Deskriptive Statistik und Wahrscheinlichkeitsrechnung

Vorlesung an der Universität des Saarlandes

PD Dr. Martin Becker

Sommersemester 2019

Organisatorisches I

- Vorlesung: Freitag, 12-14 Uhr, Gebäude B4 1, Audimax (HS 0.01)
- Übungen: nach gesonderter Ankündigung, Beginn: ab Montag, 15.04.
- Prüfung: 2-stündige Klausur nach Semesterende (1. Prüfungszeitraum)
  Anmeldung im ViPa nur vom 10.05. (8 Uhr) – 28.05. (15 Uhr)!
  (Abmeldung im ViPa bis 04.07., 12 Uhr)
- Hilfsmittel für Klausur
  - „Moderat“ programmierbarer Taschenrechner, auch mit Grafikfähigkeit
  - 2 *beliebig gestaltete* DIN A 4–Blätter (bzw. 4, falls nur einseitig)
  - Benötigte Tabellen werden gestellt, aber **keine weitere Formelsammlung!**
- Durchgefallen — was dann?
  - „Wiederholungskurs“ im kommenden (Winter-)Semester
  - „Nachprüfung“ (voraussichtlich) erst März/April 2020 (2. Prüfungszeitraum)
  - „Reguläre“ Vorlesung/Übungen wieder im Sommersemester 2020

Organisatorisches II

- Informationen und Materialien unter

  `http://www.lehrstab-statistik.de`

  bzw. spezieller

  `http://www.lehrstab-statistik.de/deskrwrss2019.html`

  (bei Problemen `http://www2.lehrstab-statistik.de` versuchen!)
- Kontakt: PD Dr. Martin Becker
  Geb. C3 1, 2. OG, Zi. 2.17
  e-Mail: `martin.becker@mx.uni-saarland.de`
- Sprechstunde nach Vereinbarung (Terminabstimmung per e-Mail)
- Vorlesungsunterlagen
  - Vorlesungsfolien (kein Kompaktskript)
  - Download (inklusive Drucker-freundlicher 2-auf-1 bzw. 4-auf-1 Versionen) in der Regel einige Tage vor der Vorlesung möglich

Organisatorisches III

- Übungsunterlagen
  - Wöchentliche Übungsblätter
  - Download i.d.R. kurz nach Ende der Vorlesung Freitag nachmittags möglich
  - Ebenfalls online: Ergebnisse *(keine Musterlösungen!)* zu einigen Aufgaben
  - Besprechung der Übungsblätter mit ausführlicheren Lösungsvorschlägen in den Übungsgruppen der folgenden Woche.
  - **Übungsaufgaben sollten unbedingt vorher selbst bearbeitet werden!**
  - Geplant: Freiwillige Bearbeitung und Abgabe von (höchstens zwei) Zusatzübungsblättern, die nach „Klausurmaßstäben“ korrigiert zurückgegeben werden.
- Alte Klausuren
  - Aktuelle Klausuren inklusive der meisten Ergebnisse unter „Klausuren“ auf Homepage des Lehrstabs verfügbar
  - Prüfungsrelevant sind (natürlich) alle in Vorlesung und Übungsgruppen besprochenen Inhalte, nicht nur die Inhalte der Altklausuren!

Was ist eigentlich „Statistik“?

- Der Begriff „Statistik“ hat verschiedene Bedeutungen, insbesondere:
  - Oberbegriff für die Gesamtheit der Methoden, die für die Erhebung und Verarbeitung empirischer Informationen relevant sind ($\rightarrow$ statistische Methodenlehre)
  - (Konkrete) Tabellarische oder grafische Darstellung von Daten
  - (Konkrete) Abbildungsvorschrift, die in Daten enthaltene Informationen auf eine „Kennzahl“ ($\rightarrow$ Teststatistik) verdichtet
- Grundlegende Teilgebiete der Statistik:
  - Deskriptive Statistik (auch: beschreibende Statistik, explorative Statistik)
  - Schließende Statistik (auch: inferenzielle Statistik, induktive Statistik)
- Typischer Einsatz von Statistik:

  Verarbeitung — insbesondere Aggregation — von (eventuell noch zu erhebenden) Daten mit dem Ziel, (informelle) Erkenntnisgewinne zu erhalten bzw. (formal) Schlüsse zu ziehen.

$\rightsquigarrow$ Bestimmte Informationen „ausblenden“, um neue Informationen zu erkennen

Vorurteile gegenüber Statistik

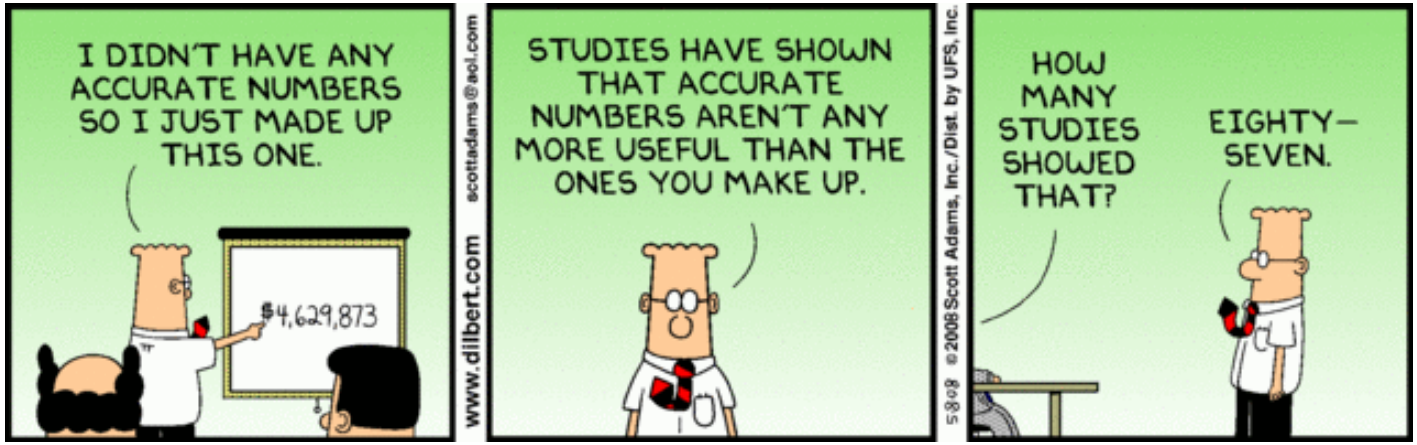
- Einige Zitate oder „Volksweisheiten“:
  - „Statistik ist pure Mathematik, und in Mathe war ich immer schlecht...“
  - „Mit Statistik kann man alles beweisen!“
  - „Ich glaube nur der Statistik, die ich selbst gefälscht habe.“ (häufig Winston Churchill zugeschrieben, aber eher Churchill von Goebbels' Propagandaministerium „in den Mund gelegt“)
  - „There are three kinds of lies: lies, damned lies, and statistics.“ (häufig Benjamin Disraeli zugeschrieben)

  $\rightsquigarrow$ negative Vorurteile gegenüber der Disziplin „Statistik“
- Tatsächlich aber
  - benötigt man für viele statistische Methoden nur die vier Grundrechenarten.
  - ist „gesunder Menschenverstand“ viel wichtiger als mathematisches Know-How.
  - sind nicht die statistischen Methoden an sich schlecht oder gar falsch, sondern die korrekte Auswahl und Anwendung der Methoden zu hinterfragen.
  - werden viele (korrekte) Ergebnisse statistischer Untersuchungen lediglich falsch interpretiert.

Kann man mit Statistik lügen? I

Und falls ja, wie (schützt man sich dagegen)?

- Natürlich kann man mit Statistik „lügen“ bzw. täuschen!
- „Anleitung“ von Prof. Dr. Walter Krämer (TU Dortmund):
  So lügt man mit Statistik, Piper, München, 2009
- Offensichtliche Möglichkeit: Daten (vorsätzlich) manipulieren/fälschen:



Kann man mit Statistik lügen? II

Und falls ja, wie (schützt man sich dagegen)?

- Weitere Möglichkeiten zur Täuschung
  - Irreführende Grafiken
  - (Bewusstes) Weglassen relevanter Information
  - (Bewusste) Auswahl ungeeigneter statistischer Methoden
- Häufiges Problem (vor allem in den Medien):
  Suggestion von Sicherheit durch hohe Genauigkeit angegebener Werte
  $\rightsquigarrow$ zusätzlich: Ablenkung vom „Adäquationsproblem“
  (misst der angegebene Wert überhaupt das „Richtige“?)
- Schutz vor Täuschung:
  - Mitdenken!
  - „Gesunden Menschenverstand“ einschalten!
  - Gute Grundkenntnisse in Statistik!

Beispiel (Adäquationsproblem) I

vgl. Walter Krämer: So lügt man mit Statistik, Piper, München, 2009

- Frage: Was ist *im Durchschnitt* sicherer, Reisen mit Bahn oder Flugzeug?
- Statistik 1:

Bahn	9 Verkehrstote pro 10 Milliarden Passagierkilometer
Flugzeug	3 Verkehrstote pro 10 Milliarden Passagierkilometer

  $\rightsquigarrow$ Fliegen sicherer als Bahnfahren!
- Statistik 2:

Bahn	7 Verkehrstote pro 100 Millionen Passagierstunden
Flugzeug	24 Verkehrstote pro 100 Millionen Passagierstunden

  $\rightsquigarrow$ Bahnfahren sicherer als Fliegen!
- Widerspruch? Fehler?

Beispiel (Adäquationsproblem) II

vgl. Walter Krämer: So lügt man mit Statistik, Piper, München, 2009

- Nein, Unterschied erklärt sich durch höhere Durchschnittsgeschwindigkeit in Flugzeugen (Annahme: ca. 800 km/h vs. ca. 80 km/h)
- Wie wird „Sicherheit“ gemessen? Welcher „Durchschnitt“ ist geeigneter? $\rightsquigarrow$ Interpretation abhängig von der Fragestellung! Hier:
  - Steht man vor der Wahl, eine gegebene Strecke per Bahn oder Flugzeug zurückzulegen, so ist Fliegen sicherer.
  - Vor einem vierstündigen Flug ist dennoch eine größere „Todesangst“ angemessen als vor einer vierstündigen Bahnfahrt.

Beispiel („Schlechte“ Statistik) I

- Studie/Pressemitteilung des ACE Auto Club Europa *anläßlich des Frauentags am 8. März 2010*: „Autofahrerinnen im Osten am besten“ (siehe `https://www.ace.de/fileadmin/user_uploads/Der_Club/Dokumente/Verkehrspolitik/Handout-Booklet-ACE-Studien.pdf`, S. 88–90)
- Untersuchungsgegenstand:
  - Regionale Unterschiede bei Unfallhäufigkeit mit Frauen als Hauptverursacher
  - Vergleich Unfallhäufigkeit mit Frau bzw. Mann als Hauptverursacher
- Wesentliche Datengrundlage ist eine Publikation des Statistischen Bundesamts (Destatis): „Unfälle im Straßenverkehr nach Geschlecht 2008“

Beispiel („Schlechte“ Statistik) II

- Beginn der Pressemitteilung des ACE:

  „Von wegen schwaches Geschlecht: Hinterm Steuer sind Frauen besonders stark.“

  Weiter heißt es:

  “Auch die durch Autofahrerinnen verursachten Unfälle mit Personenschaden liegen wesentlich hinter den von Männern verursachten gleichartigen Karambolagen zurück.“

  und in einer Zwischenüberschrift

  „Schlechtere Autofahrerinnen sind immer noch besser als Männer“

Beispiel („Schlechte“ Statistik) III

- „Statistische“ Argumentation: Laut Destatis-Quelle sind (**angeblich!**)
  - mehr als 2/3 aller Unfälle mit Personenschaden 2008 (genauer: 217 843 von etwas über 320 000 Unfällen) durch PKW-fahrende Männer verursacht worden,
  - nur 37% aller Unfälle mit Personenschaden 2008 durch PKW-fahrende Frauen verursacht worden.
- Erste Auffälligkeit: $66.6\% + 37\% = 103.6\%$ **(???)**
- Lösung: **Ablesefehler** (217 843 aller 320 614 Unfälle mit Personenschaden (67.9%) wurden mit **PKW-Fahrer** (geschlechtsunabhängig) als Hauptverursacher registriert)

Beispiel („Schlechte“ Statistik) IV

- Korrekte Werte:
  - Bei 210 905 der 217 843 Hauptunfallverursacher als PKW-Fahrzeugführer wurde Geschlecht registriert.
  - 132 757 waren männlich (62.95%), 78 148 weiblich (37.05%)
- **Also:** immer noch deutlich mehr Unfälle mit PKW-fahrenden Männern als Hauptverursacher im Vergleich zu PKW-Fahrerinnen.
- **Aber:** Absolute Anzahl von Unfällen geeignetes Kriterium für Fahrsicherheit?

Beispiel („Schlechte“ Statistik) V

- Modellrechnung des DIW aus dem Jahr 2004 schätzt
  - Anzahl Männer mit PKW-Führerschein auf 28.556 Millionen,
  - Anzahl Frauen mit PKW-Führerschein auf 24.573 Millionen.
- Weitere ältere Studie (von 2002) schätzt
  - durchschnittliche Fahrleistung von Männern mit PKW-Führerschein auf 30 km/Tag,
  - durchschnittliche Fahrleistung von Frauen mit PKW-Führerschein auf 12 km/Tag.
- Damit stehen also
  - bei Männern 132 757 verursachte Unfälle geschätzten $30 \cdot 365 \cdot 28.556 = 312688.2$ Millionen gefahrenen Kilometern,
  - bei Frauen 78 148 verursachte Unfälle geschätzten $12 \cdot 365 \cdot 24.573 = 107629.74$ Millionen gefahrenen Kilometern

  gegenüber.

Beispiel („Schlechte“ Statistik) VI

- Dies führt im Durchschnitt
  - bei Männern zu 0.425 verursachten Unfällen mit Personenschaden pro eine Million gefahrenen Kilometern,
  - bei Frauen zu 0.726 verursachten Unfällen mit Personenschaden pro eine Million gefahrenen Kilometern.
- Pro gefahrenem Kilometer verursachen (schätzungsweise) weibliche PKW-Fahrer also durchschnittlich ca. **71% mehr** Unfälle als männliche!
- Anstatt dies zu konkretisieren, räumt die Studie lediglich weit am Ende ein entsprechendes Ungleichgewicht bei der jährlichen Fahrleistung ein.

Beispiel („Schlechte“ Statistik) VII

- Welt Online (siehe `http://www.welt.de/vermischtes/article6674754/Frauen-sind-bessere-Autofahrer-als-Maenner.html`) beruft sich auf die ACE-Studie in einem Artikel mit der Überschrift

  „Frauen sind bessere Autofahrer als Männer“

  und der prägnanten Bildunterschrift

  „Männer glauben bloß, sie seien die besseren Autofahrer. Eine Unfall-Statistik beweist das Gegenteil.“

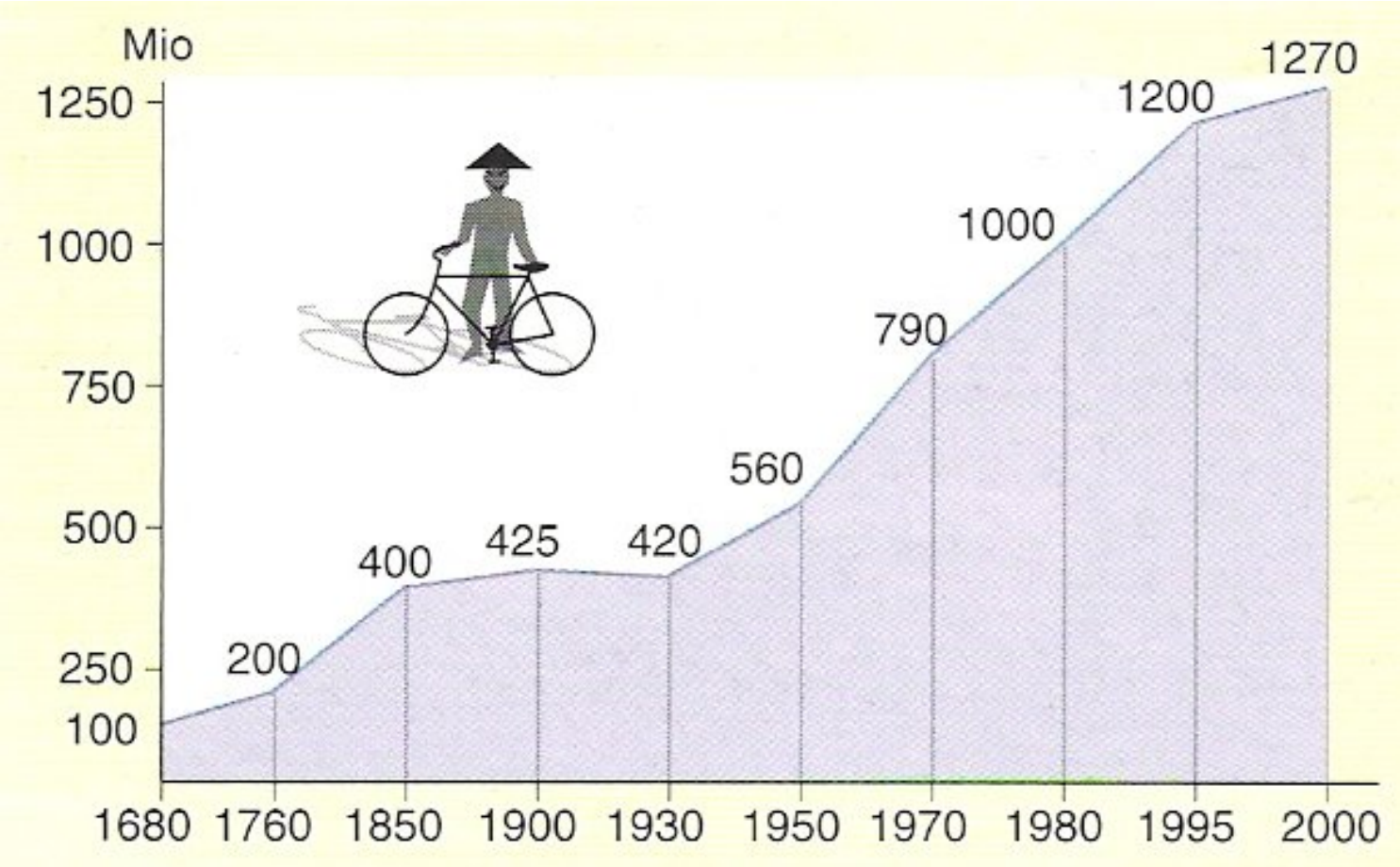
  Erst am Ende wird einschränkend erwähnt:

  „Fairerweise muss man erwähnen, dass Männer täglich deutlich mehr Kilometer zurücklegen. Und: Während 93 Prozent von ihnen einen Führerschein besitzen, sind es bei den Frauen lediglich 82 Prozent.“

Beispiel (Irreführende Grafik) I

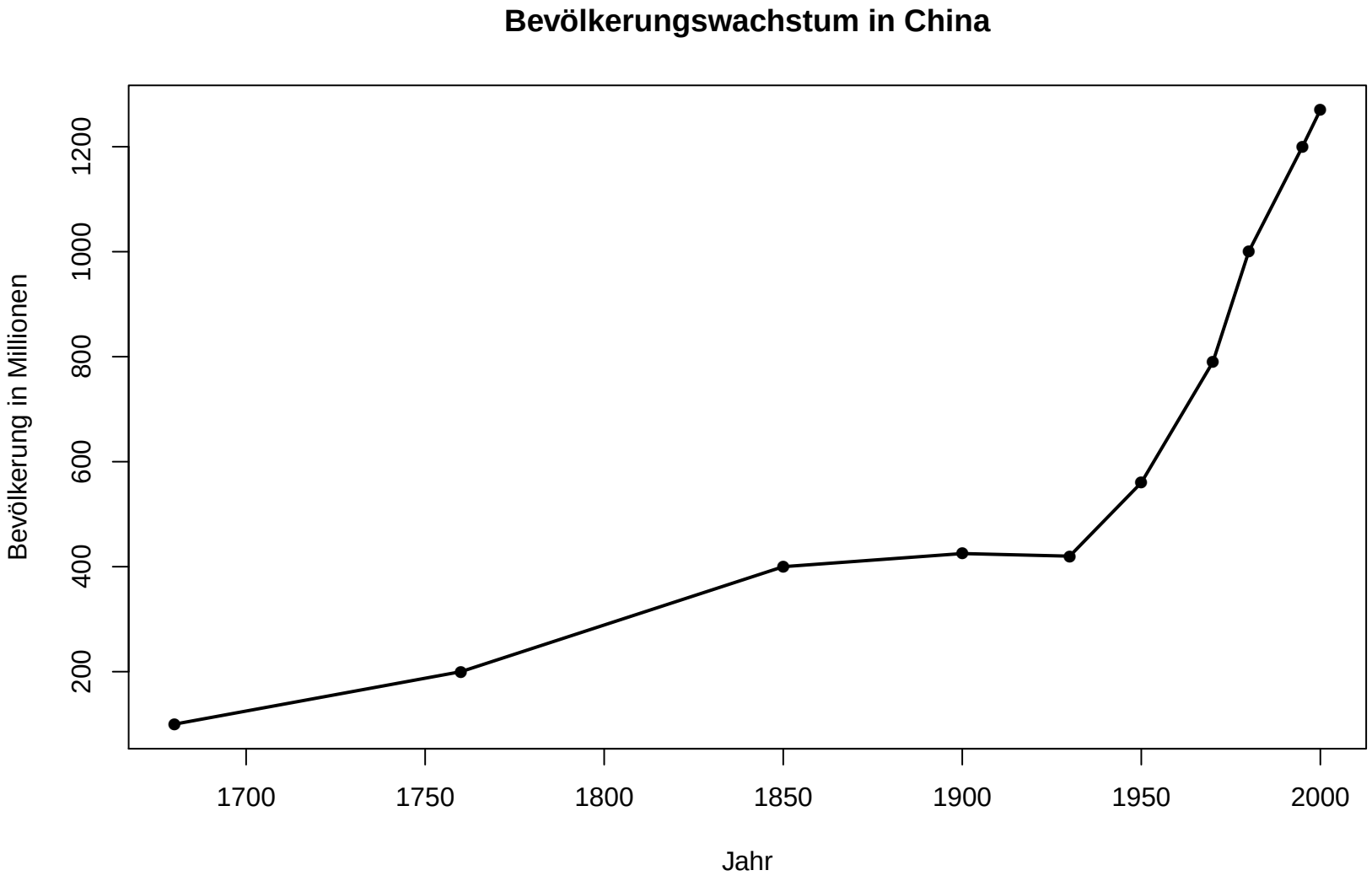
vgl. `http://www.klein-singen.de/statistik/h/Wissenschaft/Bevoelkerungswachstum.html`

Bevölkerungswachstum in China



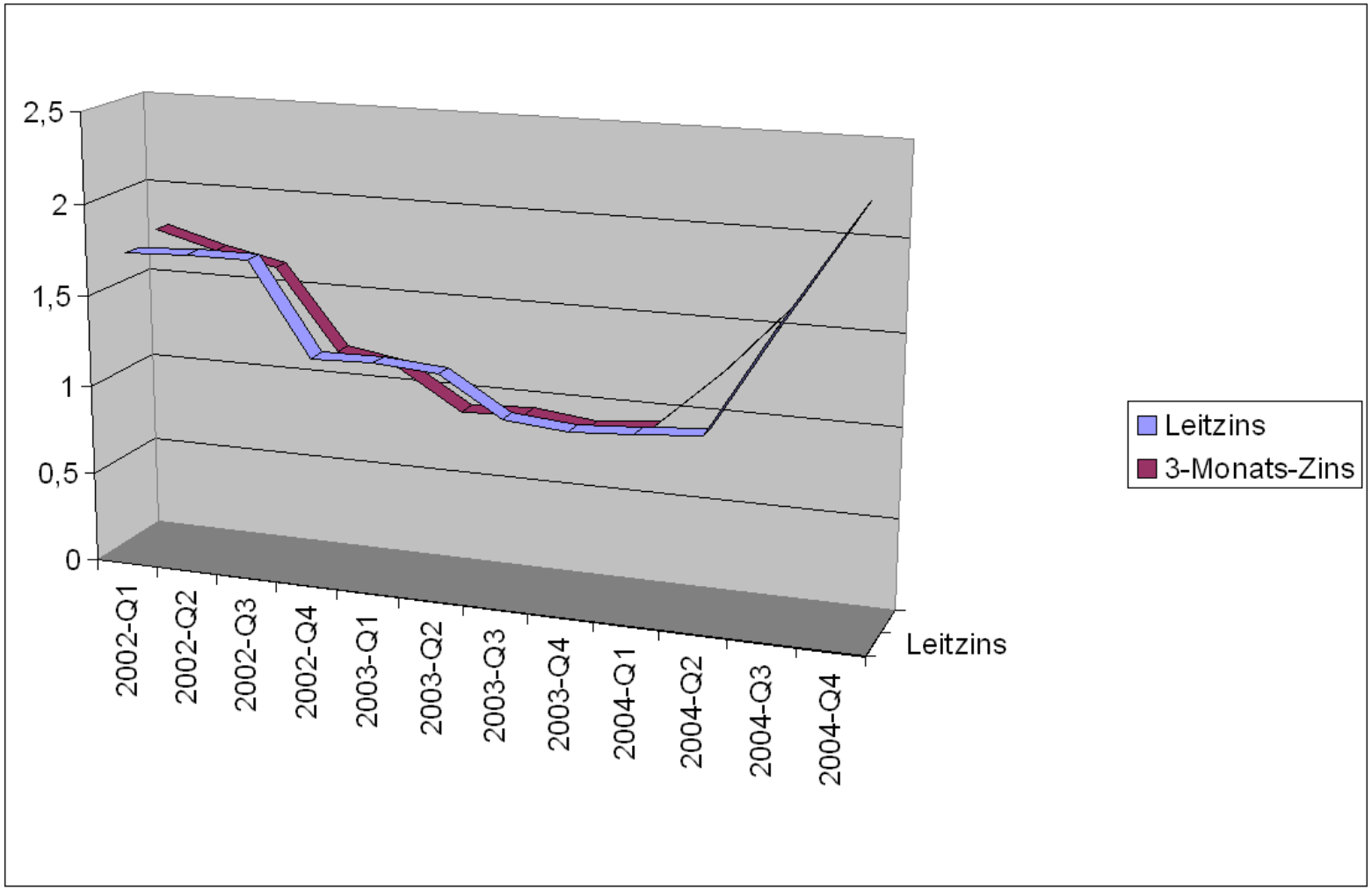
Beispiel (Irreführende Grafik) II

identischer Datensatz, angemessene Skala



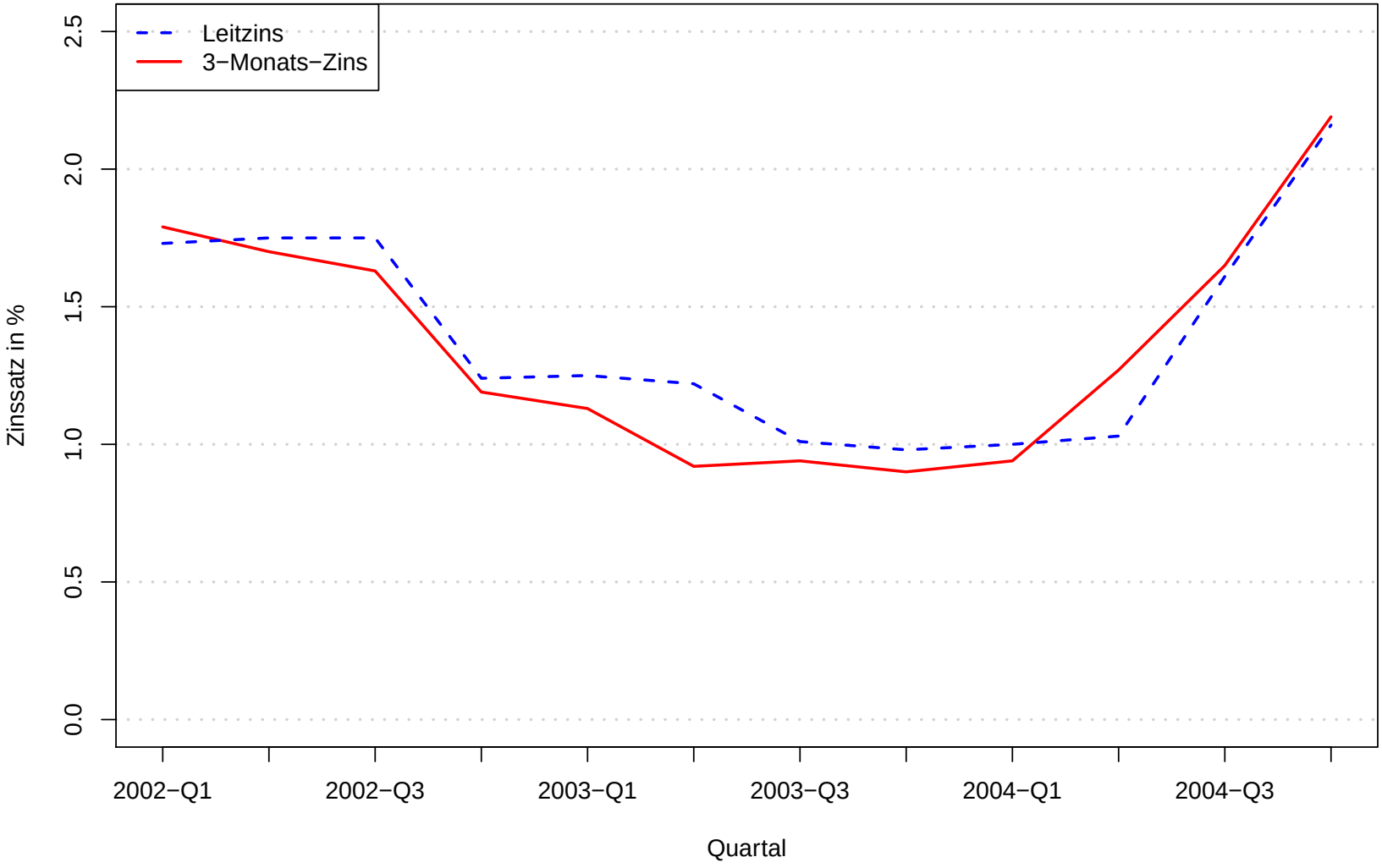
Beispiel (Chartjunk)

Microsoft Excel mit Standardeinstellung für 3D-Liniendiagramme



Beispiel (Grafik ohne Chartjunk)

Statistik-Software **R**, identischer Datensatz



Kann Statistik auch nützlich sein?

