Vom Klassenzimmer zum Flughafen: Die faszinierende Welt der Zufallsvariablen und Flugüberbuchung

Alois Pichler & Dana Uhlig

10. Januar 2025

Inhaltsverzeichnis

1 Einführung

Ziel dieser Fortbildung ist, die Kenntnisse über Zufallsvariablen aufzufrischen, die zur Wahrscheinlichkeitstheorie gehören. In Abbildung 1 ist eine grobe Aufteilung der Sochastik zu sehen. Während in der Statistik Daten in Form von Stichproben ausgewertet werden, um Rückschlüsse über das

Zufallsverhalten zu gelangen, werden in der Wahrscheinlichkeitstheorie Zufallsmodelle betrachtet, die die Spielregeln des Zufalls vorgeben.

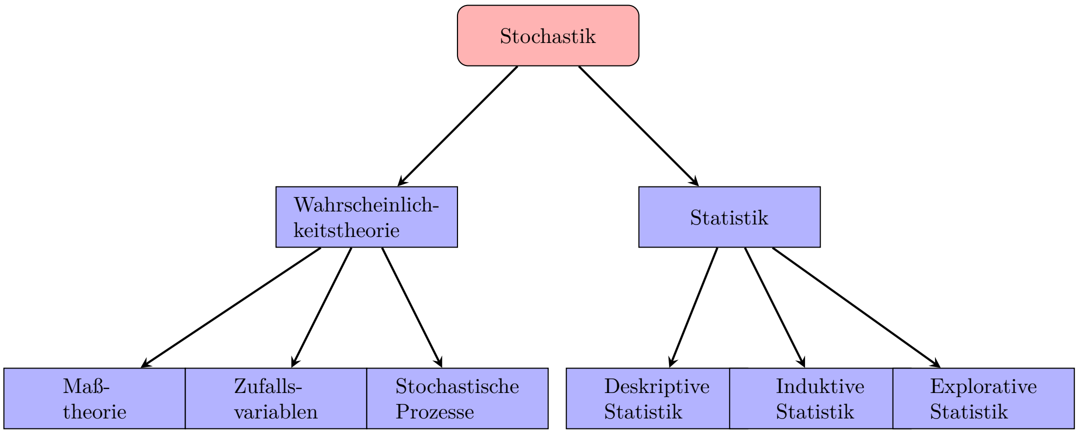


Abbildung 1: Teilgebiete der Stochastik

Statistik und Wahrscheinlichkeitstheorie sind also sehr eng miteinander verknüpft und ergänzen sich gewissermaßen. Die zur Auswertung beobachteter Daten bzw. Stichproben typischerweise benutzen Werkzeuge wie empirischer Verteilungsfunktion, empirischen Maßzahlen wie Mittelwert, Varianz und allgemeinen Momenten finden sich auch in sehr ähnlicher Weise bei der Beschreibung des Zufallscharakters von Zufallsgrößen wieder. Aus diesem Grund werden im Anhang in Abschnitt B die klassischen Werkzeuge zur Auswertung von Stichproben beginnend mit der graphischen Darstellung von Häufigkeiten bis zur Berechnung empirischer Lage- und Streuungsmaße sowie Maßen für die Schiefe und Wölbung einer Verteilung aufgefrischt. Fast alle Interpretationen der deskriptiven Statistik lassen sich einfach auf Zufallsvariablen übertragen, die wir dann in Abschnitt 2 diskutieren. Hierzu wiederholen wir die wichtigsten Konzepte im Umgang mit Zufallsgrößen und besprechen klassische diskrete und stetige Verteilungen und zeigen den Zusammenhang zur empirischen Stichprobenauswertung auf.

Entsprechend den sächsischen Lehrplänen fokussieren wir hier besonders auf die Binomial- sowie Normalverteilung, werden aber auch etwas über den Tellerrand hinausschauen und weitere Verteilungen kurz andiskutieren.

Während in der deskriptiven Statistik die konkret vorliegende Stichprobe ausgewertet wird und klassische empirische Kennwerte die beobachtete Verteilung charakterisieren, können Zufallsvariablen und deren Verteilungen als Modell für das Zufallsverhalten betrachtet werden. Anstatt Häufigkeiten und empirischen Maßzahlen werden in der Stochastik Wahrscheinlichkeiten und Momente berechnet, die uns zwar keine exakte Prognose liefern können, uns aber dennoch veranschaulichen, was zufällig passieren kann und hilft uns besser zufällige Vorgänge zu verstehen und einzuordnen.

In der induktiven Statistik geht man noch einen Schritt weiter und kann mit Hilfe solcher Wahrscheinlichkeitsmodelle das Zufallsverhalten von Stichproben beschreiben und folglich Rückschlüsse von den konkreten Daten einer (repräsentativen) Stichprobe auf die Allgemeinheit ziehen. Dies ist allerdings nicht Thema des vorliegenden Fortbildungskurses.

Aus Zeitgründen steigen wir direkt bei Zufallsgrößen bzw. Zufallsvariablen ein und setzen die Grundlagen zum Arbeiten mit Ereignissen und Wahrscheinlichkeiten voraus.

Typische Aufgaben werden besprochen und auch wie diese mit GeoGebra in der Schule erarbeitet werden können. Außerdem zeigen wir, wie mit der freien Statistiksoftware[1] R die Aufgaben einfach umgesetzt werden können. Der entsprechende Code ist für das Selbststudium mit R in grauen Boxen angegeben und kann einfach ignoriert werden. Wir versuchen durch zahlreiche Simulationsbeispiele die Zufallsmodelle zu visualisieren.

Zur Vertiefung des Stoffes gibt es zahlreiche sehr gute Lehrbücher und Tutorials im Internet. Wir empfehlen zum Einstieg

1. Software
    - Geogebra Classic `https://www.geogebra.org/classic#probability`

[1]Hierfür sollten Sie zunächst R via `http://www.r-project.org/` installieren und zur Bearbeitung empfehlen wir die Installation von RStudio `https://posit.co/download/rstudio-desktop/`

- R mit RStudio als Editor `https://posit.co/download/rstudio-desktop/`

2. Literatur
   - Statistik: Eine interdisziplinäre Einführung mit interaktiven Elementen [MS23]
   - Stochastik rezeptfrei unterrichten [HMS21]
   - Stochastik für das Lehramt [Lin14]
   - ...

3. Tutorials
   - Daniel Jung auf Youtube: [Jun14a], [Jun14f], [Jun14b], [Jun14d], [Jun14c], [Jun14e], ...
   - ...

Außerdem können Sie sich in unseren Newsletter für Mathematik-Lehrer unter `https://lists.tu-chemnitz.de/mailman3/postorius/lists/math-teachers.lists.tu-chemnitz.de/` eintragen, bei dem wir immer über aktuelle Veranstaltungen unserer Mathematikfakultät für Mathematiklehrer informieren möchten.

2 Zufallsgrößen

Die Modellierung des Zufalls geschieht mit Hilfe eines **Wahrscheinlichkeitsraumes** $(\Omega, \mathcal{A}, P)$, wobei Ω die Menge aller Versuchsausgänge des entsprechenden Zufallsexperiments, $\mathcal{A}$ das Ereignissystem ($\sigma-$Algebra) und P das Wahrscheinlichkeitsmaß bezeichnen.

Eine reelle **Zufallsvariable** bzw. **Zufallsgröße** X ist eine (messbare) Abbildung

$$X : \Omega \to \mathbb{R}\,, \tag{1}$$

die jedem Ausgang des Zufallsexperiments $\omega \in \Omega$ eine reelle Zahl $X(\omega) \in \mathbb{R}$ zuordnet, so dass für jedes $x \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A} \tag{2}$$

gilt.

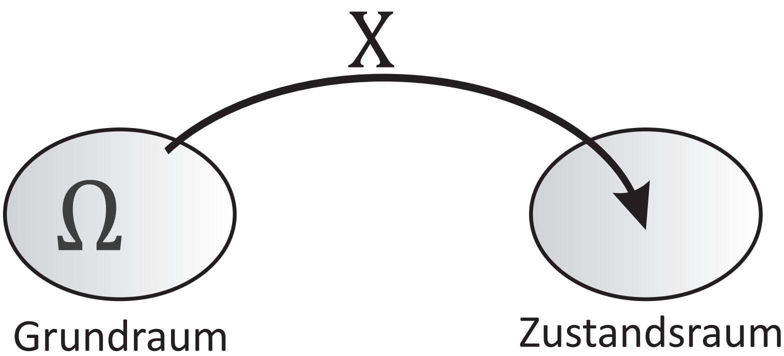
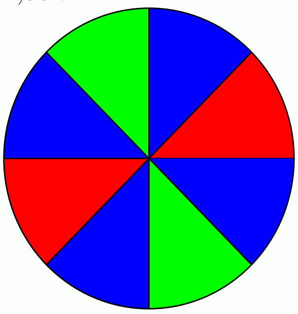


Abbildung 2: reelle Zufallsvariable: (messbare) Abbildung vom Ereignsraum auf den Zustandsraum $\mathbb{R}$

Die Bedingung (2) der Messbarkeit sichert letztendlich, dass für jedes beliebige Intervall $I \subset \mathbb{R}$ die Wahrscheinlichkeit $P(X \in I)$ berechnet werden kann, da dann die entsprechenden Urbilder der Abbildung $X^{-1}(I) = \{\omega : X(\omega) \in I\}$ Ereignisse sind, die zum Ereignissystem $\mathcal{A}$ gehören.

Beispiel 2.1: Glücksrad

Kosten 1 mal Drehen: 2,50€



Gewinnverteilung:

rot	*grün*	*blau*
5€	*2€*	*1€*

Zufallsmodell

Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit
Ereignisraum $\Omega = \{\underbrace{\text{rot}}_{=\omega_1}, \underbrace{\text{grün}}_{=\omega_2}, \underbrace{\text{blau}}_{=\omega_3}\}$
Ereignissystem $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$
Wahrscheinlichkeit P mit $P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{4}$ und $P(\{\omega_3\}) = \frac{1}{2}$

Zufallsgröße X ($\triangleq$ Gewinn)

$$X(\omega) = \begin{cases} 5, & \text{falls } \omega = \omega_1 \quad (\triangleq \text{rot}) \\ 2, & \text{falls } \omega = \omega_2 \quad (\triangleq \text{grün}) \\ 1, & \text{falls } \omega = \omega_3 \quad (\triangleq \text{blau}) \end{cases}$$

Rechnen mit Zufallsgrößen, Bsp für $I = (-\infty, 2]$

$P(X \leq 2) = P(\{\omega_2, \omega_3\}) = P(\{\omega_2\} \cup \{\omega_3\}) = P(\{\omega_2\}) + P(\{\omega_3\}) = 1/4 + 1/2 = 3/4$
$= 1 - P(\{\omega_2, \omega_3\}) = 1 - P(\{\omega_1\}) = 1 - 1/4$

Das zufällige Verhalten der Zufallsgröße X wird durch die Funktion $F : \mathbb{R} \to \mathbb{R}$ mit

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega : \ X(\omega) \leq x\}) \tag{3}$$

beschrieben.

Definition 2.1: Verteilungsfunktion

Die Funktion (3) wird als ***Verteilungsfunktion*** *der Zufallsgröße X bezeichnet und gibt für jedes Argument $x \in \mathbb{R}$ die Wahrscheinlichkeit an, dass die Zufallsvariable X höchstens den Wert x annimmt.*

Beispiel 2.2: Fortsetzung Glücksrad, Beispiel 2.1

Für die Zufallsgröße X, die den Gewinn beim Drehen des Glücksrades aus Beispiel 2.1 beschreibt, gilt

$$P(X = x) = \begin{cases} 1/4, & \text{falls } x = 2 \text{ oder } x = 5 \\ 1/2, & \text{falls } x = 1 \\ 0, & \text{sonst.} \end{cases} \tag{4}$$

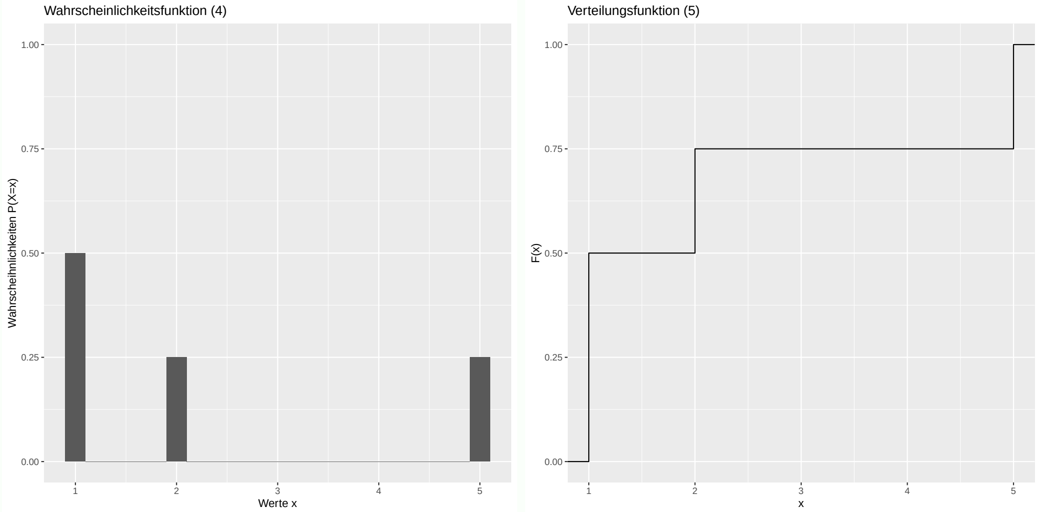
Der Wert der Verteilungsfunktion an der Stelle 3 lässt sich mit

$$F(3) = P(X \leq 3) = P(X = 1) + P(X = 2) = 1/2 + 1/4 = 3/4$$

einfach berechnen und sagt aus, dass die Wahrscheinlichkeit beim Drehen des Glücksrades höchstens 3€ zu gewinnen 75 Prozent beträgt. Allgemein lautet die Verteilungsfunktion des Gewinns

$$F(x) = \begin{cases} 0, & \text{falls } x < 1 \\ 1/2, & \text{falls } 1 \leq x < 2 \\ 3/4, & \text{falls } 2 \leq x < 5 \\ 1, & \text{falls } x \geq 5\,, \end{cases} \tag{5}$$

deren Verlauf in der folgenden rechten Grafik dargestellt ist.



Da die Zufallsgröße X nur die drei Werte $1, 2$ oder 5 annimmt, wächst die Verteilungsfunktion nur in diesen Punkten genau um die Wahrscheinlichkeiten, mit denen diese Punkte angenommen werden. Die Sprunghöhen der Verteilungsfunktion (5) entsprechen also genau den Wahrscheinlichkeiten der einzelnen Werte der Zufallsgröße. Die komplette Information der Wahrscheinlichkeitsfunktion (4) ist also in der Verteilungsfunktion (5) enthalten.

Beispiel 2.3: Augensumme zweier Würfel

Wahrscheinlichkeitsfunktion:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\sum_{k=1}^{i} p_k$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

$\Rightarrow$ *Verteilungsfunktion*

$$F(x) = \sum_{x_i \leq x} p_i = \begin{cases} 0\,, & x < 2 \\ \frac{1}{36}\,, & 2 \leq x < 3 \\ \frac{3}{36}\,, & 3 \leq x < 4 \\ \vdots & \vdots \\ \frac{35}{36}\,, & 11 \leq x < 12 \\ 1\,, & x \geq 12 \end{cases}$$

Die Funktion (4) aus Beispiel 2.2, die jedem Wert $x \in \mathbb{R}$ seine Wahrscheinlichkeit $P(X = x)$ zuordnet, nennt man auch Wahrscheinlichkeitsfunktion. Allerdings beschreibt diese nicht für jede beliebige Zufallsvariable das komplette Zufallsverhalten. Man unterscheidet je nach Wertebereich der Zufallsvariablen **diskrete** und **stetige** Zufallsgrößen. Der Wertebereich diskreter Zufallsvariablen ist abzählbar, während er für stetige Zufallsvariablen überabzählbar[2] ist. Für eine stetige Zufallsvariable mit unendlich vielen Werten, die sogar überabzählbar sind, können nur endlich viele Punkte eine positive Wahrscheinlichkeit annehmen. Für die meisten stetigen Zufallsvariablen, kann das Zufallsverhalten mit Hilfe der Dichtefunktion beschrieben werden, was als Verallgemeinerung der Zähldichte interpretiert werden kann.

Die Zufallsgröße aus Beispiel 2.1 nimmt genau drei Werte an, so dass sie zur Klasse der diskreten Zufallsgrößen zählt. Allgemein charakterisiert die Wahrscheinlichkeitsfunktion $x_i \mapsto p_i = P(X = x_i)$ jeder diskreten Zufallsgröße, die abzählbaren Werte $x_1, x_2, \ldots$ annimmt, ebenso wie die Verteilungsfunktion

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$$

das komplette Zufallsverhalten. Für stetige Zufallsgrößen ist eine Betrachtung der Wahrschein-

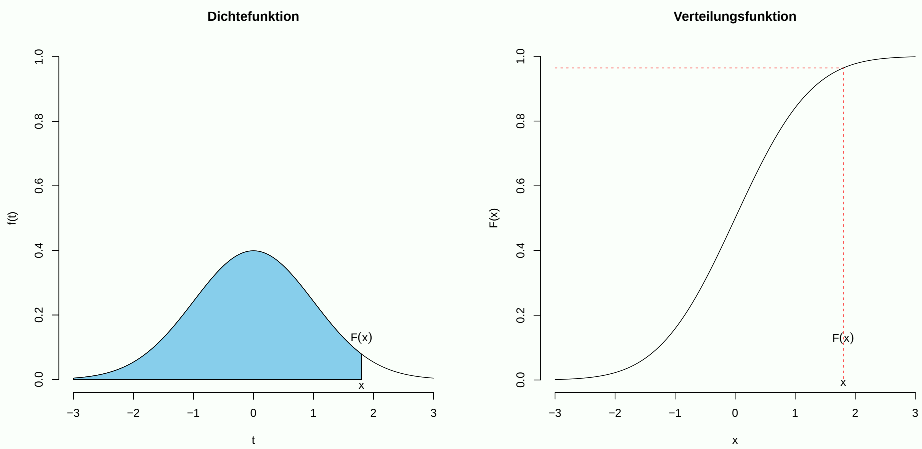
[2]Es gelingt nicht alle Werte einer stetigen Zufallsvariablen durchzunummerieren. Misst die Zufallsvariable X beispielsweise die Entfernung zweier Punkte, so kann theoretisch jeder nicht negative Wert gemessen werden, so dass $X \in \mathbb{R}_+$ gilt. Selbst wenn aus technischen Gründen nur Werte zwischen eins und zehn gemessen werden können, enthält das Intervall $[1, 10]$ unendlich viele reelle Zahlen, die überabzählbar sind.

lichkeitsfunktion allerdings nicht ausreichend, um das Zufallsverhalten vollständig zu beschreiben. Typischerweise gilt für stetige Zufallsvariablen vielmehr sogar $P(X = x) = 0$ für überabzählbar viele Punkte $x \in \mathbb{R}$. Falls für die stetige Zufallsgröße X mit Verteilungsfunktion F eine integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}_+$ mit

$$F(x) = \int_{-\infty}^{x} f(t)\,dt\,, \tag{6}$$

existiert, gehört diese Zufallsvariable zur Klasse der absolut stetigen Zufallsgrößen und die Funktion f aus Gleichung (6) wird als **Dichtefunktion** bezeichnet. Beispiel 2.4 zeigt den Zusammenhang zwischen Dichte- und Verteilungsfunktion (absolut) stetiger Zufallsgrößen am Beispiel der Standardnormalverteilung.

Beispiel 2.4: Standardnormalverteilung



Die blaue Fläche unter der Dichtefunktion f bis zum Punkt x in der linken Grafik, genauer

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

entspricht der Verteilungsfunktion im Punkt x, welche in der rechten Grafik abgetragen ist. Verschiebt man den Punkt x auf der Achse der Dichtefunktion nach links. wird die blaue Fläche kleiner. Verschiebt man in der rechten Grafik den Punkt x ebenso nach links, verringert sich der Wert der Verteilungsfunktion entsprechend. Wird der Punkt x hingegen nach rechts verschoben, wächst die blaue Fläche, so dass die Verteilungsfunktion ebenso größer wird. Der Anstieg der Verteilungsfunktion entspricht somit genau der Dichte, da die Umkehrung $F'(x) = f(x)$ für alle Punkte $x \in \mathbb{R}$ gilt.

Dichtefunktionen absolut stetiger Zufallsvariablen verallgemeinern das Konzept der Wahrscheinlichkeitsfunktion. Während für diskrete Zufallsgrößen mit Werten $x_1, x_2, \ldots$ die Wahrscheinlichkeit Werte aus dem Intervall $(a, b]$ anzunehmen

$$P(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} p_i$$

gerade der Summe über alle Wahrscheinlichkeiten $p_i = P(X = x_i)$ aller Werte x_i, die im Intervall $(a, b]$ liegen, entspricht, gilt für absolut stetige Zufallsvariablen mit Dichtefunktion f hingegen

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(t)\,dt.$$

Beispiel 2.5 veranschaulicht diesen Zusammenhang für die stetige Gleichverteilung.

Beispiel 2.5: Gleichverteilung

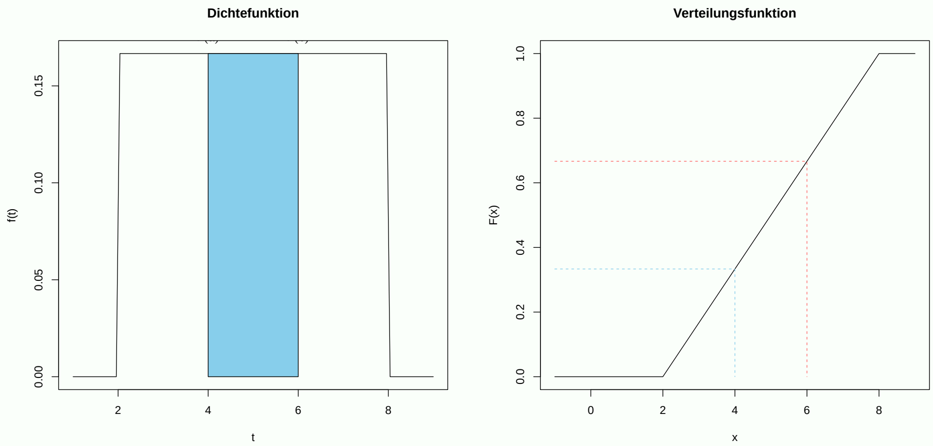
Die Zufallsgröße X mit Dichtefunktion

$$f(t) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq t \leq b \\ 0 & \text{für } t \notin [a,b] \end{cases} \tag{7}$$

heißt Gleichverteilung auf dem Intervall $[a,b]$ $(a < b)$. Die Verteilungsfunktion steigt linear auf dem Intervall mit $[a,b]$

$$F(x) = \int\limits_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases} \tag{8}$$

Die folgende Grafiken zeigen die Gleichverteilung auf dem Intervall $[2,8]$, so dass die Dichte auf diesem Intervall konstant $1/6$ ist.



Die blaue Fläche unter der Dichtefunktion f zwischen $x = 4$ und $x = 6$ entspricht der Wahrscheinlichkeit

$$P(4 \leq X \leq 6) = \int\limits_{4}^{6} f(t)\,dt = F(6) - F(4) = 0.6667 - 0.3333 = 0.3333.$$

Bemerkung 2.1: Dichtefunktion vs Wahrscheinlichkeitfunktion

Dichtefunktionen absolut stetiger Zufallsvariablen verallgemeinern das Konzept der Wahrscheinlichkeitsfunktion. Teilweise wird die Wahrscheiblichkeitsfunktion auch als diskrete Zähldichte bezeichnet. Während für diskrete Zufallsgrößen mit Werten $x_1, x_2, \ldots$ die Wahrscheinlichkeit Werte aus dem Intervall $(a,b]$ anzunehmen

$$P(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} p_i$$

gerade der Summe über alle Wahrscheinlichkeiten $p_i = P(X = x_i)$ aller Werte x_i, die im Intervall $(a,b]$ liegen, entspricht, gilt für absolut stetige Zufallsvariablen mit Dichtefunktion f hingegen

$$P(a < X \leq b) = F(b) - F(a) = \int\limits_{a}^{b} f(t)\,dt\,.$$

Beispiel 2.5 veranschaulicht diesen Zusammenhang für die stetige Gleichverteilung.

Außerdem gilt für alle Wahrscheinlichkeiten der diskreten Zufallsvariablen $0 \leq p_i \leq 1$ sowie $\sum\limits_i p_i = 1$ und in Analogie für die Dichtefunktion einer stetigen Zufallsvariablen $f(t) \geq 0$ für alle Punkte $t \in \mathbb{R}$ und $\int\limits_{-\infty}^{\infty} f(t)\,dt = 1$. Allerdings unterliegen Dichtefunktionen nicht mehr der Einschränkung $f(t) \leq 1$, da diese nicht direkt Wahrscheinlichkeiten beschreiben, sondern die Fläche unter der

Dichte Wahrscheinlichkeiten entsprechen. Für absolut stetige Zufallsgrößen gilt ferner für jedes $x \in \mathbb{R}$ die Beziehung

$$P(X = x) = \int_x^x f(t)\,dt = F(x) - F(x) = 0\,,$$

so dass kein einzelner Punkt $x \in \mathbb{R}$ mit $P(X = x) > 0$ existiert.[3]

Zusammenfassend ...

Bemerkung 2.2: diskrete vs stetige Zufallsvariablen

Diskrete Zufallsgröße mit Wahrscheinlichkeitsfunktion

*Ist der Wertebereich der Zufallsgröße X endlicher oder abzählbar, d.h. X nimmt Werte $x_1, x_2, \ldots, x_n$ bzw. $x_1, x_2, \ldots$ an, so gehört X zur Klasse der **diskreten Zufallsgrößen**. Die Funktion, die jedem Wert seine Wahrscheinlichkeit $x_i \mapsto p_i = P(X = x_i)$ zuordnet, heißt **Wahrscheinlichkeitsfunktion**, so dass die Verteilungfunktion*

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$$

treppenförmig verläuft und jede Wahrscheinlichkeitsfunktion erfüllt die folgenden Eigenschaften

1. $0 \leq p_i \leq 1$
2. $\sum\limits_i p_i = 1$
3. $P(a < X \leq b) = F(b) - F(a) = \sum\limits_{a < x_i \leq b} p_i$

Beispiele für diskrete Verteilungen sind

- *diskrete Gleichverteilung, z.B. gilt beim Werfen eines fairen Würfels für die Augenzahl X gerade $P(X = k) = 1/6$ für $k = 1, 2, \ldots, 6$*
- *Binomialverteilung, siehe Abschnitt 3*
- *Hypergeometrische Verteilung, siehe Abschnitt 5.1*
- *Geometrische Verteilung*
- *Poisson-Verteilung*
- ...

(Absolut) stetige Zufallsgröße mit Dichtefunktion

*Ist der Wertebereich der Zufallsgröße X überabzählbar, so zählt X zur Klasse der **stetigen Zufallsgrößen**. Gibt es ferner eine Dichtefunktion (6) mit*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt\,,$$

*d.h. die Wahrscheinlichkeiten werden indirekt als Fläche unter der Dichte beschrieben, so wird diese als **absolut stetige Zufallsröße** bezeichnet. In diesem Fall ist die Verteilungsfunktion immer stetig. Häufig wird allerdings die Ergänzung „absolut“ weglassen. Für stetige Zufallsvariablen X mit Dichtefunktion f gilt*

1. *Die Verteilungsfunktion F ist stetig in allen Punkten $x \in \mathbb{R}$.*
2. *Ist f an der Stelle t_0 stetig, so ist F in t_0 differenzierbar und es gilt $F'(t_0) = f(t_0)$.*
3. $P(a < X \leq b) = \int\limits_{-\infty}^{b} f(t)\,dt - \int\limits_{-\infty}^{a} f(t)\,dt = \int\limits_a^b f(t)\,dt \quad = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)$
4. $P(X = a) = P(a \leq X \leq a) = \int\limits_a^a f(t)\,dt = 0$
5. $F(x) = P(X \leq x) = P(X < x)$
6. *ferner gilt für die Dichte f*
   (a) $f(t) \geq 0$ *für alle* $t \in \mathbb{R}$
   (b) $\int\limits_{-\infty}^{\infty} f(t)\,dt = 1$

[3] Es gibt natürlich stetige Zufallsvariablen, für die einzelne Werte $x_0 \in \mathbb{R}$ mit positiver Wahrscheinlichkeit $P(X = x_0) > 0$ existieren können. Die Verteilungsfunktion springt dann in den singulären Punkt x_0 um diese Wahrscheinlichkeit $P(X = x_0) > 0$. Für solche Zufallsvariablen existiert dann allerdings keine Dichtefunktion mit der Eigenschaft (6). Häufig wird die letzte Klasse der stetigen, aber nicht absolut stetigen Zufallsvariablen nicht weiter betrachtet und dann typischerweise die Klasse der absolut stetigen Zufallsvariablen einfach nur als stetige Zufallsvariablen bezeichnet.

(c) $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\,dx$

und umgekehrt gilt, dass jede integrierbare Funktion f, die die Eigenschaften 6a) und 6b) hat, Dichtefunktion einer gewissen stetigen Zufallsvariablen X ist.

Beispiele für stetige Verteilungen sind

- *Gleichverteilung, siehe Beispiel* **??**
- *Exponentialverteilung, siehe Abschnitt 5.2*
- *Normalverteilung, siehe Abschnitt 4*
- *Gamma-Verteilung*
- ...

Gemischte Version

Neben absolut stetigen Zufallsgrößen mit stetiger Verteilungsfunktion und diskreten Zufallsgrößen mit stückweise konstanter Verteilungsfunktion gibt es noch weitere Verteilungen, die hier allerdings keine weitere Rolle spielen sollen. Daher verstehen wir im Folgenden in diesem Skript unter stetigen Zufallsvariablen immer den Fall absolut stetiger Zufallsvariablen mit stetiger Verteilungsfunktion als Zuwachs über die Dichte, da für alle Stetigkeitspunkte x der Dichte immer $f(x) = F'(x)$ gilt.

Ein allgemeiner Überblick über typische Verteilungen findet sich auf Wikipedia

- *deutsche Seite* `https://de.wikipedia.org/wiki/Liste_univariater_Wahrscheinlichkeitsverteilungen`
- *engl.* `https://en.wikipedia.org/wiki/List_of_probability_distributions`

Zudem sind die meisten Verteilungen in der Statistiksoftware R `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Distributions.html` *oder auch in Geogebra* `https://www.geogebra.org/classic#probability` *implementiert. Dort können die jeweiligen Wahrscheinlichkeitsfunktionen oder Dichten sowie deren Verteilungsfunktionen für konkrete Parameter dargestellt werden und Wahrscheinlichkeiten berechnet werden. Die geometrische Verteilung ist für den Parameter $n = 1$ ein Spezialfall[a] der* ***Pascal*** *Verteilung, die häufiger als negative Binomialverteilung bekannt ist.*

[a] siehe `https://en.wikipedia.org/wiki/Negative_binomial_distribution`

Aufgabe 2.1

Schreiben Sie für die diskrete Zufallsvariable $X \in \{0, 1, 2, \ldots\}$ mit Hilfe der Verteilungsfunktion F die folgenden Wahrscheinlichkeiten

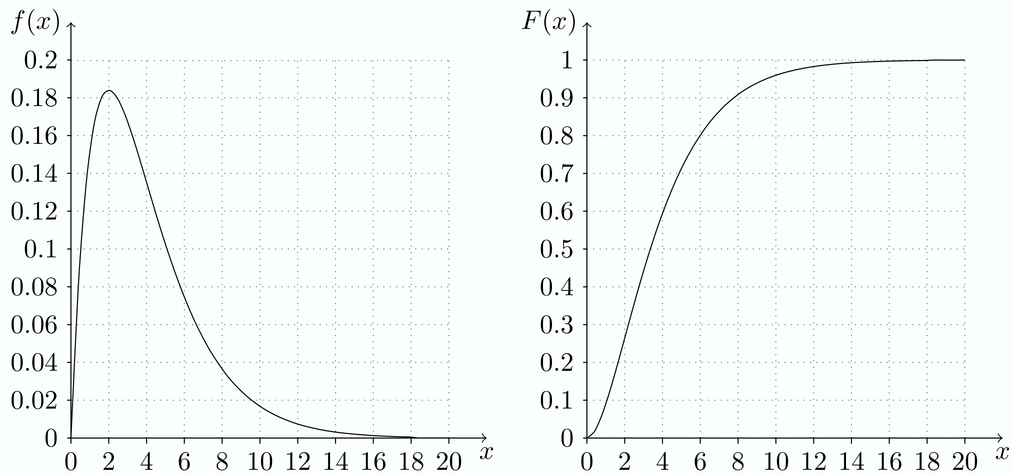
(a) $P(4 \leq X \leq 8)$
(b) $P(2 \leq X < 5)$
(c) $P(4 < X < 7)$
(d) $P(0 \leq X < 8)$

Lösung auf Seite 49

Aufgabe 2.2

Die Dichte- und die Verteilungsfunktion einer stetigen Zufallsvariablen X sei gegeben durch:

$$f(x) = \begin{cases} \frac{1}{4} x e^{-x/2} & \text{für } x > 0 \\ 0 & \text{sonst,} \end{cases} \qquad F(x) = \begin{cases} 1 - (1 + \frac{x}{2}) e^{-x/2} & \text{für } x > 0 \\ 0 & \text{sonst.} \end{cases}$$



1. *Weisen Sie nach, dass die Funktion F tatsächlich Verteilungsfunktion der Dichte f ist.*
2. *Bestimmen Sie grafisch die Wahrscheinlichkeiten $P(X \leq 2)$, $P(X > 5)$ und $P(2 \leq X \leq 5)$ mit Hilfe der Dichte- bzw. Verteilungsfunktion. Bestimmen Sie zusätzlich den exakten Wert.*
3. *Bestimmen Sie grafisch dasjenige $x \in \mathbb{R}$, so dass $F(x) = p$ mit $p = 0.25$, $p = 0.5$ und $p = 0.75$. Interpretieren Sie die Ergebnisse inhaltlich.*

Lösung auf Seite 49

2.1 Erwartungswert, Varianz und weitere Verteilungskennzahlen

Häufig betrachtet man neben der Verteilungsfunktion, welche das zufällige Verhalten von Zufallsvariablen vollständig beschreibt,[4] Kenngrößen wie Erwartungswert, Varianz, allgemeine höhere Momente oder Quantile, mit deren Hilfe die Verteilung einfacher charakterisiert werden kann.

2.1.1 Erwartungswert

Allgemein ist der Erwartungswert für alle Zufallsvariablen als (Lebesgue) Integral bezüglich dem Wahrscheinlichkeitsmaß P bzw. (Lebesgue-Stieltjes) bzgl. der Zuwächse der Verteilungsfunktion

$$\mathbb{E}X = \int_\Omega X(\omega) P(d\omega) = \int_\mathbb{R} x \, dF(x)$$

definiert. Für diskrete Zufallsvariablen reduziert sich das Integral einfach auf eine Summe bezüglich der Wahrscheinlichkeitsfunktion und für stetige zu einem gewöhnlichen Riemann Integral bzgl. der Dichte. Allerdings wollen wir diese komplizierten Lebesgue-Integrale nicht einführen und führen stattdessen den Erwartungswert für diskrete und stetige Zufallsvariablen in Abhängigkeit von Wahrscheinlichkeitsfunktion bzw. Dichte getrennt ein.

Definition 2.2: Erwartungswert

Für die diskrete Zufallsvariable mit Werten $x_1, x_2, \ldots$ und zugehörigen Wahrscheinlichkeiten $p_i = P(X = x_i)$ wird der Erwartungswert als Summe aller Werte, gewichtet mit jeweiliger Wahrscheinlichkeit definiert

$$\mathbb{E}X = \sum_i x_i P(X = x_i) = \sum_i x_i p_i \tag{9}$$

Für stetige Zufallsvariablen mit Dichtefunktion f findet die Durchschnittsbildung analog mittels Integralen statt

$$\mathbb{E}X = \int_{-\infty}^{\infty} x \, f(x) \, dx \tag{10}$$

[4] Natürlich sagt uns diese nicht voraus, welche Realisierung $X(\omega)$ im nächsten Zug auftreten wird, allerdings können wir einschätzen wie wahrscheinlich es ist, dass die nächste Realisierung beispielsweise im Intervall $(a, b]$ liegt, denn es gilt $P(X \in (a, b]) = F(b) - F(a)$. Siehe hierzu auch die Simulation in Beispiel **??** auf Seite **??**.

Falls die Zufallsvariable X unendliche viele Werte annehmen kann, muss der Erwartungswert nicht existieren, also endlich sein. Der Erwartungswert ist eine endliche reelle Zahl, falls $\sum_i |x_i| p_i < \infty$ bzw. $\int_{-\infty}^{\infty} |x|\, f(x)\, dx < \infty$.

