Rechnen mit der Binomialverteilung

X sei eine binomialverteilte Zufallsvariable mit Parametern n und π: $X \sim \text{Bin}(n, \pi)$.

- X hat Erwartungswert $\mathcal{E}(X) = n\pi$
- X hat Varianz $\text{Var}(X) = n\pi(1 - \pi)$

R-Funktionen zum Arbeiten mit Binomialverteilungen:

- `dbinom` liefert die Wahrscheinlichkeitsverteilung, `pbinom` die kumulative Verteilungsfunktion
- Namensgebung folgt allgemeinem Schema: viele Verteilungen in R implementiert! Wahrscheinlichkeitsverteilung erhält man immer mit `dxxxx`, kumulative Verteilungsfunktion mit `pxxxx`, wobei `xxxx` die Verteilung spezifiziert

Beispiel: Test eines neuen Medikaments

- Ein neues Medikament wird in einer Studie an 200 Patienten getestet. Personen mit einer seltenen genetischen Disposition, von der $\frac{1}{1000}$ der Bevölkerung betroffen ist, könnten schwere Nebenwirkungen erleiden.
- *Wie gross ist die Wahrscheinlichkeit, dass mindestens ein Patient in der Studie die genetische Disposition hat?*
- *Wie gross ist die Wahrscheinlichkeit, dass mindestens 3 Patienten die genetische Disposition haben?*

Erfolge in unbeschränkter Anzahl Versuche

- Binomialverteilung: zählt "Erfolge" in beschränkter Anzahl Versuche $\rightsquigarrow$ Wertebereich beschränkt
- Was, wenn wir Anzahl "Erfolge" in einer (potentiell) unbeschränkten Anzahl Versuche zählen wollen?
- Beispiele:
  - Anzahl Anrufe Tag in Telefonzentrale
  - Anzahl Mutationen in der DNA einer Zelle
  - Anzahl Verkehrsunfälle in der Schweiz in einem Jahr
  - etc.
- Für diese Fälle geeignet: **Poisson-Verteilung**

Poisson-Verteilung I

Definition (Poisson-Verteilung)

Eine diskrete Zufallsvariable $X \in \mathbb{N}$ hat **Poisson-Verteilung** mit Parameter λ, falls

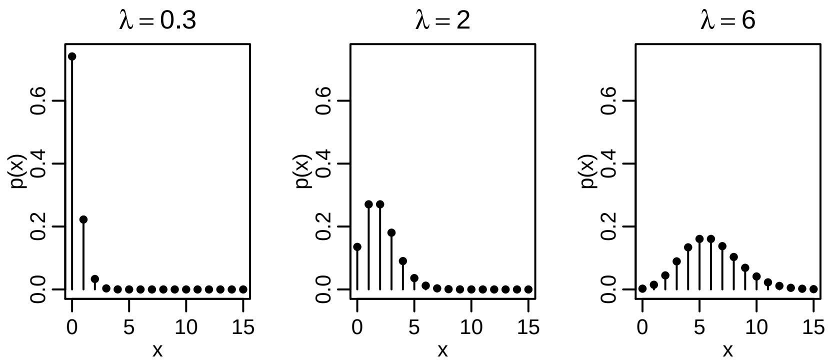
$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} .$$

Wir schreiben $X \sim \text{Pois}(\lambda)$, $\lambda > 0$.

- Poisson-Verteilung modelliert Anzahl *seltener* Ereignisse in einem festen Raum- oder Zeitbereich
- Erwartungswert: $E[X] = \lambda$, Varianz: $\text{Var}(X) = \lambda$
- R-Funktionen: `dpois` (Wahrscheinlichkeitsverteilung), `ppois` (kumulative Verteilungsfunktion)

Poisson-Verteilung II

Poisson-Verteilung für verschiedene Werte des Parameters λ:



Poisson-Verteilung: "Grenzfall" der Binomialverteilung

- Poisson-Verteilung modelliert Anzahl *seltener* Ereignisse in einem festen Raum- oder Zeitbereich
- Formaler: lässt man bei einer Binomialverteilung n ins Unendliche wachsen und π gleichzeitig schrumpfen, so dass das Produkt $n\pi$ stets konstant bleibt, strebt die Binomialverteilung gegen eine Poisson-Verteilung.
- Noch formaler: wir betrachten eine Sequenz $X_n \sim \text{Bin}(n, \pi)$ mit $n\pi = \lambda$ (konstant!). Dann gilt im Grenzwert $n \to \infty$

$$P(X_n = x) \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^x}{x!}$$

Poisson-Approximation der Binomialverteilung

- Praktische Konsequenz dieses Grenzwerts: man kann eine Binomialverteilung für grosses n und kleines π ungefähr durch eine Poisson-Verteilung annähern.
- Vage Faustregel: Poisson-Approximation ist gut für $n \gtrapprox 50$ und $\pi \lessapprox 0.05$
- Beispiel: Anzahl Druckfehler auf einer Buchseite.
  - Es hat höchstens so viele Druckfehler wie Zeichen
  - Druckfehler sind aber selten, und eine Buchseite enthält viele Zeichen
  - Daher ist die Poisson-Verteilung ein gutes Modell: man muss bloss die Fehlerrate pro Seite (λ) schätzen, nicht die Fehlerrate pro Zeichen (π) sowie die Anzahl Zeichen pro Seite (n).

Summe Poisson-verteilter Zufallsvariablen

Satz

Falls X Poisson-verteilt mit Parameter λ_1 und Y Poisson-verteilt mit Parameter λ_2 ist, ist auch $X + Y$ Poisson-verteilt mit Parameter $\lambda_1 + \lambda_2$.

Spezielle Eigenschaft der Poisson-Verteilung! Die Summe zweier gleichverteilten Variablen ist z.B. nicht gleichverteilt (Augensumme bei 2 Würfeln).

Falls $X \sim Pois(\lambda_1)$ und $X \sim Pois(\lambda_2)$: ist dann auch $\frac{1}{2}(X + Y) \sim Pois(\frac{1}{2}(\lambda_1 + \lambda_2))$?

Stetige Zufallsvariablen

Lernziele

Sie können...

- ...das Konzept der Wahrscheinlichkeitsdichte erklären
- ...zu einer Wahrscheinlichkeitsdichte die zugehörige kumulative Verteilungsfunktion berechnen und umgekehrt
- ...Erwartungswert und Varianz stetiger Zufallsvariablen berechnen
- ...Erwartungswert, Varianz und Quantile linear transformierter Zufallsvariablen berechnen
- ...mit den wichtigsten stetigen Verteilungen rechnen: uniforme, Normal- und Exponentialverteilung
- ...eine Messgrösse mit einer passenden Verteilung modellieren

Vorlesung basiert auf Kapitel 4.4 und 4.5 des Skripts.

Modellierung von Messdaten

- Bisher nur diskrete Zufallsvariablen betrachtet, bei denen man alle möglichen Werte auflisten kann
- Messdaten sind nicht von dieser Art: Länge, Masse, Geschwindigkeit etc. können Werte in ganz $\mathbb{R}$ (oder zumindest einem Intervall davon) annehmen
- Entsprechende Zufallsvariablen heissen **stetig**
- Da man bei stetigen Zufallsvariablen nicht die Werte auflisten kann, kann man offensichtlich keine "Liste" $P(X = x_i)$ angeben $\rightsquigarrow$ andere Beschreibung der Verteilung nötig

Stetige Zufallsvariablen

- Stetige Zufallsvariable: Zufallsvariable, die *alle* Werte in $\mathbb{R}$ (oder einem Intervall $[a, b] \subset \mathbb{R}$) annehmen kann.
- Charakterisiert durch ihre **Wahrscheinlichkeitsdichte** (kurz "Dichte") $f_X(x)$: eine nicht-negative Funktion auf $\mathbb{R}$ mit

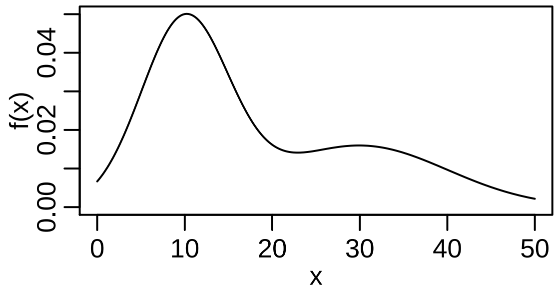
$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$$

- Verwendung der Dichtefunktion f_X: Wahrscheinlichkeiten können als Fläche unter ihrer Kurve dargestellt werden
- D.h.: für jedes Intervall $[a, b]$ gilt

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$$

Beispiel: Regenprognose

- Wie viel Regen wird morgen in Zürich fallen?
- Auf Grund der Prognosemodelle wird morgens eine Störung mit schwachem Regen erwartet.
- Eventuell bildet sich abends eine Gewitterzelle mit lokalem Starkregen.
- Mögliche Dichtefunktion für $X =$ "morgiger Niederschlag (mm)":



Rechnen mit stetigen Zufallsvariablen

X ist eine stetige Zufallsvariable mit Wahrscheinlichkeitsdichte $f_X(x)$.