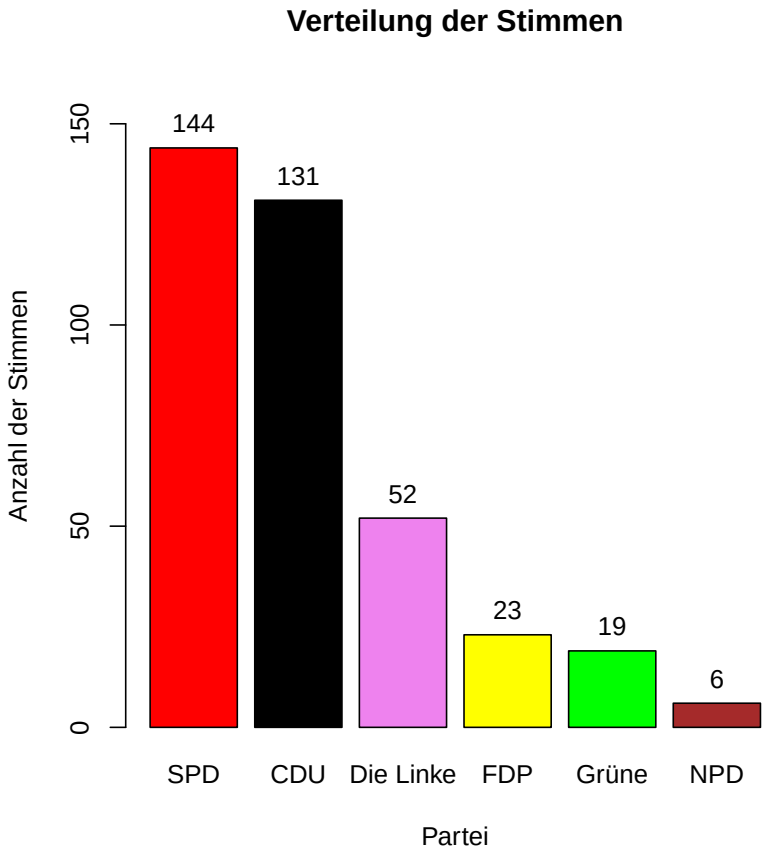
Welche Partei erhält wie viele Stimmen im Wahlbezirk 1.206 der Gemeinde Losheim am See bei den Erststimmen zur Bundestagswahl 2009? Stimmen:

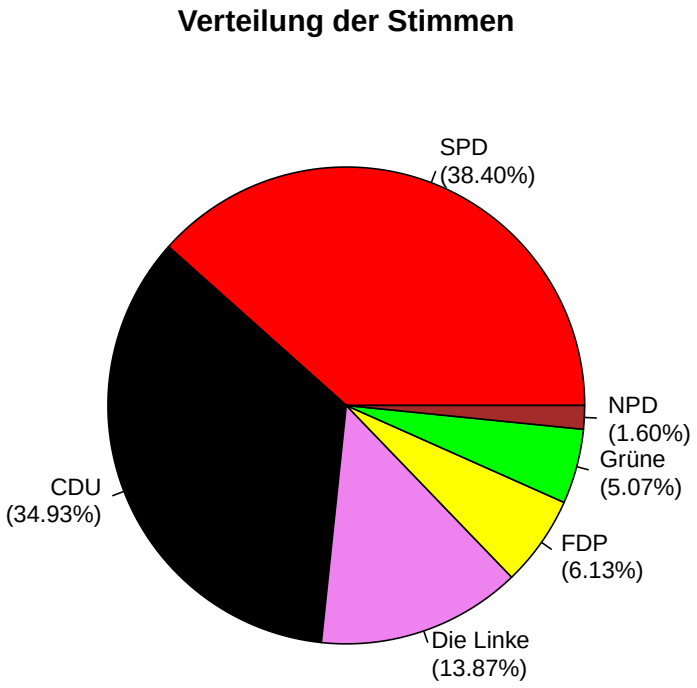
CDU, SPD, SPD, Die Linke, CDU, Die Linke, Die Linke, SPD, SPD, CDU, CDU, CDU, SPD, Grüne, FDP, SPD, SPD, NPD, SPD, FDP, Die Linke, Grüne, Grüne, CDU, CDU, Grüne, CDU, SPD, Die Linke, CDU, SPD, SPD, SPD, CDU, FDP, SPD, SPD, CDU, Grüne, SPD, FDP, SPD, SPD, SPD, SPD, Grüne, CDU, SPD, SPD, SPD, SPD, FDP, SPD, CDU, Grüne, CDU, CDU, SPD, SPD, CDU, FDP, CDU, SPD, CDU, SPD, CDU, SPD, SPD, CDU, SPD, Die Linke, SPD, CDU, Die Linke, Die Linke, Die Linke, SPD, CDU, NPD, FDP, Die Linke, Die Linke, Die Linke, CDU, SPD, CDU, SPD, CDU, CDU, CDU, CDU, CDU, SPD, SPD, SPD, SPD, CDU, CDU, SPD, Die Linke, SPD, CDU, CDU, SPD, SPD, CDU, CDU, FDP, SPD, SPD, Die Linke, SPD, NPD, CDU, CDU, CDU, SPD, CDU, Grüne, SPD, SPD, CDU, CDU, CDU, SPD, SPD, FDP, CDU, CDU, SPD, SPD, CDU, CDU, Die Linke, Die Linke, Die Linke, SPD, SPD, SPD, CDU, SPD, SPD, CDU, CDU, SPD, CDU, FDP, SPD, CDU, SPD, Die Linke, CDU, SPD, Die Linke, CDU, CDU, CDU, FDP, CDU, CDU, CDU, SPD, FDP, SPD, SPD, CDU, CDU, CDU, SPD, SPD, CDU, SPD, SPD, Die Linke, CDU, Grüne, Die Linke, SPD, SPD, SPD, Die Linke, CDU, SPD, SPD, SPD, Die Linke, Die Linke, SPD, SPD, CDU, SPD, CDU, Die Linke, FDP, FDP, CDU, CDU, Die Linke, SPD, SPD, CDU, Die Linke, CDU, SPD, CDU, CDU, SPD, CDU, CDU, SPD, SPD, SPD, SPD, SPD, SPD, CDU, Die Linke, SPD, Die Linke, CDU, SPD, Die Linke, SPD, CDU, Grüne, SPD, Die Linke, CDU, SPD, SPD, CDU, SPD, SPD, SPD, SPD, Grüne, Die Linke, Die Linke, FDP, SPD, CDU, SPD, CDU, SPD, CDU, CDU, Die Linke, Die Linke, SPD, CDU, Grüne, FDP, SPD, SPD, CDU, SPD, CDU, CDU, SPD, CDU, Die Linke, Grüne, Die Linke, Die Linke, Die Linke, SPD, Die Linke, CDU, CDU, CDU, Die Linke, CDU, SPD, Die Linke, Die Linke, SPD, SPD, SPD, SPD, SPD, CDU, SPD, CDU, SPD, Grüne, CDU, CDU, SPD, Die Linke, Grüne, CDU, FDP, Die Linke, Grüne, SPD, CDU, CDU, CDU, SPD, FDP, SPD, Die Linke, SPD, CDU, CDU, FDP, Die Linke, SPD, CDU, NPD, FDP, FDP, SPD, NPD, SPD, SPD, SPD, CDU, CDU, CDU, Grüne, SPD, SPD, SPD, FDP, CDU, CDU, SPD, Die Linke, CDU, Die Linke, SPD, CDU, SPD, Die Linke, CDU, Die Linke, CDU, CDU, CDU, SPD, SPD, SPD, Grüne, SPD, SPD, CDU, FDP, Grüne, CDU, CDU, CDU, CDU, CDU, SPD, NPD, CDU, SPD, CDU, SPD, CDU, SPD, CDU, SPD, SPD, SPD, CDU, CDU, CDU, CDU, Die Linke, CDU, CDU, SPD, CDU

- Mit etwas (deskriptiver) Statistik in tabellarischer Form:

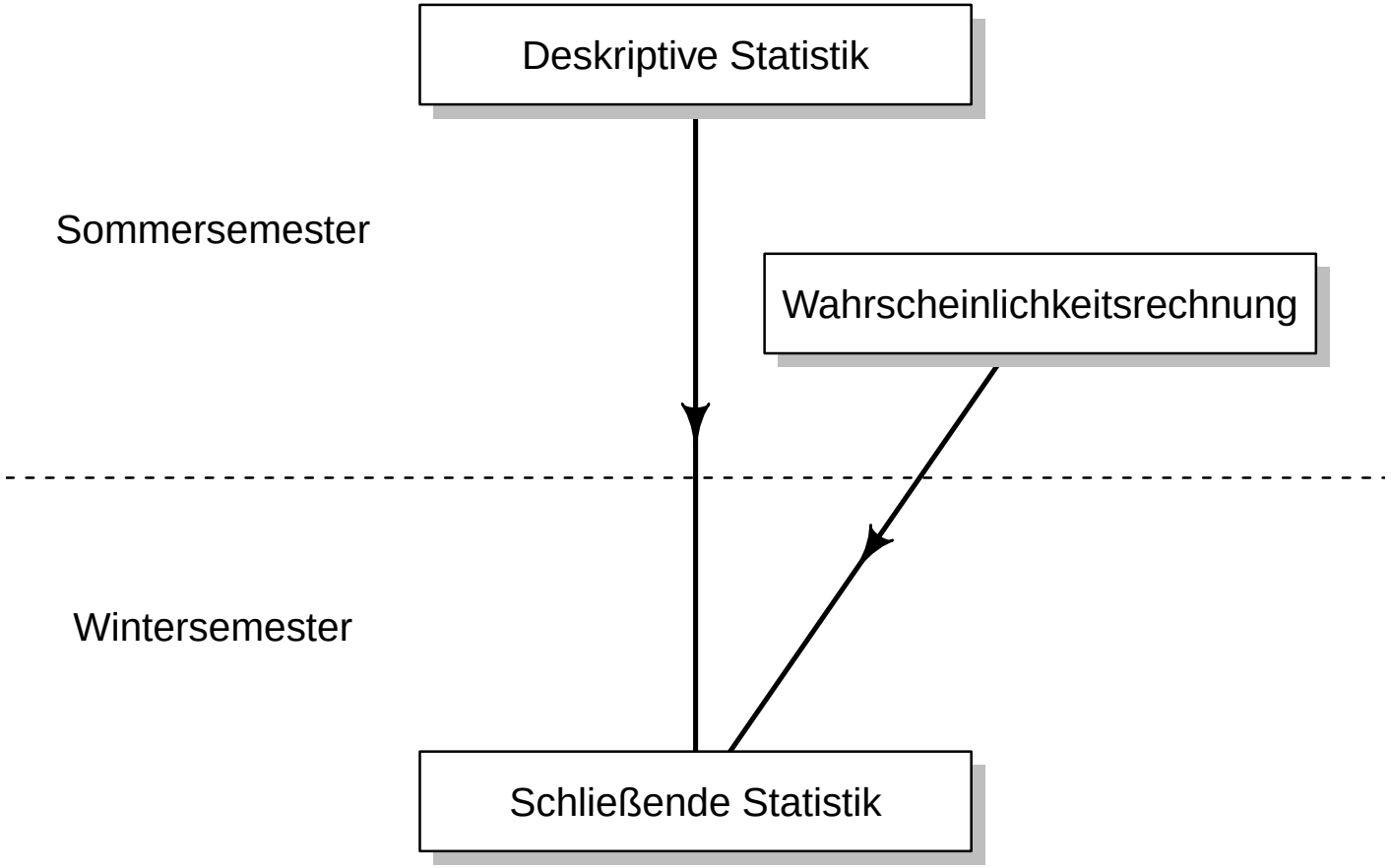
	SPD	CDU	Die Linke	FDP	Grüne	NPD	Summe
Anzahl der Stimmen	144	131	52	23	19	6	375
Stimmenanteil in %	38.40	34.93	13.87	6.13	5.07	1.60	100.00

- Grafisch aufbereitete Varianten:





Organisation der Statistik-Veranstaltungen



Teil I

Deskriptive Statistik

Datenerhebung I

- Beginn jeder (deskriptiven) statistischen Untersuchung: Datenerhebung
- Zu einer **Menge von Merkmalsträgern** (**statistische Masse**), eventuell Teil einer größeren **Grundgesamtheit**, werden ein oder mehrere **Merkmale** erhoben
- Unterscheidung nach
  - Primärerhebung $\leftrightarrow$ Sekundärerhebung:
    Neue Erhebung oder Nutzung von vorhandenem Datenmaterial
  - Vollerhebung $\leftrightarrow$ Teilerhebung:
    Erhebung der Merkmale für ganze Grundgesamtheit oder Teilgesamtheit

Datenerhebung II

- Bei Primärerhebung: Untersuchungsziel bestimmt
  - Auswahl bzw. Abgrenzung der statistischen Masse
  - Auswahl der zu erhebenden Merkmale
  - Art der Erhebung, z.B. Befragung (Post, Telefon, Internet, persönlich), Beobachtung, Experiment
- Sorgfalt bei Datenerhebung enorm wichtig:
  Fehler bei Datenerhebung sind später nicht mehr zu korrigieren!
- Ausführliche Diskussion hier aus Zeitgründen nicht möglich

Vorsicht vor „falschen Schlüssen“! I

- Deskriptive Statistik fasst lediglich Information über statistische Masse zusammen
- Schlüsse auf (größere) „Grundgesamtheit“ (bei Teilerhebung)
  $\rightsquigarrow$ Schließende Statistik
- Dennoch häufig zu beobachten:
  „Informelles“ Übertragen der Ergebnisse in der statistischen Masse auf größere Menge von Merkmalsträgern

$\rightsquigarrow$ Gefahr von falschen Schlüssen!

Vorsicht vor „falschen Schlüssen“! II

Beispiel: Bachelor-Absolventen (vgl. Krämer: So lügt man mit Statistik)

Hätte man am Ende des SS 2011 in der statistischen Masse der Absolventen des BWL-Bachelorstudiengangs in Saarbrücken die Merkmale „Studiendauer“ und „Abschlussnote“ erhoben, würde man wohl feststellen, dass alle Abschlüsse in Regelstudienzeit und im Durchschnitt mit einer guten Note erfolgt sind. Warum? Kann man dies ohne weiteres auf Absolventen anderer Semester übertragen?

$\rightsquigarrow$ Zur Interpretationsfähigkeit von Ergebnissen statistischer Untersuchungen:

- Abgrenzung der zugrundeliegenden statistischen Masse **sehr** wichtig
- (Möglichst) objektive Festlegung nach Kriterien zeitlicher, räumlicher und sachlicher Art

Definition 2.1 (Menge, Mächtigkeit, Tupel)

1. Eine (endliche) **Menge** M ist die Zusammenfassung (endlich vieler) unterschiedlicher Objekte (Elemente).
2. Zu einer endlichen Menge M bezeichnen $\#M$ oder auch $|M|$ die Anzahl der Elemente in M. $\#M$ bzw. $|M|$ heißen auch **Mächtigkeit** der Menge M.
3. Für eine Anzahl $n \geq 1$ von (nicht notwendigerweise verschiedenen!) Elementen $x_1, x_2, \ldots, x_n$ aus einer Menge M wird eine (nach ihrer Reihenfolge geordnete) Auflistung $(x_1, x_2, \ldots, x_n)$ bzw. $x_1, x_2, \ldots, x_n$ als n-**Tupel** aus der Menge M bezeichnet. 2-Tupel (x_1, x_2) heißen auch Paare.
4. Lassen sich die Elemente der Menge M (der Größe nach) ordnen, so sei (zu einer vorgegebenen Ordnung)
   1. mit $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ bzw. $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ das der Größe nach geordnete n-Tupel der n Elemente $x_1, x_2, \ldots, x_n$ aus M bezeichnet, es gelte also $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.
   2. zu einer endlichen Teilmenge $A \subseteq M$ der Mächtigkeit m mit $(a_{(1)}, a_{(2)}, \ldots, a_{(m)})$ bzw. $a_{(1)}, a_{(2)}, \ldots, a_{(m)}$ das der Größe nach geordnete m-Tupel der Elemente $a_1, a_2, \ldots, a_m$ von A bezeichnet, es gelte also $a_{(1)} < a_{(2)} < \cdots < a_{(m)}$.

Merkmalswerte, Merkmalsraum, Urliste I

- Bei der Datenerhebung werden den Merkmalsträgern zu jedem erhobenen Merkmal **Merkmalswerte** oder **Beobachtungswerte** zugeordnet.
- Man nimmt an, dass man (im Prinzip auch vor der Erhebung) eine Menge M angeben kann, die alle vorstellbaren Merkmalswerte eines Merkmals enthält.
- Das n-Tupel $(x_1, \ldots, x_n)$ der Merkmalswerte $x_1, \ldots, x_n$ (aus der Menge M) zu einem bei den n Merkmalsträgern erhobenen Merkmal X bezeichnet man als **Urliste**.
- Die Menge A der (verschiedenen) in der Urliste (tatsächlich) auftretenden Merkmalswerte, in Zeichen

$$A := \{a \in M \mid \exists i \in \{1, \ldots, n\} \text{ mit } x_i = a\} \ ,$$

heißt **Merkmalsraum**, ihre Elemente **Merkmalsausprägungen**.

Merkmalswerte, Merkmalsraum, Urliste II

Beispiel Wahlergebnis

- Urliste (siehe Folie 22) aus gewählten Parteien der 375 abgegebenen gültigen Stimmen:

  $x_1 =$ *"CDU"*, $x_2 =$ *"SPD"*, $x_3 =$ *"SPD"*, $x_4 =$ *"Die Linke"*, $x_5 =$ *"CDU"*, $x_6 =$ *"Die Linke"*, $x_7 =$ *"Die Linke"*, $x_8 =$ *"SPD"*, $x_9 =$ *"SPD"*, $x_{10} =$ *"CDU"*, $x_{11} =$ *"CDU"*, $x_{12} =$ *"CDU"*, $x_{13} =$ *"SPD"*, $x_{14} =$ *"Grüne"*, $x_{15} =$ *"FDP"*, $x_{16} =$ *"SPD"*, $x_{17} =$ *"SPD"*, $x_{18} =$ *"NPD"*, $x_{19} =$ *"SPD"*, $x_{20} =$ *"FDP"*, ...

- Merkmalsraum: $A = \{\text{SPD, CDU, Die Linke, FDP, Grüne, NPD}\}$

Merkmalstypen I

Definition 2.2 (Merkmalstypen)

1. Ein Merkmal heißt
   - **nominalskaliert**, wenn seine Ausprägungen lediglich unterschieden werden sollen,
   - **ordinalskaliert** oder **rangskaliert**, wenn (darüberhinaus) eine (Rang-)Ordnung auf den Ausprägungen vorgegeben ist,
   - **kardinalskaliert** oder **metrisch skaliert**, wenn (darüberhinaus) ein „Abstand“ auf der Menge der Ausprägungen vorgegeben ist, also wenn das Ausmaß der Unterschiede zwischen verschiedenen Ausprägungen gemessen werden kann.
2. Ein Merkmal heißt **quantitativ**, wenn es kardinalskaliert ist, **qualitativ** sonst.
3. Ein Merkmal heißt
   - **diskret**, wenn es qualitativ ist oder wenn es quantitativ ist und die Menge der möglichen Ausprägungen endlich oder abzählbar unendlich ist.
   - **stetig**, wenn es quantitativ ist und für je zwei mögliche Merkmalsausprägungen auch alle Zwischenwerte angenommen werden können.

Merkmalstypen II

- Welche der in Definition 2.2 erwähnten Eigenschaften für ein Merkmal zutreffend sind, hängt von der jeweiligen Anwendungssituation ab.
- Insbesondere ist die Abgrenzung zwischen stetigen und diskreten Merkmalen oft schwierig (allerdings meist auch nicht besonders wichtig).
- Damit ein Merkmal (mindestens) ordinalskaliert ist, muss die verwendete Ordnung — insbesondere bei Mehrdeutigkeit — eindeutig festgelegt sein.
- Häufig findet man zusätzlich zu den in 2.2 erläuterten Skalierungen auch die Begriffe **Intervallskala**, **Verhältnisskala** und **Absolutskala**. Diese stellen eine feinere Unterteilung der Kardinalskala dar.
- *Unabhängig vom Skalierungsniveau* heißt ein Merkmal **numerisch**, wenn seine Merkmalsausprägungen Zahlenwerte sind.

Merkmalstypen III

Beispiel (Merkmalstypen)

- nominalskalierte Merkmale: Geschlecht (Ausprägungen: „männlich“, „weiblich“), Parteien (siehe Wahlergebnis-Beispiel)
- ordinalskalierte Merkmale: Platzierungen, Zufriedenheit („sehr zufrieden“, „eher zufrieden“, „weniger zufrieden“, „unzufrieden“)
- kardinalskalierte Merkmale: Anzahl Kinder, Anzahl Zimmer in Wohnung, Preise, Gewichte, Streckenlängen, Zeiten
  - davon diskret: Anzahl Kinder, Anzahl Zimmer in Wohnung,
  - davon (eher) stetig: Preise, Gewichte, Streckenlängen, Zeiten

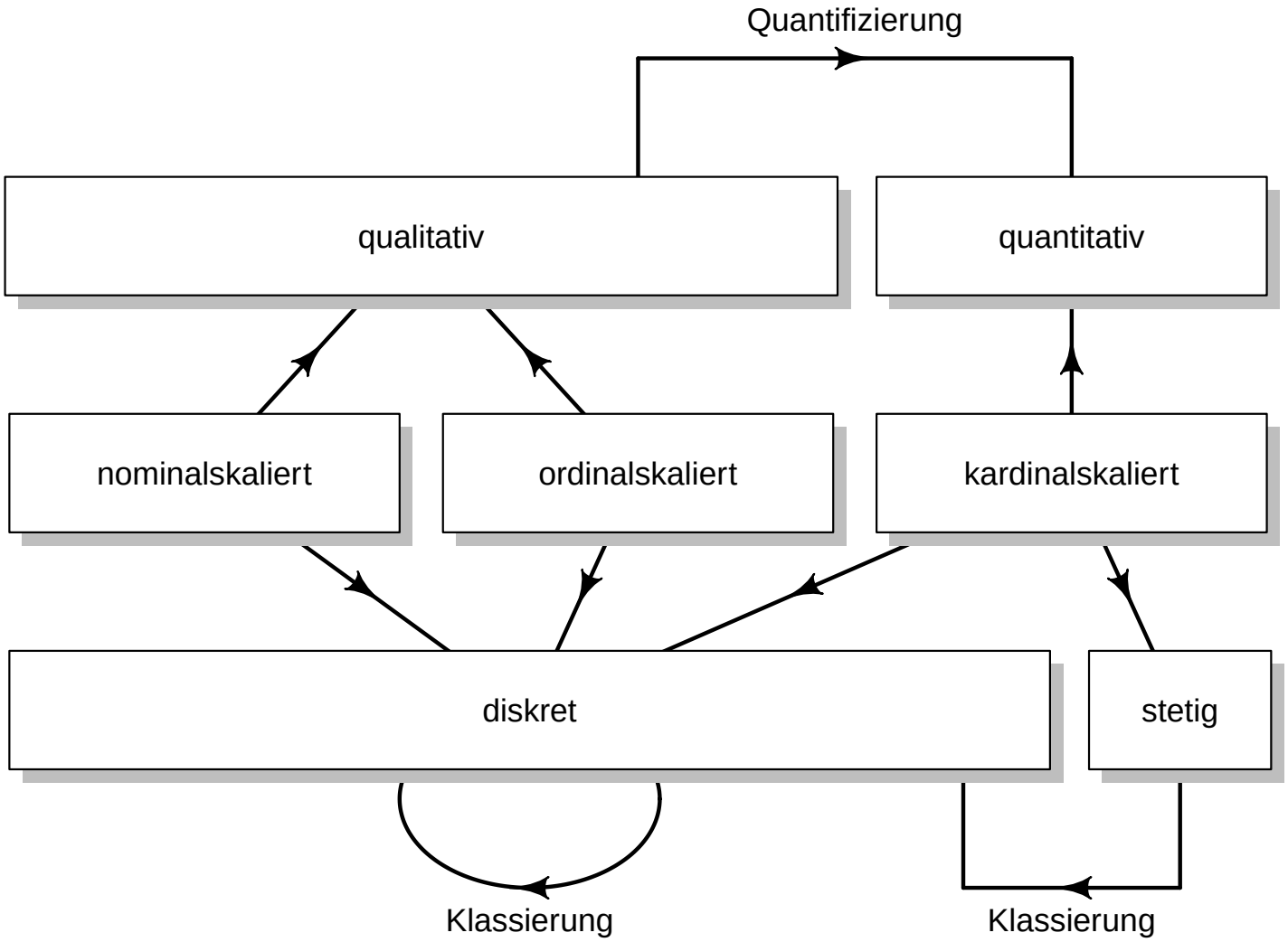
Umwandlung von Merkmalstypen I

- Umwandlung qualitativer in quantititative Merkmale durch **Quantifizierung**:
  - Ersetzen des qualitativen Merkmals „Berufserfahrung“ mit den Ausprägungen „Praktikant“, „Lehrling“, „Geselle“, „Meister“ durch quantitatives Merkmal, dessen Ausprägungen den (mindestens) erforderlichen Jahren an Berufspraxis entspricht, die zum Erreichen des Erfahrungsgrades erforderlich sind.
  - Ersetzen des qualitativen Merkmals Schulnote mit den Ausprägungen „sehr gut“, „gut“, „befriedigend“, „ausreichend“, „mangelhaft“, „ungenügend“ (eventuell feiner abgestuft durch Zusätze „+“ und „-“) durch quantitatives Merkmal, z.B. mit den Ausprägungen $15, 14, \ldots, 00$ oder den Ausprägungen $1.0, 1.3, 1.7, 2.0, 2.3, \ldots, 4.7, 5.0, 6.0$.
  - **Vorsicht:** Umwandlung nur sinnvoll, wenn Abstände tatsächlich (sinnvoll) interpretiert werden können!