Beispiel 2.6: Erwartungswert der Augensumme beim Würfeln

- *X beschreibe die Augenzahl beim Würfeln:* $\mathbb{E}X = \sum_{i=1}^{6} x_i p_i = \frac{1}{6} \sum_{i=1}^{6} i = \frac{21}{6} = 3.5$
- *X beschreibe die Augensumme zweier Würfel*

i	1	2	3	4	5	6	7	8	9	10	11
x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabelle 1: Wahrscheinlichkeitsfunktion der Augensumme zweier Würfel

$\Rightarrow \mathbb{E}X = \sum_{i=1}^{11} x_i p_i = 7 = 3.5 + 3.5$, *d.h. der Erwartungswert der Augensumme zweier Würfel entspricht der Summe der Erwartungswerte beider Würfel* X_1 *und* X_2, *d.h. für* $X = X_1 + X_2$ *folgt* $\mathbb{E}X = \mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2$

Beispiel 2.7: Erwartungswert der Gleichverteilung aus Beispiel 2.5

Für die stetige Gleichverteilung ist die Dichte außerhalb des Intervalls $[a, b]$ *immer Null, so dass das Integral außerhalb verschwindet und somit*

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x) dx = \int_{a}^{b} \frac{x}{b-a} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

der Erwartungswert genau der Intervallmitte entspricht

Die zentrale Lage einer Zufallsvariablen wird durch den Erwartungswert beschrieben und kann analog zum Stichprobenmittel $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ der Stichprobe $(x_1, x_2, \ldots, x_n)$ interpretiert werden, da nach dem Gesetz der großen Zahlen für große Stichprobenumfänge $(n \to \infty)$ eine gewisse Konvergenz besteh, siehe Gleichung (34) in Abschnitt 6.

Bemerkung 2.3: Zusammenhang Erwartungswert und arithmetisches Mittel

Um den Zusammenhang zwischen dem Erwartungswert einer diskreten Zufallsgröße mit Werten $x_1, x_2, \ldots$

$$\mathbb{E}X = \sum_k x_k P(X = x_k) = \sum_k x_k p_k$$

und dem arithmetischen Mittel einer Stichprobe mit den Ausprägungen $x_1, x_2, \ldots$ *zu sehen, müssen wir berücksichtigen wie oft jede Ausprägung in der Stichprobe vorkommt. Dies entspricht gerade der absoluten Häufigkeit* h_k, *und da beim arithmetischen Mittel jeder Stichprobenwert gleich mit dem Gewicht* $\frac{1}{n}$ *gewichtet wird, werden die Ausprägungen mit der Häufigkeit* h_k *mit den relativen Häufigkeiten* $f_k = \frac{h_k}{n}$ *gewichtet, welche wiederum gegen die Wahrscheinlichkeiten konvergieren*

$$\overline{x} = \frac{1}{n} \sum_k x_k h_k = \sum_k x_k f_k = \sum_{k=1}^{n} x_k \underbrace{H_n(X = x_k)}_{\to P(X = x_k) \text{ für } n \to \infty} \to \mathbb{E}X\,,$$

so dass für große Stichproben das arithmetische Mittel gegen den Erwartungswert konvergiert. Entsprechend kann der Erwartungswert interpretiert werden: wenn das der Zufallsgröße X *zugrundeliegende zufällige Experiment sehr oft wiederholt wird, ergibt sich im Mittel der Erwartungswert.*

Beispiel 2.8: Illustration: Augensumme 2er Würfel mit $\mathbb{E}X = 7$

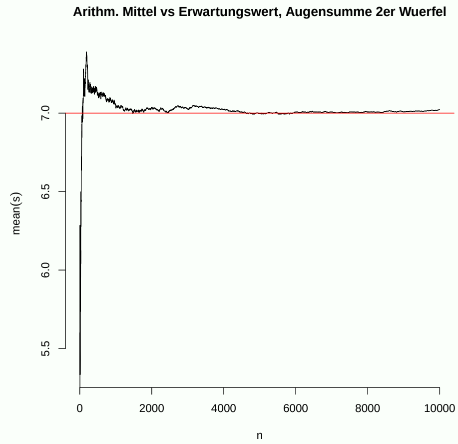
- *1. Stichprobe* $n = 10$*:* $(5,7,4,3,6,6,7,9,8,8) \Rightarrow \overline{x} = 6.3$
  10 Würfe mit 2 Würfeln:$(1,4);(1,6);(3,1);(2,1);(5,1);(2,4);(3,4);(5,4);(6,2);(4,4)$

k	1	2	3	4	5	6	7
$a_k = x_k$	3	4	5	6	7	8	9
h_k	1	1	1	2	2	2	1
$f_k = H_n(X = x_k)$	0.1	0.1	0.1	0.2	0.2	0.2	0.1

- *2. Stichprobe* $n = 500 \Rightarrow \overline{x} = 7.034$

k	1	2	3	4	5	...
x_k	2	3	4	5	6	...
f_k	0.032	0.048	0.096	0.076	0.152	...
p_k	0.028	0.056	0.083	0.111	0.139	...

Simulation mit R: für ingesamt 10 000 *Würfe wurden aus den jeweils ersten* $n = 1,\ldots,10000$ *Ergebnissen das Stichprobenmittel berechnet und auf der* y*-Achse abgetragen*



Bemerkung 2.4: Rechenregeln für den Erwartungswert

Für den Erwartungswert der Zufallsvariablen X, Y, $X+Y$ *sowie* $Z = g(X)$ *und jede Konstant* $c \in \mathbb{R}$ *gelten*

1. *Erwartungswert von Konstanten ist konstant:* $\mathbb{E}c = c$
2. *Herausziehen von Konstanten Faktoren:* $\mathbb{E}(cX) = c\mathbb{E}X$
3. *Erwartungswert der Summe zweier Zufallsgrößen entspricht der Summe der Erwartungswerte:* $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$
4. *Erwartungswert von Funktionen von Zufallsgrößen*

$$\mathbb{E}Z = \mathbb{E}(g(X)) = \begin{cases} \sum\limits_i g(x_i)\cdot P(X = x_i) & (X \text{ diskrete ZG}) \\ \int\limits_{-\infty}^{\infty} g(x)f(x)\,dx & (X \text{ stetige ZG}) \end{cases} \tag{11}$$

2.1.2 Varianz & Standardabweichung

Definition 2.3: Varianz/ Standardabweichung

Die Schwankung einer Zufallsvariablen um deren Erwartungswert wird mit der Varianz $\mathrm{Var}(X)$ *gemessen und enspricht der erwarteten quadratischen Abweichung von* X *zum Erwartungswert* $\mathbb{E}X$*. Die Standardabweichung ist* $\sigma = \sqrt{\mathrm{Var}(X)}$*.*

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\Big(X - \mathbb{E}X\Big)^2 \tag{12}$$

Beispiel 2.9: Varianz beim Würfeln

- *einfacher Würfel: (bekannt $\mathbb{E}X = 3.5$) Für die Varianz erhalten wir*

$$\operatorname{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \frac{1}{6}\sum_{i=1}^{6}(i - 3.5)^2 = 2.9167$$

  $\Rightarrow$ Standardabweichung $\sigma = 1.7078$
- *Augensumme 2er Würfel (Beispiele 2.3 & 2.6): (bekannt: $\mathbb{E}X = 7$) Für die Varianz erhalten wir mit der Wahrscheinlichkeitfunktion aus Tabelle 1 von*

$$\operatorname{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \sum_{i=1}^{11}\left((i+1) - 7\right)^2 p_i = \ldots = 5.8333$$

  $\Rightarrow$ Standardabweichung $\sigma = 2.4152$

Bemerkung 2.5: Varianzberechnung und Interpretation

1. *Die Varianz ist der Erwartungswert der quadratischen Abweichung und somit folgt mit Formel (11) und $\mu = \mathbb{E}X$ für $g(x) = (x - \mu)^2$ gerade:*

$$\operatorname{Var}(X) = \mathbb{E}g(X) = \begin{cases} \sum\limits_i (x_i - \mathbb{E}X)^2 P(X = x_i) & (X \text{ diskrete ZG}) \\ \int\limits_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f(x)\, dx & (X \text{ stetige ZG}) \end{cases}$$

2. *Die Varianz misst die Streung der Zufallsvariablen um ihren Erwartungswert, während die empirische Varianz einer Stichprobe*

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$$

   die Streuung um das Stichprobenmitel misst. Wenn man eine Stichprobe als (unabhängige) Realisierungen der Zufallsvariablen auffasst, kann man also die Interpretationen aus der deskriptiven Statistik einfach auf Zufallsvariablen übertragen. Ähnlicher Zusammenhang gilt für die Standardabweichung $\sigma = \sqrt{\operatorname{Var}(X)}$ und die empirische Standardabweichung $s = \sqrt{s^2}$.
3. *häufig lässt sich die Varianz einfacher durch das 2. Moment $\mathbb{E}X^2$ berechnen, denn*

$$\operatorname{Var}(X) = \mathbb{E}\left(X - \mathbb{E}X\right)^2 = \mathbb{E}X^2 - \left(\mathbb{E}X\right)^2$$

4. *für jede beliebige Konstante $c \in \mathbb{R}$ gilt*
   (a) *$\operatorname{Var}(c) = 0$ (Konstanten schwanken nicht)*
   (b) *$\operatorname{Var}(cX) = c^2 \operatorname{Var}(X)$ (beim Herausziehen von Konstanten aus der Varianz müssen diese quadriert werden)*
   (c) *$\operatorname{Var}(c + X) = \operatorname{Var}(X)$ (konstante Verschiebungen ändern zwar den Erwartungswert um diese Konstante, aber nicht die Schwankung)*
   (d) ***Achtung** bei der Berechnung der Varianz der Summe von Zufallsvariablen:*

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X, Y)$$

   auch wenn wir die Kovarianz $\operatorname{cov}(X, Y)$ zweier Zufallsvariablen hier nicht näher betrachten,[a] zeigt die einfache Nebenrechnung:

$$\begin{aligned} \operatorname{Var}(X+Y) &= \mathbb{E}\left(X + Y - \mathbb{E}\left(X+Y\right)\right)^2 = \mathbb{E}\left((X - \mathbb{E}X) + (Y - \mathbb{E}Y)\right)^2 \\ &= \underbrace{\mathbb{E}\left(X - \mathbb{E}X\right)^2}_{=\operatorname{Var}(X)} + 2\underbrace{\mathbb{E}\left((X - \mathbb{E}X)\left(Y - \mathbb{E}Y\right)\right)}_{=\operatorname{cov}(X,Y)} + \underbrace{\mathbb{E}\left(Y - \mathbb{E}Y\right)^2}_{=\operatorname{Var}(Y)} \\ &= \operatorname{Var}(X) + 2\operatorname{cov}(X,Y) + \operatorname{Var}(Y) \end{aligned}$$

   dass die Schwankung von $X + Y$ von dem Zusammenhang zwischen beiden Variablen abhängig ist. Sind beide positiv von einander abhängig ($\operatorname{cov}(X, Y) > 0$), so ziehen große X Werte auch große Y Werte nach sich und folglich ist die Varianz der Summe größer. Sind beide hingegen negativ abhängig ($\operatorname{cov}(X, Y) > 0$), so gleichen sich große Werte der einen Variablen mit kleinen der anderen aus und die Varianz der Summe ist enstprechend kleiner.

(e) *im Falle der* ***Unabhängigkeit*** *von X und Y gilt*

$$\operatorname{cov}(X,Y) = 0 \Rightarrow \operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$$

(im Grunde reicht schon die Unkorreliertheit[b] *von X und Y, also $\operatorname{cov}(X,Y) = 0$)*

[a] Die Kovarianz zweier Zufallsvariablen ist definiert als $\operatorname{cov}(X,Y) = \mathbb{E}\left((X-\mathbb{E}X)(Y-\mathbb{E}Y)\right) = \mathbb{E}\left(XY\right) - \mathbb{E}X \cdot \mathbb{E}Y$ und damit analog zur empirischen Kovarianz zweier Stichproben $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ definiert und misst den linearen Zusammenhang. Im Falle unabhängiger Zufallsvariablen gilt $\mathbb{E}\left(XY\right) = \mathbb{E}X \cdot \mathbb{E}Y$ und folglich $\operatorname{cov}(X,Y) = 0$.

[b] Aus Unabhängigkeit folgt stets die Unkorreliertheit, aber umgekehrt darf aus der Unkorreliertheit nicht auf die Unabhängigkeit geschlussfolgert werden.

Beim Umgang mit der Normalverteilung und auch in der Statistik benötigen wir das Konzept normierter Zufallsvariablen.

Definition 2.4: Normierte Zufallsvariablen

Eine Zufallsvariable X heißt normiert (standardisiert), falls $\mathbb{E}X = 0$ und $\operatorname{Var}(X) = 1$ gelten.

Bemerkung 2.6: Z-Transformation / Normierung / Standardisierung

Jede Zufallsgröße X mit endlichem Erwartungswert und positiver Varianz ($\operatorname{Var}(X) \neq 0$) kann mittels

$$Z = \frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var}(X)}}$$

standardisiert bzw. normiert werden. Diese lineare Transformation

$$Z = \underbrace{\frac{1}{\sqrt{\operatorname{Var}(X)}}}_{=a} \cdot X - \underbrace{\frac{\mathbb{E}X}{\sqrt{\operatorname{Var}(X)}}}_{=b} = aX - b$$

mit den Konstanten a und b verändert nicht die Charakteristik der Zufallsverteilung, sondern zentriert die Werte wegen

$$\mathbb{E}Z = a\mathbb{E}X - b = \frac{\sqrt{\operatorname{Var}(X)}}{\cdot}\mathbb{E}X - \frac{\mathbb{E}X}{\sqrt{\operatorname{Var}(X)}} = 0$$

um Null und standardisiert die Schwankung

$$\operatorname{Var} Z = \operatorname{Var}(aX - b) = a^2 \operatorname{Var}(X) = \frac{\operatorname{Var}(X)}{\operatorname{Var}(X)} = 1$$

auf Eins. Teilweise wird diese Normierung auch Z-Transformation genannt und findet sich auch in der Statistik wieder, um eine bessere Vergleichbarkeit zwischen Stichproben zu gewährleisten.

2.1.3 Höhere Momente der Zufallsvariablen und Quantile

Bei der empirischen Auswertung von Stichproben sind die empirischen Quantile sehr hilfreich (siehe Anhang B.3, DefinitionB.6 auf Seite 68) . Analog führt man Quantile für Zufallsgrößen ein.

Definition 2.5: Quantil der Ordnung p

Als Quantil der Ordnung $p \in (0,1)$ der Zufallsvariablen X mit Verteilungsfunktion F bezeichnet man die Zahl Q_p mit

$$Q_p = F^{-}(p) := \min\left\{x \in \mathbb{R} : \ F(x) \geq p\right\} .$$

Bemerkung 2.7

- *Quantil der Ordnung* $p = \frac{1}{2}$ *entspricht dem Median*
- F^- $\triangleq$ *verallgemeinerte Inverse mit* $F^-(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\} = \min\{t \in \mathbb{R} : F(t) \geq p\}$ *wegen der rechtsseitigen Stetigkeit der Verteilungfunktion* F
- *für stetige Zufallsgrößen* X *mit streng monoton wachsender Verteilungsfunktion gilt* $F^- = F^{-1}$*, d.h. die Inverse der Verteilungsfunktion existiert und stimmt mit der verallgemeinerten Inversen überein, so dass dann sogar* $F(Q_p) = p$ *gilt.*
- *Interpretation analog zu empirischen Quantilen einer Stichprobe, siehe Bemerkung B.7*

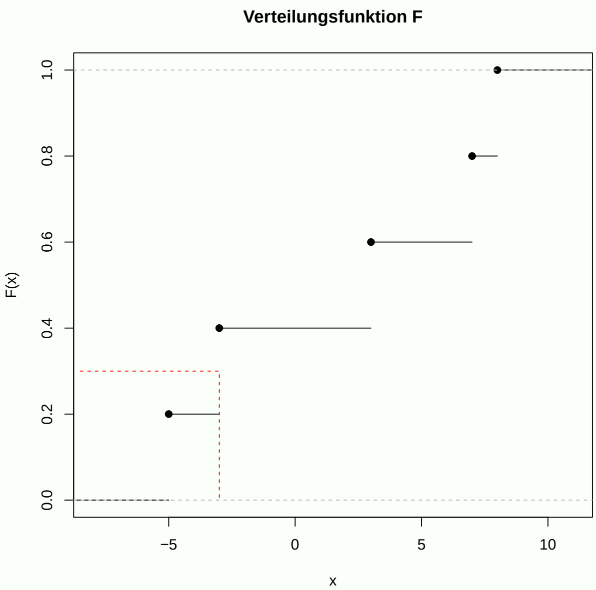
Beispiel 2.10: Quantile einer diskreten Zufallsgröße

Wir betrachten die die diskrete Zufallsgröße, die die folgenden fünf Werte

x	-5	-3	3	7	8
$P(X = x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

mit gleicher Wahrscheinlichkeit annimmt. Die folgende Grafik veranschaulicht den Zusammenhang zwischen Verteilungsfunktion und Quantilen am Beispiel des 30% *Quantils*

$$F^-(0.3) = \min\{x \in \mathbb{R} :\ F(x) \geq 0.3\} = \min\{x \in \mathbb{R} :\ x \in [3, \infty)\} = -3\,.$$



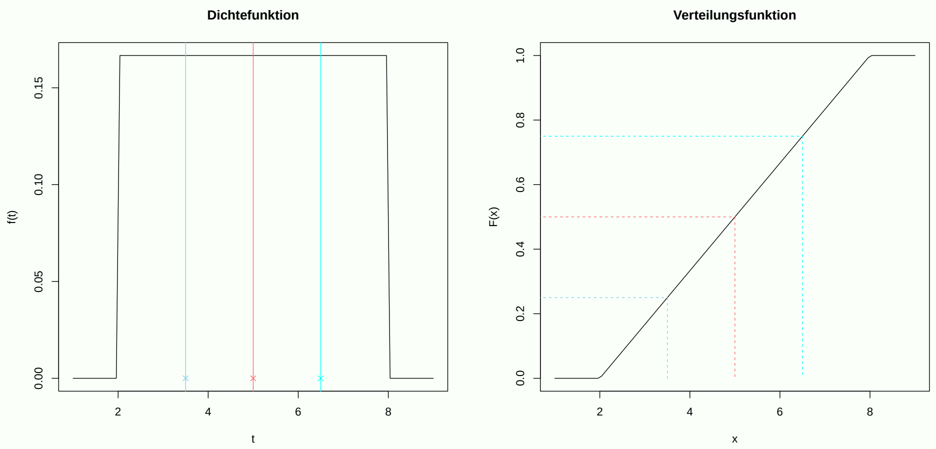
Beispiel 2.11: Quantile der stetigen Gleichverteilung $X \sim \mathrm{U}(a, b)$**, siehe Beispiel 2.5**

Die Verteilungsfunktion der stetigen Gleichverteilung lautet nach Formel (8) *für alle* $x \in [a, b]$ *gerade* $F(x) = \frac{x-a}{b-a}$*, so dass deren Umkehrfunktion* $F^{-1}(p) = p(b-a) + a$ *lautet und mit der verallgemeinerten Inversen*

$$\begin{aligned} F^-(p) &= \min\{x \in \mathbb{R} :\ F(x) \geq p\} = \min\left\{x \in \mathbb{R} :\ \frac{x-a}{b-a} \geq p\right\} \\ &= \min\{x \in \mathbb{R} :\ x \geq p(b-a) + a\} = p(b-a) + a = F^{-1}(p) \end{aligned}$$

übereinstimmt. Für $p = 0.5$ *erhalten wir den Median* $x_{med} = F^{-1}(0.5) = \frac{a+b}{2} = \mathbb{E}X$*, der genau der Intervallmitte entspricht und wegen der Symmetrie der Dichte mit dem Erwartungswert übereinstimmt.*
Die folgende Graphik zeigt die Quartile von $X \sim \mathrm{U}(2, 8)$*, d.h. die* 25%*,* 50% *sowie das* 75% *Quantile. In der linken Graphik sieht man bei der Gleichverteilung sehr schön, dass beispielsweise das* 50% *Quantil, also der Median, die Fläche unter der Dichte genau halbiert. Die zu den entsprechenden senkrechten Linien teilen die Fläche unter der Dichte in vier*

gleich große Bereiche ein.



Mit Hilfe der allgemeinen **höheren Momente** ($k \in \mathbb{N}$)

$$
\begin{aligned}
\mathbb{E}X^k &= \int_\Omega X(\omega)^k P(d\omega) = \int \mathbb{R} x dF(x) \\
&= \begin{cases} \sum_i x_i^k P(X = x_i)\,, & \text{falls } X \text{ diskrete ZV mit Werten } x_1, x_2, \ldots \\ \int\limits_{-\infty}^{\infty} x^k\, f(x)\, dx\,, & \text{falls } X \text{ absolutstetige ZV mit Dichte } f. \end{cases}
\end{aligned}
$$

kann man weitere Kennzahlen von Verteilungen wie Schiefe oder Krümmung angeben, analog zu den empirischen Maßen von Stichproben in Abschnitt B.3.3.

Definition 2.6: Höhere Momente

Es sei X eine Zufallsgröße. Dann heißen für $k = 1, 2, \ldots$ die Größen

$$
m_k = \mathbb{E}X^k \quad \textit{und} \quad \mu_k = \mathbb{E}\big(X - \mathbb{E}X\big)^k
$$

k-te Momente bzw. k-te zentrale Momente.

Speziell für $k = 3$ erhalten wir die Schiefe, welche die Symmetrie bzw. Asymmetrie einer Verteilung misst

$$
\gamma_1 := \frac{\mu_3}{\sigma^3} = \frac{\mathbb{E}\,(X - \mathbb{E}X)^3}{\sqrt{\mathrm{Var}(X)}^3} = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}\,X}}\right)^3
$$

und für $k = 4$ ein Maß für die Wölbung bzw. Krümmung

$$
\beta_2 := \frac{\mu_4}{\sigma^4} = \frac{\mathbb{E}\,(X - \mathbb{E}X)^4}{\mathrm{Var}(X)^2} = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}\,X}}\right)^4 \qquad \text{(Kurtosis)}
$$

bzw. $\gamma_2 := \beta_2 - 3$ (Exzess)

2.2 Verteilungsaussagen mittels Erwartungswert und Varianz

Das Zufallsverhalten einer Zufallsvariablen X ist vollständig durch die Verteilungsfunktion F charakterisiert. In vielen praktischen Anwendungen ist diese allerdings nicht bekannt, sondern häufig sind nur durch statistische Auswertungen Informationen über die ersten Momente bekannt. Mit Hilfe der TSCHEBYSCHEFF'schen Ungleichung kann zumindest die Wahrscheinlichkeit für symmetrische Abweichungen um den Erwartungswert $P(|X - \mathbb{E}X| \leq \varepsilon)$ mittels der Varianz abgeschätz werden, falls die Verteilungsfunktion F unbekannt[5] ist.

[5] Für F bekannt rechnet man einfach

$$
\begin{aligned}
P(|X - \mathbb{E}X| \leq \varepsilon) &= P(-\varepsilon \leq X - \mathbb{E}X \leq \varepsilon) = P(\mathbb{E}X - \varepsilon \leq X \leq \mathbb{E}X + \varepsilon) \\
&= P(\mathbb{E}X - \varepsilon < X \leq \mathbb{E}X + \varepsilon) + P(X = \mathbb{E}X - \varepsilon) \\
&= F(\mathbb{E}X + \varepsilon) - F(\mathbb{E}X - \varepsilon) + P(X = \mathbb{E}X - \varepsilon)\,,
\end{aligned}
$$

wobei für absolutstetige Zufallsvariablen die letzte Wahrscheinlichkeit $P(X = \mathbb{E}X - \varepsilon) = 0$ verschwindet und sich im allgemeinen Fall auch als Zuwachs der Verteilungsfunktion darstellen lässt.

Theorem 2.1: Tschebyscheff'sche Ungleichung

Falls $\mathbb{E}|X| < \infty$ *und* $\mathrm{Var}(X) > 0$ *gilt für jedes* $\varepsilon > 0$

$$P(|X - \mathbb{E}X| > \varepsilon) < \frac{\mathrm{Var}(X)}{\varepsilon^2} \tag{13}$$

und folglich

$$P(|X - \mathbb{E}X| \le \varepsilon) = 1 - P(|X - \mathbb{E}X| > \varepsilon) > 1 - \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

1. Begründung für Theorem 2.1

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X - \mathbb{E}X)^2 \\ &= \mathbb{E}\left(\underbrace{\mathbb{1}_{|X-\mathbb{E}X|\le\varepsilon}(X - \mathbb{E}X)^2}_{\ge 0} + \underbrace{\mathbb{1}_{|X-\mathbb{E}X|>\varepsilon}(X - \mathbb{E}X)^2}_{>\varepsilon^2}\right) \\ &> \varepsilon^2 \mathbb{E}\mathbb{1}_{|X-\mathbb{E}X|>\varepsilon} = \varepsilon^2 P(|X - \mathbb{E}X| > \varepsilon)\end{aligned}$$

2. Zusammenhang zu Markov-Ungleichung: $h : D \subset \mathbb{R} \to \mathbb{R}_+$ monoton wachsende Funktion, wobei $X(\omega) \subset D$: dann gilt $P(X \ge a) \le \frac{\mathbb{E}(h(X))}{h(a)}$
3. Beweis Markov-Ungleichung: 1. für alle $x \in [a, \infty)$ gilt wegen Monotonie $h(a) \le h(x)$, 2. $h(x) \ge 0$

$$\begin{aligned}h(a)P(X \ge a) = h(a)\int_a^\infty dF(x) &= \int_a^\infty h(a)dF(x) \le \int_a^\infty h(x)dF(x) \\ &\le \int_{\mathbb{R}} h(x)dF(x) = \mathbb{E}h(X)\end{aligned}$$

4. speziell für positive ZV $X \ge 0$ und $h(x) = x$ folgt auf $D = \mathbb{R}_+$: $P(X \ge a) \le \frac{\mathbb{E}X}{a}$
5. $|X| \ge 0$ und $h(x) = x$ folgt auf $D = \mathbb{R}_+$: $P(|X| \ge a) \le \frac{\mathbb{E}|X|}{a}$
6. Tschebyscheff für $Z = |X - \mathbb{E}X|$ und $h(x) = x^2$ auf $D = \mathbb{R}_+$: $P(|X - \mathbb{E}X| \ge a) \le \frac{\mathbb{E}|X-\mathbb{E}X|^2}{a^2} = \frac{\mathbb{E}(X-\mathbb{E}X)^2}{a^2} = \frac{\mathrm{Var}\,X}{a^2}$

Beispiel 2.12: Abschätzung mittels Tschebyscheff Ungleichung kann nicht verbessert werden

Die diskrete Zufallsvariable X*, welche die drei Werte* $-a$*,* 0 *und* $a > 0$ *mit der Wahrscheinlichkeiten*

$$P(X = k) = \begin{cases} \frac{p}{2}, & k = -a \\ 1 - p, & k = 0 \\ \frac{p}{2}, & k = a \\ 0, & \text{sonst} \end{cases}$$

für $p \in (0, 1)$ *annehmen kann, hat* $\mathbb{E}X = -a\frac{p}{2} + 0(1 - p) + a\frac{p}{2} = 0$ *und Varianz* $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = (-a)^2\frac{p}{2} + 0^2(1 - p) + a^2\frac{p}{2} = a^2p$ *und folglich liefert die Tschebyscheff Ungleichung*

$$P(|X - \mathbb{E}X| \ge \varepsilon) = P(|X| \ge \varepsilon) \le \frac{\mathrm{Var}(X)}{\varepsilon^2} = \frac{a^2p}{\varepsilon^2} \tag{14}$$

Für $\varepsilon = a > 0$ *stimmte diese obere Grenze von (14) mit* $\frac{a^2p}{a^2} = p$ *aber genau mit der Wahrscheinlichkeit* $P(|X| \ge a) = P(|X| = a) = P(X = a) + P(X = -a) = \frac{p}{2} + \frac{p}{2} = p$ *überein, d.h. die Abschätzung ist in diesem Beispiel sogar exakt und kann demnach nicht verbessert werden. Für viele Zufallsvariablen ist die Abschätzung aber dennoch grob, wie Beispiel 2.13 illustriert.*

Bemerkung 2.8: 3σ-Regel

Mit $\varepsilon := k\sqrt{\mathrm{Var}(X)} = k\sigma$ erhalten wir aus der TSCHEBYSCHEFF*-Ungleichung* (13)

$$P(|X - \mathbb{E}X| > k\sigma) < \frac{1}{k^2} \quad \text{bzw.} \quad P(|X - \mathbb{E}X| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

und somit speziell für $k = 3$ die sogenannte 3σ-Regel

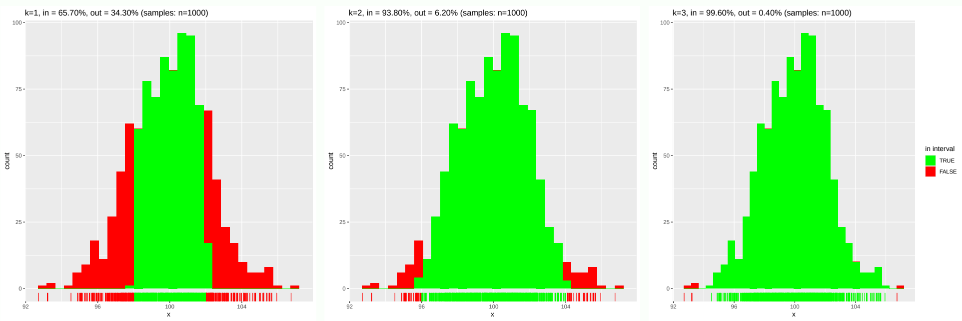
$$P(|X - \mathbb{E}X| > 3\sigma) < 0.11 \quad \text{bzw.} \quad P(|X - \mathbb{E}X| \leq 3\sigma) \geq 0.89\,,$$

d.h. mit mindestens 89% Wahrscheinlichkeit weicht jede beliebige Zufallsvariable um höchstens 3 Standardabweichungen vom Erwartungswert ab. Bemerkenswert ist, dass für diese Aussage keinerlei zusätzliche Information über die Verteilung bekannt sein muss und diese Aussage für jede beliebige Zufallsvariable (mit endlichem Erwartungswert) gilt. Allerdings ist diese Abschätzung oftmals recht grob, z.B. gilt für die Normalverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$ sogar $P(|X - \mathbb{E}X| \leq 3\sigma) = 0.9973$.

Beispiel 2.13: Illustration der $k\sigma$-Regel für Normalverteilung

In Abschnitt 4 (siehe Beispiel 4.2 zur Normalverteilung) werden wir sehen, wie wir die Wahrscheinlichkeiten für die Normalverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$ berechnen können, aber eine kleine Simulation von 1000 Realisierungen der Normalverteilung mit Erwartungswert $\mu = \mathbb{E}X = 100$ und Varianz $\sigma^2 = \mathrm{Var}(X) = 4$ zeigt die Gültigkeit der $k\sigma$-Regel für $k = 1, 2, 3$, d.h. innerhalb des symmetrischen Intervalls $(\mu - k\sigma, \mu + k\sigma)$ um den Erwartungswert μ mit einer Abweichung von $k\sigma$ nach links und rechts liegt mindestens ein Anteil von $1 - \frac{1}{k^2}$ aller Realsierungen (grün) und demnach außerhalb dieses Intervalls höchstens $\frac{1}{k^2}$ der Realisierungen (rot).

k	Mindestanteil nach Tschebyscheff	exakte Wkt für NV (in Prozent)
1	0.00	68.27
2	75.00	95.45
3	88.89	99.73
4	93.75	99.99



Im Folgenden sollen die Binomial- und die Normalverteilung näher besprochen werden.

3 Binomialverteilung

Definition 3.1: Binomialverteilung

Eine diskrete Zufallsvariable X, die die Werte $0, 1, \ldots, n$ mit der Wahrscheinlichkeit

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n \tag{15}$$

*annimmt, heißt **binomialverteilt** mit Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$. Abkürzend schreibt man $X \sim \mathrm{Bin}(n, p)$.*

Die Binomialverteilung hat vielfältige Anwendungen und lässt sich mit dem Galton-Brett beschreiben.

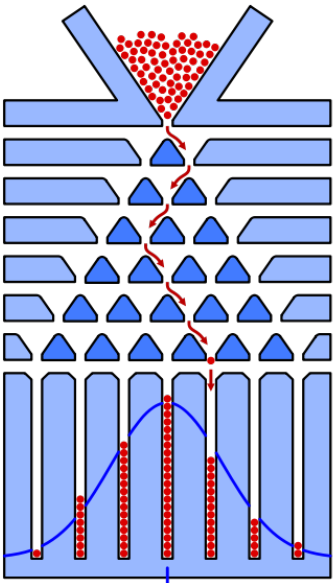


Abbildung 3: Galtonbrett

Dieses besteht aus n Ebenen wie in Abbildung 3 abgebildet. Von oben werden Kugeln eingeleitet, die sich in jeder Ebene jeweils unabhängig davon, was in der Ebene zuvor geschah, mit der Wahrscheinlichkeit p nach rechts und entsprechend mit der Wahrscheinlichkeit $1-p$ nach links bewegen. Diese zufällige Bewegung einer Kugel auf der i-ten Ebene kann durch die unabhängigen Bernoulli verteilten Zufallsgrößen modelliert werden.

Definition 3.2: Bernoulli-Verteilung

Die 0-1-verteilte Zufallsvariable mit

$$X_i = \begin{cases} 1, & \textit{Ereignis } A \textit{ tritt im } i\textit{-ten Versuch ein} \\ 0, & \textit{Ereignis } A \textit{ tritt im } i\textit{-ten Versuch nicht ein} \end{cases} \tag{16}$$

mit $P(X_i = 1) = P(A) = p$ *heißt Bernoulli verteilt.*

Hier beschreibt das Ereignis A die Bewegung der Kugel auf dem Galtonbrett nach rechts und die Bernoulli-Verteilung ist ein Spezialfall der Binomialverteilung für die Parameterwahl $n = 1$. Dann entspricht die zufällige Summe

$$X = \sum_{i=1}^{n} X_i \tag{17}$$

genau der Anzahl der „Bewegungen nach rechts“. Landet eine Kugel also nach Durchlauf des Galtonbretts ganz links, nimmt X den Wert Null an. Landet die Kugel hingegen ganz rechts, gilt $X = n$. Für $X = k$ gibt es $\binom{n}{k}$ verschiedene Pfade, die die Kugel durch das Galtonbrett nehmen kann. Da die Kugel sich in der i-ten Ebene unabhängig von der Bewegung in der vorherigen Ebene $i-1$ jeweils mit der Wahrscheinlichkeit p nach rechts und mit Wahrscheinlichkeit $1-p$ nach links bewegt, folgt somit

$$\begin{aligned} P(X = k) &= \binom{n}{k} P(\{X_1 = 1\} \cap \ldots \cap \{X_k = 1\} \cap \{X_{k+1} = 0\} \cap \ldots \cap \{X_n = 0\}) \\ &= \binom{n}{k} P(X_1 = 1) \cdot \ldots \cdot P(X_{k+1} = 0) \cdot \ldots \cdot P(X_n = 0) = \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

da für $X = k$ die Kugel sich genau k mal nach rechts und $n-k$ mal nach links bewegt.[6] Eine sehr schöne Animation in Geogebra ist auf `https://www.geogebra.org/m/eayfksqh` zu finden.

Das Prinzip des Galtonbretts lässt sich verallgemeinern. Jede Zufallsgröße X, die das Eintreten eines Ereignisses A mit $P(A) = p$ bei n-facher unabhängiger Wiederholung zählt, ist binomialverteilt. Darüber hinaus kann man sich die Verteilung auch mittels eines Urnenmodells veranschaulichen.

[6] In obiger Formel wurde unterstellt, dass die Kugel sich in der ersten k Ebenen nach rechts und den restlichen $n-k$ Ebenen nach links bewegt hat. Natürlich spielt die Anordnung der Bewegungen nach rechts und links keine Rolle, da die Wahrscheinlichkeit für k Rechtsbewegungen und $n-k$ Linksbewegungen immer $p^k(1-p)^{n-k}$ beträgt.

Bemerkung 3.1: Binomialverteilung im Urnenmodell

Aus einer Urne mit einem Anteil von $p \in [0, 1]$ schwarzen Kugeln, wird n mal mit Zurücklegen gezogen. Dann ist die Zufallsvariable X, die zählt wie oft dabei eine schwarze Kugel gezogen wird, binomialverteilt.