- Kumulative Verteilungsfunktion: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u$ ("kumulative Verteilungsfunktion ist Stammfunktion der Dichte")
- Umgekehrte Umrechnung: $f_X(x) = F'_X(x)$ ("Dichte ist Ableitung der kumulativen Verteilungsfunktion")
- Für jeden Wert $x \in \mathbb{R}$ gilt $P(X = x) = 0$
- Erwartungswert: $\mathcal{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$
- Varianz: $\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\,\mathrm{d}x$
- Andere Berechnung der Varianz: $\mathrm{Var}(X) = \mathcal{E}(X^2) - \mathcal{E}(X)^2$

Median

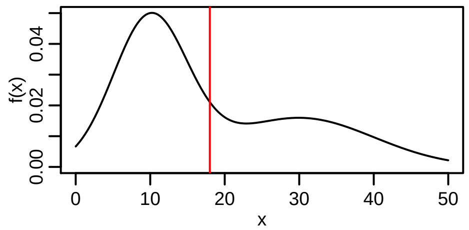
Definition (Median)

Der **Median** einer (stetigen) Zufallsvariablen X ist diejenige reelle Zahl m, für die gilt:

$$P(X \leq m) = \frac{1}{2}, \quad P(X \geq m) = \frac{1}{2}$$

Zusammenhang mit der kumulativen Verteilungsfunktion: der Median ist die Zahl m, für die $F_X(m) = \frac{1}{2}$ gilt.

Beispiel Regenprognose (Forts.): $\mathcal{E}(X) = 18$.
Ist der Median grösser oder kleiner?



Quantile

- Quantile verallgemeinern die Idee des Medians.
- X ist eine (stetige) Zufallsvariable.
- Für eine Zahl $0 < \alpha < 1$ ist das α-**Quantil** von X die Zahl q_α mit der Eigenschaft $P(X \leq q_\alpha) = \alpha$
- Zusammenhang zur kumulativen Verteilungsfunktion: $F(q_\alpha) = \alpha$ oder $q_\alpha = F^{-1}(\alpha)$
- Der Median entspricht gerade dem 0.5-Quantil.

Lineare Transformationen stetiger Zufallsvariablen

- X sei eine stetige Zufallsvariable, a und b reelle Konstanten
- Wir definieren $Y = a \cdot X + b$: lineare Transformation von X
- Wie ist Y verteilt?

Satz (Lineare Transformation einer Zufallsvariablen)

Mit X und Y wie oben gilt:

1. $F_Y(y) = F_X\left(\frac{y-b}{a}\right), \; f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$
2. $\mathcal{E}(Y) = a \cdot \mathcal{E}(X) + b$
3. $\mathrm{Var}(Y) = a^2\,\mathrm{Var}(X)$
4. *Quantile von Y: $q_{Y,\alpha} = a \cdot q_{X,\alpha} + b$ ($q_{X,\alpha}$: α-Quantil von X)*

Hinweis: 2. und 3. gelten auch für diskrete Zufallsvariablen.

Anwendung: Standardisieren einer Zufallsvariablen

- X sei eine Zufallsvariable mit Erwartungswert $\mathcal{E}(X)$ und Standardabweichung $\sigma(X)$
- Wir definieren
$$Z = \frac{X - \mathcal{E}(X)}{\sigma(X)} .$$
- Dann hat Z Erwartungswert $\mathcal{E}(Z) = 0$ und Standardabweichung $\sigma(Z) = 1$
- Transformation nennt sich **Standardisierung** von X

Stetige Verteilungen

Uniforme Verteilung

Stetiges Analogon der Gleichverteilung:

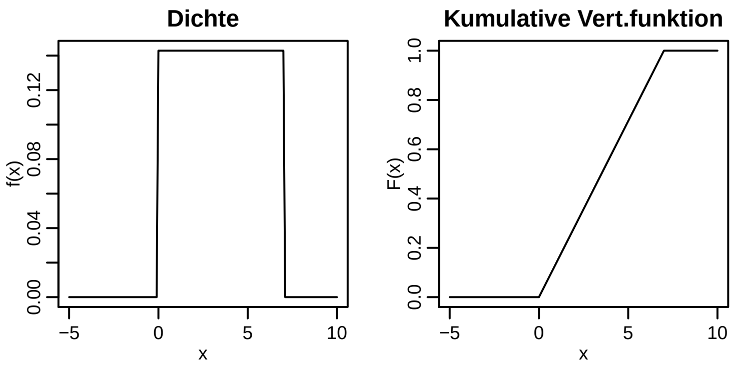
Definition (Uniforme Verteilung)

Eine stetige Zufallsvariable X hat **uniforme Verteilung** auf $[a, b]$, falls
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{sonst.} \end{cases}$$
Wir schreiben $X \sim \text{Uniform}([a, b])$.

Erwartungswert: $\mathcal{E}(X) = \frac{b+a}{2}$, Varianz: $\text{Var}(X) = \frac{(b-a)^2}{12}$
R-Funktionen: `dunif` (Wahrscheinlichkeitsdichte), `punif` (kumulative Verteilungsfunktion)

Uniforme Verteilung: Beispiel

- Auf einer Tramlinie fährt alle 7 Minuten ein Tram.
- Wenn Sie zu einer zufälligen Zeit zu einer Haltestelle kommen, ist Ihre Wartezeit X (in Minuten) aufs nächste Tram uniform verteilt auf dem Intervall $[0, 7]$:



Normalverteilung und Gauss-Verteilung

- Normalverteilung: wichtigste stetige Verteilung, häufigste Verteilung für Messwerte
- Dichte durch bekannte "Gausssche Glockenkurve" beschrieben
- Anwendungen (Beispiele): Modellierung zufälliger Messfehler, individueller physiologischer Messgrössen, etc.
- Verteilung ist wegen des "zentralen Grenzwertsatzes" (s. später) so wichtig: viele kleine, unabhängige Effekte summieren sich zu einem normalverteilten "Gesamteffekt" auf.

Normalverteilung I

Definition (Normalverteilung)

Eine stetige Zufallsvariable X ist **normalverteilt** mit Erwartungswert μ und Varianz σ^2, falls

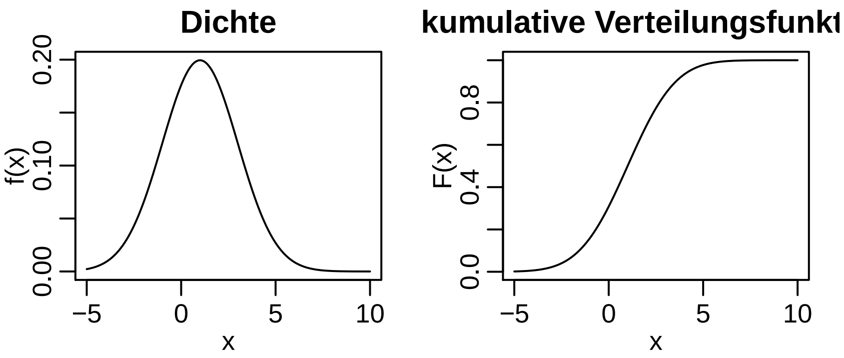
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \; x \in \mathbb{R} \; .$$

Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

- Es gilt $\mathcal{E}(X) = \mu$, $\text{Var}(X) = \sigma^2$.
- R-Funktionen: `dnorm` (Dichte), `pnorm` (kumulative Verteilungsfunktion)
- Achtung: R-Funktionen nehmen als Parameter μ und σ, nicht μ und σ^2!

Normalverteilung II

Normalverteilung mit Erwartungswert $\mu = 1$ und Varianz $\sigma^2 = 4$:



Standard-Normalverteilung

- Normalverteilung mit $\mu = 0$ und $\sigma^2 = 1$ heisst **Standard-Normalverteilung**
- Standard-Normalverteilung wird oft verwendet, z.B. in Tests; daher haben ihre Dichte und kumulative Verteilungsfunktion eigene Symbole:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \int_{-\infty}^{z} \varphi(t)\, dt \; .$$

- $\Phi(x)$ kann nicht als Verknüpfung elementarer Funktionen geschrieben werden
- Eine beliebige normalverteilte Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ kann linear in eine standard-normalverteilte Zufallsvariable umgerechnet werden:

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

Fläche unter der Glockenkurve

Wie gross ist die Wahrscheinlichkeit, höchstens 1, 2 oder 3 Standardabweichungen vom Erwartungswert entfernt zu sein?