Umwandlung von Merkmalstypen II

- Umwandlung stetiger in diskrete Merkmale durch **Klassierung** oder **Gruppierung**, d.h. Zusammenfassen ganzer Intervalle zu einzelnen Ausprägungen, z.B. Gewichtsklassen beim Boxsport.
  - Klassierung ermöglicht auch Umwandlung diskreter Merkmale in (erneut) diskrete Merkmale mit unterschiedlichem Merkmalsraum, z.B. Unternehmensgrößen kleiner und mittlerer Unternehmen nach Anzahl der Beschäftigten mit Ausprägungen „1-9“, „10-19“, „20-49“, „50-249“.
  - Klassierung erfolgt regelmäßig (aber nicht immer) bereits vor der Datenerhebung.

Übersichtsdarstellung Merkmalstypen



Inhaltsverzeichnis

(Ausschnitt)

3 Eindimensionale Daten
- Häufigkeitsverteilungen unklassierter Daten
- Häufigkeitsverteilungen klassierter Daten
- Lagemaße
- Streuungsmaße
- Box-Plot
- Symmetrie- und Wölbungsmaße

Häufigkeitsverteilungen I

- Geeignetes Mittel zur Verdichtung der Information aus Urlisten vor allem bei diskreten Merkmalen mit „wenigen“ Ausprägungen: **Häufigkeitsverteilungen**
- Zur Erstellung einer Häufigkeitsverteilung: Zählen, wie oft jede Merkmalsausprägung a aus dem Merkmalsraum $A = \{a_1, \ldots, a_m\}$ in der Urliste $(x_1, \ldots, x_n)$ vorkommt.
  - Die **absoluten Häufigkeiten** $h(a)$ geben für die Merkmalsausprägung $a \in A$ die (absolute) Anzahl der Einträge der Urliste mit der Ausprägung a an, in Zeichen

$$h(a) := \#\{i \in \{1, \ldots, n\} \mid x_i = a\} \ .$$

  - Die **relativen Häufigkeiten** $r(a)$ geben für die Merkmalsausprägung $a \in A$ den (relativen) Anteil der Einträge der Urliste mit der Ausprägung a an der gesamten Urliste an, in Zeichen

$$r(a) := \frac{h(a)}{n} = \frac{\#\{i \in \{1, \ldots, n\} \mid x_i = a\}}{n} \ .$$

Häufigkeitsverteilungen II

- Die absoluten Häufigkeiten sind natürliche Zahlen und summieren sich zu n auf (i.Z. $\sum_{j=1}^{m} h(a_j) = n$).
- Die relativen Häufigkeiten sind Zahlen zwischen 0 und 1 (bzw. zwischen 0% und 100%) und summieren sich zu 1 (bzw. 100%) auf (i.Z. $\sum_{j=1}^{m} r(a_j) = 1$).
- Ist die Anordnung (Reihenfolge) der Urliste unwichtig, geht durch Übergang zur Häufigkeitsverteilung keine relevante Information verloren.
- Häufigkeitsverteilungen werden in der Regel in tabellarischer Form angegeben, am Beispiel des Wahlergebnisses:

	SPD	CDU	Die Linke	FDP	Grüne	NPD	Summe
a_j	a_1	a_2	a_3	a_4	a_5	a_6	Σ
$h(a_j)$	144	131	52	23	19	6	375
$r(a_j)$	0.3840	0.3493	0.1387	0.0613	0.0507	0.0160	1.0000

Häufigkeitsverteilungen III

- Grafische Darstellung (insbesondere bei nominalskalierten Merkmalen) durch **Balkendiagramme** (auch: Säulendiagramme) oder **Kuchendiagramme** (siehe Folie 23).
- Balkendiagramme meist geeigneter als Kuchendiagramme (außer, wenn die anteilige Verteilung der Merkmalsausprägungen im Vordergrund steht)
- Oft mehrere Anordnungen der Spalten/Balken/Kreissegmente bei nominalskalierten Merkmalen plausibel, absteigende Sortierung nach Häufigkeiten $h(a_j)$ meist sinnvoll.
- Bei ordinalskalierten Merkmalen zweckmäßig: Sortierung der Merkmalsausprägungen nach vorgegebener Ordnung, also

$$a_1 = a_{(1)}, a_2 = a_{(2)}, \ldots, a_m = a_{(m)}$$

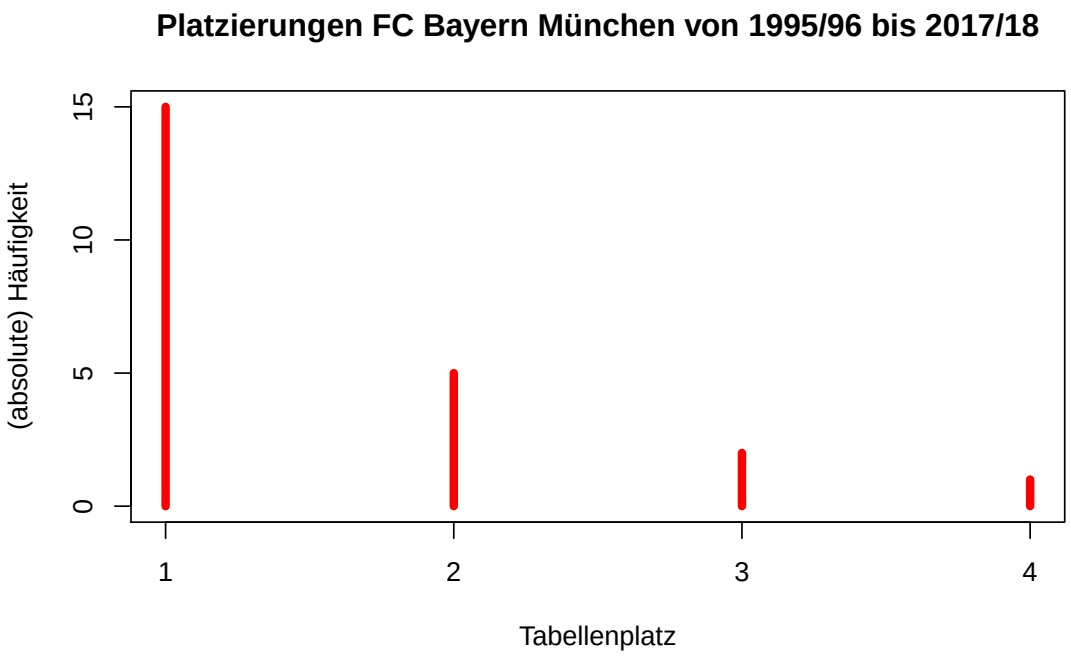
- Alternative grafische Darstellung bei (mindestens) ordinalskalierten Merkmalen mit numerischen Ausprägungen: **Stabdiagramm**

Häufigkeitsverteilungen IV

- Stabdiagramm zur Urliste

$$2, 1, 2, 1, 1, 1, 3, 1, 2, 1, 1, 4, 1, 2, 1, 3, 2, 1, 1, 1, 1, 1, 1$$

der finalen Tabellenplätze des FC Bayern München in der (ersten) Fußball-Bundesliga (Saison 1995/96 bis 2017/2018):



Empirische Verteilungsfunktion

- Bei (mindestens ordinalskalierten) numerischen Merkmalen interessante Fragestellungen:
  - Wie viele Merkmalswerte sind kleiner/größer als ein vorgegebener Wert?
  - Wie viele Merkmalswerte liegen in einem vorgegebenem Bereich (Intervall)?
- Hierzu nützlich: **(relative) kumulierte Häufigkeitsverteilung**, auch bezeichnet als **empirische Verteilungsfunktion**
- Die empirische Verteilungsfunktion $F(x)$ ordnet einer Zahl x den Anteil der Merkmalswerte $x_1, \ldots, x_n$ zu, die kleiner oder gleich x sind, also

$$F(x) := \frac{\#\{i \in \{1, \ldots, n\} \mid x_i \leq x\}}{n} \ .$$

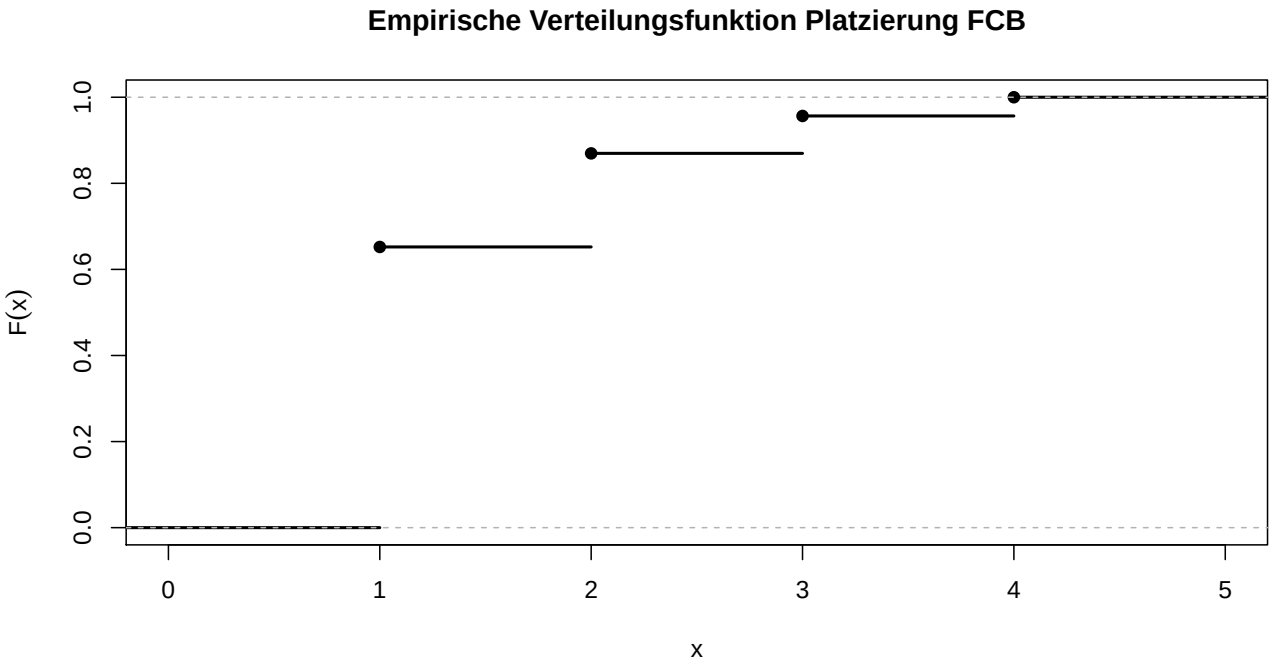
- Ein Vergleich mit den Definitionen von $h(a)$ und $r(a)$ offenbart (!), dass $F(x)$ auch mit Hilfe von $h(a)$ bzw. $r(a)$ berechnet werden kann; gibt es m Merkmalsausprägungen, so gilt:

$$F(x) = \frac{1}{n} \sum_{\substack{a_j \leq x \\ 1 \leq j \leq m}} h(a_j) = \sum_{\substack{a_j \leq x \\ 1 \leq j \leq m}} r(a_j)$$

- Beispiel: Empirische Verteilungsfunktion für FC Bayern-Platzierungen

$$F(x) = \begin{cases} 0 & \text{für} \quad x < 1 \\ \frac{15}{23} & \text{für} \quad 1 \leq x < 2 \\ \frac{20}{23} & \text{für} \quad 2 \leq x < 3 \\ \frac{22}{23} & \text{für} \quad 3 \leq x < 4 \\ 1 & \text{für} \quad x \geq 4 \end{cases} \approx \begin{cases} 0.000 & \text{für} \quad x < 1 \\ 0.652 & \text{für} \quad 1 \leq x < 2 \\ 0.870 & \text{für} \quad 2 \leq x < 3 \\ 0.957 & \text{für} \quad 3 \leq x < 4 \\ 1.000 & \text{für} \quad x \geq 4 \end{cases}$$

- Grafische Darstellung der empirischen Verteilungsfunktion:



Relative Häufigkeiten von Intervallen I

(bei numerischen Merkmalen)

- Relative Häufigkeit $r(a)$ ordnet Ausprägungen $a \in A$ zugehörigen Anteil von a an den Merkmalswerten zu.
- $r(\,\cdot\,)$ kann auch für $x \in \mathbb{R}$ mit $x \notin A$ ausgewertet werden ($\rightsquigarrow r(x) = 0$).
- „Erweiterung“ von $r(\,\cdot\,)$ auch auf Intervalle möglich:
- $F(b)$ gibt für $b \in \mathbb{R}$ bereits Intervallhäufigkeit

$$F(b) = r\left((-\infty, b]\right) = r\left(\{x \in \mathbb{R} \mid x \leq b\}\right)$$

an.

Relative Häufigkeiten von Intervallen II

(bei numerischen Merkmalen)

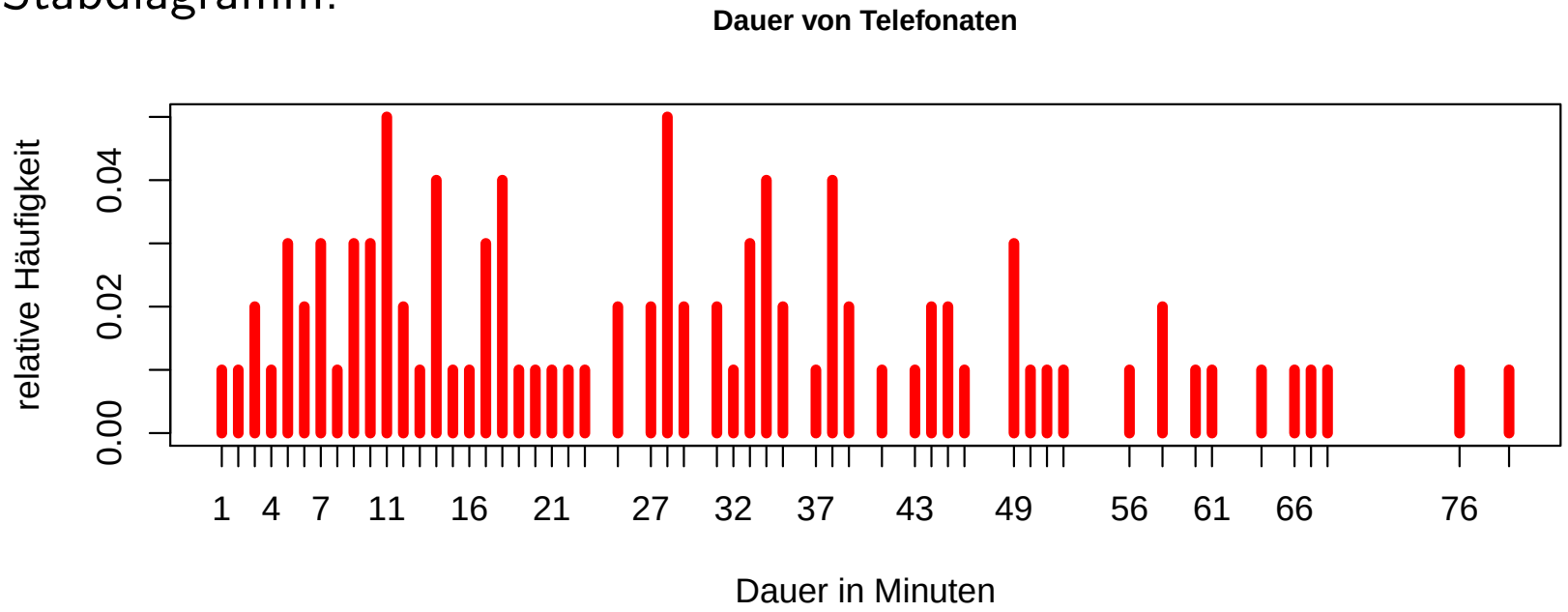
- Relative Häufigkeit des offenen Intervalls $(-\infty, b)$ als Differenz

$$r\left((-\infty, b)\right) = r\left((-\infty, b]\right) - r(b) = F(b) - r(b)$$

- Analog: relative Häufigkeiten weiterer Intervalle:
  - $r\left((a, \infty)\right) = 1 - F(a)$
  - $r\left([a, \infty)\right) = 1 - (F(a) - r(a)) = 1 - F(a) + r(a)$
  - $r\left([a, b]\right) = F(b) - (F(a) - r(a)) = F(b) - F(a) + r(a)$
  - $r\left((a, b]\right) = F(b) - F(a)$
  - $r\left([a, b)\right) = (F(b) - r(b)) - (F(a) - r(a)) = F(b) - r(b) - F(a) + r(a)$
  - $r\left((a, b)\right) = (F(b) - r(b)) - F(a) = F(b) - r(b) - F(a)$

Häufigkeitsverteilungen klassierter Daten I

- Bisherige Analysemethoden schlecht geeignet für stetige Merkmale bzw. diskrete Merkmale mit „vielen“ Ausprägungen
- (Fiktives) Beispiel: Dauer von 100 Telefonaten (in Minuten)
  - Urliste: 44, 35, 22, 5, 50, 5, 3, 17, 19, 67, 49, 52, 16, 34, 11, 27, 14, 1, 35, 11, 3, 49, 18, 58, 43, 34, 79, 34, 7, 38, 28, 21, 27, 51, 9, 17, 10, 60, 14, 32, 9, 18, 11, 23, 25, 10, 76, 28, 13, 15, 28, 7, 31, 45, 66, 61, 39, 25, 17, 33, 4, 41, 29, 38, 18, 44, 28, 12, 64, 6, 38, 8, 37, 38, 28, 5, 7, 34, 11, 2, 31, 14, 33, 39, 12, 49, 14, 58, 45, 56, 46, 68, 18, 6, 11, 10, 29, 33, 9, 20
  - Stabdiagramm:



Häufigkeitsverteilungen klassierter Daten II

- **Problem:** viele Merkmalswerte treten nur einmalig (oder „selten“) auf
  $\rightsquigarrow$ Aussagekraft von Häufigkeitstabellen und Stabdiagrammen gering
- **Lösung:** Zusammenfassen mehrerer Merkmalsausprägungen in Klassen
- Zu dieser **Klassierung** erforderlich: **Vorgabe** der Grenzen $k_0, k_1, \ldots, k_l$ von l (rechtsseitig abgeschlossenen) Intervallen

$$K_1 := (k_0, k_1],\ K_2 := (k_1, k_2],\ \ldots,\ K_l := (k_{l-1}, k_l]\ ,$$

die alle n Merkmalswerte überdecken
(also mit $k_0 < x_i \leq k_l$ für alle $i \in \{1, \ldots, n\}$)

Häufigkeitsverteilungen klassierter Daten III

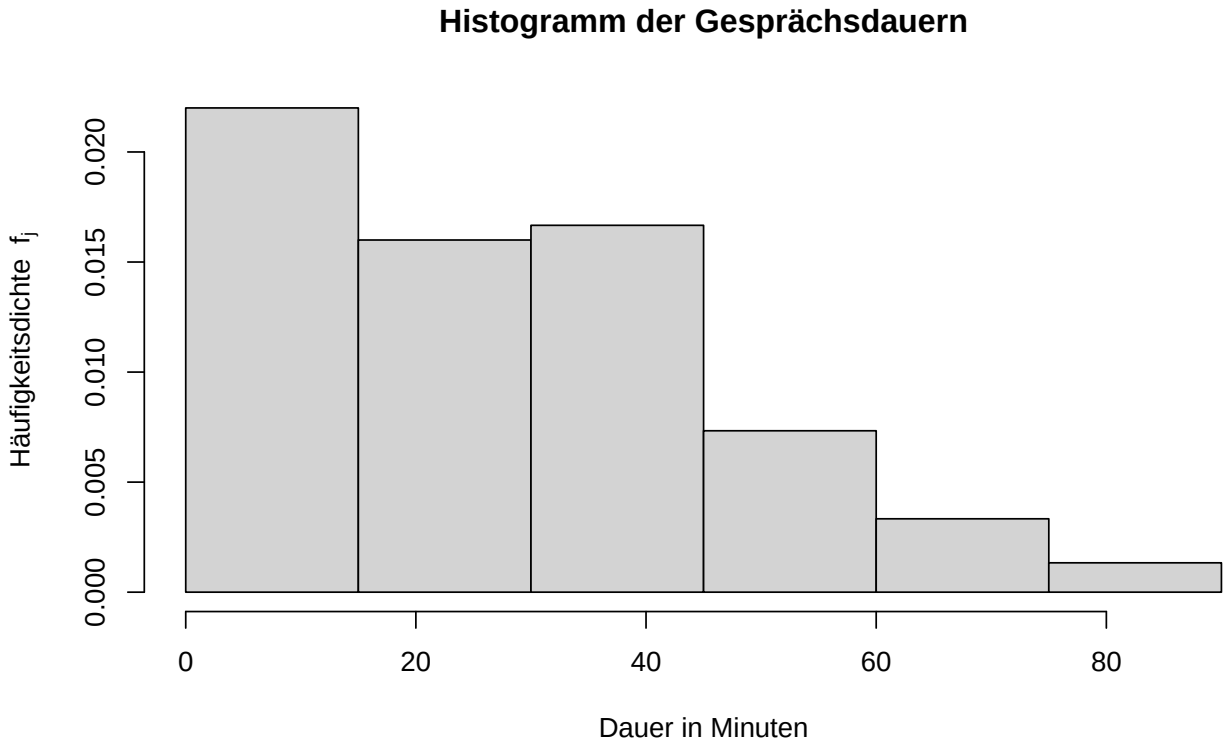
- Wichtige Kennzahlen der Klassierung (bzw. der klassierten Daten):

  Klassenbreiten $b_j := k_j - k_{j-1}$
  Klassenmitten $m_j := \frac{k_{j-1}+k_j}{2}$
  absolute Häufigkeiten $h_j := \#\{i \in \{1,\ldots,n\} \mid k_{j-1} < x_i \leq k_j\}$
  relative Häufigkeiten $r_j := \frac{h_j}{n}$
  Häufigkeitsdichten $f_j := \frac{r_j}{b_j}$

  (jeweils für $j \in \{1,\ldots,l\}$).
- Übliche grafische Darstellung von klassierten Daten: **Histogramm**
- Hierzu: Zeichnen der Rechtecke mit Höhen f_j über den Intervallen K_j (also der Rechtecke mit den Eckpunkten $(k_{j-1}, 0)$ und (k_j, f_j))

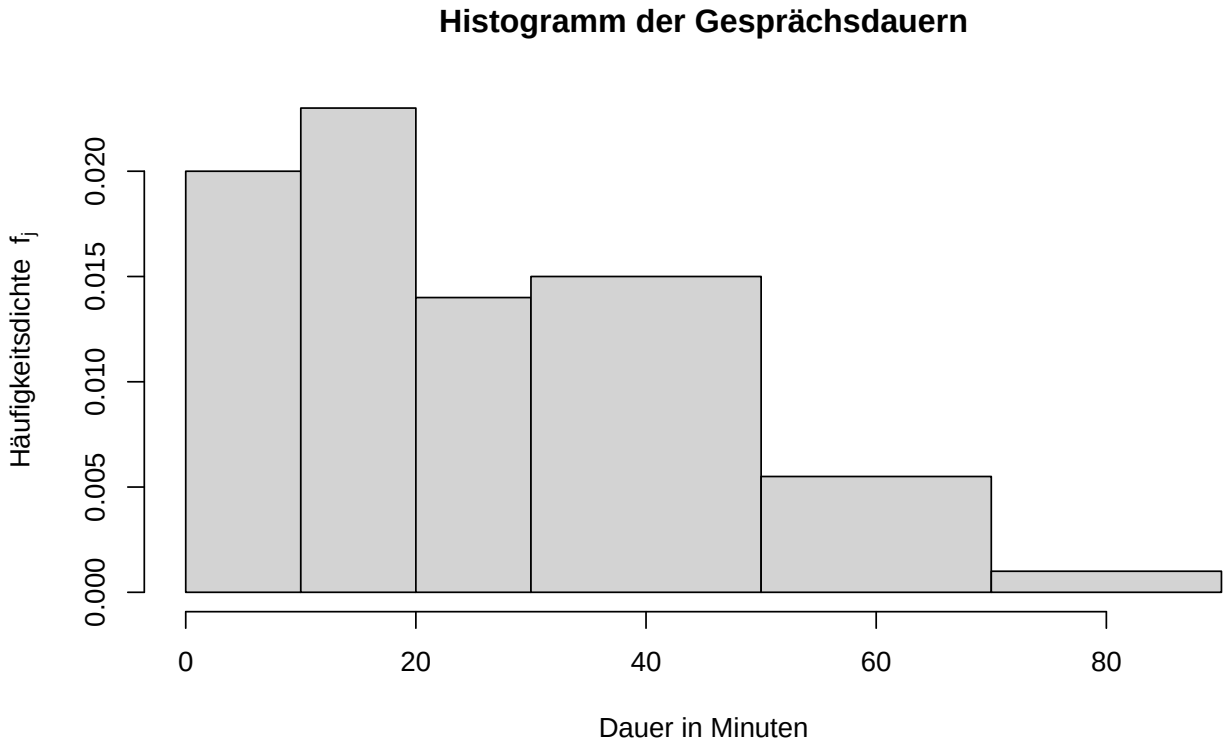
- Am Beispiel der Gesprächsdauern bei 6 Klassen zu je 15 Minuten Breite:

Nr. j	Klasse $K_j = (k_{j-1}, k_j]$	Klassenbreite b_j	Klassenmitte m_j	absolute Häufigkeit h_j	relative Häufigkeit $r_j = \frac{h_j}{n}$	Häufigkeitsdichte $f_j = \frac{r_j}{b_j}$	Verteilungsfunktion $F(k_j)$
1	$(0, 15]$	15	7.5	33	0.33	0.022	0.33
2	$(15, 30]$	15	22.5	24	0.24	0.016	0.57
3	$(30, 45]$	15	37.5	25	0.25	$0.01\overline{6}$	0.82
4	$(45, 60]$	15	52.5	11	0.11	$0.007\overline{3}$	0.93
5	$(60, 75]$	15	67.5	5	0.05	$0.00\overline{3}$	0.98
6	$(75, 90]$	15	82.5	2	0.02	$0.001\overline{3}$	1.00



- Alternativ mit 6 Klassen bei 2 verschiedenen Breiten:

Nr. j	Klasse $K_j = (k_{j-1}, k_j]$	Klassenbreite b_j	Klassenmitte m_j	absolute Häufigkeit h_j	relative Häufigkeit $r_j = \frac{h_j}{n}$	Häufigkeitsdichte $f_j = \frac{r_j}{b_j}$	Verteilungsfunktion $F(k_j)$
1	$(0, 10]$	10	5	20	0.20	0.0200	0.20
2	$(10, 20]$	10	15	23	0.23	0.0230	0.43
3	$(20, 30]$	10	25	14	0.14	0.0140	0.57
4	$(30, 50]$	20	40	30	0.30	0.0150	0.87
5	$(50, 70]$	20	60	11	0.11	0.0055	0.98
6	$(70, 90]$	20	80	2	0.02	0.0010	1.00



Bemerkungen I

- Der **Flächeninhalt** der einzelnen Rechtecke eines Histogramms entspricht der relativen Häufigkeit der zugehörigen Klasse
  - $\rightsquigarrow$ Die Summe aller Flächeninhalte beträgt 1
  - $\rightsquigarrow$ Die Höhe der Rechtecke ist nur dann proportional zu der relativen Häufigkeit der Klassen, falls alle Klassen die gleiche Breite besitzen!
- Die Klassierung ist abhängig von der Wahl der Klassengrenzen, unterschiedliche Klassengrenzen können einen Datensatz auch sehr unterschiedlich erscheinen lassen $\rightsquigarrow$ Potenzial zur Manipulation
- Es existieren verschiedene Algorithmen zur automatischen Wahl von Klassenanzahl und -grenzen (z.B. nach Scott, Sturges, Freedman-Diaconis)

Bemerkungen II

- Durch Klassierung geht Information verloren!
  - Spezielle Verfahren für klassierte Daten vorhanden
  - Verfahren approximieren ursprüngliche Daten in der Regel durch die Annahme gleichmäßiger Verteilung innerhalb der einzelnen Klassen
  - (Approximative) Verteilungsfunktion (ebenfalls mit $F(x)$ bezeichnet) zu klassierten Daten entsteht so durch lineare Interpolation der an den Klassengrenzen k_j bekannten (und auch nach erfolgter Klassierung noch exakten!) Werte der empirischen Verteilungsfunktion $F(k_j)$
  - Näherungsweise Berechnung von Intervallhäufigkeiten dann gemäß Folie 46 f. mit der approximativen empirischen Verteilungsfunktion $F(x)$.

(Approx.) Verteilungsfunktion bei klassierten Daten

Approximative Verteilungsfunktion bei klassierten Daten

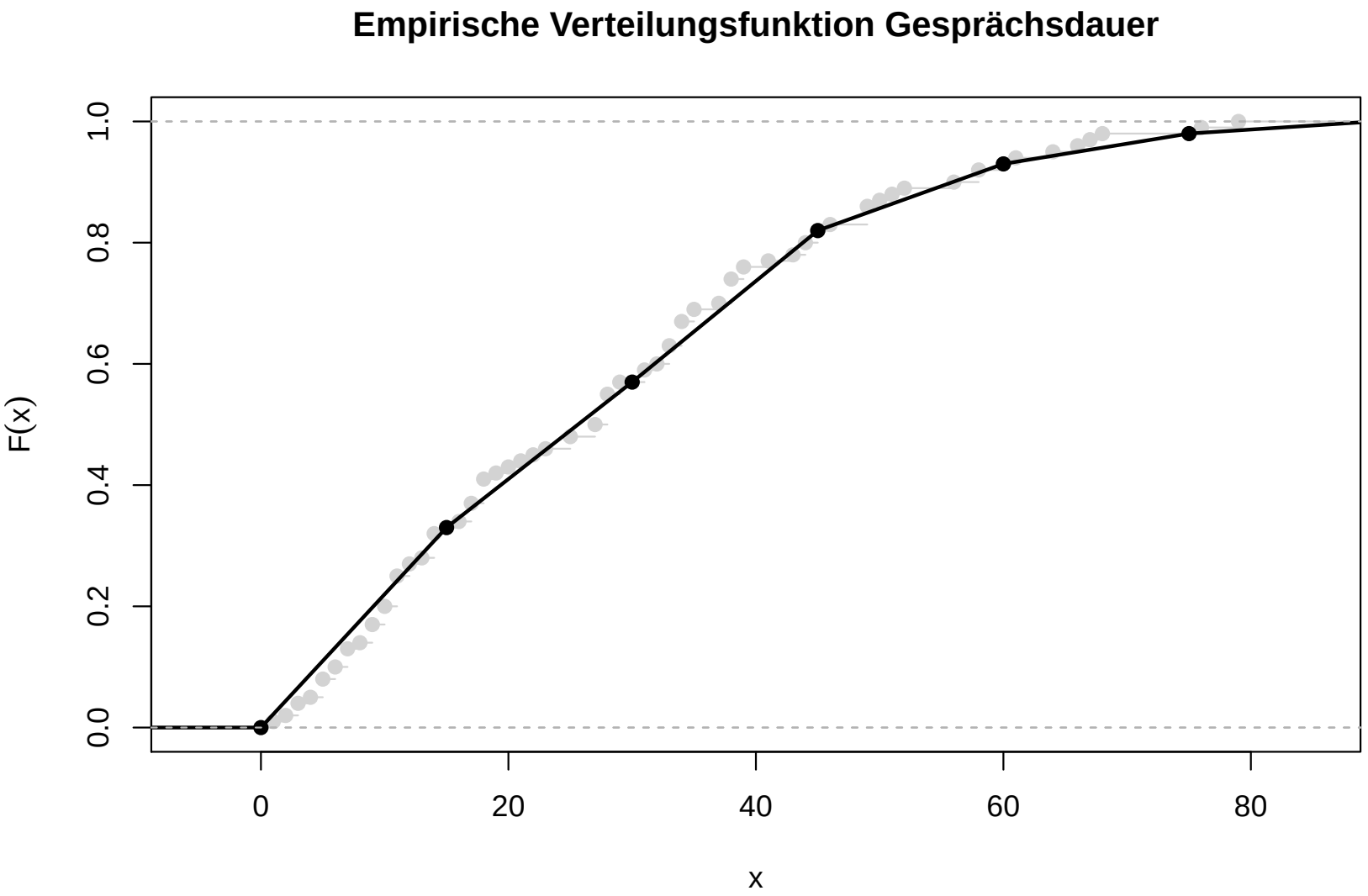
$$F(x) = \begin{cases} 0 & \text{für} \quad x \leq k_0 \\ F(k_{j-1}) + f_j \cdot (x - k_{j-1}) & \text{für} \quad k_{j-1} < x \leq k_j,\ j \in \{1, \ldots, l\} \\ 1 & \text{für} \quad x > k_l \end{cases}$$

- Am Beispiel der Gesprächdauern (Klassierung aus Folie 52)

$$F(x) = \begin{cases} 0 & \text{für} \quad x \leq 0 \\ 0.0200 \cdot (x - 0) & \text{für} \quad 0 < x \leq 10 \\ 0.20 + 0.0230 \cdot (x - 10) & \text{für} \quad 10 < x \leq 20 \\ 0.43 + 0.0140 \cdot (x - 20) & \text{für} \quad 20 < x \leq 30 \\ 0.57 + 0.0150 \cdot (x - 30) & \text{für} \quad 30 < x \leq 50 \\ 0.87 + 0.0055 \cdot (x - 50) & \text{für} \quad 50 < x \leq 70 \\ 0.98 + 0.0010 \cdot (x - 70) & \text{für} \quad 70 < x \leq 90 \\ 1 & \text{für} \quad x > 90 \end{cases}$$

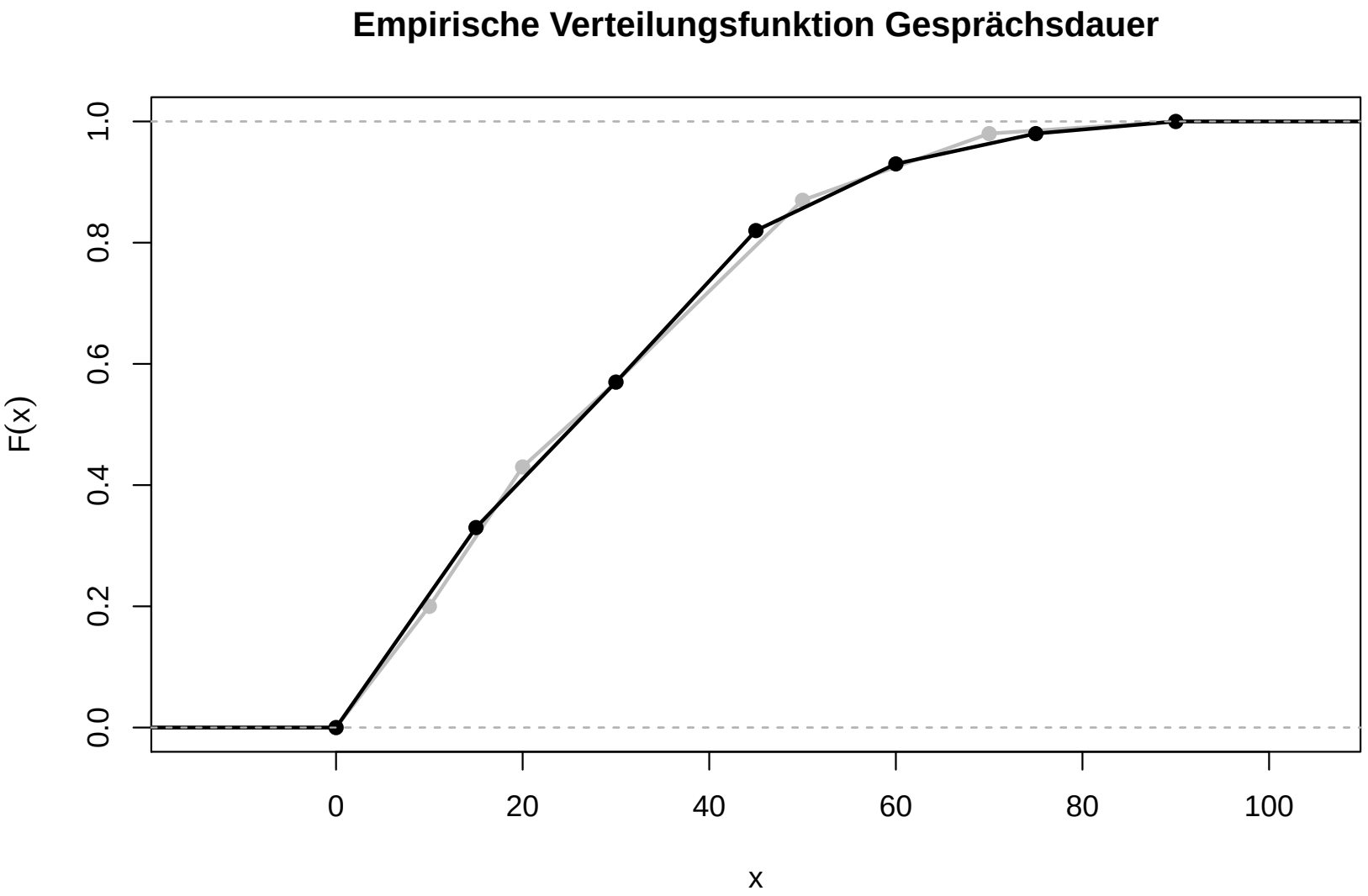
Grafik: Verteilungsfunktion bei klassierten Daten

(Empirische Verteilungsfunktion der unklassierten Daten in hellgrau)



Grafik: Verteilungsfunktion bei verschiedenen Klassierungen

(Klassierung aus Folie 51 in schwarz, Klassierung aus Folie 52 in grau)



Lagemaße

- Aggregation von Merkmalswerten zu Häufigkeitsverteilungen (auch nach erfolgter Klassierung) nicht immer ausreichend.
- Häufig gewünscht: einzelner Wert, der die Verteilung der Merkmalswerte geeignet charakterisiert $\rightsquigarrow$ „Mittelwert“
- **Aber:**
  - Gibt es immer einen „Mittelwert“?
    Was ist der Mittelwert der Merkmalswerte *rot, gelb, gelb, blau*?
    $\rightsquigarrow$ allgemeinerer Begriff: „Lagemaß“
  - Gibt es verschiedene „Mittelwerte“?
    Falls ja, welcher der Mittelwerte ist (am Besten) geeignet?

Lagemaße für nominalskalierte Merkmale

- Verschiedene Merkmalsausprägungen können lediglich unterschieden werden
- „Typische“ Merkmalswerte sind also solche, die häufig vorkommen
- Geeignetes Lagemaß: häufigster Wert (es kann mehrere geben!)

Definition 3.1 (Modus, Modalwert)

Sei X ein (mindestens) nominalskaliertes Merkmal mit Merkmalsraum $A = \{a_1, \ldots, a_m\}$ und relativer Häufigkeitsverteilung r.
Dann heißt jedes Element $a_{\text{mod}} \in A$ mit

$$r(a_{\text{mod}}) \geq r(a_j) \text{ für alle } j \in \{1, \ldots, m\}$$

Modus oder **Modalwert** von X.

- Beispiele:
  - Modus der Urliste *rot, gelb, gelb, blau*:
    $a_{\text{mod}} = \text{gelb}$
  - Modalwerte der Urliste $1, 5, 3, 3, 4, 2, 6, 7, 6, 8$:
    $a_{\text{mod},1} = 3$ und $a_{\text{mod},2} = 6$

Lagemaße für ordinalskalierte Merkmale I

- Durch die vorgegebene Anordnung auf der Menge der möglichen Ausprägungen M lässt sich der Begriff „mittlerer Wert“ mit Inhalt füllen.
- In der geordneten Folge von Merkmalswerten

$$x_{(1)}, x_{(2)}, \ldots, x_{(n-1)}, x_{(n)}$$

  bietet sich als Lagemaß also ein Wert „in der Mitte“ der Folge an.
- Ist n gerade, gibt es keine eindeutige Mitte der Folge, und eine zusätzliche Regelung ist erforderlich.

Lagemaße für ordinalskalierte Merkmale II

Definition 3.2 (Median)

Sei X ein (mindestens) ordinalskaliertes Merkmal auf der Menge der vorstellbaren Merkmalsausprägungen M und $x_{(1)}, x_{(2)}, \ldots, x_{(n-1)}, x_{(n)}$ die gemäß der vorgegebenen Ordnung sortierte Urliste zum Merkmal X.

- Ist n ungerade, so heißt $x_{\left(\frac{n+1}{2}\right)}$ der **Median** von X, in Zeichen $x_{\text{med}} = x_{\left(\frac{n+1}{2}\right)}$.
- Ist n gerade, so heißen alle (möglicherweise viele verschiedene) Elemente von M *zwischen* (bezogen auf die auf M gegebene Ordnung) $x_{\left(\frac{n}{2}\right)}$ und $x_{\left(\frac{n}{2}+1\right)}$ (einschließlich dieser beiden Merkmalswerte) **Mediane** von X.

- Bei stetigen Merkmalen kann für die Definition des Medians auch für gerades n Eindeutigkeit erreicht werden, indem spezieller der Mittelwert

$$\frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$$

der beiden „mittleren“ Merkmalswerte als Median festgelegt wird.

Lagemaße für ordinalskalierte Merkmale III

- Beispiele:
  - Ist $M = \{\text{sehr gut}, \text{gut}, \text{befriedigend}, \text{ausreichend}, \text{mangelhaft}, \text{ungenügend}\}$ als Menge der möglichen Ausprägungen eines ordinalskalierten Merkmals X mit der üblichen Ordnung von Schulnoten von „sehr gut“ bis „ungenügend“ versehen, so ist die sortierte Folge von Merkmalswerten zur Urliste

    gut, ausreichend, sehr gut, mangelhaft, mangelhaft, gut

    durch

    sehr gut, gut, gut, ausreichend, mangelhaft, mangelhaft

    gegeben und sowohl „gut“ als auch „befriedigend“ und „ausreichend“ sind Mediane von X.
  - Der oben beschriebenen Konvention für stetige Merkmale folgend ist der Median des stetigen Merkmals zur Urliste

    $$1.85,\ 6.05,\ 7.97,\ 11.16,\ 17.19,\ 18.87,\ 19.82,\ 26.95,\ 27.25,\ 28.34$$

    von 10 Merkmalsträgern durch $x_{\text{med}} = \frac{1}{2} \cdot (17.19 + 18.87) = 18.03$ gegeben.

Lagemaße für kardinalskalierte Merkmale

- Bei kardinalskalierten Merkmalen ist oft eine „klassische“ Mittelung der Merkmalswerte als Lagemaß sinnvoll, man erhält so aus der Urliste $x_1, \ldots, x_n$ das **„arithmetische Mittel“** $\overline{x} := \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- *Beispiel:*
  Die Haushalts-Nettoeinkommen (in €) von 6 Haushalten eines Mehrparteien-Wohnhauses sind:

Haushalt	1	2	3	4	5	6
Nettoeinkommen	1000	400	1500	2900	1800	2600

  Frage: Wie groß ist das durchschnittliche Nettoeinkommen?
  Antwort: $\frac{1}{6} \cdot (1000 + 400 + 1500 + 2900 + 1800 + 2600) = 1700$
- Bei klassierten Daten wird der Mittelwert als gewichtetes arithmetisches Mittel der l Klassenmitten näherungsweise berechnet:

$$\overline{x} := \frac{1}{n}\sum_{j=1}^{l} h_j \cdot m_j = \sum_{j=1}^{l} r_j \cdot m_j \ .$$

- Arithmetisches Mittel für viele (nicht alle!) Anwendungen adäquates „Mittel“
- *Beispiel:*
  Ein Wachstumssparvertrag legt folgende Zinssätze fest:

Jahr	1	2	3	4	5
Zinssatz	1.5%	1.75%	2.0%	2.5%	3.5%

  Wie groß ist der Zinssatz *im Durchschnitt*?
  - Aus Zinsrechnung bekannt: Kapital K inkl. Zinsen nach 5 Jahren bei Startkapital S beträgt

$$K = S \cdot (1 + 0.015) \cdot (1 + 0.0175) \cdot (1 + 0.02) \cdot (1 + 0.025) \cdot (1 + 0.035)$$

  - Gesucht ist (für 5 Jahre gleichbleibender) Zinssatz R, der gleiches Endkapital K produziert, also R mit der Eigenschaft

$$K \stackrel{!}{=} S \cdot (1 + R) \cdot (1 + R) \cdot (1 + R) \cdot (1 + R) \cdot (1 + R)$$

  - Ergebnis:
    $R = \sqrt[5]{(1 + 0.015) \cdot (1 + 0.0175) \cdot (1 + 0.02) \cdot (1 + 0.025) \cdot (1 + 0.035)} - 1$
    $\rightsquigarrow \quad R = 2.2476\%.$
- Der in diesem Beispiel für die Bruttorenditen (1+Zinssatz) sinnvolle Mittelwert heißt **„geometrisches Mittel“**.

- *Beispiel:*
  Auf einer Autofahrt von insgesamt 30 [km] werden $s_1 = 10$ [km] mit einer Geschwindigkeit von $v_1 = 30$ [km/h], $s_2 = 10$ [km] mit einer Geschwindigkeit von $v_2 = 60$ [km/h] und $s_3 = 10$ [km] mit einer Geschwindigkeit von $v_3 = 120$ [km/h] zurückgelegt.
  Wie hoch ist die durchschnittliche Geschwindigkeit?
  - Durchschnittliche Geschwindigkeit: Quotient aus Gesamtstrecke und Gesamtzeit
  - Gesamtstrecke: $s_1 + s_2 + s_3 = 10\ [\text{km}] + 10\ [\text{km}] + 10\ [\text{km}] = 30\ [\text{km}]$
  - Zeit für Streckenabschnitt: Quotient aus Streckenlänge und Geschwindigkeit
  - Einzelzeiten also:

$$\frac{s_1}{v_1} = \frac{10\ [\text{km}]}{30\ [\text{km/h}]},\ \frac{s_2}{v_2} = \frac{10\ [\text{km}]}{60\ [\text{km/h}]} \text{ und } \frac{s_3}{v_3} = \frac{10\ [\text{km}]}{120\ [\text{km/h}]}$$

  $\rightsquigarrow$ Durchschnittsgeschwindigkeit

$$\frac{s_1 + s_2 + s_3}{\frac{s_1}{v_1} + \frac{s_2}{v_2} + \frac{s_3}{v_3}} = \frac{30\ [\text{km}]}{\frac{10}{30}\ [\text{h}] + \frac{10}{60}\ [\text{h}] + \frac{10}{120}\ [\text{h}]} = \frac{30}{\frac{7}{12}}\ [\text{km/h}] = 51.429\ [\text{km/h}]$$

- Der in diesem Beispiel für die Geschwindigkeiten sinnvolle Mittelwert heißt **„harmonisches Mittel“**.

Zusammenfassung: Mittelwerte I

Definition 3.3 (Mittelwerte)

Seien $x_1, x_2, \ldots, x_n$ die Merkmalswerte zu einem kardinalskalierten Merkmal X. Dann heißt

1. $\overline{x} := \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ das arithmetische Mittel,
2. $\overline{x}^{(g)} := \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ das geometrische Mittel,
3. $\overline{x}^{(h)} := \dfrac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)} = \dfrac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}}$ das harmonische Mittel

von $x_1, \ldots, x_n$.

Zusammenfassung: Mittelwerte II

Bemerkung 3.4

Liegt die absolute (bzw. relative) Häufigkeitsverteilung h (bzw. r) eines kardinalskalierten Merkmals X mit Merkmalsraum $A = \{a_1, \ldots, a_m\}$ vor, so gilt

1. $\overline{x} = \frac{1}{n} \sum_{j=1}^{m} h(a_j) \cdot a_j = \sum_{j=1}^{m} r(a_j) \cdot a_j$
2. $\overline{x}^{(g)} = \sqrt[n]{\prod_{j=1}^{m} a_j^{h(a_j)}} = \prod_{j=1}^{m} a_j^{r(a_j)}$
3. $\overline{x}^{(h)} = \dfrac{1}{\frac{1}{n} \sum_{j=1}^{m} \frac{h(a_j)}{a_j}} = \dfrac{n}{\sum_{j=1}^{m} \frac{h(a_j)}{a_j}} = \dfrac{1}{\sum_{j=1}^{m} \frac{r(a_j)}{a_j}}$

- Die in Bemerkung 3.4 berechneten Mittelwerte können als sogenannte *gewichtete Mittelwerte* der aufgetreten Merkmalswerte $a_1, \ldots, a_m$ aufgefasst werden, wobei die Gewichte durch die absoluten Häufigkeiten $h(a_1), \ldots, h(a_m)$ (bzw. durch die relativen Häufigkeiten $r(a_1), \ldots, r(a_m)$) der aufgetretenen Merkmalswerte gegeben sind.

Weitere Beispiele I

- Pauschale Aussagen, wann welcher Mittelwert geeignet ist, nicht möglich!
- *Beispiel Zinssätze:*
  Aufgrund Begrenzungen der europäischen Einlagensicherung möchte ein Anleger Kapital von 100 000 € gleichmäßig auf 5 Banken verteilen, die für die vorgegebene Anlagedauer folgende Zinsen anbieten:

Bank	1	2	3	4	5
Zinssatz	2.5%	2.25%	2.4%	2.6%	2.55%

  Frage: Wie groß ist der durchschnittliche Zinssatz?
  Antwort: $\frac{1}{5} \cdot (2.5\% + 2.25\% + 2.4\% + 2.6\% + 2.55\%) = 2.46\%$

Weitere Beispiele II

- *Beispiel Geschwindigkeiten:*
  Auf einer Autofahrt von insgesamt 30 [Min.] Fahrzeit werden $t_1 = 10$ [Min.] mit einer Geschwindigkeit von $v_1 = 30$ [km/h], $t_2 = 10$ [Min.] mit $v_2 = 60$ [km/h] und $t_3 = 10$ [Min.] mit $v_3 = 120$ [km/h] zurückgelegt.
  Wie hoch ist die durchschnittliche Geschwindigkeit?
  - Durchschnittliche Geschwindigkeit: Quotient aus Gesamtstrecke und -zeit
  - Gesamtzeit: $t = t_1 + t_2 + t_3 = 10\ [\text{Min.}] + 10\ [\text{Min.}] + 10\ [\text{Min.}] = 30\ [\text{Min.}]$
  - Länge der Streckenabschnitte: Produkt aus Geschwindigkeit und Fahrzeit

  $\rightsquigarrow$ Durchschnittsgeschwindigkeit

$$\frac{v_1 \cdot t_1 + v_2 \cdot t_2 + v_3 \cdot t_3}{t} = \frac{1}{3} \cdot 30\ [\text{km/h}] + \frac{1}{3} \cdot 60\ [\text{km/h}] + \frac{1}{3} \cdot 120\ [\text{km/h}] = 70\ [\text{km/h}]$$

Bemerkungen I

- Insbesondere bei diskreten Merkmalen wie z.B. einer Anzahl muss der erhaltene (arithmetische, geometrische, harmonische) Mittelwert weder zum Merkmalsraum A noch zur Menge der vorstellbaren Merkmalsausprägungen M gehören (z.B. „im Durchschnitt 2.2 Kinder pro Haushalt“).
- Auch der/die Median(e) gehören (insbesondere bei numerischen Merkmalen) häufiger nicht zur Menge A der Merkmalsausprägungen; lediglich der/die Modalwert(e) kommen stets auch in der Liste der Merkmalswerte vor!
- **Vorsicht** vor falschen Rückschlüssen vom Mittelwert auf die Häufigkeitsverteilung!