Tatsächlich ist die Zuordnung $p_k = P(X = k)$ für $k = 0, 1, \ldots, n$ eine Wahrscheinlichkeitsfunktion, denn es gilt $0 \leq p_k \leq 1$ sowie

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Beispiel 3.1: 20 maliger Würfelwurf

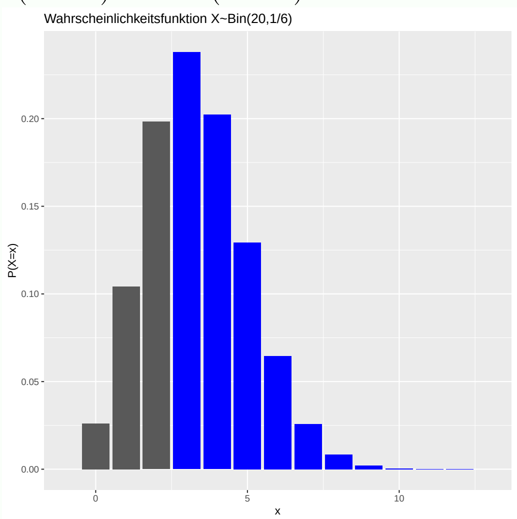
Ein (idealer) Würfel wird $n = 20$ mal geworfen. Wie hoch ist die Wahrscheinlichkeit, dass mindestens dreimal eine 6 gewürfelt wird? Zur Lösung betrachten wir die Zufallsvariable X, die die Anzahl der geworfenen Sechsen beim 20 maligen Würfelwurf zählt. Dann ist $X \sim \mathrm{Bin}(n = 20, p = 1/6)$ und es folgt

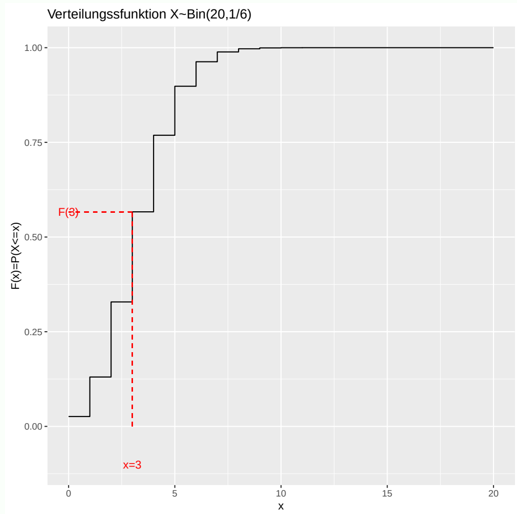
$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) = 1 - F(2) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) \\ &= 1 - \left(\binom{20}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{20} + \binom{20}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{19} + \binom{20}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{18} \right) \\ &= 0.6713\,, \end{aligned}$$

d.h. bei 20 Würfen fällt mit 67%−iger Wahrscheinlichkeit mindestens dreimal eine 6. Die Berechnung der gesuchten Wahrscheinlichkeit

$$P(X \geq 3) = \sum_{k=3}^{20} P(X = k)$$

als Summe über alle Wahrscheinlichkeiten der Wahrscheinlichkeitsfunktion, für die X den Wert $3, 4, \ldots, 20$ (blau eingefärbt) annimmt, ist in der folgenden linken Grafik dargestellt. Die rechte Grafik zeigt hingegen die Berechnung $P(X \geq 3) = 1 - F(2)$ mit Hilfe des Gegenereignisses $\overline{\{X \geq 3\}} = \{X < 3\} = \{X \leq 2\}$ sowie der Verteilungsfunktion $F(2) = P(X \leq 2)$. Die Wahrscheinlichkeit $P(X \leq 2) = F(2) = P(X = 0) + P(X = 1) + P(X = 2)$ entspricht natürlich der Summe der grau eingefärbten Einzelwahrscheinlichkeiten $P(X = 0)$, $P(X = 1)$ sowie $P(X = 2)$ in der linken Grafik.



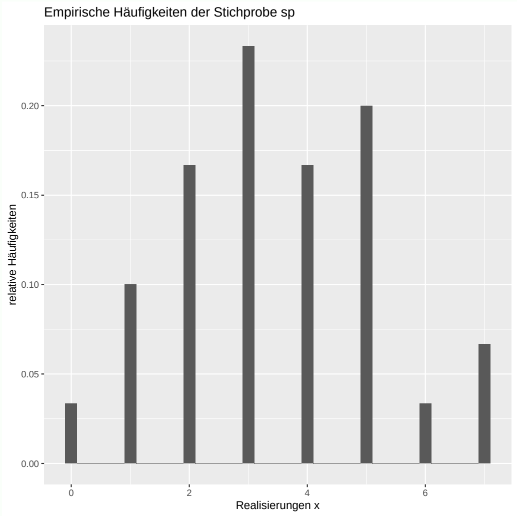


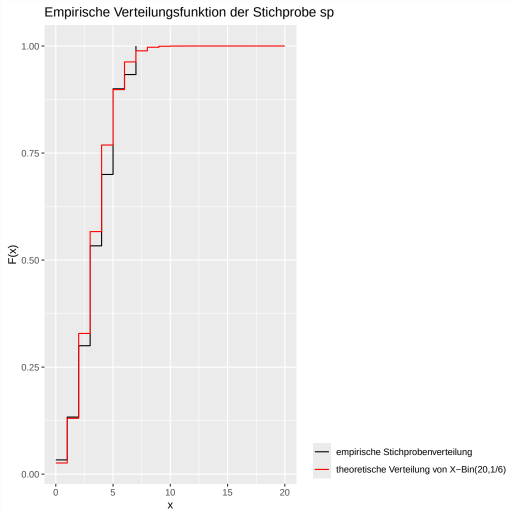
An dieser Stelle wollen wir nochmals betonen, dass die Verteilungsfunktion den kompletten Zufallscharakter beschreibt. Wenn wir jetzt 20 mal nacheinander mit einem fairen Würfel würfeln, dann beträgt die Wahrscheinlichkeit mindestens drei Sechsen zu beobachten gerade $P(X \geq 3) = 0.6713$. Das erlaubt natürlich keine Voraussage, dass bei unserem Experiment tatsächlich dieses Ereignis eintritt, sondern nur mit welcher Wahrscheinlichkeit. Würden wir das Experiment allerdings sehr oft wiederholen, so würde in etwas 67.1341 Prozent aller Experimente tatsächlich das Ereignis $\{X \geq 3\}$ auftreten.
Zum Abschluss dieses Beispiels wollen wir dies demonstrieren, indem wir das Würfelexperiment zunächst 30 mal durchführen bzw. simulieren. Dabei erhalten wir die folgende

Stichprobe

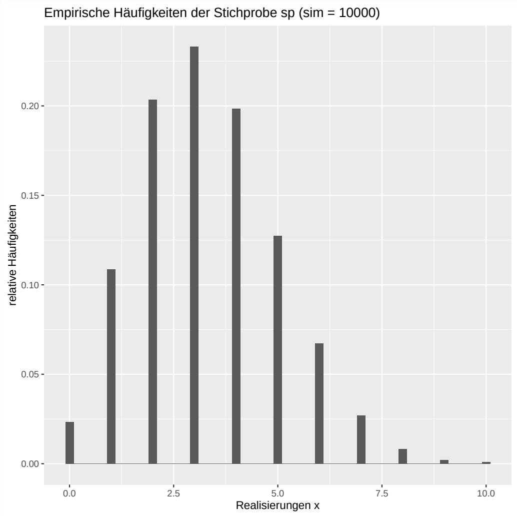
$$sp = (0,3,2,3,5,3,6,2,2,7,5,1,4,1,7,2,3,4,3,4,2,4,5,3,3,5,1,5,4,5)\,,$$

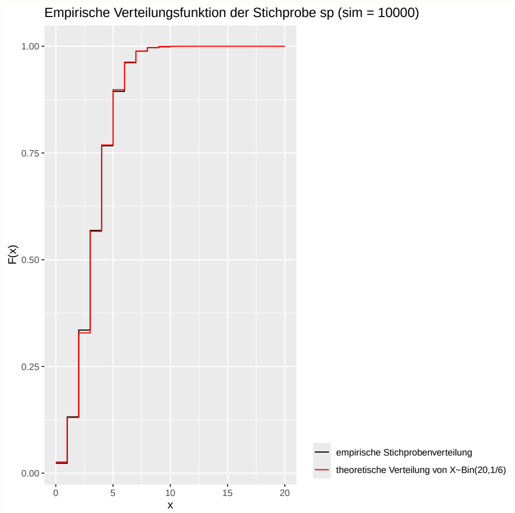
d.h. beim ersten Versuch haben wir genau 0 *und beim zweiten genau* 3 *Sechsen beim* 20 *maligen Würfeln beobachtet.*





Obige Grafiken zeigen die relativen und kumulierten Häufigkeiten dieser Simulation und stellen die empirische Wahrscheinlichkeits- sowie die empirische Verteilungsfunktion der Stichprobe dar. So haben wir in unserer Stichprobe 21 *mal das Ereignis* $\{X \geq 3\}$ *in* 30 *Durchgängen beobachtet, so dass die relative Häufigkeit* $^{21}/_{30} = 0.7$ *die tatsächliche Wahrscheinlichkeit* $P(X \geq 3) = 0.6713$ *gut approximiert.*
Mit Erhöhung der Simulationsanzahl auf 10000 *verbessert sich auch die Güte dieser Approximation auf* $^{6647}/_{10000} = 0.6647$*. Auch die Konvergenz der schwarzen empirischen Verteilungsfunktion gegen die rote theoretische Verteilungsfunktion ist dann zu beobachten, wie die zugehörigen Grafiken zeigen.*





Bemerkung 3.2: Berechnung mit Geogebra

Abbildung 4 zeigt die Berechnung von $P(X \geq 3) = 0.6713$ *mit Geogebra* `https://www.geogebra.org/classic#probability` *aus obigen Beispiel 3.1 für* $X \sim \mathrm{Bin}(n = 20, p = \frac{1}{6})$

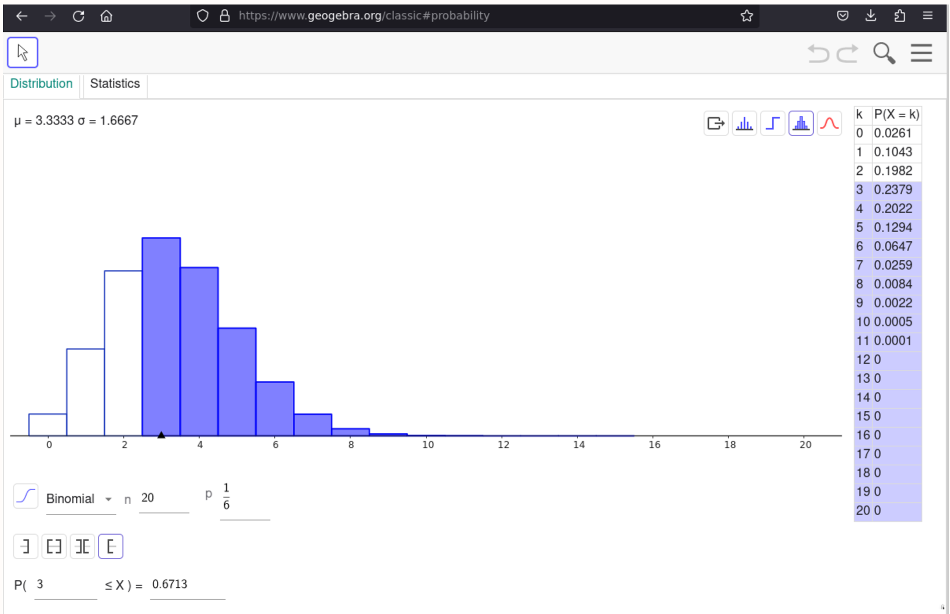


Abbildung 4: Berechnung mit Geogebra

Die Berechnung des Erwartungswertes für $X \sim \text{Bin}(n,p)$ geschieht über

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n} \frac{kn!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!}pp^{k-1}(1-p)^{n-1-(k-1)} \\
&= np\sum_{j=0}^{n-1} \frac{n!}{j!(n-1-j)!}p^j(1-p)^{n-1-j} = np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j} = np
\end{aligned}$$

und für die Berechnung der Varianz nutzen wir zunächst $\mathbb{E}X^2$ (analog zu $\mathbb{E}X$):

$$\begin{aligned}
\mathbb{E}X^2 &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} + \mathbb{E}X \\
&= n(1-n)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{n-k} + \mathbb{E}X \\
&= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{(n-2)-j} + \mathbb{E}X \\
&= n(n-1)p^2(p+(1-p))^{n-2} + np = n(n-1)p^2 + np
\end{aligned}$$

Folglich erhalten wir für die Varianz

$$\text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

Diese etwas umständliche Berechnung lässt sich allerdings deutlich abkürzen, wenn wir die Darstellung (17) von X als Summe unabhängiger Bernoulli verteilter Null-Eins-Zufallsvariablen $X = \sum_{i=1}^{n} X_i$ mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$ sowie die Additivität des Erwartungswertes nutzen.

Bemerkung 3.3: Erwartungswert und Varianz der Bernoulli-Verteilung

ür $X_i \sim \text{Bin}(1,p)$ gilt $\mathbb{E}X_i = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = p$ sowie $\mathbb{E}X_i^2 = 1^2 \cdot P(X_i = 1) + 0^2 \cdot P(X_i = 0) = p$ und folglich $\text{Var}(X_i) = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = p - p^2 = p(1-p)$.

Folglich erhatlter wir den Erwartungswert für die Summe $X = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$

$$\mathbb{E}X = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} (\mathbb{E}X_i) = \sum_{i=1}^{n} p = np \tag{18}$$

und wegen der Unabhängigkeit von $X_1, X_2, \ldots, X_n$ für die Varianz von $X \sim \text{Bin}(n, p)$

$$\text{Var}\, X = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} (\text{Var}\, X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)\,. \tag{19}$$

Theorem 3.1: Erwartungswert und Varianz von $X \sim \text{Bin}(n, p)$

Für Zufallsvariable $X \sim \text{Bin}(n, p)$ gilt

$$\mathbb{E}(X) = np$$
$$\text{Var}(X) = np(1-p)$$

Bemerkung 3.4: R-Code zur Binomialverteilung

Zur Illustration betrachten wir eine binomialverteilte Zufallsgröße $X \sim \text{Bin}(n, p)$ mit den Parametern $n = 10$ und $p = 0.3$. Mit der R-Funktion dbinom() *kann man Wahrscheinlichkeiten berechnen, z.B.*

$$P(X = 3) = \binom{n}{3} \cdot p^3 \cdot (1-p)^{n-3} = \binom{10}{3} \cdot 0.3^3 \cdot 0.7^7 = 0.2668\,.$$

Mit der R-Funktion pbinom() *kann man auch Wahrscheinlichkeiten mittels der Verteilungsfunktion*

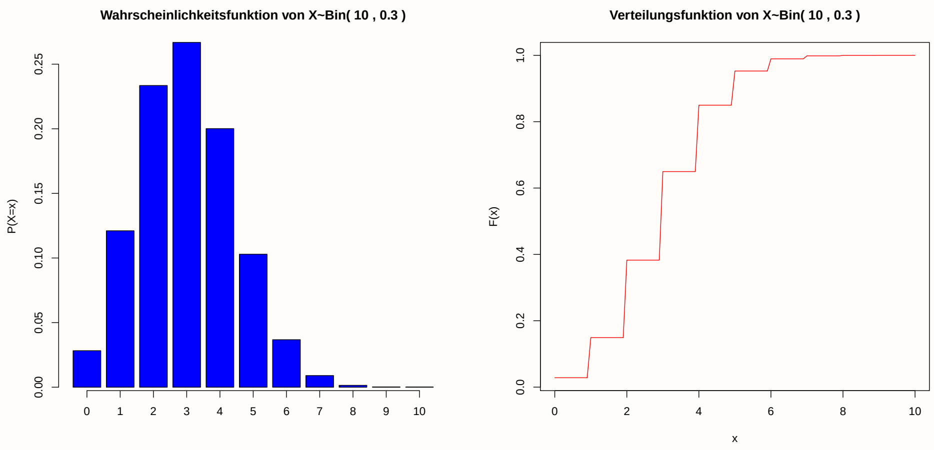
$$P(X \le 5) = F(5) = \sum_{k=0}^{5} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \sum_{k=0}^{5} \binom{10}{k} \cdot 0.3^k \cdot 0.7^{10-k} = 0.9527\,.$$

bequem berechnen, wie der folgende R-Code demonstriert.

```r
#Binomialverteilung
n = 10
p = 0.3

#berechnen Wahrscheinlichkeit, dass X=3
dbinom(x=3,size=n,prob=p)
## [1] 0.2668
#berechnen Wahrscheinlichkeit, dass X<=5
pbinom(q=5,size=n,prob=p)
## [1] 0.9527
#berechnen Wahrscheinlichkeit, dass X>5
#1. Variante: mittels Option lower.tail: P(X>5) direkt
pbinom(q=5,size=n,prob=p,lower.tail=FALSE)
## [1] 0.04735
#2. Variante: durch Gegenereignis: P(X>5)=1-P(X<=5)=1-F(5)
1-pbinom(q=5,size=n,prob=p)
## [1] 0.04735
```

Damit kann man natürlich in R sich auch die Wahrscheinlichkeits- und Verteilungsfunktion grafisch veranschaulichen.

```r
bp = barplot(dbinom(x=0:n,size=n,prob=p),col="blue",
      main=paste("Wahrscheinlichkeitsfunktion von X~Bin(",n,",",p,")"),
      ylab="P(X=x)")
      axis(side=1, at=bp, labels = 0:n)
curve(pbinom(x,n,p),from=0,to=n,col="red",
      main=paste("Verteilungsfunktion von X~Bin(",n,",",p,")"),
      ylab="F(x)")
```



Aufgabe 3.1

Im Schnitt gibt es in einem von vier Haushalten mindestens zwei Fernsehgeräte. In einer Siedlung befinden sich 12 Haushalte. X sei die Anzahl der Haushalte mit mehr als einem Fernsehgerät. Bestimmen Sie die Verteilungsfunktion und berechnen Sie die Wahrscheinlichkeit, dass

(a) *höchstens die Hälfte der Haushalte maximal ein Fernsehgerät besitzt.*
(b) *mindestens die Hälfte der Haushalte mehr als ein Fernsehgerät besitzt.*
(c) *weniger als zwei Drittel der haushalte maximal ein Fernsehgerät besitzt.*
(d) *höchstens zwei Haushalte mehr als ein Fernsehgerät besitzen.*

Lösung auf Seite 50

Aufgabe 3.2

Ein Multiple-Choice Test besteht aus 30 Aufgaben mit jeweils 5 Antworten, von denen jeweils nur eine richtig ist. Mit welcher Wahrscheinlichkeit wird man durch bloßes Raten folgende Anzahl von Aufgaben richtig beantworten?

(a) *mehr als 5 Aufgaben*
(b) *mindestens 4 und höchstens 10 Aufgaben*
(c) *weniger als 3 Aufgaben*
(d) *genau 6 Aufgaben*

Lösung auf Seite 51

Aufgabe 3.3

Einer Lieferung von Kondensatoren werden 50 Kondensatoren entnommen. Erfahrungsgemäß sind 15% der Kondensatoren schadhaft. Es beschreibe X die Anzahl schadhaft entnommener Kondensatoren. Bestimmen Sie den Erwartungswert und die Varianz und interpretieren Sie beide Größen!

Lösung auf Seite 51

4 Normalverteilung

Eine der wichtigsten stetigen Verteilungen, ist die Normal- bzw. Gauß-Verteilung und wird durch zwei Parameter bestimmt:

Definition 4.1: Normalverteilung

Eine stetige Zufallsvariable mit Dichte[a]

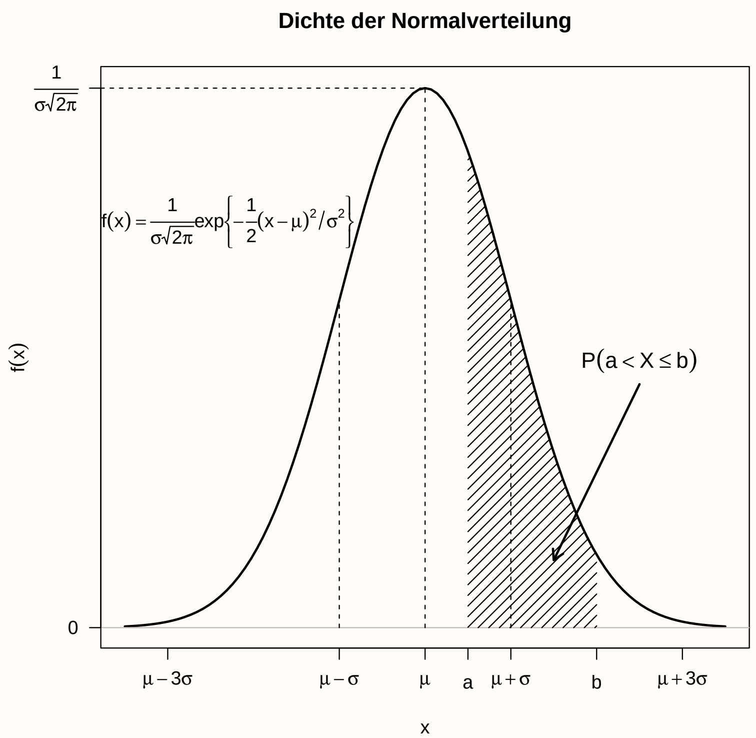
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{20}$$

heißt normalverteilt mit Parametern μ und σ, Schreibweise $X \sim \mathcal{N}(\mu, \sigma^2)$

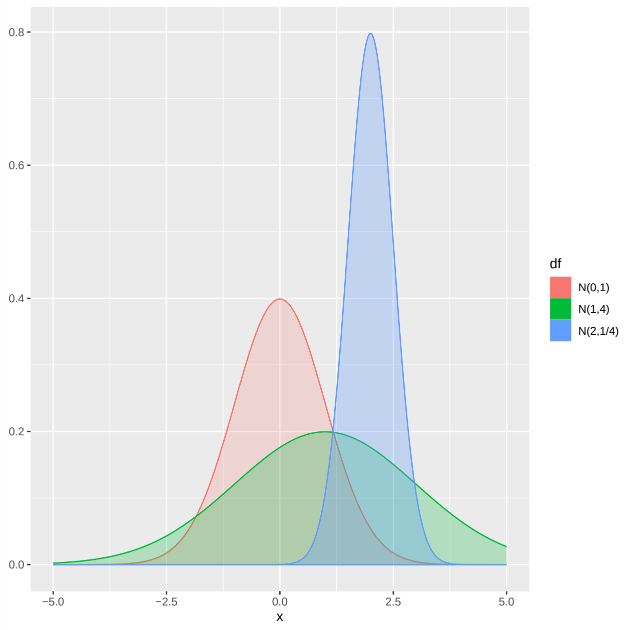
[a] Wegen $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}$ handelt es sich bei (20) tatsächlich um eine Dichte.

Der Parameter μ beschreibt das Zentrum der Verteilung und stimmt sowohl mit dem Erwartungswert als auch dem Median überein. Der zweite Parameter σ entspricht der Standardabweichung, d.h. die Varianz entspricht gerade σ^2.

Bemerkung 4.1: Interpretation der Parameter



Das Maximum der glockenförmigen Dichte wird im Punkt $x = \mu$ angenommen. Zudem geht durch diesen Punkt die Symmetrieachse der symmetrischen Dichte. Die Punkte $x = \mu \pm \sigma$ entsprechen Wendestellen. Daher bestimmt σ^2 den Breitenverlauf der Glockenkurve, d.h. ein großes σ liefert eine breite Glockenkurve, da die Werte von X stark um den Erwartungswert μ schwanken. Ein kleines σ führt hingegen zu einer nadelförmigen Glockenkurve, so dass die Werte nur wenig um μ streuen.



Die Normalverteilung findet in der Praxis eine Vielzahl von Anwendungen und nimmt auch wegen dem zentralen Grenzwertsatz, den wir in Abschnitt **??** näher bsprechen werden, eine besondere Bedeutung ein.

Bemerkung 4.2: Gaußverteilung auf alten 10 DM Noten

Abbildung 5: 10 DM Schein

Überlagern sich mehrer unabhängige ähnlich verteilte zufällige Einflussfaktoren, entsteht natürlicher Weise eine Normalverteilung.

Bemerkung 4.3: Standardnormalverteilung

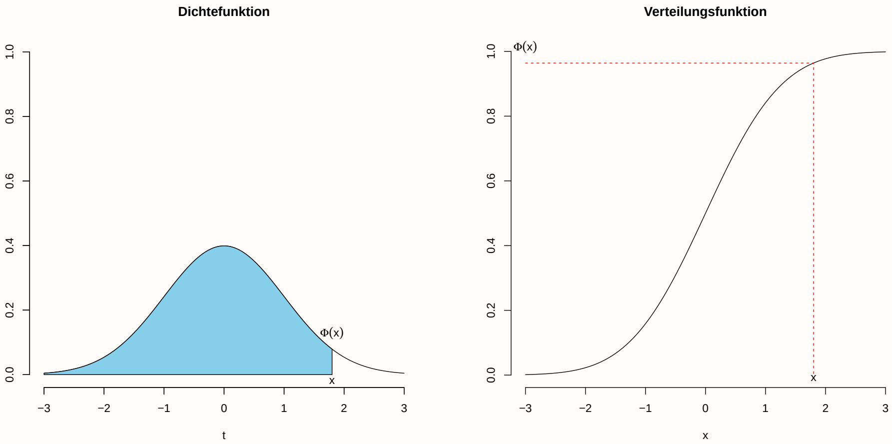
Für $\mu = 0$ und $\sigma = 1$ ergibt sich der Spezialfall der ***Standardnormalverteilung*** *$Z \sim \mathcal{N}(0,1)$, deren Dichtefunktion häufig mit φ*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{21}$$

und Verteilungsfunktion mit Φ

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \varphi(t)\, dt \tag{22}$$

bezeichnet werden.



Die Standardnormalverteilung spielt für die Berechnungen rund um die Normalverteilung eine zentrale Rolle, da es für die Verteilungsfunktion (23) von $X \sim \mathcal{N}(\mu, \sigma^2)$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{(t-\mu)}{\sigma}\right)^2} dt \tag{23}$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{24}$$

keine geschlossenen Formel gibt und daher $F(x)$ numerische berechnet werden muss. Dies kann aber für beliebige Parameterpaare $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ immer auf die Standardnormalverteilung (siehe Formel (24)) zurückgeführt werden, da die standardisierte normalverteilte Zufallsvariable

$$Z = \frac{X - \mathbb{E}X}{\sqrt{\text{Var}(X)}} = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

wegen[a]

$$F(x) = P\left(X \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \tag{25}$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(s) ds = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$= P\left(Z \leq \frac{x-\mu}{\sigma}\right) = P\left(\sigma Z + \mu \leq x\right)$$

und $F(x) = P\left(X \leq x\right) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq \frac{x-\mu}{\sigma}) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ immer standardnormalverteilt ist. Aufgrund der zentralen Bedeutung der Standardnormalverteilung ist diese auch in vielen statistischen Formelsammlungen tabelliert und in Taschenrechnern sowie Statistiksoftware implementiert.

[a] Substitution in (25) $s = \frac{t-\mu}{\sigma}$: wenn t von $-\infty$ bis x läuft, so läuft s ebenfalls von $-\infty$ bis zur oberen Grenze $\frac{x-\mu}{\sigma}$ und wegen $\frac{ds}{dt} = \frac{1}{\sigma}$ gilt $dt = \sigma ds$.

Theorem 4.1: Berechnung der Verteilungsfunktion der Normalverteilung

Die Verteilungsfunktion der normalverteilten Zufallsvariablen $X \sim \mathcal{N}(\mu, \sigma^2)$ kann für beliebige Parameter $\mu \in R$ und $\sigma > 0$ auf die Standardnormalverteilungsfunktion Φ zurückgeführt werden:

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Beispiel 4.1: Rechnen mit Normalverteilung

Wir illustrieren hier die Berechnungen für die Standardnormalverteilung $Z \sim \mathcal{N}(0,1)$ sowie für $X \sim \mathcal{N}(1,4)$ das Zurückführen auf die Standardnormalverteilung.

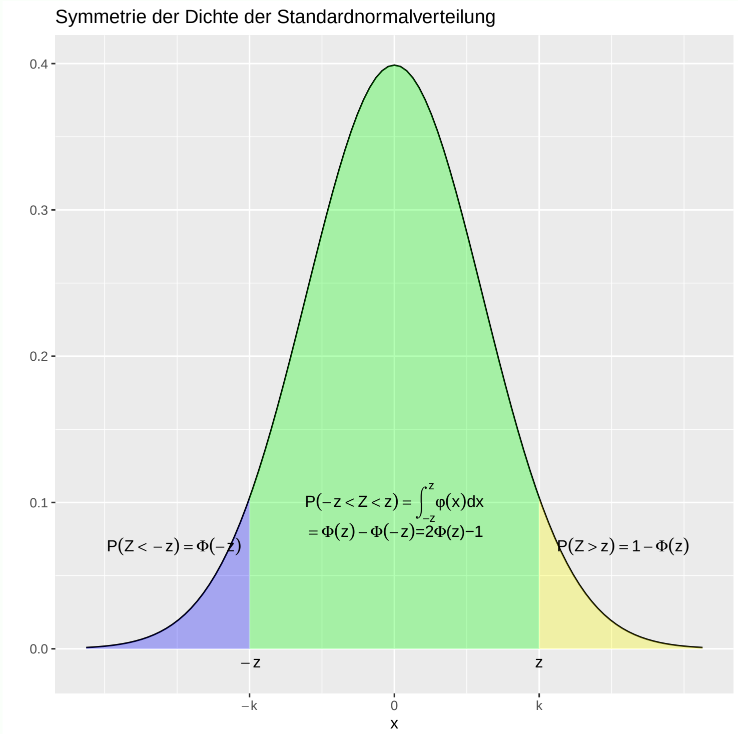
1. *Für Berechnungen von Wahrscheinlichkeiten rund um die Standardnormalverteilung, beispielsweise*

$$P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$$

nutzt man entweder Tabellen der Standardnormalverteilung und liest $\Phi(1.5) = 0.9332$ dort ab oder verwendet Statistiksoftware[a]. Viele wissenschaftliche Taschenrechner haben ebenfalls die Standardnormalverteilungsfunktion implentiert. In R setzt man obige Rechnungen wie folgt um

```
        #P(Z >1.5) = 1-P(Z<=1.5)=1-Phi(1.5)
        1-pnorm(1.5)
## [1] 0.06681
        #oder direkt P(Z>1.5) mittels Option lower.tail=FALSE
        pnorm(1.5, lower.tail = FALSE)
## [1] 0.06681
```

Bei der Verwendung von Normalverteilungstabellen sind typischerweise nur nicht-negative Werte tabelliert und gegebenenfalls muss man auf das nächste Argument runden, da natürlich nicht jeder beliebige Wert tabelliert sein kann. Die Berechnung von negativen Werten, kann man aber wegen der Symmetrie der Dichte $\varphi(-z) = \varphi(z)$ und daraus resultierent $\Phi(-z) = 1 - \Phi(z)$, wie in der folgenden Abbildung illustriert, immer umgehen. Die blau eingefärbte Fläche unter der Dichte beschreibt die Wahrscheinlichkeit $P(Z \leq -z) = \Phi(-z)$ und stimmt mit der gelb eingefärbten Fläche $P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$ überein.



Somit gilt also automatisch $P(Z < -1.5) = P(Z > 1.5) = 0.0668$.

```r
#Phi(-1.5)=P(Z <-1.5) =
pnorm(-1.5)
## [1] 0.06681
#=P(Z>1.5) = 1-Phi(1.5)
1-pnorm(1.5)
## [1] 0.06681
```

2. *Rechnungen im allgemeinen Fall $X \sim \mathcal{N}(\mu, \sigma^2)$ führt man einfach auf die Standardnormalverteilung durch die Z-Transformation $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ zurück bzw. nutzt den Zusammenhang der Verteilungsfunktion $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ zur Standardnormalverteilungsfunktion. Speziell für $X \sim \mathcal{N}(1,4)$ erhält man beispielsweise*

$$P(X \leq 2) = P\left(\frac{X-\mu}{\sigma} \leq \frac{2-\mu}{\sigma}\right) = P\left(Z < \frac{2-1}{2}\right) = \Phi(0.5) = 0.6915$$

```r
#X ~ N(1,4), d.h. mu=1 und sigma = 2 / sigma^2 =4
mu = 1
sigma = 2
#1. direkte Rechnung durch Angabe von mu / sigma
pnorm(q = 2, mean = mu, sd = sigma)
## [1] 0.6915
#2. selbständig Standardisieren und Rechnung mit
#Standardnormalverteilung
pnorm(q=(2-mu)/sigma, mean = 0, sd = 1)
## [1] 0.6915
#werden keine Parameter angegeben, rechnet R immer mit Z(0,1)
pnorm(q=(2-mu)/sigma)
## [1] 0.6915
```

Die Wahrscheinlichkeit $P(-2 \leq X < 3)$ berechnet sich mit $\mu = 1$ und $\sigma = 2$

$$\begin{aligned}P(-2 \leq X < 3) = F(3) - F(-2) &= \Phi\left(\frac{3-1}{2}\right) - \Phi\left(\frac{-2-1}{2}\right)\\ &= \Phi(1) - \Phi(-1.5) = \Phi(1) - (1 - \Phi(1.5)) = \Phi(1) - 1 + \Phi(1.5)\\ &= 0.8413 - 1 + 0.9332 = 0.7745\end{aligned}$$

```r
#1. direkte Rechnung durch Angabe von mu / sigma
pnorm(q = 3, mean = mu, sd = sigma) - pnorm(q=-2, mean = mu, sd = sigma)
## [1] 0.7745
#2. selbständig Standardisieren
pnorm(q=(3-mu)/sigma) - pnorm(q=(-2-mu)/sigma)
## [1] 0.7745
pnorm(1) -1 + pnorm(1.5)
## [1] 0.7745
```

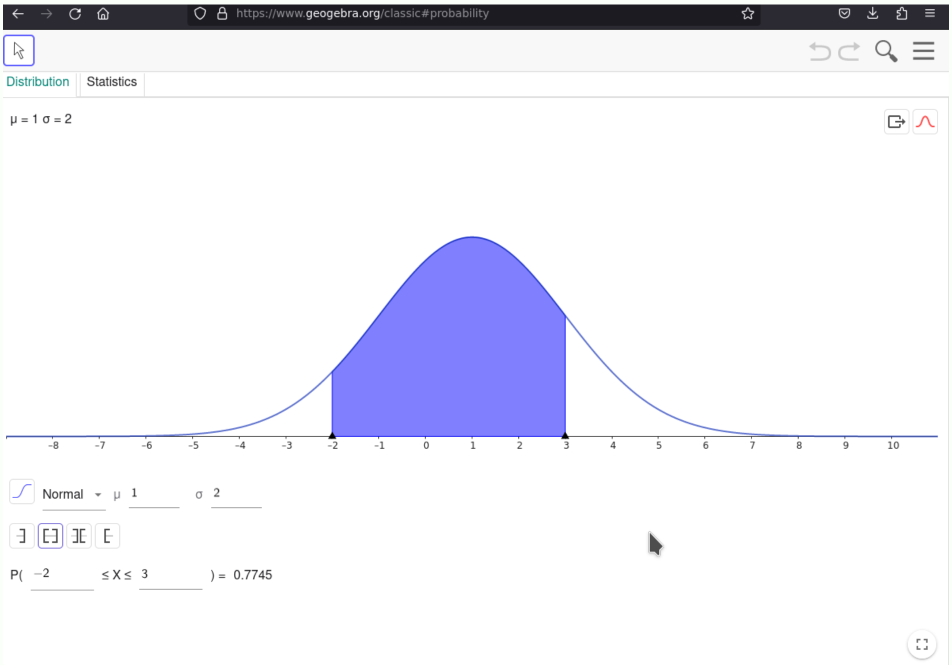


Abbildung 6: Berechnung mit Geogebra

[a] Beipsielsweise können auf `https://www.geogebra.org/classic#probability` Wahrscheinlichkeiten für Intervalle der häufigsten Verteilungen berechnet werden, wie in Abbildung 6 illustriert.

Erwartungswert und Varianz der Normalverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$ können mittels Variablensubstitutionen[7]

$$\begin{aligned}
\mathbb{E}X &= \int\limits_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\mu + \sigma z) \exp\left(-\frac{1}{2}z^2\right) dz \\
&= \mu \underbrace{\int\limits_{-\infty}^{\infty} \varphi(z) dz}_{=1} + \sigma \underbrace{\int\limits_{-\infty}^{\infty} \varphi(z) dz}_{=0} = \mu
\end{aligned} \tag{26}$$

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}\left(X - \mathbb{E}X\right)^2 &= \int\limits_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \sigma^2 z^2 \exp\left(-\frac{1}{2}z^2\right) \\
&= 2\sigma^2 \int\limits_{0}^{\infty} \underbrace{z}_{=v} \underbrace{\frac{z}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)}_{=u' \Rightarrow u = -\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}z^2} \; dz \\
&= 2\sigma^2 \left(\underbrace{\left[-z\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)\right]_0^\infty}_{=0-0=0} + \underbrace{\int\limits_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz}_{=\frac{1}{2}} \right) = \sigma^2
\end{aligned} \tag{27}$$

[7] Hierfür nutzt man wiederum $\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi} = \int\limits_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\,dx$ sowie die Substitution $z = \frac{x-\mu}{\sigma}$. Das letzte Integral in (26) verschwindet wegen der Symmetrie $\varphi(-z) = -\varphi(z)$ und in (27) stimmt wegen $(-z)^2\varphi(-z) = z^2\varphi(z)$ das Integral von $-\infty$ bis 0 mit dem Integral von 0 bis ∞ überein.

berechnet werden.