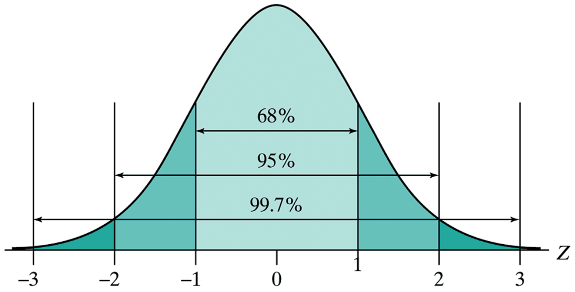


Figure 4.3.5 Areas under a standard normal curve between −1 and +1, between −2 and +2, and between −3 and +3

Quelle: Samuels et al. (2012)

Beispiel: Fertigungstoleranzen bei Widerständen

- Wegen nicht vollständig kontrollierbarer Prozesse bei der Herstellung haben elektrische Widerstände immer Fertigungstoleranzen
- Z.B. kann der tatsächliche elektrische Widerstand eines Bauteils mit Nennwert 220 Ω als normalverteilte Zufallsvariable mit $\mu = 220\ \Omega$ und $\sigma = 11\ \Omega$ aufgefasst werden.
- *Wie gross ist die Wahrscheinlichkeit, dass ein produziertes Bauteil Widerstand von weniger als 200 Ω hat?*
- *Wie gross ist die Wahrscheinlichkeit, dass der Widerstand grösser als 250 Ω ist?*

Exponentialverteilung I

Definition (Exponentialverteilung)

Eine stetige Zufallsvariable $X \geq 0$ ist **exponentialverteilt** mit Parameter λ, falls

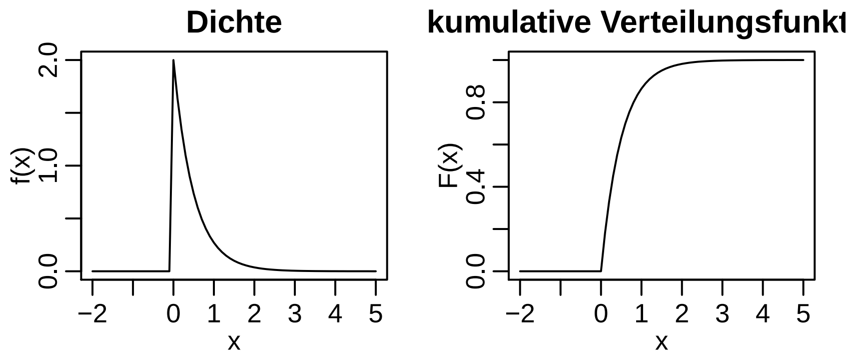
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{sonst.} \end{cases}$$

Wir schreiben $X \sim \text{Exp}(\lambda)$, $\lambda > 0$.

Erwartungswert: $\mathcal{E}(X) = \frac{1}{\lambda}$, Varianz: $\text{Var}(X) = \frac{1}{\lambda^2}$
R-Funktionen: `dexp` (Dichte), `pexp` (kumulative Verteilungsfunktion)

Exponentialverteilung II

Exponentialverteilung für $\lambda = 2$:



Exponentialverteilung: Verteilung “ohne Gedächtnis”

- Exponentialverteilung verwendet zur Modellierung bestimmter “Wartezeiten”:
  - radioaktiver Zerfall eines Atoms
  - Wartezeit auf nächsten Anruf in Telefonzentrale
  - Ausfallzeit von Maschinen, falls Alterung keine Rolle spielt
- T sei exponentialverteilt mit Parameter λ
- $P(T > t) = e^{-\lambda t}$
- $P(T > t + s \mid T > s) = e^{-\lambda t} = P(T > t)$: “erwartete Wartezeit verkürzt sich nicht, wenn man schon Zeit s gewartet hat” $\rightsquigarrow$ gedächtnislose Wartezeit

Literatur

Myra L Samuels, Jeffrey A Witmer, and Andrew Schaffner. *Statistics for the life sciences*. Pearson Education, 2012.