Bemerkungen II

Mobilfunknutzung Europa in 2006

In einem Online-Artikel der Zeitschrift „Computerwoche“ vom 03.04.2007 (siehe `http://www.computerwoche.de/a/statistik-jeder-europaeer-telefoniert-mobil,590888`) wird aus der Tatsache, dass die Anzahl der Mobiltelefone in Europa größer ist als die Anzahl der Europäer, also das arithmetische Mittel des Merkmals *Anzahl Mobiltelefone pro Person* in Europa größer als 1 ist, die folgende Aussage in der Überschrift abgeleitet:

Statistik: Jeder Europäer telefoniert mobil

Zusammenfassend heißt es außerdem:

Laut einer aktuellen Studie telefoniert jeder Europäer mittlerweile mit mindestens einem Mobiltelefon.

Wie sind diese Aussagen zu beurteilen? Welcher Fehlschluss ist gezogen worden?

Optimalitätseigenschaften

einiger Lagemaße bei kardinalskalierten Daten

- Für kardinalskalierte Merkmale besitzen Mediane und arithmetische Mittelwerte spezielle (Optimalitäts-)Eigenschaften.
- Für jeden Median x_{med} eines Merkmals X mit den n Merkmalswerten $x_1, \ldots, x_n$ gilt:

$$\sum_{i=1}^{n} |x_i - x_{\text{med}}| \leq \sum_{i=1}^{n} |x_i - t| \quad \text{für alle } t \in \mathbb{R}$$

- Für das arithmetische Mittel $\overline{x}$ eines Merkmals X mit den n Merkmalswerten $x_1, \ldots, x_n$ gilt:
  1. $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$
  2. $\sum_{i=1}^{n} (x_i - \overline{x})^2 \leq \sum_{i=1}^{n} (x_i - t)^2$ für alle $t \in \mathbb{R}$

Weitere Lagemaße: Quantile/Perzentile I

- Für jeden Median x_{med} gilt: Mindestens 50% der Merkmalswerte sind kleiner gleich x_{med} und ebenso mindestens 50% größer gleich x_{med}.
- Verallgemeinerung dieser Eigenschaft auf beliebige Anteile geläufig, also auf Werte, zu denen mindestens ein Anteil p kleiner gleich und ein Anteil $1-p$ größer gleich ist, sog. p-**Quantilen** (auch p-**Perzentile**) x_p.
- Mediane sind dann gleichbedeutend mit 50%-Quantilen bzw. 0.5-Quantilen, es gilt also insbesondere bei eindeutigen Medianen

$$x_{\text{med}} = x_{0.5} \ .$$

Weitere Lagemaße: Quantile/Perzentile II

Definition 3.5 (Quantile/Perzentile, Quartile)

Sei X ein (mindestens) ordinalskaliertes Merkmal auf der Menge der vorstellbaren Merkmalsausprägungen M mit den Merkmalswerten $x_1, \ldots, x_n$.
Für $0 < p < 1$ heißt jeder Wert $x_p \in M$ mit der Eigenschaft

$$\frac{\#\{i \in \{1, \ldots, n\} \mid x_i \leq x_p\}}{n} \geq p \quad \text{und} \quad \frac{\#\{i \in \{1, \ldots, n\} \mid x_i \geq x_p\}}{n} \geq 1 - p$$

p-**Quantil** (auch p-**Perzentil**) von X. Man bezeichnet spezieller das 0.25-Quantil $x_{0.25}$ als **unteres Quartil** sowie das 0.75-Quantil $x_{0.75}$ als **oberes Quartil**.

Weitere Lagemaße: Quantile/Perzentile III

- p-Quantile kann man auch mit der emp. Verteilungsfunktion F bestimmen:
- Mit der Abkürzung

$$F(x-0) := \lim_{\substack{h \to 0 \\ h > 0}} F(x-h), \qquad x \in \mathbb{R},$$

für linksseitige Grenzwerte empirischer Verteilungsfunktionen F ist x_p ist genau dann ein p-Quantil, wenn gilt:

$$F(x_p - 0) \leq p \leq F(x_p)$$

- Spezieller ist x_p genau dann ein p-Quantil, wenn
  - bei Vorliegen der exakten Häufigkeitsverteilung r und Verteilungsfunktion F

$$F(x_p) - r(x_p) \leq p \leq F(x_p) \ ,$$

  - bei Verwendung der approximativen Verteilungsfunktion F bei klassierten Daten (wegen der Stetigkeit der Approximation!)

$$F(x_p) = p$$

gilt.

Weitere Lagemaße: Quantile/Perzentile IV

- Genauso wie der Median muss ein p-Quantil nicht eindeutig bestimmt sein.
- Bei stetigen Merkmalen kann Eindeutigkeit *zum Beispiel* durch die gängige Festlegung

$$x_p = \begin{cases} x_{(\lfloor n \cdot p \rfloor + 1)} & \text{für} \quad n \cdot p \notin \mathbb{N} \\ \frac{1}{2} \cdot \left(x_{(n \cdot p)} + x_{(n \cdot p + 1)} \right) & \text{für} \quad n \cdot p \in \mathbb{N} \end{cases}$$

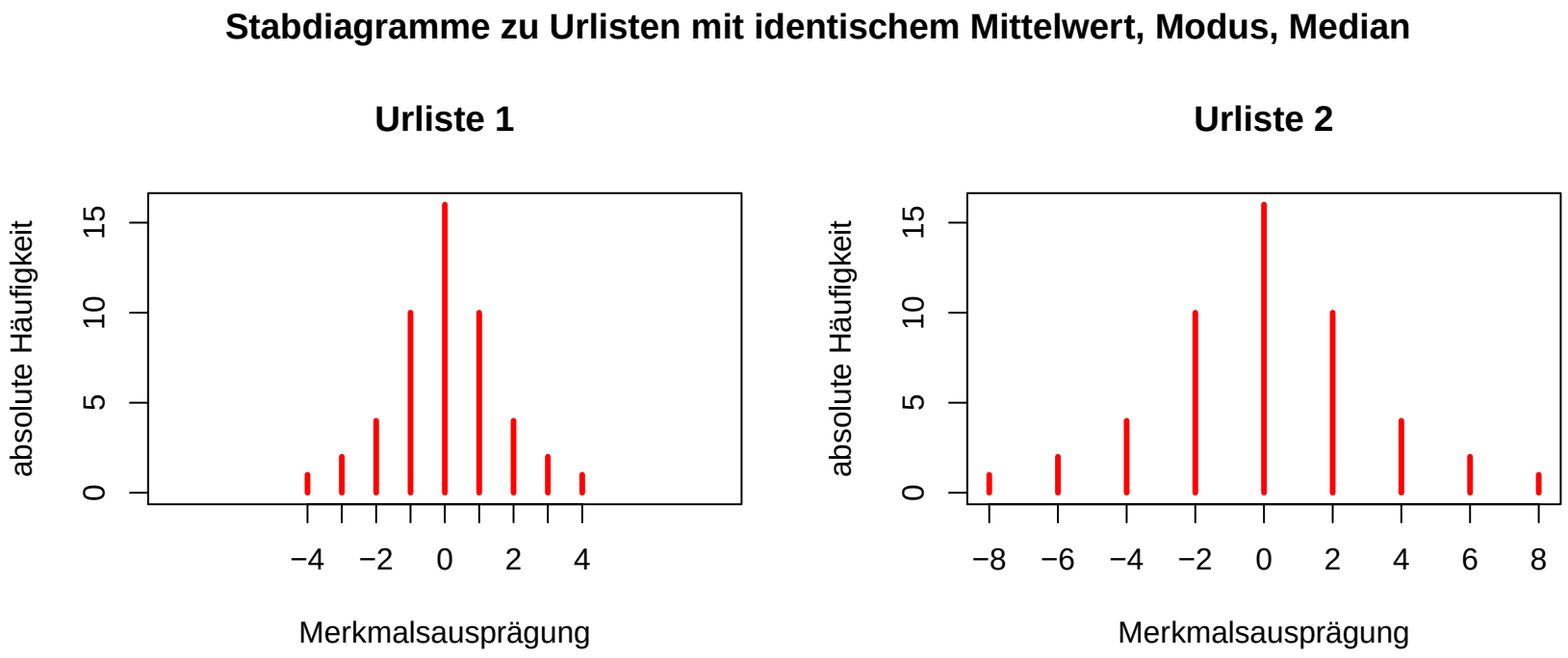
  erreicht werden, wobei $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ die gemäß der vorgegebenen Ordnung sortierte Urliste ist und mit $\lfloor y \rfloor$ für $y \in \mathbb{R}$ die größte ganze Zahl kleiner gleich y bezeichnet wird.
- Zum Beispiel ist für die (bereits sortierte) Urliste

$$6.77, 7.06, 8.84, 9.98, 11.87, 12.18, 12.7, 14.92$$

  der Länge $n = 8$ das 0.25-Quantil $x_{0.25}$ wegen $n \cdot p = 8 \cdot 0.25 = 2 \in \mathbb{N}$ nicht eindeutig bestimmt, sondern alle Werte $x_{0.25} \in [7.06, 8.84]$ sind 0.25-Quantile. Die eindeutige Festlegung nach obiger Konvention würde dann die „Auswahl“ $x_{0.25} = \frac{1}{2}(7.06 + 8.84) = 7.95$ treffen.

Streuungsmaße I

- Verdichtung der Merkmalswerte auf einen Lageparameter als einzige Kennzahl recht unspezifisch.
- Starke Unterschiede trotz übereinstimmender Lagemaße möglich:



Streuungsmaße II

- Bei kardinalskalierten Merkmalen: zusätzliche Kennzahl für Variation bzw. Streuung der Merkmalswerte von Interesse
- Ähnlich wie bei Lagemaßen: verschiedene Streuungsmaße gängig
- Allen Streuungsmaßen gemeinsam:
  Bezug zu „Abstand“ zwischen Merkmalswerten
- *Ein* möglicher Abstand: (Betrag der) Differenz zwischen Merkmalswerten

Streuungsmaße III

Definition 3.6 (Spannweite, IQA, mittlere abs. Abweichung)

Seien $x_1, \ldots, x_n$ die Urliste zu einem kardinalskalierten Merkmal X, x_{med} der Median und $x_{0.25}$ bzw. $x_{0.75}$ das untere bzw. obere Quartil von X.
Dann heißt

1. $SP := \left(\max_{i \in \{1,\ldots,n\}} x_i\right) - \left(\min_{i \in \{1,\ldots,n\}} x_i\right) = x_{(n)} - x_{(1)}$ die **Spannweite** von X,
2. $IQA := x_{0.75} - x_{0.25}$ der **Interquartilsabstand (IQA)** von X,
3. $MAA := \frac{1}{n} \sum_{i=1}^{n} |x_i - x_{\text{med}}|$ die **mittlere absolute Abweichung** von X.

Streuungsmaße IV

- Die Betragsstriche in Teil 1 und 2 von Definition 3.6 fehlen, da sie überflüssig sind.
- Um Eindeutigkeit in Teil 2 und 3 von Definition 3.6 zu erhalten, sind die für kardinalskalierte Merkmale üblichen Konventionen zur Berechnung von Median und Quantilen aus Folie 61 bzw. 76 anzuwenden.
- Verwendung von $\overline{x}$ statt x_{med} in Teil 3 von Definition 3.6 prinzipiell möglich, aber: Beachte Folie 72!
- Weiterer möglicher Abstand: Quadrate der Differenzen zwischen Merkmalswerten

Streuungsmaße V

Definition 3.7 (empirische Varianz, empirische Standardabweichung)

Seien $x_1, \ldots, x_n$ die Urliste zu einem kardinalskalierten Merkmal X, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ das arithmetische Mittel von X. Dann heißt

1. $s^2 := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$ die **(empirische) Varianz** von X,
2. die (positive) Wurzel $s = \sqrt{s^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ die **(empirische) Standardabweichung** von X.

Streuungsmaße VI

- Empirische Varianz bzw. Standardabweichung sind die gebräuchlichsten Streuungsmaße.
- Standardabweichung s hat dieselbe Dimension wie die Merkmalswerte, daher i.d.R. besser zu interpretieren als Varianz.
- Für Merkmale mit positivem Mittelwert $\overline{x}$ als relatives Streuungsmaß gebräuchlich: **Variationskoeffizient** $VK := \frac{s}{\overline{x}}$
- „Rechenregeln“ zur alternativen Berechnung von s bzw. s^2 vorhanden.

Satz 3.8 (Verschiebungssatz)

Seien $x_1, \ldots, x_n$ die Urliste zu einem kardinalskalierten Merkmal X, $\overline{x}$ das arithmetische Mittel und s^2 die empirische Varianz von X. Dann gilt

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

Streuungsmaße VII

- Mit der Schreibweise $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ erhält man aus Satz 3.8 die kürzere Darstellung $s^2 = \overline{x^2} - \overline{x}^2$.
- Liegt zum Merkmal X die absolute Häufigkeitsverteilung $h(a)$ bzw. die relative Häufigkeitsverteilung $r(a)$ auf der Menge der Ausprägungen $A = \{a_1, \ldots, a_m\}$ vor, so kann s^2 auch durch

$$s^2 = \frac{1}{n}\sum_{j=1}^{m} h(a_j)\cdot(a_j - \overline{x})^2 = \sum_{j=1}^{m} r(a_j)\cdot(a_j - \overline{x})^2$$

  berechnet werden. (Berechnung von $\overline{x}$ dann mit Häufigkeiten als $\overline{x} = \frac{1}{n}\sum_{j=1}^{m} h(a_j)\cdot a_j = \sum_{j=1}^{m} r(a_j)\cdot a_j$, siehe Bemerkung 3.4 auf Folie 67)
- Natürlich kann alternativ auch Satz 3.8 verwendet und $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ mit Hilfe der Häufigkeitsverteilung durch

$$\overline{x^2} = \frac{1}{n}\sum_{j=1}^{m} h(a_j)\cdot a_j^2 = \sum_{j=1}^{m} r(a_j)\cdot a_j^2$$

  berechnet werden.

Empirische Varianz bei klassierten Daten

- Bei klassierten Daten: auch für empirische Varianz nur Approximation möglich.
- Analog zur Berechnung von s^2 aus Häufigkeitsverteilungen:
  - Näherungsweise Berechnung von s^2 aus Klassenmitten m_j und absoluten bzw. relativen Klassenhäufigkeiten h_j bzw. r_j der l Klassen als

$$s^2 = \frac{1}{n}\sum_{j=1}^{l} h_j \cdot (m_j - \overline{x})^2 \quad \text{mit} \quad \overline{x} = \frac{1}{n}\sum_{j=1}^{l} h_j \cdot m_j$$

    bzw.

$$s^2 = \sum_{j=1}^{l} r_j \cdot (m_j - \overline{x})^2 \quad \text{mit} \quad \overline{x} = \sum_{j=1}^{l} r_j \cdot m_j \ .$$

  - Alternativ: Verwendung von Satz 3.8 mit

$$\overline{x} := \frac{1}{n}\sum_{j=1}^{l} h_j \cdot m_j = \sum_{j=1}^{l} r_j \cdot m_j$$

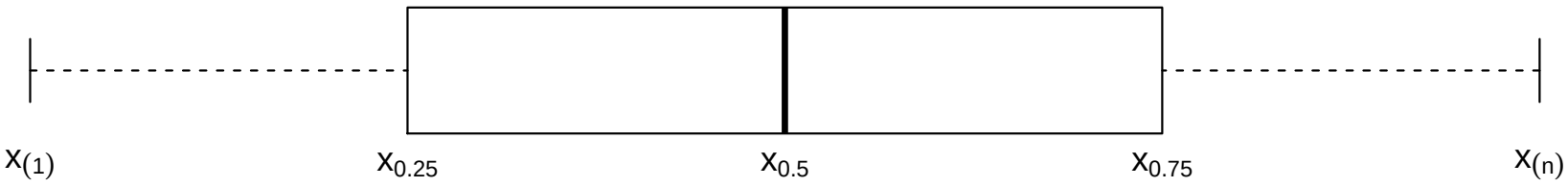
    und

$$\overline{x^2} := \frac{1}{n}\sum_{j=1}^{l} h_j \cdot m_j^2 = \sum_{j=1}^{l} r_j \cdot m_j^2 \ .$$

Box-and-whisker-Plot I

- Häufig von Interesse:
  Visueller Vergleich **eines** Merkmals für **verschiedene** statistische Massen
- Dazu nötig: Grafische Darstellung mit Ausdehnung (im Wesentlichen) nur in einer Dimension (2. Dimension für Nebeneinanderstellung der Datensätze)

$\rightsquigarrow$ **Box-and-whisker-Plot** oder kürzer **Box-Plot**:
Zur Urliste $x_1, \ldots, x_n$ eines kardinalskalierten Merkmals werden *im Prinzip* die 5 Kennzahlen $x_{(1)}, x_{0.25}, x_{0.5}, x_{0.75}, x_{(n)}$ in Form eines durch $x_{0.5}$ geteilten „Kästchens“ (Box) von $x_{0.25}$ bis $x_{0.75}$ und daran anschließende „Schnurrhaare“ (Whisker) bis zum kleinsten Merkmalswert $x_{(1)}$ und zum größten Merkmalswert $x_{(n)}$ dargestellt:

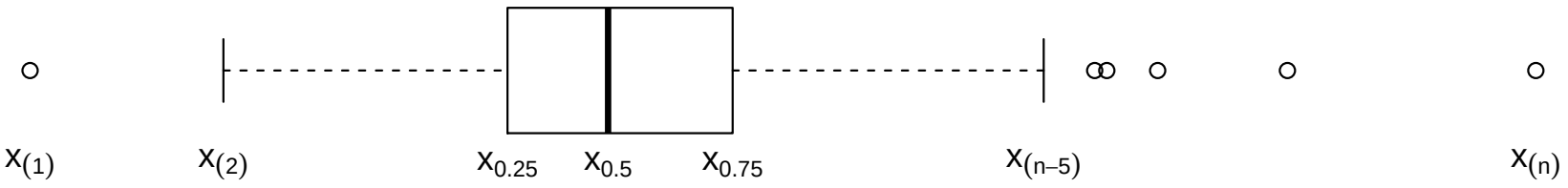


Box-and-whisker-Plot II

- (Häufig auftretende!) Ausnahme:
  $x_{(1)}$ und/oder $x_{(n)}$ liegen weiter als der 1.5-fache Interquartilsabstand (IQA) $x_{0.75} - x_{0.25}$ von der Box entfernt (also weiter als die 1.5-fache Breite der Box)

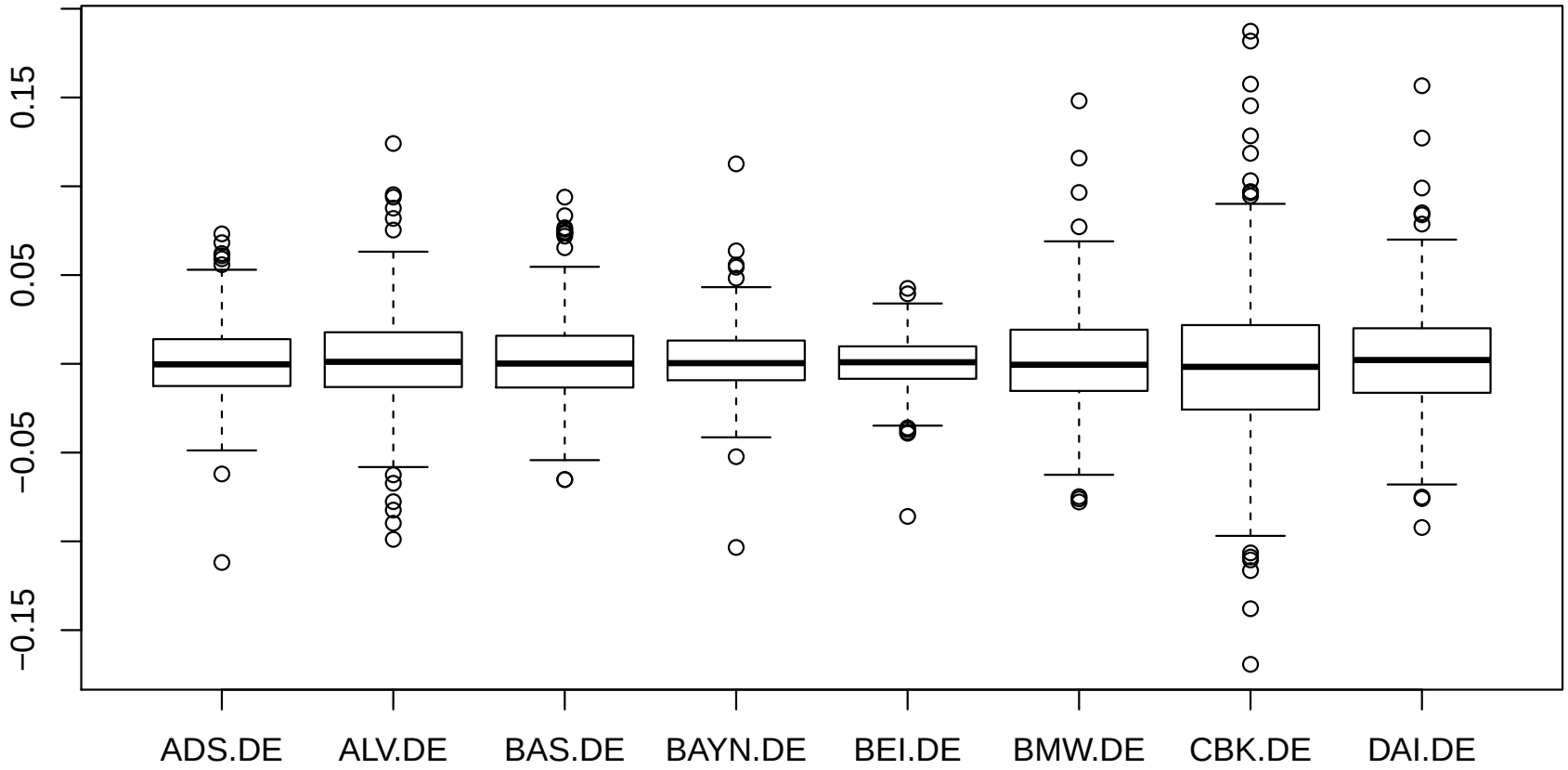
$\rightsquigarrow$ Dann: Whiskers nur bis zu äußersten Merkmalswerten innerhalb dieser Distanz und separates Eintragen der „Ausreißer“, d.h. aller Urlisteneinträge, die nicht von der Box und den Whiskers abgedeckt werden.

- Beispiel mit „Ausreißern“:



Box-and-whisker-Plot III

- Beispiel für Gegenüberstellung mehrerer Datensätze (Diskrete Tagesrenditen verschiedener DAX-Papiere)



Symmetrie(-maß), Schiefe I

- Neben Lage und Streuung bei kardinalskalierten Merkmalen auch interessant: **Symmetrie** (bzw. Asymmetrie oder Schiefe) und **Wölbung**
- Ein Merkmal X ist symmetrisch (um $\overline{x}$), wenn die Häufigkeitsverteilung von $X - \overline{x}$ mit der von $\overline{x} - X$ übereinstimmt.
  (Dabei ist mit $X - \overline{x}$ das Merkmal mit den Urlistenelementen $x_i - \overline{x}$ für $i \in \{1, \ldots, n\}$ bezeichnet, dies gilt analog für $\overline{x} - X$.)
- Symmetrie eines Merkmals entspricht also der Achsensymmetrie des zugehörigen Stabdiagramms um $\overline{x}$.
- Ist ein Merkmal nicht symmetrisch, ist die **empirische Schiefe** bzw. **empirische Skewness** ein geeignetes Maß für die Stärke der Asymmetrie.

Symmetrie(-maß), Schiefe II

Definition 3.9 (empirische Schiefe, Skewness)

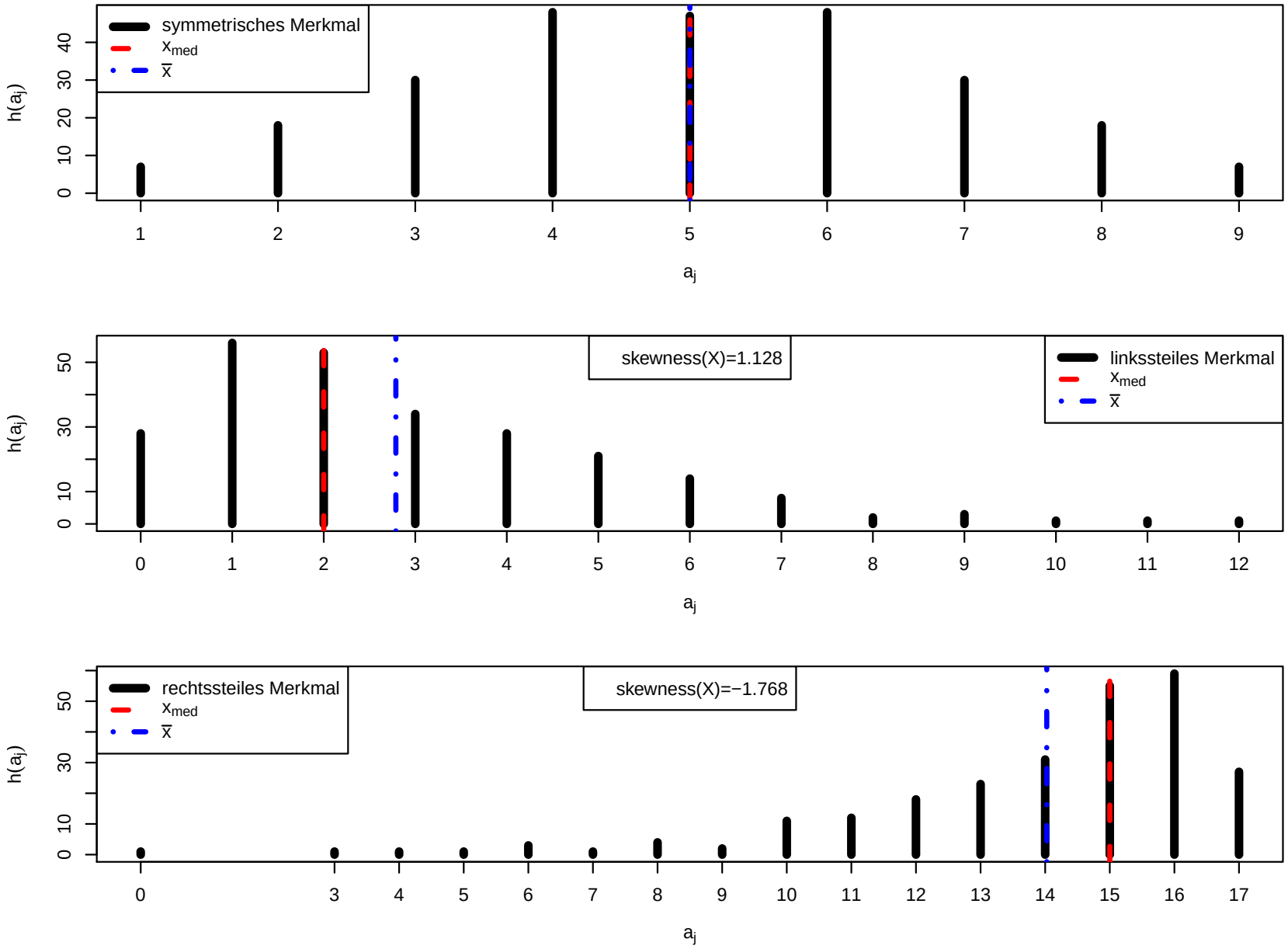
Sei X ein Merkmal mit der Urliste $x_1, \ldots, x_n$. Dann heißt

$$\text{skewness}(X) := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s} \right)^3$$

mit $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ und $s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2}$ die **empirische Schiefe (Skewness)** von X.