Theorem 4.2: Erwartungswert und Varianz der Normalverteilung

Für eine normalverteilte Zufallsgröße $X \sim \mathcal{N}(\mu, \sigma^2)$ mit dem Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$ gilt

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

Bemerkung 4.4: Höhere Momente der Normalverteilung

Die höheren Momente von $X \sim \mathcal{N}(\mu, \sigma^2)$ für $k = 1, 2 \ldots$ erfüllen

$$\mu_k = \mathbb{E}(X - \mathbb{E}X)^k = \begin{cases} 0 & \text{, falls } k \text{ ungerade} \\ \frac{(k)!}{2^{\frac{k}{2}}\left(\frac{k}{2}\right)!}\sigma^k & \text{, falls } k \text{ gerade} \end{cases}$$

Somit sind Schiefe und Exzess der Normalverteilung immer Null, denn

$$\Rightarrow \text{Schiefe:} \quad \gamma_1 = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}\right)^3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{(\mathrm{Var}\, X)^{\frac{3}{2}}} = \frac{\mu_3}{\sigma^3} = 0$$

$$\Rightarrow \text{Exzess:} \quad \gamma_2 = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}\right)^4 - 3 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(\mathrm{Var}(X))^2} - 3 = \frac{\mu_4}{\sigma^4} - 3 = 0$$

In Beispiel 2.13 hatten wir die $k - \sigma$-Abweichungen zum Erwartungswert bereits mittels Tschebyscheffungleichung abgeschätzt. Nun können wir die Abweichung auch exakt berechnen.

Beispiel 4.2: Sigma-Regel für die Normalverteilung, Fortsetzung Beispiel ??

Das symmetrische Intervall

$$[\mu - k\sigma, \mu + k\sigma]$$

um den Erwartungswert $\mu = \mathbb{E}X$ der normalverteilten Zufallsvariablen $X \sim \mathcal{N}(\mu, \sigma^2)$ nennt man auch k-Sigma-Intervall. Die Wahrscheinlichkeit

$$\begin{aligned} p = P\left(X \in [\mu - k\sigma, \mu + k\sigma]\right) = P\left(\mu - k\sigma \leq X \leq \mu + k\sigma\right) = P\left(-k \leq \frac{X - \mu}{\sigma} \leq k\right) \\ = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1\,, \end{aligned} \tag{28}$$

dass die Zufallsvariable Werte innerhalb dieses k-Sigma-Intervalls annimmt, hängt nur von k ab. Die folgende Tabelle zeigt die Wahrscheinlichkeiten aus Formel (28) für klassiche Werte

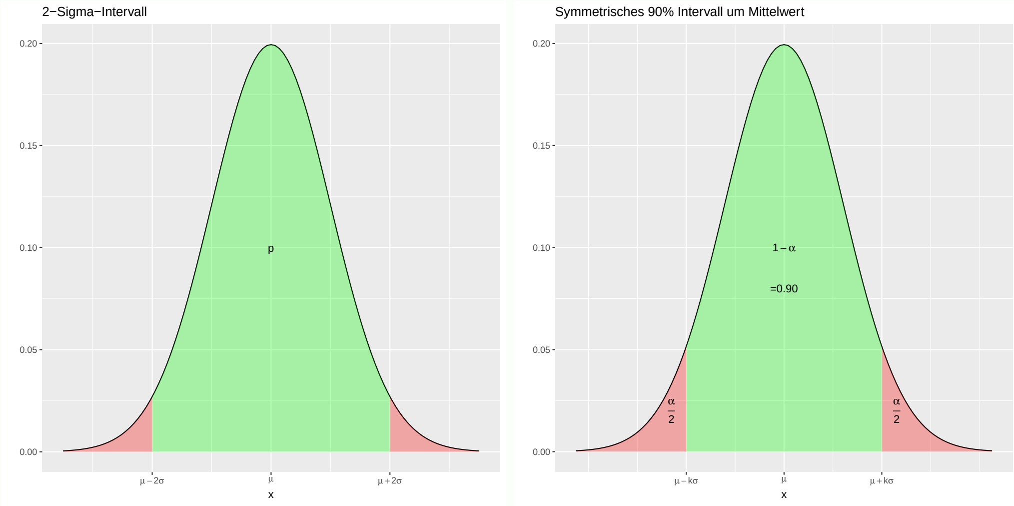
k	1	2	3	4
p	0.6827	0.9545	0.9973	0.9999

So ergibt sich für $k = 2$, dass die Wahrscheinlichkeit, dass sich die Werte der Normalverteilung um maximal 2σ vom Erwartungswert entfernen, gerade 0.9545 beträgt. Dies ist in der linken Grafik dargestellt. Man kann allerdings auch sich die Wahrscheinlichkeit $p = 1 - \alpha$ vorgeben[a] und daraus invers wegen

$$k = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

den Wert $k \in \mathbb{R}_+$ ermitteln. Für typische α Werte, die bei der Konstruktion von Konfidenzintervallen eine wichtige Rolle spielen, zeigt die folgende Tabelle die entsprechenden Werte für k. Die rechte Grafik zeigt das Intervall, dass sich für $\alpha = 0.1$ ergibt.

α	0.01	0.05	0.1	0.2
k	2.5758	1.96	1.6449	1.2816



[a] In der induktiven Statistik, die wir näher in Abschnitt **??** diskutieren werden, weist man typischerweise dem Intervall $[\mu - k\sigma, \mu + k\sigma]$ die Wahrscheinlichkeit $1 - \alpha$ zu, so dass die Wahrscheinlichkeit, die rot dargestellten Wahrscheinlichkeiten zusammen α betragen.

Die Normalverteilung ist in vielerlei Hinsicht eine besondere Verteilung und zählt zu den sogenannten stabilen Verteilungen, was bedeuted dass die Summe von unabhängigen normalverteilten Zufallsvariablen wieder normalverteilt ist. Bei einer Vielzahl anderer Verteilungen, wie beispielsweise der Exponentialverteilung oder Gleichverteilung, verlässt man durch Summenbildung hingegen die Verteilungsklassen.

Theorem 4.3: Summe unabhängiger normalverteilter Zufallsvariablen ist wieder normalverteilt

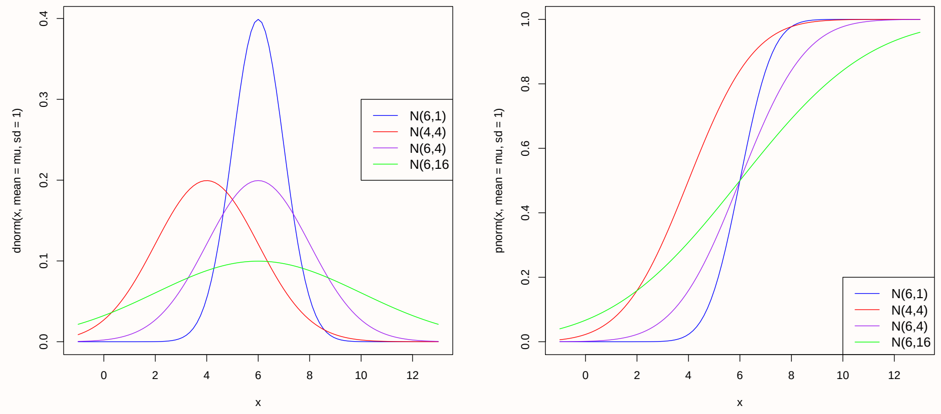
Die Summe $Z = X_1 + X_2 + \ldots + X_n$ der paarweise unabhängigen normalverteilten Zufallsvariablen $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ für $i = 1, 2, \ldots, n$, ist wieder normalverteilt mit Erwartungswert $\mu_Z = \mu_1 + \ldots + \mu_n$ und Varianz $\sigma_Z^2 = \sigma_1^2 + \ldots + \sigma_n^2$, d.h.

$$Z = \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

Bemerkung 4.5: R-Code zur Normalverteilung

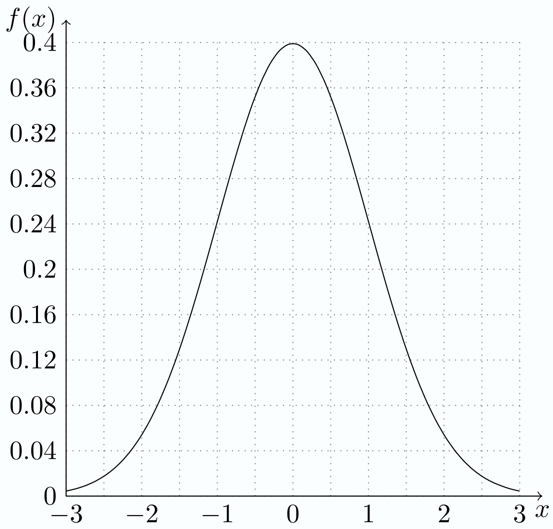
```r
#Normalverteilung
mu = 6
sigma = 2

#berechnen Dichte in x=3 (Achtung P(X=3)=0!!!)
dnorm(x=3,mean = mu, sd = sigma)
## [1] 0.06476
#berechnen Wahrscheinlichkeit, dass X<=5
pnorm(q = 5, mean =mu, sd = sigma)
## [1] 0.3085
#berechnen Wahrscheinlichkeit, dass X>5
#1. Variante: mittels Option lower.tail: P(X>5) direkt
pnorm(q=5, mean =mu, sd=sigma, lower.tail=FALSE)
## [1] 0.6915
#standardisierte Version
pnorm(q=(5-mu)/sigma, mean=0, sd=1, lower.tail = FALSE)
## [1] 0.6915
#mean = 0, sd =1 default, koennen weg gelassen werden
pnorm(q=(5-mu)/sigma, lower.tail = FALSE)
## [1] 0.6915
#2. Variante: durch Gegenereignis: P(X>5)=1-P(X<=5)=1-F(5)
1-pnorm(q=5,mean=mu, sd = sigma)
## [1] 0.6915
#Berechnung Quantile mittels inverser Verteilung
qnorm(p = 0.5, mean = mu, sd =sigma)
## [1] 6
qnorm(p=c(0.05,.95))
## [1] -1.645  1.645
#Kontrolle
pnorm(q=-1.644854)
## [1] 0.05
pnorm(q=1.644854)
## [1] 0.95
```

```r
#Dichten = density (pdf)
curve(dnorm(x, mean=mu, sd = 1), from=-1, to=13, col='blue')
curve(dnorm(x, mean =4, sd = sigma), from=-1, to=13, col='red',
           add=TRUE)
curve(dnorm(x, mean = mu, sd = sigma) , from=-1, to=13, col='purple',
           add=TRUE)
curve(dnorm(x, mean = mu, sd = 4) , from=-1, to=13, col='green',
           add=TRUE)

legend(10, .3, legend=c("N(6,1)", "N(4,4)", "N(6,4)","N(6,16"),
       col=c("blue", "red", "purple","green"), lty=1, cex=1.2)

#Verteilungsfunktion = cumulative distribution = cdf
curve(pnorm(x, mean=mu, sd = 1), from=-1, to=13, col='blue')
curve(pnorm(x, mean =4, sd = sigma), from=-1, to=13, col='red',
           add=TRUE)
curve(pnorm(x, mean = mu, sd = sigma) , from=-1, to=13, col='purple',
           add=TRUE)
curve(pnorm(x, mean = mu, sd = 4) , from=-1, to=13, col='green',
           add=TRUE)

legend(10, .2, legend=c("N(6,1)", "N(4,4)", "N(6,4)","N(6,16"),
       col=c("blue", "red", "purple","green"), lty=1, cex=1.2)
```



Aufgabe 4.1

Es sei $Z \sim \mathcal{N}(0,1)$ eine standardnormalverteilte Zufallsgröße. Schraffieren und berechnen Sie die folgenden Wahrscheinlichkeiten.



(a) $P(-2 < Z < 1)$
(b) $P(-1 < Z < 2)$
(c) $P(-1 < Z < 1)$
(d) $P(-2 < Z < 2)$
(e) $P(Z > -1)$
(f) $P(Z > 1)$

Lösung auf Seite 52

Aufgabe 4.2

Der Interlligenzquotient (IQ) ist normalverteilt mit $\mu = 100$ und $\sigma = 15$.

(a) *Welchen IQ muss man haben, um zu den intelligentesten 2% der Bevölkerung zu zählen?*
(b) *In welchem symmetrischen Bereich um den mittleren IQ liegen 80% der Bevölkerung?*
(c) *Ein einfacher Intelligenztest ist so ausgelegt, dass ihn 85% aller Personen bestehen. Welcher IQ ist dazu notwendig?*

Lösung auf Seite 52

Aufgabe 4.3

Die Auswertung eines Aufnahmetests ergibt, dass die erreichte Punktanzahl in etwa normalverteilt mit Erwartungswert $\mu = 80$ und Standardabweichung $\sigma = 12$ ist.

(a) *Wie viele Punkte muss man erreichen, um zu den besten 10% zu gehören?*
(b) *In welchem symmetrischen Intervall um den Erwartungswert liegen 95% aller Ergebnisse?*
(c) *Wie hoch ist die Punktzahl, die von 95% aller Antretenden mindestens erreicht wird?*
(d) *Es bewerben sich 2000 Personen, aber nur die besten 40 Bewerber können aufgenommen werden. Welche Punktzahl muss man dazu erreichen?*

Lösung auf Seite 52

5 Weitere Verteilungen

5.1 Hypergeometrische Verteilung

Die hypergeometrische Verteilung ähnelt der Binomialverteilung und zählt auch wie oft bei der Durchführung von n Zufallsexperimenten ein enstprechendes Ereignis eintritt. Der Unterschied zur Binomialverteilung liegt darin, dass bei der hypergeometrischen Verteilung die einzelnen Experimente nicht mehr unabhängig voneinander sind, sondern sich gegenseitig beeinflussen und dies auf ein Urnenmodell **ohne Zurücklegen** zurückgeführt werden kann, während die Binomialverteilung einem Urnenmodell **mit Zurücklegen** entspricht.

Definition 5.1: Hypergeometrische Verteilung

Eine Zufallgröße X für welche

$$P(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}},$$

für alle $m = 0,1,\ldots,m$, heißt hypergeometrisch verteilt, Kurzschreibweise $X \sim \mathrm{Hyp}(n,N,M)$.

Bemerkung 5.1: Referenzmodell für die hypergeometrische Verteilung

Wir betrachten eine Urne mit N Kugeln, davon

- *M schwarze Kugeln*
- *$N-M$ weiße Kugeln*
- *ziehen ohne Zurücklegen n Kugeln*
- *betrachten Zufallsgröße: $X \triangleq$ Anzahl der entnommenen schwarzen Kugeln*
- *falls ausreichend schwarze und weiße Kugeln in der Urne liegen, sind prinzipiell alle Werte $0,1,2,\ldots,n$ für X möglich, aber wir können natürlich nur so viele schwarze Kugeln ziehen, wie zu Beginn maximal enthalten sind. Ebenso können wir nur maximal $N-M$ weiße Kugeln ziehen und falls beispielsweise nur 2 weiße Kugeln enthalten sind, wir bei $n>2$ Ziehungen dann mindestens $n-2$ schwarze Kugeln ziehen werden. Allgemein gilt*
  - *maximale Anzahl schwarzer Kugeln: $X \leq n$ und $X \leq M$ Bsp. $n=5$, $M=4 \Rightarrow X \leq 4 = \min(n,M) \Rightarrow X \leq \min(n,M)$*
  - *minimale Anzahl schwarzer Kugeln: $X \geq 0$ und $X \geq n-(N-M)$ Bsp. $n=5$, $N-M=1 \Rightarrow X \geq 4 = \max(0,n-(N-M)) \Rightarrow X \geq \max(0,n-(N-M))$*
- *Wahrscheinlichkeit, genau m schwarze (und somit $n-m$ weiße Kugeln) zu ziehen*

$$P(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}, \tag{29}$$

  wobei $\max(0,n-(N-M)) \leq m \leq \min(n,M)$
- *Mit der natürlichen Erweiterung des Binoimalkoeffizienten $\binom{n}{k}=0$ für $k>n$ gilt Formel (29) sogar für alle $m=0,1,\ldots,n$, da die Fälle $m<n-(N-M) \Leftrightarrow n-m>N-M$ wegen $\binom{N-M}{n-m}=0$ sowie $m>M$ wegen $\binom{M}{m}=0$ tatsächlich die Wahrscheinlichkeit Null zugeordnet bekommen.*
- *die Farbzuordnung ist absolut willkürlich, hier steht das Ziehen einer schwarzen Kugel symbolisch für das Eintreten des Ereignisses A, welches von Interesse ist. Die restlichen nicht schwarzen Kugeln könnten auch verschiedene Farben annehmen, und werden hier unter weiß zusammengefassst.*

Beispiel 5.1: Urne mit roten Kugeln

In einer Urne befinden sich 12 Kugeln ($N=12$), davon sind 8 Kugeln rot ($M=8$) und 4 Kugeln blau. Wir entnehmen 5 Kugeln ($n=5$) ohne Zurücklegen. Wie groß ist die

Wahrscheinlichkeit, dass in der Stichprobe 2 rote Kugeln ($m = 2$) vorhanden sind?

$$P(X=2) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} = \frac{\binom{8}{2}\cdot\binom{4}{3}}{\binom{12}{5}} = \frac{\frac{8!}{(8-2)!\cdot 2!}\cdot\frac{4!}{(4-3)!\cdot 3!}}{\frac{12!}{(12-5)!\cdot 5!}} = \frac{\frac{8!}{6!\cdot 2!}\cdot\frac{4!}{1!\cdot 3!}}{\frac{12!}{7!5!}}$$

$$= \frac{\frac{\cancel{8}\cdot 7\cdot\cancel{6}\cdot\cancel{5}\cdot\ldots\cdot\cancel{1}}{2\cdot 1\cdot\cancel{6}\cdot\cancel{5}\cdot\ldots\cdot\cancel{1}}\cdot\frac{4\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}{\cancel{3}\cdot\cancel{2}\cdot\cancel{1}\cdot 1}}{\frac{\cancel{12}\cdot 11\cdot\cancel{10}\cdot 9\cdot\cancel{8}\cdot\cancel{7}\cdot\cancel{6}\cdot\ldots\cdot\cancel{1}}{\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot 1\cdot\cancel{7}\cdot\cancel{6}\cdot\ldots\cdot\cancel{1}}} = \frac{7\cdot 2}{11\cdot 9} = \frac{14}{99} = 0.1414$$

Achtung: bei der Verwendung von Statistiksoftware kann die Parametrisierung leicht abweichen, wie die Umsetzung mit R und Geogebra zeigen

```r
## ?dhyper #Aufruf Hilfefunktion fuer genaue Parametrisierung
## #m = Kugeln die gezaehlt werden, n = Kugeln andere Farbe, k = Anzahl Ziehen
dhyper(x = 2, m = 8, n = 4, k = 5)
## [1] 0.1414
```

Abbildung 7: Berechnung mit Geogebra

Die Berechnung von Erwartungswert und Varianz der hypergeometrischen Zufallsvariablen $X \sim \text{Hyp}(n, N, M)$ ist etwas komplizierter. Falls man $n \leq M$ und $n \leq N - M$ annimmt, d.h. es sind ausreichend weiße und schwarze Kugeln in der Urne, so dass X tatsächlich alle Werte von $0, 1, \ldots, n$ annehmen kann, vereinfacht sich die Rechnung des Erwartungswertes auf[8]

$$\begin{aligned}
\mathbb{E}X &= \sum_{m=0}^{n} mP(X=m) = \sum_{m=0}^{n} m\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} = \sum_{m=1}^{n} m\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \\
&= \sum_{m=1}^{n} M\frac{\binom{M-1}{m-1}\binom{(N-1)-(M-1)}{(n-1)-(m-1)}}{\frac{N}{n}\binom{N-1}{n-1}} = \frac{M}{N}n\sum_{m=1}^{n}\frac{\binom{M-1}{m-1}\binom{N-1-(M-1)}{(n-1)-(m-1)}}{\binom{N-1}{n-1}} \\
&= \frac{M}{N}n\underbrace{\sum_{m=0}^{n-1}\frac{\binom{M-1}{m}\binom{N-1-(M-1)}{(n-1)-m}}{\binom{N-1}{n-1}}}_{=1} = n\cdot\frac{M}{N}
\end{aligned}$$

[8] Wir nutzen $m\binom{M}{m} = M\binom{M-1}{m-1}$ sowie $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$. Außerdem nutzen wir die Vandermondsche Identität `https://en.wikipedia.org/wiki/Vandermonde%27s_identity` $\binom{m+n}{r} = \sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}$.

und die Berechnung des 2. Momentes

$$\begin{aligned}\mathbb{E}X^2 = \mathbb{E}\left(X(X-1)\right) + \mathbb{E}X &= \sum_{m=2}^{n} m(m-1)\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} + n\frac{M}{N} \\ &= \frac{M(M-1)n(n-1)}{N(N-1)} + n\frac{M}{N}\end{aligned}$$

so dass

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 &= \frac{M(M-1)n(n-1)}{N(N-1)} + n\frac{M}{N} - \left(n\frac{M}{N}\right)^2 \\ &= n\frac{M}{N}\left(\frac{(M-1)(n-1)}{N-1} + 1 - n\frac{M}{N}\right) = n\frac{M}{N}\left(\frac{N(M-1)(n-1) + N(N-1) - nM(N-1)}{N(N-1)}\right) \\ &= n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)\end{aligned}$$

Da wie in Bemerkung 5.1 erwähnt auch die Fälle $n-m > N-M$ und $m > M$ wegen $\binom{N-M}{n-m} = 0$ bzw. $\binom{M}{m} = 0$ die Wahrscheinlichkeit 29 für alle $m = 0, 1, \ldots, n$ gilt, können demnach mit obigen Formeln auch in diesen Fällen die Momente berechnet werden. Zudem kann auch wie im Falle der Binomilaverteilung mittels Summe unabhängiger 0-1-Bernoulli-Variablen auch jede hypergeometrisch Verteilte Zufallsvariable als Summe abhängiger Variablen gebildet werden. Allerdings ist die Berechnung in diesem Fall wegen der Abhängigkeit der Bernoulli-Variablen deutlich komplizierter, wie in der folgenden Bemerkung 5.2 illustriert und kann auch übergangen werden.[9]

Bemerkung 5.2: Hypergeometrische Verteilung als Summe abhängiger Bernoulli Zufallsvariablen

Wir bezeichnen mit

$$X_i = \begin{cases} 1, & \text{falls } i\text{-te Kugel schwarz (bzw. Ereignis } A \text{ im } i\text{-ten Versuch eingetreten ist)} \\ 0, & \text{sonst} \end{cases}$$

die Bernoulli verteilte Zufallsvariablen und nutzen $X = \sum_{i=1}^{n} X_i$. *Im Gegensatz zur Binomialverteilung sind jetzt aber die Bernoulli Zufallsvariablen nicht mehr unabhängig, da sich das Verhältnis von schwarzen zu weißen Kugeln nach jedem Ziehen ohne Zurücklegen in Abhängigkeit des Experimentverlaufs verändert. Nach dem ersten Ziehen stellt sich die Situation wie in Abbildung 8 illustriert dar*

[9] Zur Berechnung der Varianz $\mathrm{Var}(X+Y) = \mathbb{E}\left(X+Y-\mathbb{E}(X+Y)\right)^2$ abhängiger Zufallsvariablen benötigen wir zusätzlich die Kovarianz $\mathrm{cov}(X,Y) = \mathbb{E}\left(X-\mathbb{E}X\right)\left(Y-\mathbb{E}Y\right)$, die allerdings multivariate Verteilung benötigt, da

$$\begin{aligned}\mathrm{Var}(X+Y) = \mathbb{E}\left(X+Y-\mathbb{E}(X+Y)\right)^2 &= \mathbb{E}\left([X-\mathbb{E}X]+[Y-\mathbb{E}Y]\right)^2 \\ &= \mathbb{E}\left[X-\mathbb{E}X\right]^2 + 2\mathbb{E}\left[X-\mathbb{E}X\right]\left[Y-\mathbb{E}Y\right] + \mathbb{E}\left[Y-\mathbb{E}Y\right]^2 \\ &= \mathrm{Var}(X) + 2\,\mathrm{cov}(X,Y) + \mathrm{Var}(Y)\end{aligned}$$

Für unabängige Zufallsvariablen X und Y gilt allerdings immer $\mathrm{cov}(X,Y) = 0$, so dass sich dann die Varianz der Summe unabhängiger Zufallsvariablen vereinfachend als Summe der Varianzen ergibt.

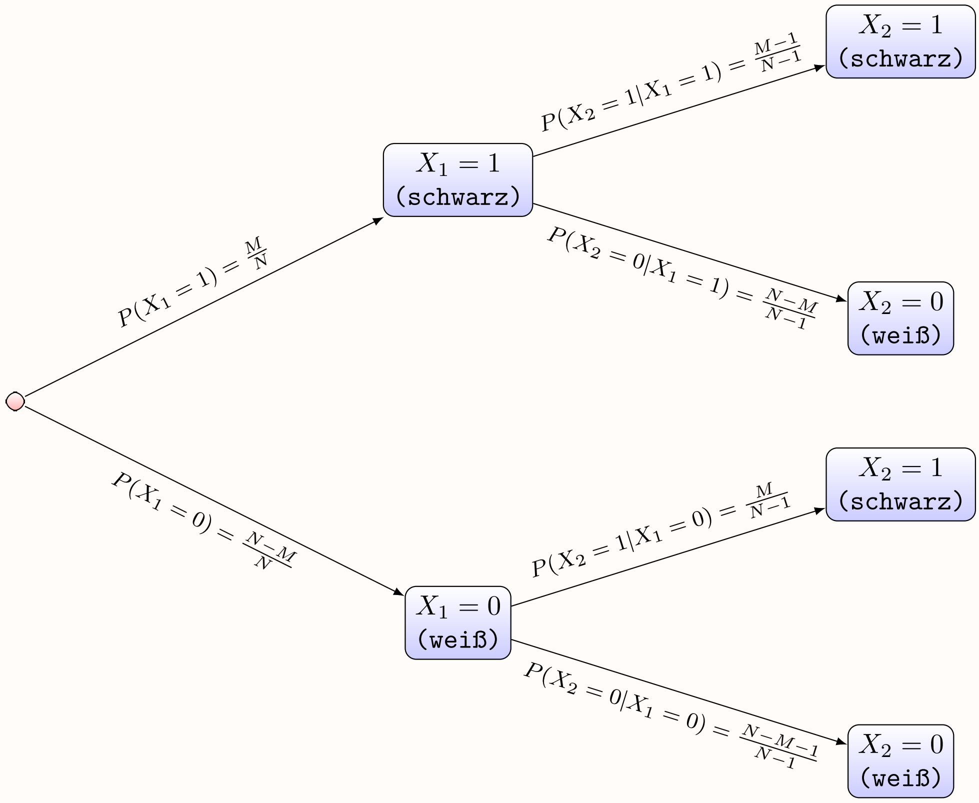


Abbildung 8: Abhängigkeiten der Bernoulli Zufallsvariablen der hypergeometrischen Verteilung

und mit dem Satz der totalen Wahrscheinlichkeit erhalten wir

$$\begin{aligned} P(X_2 = 1) &= P(X_2 = 1|X_1 = 0)P(X_1 = 0) + P(X_2 = 1|X_1 = 1)P(X_1 = 1) \\ &= \frac{M}{N}\frac{M-1}{N-1} + \frac{N-M}{N}\frac{M}{N-1} = \frac{M}{N} \end{aligned}$$

Dies lässt sich verallgemeinern[a] *(siehe auch* `https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/12%3A_Finite_Sampling_Models/12.02%3A_The_Hypergeometric_Distribution#mom1`*) und man erhält*

$$P(X_i = 1) = \frac{M}{N}$$

für alle $i = 1, \ldots, n$ aus dem Satz der totalen Wahrscheinlichkeit

$$\begin{aligned} P(X_j = 1) &= \sum_{k=0}^{j-1} \underbrace{P(X_j = 1|X_1 + \ldots + X_{j-1} = k)}_{=\frac{M-k}{N-(j-1)}} \cdot \underbrace{P(X_1 + \ldots + X_{j-1} = k)}_{=\frac{\binom{M}{k}\binom{N-M}{j-1-k}}{\binom{N}{j-1}}} \\ &= \sum_{k=0}^{j-1} \frac{M-k}{N-(j-1)} \cdot \frac{\binom{M}{k}\binom{N-M}{j-1-k}}{\binom{N}{j-1}} = \frac{M}{N} \end{aligned}$$

denn die erste Wahrscheinlichkeit entspricht dem einmaligen Ziehen aus einer Urne mit $N-(j-1)$ Kugeln, davon $M-k$ schwarze und die zweite dem $(j-1)$-maligen Ziehen ohne Zurücklegen aus einer Urne mit N Kugeln, davon M schwarzen.
Somit ergibt sich für den Erwartungswert jeder Bernoulli Zufallsvariablen

$$\mathbb{E}X_i = 1 \cdot \frac{M}{N} + 0 \cdot \frac{N-M}{N} = \frac{M}{N}.$$

Damit erhalten wir dann auch sofort den Erwartungswert für die Summe

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}X_i = n\frac{M}{N}.$$

Analog bekommt man für das zweite Moment der Bernoulli Variablen $\mathbb{E}X_i^2 = \frac{M}{N}$ und somit die Varianz $\mathrm{Var}(X_i) = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = \frac{M}{N}\frac{N-M}{N}$. Für die Varianzberechnung der Summe müssen nun aber wegen der Abhängigkeiten der Bernoulli Variablen deren Kovarianzen berücksichtigt werden Für die Summe der Zufallsvariablen $X_1, \ldots, X_n$ berechnet sich die Varianz als Summe allermöglichen Kovarianzen, genauer

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\right)^2 = \mathbb{E}\left(\sum_{i=1}^{n} (X_i - \mathbb{E}X_i)\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{n} (X_i - \mathbb{E}X_i) \sum_{j=1}^{n} (X_j - \mathbb{E}X_j)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left(X_i - \mathbb{E}X_i\right)\left(X_j - \mathbb{E}X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{cov}(X_i, X_j)\,.
\end{aligned}
$$

Für die abhängigen Bernoulli Variablen mit den Werten Null oder Eins ist das Produkt X_iX_j

$$
X_iX_j = \begin{cases} 1, & X_i = 1, X_j = 1 \\ 0, & \textit{sonst} \end{cases}
$$

genau dann 1, wenn beide jeweils den Wert 1 annehmen. Somit ist für diesen einfachen Fall $\mathbb{E}\left(X_iX_j\right) = 1 \cdot P(X_i = 1, X_j = 1)$ und für $i \neq j$ erhalten wir[b]

$$
\mathbb{E}\left(X_iX_j\right) = P(X_i = 1, X_j = 1) = \frac{M}{N}\frac{M-1}{N-1}\,,
$$

und deswegen

$$
\begin{aligned}
\mathrm{cov}(X_i, X_j) &= \mathbb{E}(X_iX_j) - \mathbb{E}X_i\mathbb{E}X_j = 1 \cdot P(X_i = 1, X_j = 1) - \frac{M^2}{N^2} \\
&= \frac{M-1}{N-1} \cdot \frac{M}{N} - \frac{M^2}{N^2} \\
&= -\frac{M}{N}\frac{N-M}{N(N-1)} < 0
\end{aligned}
$$

und letztendlich die Varianz der Summe

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, X_j) = n\frac{M}{N}\frac{N-M}{N} - (n^2 - n)\frac{M}{N}\frac{N-M}{N(N-1)} \\
&= \ldots = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1} = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}
\end{aligned}
$$

[a]Nebenrechnung siehe `https://www.wolframalpha.com/input?i=sum%28%28M-k%29%2F%28N-%28j-1%29%29*binom%28M%2Ck%29*binom%28N-M%2Cj-1-k%29%2Fbinom%28N%2Cj-1%29%2Ck%3D0..j-1%29`

[b]Einfach zu sehen ist zunächst der Fall $i = 1$ und $j = 2$, danach verallgemeinert man für i und $j = i + 1$ mittels $P(X_i = 1, X_{i+1} = 1) = P(X_{i+1} = 1 | X_i = 1)P(X_i = 1) = \frac{M-1}{N-1} \cdot \frac{M}{N}$ für $i = 1$ und $j = i + 1 = 2$ und dannach auf $j > i$.

5.2 Exponentialverteilung

Die Exponentialverteilung wird häufig zur Modellierung von Wartezeiten oder Lebensdauern verwendet. Die Wartezeiten eines Poisson-Prozesses[10] sind exponentialverteilt.

[10]Der Poisson-Prozess zum Zeitpunkt t zählt wie oft ein Ereignis bis zum Zeitpunkt t eingetreten ist.

Definition 5.2: Exponentialverteilung

Eine stetige Zufallsvariable X mit Dichte

$$f(x) = \begin{cases} \mu e^{-\mu x}, & \text{für } x \geq 0 \\ 0, & \text{für } x < 0 \end{cases} \tag{30}$$

und Verteilungsfunktion

$$F(x) = \begin{cases} 0, x < 0 \\ 1 - e^{\mu x}, & x \geq 0 \\ 1, & x > b \end{cases} \tag{31}$$

heißt exponentialverteilt mit Parameter $\mu > 0$. Abkürzend schreiben wir $X \sim \mathrm{Exp}(\mu)$.

Die Berechnung von Erwartungswert

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{0}^{\infty} \mu x e^{-\mu x}\, dx \quad \text{Partielle Integration} \\
&= \mu \int_0^\infty \underbrace{x}_{=v(x)} \underbrace{e^{-\mu x}}_{=u'(x)}\, dx \quad \int v(x)u'(x) = u(x)v(x) - \int v'(x)u(x)\, dx \\
&= \mu \Big(\Big[\underbrace{-\frac{1}{\mu} e^{-\mu x}}_{u(x)} \underbrace{x}_{v(x)} \Big]_0^\infty - \int_0^\infty \underbrace{1}_{v'(x)} \underbrace{\Big(-\frac{1}{\mu} \Big) e^{-\mu x}}_{u(x)}\, dx \Big) \\
&= \mu \Big(0 - 0 - \Big[\frac{1}{\mu^2} e^{-\mu x} \Big]_0^\infty \Big) = \mu \frac{1}{\mu^2} = \frac{1}{\mu}
\end{aligned}$$

und 2. Moment $\mathbb{E}X^2$ lässt sich mittels partieller Integration durchführen

$$\begin{aligned}
\mathbb{E}X^2 &= \int_0^\infty x^2 \mu e^{-\mu x}\, dx \\
&= \mu \Big(\Big[\underbrace{-\frac{1}{\mu} e^{-\mu x}}_{u(x)} \underbrace{x^2}_{v(x)} \Big]_0^\infty - \int_0^\infty \underbrace{2x}_{v'(x)} \underbrace{\Big(-\frac{1}{\mu} \Big) e^{-\mu x}}_{u(x)}\, dx \Big) \\
&= \mu \Big(0 - 0 - 2\big(-\frac{1}{\mu} \big) \int_0^\infty x e^{-\mu x}\, dx \Big) \\
&= \frac{2}{\mu} \int_0^\infty x \mu e^{-\mu x}\, dx \quad \text{vgl. Erwartungswert} \\
&= \frac{2}{\mu} \frac{1}{\mu} = \frac{2}{\mu^2},
\end{aligned}$$

so dass sich für die Varianz

$$\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\mu^2} - \Big(\frac{1}{\mu} \Big)^2 = \frac{1}{\mu^2}$$

ergibt.