- Man kann zeigen: X symmetrisch $\Rightarrow$ skewness$(X) = 0$
- X heißt **linkssteil** oder **rechtsschief**, falls skewness$(X) > 0$.
- X heißt **rechtssteil** oder **linksschief**, falls skewness$(X) < 0$.
- Für symmetrische Merkmale ist $\overline{x}$ gleichzeitig Median von X, bei linkssteilen Merkmalen gilt *tendenziell* $\overline{x} > x_{\text{med}}$, bei rechtssteilen *tendenziell* $\overline{x} < x_{\text{med}}$.

Beispiele für empirische Schiefe von Merkmalen



Wölbungsmaß (Kurtosis) I

Definition 3.10 (empirische Wölbung, Kurtosis)

Sei X ein Merkmal mit der Urliste $x_1, \ldots, x_n$. Dann heißt

$$\text{kurtosis}(X) := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s} \right)^4$$

mit $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ und $s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2}$ die **empirische Wölbung (Kurtosis)** von X.

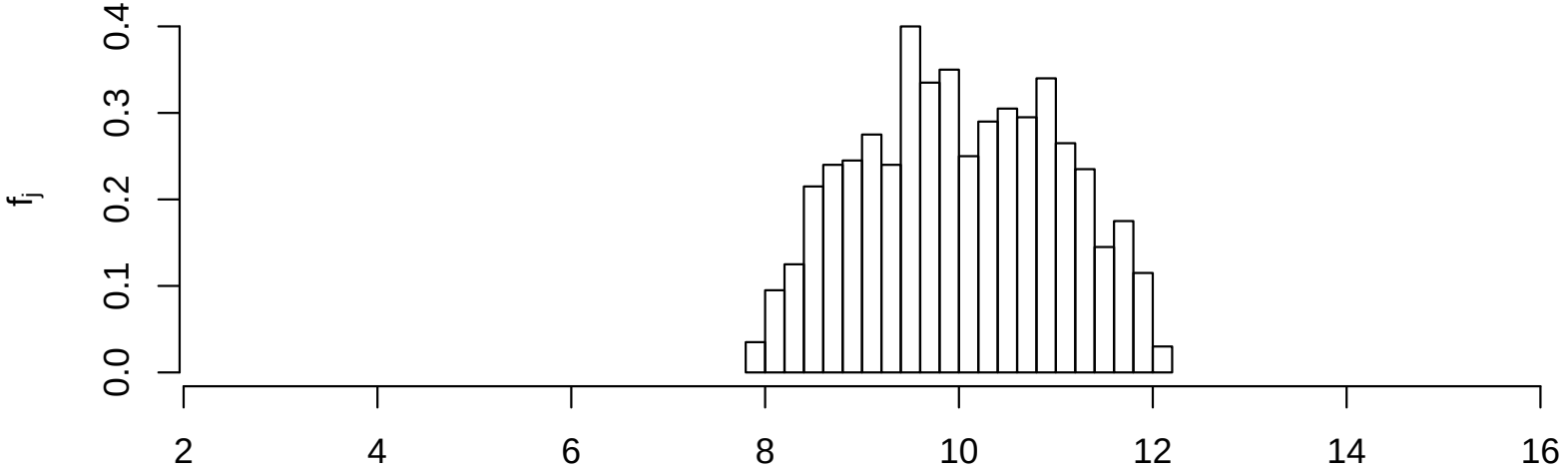
- Kurtosis misst bei Merkmalen mit *einem* Modalwert, wie „flach“ (kleiner Wert) bzw. „spitz“ (großer Wert) der „Gipfel“ um diesen Modalwert ist.

Wölbungsmaß (Kurtosis) II

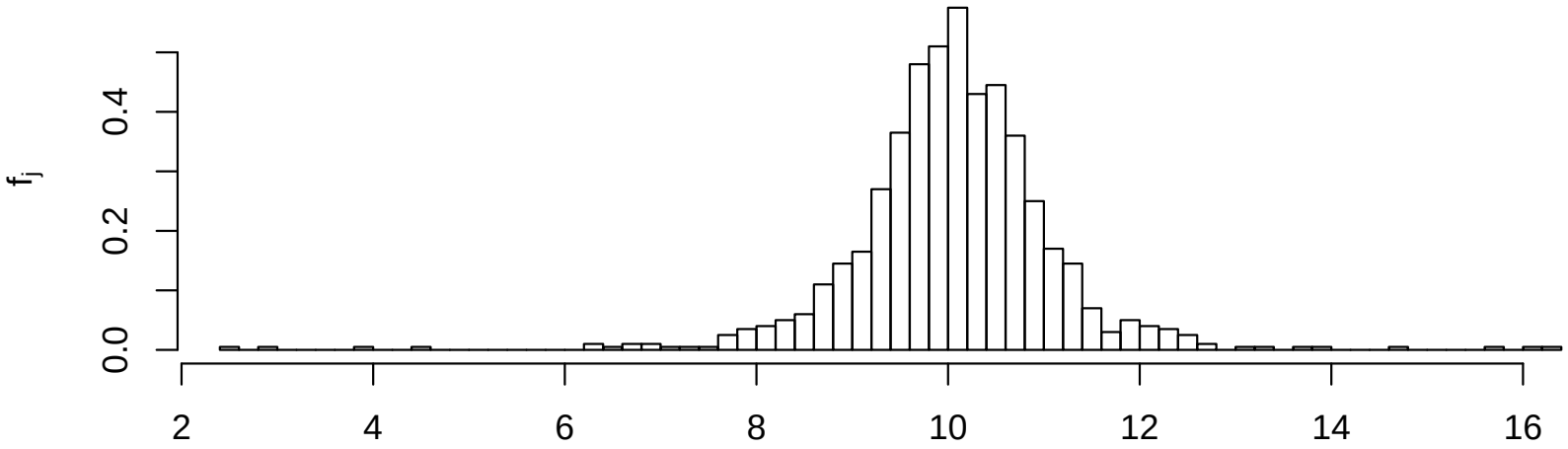
- Bei gleicher mittlerer quadratischer Abweichung vom Mittelwert ($\rightsquigarrow$ Varianz) müssen Merkmale mit größerer emp. Kurtosis (mehr Werten in der Nähe des Gipfels) auch mehr weit vom Gipfel entfernte Merkmalswerte besitzen.
- Der Wert 3 wird als „normaler“ Wert für die empirische Kurtosis angenommen, Merkmale mit $1 \leq \text{kurtosis}(X) < 3$ heißen platykurtisch, Merkmale mit $\text{kurtosis}(X) > 3$ leptokurtisch.
- *Vorsicht:* Statt der Kurtosis von X wird oft die **Exzess-Kurtosis** von X angegeben, die der um den Wert 3 verminderten Kurtosis entspricht.

Beispiele für Merkmale mit unterschiedlicher empirischer Kurtosis

Merkmal mit kleiner empirischer Kurtosis (2.088)

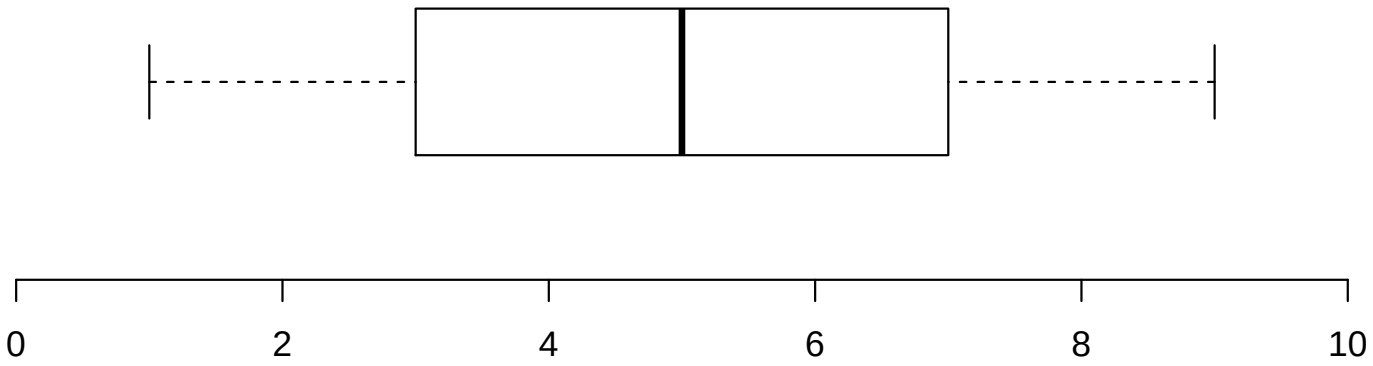


Merkmal mit großer empirischer Kurtosis (12.188)



Schiefe und Wölbung in grafischen Darstellungen I

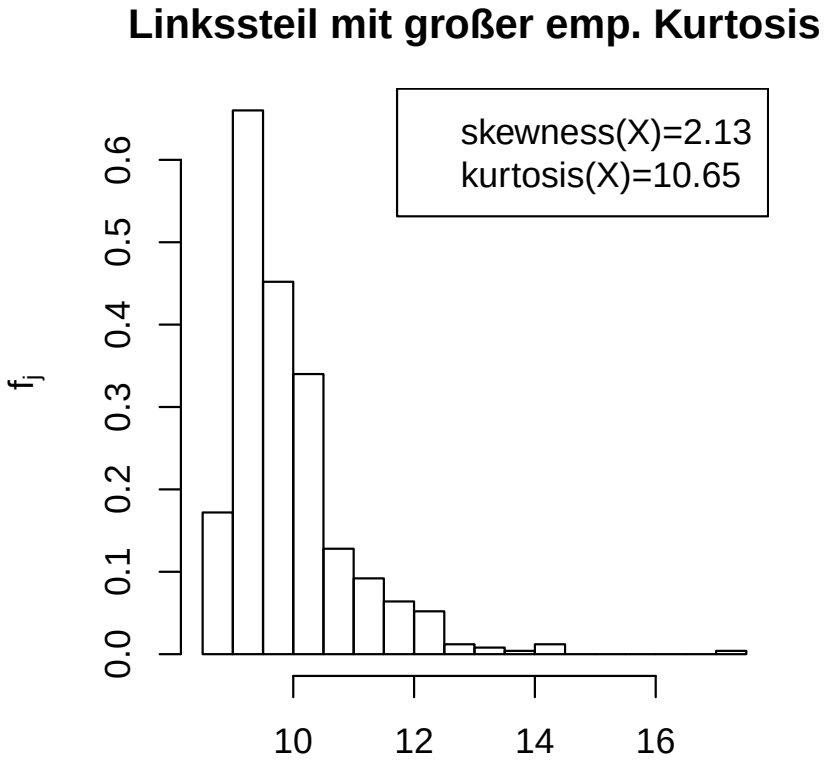
- Box-Plots lassen auch auf empirische Schiefe und Kurtosis schließen.
- Bei symmetrischen Merkmalen sind auch die Box-Plots symmetrisch. Beispiel: Box-Plot zur Urliste $1, 2, 3, 4, 5, 6, 7, 8, 9$:

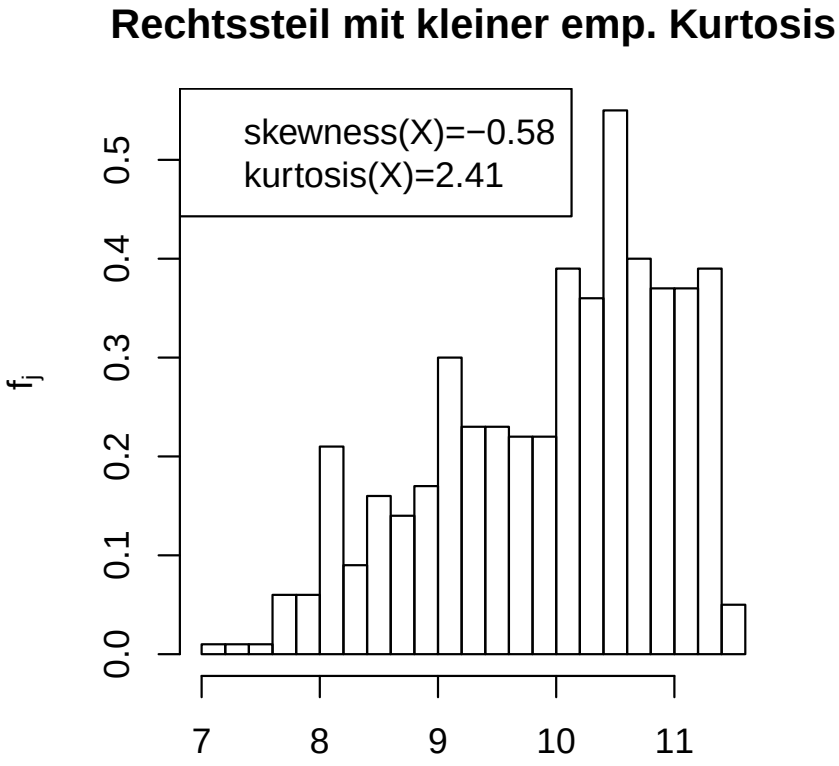


Schiefe und Wölbung in grafischen Darstellungen II

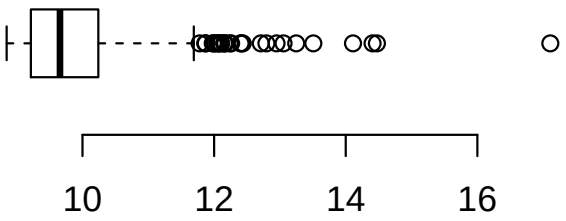
- Bei linkssteilen Merkmalen hat *tendenziell* der rechte/obere Teil (rechter/oberer Teil der Box und rechter/oberer Whisker) eine **größere** Ausdehnung als der linke/untere Teil.
- Bei rechtssteilen Merkmalen hat *tendenziell* der rechte/obere Teil (rechter/oberer Teil der Box und rechter/oberer Whisker) eine **kleinere** Ausdehnung als der linke/untere Teil.
- Bei Merkmalen mit **großer** empirischer Kurtosis gibt es *tendenziell* **viele** „Ausreißer“, also separat eingetragene Merkmalswerte außerhalb der Whiskers (wenigstens auf einer Seite).
- Bei Merkmalen mit **kleiner** empirischer Kurtosis gibt es häufig **wenige** oder **gar keine** „Ausreißer“.

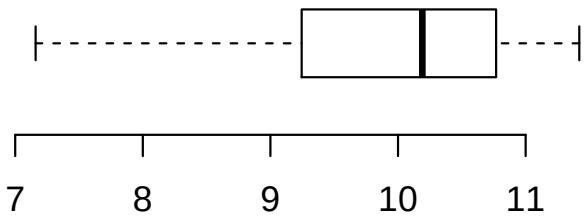
- Beispiele für Merkmale mit unterschiedlicher empirischer Schiefe/Kurtosis





- Zugehörige Box-Plots:





Inhaltsverzeichnis

(Ausschnitt)

4 Zweidimensionale Daten
- Häufigkeitsverteilungen unklassierter Daten
- Häufigkeitsverteilungen klassierter Daten
- Bedingte Häufigkeitsverteilungen und Unabhängigkeit
- Abhängigkeitsmaße

Auswertungsmethoden für mehrdimensionale Daten I

- Werden zu einer statistischen Masse mehrere Merkmale erhoben, so können diese natürlich individuell mit den Methoden für einzelne Merkmale ausgewertet werden.
- Eine Menge von Kennzahlen in den Spalten kann zum Beispiel gegen eine Menge von Merkmalen in den Zeilen tabelliert werden:

BMW.DE	$x_{(1)}$	$x_{0.5}$	$x_{(n)}$	$\overline{x}$	s	IQA	Schiefe	Kurt.
Preise	17.610	28.040	35.940	27.967	4.974	8.015	−0.383	1.932
log-Preise	2.868	3.334	3.582	3.314	0.189	0.286	−0.618	2.258
Renditen	−0.078	−0.001	0.148	0.002	0.030	0.034	0.672	5.941
log-Renditen	−0.081	−0.001	0.138	0.001	0.029	0.034	0.484	5.396

- Liegen die Merkmalswerte jeweils in ähnlichen Wertebereichen, ist auch ein Box-Plot verschiedener Merkmale nützlich.

Auswertungsmethoden für mehrdimensionale Daten II

- Isolierte Betrachtung der einzelnen Merkmale kann allerdings Abhängigkeiten zwischen mehreren Merkmalen nicht erkennbar machen!
- Zur Untersuchung von Abhängigkeiten mehrerer Merkmale „simultane“ Betrachtung der Merkmale erforderlich.
- Gemeinsame Betrachtung von mehr als 2 Merkmalen allerdings technisch schwierig.

$\rightsquigarrow$ Spezielle Methoden für **zweidimensionale** Daten (2 Merkmale simultan)

Häufigkeitsverteilungen zweidimensionaler Daten I

- Im Folgenden wird angenommen, dass den Merkmalsträgern zu **zwei** Merkmalen X und Y Merkmalswerte zugeordnet werden, also ein **zweidimensionales Merkmal** (X, Y) vorliegt.
- Analog zum eindimensionalen Fall geht man davon aus, auch vor der Erhebung schon Mengen M_1 bzw. M_2 angeben zu können, die alle vorstellbaren Merkmalswerte des Merkmals X bzw. Y enthalten.
- Die Urliste der Länge n (zur statistischen Masse der Mächtigkeit n) besteht nun aus den n Paaren

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

mit $x_m \in M_1$ und $y_m \in M_2$ bzw. $(x_m, y_m) \in M_1 \times M_2$ für $m \in \{1, \ldots, n\}$.

Häufigkeitsverteilungen zweidimensionaler Daten II

- *Unverzichtbare Eigenschaft der Urliste ist, dass die Paare von Merkmalswerten jeweils demselben Merkmalsträger zuzuordnen sind!*
- Wie im eindimensionalen Fall wird der Merkmalsraum zu X mit $A = \{a_1, \ldots, a_k\}$ bezeichnet, darüberhinaus der Merkmalsraum zu Y mit $B = \{b_1, \ldots, b_l\}$.
- Es muss nicht jede der $k \cdot l$ Kombinationen (a_i, b_j) in der Urliste auftreten!
- Geeignetes Mittel zur Aggregation der Merkmalswerte, wenn sowohl $k = \#A$ als auch $l = \#B$ „klein“ sind: **Häufigkeitsverteilungen**

Häufigkeitsverteilungen zweidimensionaler Daten III

- Zur Erstellung einer Häufigkeitsverteilung: Zählen, wie oft jede Kombination (a_i, b_j) der Merkmalsausprägung a_i von X und b_j von Y, $i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, l\}$, in der Urliste $(x_1, y_1), \ldots, (x_n, y_n)$ vorkommt.
  - Die **absoluten Häufigkeiten** $h_{ij} := h(a_i, b_j)$ geben für die Kombination (a_i, b_j), $i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, l\}$, die (absolute) Anzahl der Einträge der Urliste mit der Ausprägung (a_i, b_j) an, in Zeichen

$$h_{ij} := h(a_i, b_j) := \#\{m \in \{1, \ldots, n\} \mid (x_m, y_m) = (a_i, b_j)\} \ .$$

  - Die **relativen Häufigkeiten** $r_{ij} := r(a_i, b_j)$ geben für die Kombination (a_i, b_j), $i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, l\}$, den (relativen) Anteil der Einträge der Urliste mit der Ausprägung (a_i, b_j) an der gesamten Urliste an, in Zeichen

$$r_{ij} := r(a_i, b_j) := \frac{h(a_i, b_j)}{n} = \frac{\#\{m \in \{1, \ldots, n\} \mid (x_m, y_m) = (a_i, b_j)\}}{n} \ .$$

Häufigkeitsverteilungen zweidimensionaler Daten IV

- Natürlich gilt auch hier $\sum_{i=1}^{k}\sum_{j=1}^{l} h(a_i, b_j) = n$ und $\sum_{i=1}^{k}\sum_{j=1}^{l} r(a_i, b_j) = 1$.
- Tabellarische Darstellung zweidimensionaler Häufigkeitsverteilungen in **Kontingenztabellen**:

$X \setminus Y$	b_1	b_2	$\cdots$	b_l
a_1	h_{11}	h_{12}	$\cdots$	h_{1l}
a_2	h_{21}	h_{22}	$\cdots$	h_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_k	h_{k1}	h_{k2}	$\cdots$	h_{kl}

- Statt absoluter Häufigkeiten h_{ij} hier auch relative Häufigkeiten r_{ij} üblich.

Häufigkeitsverteilungen zweidimensionaler Daten V

- Zu den absoluten Häufigkeiten h_{ij} und relativen Häufigkeiten r_{ij} definiert man die **absoluten Randhäufigkeiten**

$$h_{i\cdot} := \sum_{j=1}^{l} h_{ij} \text{ für } i \in \{1, \ldots, k\} \qquad \text{und} \qquad h_{\cdot j} := \sum_{i=1}^{k} h_{ij} \text{ für } j \in \{1, \ldots, l\}$$

sowie die **relativen Randhäufigkeiten**

$$r_{i\cdot} := \sum_{j=1}^{l} r_{ij} \text{ für } i \in \{1, \ldots, k\} \qquad \text{und} \qquad r_{\cdot j} := \sum_{i=1}^{k} r_{ij} \text{ für } j \in \{1, \ldots, l\} \ .$$

- Diese Randhäufigkeiten stimmen offensichtlich (!) mit den (eindimensionalen) individuellen Häufigkeitsverteilungen der Merkmale X bzw. Y überein.

Häufigkeitsverteilungen zweidimensionaler Daten VI

- Kontingenztabellen werden oft durch die Randhäufigkeiten, die sich dann als Zeilen- bzw. Spaltensummen ergeben, in der Form

$X \setminus Y$	b_1	b_2	$\cdots$	b_l	$h_{i\cdot}$
a_1	h_{11}	h_{12}	$\cdots$	h_{1l}	$h_{1\cdot}$
a_2	h_{21}	h_{22}	$\cdots$	h_{2l}	$h_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_k	h_{k1}	h_{k2}	$\cdots$	h_{kl}	$h_{k\cdot}$
$h_{\cdot j}$	$h_{\cdot 1}$	$h_{\cdot 2}$	$\cdots$	$h_{\cdot l}$	n

oder

$X \setminus Y$	b_1	b_2	$\cdots$	b_l	$r_{i\cdot}$
a_1	r_{11}	r_{12}	$\cdots$	r_{1l}	$r_{1\cdot}$
a_2	r_{21}	r_{22}	$\cdots$	r_{2l}	$r_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_k	r_{k1}	r_{k2}	$\cdots$	r_{kl}	$r_{k\cdot}$
$r_{\cdot j}$	$r_{\cdot 1}$	$r_{\cdot 2}$	$\cdots$	$r_{\cdot l}$	1

ergänzt.

- Zur besseren Abgrenzung von Randhäufigkeiten nennt man h_{ij} bzw. r_{ij} oft auch **gemeinsame absolute** bzw. **relative Häufigkeiten**.