Theorem 5.1: Erwartungswert und Varianz der Exponentialverteilung

Für eine exponentialverteilte Zufallsgröße $X \sim \mathrm{Exp}(\mu)$ mit dem Parameter $\mu > 0$ gilt

$$\mathbb{E}X = \frac{1}{\mu} \quad \text{und} \quad \mathrm{Var}(X) = \frac{1}{\mu^2}$$

Beispiel 5.2: PKW-Inspektion

Die durchschnittliche Dauer einer PKW-Inspektion in einer Werkstatt dauert 2 Stunden. Wie groß ist die Wahrscheinlichkeit, dass eine Inspektion länger als 3 Stunden dauert?
Lösung:

- *Zufallsvariable* $X \triangleq$ *Zeitdauer (in Stunden) für eine PKW-Inspektion, Annahme:* $X \sim \mathrm{Exp}(\mu)$
- *Gesucht:* $P(X > 3)$
- *Bestimmung des Parameters* μ: $\mathbb{E}X = 2 = \frac{1}{\mu} \Rightarrow \mu = \frac{1}{2}$
- *Rechnung:*

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - F(3) \\ &= 1 - \left(1 - e^{-3\mu}\right) = e^{-\frac{3}{2}} = 0.223 \end{aligned}$$

d.h. in durchschnittlich 22.3% aller Fälle dauert die Inspektion länger als 3 Stunden

Bemerkung 5.3: Gedächtnislosigkeit der Exponentialverteilung

Die Exponentialverteilung wird aufgrund ihres Zusammenhangs zu Poisson-Prozessen häufig zur Modellierung von Lebensdauern verwendet. Allerdings werden dann wegen

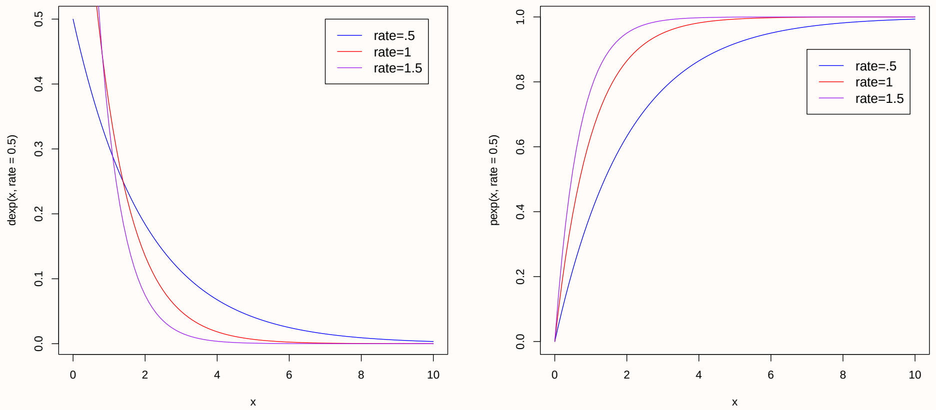
$$\begin{aligned} P(X \leq x_0 + x | X > x_0) &= \frac{P\left(\{X \leq x_0 + x\} \cap \{X > x_0\}\right)}{P(X > x_0)} \\ &= \frac{P(x_0 < X \leq x_0 + x)}{1 - P(X \leq x_0)} = \frac{F(x_0 + x) - F(x_0)}{1 - F(x_0)} \\ &= \frac{1 - e^{-\mu(x_0 + x)} - (1 - e^{-\mu x_0})}{1 - (1 - e^{-\mu x_0})} \\ &= \frac{e^{-\mu x_0} - e^{-\mu(x_0 + x)}}{e^{-\mu x_0}} = 1 - e^{-\mu x} = F(x) \\ &= P(X \leq x) \end{aligned}$$

keine „Spätausfälle“ aufgrund von Alterungserscheinungen berücksichtigt.

Bemerkung 5.4: R-Code zur Exponentialverteilung

```
#Exponential-Verteilung
mu = 1/5

#berechnen Dichte in x=3 (Achtung P(X=3)=0!!!)
dexp(x=3,rate = mu)
## [1] 0.1098
#berechnen Wahrscheinlichkeit, dass X<=5
pexp(q = 5,rate = mu)
## [1] 0.6321
#berechnen Wahrscheinlichkeit, dass X>5
#1. Variante: mittels Option lower.tail: P(X>5) direkt
pexp(q=5,rate = mu, lower.tail=FALSE)
## [1] 0.3679
#2. Variante: durch Gegenereignis: P(X>5)=1-P(X<=5)=1-F(5)
1-pexp(q=5,rate=mu)
## [1] 0.3679
```

```
#Dichten = density (pdf)
curve(dexp(x, rate = .5), from=0, to=10, col='blue')
  curve(dexp(x, rate = 1), from=0, to=10, col='red', add=TRUE)
  curve(dexp(x, rate = 1.5), from=0, to=10, col='purple', add=TRUE)

legend(7, .5, legend=c("rate=.5", "rate=1", "rate=1.5"),
       col=c("blue", "red", "purple"), lty=1, cex=1.2)

#Verteilungsfunktion = cumulative distribution = cdf

curve(pexp(x, rate = .5), from=0, to=10, col='blue')
  curve(pexp(x, rate = 1), from=0, to=10, col='red', add=TRUE)
  curve(pexp(x, rate = 1.5), from=0, to=10, col='purple', add=TRUE)
  #add legend
  legend(7, .9, legend=c("rate=.5", "rate=1", "rate=1.5"),
       col=c("blue", "red", "purple"), lty=1, cex=1.2)
```



6 Das Gesetz der großen Zahlen

Ein Grundlage um typische Fragestellungen der beurteilenden Statistik bearbeiten zu können, bildet die Frage nach der Verteilung von $Y_n = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ für identisch verteilte und unabhängige Zufallsvariablen $X_1, X_2, \ldots, X_n$ mit $\mathbb{E}X_k = \mu$ und $\mathrm{Var}(X_k) = \sigma^2$. Insbesondere der Fall $a_1 = \ldots = a_n = \frac{1}{n}$ spielt in der beurteilenden Statistik eine zentrale Rolle, da er auf das arithmetische Mittel

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

führt und dieses nach dem schwachen Gesetz der großen Zahlen gegen den Erwartungswert μ konvergiert, falls n hinreichend groß gewählt wird. Somit kann man aus einer repräsentativen Stichprobe mit unabhängigen Beobachtungen durch die Bildung des Stichprobenmittels Rückschlüsse auf den Parameter $\mu = \mathbb{E}X$ der beobachteten Zufallsvariablen ziehen.

Egal wie die identisch verteilten unabhängigen Zufallsvariablen $X_1, X_2, \ldots, X_n$ verteilt sind, gilt für den Erwartungswert des arithmetischen Mittels stets

$$\mathbb{E}\overline{X} = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = \mu \tag{32}$$

und wegen der Unabhängigkeit[11] für die Varianz

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) = \frac{\sigma^2}{n} \underset{n\to\infty}{\longrightarrow} 0. \tag{33}$$

[11] Aus der Unabhängigkeit von X_k und X_j folgt $\mathbb{E}(X_kX_j) = \mathbb{E}X_k\mathbb{E}X_j$ und somit $\mathrm{cov}(X_k, X_j) = \mathbb{E}((X_k - \mathbb{E}X_k)(X_j - \mathbb{E}X_j)) = \mathbb{E}(X_kX_j) - \mathbb{E}X_k\mathbb{E}X_j = 0$ für alle $k \neq j$, so dass die Varianz der Summe unabhängiger Zufallsvariablen $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{cov}(X_j, X_k) = \sum_{j=1}^{n}\mathrm{cov}(X_j, X_j) = \sum_{j=1}^{n}\mathrm{Var}(X_j)$ die Summe aller Einzelvarianzen ist.

Somit folgt aus der Tschebyscheff'schen Ungleichung für beliebiges $\varepsilon > 0$ die Ungleichung

$$P\left(|\overline{X}_n - \mu| > \varepsilon\right) = P\left(|\overline{X}_n - \mathbb{E}\overline{X}_n| > \varepsilon\right) < \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

und folglich gilt für jedes $n \in \mathbb{N}$

$$1 \geq P\left(|\overline{X}_n - \mu| \leq \varepsilon\right) > 1 - \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 1\,,$$

womit letztendlich der Grenzfall $n \to \infty$ dem schwachen Gesetz der großen Zahlen

$$\lim_{n\to\infty} P\left(|\overline{X}_n - \mu| \leq \varepsilon\right) = 1 \tag{34}$$

entspricht. Dieses besagt, dass die Wahrscheinlichkeit einer Abweichung größer als $\varepsilon > 0$ zwischen arithmetischen Mittel und Erwartungswert für hinreichend große n gegen Null konvergiert.

Einen Spezialfall der schwachen Konvergenz der relativen Häufigkeit $H_n(A)$ gegen die Wahrscheinlichkeit $p = P(A)$ des Ereignisses A, erhält man durch n unabhängige Wiederholungen des Experimentes. Hierfür betrachten wir die Bernoulli verteilten Zufallsgrößen

$$X_i = \begin{cases} 1, & \text{wenn } A \text{ im } i-\text{ten Versuch eintritt} \\ 0, & \text{wenn } A \text{ im } i-\text{ten Versuch nicht eintritt,} \end{cases}$$

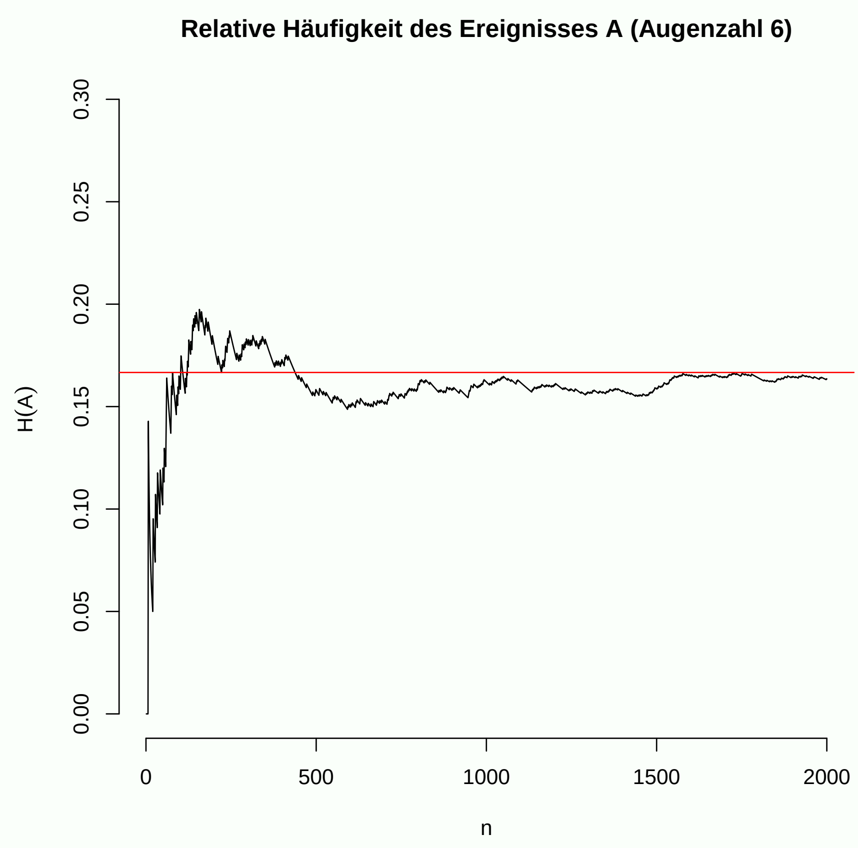
die unabhängig und identisch verteilt sind. Dann gilt $\mu = \mathbb{E}X_i = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = P(A) = p$ und das arithmetische Mittel $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ entspricht der relativen Häufigkeit $H_n(A)$ des Ereignisses A. Daher besagt das schwache Gesetz der großen Zahlen, dass für wachsende Versuchsanzahl n die relative Häufigkeit $H_n(A) = \overline{X}$ schwach gegen die Wahrscheinlichkeit $p = P(A)$, genauer

$$\lim_{n\to\infty} P(|H_n(A) - p| \leq \varepsilon) = 1$$

konvergiert.

Beispiel 6.1: Konvergenz der relativen Häufigkeit beim Würfeln

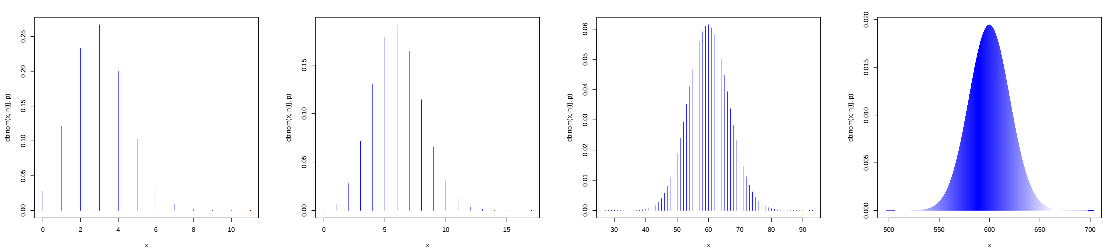
Das schwache Gesetz der großen Zahlen gewährleistet, dass relative Häufigkeiten gegen die Wahrscheinlichkeiten konvergieren. Dies wollen wir anhand einer Simulation verdeutlichen und würfeln insgesamt $n = 2000$ mal mit einem fairen Würfel und zählen wie oft das Ereignis A, dass eine Sechs gewürfelt wird, eintritt und berechnen daraus die relative Häufigkeit. Diese tragen wir in der folgenden Grafik gegen n ab, d.h. für $n = 10$ werden bei der Berechnung der realtiven Häufigkeit nur die ersten zehn Simulationen berücksichtigt, während für $n = 2000$ alle Ergebnisse Berücksichtigung finden.

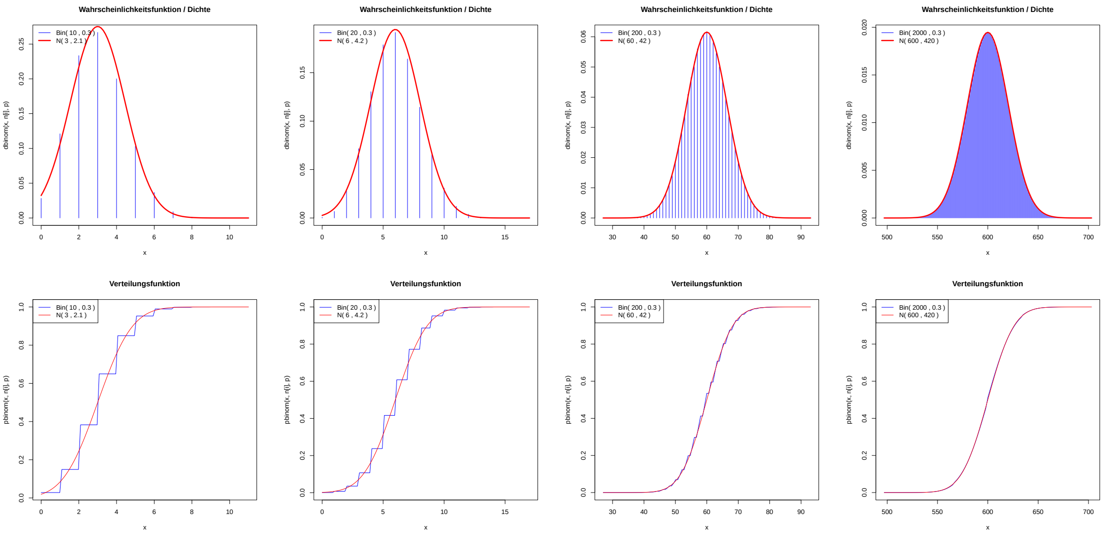


Beispiel 6.1 zeigt sehr deutlich, dass für größere n die relative Häufigkeit $H_n(A)$ des Ereignisses A gegen die Wahrscheinlichkeit $p = P(A) = 1/6$ konvergiert. Es ist allerdings auch zu erkennen, dass die relative Häufigkeit schwankt und selbst zufällig ist. Die Frage nach der Verteilung des arithmetischen Mittels beantwortet der zentrale Grenzwertsatz, den wir in Abschnitt 6.2 kurz etwas näher betrachten wollen. Der Grenzwertsatz von Moivre und Laplace kann als Spezialfall des zentralen Grenzwertsatzes betrachtet werden und beschreibt die Normalapproximation der Binomialverteilung für große n.

6.1 Grenzwertsatz von Moivre und Laplace: Approximation der Binomialverteilung durch die Normalverteilung

Der Grenzwertsatz von Moivre und Laplace, nachdem sich die Binomialverteilung von $X \sim \text{Bin}(n, p)$ durch die Normalverteilung approximieren lässt, sofern der Parameter n hinreichend groß ist, stellt einen Spezialfall des zentralen Grenzwertsatzes dar und soll hier nur kurz anhand eines Beispiels demonstriert werden. Hierfür betrachten wir für $p = 0.3$ und vier verschiedene Parameter $n \in \{10, 20, 200, 2000\}$ jeweils die Zufallsgrößen $X = X_1 + X_2 + \ldots + X_n = n\overline{X} \sim \text{Bin}(n, p)$ als Summe Bernoulli verteilter Zufallsvariablen $X_1, \ldots, X_n$ (siehe Formel (16)) und stellen grafisch die Wahrscheinlichkeitsfunktion der entsprechenden Binomialverteilung und die Dichte der approximierten Normalverteilung sowie deren Verteilungsfunktionen dar. Historisch wurde diese Approximation 1730 von Abraham de Moivre für $p = 1/2$ und im Jahre 1812 von Pierre-Simon Laplace für allgemeines $0 < p < 1$ gezeigt, bevor dann später Lindberg und Lévy den zentralen Grenzwertsatz für allgemeine Verteilungen formulierten.





Es ist deutlich zu erkennen, dass die Normalapproximation für $n = 200$ gut und für $n = 2000$ sehr gut die Binomialverteilung beschreibt, während für $n = 10$ und $n = 20$ die Annäherungen der stetigen Normalverteilung an die diskrete Binomialverteilung noch recht grob sind. Der Grenzwertsatz macht nur für hinreichend große n Aussagen. Für praktische Anwendungen stellt sich natürlich die Frage, ob n bereits hinreichend groß ist. Als Faustregel gilt, dass unter Erfüllung der Forderung $\mathrm{Var}(X) = np(1-p) > 9$ die Normalapproximation gute Näherungen liefert.

Zum Abschluss dieses Abschnittes wollen wir noch die sogenannte Stetigkeitskorrektur besprechen, die bei der Normalapproximation in der Regel bessere Ergebnisse liefert. Wie obige Grafik zeigt, ist gerade für kleiner n der Übergang von der treppenförmigen Verteilungsfunktion der Binomialverteilung zur stetigen Verteilungsfunktion der Normalverteilung teilweise recht grob. Somit weichen dann auch die Ergebnisse der Normalapproximation recht stark von dem exakten Ergebnis ab. Durch die Stetigkeitskorrektur versucht man die Approxiamtion zu verbessern, indem die Grenzen leicht verschoben werden. Zur Illustration der Probleme beim Übergang von diskreten zu stetigen Zufallsvariablen betrachten wir für beliebige Intervallgrenzen $a < b \in \mathbb{N}$ die folgende Wahrscheinlichkeiten

$$P\left(a \leq X \leq b\right) = P\left(a-1 < X < b+1\right) \tag{35}$$

Für die diskrete Zufallsvariable $X \sim \mathrm{Bin}(n,p)$ gilt tatsächlich obige Gleichung, während sie für eine stetige Zufallsvariable im Allgemeinen nicht wahr ist. Die Normalapproximation auf der linken und rechten Seite von (35) liefert daher verschiedene Ergebnisse

$$\begin{aligned}
&\Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \\
&\qquad \neq \Phi\left(\frac{b+1-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-1-np}{\sqrt{np(1-p)}}\right).
\end{aligned}$$

Mit der Stetigkeitskorrektur

$$\begin{aligned}
P\left(a \leq X \leq b\right) &= P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right) \\
&\approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)
\end{aligned} \tag{36}$$

deutlich bessere Ergebnisse, wie das folgende Beispiel illustriert.

Beispiel 6.2: Stetigkeitskorrektur bei der Normalapproximation von $X \sim \mathrm{Bin}(100, 0.25)$ für $a = 15$ und $b = 30$

Die exakte Berechnung von

$$P(15 \leq X \leq 30) = \sum_{k=15}^{30} \binom{100}{k} 0.25^k 0.75^{100-k} = F(30) - F(14) = 0.8908$$

ist zwar leicht mit Statistiksoftware zu berechnen. Mit einem einfachen Taschenrechner ist *die Berechnung allerdings recht umständlich, so dass die Normalapproximation in Verbindung mit der Standardnormalverteilungstabelle hier einen einfachen Weg zur Berechnung liefert. Die Faustregel $np(1-p) = 18.75 > 9$ ist erfüllt, so dass die Normalapproximation der linken Seite von Gleichung (35)*

$$P(15 \leq X \leq 30) \approx \Phi\left(\frac{30-25}{\sqrt{18.75}}\right) - \Phi\left(\frac{15-25}{\sqrt{18.75}}\right) = 0.8654$$

ebenso wie

$$\begin{aligned} P(15 \leq X \leq 30) &= P(14 < X \leq 30) = F(30) - F(14) \\ &\approx \Phi\left(\frac{30-25}{\sqrt{18.75}}\right) - \Phi\left(\frac{14-25}{\sqrt{18.75}}\right) = 0.8704 \end{aligned}$$

oder die rechte Seite von Gleichung (35)

$$\begin{aligned} P(15 \leq X \leq 30) &= P(14 < X < 31) \\ &\approx \Phi\left(\frac{31-25}{\sqrt{18.75}}\right) - \Phi\left(\frac{14-25}{\sqrt{18.75}}\right) = 0.9115 \end{aligned}$$

eine gute Näherung liefern. Die Stetigkeitskorrektur

$$P(15 \leq X \leq 30) = P(14.5 < X < 30.5) \approx \Phi\left(\frac{30.5-25}{\sqrt{18.75}}\right) - \Phi\left(\frac{14.5-25}{\sqrt{18.75}}\right) = 0.8903$$

liefert allerdings eine deutlich bessere Approximation.

Aufgabe 6.1

Ein Würfel wird

(a) *42 mal,*
(b) *72 mal,*
(c) *114 mal*

geworfen. X mit $\mu = \mathbb{E}X$ und $\sigma^2 = \mathrm{Var}(X)$ gibt an, wie of die Augenzahl 5 oder 6 gewürfelt wird. Berechnen sie die Wahrscheinlichkeit, dass X um maximal σ von μ abweicht exakt und näherungsweise.

Lösung auf Seite 53

Aufgabe 6.2

Schätzungen zufolge ist der Anteil der Mopedfahrer, die ein auffrisiertes Moped fahren, 40%.

(a) *Die Polizei kontrolliert 120 Mopedfahrer. Mit welcher Wahrscheinlichkeit werden weniger als 40 mit einem auffrisierten Moped angetroffen?*

(b) *Bei einer weiteren Kontrolle sind von 100 Mopedfahrern 45 mit einem auffrisierten unterwegs. Wie groß ist die Wahrscheinlichkeit für eine so starke oder eine noch stärkere Abweichung*[a] *vom erwarteten Wert?*

Lösung auf Seite 53

[a] Es interessieren an dieser Stelle nur Abweichungen nach oben, d.h. Werte über dem Erwartungswert.

Aufgabe 6.3

Im letzten Jahr hatte die Computerfirma „Digitoll" einen Marktanteil von 11 %.

(a) *Berechnen Sie die Wahrscheinlichkeit, dass von 400 Computerkäufern sich mehr als 50 für die Marke „Digitoll" entscheiden.*

(b) *Auf einer Computermesse, die dieses Jahr stattfindet, werden 8 000 Besucher erwartet, die einen Computer kaufen werden. Wie viele davon werden einer Computer der Marke „Digitoll" kaufen? Bestimmen Sie dazu eine minimale bzw. maximale Anzahl symmetrisch um den Erwartungswert, so dass die tatsächliche Anzahl mit 95 %-iger Wahrscheinlichkeit zwischen Minimal- und Maximalwert liegt.*

Lösung auf Seite 54

6.2 Zentraler Grenzwertsatz

Wir betrachten das arithmetische Mittel $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ von beliebig, aber identisch verteilten und unabhängigen Zufallsvariablen $X_1, X_2, \dots, X_n$ mit $\mu = \mathbb{E}X_i \in \mathbb{R}$ sowie $0 < \operatorname{Var}(X_i) = \sigma^2 < \infty$. Dann konvergiert die Verteilung des standardisierten arithmetischen Mittels

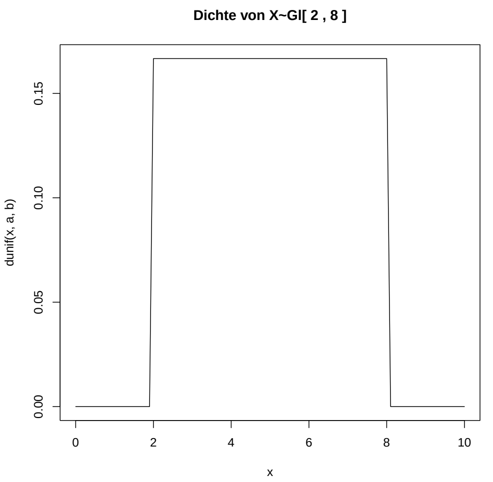
$$Z_n = \frac{\overline{X}_n - \mathbb{E}\overline{X}_n}{\sqrt{\operatorname{Var}(\overline{X}_n)}} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}$$

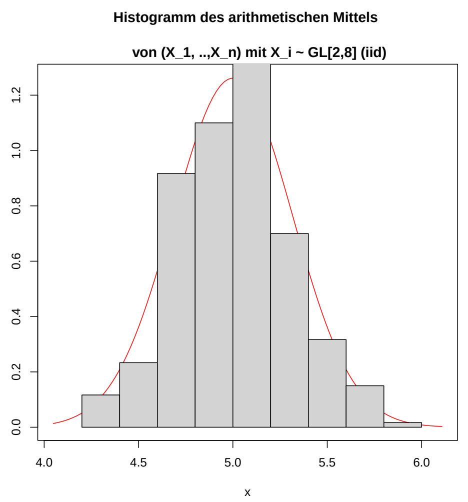
für alle $x \in \mathbb{R}$ gegen die Standardnormalverteilung, d.h. es gilt

$$\lim_{n\to\infty} P\left(Z_n \le x\right) = \lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \le x\right) = \Phi(x).$$

Für große Anzahl n ist somit das standardisierte arithmetische $Z_n \approx \mathcal{N}(0,1)$ näherungsweise standardnormalverteilt. Daher ist das arithmetische Mittel $\overline{X}_n$ für große n näherungsweise $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ verteilt und die Summe $X_1 + \ldots + X_n$ für große n näherungsweise $\mathcal{N}(n\mu, n\sigma^2)$ verteilt.

Hier wählen wir zur Illustration die Gleichverteilung und mitteln über $n = 30$ unabhängige und gleichverteilte Zufallsgrößen. Die linke Grafik zeigt die Dichtefunktion der Gleichverteilung und die rechte Grafik die Verteilung des arithmetischen Mittels.





7 Gemischte Aufgaben

Aufgabe 7.1

Eine Fluggesellschaft verkauft Tickets an Geschäftsleute für ein Flugzeug mit 100 Sitzplätzen. Aus Erfahrung ist allerdings bekannt, dass durchschnittlich nur 85 von 100 Geschäftskunden tatsächlich den Flug antreten und überbuchen daher den Flug. Tatsächlich verkaufen Sie 120 Tickets. Gehen Sie davon aus, dass alle Geschäftsleute unabhängig voneinander entscheiden, den Flug anzutreten und verfallen lassen.

(a) *Mit welcher Wahrscheinlichkeit möchten mehr Passagiere den Flug antreten als Sitzplätze verfügbar sind?*

(b) *Angenommen ein Ticket ...*

(c) *Wie viele Sitze müsste das Flugzeug haben, damit bei $n = 120$ verkauften Tickets mit mindestens 95%-iger Wahrscheinlichkeit keine Überbuchung auftritt?*

Lösung auf Seite 54

Aufgabe 7.2

Die Masse der Packungen des Waschmittels „Riesenweiß" sei normalverteilt mit Erwartungswert 3200 Gramm und Standardabweichung 100 Gramm.

(a) *Berechnen Sie die Wahrscheinlichkeit, dass eine Packung weniger als 3150 Gramm Masse besitzt.*

(b) *Der Produzent möchte garantieren, dass höchsten 9% aller Packungen um mehr als c Gramm vom Erwartungswert abweichen. Wie muss c gewählt werden?*

(c) *Der Produzent möchte eine neue Abfüllmaschine kaufen, um sicherzustellen, dass bei unveränderten Erwartungswert höchsten 4% aller Packungen eine Masse von mehr als 3350 Gramm besitzen. Welche Standardabweichung muss die neue Maschine haben?*

Lösung auf Seite 54

Aufgabe 7.3

Analysieren Sie die folgenden Fragestellungen mit Hilfe geeigneter statistischer Testverfahren. Als Irrtumswahrscheinlichkeit soll $\alpha = 0.05$ gewählt werden.

(a) *In einer Umfrage während einer Vorlesung gaben 55% der männlichen Studenten an, dass Sie in Sport die Note sehr gut hatten. Von 104 weiblichen Studenten hatten 35 die Note sehr gut. Lässt sich statistisch absichern, dass der Anteil unter den weiblichen Studenten mit der Note sehr gut kleiner als 55% ist?*

(b) *Was ändert sich in (a), wenn 52 von 104 weiblichen Studenten angaben, in Sport die Note sehr gut erhalten zu haben?*

(c) *In der letzten Klausur hatten 47% der männlichen Studenten die Note gut oder sehr gut. Bei den weiblichen Studenten waren es 64 von 106. Lässt sich statistisch absichern, dass der Anteil unter den weiblichen Studenten mit der Note gut oder sehr gut größer als 47% ist?*

(d) *In Deutsch hatten 50% der männlichen Studenten die Note gut oder sehr gut. Bei den weiblichen Studenten 68 von 106. Lässt sich statistisch absichern, dass der Anteil unter den weiblichen Studenten mit der Note gut oder sehr gut verschieden von 50% ist?*

Lösung auf Seite 55

Aufgabe 7.4

Eine Fahrschule behauptet, dass 25% aller Fahrschüler beim ersten Versuch durch die Fahrprüfung fallen. Berechnen Sie die Wahrscheinlichkeit, dass bei 25%-tiger Durchfallquote dann bei einer Gruppe von 50 Fahrschülern

(a) *höchstens 14 durchfallen,*

(b) *genau 37 durchkommen,*

(c) *höchsten 18 und mindestens 8 durchfallen.*

(d) *Wie groß ist die Irrtumswahrscheinlichkeit α, wenn obige Hypothese anhand von 50 zufällig ausgewählten Fahrschülern getestet wird? Die Hypothese wird dabei verworfen, wenn mehr als 16 Personen durchfallen.*

Lösung auf Seite 55

A Lösungen

Lösung von Aufgabe 2.1 auf Seite 9

- *X diskret $\Rightarrow$ Verteilungsfunktion F mit* $F(x) = P(X \leq x) = \sum\limits_{k=0}^{x} P(X = k)$
- *Wertebereich:* $W = \{0, 1, 2, \ldots\} \Rightarrow$ *für alle* $a, b \in \mathbb{Z}$ *gilt* $X \in (a, b) \Leftrightarrow X \in [a+1, b-1]$

(a) $P(4 \leq X \leq 8) = P(3 < X \leq 8) = F(8) - F(3)$ *denn* $P(4 \leq X \leq 8) = \sum\limits_{k=4}^{8} P(X = x) = \sum\limits_{k=0}^{8} P(X \leq x) - \sum\limits_{k=0}^{3} P(X \leq x) = F(8) - F(3)$

(b) $P(2 \leq X < 5) = P(1 < X \leq 4) = F(4) - F(1) = \sum\limits_{k=2}^{4} P(X = k)$

(c) $P(4 < X < 7) = P(4 < X \leq 6) = F(6) - F(4) = \sum\limits_{k=5}^{6} P(X = k)$

(d) $P(1 \leq X < 8) = P(0 < X \leq 7) = F(7) - F(0) = \sum\limits_{k=1}^{7} P(X = k)$

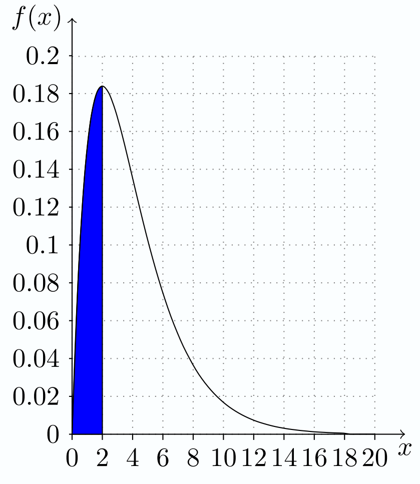
- *Folgerung: allgemein gilt* $P(a < X \leq b) = F(b) - F(a)$, $P(a \leq X \leq b) = F(b) - F(a-1)$, $P(a < X < b) = F(b-1) - F(a)$ *und* $P(a \leq X < b) = F(b-1) - F(a-1)$

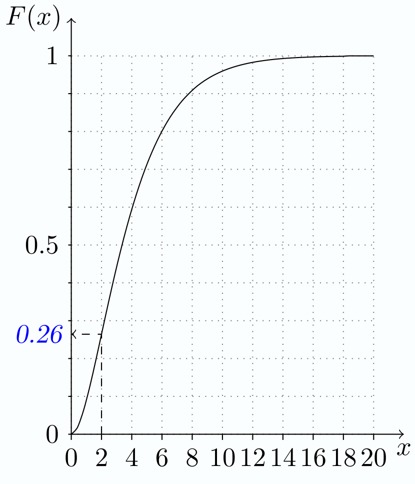
Lösung von Aufgabe 2.2 auf Seite 9

1. *Für* $x \leq 0$ *gilt* $F'(x) = 0 = f(x)$ *und für* $x > 0$ *gilt* $F'(x) = \ldots = \frac{x}{4}e^{-x/2}$. *Man könnte natürlich auch die Dichte partiell integrieren und erhält dann für* $x > 0$

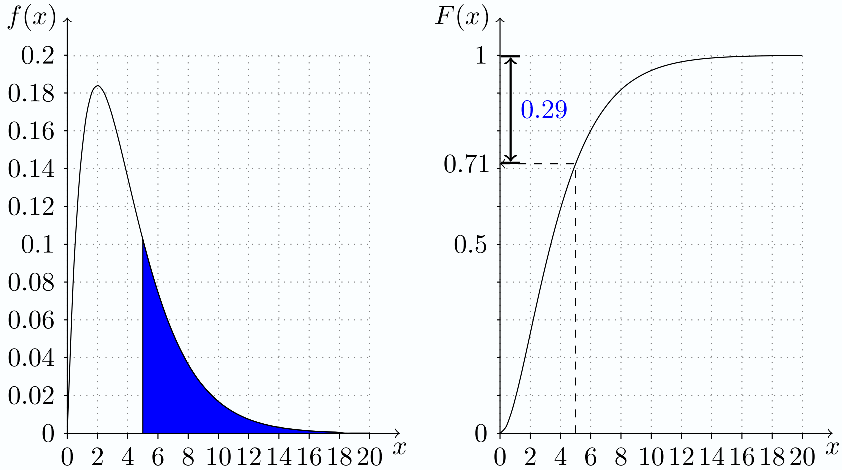
$$\int_{-\infty}^{x} f(t)\,dt = \frac{1}{4}\int_{0}^{x} \underbrace{t}_{u}\,\underbrace{e^{-t/2}}_{v'}\,dt = \frac{1}{4}\left(\left[-2te^{-t/2}\right]_0^x + 2\int_0^x e^{-t/2}\,dt\right)$$
$$= \frac{1}{4}\left(-2xe^{-x/2} - (2e^{-x/2} + 2)\right) = 1 - (1 + \frac{x}{2})e^{-x/2} = F(x)$$

2. - *exakt:* $P(X \leq 2) = F(2) = 0.2642$, *grafische Lösung:*

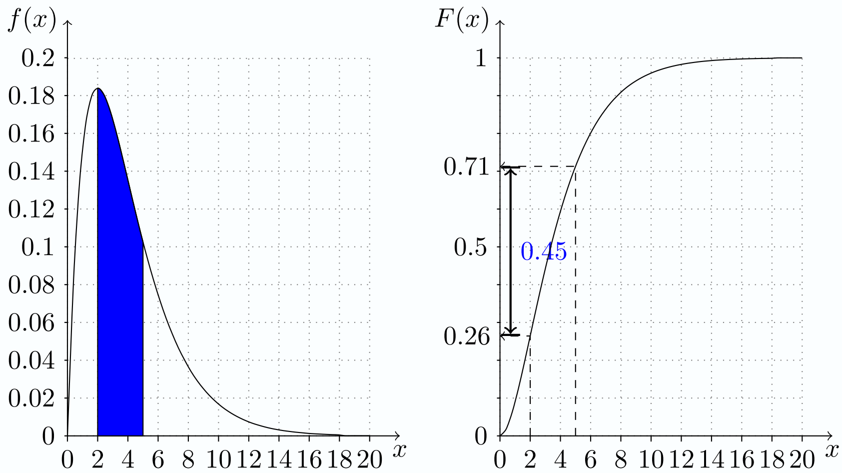




   - *exakt:* $P(X > 5) = 1 - F(5) = 0.2873$, *grafische Lösung:*

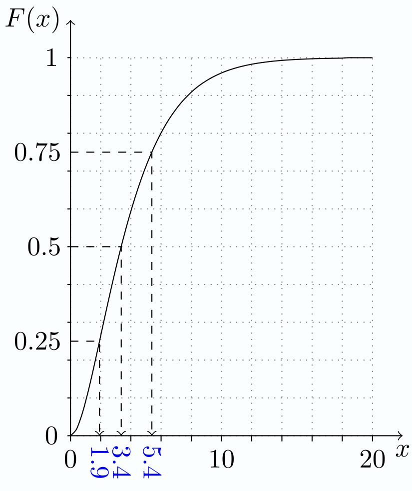


- *exakt:* $P(2 < X < 5) = F(5) - F(2) = 0.4485$, *grafische Lösung:*



3. ***Interpretation am Beispiel*** $\mathbf{p = 0.25}$***:*** *Die Fragestellung verlangt die Berechnung des q-Quantils. Das 25%-Quantil der Zufallsvariablen X beträgt* 1.92. *Für x muss der Wert* 1.92 *gewählt werden, damit $P(X \leq x) = 0.25$ gilt, d. h., es gilt $P(X \leq 1.92) = 0.25$.*

   grafische Lösung:



Lösung von Aufgabe 3.1 auf Seite 24

- *Zufallsgröße $X \triangleq$ Anzahl der Haushalte mit mind. 2 Geräten*
- $X \sim \mathrm{Bin}(n = 12, p = 1/4)$ *vgl*[a]*: Ereignis $A \triangleq$ Haushalt hat mind. 2 Fernsehgeräte mit $P(A) = p = 1/4$;*

$$X_i = \begin{cases} 1 & \textit{Ereignis } A \textit{ tritt in } i\textit{-ten Versuch ein, d.h. } i\textit{-te H. mind. 2 Fernseher} \\ 0 & \textit{Ereignis } A \textit{ tritt nicht ein in } i\textit{-ten Versuch, d.h. } i\textit{-te H. max. 1 Fernseher} \end{cases}$$

falls X_i $(i = 1, \dots, n)$ unabhängig, gilt $X = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p)$, *da* $X_i \sim \mathrm{Bin}(1, p)$

(a) *höchstens die Hälfte der Haushalte maximal ein Fernsehgerät besitzt:* $P(Y \leq 6) = P(X \geq 6) = \sum_{k=6}^{12} \binom{12}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{12-k} = 0.05440223$

(b) *mindestens die Hälfte der Haushalte mehr als ein Fernsehgerät besitzt: (siehe a)* $P(X \geq 6) = P(Y \leq 6) = 0.05440223$

(c) *weniger als zwei Drittel der Haushalte maximal ein Fernsehgerät besitzt:* $P(Y < 8) = P(X > 4) = 1 - P(X \leq 4) = 1 - F(4) = 0.1576437$

(d) *höchstens zwei Haushalte mehr als ein Fernsehgerät besitzen:* $P(X \leq 2) = 0.390675 = P(Y \geq 10)$

---

[a] Mit $Y_i = 1 - X_i \sim \mathrm{Bin}(1, 1-p)$ gilt für $Y = \sum_{i=1}^{n} Y_i = n - X \sim \mathrm{Bin}(n, 1-p)$ zählt die Anzahl der Hausahlte mit höchstens einem Gerät.