Beispiel (Kontingenztabelle)

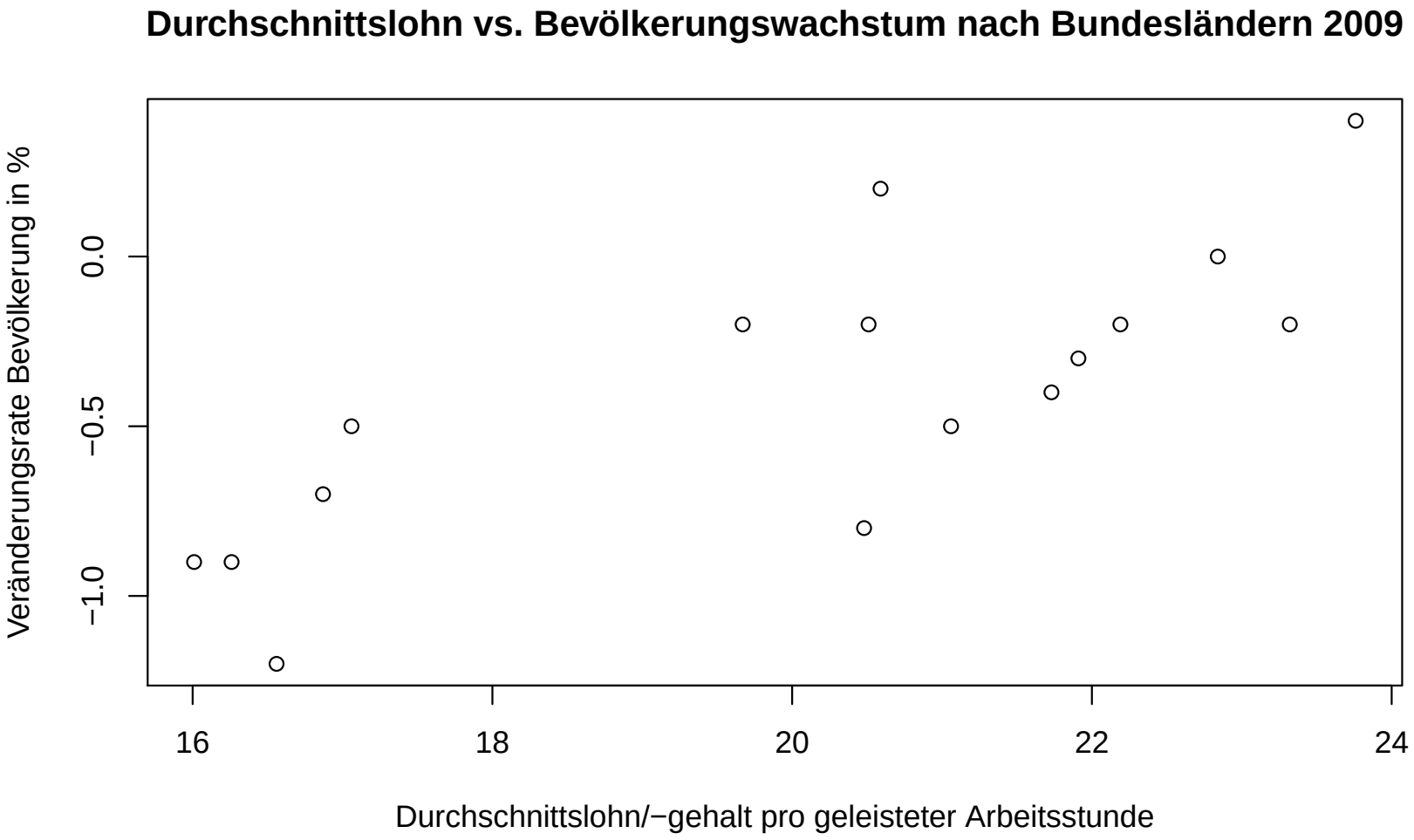
- Merkmal X: Mathematiknote,
  Merkmal Y: Physiknote,
  Urliste zum zweidimensionalen Merkmal (X, Y):

  $$(2,2), (2,3), (3,3), (5,3), (2,3), (5,4), (5,5), (4,2), (4,4), (1,2), (2,3), (1,3), (4,4), (2,3), (4,4), (3,4), (4,2), (5,4), (2,3), (4,4), (5,4), (2,3), (4,3), (1,1), (2,1), (2,2), (1,1), (2,3), (5,4), (2,2)$$

- Kontingenztabelle (mit Randhäufigkeiten)

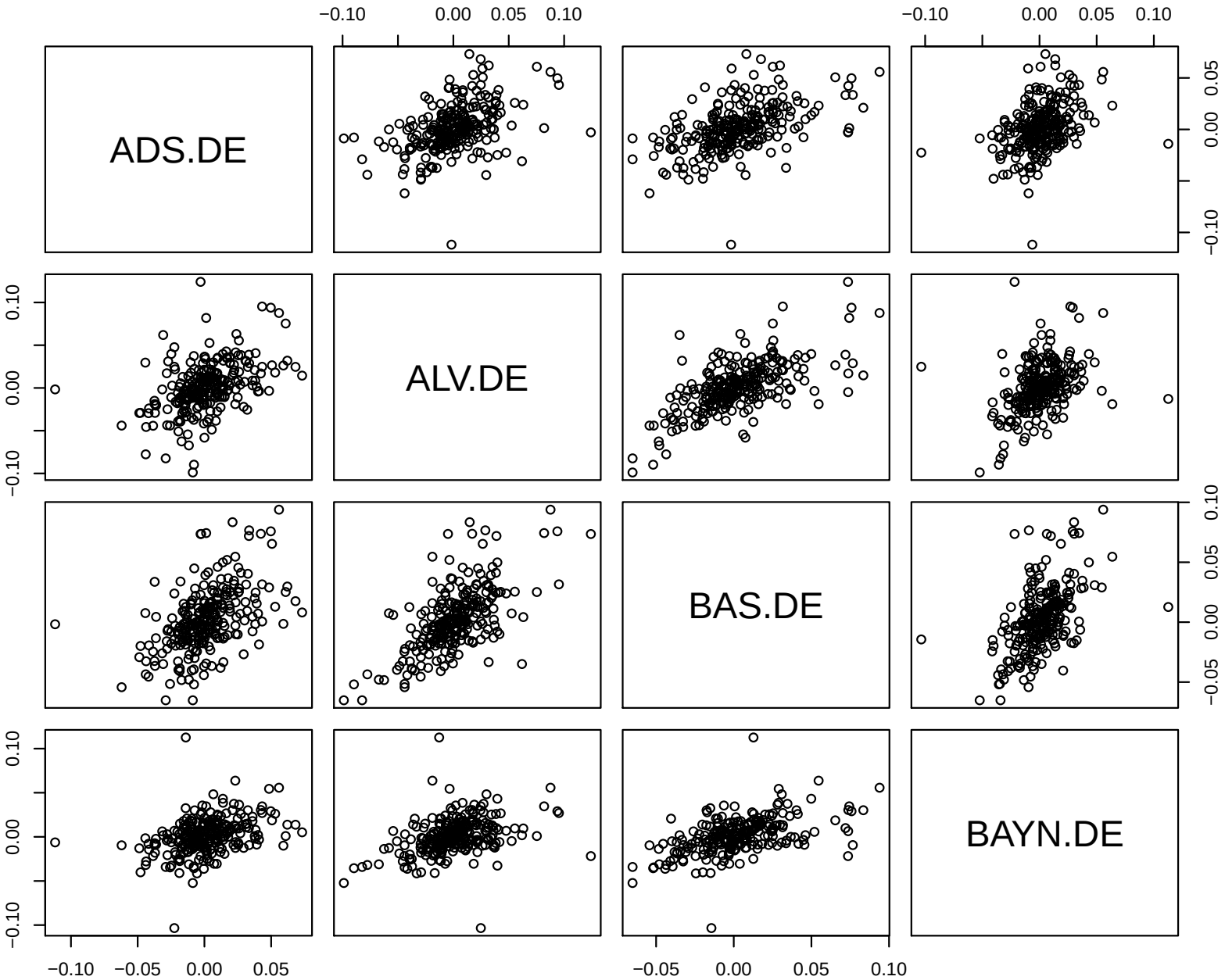
$X \backslash Y$	1	2	3	4	5	$h_{i.}$
1	2	1	1	0	0	4
2	1	3	7	0	0	11
3	0	0	1	1	0	2
4	0	2	1	4	0	7
5	0	0	1	4	1	6
$h_{.j}$	3	6	11	9	1	30

- Zur Visualisierung zweidimensionaler Daten mit (überwiegend) paarweise verschiedenen Ausprägungen (z.B. bei zwei stetigen Merkmalen): **Streudiagramm** bzw. **Scatter-Plot**



- Bei mehr als zwei Merkmalen: Paarweise Streudiagramme üblich.

Tagesrenditen (2008) verschiedener DAX−Papiere



Klassierung mehrdimensionaler Daten

- Analog zu eindimensionalen Daten:
  Häufigkeitstabellen schlecht geeignet für Merkmale mit „vielen“ verschiedenen Ausprägungen, also insbesondere stetige Merkmale.
- Genauso wie bei eindimensionalen Daten möglich: **Klassierung**
- Klassierung für mehrdimensionale Daten wird hier nicht mehr im Detail ausgeführt
- Allgemeine „Anleitungen“ für Klassierungen mehrdimensionaler Daten:
  - Oft werden nicht alle Merkmale klassiert, sondern nur einzelne.
  - Klassierung erfolgt individuell für jedes zu klassierende Merkmal.
  - Anwendung von Verfahren für nominalskalierte und ordinalskalierte Merkmale klar.
  - Bei Verfahren für kardinalskalierte Daten: Annahme gleichmäßiger Verteilung der Merkmalswerte auf Klassen, ggf. Klassenmitte als Näherung für die Merkmalsausprägungen (wie bisher!)

Bedingte Häufigkeitsverteilungen I

- Ziel einer gemeinsamen Betrachtung von mehreren Merkmalen: *Untersuchung von Abhängigkeiten und Zusammenhängen*
- **Zentrale Frage:** Hängen die jeweils angenommenen Merkmalswerte eines Merkmals X mit denen eines anderen Merkmals Y zusammen?
- Untersuchungsmöglichkeit auf Grundlage gemeinsamer Häufigkeiten zu den Merkmalen X mit Merkmalsraum $A = \{a_1, \ldots, a_k\}$ und Y mit Merkmalsraum $B = \{b_1, \ldots, b_l\}$: Bilden der **bedingten relativen Häufigkeiten**
  - $r(a_i | Y = b_j) := \dfrac{h_{ij}}{h_{\cdot j}}$
  - $r(b_j | X = a_i) := \dfrac{h_{ij}}{h_{i\cdot}}$

  für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.

Bedingte Häufigkeitsverteilungen II

- Für festes $j \in \{1, \ldots, l\}$ entsprechen die bedingten relativen Häufigkeiten $r(a_i|Y = b_j)$ also den relativen Häufigkeiten von Merkmal X *bei Einschränkung der statistischen Masse auf die Merkmalsträger, für die das Merkmal Y die Ausprägung b_j annimmt*.
- Umgekehrt enstprechen für festes $i \in \{1, \ldots, k\}$ die bedingten relativen Häufigkeiten $r(b_j|X = a_i)$ den relativen Häufigkeiten von Merkmal Y *bei Einschränkung der statistischen Masse auf die Merkmalsträger, für die das Merkmal X die Ausprägung a_i annimmt*.
- Man nennt die Merkmale X und Y *unabhängig*, wenn diese Einschränkungen keinen Effekt auf die relativen Häufigkeiten haben, d.h. alle bedingten relativen Häufigkeiten mit den zugehörigen relativen Randhäufigkeiten übereinstimmen.

Beispiel (bedingte Häufigkeitsverteilungen)

- Von Folie 106: Merkmal X: Mathematiknote, Merkmal Y: Physiknote, Kontingenztabelle

$X \backslash Y$	1	2	3	4	5	$h_{i\cdot}$
1	2	1	1	0	0	4
2	1	3	7	0	0	11
3	0	0	1	1	0	2
4	0	2	1	4	0	7
5	0	0	1	4	1	6
$h_{\cdot j}$	3	6	11	9	1	30

- Tabelle mit bedingten Häufigkeitsverteilungen für $Y|X = a_i$:

b_j	1	2	3	4	5	Σ
$r(b_j\|X=1)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	1
$r(b_j\|X=2)$	$\frac{1}{11}$	$\frac{3}{11}$	$\frac{7}{11}$	0	0	1
$r(b_j\|X=3)$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	1
$r(b_j\|X=4)$	0	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{4}{7}$	0	1
$r(b_j\|X=5)$	0	0	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	1

Unabhängigkeit von zwei Merkmalen I

Definition 4.1 (Unabhängigkeit von zwei Merkmalen)

Die Merkmale X mit Merkmalsraum $A = \{a_1, \ldots, a_k\}$ und Y mit Merkmalsraum $B = \{b_1, \ldots, b_l\}$ eines zweidimensionalen Merkmals (X, Y) zu einer Urliste der Länge n heißen **unabhängig**, falls

$$r(a_i | Y = b_j) = \frac{h_{ij}}{h_{.j}} \stackrel{!}{=} \frac{h_{i.}}{n} = r(a_i)$$

bzw. (gleichbedeutend)

$$r(b_j | X = a_i) = \frac{h_{ij}}{h_{i.}} \stackrel{!}{=} \frac{h_{.j}}{n} = r(b_j)$$

für alle $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$ gilt.

Unabhängigkeit von zwei Merkmalen II

- Die Bedingungen in Definition 4.1 sind offensichtlich genau dann erfüllt, wenn $h_{ij} = \frac{h_{i\cdot} \cdot h_{\cdot j}}{n}$ bzw. $r_{ij} = r_{i\cdot} \cdot r_{\cdot j}$ für alle $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$ gilt, die gemeinsamen relativen Häufigkeiten sich also als Produkt der relativen Randhäufigkeiten ergeben.
- Unabhängigkeit im Sinne von Definition 4.1 ist eher ein idealtypisches Konzept und in der Praxis kaum erfüllt.
- Interessant sind daher Maße, die vorhandene Abhängigkeiten zwischen zwei Merkmalen näher quantifizieren.

Abhängigkeitsmaße

- Je nach Skalierungsniveau der Merkmale X und Y können verschiedene Verfahren zur Messung der Abhängigkeit verwendet werden, das niedrigste Skalierungsniveau (nominal $\prec$ ordinal $\prec$ kardinal) ist dabei für die Einschränkung der geeigneten Verfahren maßgeblich:
  - Verfahren für ordinalskalierte Merkmale können nur dann eingesetzt werden, wenn beide Merkmale X und Y mindestens ordinalskaliert sind.
  - Verfahren für kardinalskalierte Merkmale können nur dann eingesetzt werden, wenn beide Merkmale X und Y kardinalskaliert sind.
- Trotz unterschiedlicher Wertebereiche der Abhängigkeitsmaße besteht die Gemeinsamkeit, dass die Abhängigkeit von X und Y stets mit dem Wert 0 gemessen wird, falls X und Y unabhängig gemäß Definition 4.1 sind.
- *Vorsicht beim Ableiten von Kausalitätsbeziehungen (Wirkungsrichtungen) aus entdeckten Abhängigkeiten!*

Kardinalskalierte Merkmale

Definition 4.2 (emp. Kovarianz, Pearsonscher Korrelationskoeffizient)

Gegeben sei das zweidimensionale Merkmal (X, Y) mit der Urliste $(x_1, y_1), \ldots, (x_n, y_n)$ der Länge n, X und Y seien kardinalskaliert. Mit $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ bzw. $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ seien wie üblich die arithmetischen Mittelwerte von X bzw. Y bezeichnet, mit

$$s_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad \text{bzw.} \quad s_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

die jeweiligen empirischen Standardabweichungen. Dann heißen

- $s_{X,Y} := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$

  die **empirische Kovarianz** von X und Y und
- $r_{X,Y} := \dfrac{s_{X,Y}}{s_X \cdot s_Y} = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$

  der **(Bravais-)Pearsonsche Korrelationskoeffizient** von X und Y.

Bemerkungen I

- $s_{X,Y}$ kann meist einfacher gemäß $s_{X,Y} = \overline{xy} - \overline{x} \cdot \overline{y}$ mit

$$\overline{xy} := \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i$$

  berechnet werden.
- Bei Vorliegen der Häufigkeitsverteilung kann $\overline{xy}$ einfacher gemäß

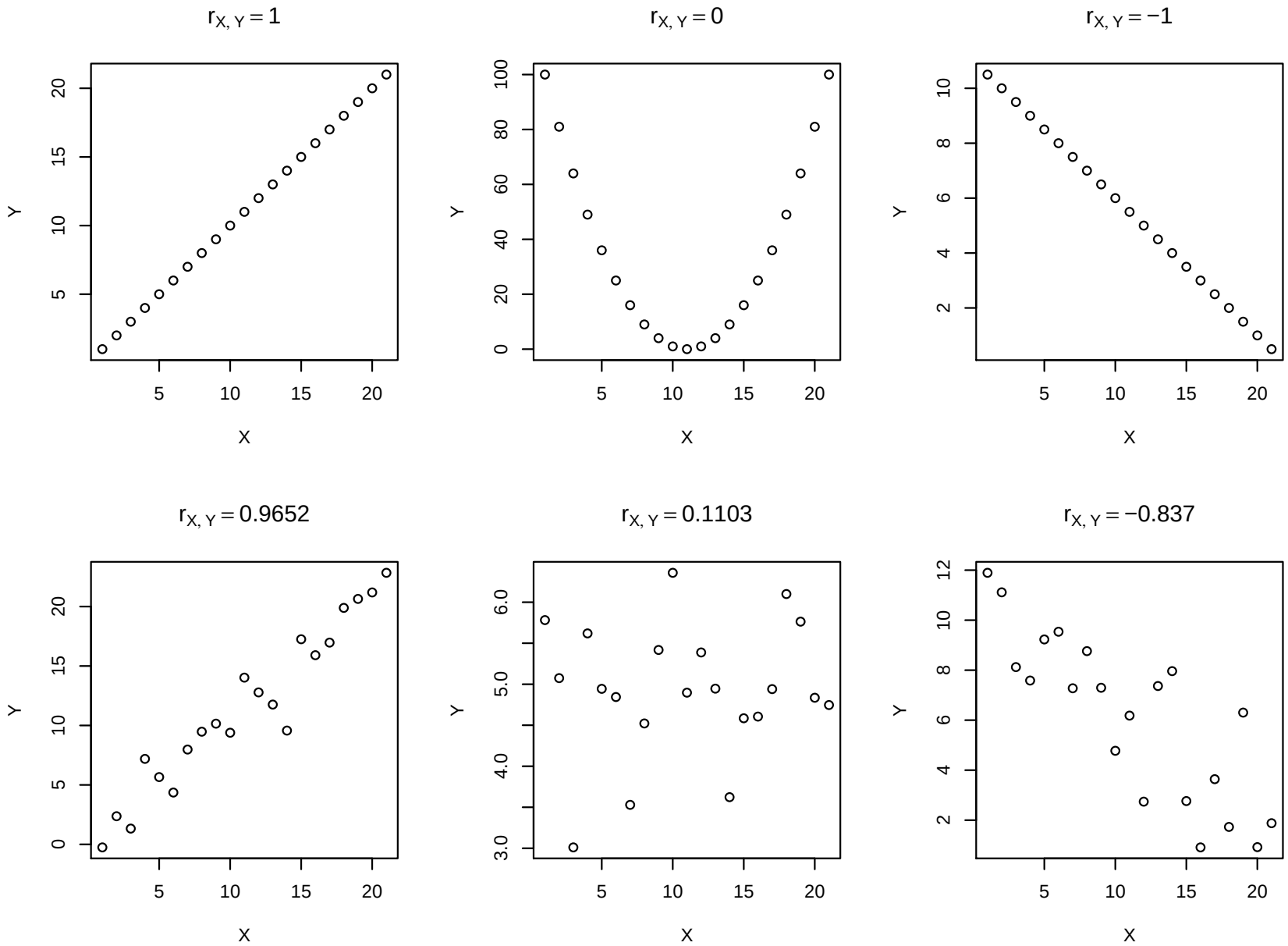
$$\overline{xy} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{l} a_i b_j \cdot h_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{l} a_i b_j \cdot r_{ij}$$

  berechnet werden ($\overline{x}$ und $\overline{y}$ werden zur Berechnung von $s_{X,Y}$ dann zweckmäßigerweise ebenfalls mit Hilfe der Häufigkeitsverteilungen berechnet, siehe dazu Folie 67).

Bemerkungen II

- Es gilt **stets** $-1 \leq r_{X,Y} \leq 1$.
- $r_{X,Y}$ misst **lineare** Zusammenhänge, spezieller gilt
  - $r_{X,Y} > 0$ bei positiver „Steigung“ („X und Y sind **positiv korreliert**“),
  - $r_{X,Y} < 0$ bei negativer „Steigung“ („X und Y sind **negativ korreliert**“),
  - $|r_{X,Y}| = 1$, falls alle (x_i, y_i) auf einer Geraden (mit Steigung $\neq 0$) liegen.
- $r_{X,Y}$ ist nur definiert, wenn X und Y jeweils mindestens zwei verschiedene Merkmalsausprägungen besitzen.

Beispiel: Pearsonscher Korrelationskoeffizient



(Mindestens) ordinalskalierte Merkmale I

- Messen *linearer* Zusammenhänge bei Ordinalskala nicht (mehr) möglich, stattdessen: Messen *monotoner* Zusammenhänge.
- Hierzu für X und Y erforderlich: Bilden der **Ränge** der Merkmalswerte (gemäß der vorgegebenen Ordnung).
- Aus den Merkmalen X und Y mit Merkmalswerten $x_1, \ldots, x_n$ bzw. $y_1, \ldots, y_n$ werden dabei neue Merkmale $\mathrm{rg}(X)$ und $\mathrm{rg}(Y)$ mit Merkmalswerten $\mathrm{rg}(X)_1, \ldots, \mathrm{rg}(X)_n$ bzw. $\mathrm{rg}(Y)_1, \ldots, \mathrm{rg}(Y)_n$.
- Bilden der Ränge wird exemplarisch für Merkmal X beschrieben (Bilden der Ränge für Y ganz analog).

(Mindestens) ordinalskalierte Merkmale II

1. Einfacher Fall: *Alle n Merkmalswerte sind verschieden.*
   $\rightsquigarrow$ Ränge von 1 bis n werden den Merkmalswerten nach der Position in der gemäß der vorgegebenen Ordnung sortierten Urliste zugewiesen:
   $$x_{(1)} \mapsto 1, \ldots, x_{(n)} \mapsto n$$
2. Komplizierter Fall: Es existieren mehrfach auftretende Merkmalswerte (sog. *Bindungen*), d.h. es gilt $x_i = x_j$ für (mindestens) ein Paar (i, j) mit $i \neq j$.
   $\rightsquigarrow$ Prinzipielle Vorgehensweise wie im einfachen Fall, Ränge übereinstimmender Merkmalswerte müssen aber (arithmetisch) gemittelt werden.

(Mindestens) ordinalskalierte Merkmale III

- „Berechnungsvorschrift“ für beide Fälle in folgender Definition:

Definition 4.3 (Rang eines Merkmals X, $\mathrm{rg}(X)_i$)

Gegeben sei ein Merkmal X mit Urliste $x_1, \ldots, x_n$. Dann heißt für $i \in \{1, \ldots, n\}$

$$\begin{aligned}\mathrm{rg}(X)_i &:= \#\{j \in \{1, \ldots, n\} \mid x_j \leq x_i\} - \frac{\#\{j \in \{1, \ldots, n\} \mid x_j = x_i\} - 1}{2} \\ &= \sum_{\substack{a_j \leq x_i \\ 1 \leq j \leq m}} h(a_j) - \frac{h(x_i) - 1}{2} \\ &= n \cdot F(x_i) - \frac{h(x_i) - 1}{2}\end{aligned}$$

der **Rang** von x_i. Die Werte $\mathrm{rg}(X)_1, \ldots, \mathrm{rg}(X)_n$ können als Urliste zu einem neuen Merkmal $\mathrm{rg}(X)$ aufgefasst werden.

(Mindestens) ordinalskalierte Merkmale IV

- Der zweite (subtrahierte) Term $\frac{\#\{j\in\{1,\ldots,n\}\mid x_j=x_i\}-1}{2}$ bzw. $\frac{h(x_i)-1}{2}$ in Definition 4.3 dient der Berechnung des arithmetischen Mittels bei Vorliegen von Bindungen.
- Liegen keine Bindungen vor (sind also alle Merkmalswerte verschieden), ist der zweite (subtrahierte) Term in Definition 4.3 immer gleich 0.
- Idee zur Konstruktion eines Abhängigkeitsmaßes für (mindestens) ordinalskalierte zweidimensionale Merkmale (X, Y):
  1. Übergang von X zu $\mathrm{rg}(X)$ sowie von Y zu $\mathrm{rg}(Y)$
  2. Berechnung des Pearsonschen Korrelationskoeffizienten von $\mathrm{rg}(X)$ und $\mathrm{rg}(Y)$

Spearmanscher Rangkorrelationskoeffizient I

Definition 4.4 (Spearmanscher Rangkorrelationskoeffizient)

Gegeben sei das zweidimensionale Merkmal (X, Y) mit der Urliste $(x_1, y_1), \ldots, (x_n, y_n)$ der Länge n, X und Y seien (mindestens) ordinalskaliert. Zu X und Y seien die Ränge $\mathrm{rg}(X)$ und $\mathrm{rg}(Y)$ gemäß Definition 4.3 gegeben. Dann heißt

$$r_{X,Y}^{(S)} := r_{\mathrm{rg}(X),\mathrm{rg}(Y)} = \frac{s_{\mathrm{rg}(X),\mathrm{rg}(Y)}}{s_{\mathrm{rg}(X)} \cdot s_{\mathrm{rg}(Y)}}$$

der **Spearmansche Rangkorrelationskoeffizient** von X und Y.

- Wegen des Zusammenhangs mit dem Pearsonschen Korrelationskoeffizienten gilt offensichtlich stets

$$-1 \leq r_{X,Y}^{(S)} \leq 1 \ .$$

Spearmanscher Rangkorrelationskoeffizient II

- Bei der Berechnung von $r_{X,Y}^{(S)}$ kann die Eigenschaft

$$\overline{\mathrm{rg}(X)} = \overline{\mathrm{rg}(Y)} = \frac{n+1}{2}$$

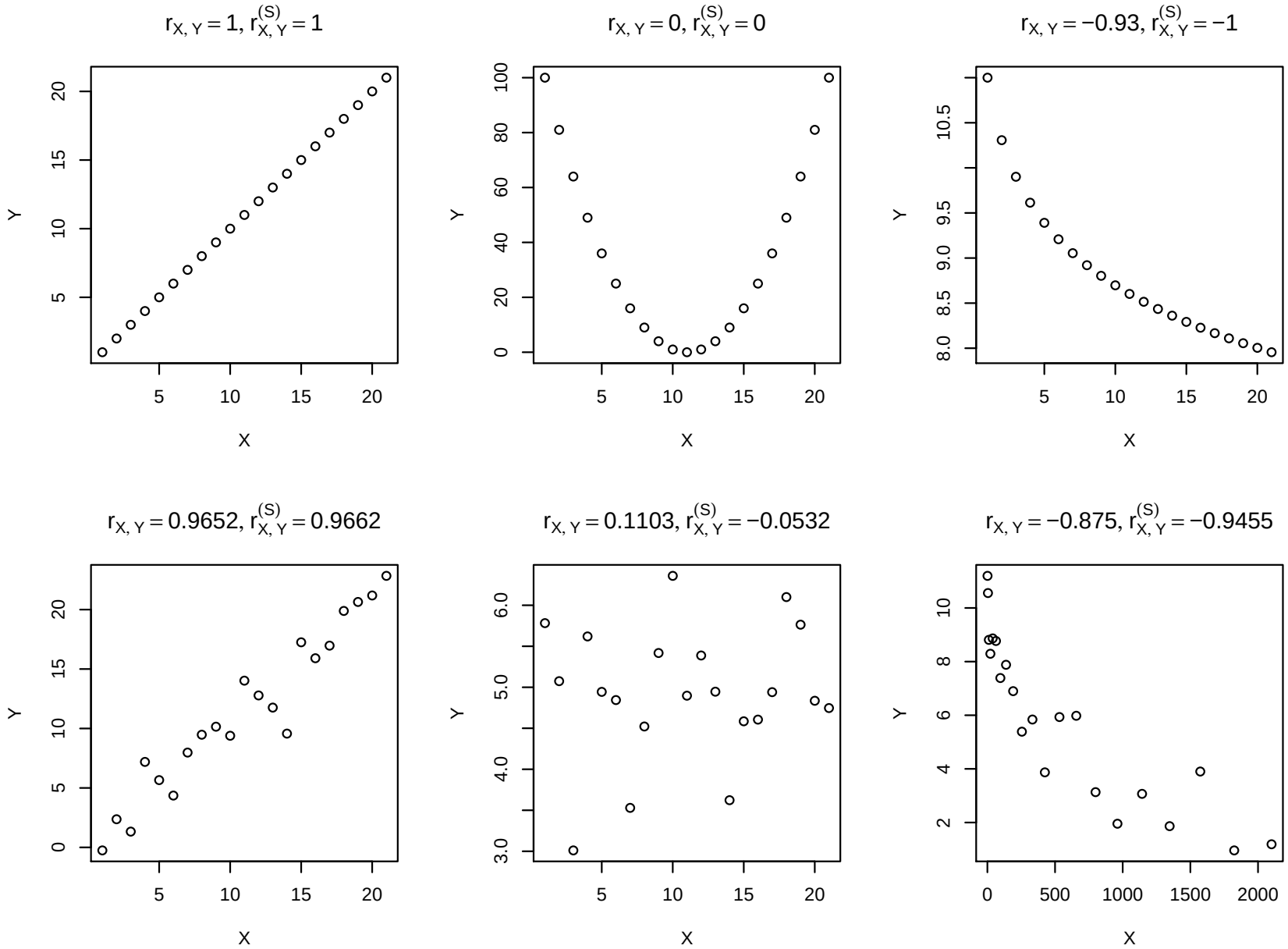
  ausgenutzt werden.
- Damit gilt für $r_{X,Y}^{(S)}$ stets:

$$r_{X,Y}^{(S)} = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathrm{rg}(X)_i \cdot \mathrm{rg}(Y)_i - \frac{(n+1)^2}{4}}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(\mathrm{rg}(X)_i)^2 - \frac{(n+1)^2}{4}\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}(\mathrm{rg}(Y)_i)^2 - \frac{(n+1)^2}{4}\right]}}$$

- **Nur** wenn $x_i \neq x_j$ und $y_i \neq y_j$ für alle $i \neq j$ gilt (also **keine** Bindungen vorliegen), gilt die wesentlich leichter zu berechnende „Formel“

$$r_{X,Y}^{(S)} = 1 - \frac{6 \cdot \sum_{i=1}^{n}(\mathrm{rg}(X)_i - \mathrm{rg}(Y)_i)^2}{n \cdot (n^2 - 1)} \ .$$

Beispiel: Spearmanscher Rangkorrelationskoeffizient



(Mindestens) nominalskalierte Merkmale I

- Da bei nominalskalierten Merkmalen keine Ordnung vorgegeben ist, kann hier lediglich die *Stärke*, nicht aber die *Richtung* der Abhängigkeit zwischen X und Y gemessen werden.
- Idee zur Konstruktion eines Abhängigkeitsmaßes auf Basis der gemeinsamen Häufigkeitstabelle zu (X, Y):
  - Bei Unabhängigkeit der Merkmale X und Y müsste nach Definition 4.1 auf Folie 113

    $$h_{ij} = \frac{h_{i.} \cdot h_{.j}}{n} \qquad \text{für alle } i \in \{1, \ldots, k\}, j \in \{1, \ldots, l\}$$

    gelten.
  - Abweichungen zwischen h_{ij} und $\frac{h_{i.} \cdot h_{.j}}{n}$ können also zur Messung der Abhängigkeit eingesetzt werden.
- Hier verwendetes Abhängigkeitsmaß entsteht aus geeigneter Zusammenfassung und Normierung dieser Abweichungen.