Lösung von Aufgabe 3.2 auf Seite 24

- *Zufallsgröße* $X \triangleq$ *Anzahl der richtig beantworteten Fragen beim MC Test mit 30 Aufgaben (mit jeweils einer richtigen Auswahl von fünf)*
- $X \sim \mathrm{Bin}(n = 30, p = 1/5)$ *da alle MC-Antworten unabhängig voneinander und zufällig beantwortet werden*

(a) $P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^{5} \binom{30}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{30-k} = 0.5725$

(b) $P(4 \leq X \leq 10) = P(3 < X \leq 10) = F(10) - F(3) = \sum_{k=4}^{10} \binom{30}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{30-k} = 0.9744 - 0.1227 = 0.8517$

(c) $P(X < 3) = P(X \leq 2) = F(2) = \sum_{k=0}^{2} \binom{30}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{30-k} = 0.04418$

(d) $P(X = 6) = \binom{30}{6} \left(\frac{1}{5}\right)^6 \left(\frac{4}{5}\right)^{24} = 0.1795$

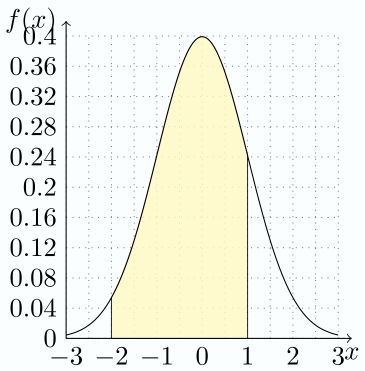
Lösung von Aufgabe 3.3 auf Seite 24

- *Zufallsgröße* $X \triangleq$ *Anzahl schadhafter Kondensatoren*
- *Verteilung:* $X \sim \mathrm{Bin}(n = 50, p = 0.15)$
- $\mathbb{E}X = np = 7.5 \triangleq$ *im Mittel sind 7.5 von 50 Kondensatoren schadhaft. Genauer: würden wir sehr oft 50 Kondensatoren der Lieferung entnehmen*[a]
- $\mathrm{Var}(X) = np(1-p) = 6.375 \triangleq$ *beschreibt Schwankung*

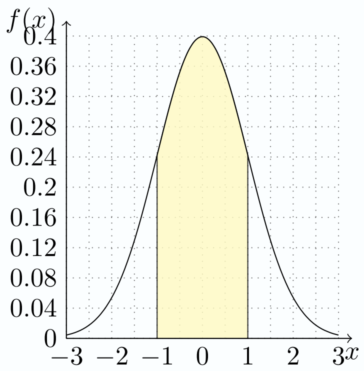
---

[a] Streng genommen müssten wir jeden Kondensator nach der Untersuchung wieder zurücklegen, so dass theoretisch ein Kondensator mehrfach untersucht werden könnte. Andernfalls liegt kein Versuchsaufbau wie beim Galtonbrett vor, bei dem jeder einzeln unteruschte Kondensator, unabhängig von den Untersucheungen der restlichen Kondensatoren, mit der Wahrscheinlichkeit $p = 0.15$ defekt ist. Dann würde eine hypergeometrische Verteilung vorliegen. Der Erwartungswert dieser hypergeometrischen Verteilung wäre aber identisch zur Binomialverteilung. Nur die Varianz würde sich im Vergleich zur Binomialverteilung verkleinern. Für sehr große Liefrungen wäre der Unterschied zwischen beiden Verteilungen übrigens kaum spürbar.

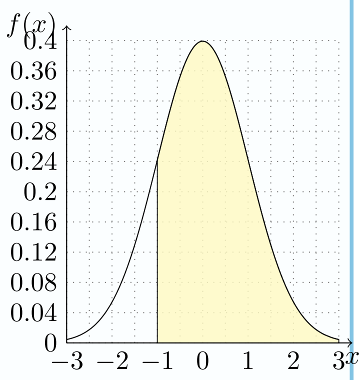
Lösung von Aufgabe 4.1 auf Seite 34



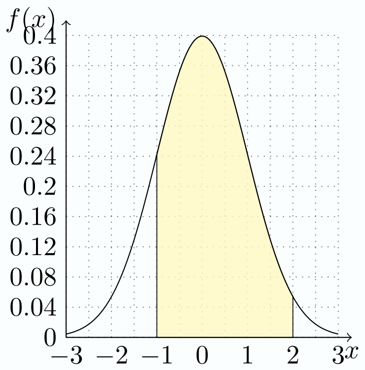
(a) $P(-2 < Z < 1) = \Phi(1) - \Phi(-2) = 0.8185946$



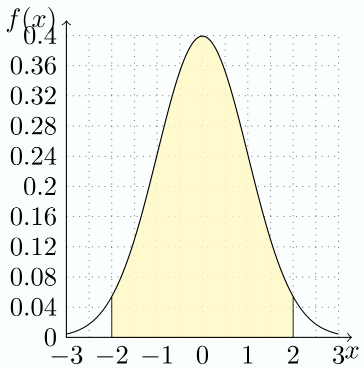
(c) $P(-1 < Z < 1) = 2\Phi(1) - 1 = 0.6826895$



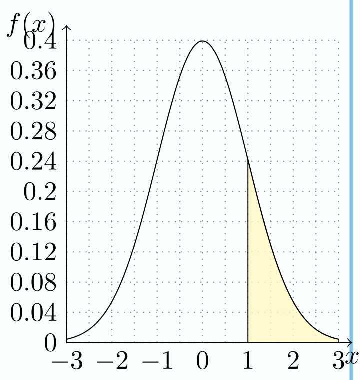
(e) $P(Z > -1) = 1 - \Phi(-1) = \Phi(1) = 0.8413447$



(b) $P(-1 < Z < 2) = \Phi(2) - \Phi(-1) = 0.8185946$



(d) $P(-2 < Z < 2) = 2\Phi(2) - 1 = 0.9544997$



(f) $P(Z > 1) = 1 - \Phi(1) = \Phi(-1) = 0.1586553$

Nebenrechnungen:

(a) $P(-2 < Z < 1) = \Phi(1) - \Phi(-2) = 0.8413447 - 0.02275013 = 0.8185946$
(b) $P(-1 < Z < 2) = \Phi(2) - \Phi(-1) = 0.9772499 - 0.1586553 = 0.8185946$
(c) $P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2 \cdot 0.8413447 - 1 = 0.6826895$
(d) $P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 2 \cdot 0.9772499 - 1 = 0.9544997$
(e) $P(Z > -1) = 1 - \Phi(-1) = \Phi(1) = 0.8413447$
(f) $P(Z > 1) = 1 - \Phi(1) = \Phi(-1) = 0.1586553$

Lösung von Aufgabe 4.2 auf Seite 34

- *Zufallsgröße* $X \triangleq$ *IQ*, $X \sim \mathcal{N}(\mu = 100, \sigma^2 = 15^2)$

(a) *gesucht* $x \in \mathbb{R}$ *mit* $P(X > x) = 0.02 \Leftrightarrow F(x) = P(X \leq x) = \Phi(\frac{x-100}{15}) = 0.98$ $\Rightarrow \Phi^{-1}(0.98) = 2.053749 = \frac{x-100}{15} \Rightarrow x = 130.8062$

(b) *Gesucht ist* $k \in \mathbb{R}_+$ *mit* $P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(-k \leq \frac{X-\mu}{\sigma} \leq +k) = 2\Phi(k) - 1 = 0.8 \Rightarrow k = \Phi^{-1}(0.9) = 1.281552$ *und somit* $\mu - k\sigma, \mu + k\sigma] = [80.77673, 119.22327]$ *(siehe auch Sigma-Regel der Normalverteilung für* $1 - \alpha = 0.8$ *– siehe Beispiel 4.2 auf Seite 31)*

(c) *gesucht* $x \in \mathbb{R}$ *mit* $P(X > x) = 0.85 \Leftrightarrow F(x) = P(X \leq x) = \Phi(\frac{x-100}{15}) = 0.15$ $\Rightarrow \Phi^{-1}(0.15) = -1.036433 = \frac{x-100}{15} \Rightarrow x = 84.4535$

Lösung von Aufgabe 4.3 auf Seite 34

- *Zufallsgröße* $X \triangleq$ *Punktanzahl beim Test*, $X \sim \mathcal{N}(\mu = 80, \sigma^2 = 12^2)$

(a) *gesucht* $x \in \mathbb{R}$ *mit* $P(X > x) = 0.1 \Leftrightarrow F(x) = P(X \leq x) = \Phi(\frac{x-80}{12}) = 0.9 \Rightarrow \Phi^{-1}(0.9) = 1.281552 = \frac{x-80}{12} \Rightarrow x = 95.37862$

(b) *Gesucht ist* $k \in \mathbb{R}_+$ *mit* $P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(-k \leq \frac{X-\mu}{\sigma} \leq +k) = 2\Phi(k) - 1 = 0.95 \Rightarrow k = \Phi^{-1}(0.975) = 1.959964$ *und somit* $\mu - k\sigma, \mu + k\sigma] = [56.48043, 103.51957]$ *(siehe auch Sigma-Regel der Normalverteilung für* $1 - \alpha = 0.95$ *– siehe Beispiel 4.2 auf Seite 31)*

(c) *gesucht* $x \in \mathbb{R}$ *mit* $P(X > x) = 0.95 \Leftrightarrow F(x) = P(X \leq x) = \Phi(\frac{x-80}{12}) = 0.05$

$\Rightarrow \Phi^{-1}(0.05) = -1.644854 = \frac{x-80}{12} \Rightarrow x = 60.26176$

(d) gesucht $x \in \mathbb{R}$ mit $P(X > x) = \frac{40}{2000} = 0.02 \Leftrightarrow F(x) = P(X \leq x) = \Phi(\frac{x-80}{12}) = 0.98$
$\Rightarrow \Phi^{-1}(0.98) = 2.053749 = \frac{x-80}{12} \Rightarrow x = 104.645$

Lösung von Aufgabe 6.1 auf Seite 46

- *Zufallsgröße* $X \triangleq$ *Anzahl der 5er und 6er*
- $X \sim \mathrm{Bin}(n, p = 1/3)$ *mit (a)* $n = 42$, *(b)* $n = 72$ *und (c)* $n = 114$
- *gesucht:* $P(X \in [\mu - \sigma, \mu + \sigma]) = \sum_{k=\mu-\sigma}^{\mu+\sigma} \binom{n}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{n-k} \approx$

(a) $n = 42 \Rightarrow \mathbb{E}X = 14$, $\sigma = \sqrt{var(X)} = 3.055050 \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(11 \leq X \leq 17) = P(10 < X \leq 17) = F(17) - F(10) = 0.7486701$
Normalapproxiamtion $X \approx \mathcal{N}(14, 9.333333) \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(11 \leq X \leq 17) \approx \Phi(\frac{17-\mu}{\sigma}) - \Phi(\frac{11-\mu}{\sigma}) = 0.6738905$
(Normalapproximation mit Stetigkeitskorrektur $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(11 \leq X \leq 17) \approx \Phi(\frac{17.5-\mu}{\sigma}) - \Phi(\frac{10.5-\mu}{\sigma}) = 0.7480575$

(b) $n = 72 \Rightarrow \mathbb{E}X = 24$, $\sigma = \sqrt{var(X)} = 4 \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(20 \leq X \leq 28) = P(19 < X \leq 28) = F(28) - F(19) = 0.7397422$
Normalapproxiamtion $X \approx \mathcal{N}(24, 16) \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(20 \leq X \leq 28) \approx \Phi(\frac{28-\mu}{\sigma}) - \Phi(\frac{20-\mu}{\sigma}) = 0.6826895$
(Normalapproximation mit Stetigkeitskorrektur $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(20 \leq X \leq 28) \approx \Phi(\frac{28.5-\mu}{\sigma}) - \Phi(\frac{19.5-\mu}{\sigma}) = 0.7394110$

(c) $n = 114 \Rightarrow \mathbb{E}X = 38$, $\sigma = \sqrt{var(X)} = 5.033223 \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(33 \leq X \leq 43) = P(32 < X \leq 43) = F(43) - F(33) = 0.7256700$
Normalapproxiamtion $X \approx \mathcal{N}(38, 25.333333) \Rightarrow P(\mu - \sigma \leq X \leq \mu + \sigma) = P(33 \leq X \leq 43) \approx \Phi(\frac{43-\mu}{\sigma}) - \Phi(\frac{33-\mu}{\sigma}) = 0.6794846$
(Normalapproximation mit Stetigkeitskorrektur $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(33 \leq X \leq 43) \approx \Phi(\frac{43.5-\mu}{\sigma}) - \Phi(\frac{32.5-\mu}{\sigma}) = 0.7254917$

- *Fazit: 1. Normalapproximation liefert nach Faustregel des Grenzwertsatzes von Moivre / Laplace* ($\mathrm{Var}(X) = np(1-p) > 9$) *brauchbare Näherungen. Mit Stetigkeitskorrektur*

$$P(a \leq X \leq b) = P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right)$$
$$\approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

deutlich bessere Ergebnisse
2. vgl. Sigma-Regel der Normalverteilung für $k = 1$ ($Y \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow P(\mu - \sigma \leq Y \leq \mu + \sigma) = 0.6827$), *siehe Beispiel 4.2 auf Seite 31*

Lösung von Aufgabe 6.2 auf Seite 47

- *Zufallsgröße* $X \triangleq$ *Anzahl auffrisierter Mopeds*
- $X \sim \mathrm{Bin}(n, p = 0.4)$ *mit (a)* $n = 120$ *und (b)* $n = 100$

(a) $n = 120$ *gesucht* $P(X < 40) = P(X \leq 39) = F(39) = \sum_{k=0}^{39} \binom{120}{k} 0.4^k 0.6^{120-k} = 0.05531556$
Normalapproxiamtion: $\mathbb{E}X = np = 48$, $\sigma = \sqrt{var(X)} = \sqrt{28.8} = 5.366563 \Rightarrow X \approx \mathcal{N}(48, 5.366563) \Rightarrow P(X \leq 39) \approx \Phi(\frac{39-\mu}{\sigma}) = 0.04676626$ *(mit Stetigkeitskorrektur:* $P(X \leq 39) = P(X < 39 + 1/2) \approx \Phi(\frac{39.5-\mu}{\sigma})) = 0.05661035$

(b) $n = 100 \Rightarrow \mathbb{E}X = np = 40$ *gesucht* $P(X - \mu \geq 5) = P(X \geq 45) = 1 - P(X < 45) = 1 - P(X \leq 44) = 1 - F(44) = 1 - \sum_{k=0}^{44} \binom{100}{k} 0.4^k 0.6^{120-k} = \sum_{k=45}^{100} \binom{100}{k} 0.4^k 0.6^{120-k} = 0.1789016$
Normalapproxiamtion: $\mathbb{E}X = np = 48$, $\sigma = \sqrt{var(X)} = \sqrt{24} = 4.898979 \Rightarrow X \approx \mathcal{N}(40, 4.898979) \Rightarrow 1 - P(X \leq 44) = 1 - P(X < 44 + 1/2) \approx 1 - \Phi(\frac{44-\mu}{\sigma}) = 0.2071081$
(mit Stetigkeitskorrektur: $1 - P(X \leq 44) = 1 - P(X < 44 + 1/2) \approx 1 - \Phi(\frac{44.5-\mu}{\sigma})) = 0.1791632$

Lösung von Aufgabe 6.3 auf Seite 47

- *Zufallsgröße* $X \triangleq$ *Anzahl gekaufter Computer der Marke „Digitoll"*
- $X \sim \mathrm{Bin}(n, p = 0.11)$ *mit (a)* $n = 400$ *und (b)* $n = 8000$

(a) $n = 400$ *gesucht* $P(X > 50) = 1 - P(X \leq 50) = 1 - F(50) = 1 - \sum_{k=0}^{50} \binom{400}{k} 0.11^k 0.89^{400-k} = 0.1496761$
Normalapproxiamtion: $\mathbb{E}X = np = 44, \sigma = \sqrt{var(X)} = 6.257795 \Rightarrow X \approx \mathcal{N}(44, 39.16) \Rightarrow P(X > 50) = 1 - P(X \leq 50) = 1 - F(50) \approx 1 - \Phi(\frac{50-\mu}{\sigma}) = 1 - \Phi(0.9588042) = 0.1688287$
bzw. Normalapproximation mit Stetigkeitskorrektur liefert $P(X > 50) = 1 - P(X \leq 50) = 1 - P(X \leq 50.5) = 1 - F(50.5) \approx 1 - \Phi(\frac{50.5-\mu}{\sigma}) = 1 - \Phi(1.038705) = 0.1494711$

(b) $n = 8000 \Rightarrow \mathbb{E}X = 880,\ \mathrm{Var}(X) = 738.2 \rightarrow X \approx \mathcal{N}(880, 783.2)$
mit der Sigma-Regel der Normalverteilung für $p = 1 - \alpha = 0.95$ *(*$Y \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow P(\mu - k\sigma \leq Y \leq \mu + k\sigma) = p = 1 - \alpha \Rightarrow k = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) = 1.959964 \Rightarrow (\mu - k \cdot \sigma,\ \mu + k \cdot \sigma) = (825.149,\ 934.851) \rightarrow P(Y \in (825.149,\ 934.851)) = 0.95$*), siehe Beispiel 4.2 auf Seite 31*
Mit Statistiksoftware kann man auch die Quantile der Ordnung $\alpha/2 = 0.025$ *sowie* $1 - \alpha/2 = 0.975$ *der Binomialverteilung exakt bestimmen: es lässt sich leicht überprüfen, dass*
1. $F_X^{-1}(0.025) = \min\{x : F_X(x) \geq 0.025\} = 826$, *denn* $F_X(826) = 0.0271508 \geq 0.025$, *aber* $F_X(825) = 0.02494992 < 0.025$
2. $F_X^{-1}(0.975) = 935$, *denn* $F_X(935) = 0.9755782 \geq 0.975$, *aber* $F_X(934) = 0.9734957 < 0.975$
3. *folglich gilt* $P(826 < X \leq 935) = F_X(935) - F_X(826) = 0.9484274 \approx 0.95$, *das Intervall* $(826,\ 935)$ *liegt (fast) symmetrisch um* $\mathbb{E}X = 880$

- *wegen der Stufenform der Verteilungsfunktion (diskrete Zufallsvariable* $X \sim \mathrm{Bin}(880, 0.11)$ *ist die Berechnung mit Hilfe der Binomialverteilung auch mit Statistiksoftware (deutlich) komplizierter als mit der Normalapproximation; klassische Berechnung mit Taschenrechner und Formelsammlung kann man in vertretbaren Aufwand nur mittels Normalapproximation lösen*

Lösung von Aufgabe 7.1 auf Seite 48

- $X \triangleq$ *Anzahl tatsächlicher Fluggäste* $\sim \mathrm{Bin}(n = 120, p = 0.85)$

(a) *Flug ist überbucht:* $P(X > 100) = P(X \geq 101) = \sum_{k=101}^{120} \binom{120}{k} (0.85)^k (0.15)^{120-k} =$

(b) *to do*

(c) *to do*

Lösung von Aufgabe 7.2 auf Seite 48

- *Zufallsgröße* $X \triangleq$ *Masse, geg.* $X \sim \mathcal{N}(\mu, \sigma^2)$ *mit* $\mu = 3200$ *und* $\sigma = 100$

(a) $P(X < 3150) = P(X \leq 3150) = F(3150) = \Phi\left(\frac{3150-3200}{100}\right) = \Phi(-0.5) = 0.3085375$

(b) *gesucht* c *mit* $P(|X - \mu| \leq c) \leq 0.09$, *berechnen hierfür*

$$0.09 \stackrel{!}{=} P(|X - \mu| \leq c) = P(-c \leq X - \mu \leq c) = P\left(-\frac{c}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{c}{\sigma}\right)$$
$$= \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(-\frac{c}{\sigma}\right) = \Phi\left(\frac{c}{\sigma}\right) - \left(1 - \Phi\left(\frac{c}{\sigma}\right)\right) = 2\Phi\left(\frac{c}{\sigma}\right) - 1$$

$\Rightarrow \Phi\left(\frac{c}{\sigma}\right) = 1.09/2 = 0.545 \Rightarrow \frac{c}{\sigma} = \Phi^{-1}(0.545) = 0.1130385 \Rightarrow c = 11.303855$

(c) *gesucht ist* σ *mit* $P(X > 3350) \leq 0.04$, *berechnen hierfür*

$$0.04 \stackrel{!}{=} P(X > 3350) = 1 - P(X \leq 3350) = 1 - P\left(X \leq \frac{3350 - 3500}{\sigma}\right)$$

$\Rightarrow 0.96 = \Phi\left(\frac{150}{\sigma}\right) \Rightarrow \Phi^{-1}(0.96) = 1.750686 \Rightarrow \sigma = \frac{150}{\Phi^{-1}(0.96)} = 85.6807$

Lösung von Aufgabe 7.3 auf Seite 48

Binomialtest: Es wird auf den Anteilwert p getestet, da p unbekannt ist.

(a)
- $n = 104$, *Punktschätzer (relative Häufigkeit):* $\hat{p} = 35/104 = 0.3365$. *Da n groß (vgl. Faustregel Satz von Moivre-Laplace $np(1-p) > 9$ für gute Approximation: hier $np(1-p) = 25.74$ ✓) Normalapproximation (alternative Berechnung mit Statistiksoftware auch direkt für Binomialverteilung möglich)*
- $H_0 : p = p_0 = 0.55$, $H_1 : p < p_0 = 0.55$, *standardisierte Testgröße* $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.3365-0.55}{\sqrt{0.55\cdot 0.45}}\sqrt{104} = -4.3757$, *kritischer Bereich:* $K^\star = \{z \in \mathbb{R} : z < -z_{1-\alpha} = -z_{0.95} = -1.6449\}$ $\Rightarrow K^\star = (-\infty, -1.6449)$ $\Rightarrow z \in K^\star$, *$H_0$ wird abgelehnt, H_1 wird angenommen, d.h. es lässt sich statistisch absichern, dass der Anteil der weiblichen Studenten mit der Note sehr gut in Sport kleiner als 55% ist.*

(b)
- *Hypothesen wie in (a), Testgröße* $z = -1.02$, *kritischer Bereich wie in (a), H_0 wird diesmal angenommen* $\Rightarrow$ *kein Nachweis!*

(c)
- *Voraussetzung* $np(1-p) = 26.4046$ ✓
- $H_0 : p = p_0 = 0.47$, $H_1 : p > p_0 = 0.47$, $T = 2.76$, *kritischer Bereich:* $T > z_{1-\alpha} = z_{0.95} = 1.6449$ $\Rightarrow K_* = (1.6449, \infty)$ $\Rightarrow T \in K_*$, *H_0 wird abgelehnt, H_1 wird angenommen*

(d)
- *Voraussetzung* $np(1-p) = 26.5$ ✓
- $H_0 : p = p_0 = 0.5$, $H_1 : p \neq p_0 = 0.5$, $T = 2.914$, *kritischer Bereich:* $K^\star = \{z \in \mathbb{R} : |z| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$ $\Rightarrow K^\star = (-\infty, -1.96) \cup (1.96, \infty)$ $\Rightarrow z \in K^\star$, *$H_0$ wird abgelehnt, H_1 wird angenommen*

Lösung von Aufgabe 7.4 auf Seite 48

- *Zufallsgröße* $X_i = \begin{cases} 1 & \text{Fahrschüler } i \text{ fällt durch} \\ 0 & \text{... besteht ...} \end{cases} \sim \mathrm{Bin}(1, p)$ *mit unbekannten Parameter $p \triangleq$ wahre Durchfallquote*
- *rechtsseitiger Binomialtest:* $H_0 : p = p_0 = 0.25$ *gegen* $H_1 : p > p_0 = 0.25$, *da* $K^\star = t : t \geq 16$ *gegeben*
- *Testgröße unter H_0 ist binomialverteilt* $T = \sum_{50} X_i \sim \mathrm{Bin}(50, p_0) \triangleq$ *Anzahl Durchfaller*
- *mit Binomialverteilung* $T \sim \mathrm{Bin}(50, p_0)$ *folgt*
  (a) $P(T \leq 14) = F(14) = \sum_{k=0}^{14} \binom{50}{k} 0.25^k 0.75^{50-k} = 0.7481$
  (b) $P(T = 13) = \binom{50}{13} 0.25^{13} 0.75^{37} = 0.1261$
  (c) $P(8 \leq T \leq 18) = F(18) - F(7) = \sum_{k=8}^{18} \binom{50}{k} 0.25^k 0.75^{50-k} = 0.926$

(d) *kritischer Bereich des Tests:* $K^\star = \{t : t \geq 16\} \Rightarrow \alpha = P(T \in K^\star) = P(T \geq 16) = 1 - P(T \leq 15) = 1 - F(15) = 0.1631$

B Exkurs deskriptive Statistik: empirische Auswertung der Daten einer Stichprobe

Das Ziel der deskriptiven Statistik ist es, eine große Datenmenge anschaulich Mithilfe von Tabellen, Grafiken und statistischen Kennzahlen zu beurteilen und zu beschreiben. Wir besprechen hier nur den univariaten, d.h. eindimensionalen Fall und werten ein Merkmal (Variable) X , welches wiederholt (unabhängig) beobachtet bzw. gemessen und dessen Ergebnisse in der Stichprobe (SP) $(x_1, x_2, \ldots, x_n)$ zusammengefasst wurde. Da diese in der Regel sehr unübersichtlich für große Stichproben ist, versuchen wir die Informationen zu komprimieren. Werden die Daten auf einer Ordinalskala[12] oder höher gemessen, können wir zur **geordneten Stichprobe** $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ mit $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ übergehen und bezeichnen die beobachteten (geordneten) **Ausprägungen** der Stichprobe mit $a_1, a_2, \ldots a_k$, d.h. $a_1 < a_2 < \ldots < a_k$ und $k \leq n$.

Da jede tatsächlich beobachtete Stichprobe natürlich nur aus endlichen vielen Beobachtungen ($n \in \mathbb{N}$) bestehen kann, sind die Daten der Stichprobe prinzipiell endlich und somit abzählbar, also diskret. Allerdings unterscheidet man je nach Art der möglichen Ausprägungen diskrete und stetige

[12] Alle Ausprägungen lassen sich also sinnvoll sortieren. Für rein qualitative Merkmale, die nur auf einer Nominalskala gemessen werden, wie beispielsweise Farben oder Marken oder Hersteller ist dies nicht möglich.

Merkmale, je nachdem ob der Raum aller (theoretisch) Merkmalsausprägungen abzählbar (diskrete Variablen) oder überabzählbar[13] (stetige Variablen) ist, da dies unmittelbare Auswirkung auf die Auswertung der Häufigkeiten und deren grafische Darstellung hat. Im Folgenden lassen wir die Einheiten bei der Auflistung der Stichprobenwerte der Einfachheit halber weg.

Beispiel B.1: verschiedene Ausprägungen eines stetigen Merkmals

- *Restaurant bietet am Steak-Tag ein Gericht mit einem 100 Gramm Steak zum Preis von 9.55€ an $\curvearrowright$ Frage: wird das Gewicht eingehalten?*
- *7 Messungen: SP* $(100.9, 98.7, 100.6, 98.9, 100.0, 99.8, 99.2)$
- *geordnete SP* $(98.7, 98.9, 99.2, 99.8, 100.0, 100.6, 100.9)$
- *Ausprägungen:* $a_1 = x_{(1)} = 98.7 = x_2$, $a_2 = x_{(2)} = 98.9 = x_4$ *usw.*

Beispiel B.2: vermehrt auftretende Ausprägungen eines diskreten Merkmals

- *Noten von 10 Schülern: SP* $(2, 4, 5, 2, 3, 4, 2, 6, 5, 4)$
- *geordnete SP:* $(2, 2, 2, 3, 4, 4, 4, 5, 5, 6)$
- *Merkmalsausprägungen:* $a_1 = 2$, $a_2 = 3$, $a_3 = 4$, $a_4 = 5$, $a_5 = 6$

B.1 Häufigkeitsverteilung und deren graphische Darstellung

Das folgende Beispiel illustriert die Häufigkeitsanalyse, bei der wir einfach zählen, wie oft jede Ausprägung in einer Stichprobe auftritt.

Beispiel B.3: Häufigkeitsverteilung ohne Formeln

- *Stichprobe* $(3, 9, 10, 4, 8, 1, 6, 5, 9, 9, 4, 3, 7, 4, 9)$, $n = 15$
- *geordnete Stichprobe* $(1, 3, 3, 4, 4, 4, 5, 6, 7, 8, 9, 9, 9, 9, 10)$
- *Merkmalsausprägungen*
  - $a_1 = 1$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_1 = 1$, *relative H.* $f_1 = 1/15 = \frac{h_1}{n}$
  - $a_2 = 3$, *Häufigkeit in SP:* $2 \Rightarrow$ *absolute H.* $h_2 = 2$, *relative H.* $f_2 = 2/15 = \frac{h_2}{n}$
  - $a_3 = 4$, *Häufigkeit in SP:* $3 \Rightarrow$ *absolute H.* $h_3 = 3$, *relative H.* $f_3 = 3/15 = \frac{h_3}{n}$
  - $a_4 = 5$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_4 = 1$, *relative H.* $f_4 = 1/15 = \frac{h_4}{n}$
  - $a_5 = 6$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_5 = 1$, *relative H.* $f_5 = 1/15 = \frac{h_5}{n}$
  - $a_6 = 7$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_6 = 1$, *relative H.* $f_6 = 1/15 = \frac{h_6}{n}$
  - $a_7 = 8$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_7 = 1$, *relative H.* $f_7 = 1/15 = \frac{h_7}{n}$
  - $a_8 = 9$, *Häufigkeit in SP:* $4 \Rightarrow$ *absolute H.* $h_8 = 4$, *relative H.* $f_8 = 4/15 = \frac{h_8}{n}$
  - $a_9 = 10$, *Häufigkeit in SP:* $1 \Rightarrow$ *absolute H.* $h_9 = 1$, *relative H.* $f_9 = 1/15 = \frac{h_9}{n}$
- *Beobachtung*
  1. *Summe aller absoluten H.:* $h_1 + h_2 + \ldots + h_9 = 15$ *allgemein* $\sum_{j=1}^{k} h_j = n$
  2. *Summe aller relativen H.:* $f_1 + f_2 + \ldots + f_9 = 1$ *allgemein* $\sum_{j=1}^{k} f_j = 1$

Definition B.1: absolute und relative Häufigkeiten

- ***absolute Häufigkeit*** *der Merkmalsausprägung* a_j $(j = 1, \ldots, k)$*: Anzahl in der Stichprobe, d.h.*

$$h_j = h(a_j) = \sum_{i=1}^{n} \mathbf{1}_{\{a_j\}}(x_i) \tag{37}$$

- ***relative Häufigkeit*** *von* a_j*:*

$$f_j = f(a_j) = \frac{h(a_j)}{n} \tag{38}$$

[13] Beim Messen des Gewichts verschiedener Steaks sind beispielsweise alle Werte innerhalb eines Intervalls von $(0, b)$ möglich und wenn wir eine beliebig feine Waage haben auf der wir für ein Steak $x_i = 100.1$g und ein weiteres $x_j = 100.2$g messen können, dann ist natürlich ein weiteres Steak mit jedem beliebigen Wert dazwischen, also $x_k \in (100.1\text{g}, 100.2\text{g})$ möglich, was unmittelbar auf überabzählbar unendlich viele mögliche Ausprägungen hinausläuft.

Bemerkung B.1: Eigenschaften von Häufigkeiten

1. *absolute Häufigkeiten summieren sich zum Stichprobenumfang:*

$$\sum_{j=1}^{k} h(a_j) = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} \mathbf{1}_{\{a_j\}}(x_i) \right) \tag{39}$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n} \mathbf{1}_{\{a_j = x_i\}} = \sum_{i=1}^{n} \underbrace{\sum_{j=1}^{k} \mathbf{1}_{\{x_i = a_j\}}}_{=1} = n \tag{40}$$

denn jeder Stichprobenwert x_i *entspricht genau einer der Ausprägungen* $a_1 < a_2 < \ldots < a_k$

2. *Folglich ist die Summer der relativen Häufigkeiten immer 1*

$$\sum_{j=1}^{k} f(a_j) = \sum_{j=1}^{k} \frac{h_j}{n} = \frac{\sum_{j=1}^{k} h_j}{n} = 1 \tag{41}$$

und somit wobei $0 \leq f_j \leq 1$

3. *relative Häufigkeiten hängen eng mit Wahrscheinlichkeiten zusammen*

Beispiel B.4: Fortsetzung von Beispiel B.1 Steakgewicht

- *Stichprobe* $(100.9, 98.7, 100.6, 98.9, 99.8, 99.2)$
- *geordnete SP:* $(98.7, 98.9, 99.2, 99.8, 100.6, 100.9)$
- *jede Ausprägung kommt genau einmal vor:* $h_i = 1$ *und* $f_i = \frac{1}{6}$
- *Umsetzung mit Statistiksoftware R*

```
steaks = c(98.7,  98.9,  99.2,  99.8, 100.0, 100.6, 100.9)
#sortierte Stichprobe
sort(steaks)
## [1]  98.7  98.9  99.2  99.8 100.0 100.6 100.9
n=length(steaks)

#absolute Haeufigkeiten
table(steaks)
## steaks
##  98.7  98.9  99.2  99.8   100 100.6 100.9
##     1     1     1     1     1     1     1
#relative Haeufigkeiten
table(steaks)/n
## steaks
##   98.7   98.9   99.2   99.8    100  100.6  100.9
## 0.1429 0.1429 0.1429 0.1429 0.1429 0.1429 0.1429
#kumulierte rel. Haeufigkeiten
cumsum(table(steaks)/n)
##   98.7   98.9   99.2   99.8    100  100.6  100.9
## 0.1429 0.2857 0.4286 0.5714 0.7143 0.8571 1.0000
```

Beispiel B.5: Fortsetzung Beispiel B.2 Noten

- *SP* $(2, 4, 5, 2, 3, 4, 2, 6, 5, 4)$
- *sortierte SP:* $(2, 2, 2, 3, 4, 4, 4, 5, 5, 6)$
- *Häuigkeiten*

i	1	2	3	4	5
a_i	2	3	4	5	6
h_i	3	1	3	2	1
f_i	0.3	0.1	0.3	0.2	0.1

- *Umsetzung in R:*

```r
#beispiel: Noten, diskrete Daten
noten = c(2, 4, 5, 2, 3, 4, 2, 6, 5, 4)
#Stichprobenlaenge
(n.noten = length(noten))
## [1] 10
#sortierte SP
sort(noten)
##  [1] 2 2 2 3 4 4 4 5 5 6
#Haeufigkeiten
table(noten)
## noten
## 2 3 4 5 6
## 3 1 3 2 1
#relative Haeufigkeiten
table(noten)/n.noten
## noten
##   2   3   4   5   6
## 0.3 0.1 0.3 0.2 0.1
#kumulierte rel. Haeufigkeiten
cumsum(table(noten)/n.noten)
##   2   3   4   5   6
## 0.3 0.4 0.7 0.9 1.0
```

Bemerkung B.2: Fazit Häufigkeitstabellen

- *Darstellung absoluter und relativer Häufigkeiten*
- *bei großen Stichproben nur sinnvoll bzw. übersichtlich, wenn $k \ll n$, was typisch für* ***diskrete*** *Merkmale ist und dann entsteht durch Häufigkeitstabelle hohe Datenkomprimierung*
- *für* ***stetige*** *Merkmale ist die Darstellung absoluter und relativer Häufigkeiten oft nicht sinnvoll, da häufig $k \approx n$ und somit $h(a_j) \approx 1$ bzw. $f(a_j) \approx \frac{1}{n}$ gelten*
- *für stetige Merkmale bietet die Häufigkeitstabelle daher nur eine geringe Datenkomprimierung*
- *eine Möglichkeit Häufigkeiten stetiger Merkmale übersichtlich darzustellen ist die Gruppierung von Daten und Berechnung und Darstellung von Gruppenhäufigkeiten*

Beispiel B.6: US economics (economics)

In R sind viele Datensätze implementiert und wir betrachten im Folgenden den economics *Datensatz, der die folgenden Variablen / Merkmale zur US-Wirtschaft enthält*

pce *personal consumption expenditures (in billions of dollars)*
pop *total population (in thousands)*
psavert *personal savings rate*
uempmed *median duration of unemployment, in weeks*
unemploy *number of unemployed in thousands*

```r
library(tidyverse)
economics %>% tail(n = 8)
## # A tibble: 8 x 6
##   date          pce     pop psavert uempmed unemploy
##   <date>      <dbl>   <dbl>   <dbl>   <dbl>    <dbl>
## 1 2014-09-01 11957. 319125.     7.4    13.4     9262
## 2 2014-10-01 12023  319354.     7.2    13.6     8990
## 3 2014-11-01 12051. 319564.     7.3    13       9090
## 4 2014-12-01 12062  319746.     7.6    12.9     8717
## 5 2015-01-01 12046  319929.     7.7    13.2     8903
## 6 2015-02-01 12082. 320075.     7.9    12.9     8610
## 7 2015-03-01 12158. 320231.     7.4    12       8504
## 8 2015-04-01 12194. 320402.     7.6    11.5     8526
```

Um die Häufigkeiten der Konsumausgaben (stetiges Merkmal pce*) zu analysieren, filtern wir der Überischtlichkeit hier nur die Daten des Jahres 2014*

```
economics  %>% filter(lubridate::year(date)==2014) %>%
  group_by(pce) %>%
  summarise(h = n()) %>% mutate(freq = h /sum(h))
## # A tibble: 12 x 3
##        pce     h   freq
##      <dbl> <int>  <dbl>
##   1 11512.     1 0.0833
##   2 11566.     1 0.0833
##   3 11643      1 0.0833
##   4 11703.     1 0.0833
##   5 11748.     1 0.0833
##   6 11817      1 0.0833
##   7 11860.     1 0.0833
##   8 11944.     1 0.0833
##   9 11957.     1 0.0833
##  10 12023      1 0.0833
##  11 12051.     1 0.0833
##  12 12062      1 0.0833
```

Ähnliches gilt für die anderen stetigen Variablen, beispielsweise die Sparrate (psavert, Bezeichnung der Stichprobenwerte mit x_i). Betrachten wir hingegen die Häufigkeit des klassierten Merkmals class $\triangleq y_i$ *mit*

$$y_i = \begin{cases} 1 \triangleq \text{ (small)}, & x_i \leq 6 \\ 2 \triangleq \text{ (medium)}, & 6 < x_i \leq 12 \\ 3 \triangleq \text{ (high)}, & x_i > 12 \end{cases} \tag{42}$$

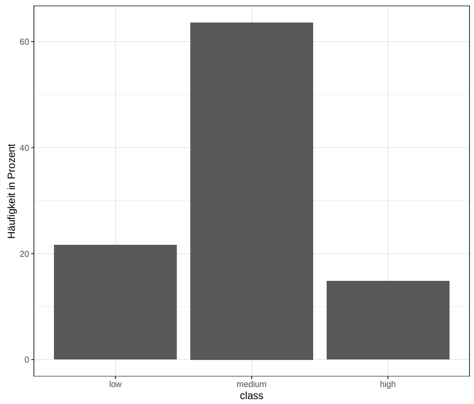
erhalten wir folgende Häufigkeiten

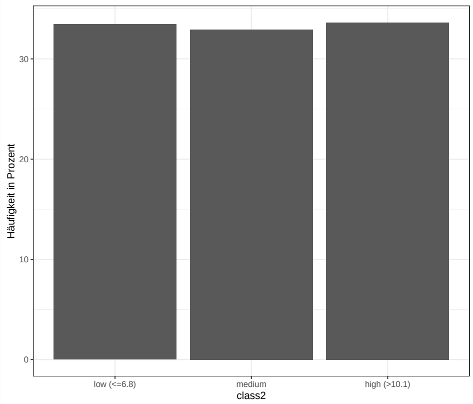
```
## # A tibble: 3 x 3
##   class      h  freq
##   <fct>  <int> <dbl>
## 1 low      124 0.216
## 2 medium   365 0.636
## 3 high      85 0.148
## # A tibble: 3 x 3
##   class2           h  freq
##   <fct>        <int> <dbl>
## 1 low (<=6.8)    192 0.334
## 2 medium         189 0.329
## 3 high (>10.1)   193 0.336
```

Um zu illustrieren, dass die Klassenaufteilung willkürlich ist, wurde oben zusätzlich die Variable class2 $\triangleq \tilde{y}_i$ *mit*

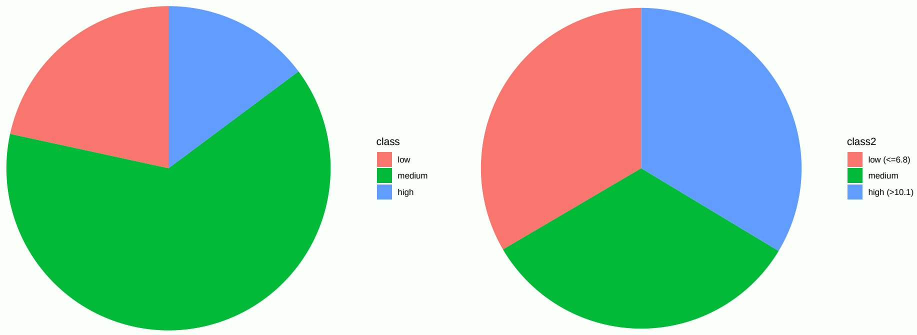
$$\tilde{y}_i = \begin{cases} 1 \triangleq \text{ (small)}, & x_i \leq 6.8 \\ 2 \triangleq \text{ (medium)}, & 6.8 < x_i \leq 10.1 \\ 3 \triangleq \text{ (high)}, & x_i > 10.1 \end{cases} \tag{43}$$

klassiert. Grafische Darstellung der Häufigkeiten beider klassierter Variablen mittels Balkendiagramm

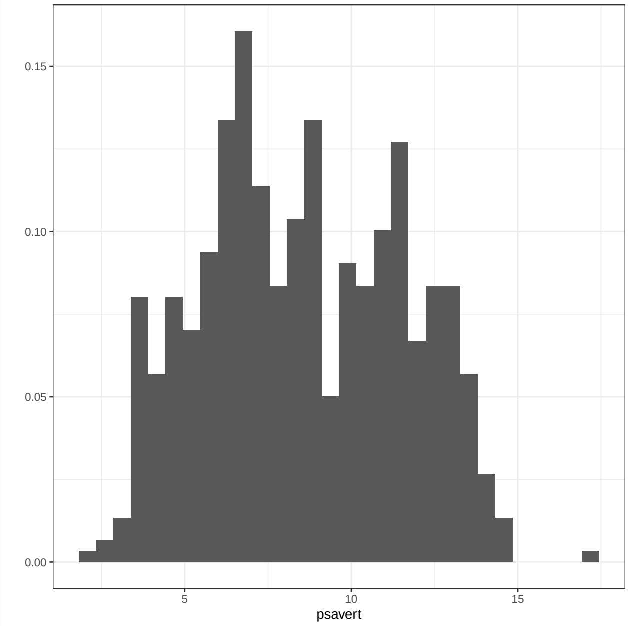




sowie mit Hilfe von Kreisdiagrammen



Da die Klassenaufteilung maßgeblich die Häufigkeiten beeinflusst, sollte diese entsprechend entsprechend der inhaltlichen Fragestellung vorsichtig gewählt werden. Um die Häufigkeiten stetiger Merkmale direkt darzustellen, verwendet man daher Histogramme, welche in Abschnitt B.1.4 nochmals näher beschrieben werden.



B.1.1 Anmerkungen zur grafischen Darstellung von Häufigkeiten

Allgemeines zu Grafiken:

- übersichtliche Darstellung von Datenstrukturen „auf einen Blick“
- **Achtung:** Gefahr der Fehlinterpretation durch weglassen wichtiger Informationen (z.B. Achsenskalierung)
- Diagrammtypen:
  - Stab- oder Balkendiagramme
  - Kreisdiagramme
  - Histogramme
  - Stamm-Blatt-Diagramme ↷ Literatur
- Einsatz abhängig vom Variablentyp und Skala

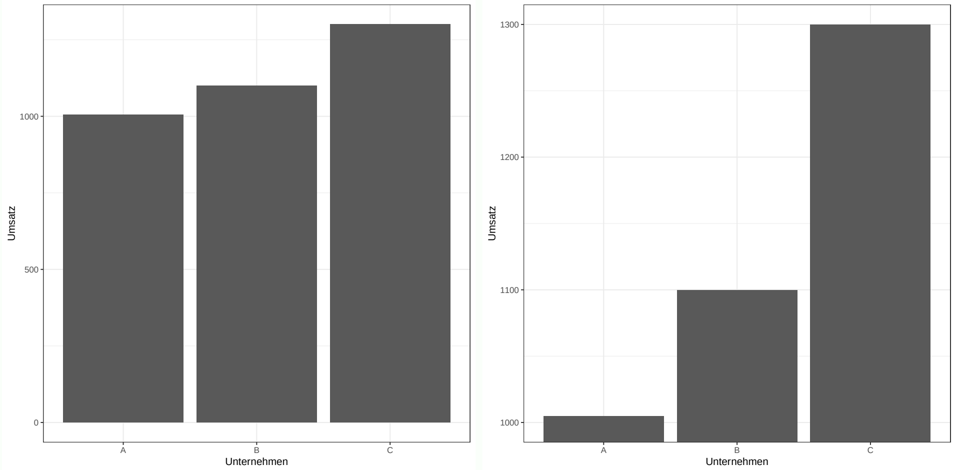
B.1.2 Stab- oder Balkendiagramme

- einfachste grafische Darstellungsmöglichkeit
- Höhe der Balken / Stäbe proportional zu Häufigkeiten
- Breite der Balken immer konstant, so dass die Flächeninhalte der Balken ebenfalls proportional zu Häufigkeiten sind

- sinnvoll für diskrete Merkmale mit wenigen Ausprägungen
- Reihenfolge auf X-Achse nur ab Ordinalskala interpretierbar
- die Reihenfolge der Balken für nominal skalierte Merkmale sollte nach Häufigkeiten (auf- oder absteigend) gewählt werden, um Fehlinterpretationen zu vermeiden
- Abstände auf X-Achse nur für metrisch messbare Merkmale interpretierbar

Beispiel B.7: Achtung: fehlende Achsenbeschriftung kann Aussagen verzerren

3 Unternehmen mit Umsätzen $x = (1005, 1100, 1300)$



B.1.3 Kreisdiagramme

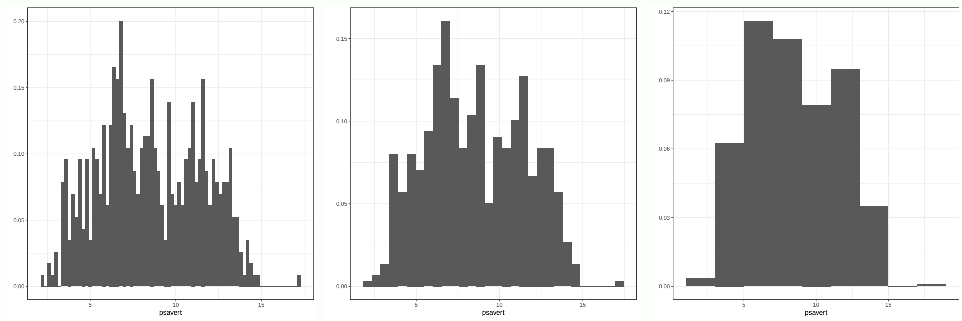
- Winkel sowie Fläche des Kreissektors proportional zur Häufigkeit
- sinnvoll für diskrete Merkmale mit wenigen Ausprägungen und Nominalskala
- Ordnungen nicht darstellbar

B.1.4 Histogramme

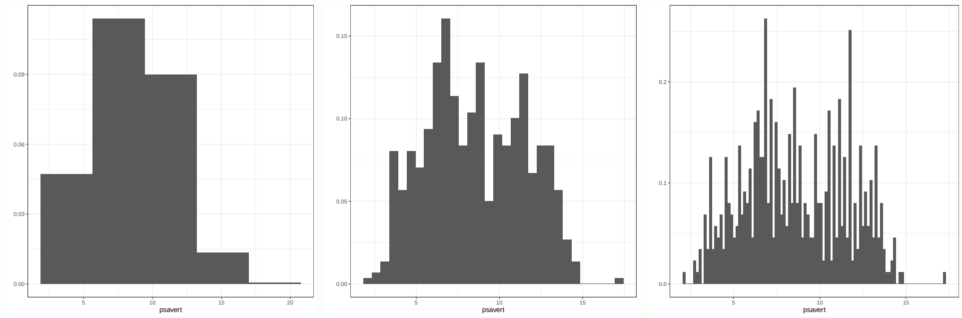
- mind. ordinalskalierte Merkmale mit vielen Ausprägungen, insbesondere geeignet für stetige Merkmale auf einer metrischen Skala
- Daten werden gruppiert, häufig ist die Klassenbreite / Bandbreite gleich, darf aber auch variieren
- Fläche des Rechtecks jeder Gruppe proportional zur (relativen oder absoluten) Häufigkeit
- Falls die Bandbreite für alle Klassen konstant ist, sind natürlich auch die Höhen proportional zu den Häufigkeiten
- Falls die Flächeninhalte relativen Häufigkeiten entsprechen, ist die Gesamtfläche des Histogramms 1 und falls die Flächeninhalte den absoluten Häufigkeiten entsprechen, beträgt die Gesamtfläche gerade dem Stichprobenumfang n.
- Falls die Bandbreite für alle Klassen gleich sein soll, muss nur noch die Anzahl der Klassen / Gruppen bzw. deren Breite festgelegt werden. In R und anderen Statistik-Programmen wird diese üblicherweise automatisch gewählt, kann aber auch vom Nutzer angepasst werden, siehe Beispiel B.8

Beispiel B.8: Fortsetzung Beispiel B.6

verschiedene Bandbreiten (0.2, automatisch, 2)



bzw. Festlegung der Klassenanzahl (5, 30, 100)



B.2 Empirische Verteilungsfunktion

Wir haben gesehen, dass wir bei der Häufigkeitsanalyse für diskrete und stetige Merkmale unterschiedlich vorgehen müssen, da sowohl die relativen als auch die absoluten Häufigkeiten typischerweise nur für diskrete Merkmale aussagekräftig sind und stetige Merkmale klassiert werden müssen. Die empirische Verteilungsfunktion liefert hingegen Information über die Verteilung einer Stichprobe sowohl für diskrete als auch stetige Merkmale auf einen Blick, indem Sie die relativen Häufigkeiten kumuliert.

Definition B.2: Empirische Verteilungfunktion

Die Funktion $F : \mathbb{R} \to \mathbb{R}$ mit

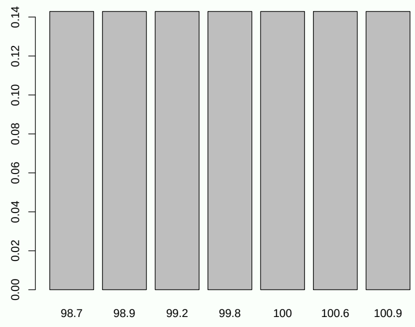
$$F(x) = \sum_{j:\, a_j \leq x} f(a_j), \quad x \in \mathbb{R} \tag{44}$$

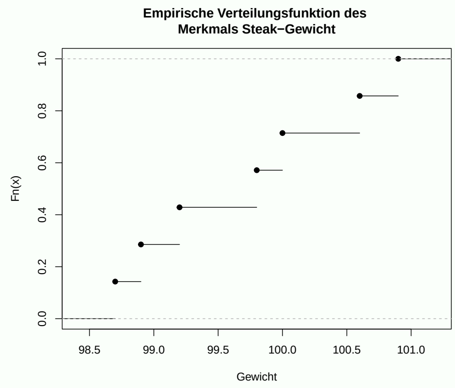
*heißt **empirische Verteilungsfunktion**. Formel (44) lässt sich alternativ schreiben als:*

$$F(x) = \begin{cases} 0, & \text{für } x < a_1 \\ f(a_1), & \text{für } a_1 \leq x < a_2 \\ f(a_1) + f(a_2), & \text{für } a_2 \leq x < a_3 \\ \vdots & \quad \vdots \\ f(a_1) + f(a_2) + \ldots + f(a_{k-1}), & \text{für } a_{k-1} \leq x < a_k \\ 1, & \text{für } x \geq a_k \end{cases}$$

Beispiel B.9: Fortsetzung Steakgewicht, Beispiel B.1

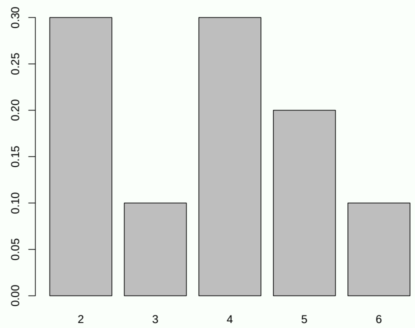
Die linke Grafik zeigt die relativen Häufigkeiten und die rechte Grafik die kumulierten Häufigkeiten. Beide Grafiken enthalten die selbe Information, aufbereitet aus verschiedenen Blickwinkeln. Für steige Merkmale ist die linke Grafik in der Regel nicht sehr aussagekräftig, da alle Werte hier genau einmal aufgetreten sind.

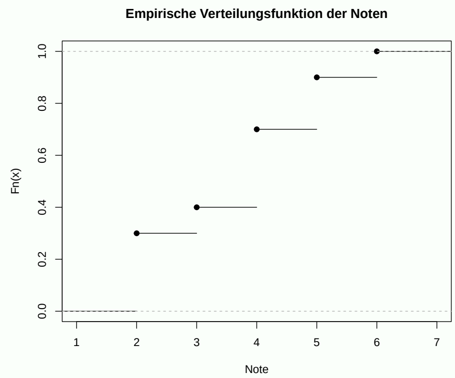




Ais diesem Grund betrachtet man für stetige Merkmale Histgramme, welche ihre vollen Vorteile allerdings erst für größere Stichproben ausspielen, wie in Beispiel B.11 anhand echter Daten illustriert.

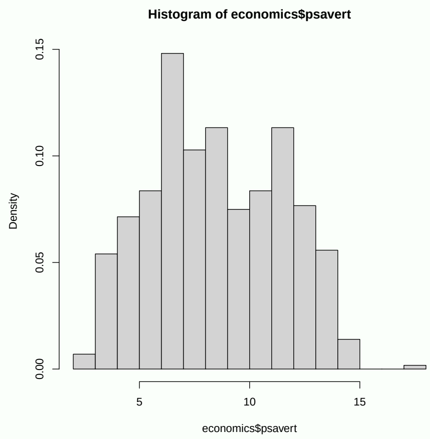
Beispiel B.10: Fortsetzung Noten, Beispiel B.2

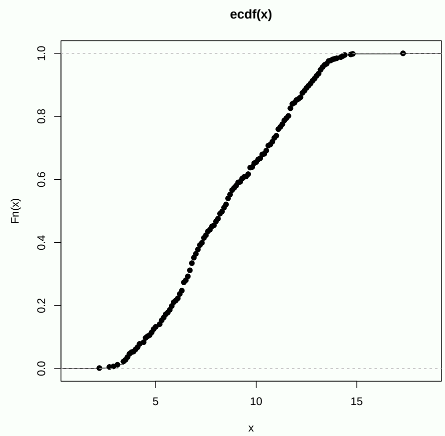




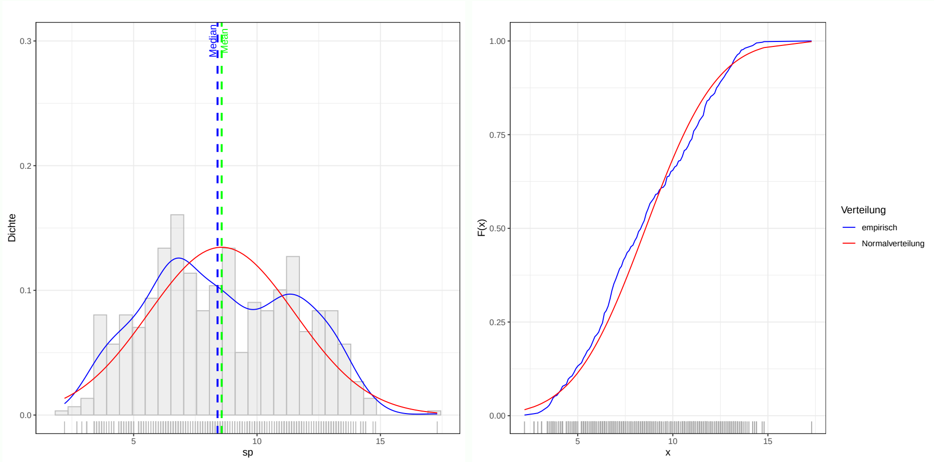
Beispiel B.11: Fortsetzung Beispiel B.6

Verteilung des stetigen Merkmals Sparrate (psavert) - links Histogramm und rechts die Verteilungsfunktion





Die folgendende Grafik zeigt zusätzlich zur empirischen Verteilung noch Median und Mittelwert sowie die Normalverteilung und zeigt auf der x-Achse nochmals die konkreten Stichprobewerte an



Bemerkung B.3: Eigenschaften der empirischen Verteilungsfunktion

- *F ist monoton wachsend,*
- $\lim_{x\to-\infty} F(x) = 0$ *sowie* $\lim_{x\to\infty} F(x) = 1$,
- $0 \leq F(x) \leq 1$ *für alle* $x \in \mathbb{R}$ *und*
- *F ist rechtsseitig stetig.*

B.3 Empirische Maßzahlen

Mittels Grafiken oder Häufigkeitstabellen können die Verteilungen dargestellt werden, so dass Strukturen („Zentrum“, „Schwerpunkt“, „Streuung“, ...) sichtbar werden. Maßzahlen beschreiben eine Verteilung durch einen numerischen Wert. Welches Maß in einer bestimmten Fragestellung sinnvoll ist, hängt vom Kontext, von der Datensituation und vom Skalenniveau des Merkmals ab. Dieser Abschnitt behandelt Maßzahlen oder Parameter von Verteilungen, mit denen sich deren Eigenschaften in komprimierter Form durch numerische Werte formal quantifizieren lassen und somit eine Vergleichbarkeit der Verteilungen verschiedener Merkmale und folglich natürlich auch der Stichproben selbst möglich machen. Es gibt verschiedene Maße, um spezielle Eigenschaften, beispielsweise

- Zentrum / Lage,
- Streuung,
- Schiefe und Wölbung und
- Konzentration,

zu quantifizieren. Diese Maßzahlen lassen sich auch in geeigneter Form, wie etwa im Box-Plot, visualisieren.

B.3.1 Lagemaße

Ein Lagermaß oder Lageparameter beschreibt in bestimmter Weise ausgezeichnete Werte, wie z.B. das Zentrum (Schwerpunkt) einer Stichprobe bzw. deren zugehöriger Häufigkeitsverteilung. Sie dienen zur Beschreibung des mittleren Niveaus eines Merkmals. Beispiele sind das Durchschnitt-seinkommen, die mittlere Lebensdauer eines technischen Geräts, das normale Heiratsalter oder das am häufigsten genannte Studienfach.

Definition B.3: Lagemaße

Ein Lagemaß dient zur Beschreibung des Zentrums (mittleren Niveaus) einer Stichprobe.

Bemerkung B.4: Forderung: Translationsäquivarianz

Es sollte keine Rolle spielen, ob ein Merkmal erst linear transformiert wird und dann das Lagemaß berechnet wird, oder anders herum. Wird eine Längenangabe beispielsweise von Meilen in Kilometer umgerechnet, so soll die Maßzahl in Kilometer mit der entsprechend transformierten Maßzahl aus den Meilenangaben übereinstimmen. Formal soll also für die mittels $y = ax + b$ linear transformierten Daten die Maßzahl $L(y) = aL(x) + b$ sich ebenso linear aus der Maßzahl $L(x)$ ergeben.

Es gibt verschiedene Lagemaße, die unterschiedliche Mindestanforderungen an das Skalenniveau stellen und unterschiedliche Interpretationen der „Mitte " eines Datensatzes geben:

- Modus bzw. Modalwert (mind. Nominalskala)
- Median (mind. Ordinalskala)
- arithmetisches Mittel (metrische Skala)
- geometrisches Mittel (Verhältnisskala)
- harmonisches Mittel
- getrimmtes Mittel $\curvearrowright$ Literatur

Welches Lagemaß ist nun geeignet? Die Auswahl ist abhängig von der konkreten Fragestellung sowie vom Skalenniveau des untersuchten Merkmals. Häufig verwendet man den Modus ($\triangleq$ häufigster Wert der Stichprobe) für die Nominalskala, den Median ($\triangleq$ Mitte der Stichprobe) für ordinalskalierte Merkmale und das arithmetisches Mittel auf der metrischen Skala. Natürlich ist es auch sinnvoll für metrische Merkmale den Median zu brechnen, insbesondere für schiefe Verteilungen. Beispielsweise ist das Medianeinkommen häufig aussagekräftiger als das Durschschnittseinkommen, da die wenigen hohen Einkommen, den Durchschnitt nach oben ziehen. Manchmal sind aber trotz metrischer Skala Median und arithmetisches Mittel teilweise ungeeignet, z.B. bei

- sich multiplikativ verändernden Werten (Zinsen, Wachstumsfaktoren, ...) $\Rightarrow$ geometrisches Mittel
- Durchschnittsgeschwindigkeiten $\Rightarrow$ harmonisches Mittel

Definition B.4: Modus bzw. Modalwert

Der Modus bzw. Modalwert entspricht der am häufigsten auftretenden Ausprägung. Er wird insbesondere bei nominalskalierte Merkmale berechnet. Mathematische Darstellung:

$$x_{mod} = a_j \Leftrightarrow h_j = \max\{h_1, h_2, \dots, h_k\} \tag{45}$$

Der Modus ist nicht immer eindeutig, falls es mehrere Maxima gibt, jedoch ist er sehr robust gegenüber Ausreißern, d.h. besonders großen oder kleinen Werten. Bei linearer Transformation (mit $a \neq 0$) der Daten ist der Modus translationsäquvariant. Die Verwendung des Modus ist bei jedem Skalenniveau möglich. Für nominalskalierte Daten ist der Modus der einzige zulässige Lageparameter. Eine sinnvolle Beschreibung der Daten mit Hilfe des Modus ergibt sich bei jedem Datenniveau aber nur für den Fall einer eingipfligen (unimodalen) Verteilung.

Beispiel B.12: Modus der Stichprobe $(5, 3, 2, 7, 1, 3, 2, 7)$

- $n = 8$, $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 5$, $a_5 = 7 \Rightarrow k = 5$
- $h_1 = h(a_1) = h(1) = 1 \Rightarrow f_1 = \frac{1}{8}$,
- $h_2 = h(a_2) = h(2) = 2 \Rightarrow f_2 = \frac{2}{8}$
- $h_3 = h(a_3) = h(3) = 2 \Rightarrow f_3 = \frac{2}{8}$
- $h_4 = h(a_4) = h(5) = 1 \Rightarrow f_4 = \frac{1}{8}$
- $h_5 = h(a_5) = h(7) = 2 \Rightarrow f_5 = \frac{2}{8}$

$\Rightarrow \quad x_{mod} \in \{2, 3, 7\}$ *(nicht eindeutig!)*

Als Median wird die „Mitte" eines geordneten Datensatzes bezeichnet. Vorraussetzung für die Bestimmung des Median ist eine Ordnungsskala, so die Stichprobe aufsteigend sortiert werden kann und wir die geordnete Stichprobe $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ erhalten.

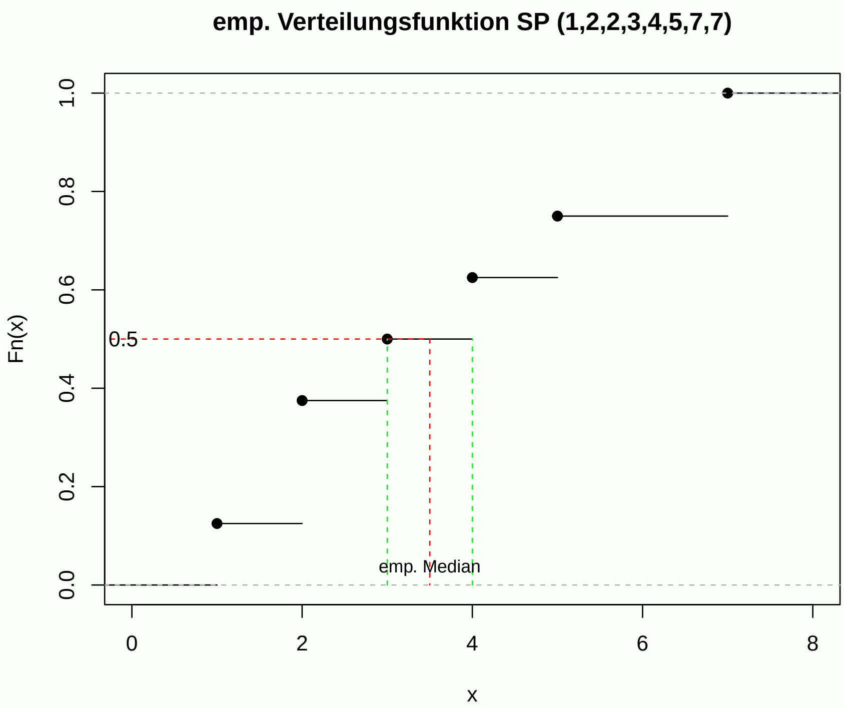
Definition B.5: Median

Der empirische Median der Stichprobe erfüllt die Forderung, dass mindestens 50% der beobachteten Werte kleiner oder gleich und mindestens 50% der beobachteten Werte größer oder gleich dem Median sind.

Der Median muss nicht eindeutig sein, wie das folgende Beispiel zeigt. Mit der empirischen Verteilungsfunktion kann man dies auch grafisch veranschaulichen.

Beispiel B.13: uneindeutiger Median und Illustration mittels Verteilungsfunktion

```
## [1] 3.5
## 50%
##   3
## 50%
## 3.5
## [1] 0.5
## [1] 0.625
## [1] 0.5
## [1] 0.5
## [1] 0.625
## [1] 0.5
```



Bemerkung B.5: Berechnung des Medians

Falls der Umfang der Stichprobe n gerade ist, gibt es keine eindeutige Lösung und sowohl $x_{(\frac{n}{2})}$ als auch $x_{(\frac{n}{2}+1)}$ und sogar jeder Wert dazwischen erfüllen diese Forderung. Bei einer metrischen Skala mittelt man dann häufig über $x_{(\frac{n}{2})}$ und $x_{(\frac{n}{2}+1)}$, d.h.

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right), & \text{falls } n \text{ gerade} \end{cases} \tag{46}$$

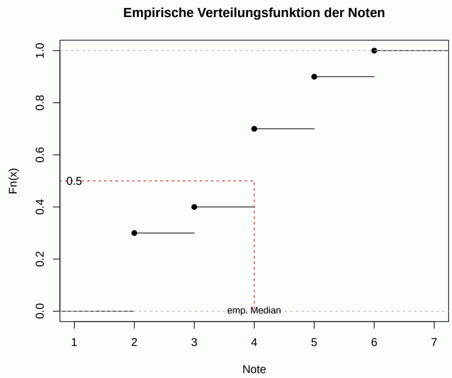
Eine andere Alternative den empirischen Median eindeutig zu wählen, ist immer den kleinsten Wert, der die Bedingungen aus Definition B.5 erfüllt, zu wählen. Dies kann mit Hilfe der verallgemeinerten Inversen der empirischen Verteilungfunktion

$$x_{med} = F^{-}(0.5) = \min\{x :\ F(x) \geq 0.5\} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade} \\ x_{(\frac{n}{2})}, & \text{falls } n \text{ gerade} \end{cases} \tag{47}$$

geschrieben werden und in Beispiel B.14 illustriert wird.

Beispiel B.14: Fortsetzung Noten, Beispiel B.2 & B.5

```
#beispiel: Noten, diskrete Daten
#noten = c(2, 4, 5, 2, 3, 4, 2, 6, 5, 4)
noten.sort = sort(noten)

plot.ecdf(noten, main="Empirische Verteilungsfunktion der Noten",
        xlab="Note")

#empirischer Median als Mittelung
(noten.median = median(noten))
## [1] 4
#empirischen Median enzeichnen
lines(c(0,noten.median,noten.median),c(0.5,0.5,0),lty=2,col="red")

text(4, 0.0, "emp. Median",cex = .8)
text(1, 0.5, "0.5")

#4 ist der erste Wert, für den F(4)>=0.5
F.noten(4)
## [1] 0.7
F.noten(3.99)
## [1] 0.4
#Mittelung des 5. und 6. kleinsten Wertes
noten.sort[n/2]
## [1] 2
noten.sort[n/2+1]
## [1] 3
```



In diesem Beispiel ist der Median eindeutig, so dass auch die Mittelung des 5. und 6. kleinsten Stichprobenwertes

$$x_{med} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2}\left(x_{(5)} + x_{(6)}\right) = \frac{1}{2}(4+4) = 4 \tag{48}$$

auf den eindeutigen Median führt. Ebenso ist im Graph zur Verteilungsfunktion zu erkennen, dass alle Werte $x \geq 4$ *die Forderung* $F(x) \geq 0.5$ *erfüllen, aber kein* $x < 4$*. Demnach ist der kleinste dieser Werte tatsächlich 4 und somit*

$$x_{med} = 4 = \min\{x : F(x) \geq 0.5\}.$$

Bemerkung B.6: Umgang mit uneindeutigen Median für ordinalskalierte Merkmale

Falls das betrachtete Merkmal nur ordinal skaliert ist, so ist bei der Berechnung des Medians zu beachten, dass die Mittelung von $x_{\left(\frac{n}{2}\right)}$ *und* $x_{\left(\frac{n}{2}+1\right)}$ *für den Fall* n *gerade nicht sinnvoll ist, es sei denn beide Werte sind gleich. Im Falle verschiedener Werte erfüllt sowohl* $x_{\left(\frac{n}{2}\right)}$ *als auch* $x_{\left(\frac{n}{2}+1\right)}$ *die Forderung an den Median (höchstens 50% der Werte kleiner oder gleich und höchstens 50% der Werte größer oder gleich dem Median), so dass dieser nicht mehr eindeutig bestimmt werden kann. Bsp.: Zahlen in obiger Stichprobe entsprechen Schadstoffklassen, dann kann der Wert 3.5 nicht sinnvoll interpretiert werden. Allerdings* *gilt für geraden Stichprobenumfang automatisch immer*

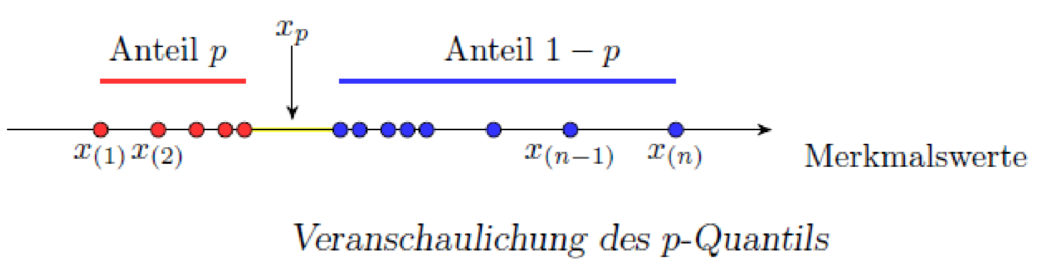
$$x_{med} = F^{-}(0.5) = \min\{x : F(x) \geq 0.5\} = x_{(\frac{n}{2})} \tag{49}$$

Ein Quantil ist die Verallgemeinerung eines Medians.

Definition B.6: (empirische) Quantile

Das p-Quantil $\tilde{x}_p$ der Stichprobe $x_1, \ldots, x_n$ mit $p \in (0,1)$ erfüllt die folgende Forderung: mindestens np Werte sind kleiner oder gleich $\tilde{x}_p$ und mindestens $n(1-p)$ Werte größer oder gleich $\tilde{x}_p$

Für $p = 0.5$ erhalten wir gerade den Median. Die folgende Abbildung illustriert das p-Quantil



Bemerkung B.7: Berechnung Quantile

Ebenso wie der Median, sind auch die Quantile nicht immer eindeutig. Möchte man eine eindeutige Version, verwendet man häufig, wie für den Median illustriert, einfach den kleinsten Wert aller Werte, die die Forderung aus Definition B.6 und dann gilt $F(\tilde{x}_p) \approx p$

$$\tilde{x}_p = F^{-}(p) = \min\{x : F(x) \geq p\},$$

d.h. es gilt $F(\tilde{x}_p) \geq p$ und von allen x mit $F(x) \geq p$ ist $\tilde{x}_p$ der kleinste Wert

Häufig sind das 5%-Quantil und 95%-Quntil von besonderen Interesse, die die Ränder der Verteilung näher beschreiben und im Bank- und Risikomanagement auch als **Value at Risk** bekannt sind. Die Quartile $\tilde{x}_{0.25}$, $\tilde{x}_{0.5}$ und $\tilde{x}_{0.75}$ werden ebenfalls häufig betrachtet, da diese die Stichprobe vierteln (analog zum Median, der die Daten in Hälften unterteilt). Die Quartile bilden die Grundlage der Box des sogenannten **Boxplots**, wobei in der einfachsten Basisversion (siehe Abbildung 9) die Whiskers bzw. Antennen durch den kleinsten bzw. größten Stichprobenwert dargestellt werden.

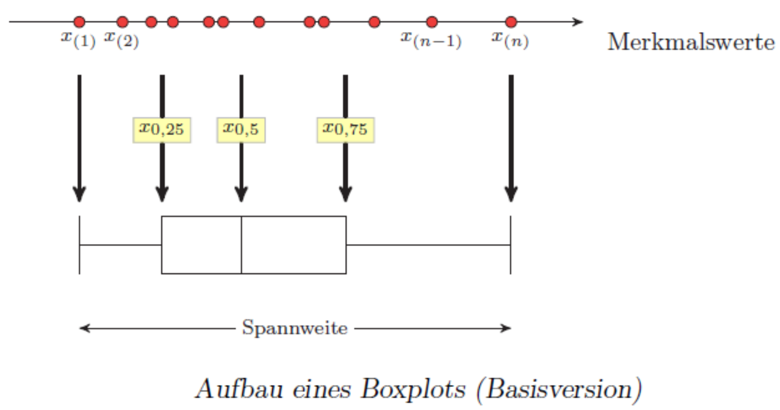


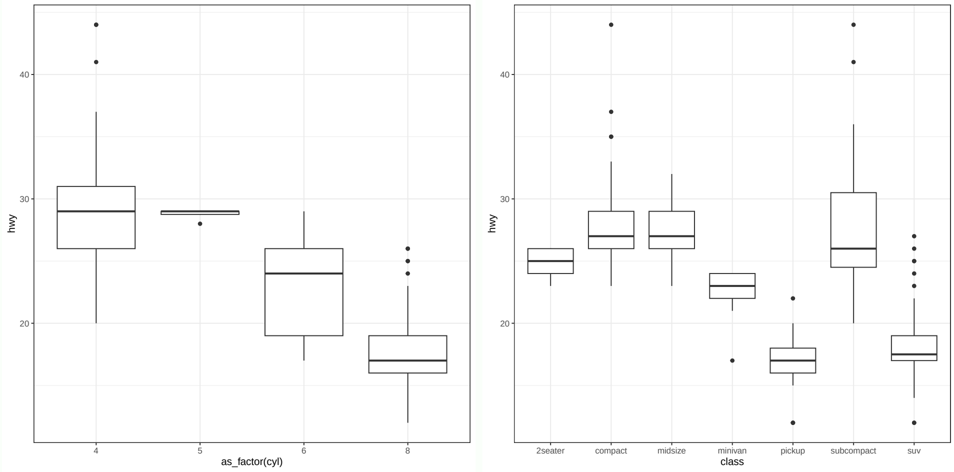
Abbildung 9: Grundversion des Boxplots, Quelle [**Mittag2020**]

Boxplots sind in allen gängigen Statistikprogrammen implementiert und beschränken typischerweise die Whiskers auf das anderthalbfache der Länge der Box. Diese Darstellung eignet sich besonders zum Vergleich verschiedener Gruppen.

Beispiel B.15: Verbrauch

In R stellt das Grafik-Paket ggplot2 *den Datensatz* mpg*, der öknomische Daten für verschiedene Automodelle bereithält. Beispielsweise misst die Variable* hwy *(highway miles per gallon) indirekt den Verbrauch. Die beiden Boxplots vergleichen die zurückgelegten Strecken je Gallone in Abhängigkeit von der Zylinderanzahl bzw. der Klasse.*

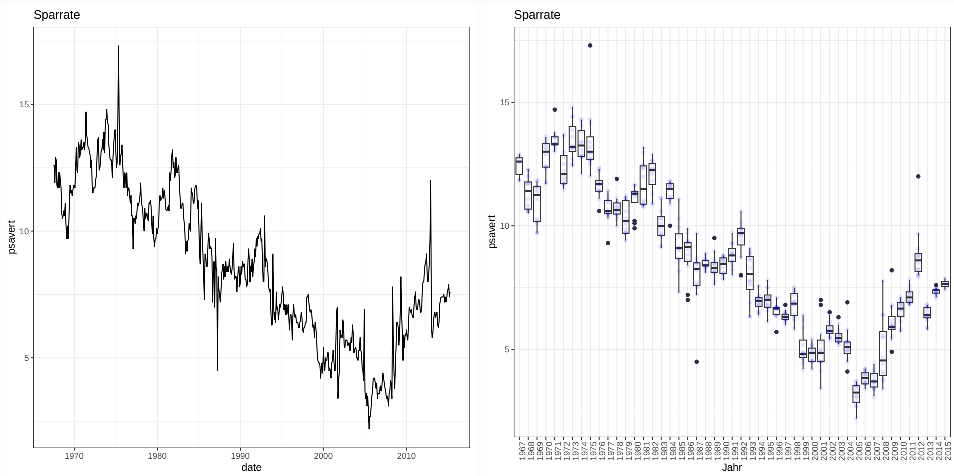
```
## ?mpg #Aufruf der Hilfe zur Beschreibung des Datensatzes
head(mpg)

#Verbrauch in Abhängigkeit der Zylinder-Anzahl
ggplot(mpg, aes(as_factor(cyl), hwy)) +
  geom_boxplot()

#Verbrauch in Abhängigkeit der Klasse
ggplot(mpg, aes(class, hwy)) +
  geom_boxplot()
```



Beispiel B.16: Fortsetzung Beispiel B.6

Auch der zeitliche Verlauf der Schwankungen kann mit einem Boxplot dargestellt werden.

```
eco = eco %>% mutate(year = lubridate::year(date),
                     month = lubridate::month(date),
                     year.fac = as_factor(year),
                                         month.fac = as_factor(month))
#zeitliche Entwicklung
eco %>% ggplot(aes(x=date, y=psavert))+ geom_line() +
     ggtitle("Sparrate")

#Boxplot zum Vergleich der Jahre (jeweils 12 Monatswerte)
eco %>% ggplot(aes(x=year.fac, y=psavert))+ geom_boxplot() +
 xlab("Jahr") +   geom_jitter(width=0.03, alpha = 0.1, col="blue") +
 theme(axis.text.x = element_text(angle = 90), ) + ggtitle("Sparrate")
```



Bemerkung B.8: Berechnung der Quantile in R

In R gibt es 9 verschiedene Berechnungsmethoden für die empirischen Quantile einer Stichprobe, wobei die Methode 7 standardmäßig verwendet wird. Die verschiedenen Methoden basieren auf verschiedenen Schätzfunktionen, die später im Kapitel zur induktiven Statistik besprochen werden. Der folgende R-Code illustriert die verschiedenen Methoden.

```
#Beispiele zu Quantilen und deren verschiedenen Umsetzungen
# im Falle der Nicht-Eindeutigkeit
daten1 = c(1,1,1,3,4,7,9,11,13,13)
n=length(daten1)

p = c(0.1,0.3,0.35,0.5,0.75,0.9)

#Standardvariante in R
quantile(daten1,probs=p)
##   10%   30%   35%   50%   75%   90%
##  1.00  2.40  3.15  5.50 10.50 13.00
quantile(daten1,probs=p,type = 7)
##   10%   30%   35%   50%   75%   90%
##  1.00  2.40  3.15  5.50 10.50 13.00
#alternative Algorithmen
quantile(daten1,probs=p,type = 1)
## 10% 30% 35% 50% 75% 90%
##   1   1   3   4  11  13
quantile(daten1,probs=p,type = 2)
##  10%  30%  35%  50%  75%  90%
##  1.0  2.0  3.0  5.5 11.0 13.0
quantile(daten1,probs=p,type = 3)
## 10% 30% 35% 50% 75% 90%
##   1   1   3   4  11  13
quantile(daten1,probs=p,type = 4)
## 10% 30% 35% 50% 75% 90%
##   1   1   2   4  10  13
quantile(daten1,probs=p,type = 5)
##  10%  30%  35%  50%  75%  90%
##  1.0  2.0  3.0  5.5 11.0 13.0
quantile(daten1,probs=p,type = 6)
##  10%  30%  35%  50%  75%  90%
##  1.0  1.6  2.7  5.5 11.5 13.0
quantile(daten1,probs=p,type = 7)
##   10%   30%   35%   50%   75%   90%
##  1.00  2.40  3.15  5.50 10.50 13.00
quantile(daten1,probs=p,type = 8)
##    10%    30%    35%    50%    75%    90%
##  1.000  1.867  2.900  5.500 11.167 13.000
quantile(daten1,probs=p,type = 9)
##    10%    30%    35%    50%    75%    90%
##  1.000  1.900  2.925  5.500 11.125 13.000
```

Am häufigsten wird das arithmetische Mittel, welches als Durchschnitt aller Werte einer Stichprobe berechnet wird, als zentrales Lagemaß berechnet. Die Voraussetzung zur Bestimmung des arithmetischen Mittel sind metrisch messbare Merkmale.

Definition B.7: Arithmetisches Mittel

Bei dem arithmetischen Mittel

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

handelt es sich um den Durchschnitt aller Werte.

Beispiel B.17: arithmetisches Mittel

- *SP* $(5, 3, 2, 7, 1, 3, 2, 7)$
- $\bar{x} = \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{30}{8} = 3.75$
- *Berechnung in R*

```
## [1] 3.75
```

Bemerkung B.9: Eigenschaften des arithmetischen Mittels

Das arithmetische Mittel muss nicht angenommen werden. Bei linearer Transformation der Daten ist das arithmetische Mittel translationsäquvariant. Es ist sehr anfällig gegen Ausreißer. Ferner enstpricht wegen

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

das arithmetische Mittel dem Schwerpunkt. Suchen wir für die Funktion

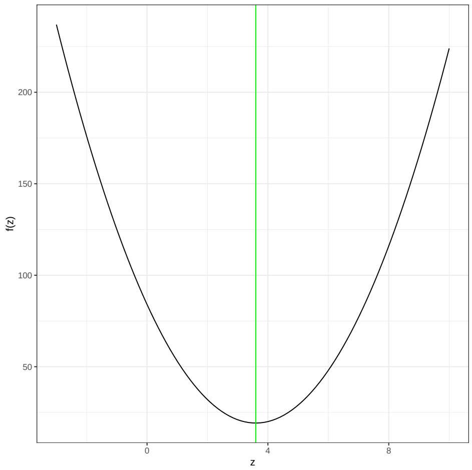
$$f(z) = \sum_{i=1}^{n} (x_i - z)^2 \to \min_{z \in \mathbb{R}} \tag{50}$$

das Minimum wird dies gerade für $z^\star = \overline{x}$ angenommen, d.h. es gilt $\forall z \in \mathbb{R}$ immer

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 \leq \sum_{i=1}^{n} (x_i - z)^2$$

Illustration der Minimaleigenschaft:

```r
sp = c(3,1,4,3,7)
f = function(x){
  return(sum((sp-x)^2))
}

f(4)
## [1] 20
f(5)
## [1] 29
f(mean(sp))
## [1] 19.2
ggplot() + xlim(-3,10) + geom_function(fun=Vectorize(f)) + xlab("z") + ylab("f(z)") +
  geom_vline(xintercept = mean(sp), col = "green")
```



In einigen Situationen berechnet man darüber hinaus auch das geometrische oder harmonische Mittel, da das arithmetische Mittel nicht automtisch immer den Durchschnitt angibt, den man tatsächlich meint. Möchte man beispielsweise die durchschnittliche Wachstumsrate oder Durchschnittsrenditen ermitteln, führt dies zum geometrischen Mittel.

Definition B.8: geometrisches Mittel

Das geometrische Mittel ist derjenige Mittelwert, den man Mithilfe der n-ten Wurzel aus dem Produkt der betrachteten n positiven Zahlen erhält.

$$x_{geo} = (x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{51}$$

Bemerkung B.10: Berechnung des geometrischen Mittels

Mit der Transformation $y_i = \log(x_i)$ sowie den Umformungen

$$\begin{aligned} x_{geo} &= (x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{\frac{1}{n}} = e^{\frac{1}{n}\log(x_1 \cdot x_2 \cdot \ldots \cdot x_n)} \\ &= e^{\frac{1}{n}(\log(x_1)+\log(x_2)+\ldots+\log(x_n))} = e^{\frac{1}{n}(y_1+y_2+\ldots+y_n)} = e^{\overline{y}} = \exp\left\{\overline{\log x}\right\} \end{aligned} \tag{52}$$

ist gut zu erkennen, dass wir das geometrische Mittel auch berechnen können, indem wir die Exponentialfunktion des arithmetischen Mittels der logarithmierten Stichprobe berechnen.

```
sp = c(3,1,4,3,7)

#geometrisches Mittel via Definition
(prod(sp))^(1/length(sp))
## [1] 3.022
#mittels obiger Transformation
exp(mean(log(sp)))
## [1] 3.022
#zum Vergleich: arithmetisches Mittel (immer groesser!)
mean(sp)
## [1] 3.6
```

Außerdem gilt immer

$$x_{geo} = \sqrt[n]{\prod_{i=1}^{n} x_i} \leq \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,,$$

was sich für $n = 2$ sehr einfach mittels Binomischer Formel beweisen lässt oder sogar geometrisch (siehe **https://de.wikipedia.org/wiki/Ungleichung_vom_arithmetischen_und_geometrischen_Mittel***) interpretiert werden kann und im Allgemeinen auf die Jensen-Ungleichung zurückzuführen ist.*

Beispiel B.18: Wertsteigerung von Wertpapieren

Sie besitzen ein Wertpapier, welches sich in den letzten Jahren wie folgt entwickelt hat:

to do

Jahr	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*
Wertsteigerung (in %)	*1.5*	*2.0*	*2.4*	*3.5*	*7.4*	*-0.5*	*-2.5*	*-0.2*

Berechnen Sie die durchschnittliche jährliche Wertsteigerung!

Lösung

- *Geometrisches Mittel: $x_{geo} = \sqrt[n]{\prod_{i=1}^{n} x_i}$ (Hinweis: $x_1 = 1.015$, denn wenn man Startkapital K_0 Anfang 2014 investierte, wuchs das Kapital um 1.5% nach einem Jahr auf $K_1 = K_0 \cdot (1 + 0.015) = K_0 \cdot 1.015 = K_0 \cdot x_1$ und das Kapital nach 2 Jahren auf $K_2 = K_0 \cdot x_1 \cdot (1.02) = K_0 = \cdot x_1 \cdot x_2$.) Allgemein wächst das Kapital nach n Jahren dann auf $K_n \cdot x_1 \cdot x_2 \cdot \ldots \cdot x_n$ welches bei einer durchschnittlichen Rendite $\tilde{x}$ über n Jahre dann mit $K_n = K_0 d^n$ übereinstimmen muss. Folglich entspricht die durchschnittliche Rendite $d = x_{geo}$ dem geometrischen Mittel der einzelnen Jahres-Renditen.*
- $x_{geo} = \sqrt[8]{1.1410} = 1.1410^{1/8} = 1.0166$
- *Die durchschnittliche jährliche Wertsteigerung beträgt 1.66%.*

Die Berechnung von Durchschnittsgeschwindigkeiten führt zum harmonischen Mittel.

Definition B.9: harmonisches Mittel

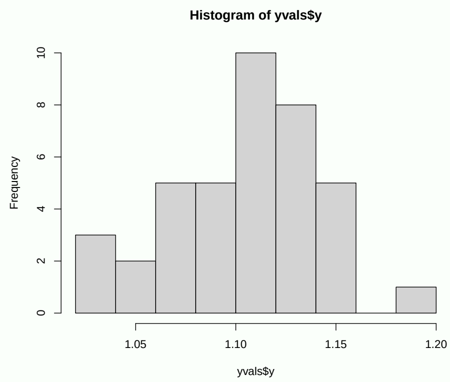
Das harmonische Mittel entspricht dem Kehrwert des arithmetischen Mittels der Kehrwerte der Ausgangsdaten.

$$x_{\text{harm}} = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\sum\limits_{i=1}^{n}\frac{1}{x_i}} = \left(\overline{\left(\frac{1}{x}\right)}\right)^{-1}$$

Das folgende Beispiel ist angelehnt an [Cra12][Kapitel 3].

Beispiel B.19: arithmetisches vs geometrisches vs harmonisches Mittel, I

```
yvals = read.table("./data/yvalues.txt",header=T)

hist(yvals$y)
#Median
median(yvals$y)
## [1] 1.109
#Mittelwert
mean(yvals$y)
## [1] 1.103
#geometrisches Mittel
exp(mean(log(yvals$y)))
## [1] 1.103
#harmonisches Mittel
1/mean(1/yvals$y)
## [1] 1.102
```



In obigen Beispiel ist die Verteilung recht symmetrisch, so dass alle Lagemaße sehr nahe beieinander liegen. Die folgenden Beispiele zeigen, dass die verschiedenen Maße für die zentrale Lage sich auch deutlich unterscheiden können.

Beispiel B.20: arithmetisches vs geometrisches vs harmonisches Mittel, II

```
#weitere Beispiele
data = c(3,4,6,7)
mean(data)
## [1] 5
median(data)
## [1] 5
exp(mean(log(data)))
## [1] 4.738
1/mean(1/data)
## [1] 4.48
insects = c(1,10,1000,10,1)
mean(insects)
## [1] 204.4
median(insects)
## [1] 10
exp(mean(log(insects)))
## [1] 10
1/mean(1/insects)
## [1] 2.272
```

```
#Bsp
v = c(1,2,4,1)
mean(v) #entspricht nicht der Durchschnittsgeschwindigkeit
## [1] 2
median(v)
## [1] 1.5
exp(mean(log(v)))
## [1] 1.682
1/mean(1/v) #tatsaechliche Durchschnittsgewschwindigkeit
## [1] 1.455
```

B.3.2 Streuungsmaße

Zusätzlich zur Angabe eines Lagermaßes, welches Information über die Mitte der Beobachtungswerte angibt, wird eine Verteilung durch die Angabe von Streuungsmaßen charakterisiert. Wir betrachten im Folgenden

- Spannweite
- Quartilsabstand
- mittlere absolute Abweichung vom Median
- Standardabweichung bzw. Varianz
- Variationskoeffizient

Das Streuungsmaß gibt Auskunft in welcher „Bandbreite “die Werte liegen. In der Regel ist die Berechnugn der Streuungsmaße nur für metrisch skalierte Merkmale sinnvoll, da fast alle Maße auf Abständen basieren.

Die Spannbreite gibt einfach die Breite des Streubereichs wieder.

Definition B.10: Spannweite

$$R = x_{(n)} - x_{(1)} \tag{53}$$

Allerdings ist die Spannweite sehr anfällig gegenüber Ausreißern. Betrachtet man stattdessen den Quartilsabstand, erhält man ein deutlich robusteres Maß für die Schwankung

Definition B.11: Quantilsabstand

$$Q = \tilde{x}_{0.75} - \tilde{x}_{0.25} \tag{54}$$

Definition B.12: Mittlere Abweichung vom Median

$$d = \frac{1}{n}\sum_{i=1}^{n} |x_i - x_{med}| \tag{55}$$

(55) ist ähnlich wie der Median aus Definition B.5 unanfällig gegen Ausreißer, wird aber seltenener verwendet. Ein Grund hierfür liegt in der Nichtdifferenzierbarkeit der Betragsfunktion.

Definition B.13: Empirische Varianz s^2 und Standardabweichung s

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{56}$$

$$s = \sqrt{s^2} \tag{57}$$

$$\tag{58}$$

Die empirische Standardabweichung (57) hat im Gegensatz zur empirischen Varianz (56) die selbe Masseinheit wie das untersuchte Merkmal.

Bemerkung B.11: Bessel-Korrektur der empirischen Varianz und Freiheitsgrade

Warum in Formel (56) bei der Mittelung über alle n quadratischen Abweichungen $(x_1 - \overline{x})^2, \dots, (x_n - \overline{x})^2$ mit dem Faktor $\frac{1}{n-1}$ anstatt mit $\frac{1}{n}$ stattfindet, können wir erst mit Argumenten der induktiven Statistik begründen. Es zeigt sich, dass die Ergebnisse dann einfach zuverlässiger sind, da (56) erwartungstreu bzw. unverzerrt die Varianz schätzt, während mittels (59) die tatsächliche Varianz immer etwas zu klein geschätzt und damit verzerrt geschätzt wird. Der Zusammenhang zwischen der sogenannten unkorrigierten Version

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \tag{59}$$

$$\tag{60}$$

und (56) ist durch

$$s^2 = \frac{n}{n-1}\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2 \tag{61}$$

gegeben und insbesondere für große Stichproben (n groß) wegen $\frac{n}{n-1} \approx 1$ nur marginal und deswegen nur für kleine Stichproben von Bedeutung. Wegen $\frac{n}{n-1} > 1$ ist die Bessel-korrigierte empirische Varianz mit $s^2 = \frac{n}{n-1}\hat{s}^2 > \hat{s}^2$ immer leicht größer als die unkorrigierte empirische Varianz, wobei der Unterschied mit wachsender Stichprobengröße immer kleiner wird. Das korrigierte Gewicht $\frac{1}{n-1}$ steht eng in Verbindung mit den Freiheitsgraden, die wir bei der Verwendung des arihtmetischen Mittels zur Berechnung von (56) haben. Wenn wir für die Stichprobe $x_1, \ldots, x_n$ das arithmetische Mittel $\overline{x} = c$ vorgeben, können nur $n-1$ der n Werte frei gewählt werden. Wählen wir beispielsweise die ersten $n-1$ Werte frei, so muss der letzte Wert

$$x_n = nc - \sum_{i=1}^{n-1} x_i \tag{62}$$

sein, damit das Stichprobenmittel $\overline{x} = c$ enstpricht. Schätzt man in der Statistik einen Parameter, hier die Varianz, dividiert man durch die Anzahl der Freiheitsgrade der Werte, die auf Grundlage der Daten geschätzt werden müssen, hier das arithmetische Mittel.

Beispiel B.21: Berechnung empirische Varianz und Standardabweichung in R

```
data = c(3,4,6,7)
n.data = length(data)

#Varianz und Standardabweichung in R
var(data)
## [1] 3.333
sd(data)
## [1] 1.826
sqrt(var(data))
## [1] 1.826
#eigene Berechnung mit Besselkorrektur
sum((data-mean(data))^2)/(n.data-1)
## [1] 3.333
#ohne Besselkorrektur
sum((data-mean(data))^2)/n.data
## [1] 2.5
var(data)*(n.data-1)/n.data
## [1] 2.5
#einzelne Komponenten
sum(data^2)
## [1] 110
mean(data)
## [1] 5
```

Der Variatinskoeffizient dient zur Berschreibung der Steuung der Daten um ihren Mittelwert. Im Gegensatz zu der Standardabweichung können mittels dem Variationskoeffizienten verschiedene Merkmale der Stichprobe verglichen werden.

Definition B.14: Variatonskoeffizient

$$v = \frac{s}{\overline{x}} \tag{63}$$

Neben grafischer Auswertung (Box-Plots) und Vergleich der Lage- (Mittelwert, Median) und Streuungsmaße, gibt es weitere Maßzahlen, die auf höheren Momenten basieren.

B.3.3 Maße für Schiefe und Wölbung

Schiefe und Wölbung sind weitere Maßzahlen, die die Form der Verteilung charakterisieren. Eine sinnvolle Verwendung ergibt sich jedoch nur im Fall eingipfliger Verteilungen. Eingipflige Verteilungen können symmetrisch, links- oder rechtsschief sein.

Definition B.15: Schiefe

$$g = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^3}} \tag{64}$$

Bemerkung B.12: Schiefe einer Verteilung

- *symmetrische Verteilung* ($\bar{x} \approx x_{med} \approx x_{mod}$, $g \approx 0$)
- *linksschiefe (rechtssteil) Verteilung* ($\bar{x} < x_{med} < x_{mod}$, $g < 0$)
- *rechtsschiefe (linkssteil) Verteilung* ($\bar{x} > x_{med} > x_{mod}$, $g > 0$)

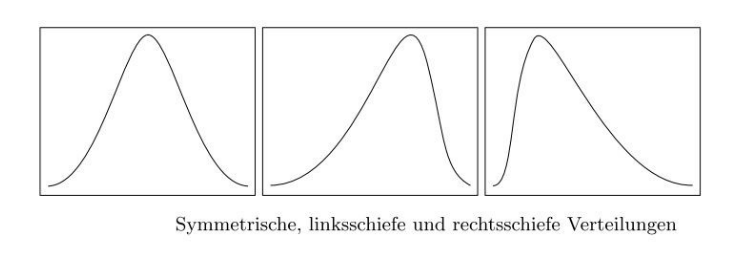


Abbildung 10: *
Quelle: Helge Toutenburg & Christian Heumann [**2008deskriptive**]

Maßzahlen für die Wölbung sollen charakterisieren, wie stark oder schwach der zentrale Bereich und damit zusammenhängend die Randbereiche der Daten besetzt sind. Dabei ist zu beachten, dass Verteilungen mit gleicher Streuung unterschiedliche Wölbungen in der Mitte bzw. unterschiedliche linke und rechte Enden in den Randbereichen besitzen können.

Definition B.16: Wölbung

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} - 3 \tag{65}$$

misst wie stark zentraler Bereich und folglich auch wie schwach die „tails“ im Vergleich zur Normalverteilung besetzt sind.

Bemerkung B.13: Interpretation Wölbung

- *Normalverteilung* ($\gamma = 0$)
- *spitzere Verteilung* ($\gamma > 0$)
- *breitere Verteilung* ($\gamma < 0$)

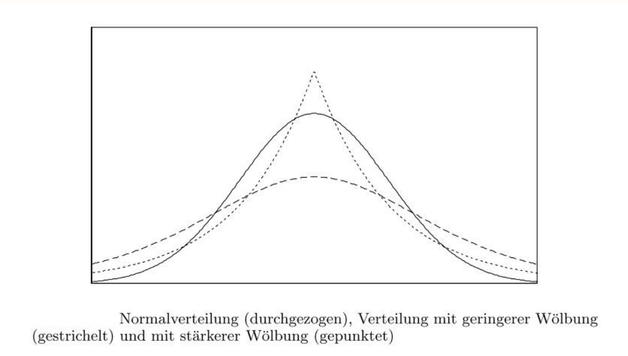
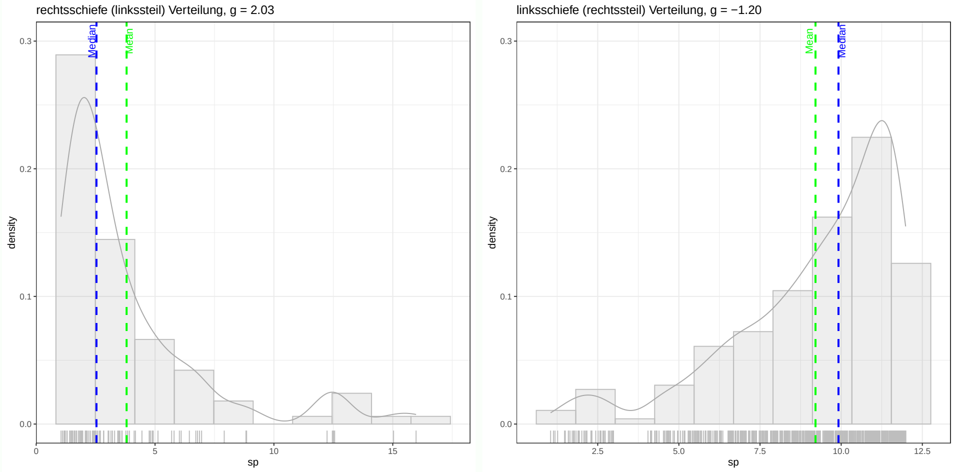


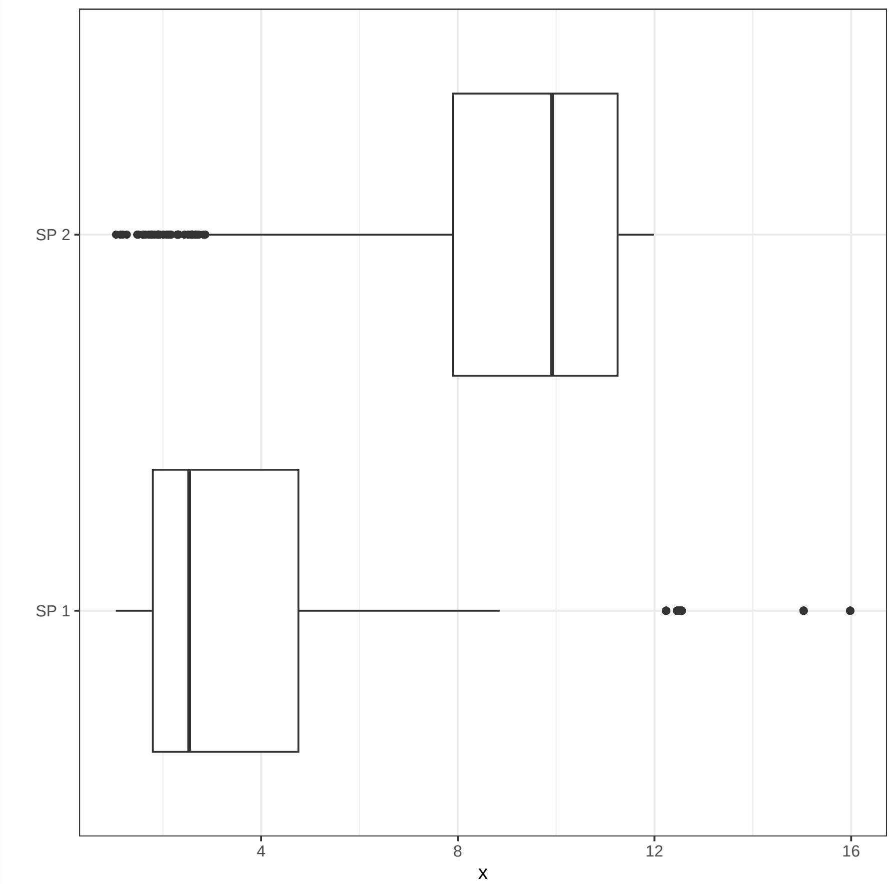
Abbildung 11: *
Quelle: Helge Toutenburg & Christian Heumann [**2008deskriptive**]

Beispiel B.22: Illustration der Schiefe mittels Dichten

Die folgenden beiden Grafiken zeigen zwei unsymmetrisch verteilte Stichproben und die zugehörigen Werte für die Schiefe.



Für den direkten Vergleich beider Verteilungen eignet sich ein Boxplot.



B.3.4 Zusammenfassung Stichprobenmomente

Mit Hilfe der empirischen Momente kann die Stichprobenverteilungen charakterisiert werden.

Definition B.17: Stichprobenmomente

Das allgemeine $k-$te Moment M_k sowie zentrale Moment $\tilde{M}_k$ ($k = 1, 2, \ldots$) der Stichprobe $x_1, \ldots, x_n$ enstprechen

$$M_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k \tag{66}$$

$$\tilde{M}_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^k \tag{67}$$

Das **1. Moment** $M_1 = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$ enstpricht gerade dem arithmetische Mittel. Das zentrierte 1. Moment verschwindet wegen $\tilde{M}_1 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}) = 0$ immer. Das **2. zentrierte Moment** ist wegen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right) = \frac{n}{n-1} \tilde{M}_2 \tag{68}$$

eng mit der empirischen Varianz (56) verbunden, wobei $\hat{s}^2 = \tilde{M}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2$, siehe auch Bemerkung B.11.

Das **3. zentrierte Moment**

$$\tilde{M}_3 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^3$$

ist wesentlicher Bestandteil der Schiefe

$$g = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^3}{\sqrt{\left(\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^2 \right)^3}} = \frac{\tilde{M}_3}{\left(\tilde{M}_2\right)^{3/2}}$$

und das **4. zentrierte Moment**

$$\tilde{M}_4 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^4$$

bestimmt die Wölbung

$$\gamma = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \bar{x})^2 \right)^2} - 3 = \frac{\tilde{M}_4}{\tilde{M}_2^2} - 3$$

Literatur

[Cra12] Michael J. Crawley. *The R Book*. 2nd edition. New York: John Wiley & Sons, 2012. ISBN: 978-1-118-44896-0.

[HMS21] Norbert Henze, Kai Müller und Judith Schilling. *Stochastik rezeptfrei unterrichten*. Springer, 2021. URL: https://link.springer.com/book/10.1007/978-3-662-62744-0.

[Jun14a] Daniel Jung. *Grundlagen der Wahrscheinlichkeit*. 2014. URL: https://www.youtube.com/watch?v=EdKjU6wGwVc&list=PLLTAHuUj-zHhB8UnwPOJzUVHV6_AJ9tz3.

[Jun14b] Daniel Jung. *Normalverteilung & Gaußverteilung*. 2014. URL: https://www.youtube.com/watch?v=_f1vgWUiavY&list=PLLTAHuUj-zHj7qBEx4VYSTjzTEtntuycU.

[Jun14c] Daniel Jung. *Statistik*. 2014. URL: https://www.youtube.com/watch?v=kIZ9-mGbuN8&list=PLLTAHuUj-zHifw_3OhBTvQq2EGX5NedOy.

[Jun14d] Daniel Jung. *Überbuchungen*. 2014. URL: https://www.youtube.com/playlist?list=PLLTAHuUj-zHi2dkHRtQvJ2AwLed917IbL.

[Jun14e] Daniel Jung. *Wahrscheinlichkeitsrechnung Klasse 5-10*. 2014. URL: https://www.youtube.com/watch?v=NxJAgOxhddE&list=PLLTAHuUj-zHhz-bGujv5ZYKVqUAXRp1fQ.

[Jun14f] Daniel Jung. *Zufallsgröße und Wahrscheinlichkeitsverteilung, Grundlagen mit Beispiel*. 2014. URL: https://www.youtube.com/watch?v=unwJSkloq8M&list=PLLTAHuUj-zHjYzRbxle5HNBzn9vOdzZro.

[Lin14] Werner Linde. *Stochastik für das Lehramt*. München: De Gruyter Oldenbourg, 2014. ISBN: 9783110362411. DOI: doi:10.1524/9783110362411. URL: https://doi.org/10.1524/9783110362411.

[MS23] H.J. Mittag und K. Schüller. *Statistik: Eine interdisziplinäre Einführung mit interaktiven Elementen*. Springer Berlin Heidelberg, 2023. ISBN: 9783662682241. URL: https://link.springer.com/book/10.1007/978-3-662-68224-1.