(Mindestens) nominalskalierte Merkmale II

Definition 4.5 (Pearsonscher Kontingenzkoeffizient)

Gegeben sei das zweidimensionale Merkmal (X, Y) zu einer Urliste der Länge n mit den zugehörigen absoluten gemeinsamen Häufigkeiten h_{ij} sowie den Randhäufigkeiten $h_{i\cdot}$ und $h_{\cdot j}$ für $i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, l\}$.
Dann heißt

$$C_{X,Y}^{\text{korr}} := \sqrt{\frac{\min\{k, l\}}{\min\{k, l\} - 1} \cdot \frac{\chi^2}{n + \chi^2}}$$

mit

$$\chi^2 := \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(h_{ij} - \frac{h_{i\cdot} \cdot h_{\cdot j}}{n}\right)^2}{\frac{h_{i\cdot} \cdot h_{\cdot j}}{n}}$$

korrigierter Pearsonscher Kontingenzkoeffizient der Merkmale X und Y.

- Es gilt stets $0 \leq C_{X,Y}^{\text{korr}} \leq 1$.

Teil II

Wahrscheinlichkeitsrechnung

Inhaltsverzeichnis

(Ausschnitt)

5 Zufallsexperimente
- Ergebnisse
- Ereignisse
- Wahrscheinlichkeiten

Zufallsexperimente (Zufallsvorgänge)

- Unter Zufallsexperimenten versteht man Geschehnisse,
  - deren *mögliche* (sich gegenseitig ausschließende!) Ausgänge bekannt sind,
  - deren (konkreter) Ausgang aber ungewiss ist (oder zumindest ungewiss zu sein scheint bzw. in der Anwendungssituation a priori nicht bestimmt werden kann!).
- Beispiele für Zufallsexperimente: Würfelwurf, Münzwurf, Lottoziehung, gefallene Zahl beim Roulette, ausgeteiltes Blatt bei Kartenspielen
- Entscheidend ist weniger, ob der Ausgang eines Vorgangs tatsächlich „zufällig“ ist, sondern vielmehr, ob der Ausgang als zufällig angesehen werden soll!
- Zur Modellierung von Zufallsexperimenten mehrere Komponenten erforderlich.

Ergebnisse

Definition 5.1

Zu einem Zufallsexperiment nennt man die nichtleere Menge

$$\Omega = \{\omega \,|\, \omega \text{ ist ein möglicher Ausgang des Zufallsexperiments}\}$$

der möglichen (sich gegenseitig ausschließenden) Ausgänge **Ergebnisraum**, ihre Elemente ω auch **Ergebnisse**.

- Nach Definition 5.1 kann also **nach** Durchführung des Zufallsexperiments **genau ein** Ergebnis $\omega \in \Omega$ angegeben werden, das den Ausgang des Zufallsexperiments beschreibt.

Ereignisse I

- Interesse gilt nicht nur einzelnen Ausgängen des Zufallsexperiments, sondern oft auch Zusammenschlüssen von Ausgängen, sog. **Ereignissen**.
  Beispiele
  - beim Roulette: Es ist Rot gefallen
  - beim Würfeln: Es ist eine gerade Zahl gefallen
  - beim Skatspiel: Das eigene Blatt enthält mindestens 2 Buben.
  - beim Lottospiel: Eine vorgegebene Tippreihe enthält (genau) 3 „Richtige“
- Ereignisse sind also Teilmengen des Ergebnisraums; ist Ω der Ergebnisraum eines Zufallsexperiments, gilt für jedes Ereignis A die Beziehung $A \subseteq \Omega$.
- Ereignisse, die nur aus einem Element des Ergebnisraums (also einem einzigen Ergebnis) bestehen, heißen auch **Elementarereignisse**.

Ereignisse II

- Ob ein Ereignis A eingetreten ist oder nicht, hängt vom tatsächlichen Ausgang des Zufallsexperiments ab:
  - Ein Ereignis $A \subseteq \Omega$ ist eingetreten, falls für den Ausgang ω gilt: $\omega \in A$
  - Ein Ereignis $A \subseteq \Omega$ ist nicht eingetreten, falls für den Ausgang ω gilt: $\omega \notin A$

$\rightsquigarrow$ Es gilt insbesondere **nicht** (wie bei Ergebnissen), dass jeweils (nur) genau ein Ereignis eintritt!

- Das Ereignis Ω wird als **sicheres Ereignis** bezeichnet, da es offensichtlich immer eintritt.
- Das Ereignis $\emptyset = \{\}$ (die leere Menge) wird als **unmögliches Ereignis** bezeichnet, da es offensichtlich nie eintritt.
  (Für kein Element $\omega \in \Omega$ kann $\omega \in \emptyset$ gelten.)

Ereignisse III

- Ereignisse können verknüpft werden. Dabei lassen sich aussagenlogische Verknüpfungen in mengentheoretische übersetzen, z.B.:
  - Wenn Ereignis A eintritt, dann auch Ereignis $B \Leftrightarrow$ Es gilt $A \subseteq B$.
  - Die Ereignisse A und B treten nie gleichzeitig ein $\Leftrightarrow$ Es gilt $A \cap B = \emptyset$.
  - Ereignis A und Ereignis B treten ein $\Leftrightarrow$ Ereignis $A \cap B$ tritt ein.
  - Ereignis A oder Ereignis B treten ein $\Leftrightarrow$ Ereignis $A \cup B$ tritt ein.

Vorsicht!

„Oder“ (für sich betrachtet) wird stets nicht-exklusiv verwendet, also **nicht** im Sinne von „entweder–oder“. „A oder B tritt ein“ bedeutet also, dass A eintritt oder B eintritt oder beide Ereignisse eintreten.

Ereignisse IV

Definition 5.2

Sei Ω eine Menge, seien A und B Teilmengen von Ω, es gelte also $A \subseteq \Omega$ und $B \subseteq \Omega$. Dann heißen

- $A \cup B := \{\omega \in \Omega \,|\, \omega \in A \vee \omega \in B\}$ die **Vereinigung** von A und B,
- $A \cap B := \{\omega \in \Omega \,|\, \omega \in A \wedge \omega \in B\}$ der **Durchschnitt** von A und B,
- $A \backslash B := \{\omega \in \Omega \,|\, \omega \in A \wedge \omega \notin B\}$ die **Differenz** von A und B,
- $\overline{A} := \Omega \backslash A = \{\omega \in \Omega \,|\, \omega \notin A\}$ das **Komplement** von A,
- A und B **disjunkt**, falls $A \cap B = \emptyset$,
- A und B **komplementär**, falls $B = \overline{A}$.

Ereignisse V

- Das Komplement $\overline{A}$ eines Ereignisses A heißt auch **Gegenereignis** von A. Es tritt offensichtlich genau dann ein, wenn A nicht eintritt.
- Vereinigungen und Durchschnitte werden auch für mehr als 2 (Teil-)Mengen bzw. Ereignisse betrachtet, zum Beispiel
  - die Vereinigung $\bigcup_{i=1}^{n} A_i$ bzw. der Durchschnitt $\bigcap_{i=1}^{n} A_i$ der n Mengen bzw. Ereignisse $A_1, \ldots, A_n$,
  - die Vereinigung $\bigcup_{i=1}^{\infty} A_i$ bzw. der Durchschnitt $\bigcap_{i=1}^{\infty} A_i$ der (abzählbar) unendlichen vielen Mengen bzw. Ereignisse $A_1, A_2, \ldots$.

Rechenregeln für Mengenoperationen I

Für Teilmengen bzw. Ereignisse A, B, C von Ω gelten die „Rechenregeln“:

- Idempotenzgesetze, neutrale und absorbierende Elemente

$$A \cup \emptyset = A \qquad A \cap \emptyset = \emptyset \qquad A \cup \Omega = \Omega \qquad A \cap B \subseteq A$$
$$A \cup A = A \qquad A \cap A = A \qquad A \cap \Omega = A \qquad A \cap B \subseteq B$$

- Kommutativgesetze

$$A \cup B = B \cup A \qquad A \cap B = B \cap A$$

- Assoziativgesetze

$$A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C$$

Rechenregeln für Mengenoperationen II

- Distributivgesetze

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

- De Morgansche Gesetze

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

- Disjunkte Zerlegung von A durch B

$$A = (A \cap B) \cup (A \cap \overline{B}) \qquad \text{mit} \qquad (A \cap B) \cap (A \cap \overline{B}) = \emptyset$$

Wahrscheinlichkeiten I

- Zur Modellierung von Zufallsexperimenten: „Quantifizierung“ des Zufalls durch *Angabe von* **Wahrscheinlichkeiten** *für Ereignisse.*
- Spezieller:
  Gesucht ist eine Abbildung, die *einer bestimmten Menge von* Ereignissen (Eintritts-)Wahrscheinlichkeiten zuordnet.
- Aber:
  - Was sind Wahrscheinlichkeiten?
  - Woher kommen diese Wahrscheinlichkeiten?
  - Was sagen diese Wahrscheinlichkeiten aus?

Wahrscheinlichkeiten II

- Es gibt verschiedene geläufige Wahrscheinlichkeitsbegriffe, insbesondere:
  - Klassischer oder Laplacescher Wahrscheinlichkeitsbegriff:
    Ω ist so konstruiert, dass alle Elementarereignisse die gleiche Eintrittswahrscheinlichkeit haben.
  - Häufigkeitsbasierter Wahrscheinlichkeitsbegriff:
    Zufallsexperiment wird als unendlich oft wiederholbar vorausgesetzt, Eintrittswahrscheinlichkeiten der Ereignisse als „Grenzwert“ der relativen Eintrittshäufigkeiten in (unendlicher) Folge von Durchführungen.
  - Subjektiver Wahrscheinlichkeitsbegriff:
    Persönliche Einschätzungen geben Eintrittswahrscheinlichkeiten an. Festlegung zum Beispiel durch *subjektive* Angabe von „Grenz“-Wettquoten auf das Eintreten von Ereignissen.
- Unabhängig vom konkreten Wahrscheinlichkeitsbegriff:
  Bestimmte „Konsistenzbedingungen“ für
  - Wahrscheinlichkeiten und
  - Menge der Ereignisse, denen Wahrscheinlichkeiten zugeordnet werden können

  sinnvoll.

Wahrscheinlichkeiten III

- Folgender „Minimalsatz“ von Bedingungen geläufig:
  - Ω muss eine Wahrscheinlichkeit zugeordnet werden können; genauer muss diese Wahrscheinlichkeit 1 betragen.
  - Wenn einem Ereignis A eine (Eintritts-)Wahrscheinlichkeit zugeordnet werden kann, dann muss auch dem Gegenereignis $\overline{A}$ (also dem Nichteintreten des Ereignisses A) eine Wahrscheinlichkeit zugeordnet werden können; genauer muss die Summe dieser beiden Wahrscheinlichkeiten 1 betragen.
  - Wenn einer Menge von Ereignissen $A_1, A_2, \ldots$ Wahrscheinlichkeiten zugeordnet werden können, dann muss auch eine Wahrscheinlichkeit dafür, dass mindestens eines dieser Ereignisse eintritt, angegeben werden können; sind die Ereignisse $A_1, A_2, \ldots$ (paarweise) disjunkt, dann muss diese Wahrscheinlichkeit mit der Summe der Einzelwahrscheinlichkeiten übereinstimmen.
  - Wahrscheinlichkeiten von Ereignissen dürfen nicht negativ sein.
- Mengensysteme, die diesen Bedingungen genügen, heißen σ-Algebren.

σ-Algebra I

Definition 5.3 (σ-Algebra)

Sei $\Omega \neq \emptyset$ eine Menge. Eine Menge $\mathcal{F}$ von Teilmengen von Ω heißt σ-**Algebra** über Ω, falls die folgenden Bedingungen erfüllt sind:

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$,
3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

- Ist Ω endlich (oder abzählbar unendlich), wird für $\mathcal{F}$ häufig die Menge *aller* Teilmengen von Ω, die sog. **Potenzmenge** $\mathcal{P}(\Omega)$, gewählt.
- Gilt $\#\Omega = n$, so ist $\#\mathcal{P}(\Omega) = 2^n$.

σ-Algebra II

- Aus Definition 5.3 folgen zahlreiche weitere Eigenschaften, z.B.:
  - $\emptyset \in \mathcal{F}$,
  - $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i \in \mathbb{N}} A_i \in \mathcal{F}$,
  - $A_1, \ldots, A_n \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$ und $\bigcap_{i=1}^{n} A_i \in \mathcal{F}$,
  - $A, B \in \mathcal{F} \Rightarrow A \backslash B \in \mathcal{F}$.
- Das Paar $(\Omega, \mathcal{F})$ wird auch „Messraum“ genannt.

Wahrscheinlichkeitsmaß

- Abbildungen, die den Elementen von $\mathcal{F}$ (!) Wahrscheinlichkeiten zuordnen, heißen Wahrscheinlichkeitsmaße.

Definition 5.4 (Wahrscheinlichkeitsmaß)

Es seien $\Omega \neq \emptyset$ eine Menge, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ eine σ-Algebra über Ω und $P : \mathcal{F} \to \mathbb{R}$ eine Abbildung. P heißt **Wahrscheinlichkeitsmaß** auf $\mathcal{F}$, falls gilt:

1. $P(A) \geq 0$ für alle $A \in \mathcal{F}$,
2. $P(\Omega) = 1$,
3. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ für alle $A_1, A_2, \ldots \in \mathcal{F}$ mit $A_j \cap A_k = \emptyset$ für $j \neq k$.

- Die Eigenschaften 1–3 in Definition 5.4 heißen auch *Axiome von Kolmogorov*.

Wahrscheinlichkeitsraum

- Insgesamt wird ein Zufallsexperiment vollständig durch einen Wahrscheinlichkeitsraum beschrieben (und mit ihm identifiziert):

Definition 5.5 (Wahrscheinlichkeitsraum)

Seien $\Omega \neq \emptyset$ ein Ergebnisraum, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ eine σ-Algebra über Ω und P ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$. Dann heißt das Tripel $(\Omega, \mathcal{F}, P)$ **Wahrscheinlichkeitsraum**. $\mathcal{F}$ wird auch **Ereignisraum** genannt.

Beispiel (Zufallsexperiment)

- (Einmaliges) Werfen eines (fairen!) Würfels als Zufallsexperiment.
  - Geeigneter Ergebnisraum: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Geeigneter Ereignisraum: $\mathcal{F} = \mathcal{P}(\Omega)$ mit $\#\mathcal{F} = 2^{\#\Omega} = 2^6 = 64$
  - Geeignetes Wahrscheinlichkeitsmaß: $P : \mathcal{F} \to \mathbb{R}; P(A) = \dfrac{\#A}{6}$
- Beispiele für Ereignisse:
  - A: Eine vier ist gefallen; $A = \{4\}$
  - B: Eine gerade Zahl wurde gewürfelt; $B = \{2, 4, 6\}$
  - C: Eine Zahl ≤ 3 wurde gewürfelt; $C = \{1, 2, 3\}$
  - D: Eine Zahl > 2 wurde gewürfelt; $D = \{3, 4, 5, 6\}$
- Es gilt: $P(A) = \frac{1}{6}, \qquad P(B) = P(C) = \frac{3}{6} = \frac{1}{2}, \qquad P(D) = \frac{4}{6} = \frac{2}{3}.$
- Es gilt zum Beispiel, dass A und C disjunkt sind, oder dass $\overline{B}$ das Ereignis „ungerade Zahl gewürfelt“ beschreibt.

Rechnen mit Wahrscheinlichkeiten

Es lassen sich die folgenden Rechenregeln für Wahrscheinlichkeitsräume $(\Omega, \mathcal{F}, P)$ und beliebige Ereignisse $A, B, A_1, A_2, \ldots \in \mathcal{F}$ herleiten:

1. $P(A) \leq 1$,
2. $P(\emptyset) = 0$,
3. $A \subseteq B \Rightarrow P(A) \leq P(B)$,
4. $P(\overline{A}) = 1 - P(A)$,
5. $P(\bigcup_{i=1}^{n} A_i) = P(A_1) + P(\overline{A_1} \cap A_2) + P(\overline{A_1} \cap \overline{A_2} \cap A_3) + \cdots + P(\overline{A_1} \cap \cdots \overline{A_{n-1}} \cap A_n)$

   und für $n = 2$ und $n = 3$ weiter aufgeschlüsselt:
   1. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
   2. $P(A_1 \cup A_2 \cup A_3) =$
      $P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$
6. Bilden die A_i eine **Zerlegung** von Ω, d.h. gilt $A_i \cap A_j = \emptyset$ für alle $i \neq j$ und $\bigcup_{i \in \mathbb{N}} A_i = \Omega$, so gilt

   $$P(B) = \sum_{i \in \mathbb{N}} P(B \cap A_i) \ .$$

   Insbesondere gilt stets $P(B) = P(B \cap A) + P(B \cap \overline{A})$.

Inhaltsverzeichnis

(Ausschnitt)

6 Diskrete Wahrscheinlichkeitsräume
- Laplacesche Wahrscheinlichkeitsräume
- Kombinatorik
- Allgemeine diskrete Wahrscheinlichkeitsräume

Laplacesche Wahrscheinlichkeitsräume I

- Einfachster Fall für $(\Omega, \mathcal{F}, P)$ (wie in Würfel-Beispiel):
  - Ω endlich,
  - Eintritt aller Ergebnisse $\omega \in \Omega$ gleichwahrscheinlich.
- Wahrscheinlichkeitsräume mit dieser Eigenschaft heißen **Laplacesche Wahrscheinlichkeitsräume**.
- Als σ-Algebra $\mathcal{F}$ kann stets $\mathcal{P}(\Omega)$ angenommen werden.
  Insbesondere ist also jede beliebige Teilmenge von Ω ein Ereignis, dessen Eintrittswahrscheinlichkeit berechnet werden kann.

Laplacesche Wahrscheinlichkeitsräume II

- Das **Laplacesche Wahrscheinlichkeitsmaß** P ergibt sich dann als:

$$P : \mathcal{F} \to \mathbb{R}; P(A) = \frac{\#A}{\#\Omega} = \frac{|A|}{|\Omega|}$$

- Wahrscheinlichkeit $P(A)$ eines Ereignisses A ist also der Quotient

$$\frac{\text{Anzahl der im Ereignis } A \text{ enthaltenen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}}$$

bzw.

$$\frac{\text{Anzahl der (für Ereignis } A\text{) günstigen Fälle}}{\text{Anzahl der (insgesamt) möglichen Fälle}} .$$

- Einzige Schwierigkeit: **Zählen!**

Kombinatorik

- Disziplin, die sich mit „Zählen“ beschäftigt: **Kombinatorik**
- Verwendung allgemeiner Prinzipien und Modelle als Hilfestellung zum Zählen in konkreten Anwendungen.

Satz 6.1 (Additionsprinzip, Multiplikationsprinzip)

Sei $r \in \mathbb{N}$, seien $M_1, M_2, \ldots, M_r$ (jeweils) endliche Mengen.

- *Ist $M_i \cap M_j = \emptyset$ für alle $i \neq j$, dann gilt das* **Additionsprinzip**

$$|M_1 \cup M_2 \cup \cdots \cup M_r| = |M_1| + |M_2| + \cdots + |M_r| \ .$$

- *Mit $M_1 \times M_2 \times \cdots \times M_r := \{(m_1, \ldots, m_r) \mid m_1 \in M_1, \ldots, m_r \in M_r\}$ gilt das* **Multiplikationsprinzip**

$$|M_1 \times M_2 \times \cdots \times M_r| = |M_1| \cdot |M_2| \cdot \ldots \cdot |M_r|$$

und im Fall $M = M_1 = M_2 = \cdots = M_r$ mit $M^r := \underbrace{M \times M \times \cdots \times M}_{r-\text{mal}}$ spezieller

$|M^r| = |M|^r$.

Definition 6.2 (Fakultät, Binomialkoeffizient)

1. Mit $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ sei die Menge der natürlichen Zahlen einschließlich Null bezeichnet.
2. Für jedes $n \in \mathbb{N}_0$ definiert man die Zahl $n! \in \mathbb{N}$ (gelesen „n-**Fakultät**“) rekursiv durch
   - $0! := 1$ und
   - $(n+1)! := (n+1) \cdot n!$ für alle $n \in \mathbb{N}_0$.
3. Für $n, r \in \mathbb{N}_0$ mit $0 \leq r \leq n$ definiert man die Zahl $(n)_r$ (gelesen „n tief r“) durch
$$(n)_r := \frac{n!}{(n-r)!} = n \cdot (n-1) \cdot \ldots \cdot (n-r+1) \ .$$
4. Für $n, r \in \mathbb{N}_0$ und $0 \leq r \leq n$ definiert man den **Binomialkoeffizienten** $\binom{n}{r}$ (gelesen „n über r“) durch
$$\binom{n}{r} := \frac{n!}{(n-r)! \cdot r!} \ .$$

Modelle zum Zählen I

- Gebräuchliches (Hilfs-)Modell zum Zählen: **Urnenmodell**:

  Wie viele Möglichkeiten gibt es, r mal aus einer Urne mit n unterscheidbaren (z.B. von 1 bis n nummerierten) Kugeln zu ziehen?

- Varianten:
  - Ist die Reihenfolge der Ziehungen relevant?
  - Wird die Kugel nach jeder Ziehung wieder in die Urne zurückgelegt?

Modelle zum Zählen II

- *Alternatives Modell:*
  Wie viele Möglichkeiten gibt es, r Murmeln in n unterscheidbare (z. B. von 1 bis n nummerierte) Schubladen zu verteilen.

  Achtung: Auch als weiteres Urnenmodell (Verteilen von r Kugeln auf n Urnen) geläufig!
- Varianten:
  - Sind (auch) die Murmeln unterscheidbar?
  - Dürfen mehrere Murmeln pro Schublade (Mehrfachbelegungen) vorhanden sein?
- Beide Modelle entsprechen sich (einschließlich ihrer Varianten)!
- Im Folgenden (zur Vereinfachung der Darstellung):
  Identifizieren von endlichen Mengen der Mächtigkeit n mit der Menge $\mathbb{N}_n := \{1, 2, \ldots, n\}$ der ersten n natürlichen Zahlen.

Variante I

„geordnete Probe (Variation) mit Wiederholung“

- Ziehen **mit Zurücklegen** und **mit** Berücksichtigung der **Reihenfolge**.
  Für jede der r Ziehungen n Möglichkeiten, Multiplikationsprinzip anwenden.
- Anzahl der Möglichkeiten:

$$ {}^{w}_{n}V_r := \underbrace{n \cdot n \cdot \ldots \cdot n}_{r \text{ Faktoren}} = n^r $$

- Formale Darstellung aller Möglichkeiten:

$$ \{(m_1, \ldots, m_r) \mid m_1, \ldots, m_r \in \mathbb{N}_n\} = \mathbb{N}_n^r $$

- gleichbedeutend: Verteilen von **unterscheidbaren** Murmeln **mit** Zulassung von **Mehrfachbelegungen**

Variante II

„geordnete Probe (Variation) ohne Wiederholung“

- Ziehen **ohne Zurücklegen** und **mit** Berücksichtigung der **Reihenfolge**.

  Für erste Ziehung n Möglichkeiten, für zweite $n-1$, ..., für r-te Ziehung $n-r+1$ Möglichkeiten, Multiplikationsprinzip anwenden.
- Anzahl der Möglichkeiten:

$$ {}_nV_r := n \cdot (n-1) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!} = (n)_r $$

- Formale Darstellung aller Möglichkeiten:

$$ \{(m_1, \ldots, m_r) \in \mathbb{N}_n^r \mid m_i \neq m_j \text{ für } i \neq j, 1 \leq i, j \leq r\} $$

- gleichbedeutend: Verteilen von **unterscheidbaren** Murmeln **ohne** Zulassung von **Mehrfachbelegungen**
- **Spezialfall:** $n = r$
  - $\rightsquigarrow$ $n!$ verschiedene Möglichkeiten
  - entspricht (Anzahl der) möglichen Anordnungen (**Permutationen**) von n unterscheidbaren Kugeln bzw. n unterscheidbaren Murmeln

Variante III

„ungeordnete Probe (Kombination) ohne Wiederholung“

- Ziehen **ohne Zurücklegen** und **ohne** Berücksichtigung der **Reihenfolge**.

  Wie in Variante 2: Für erste Ziehung n Möglichkeiten, für zweite $n-1$, ..., für r-te Ziehung $n-r+1$ Möglichkeiten, Multiplikationsprinzip anwenden; **aber:** *je $r!$ Möglichkeiten unterscheiden sich nur durch die (nicht zu berücksichtigende!) Reihenfolge.*
- Anzahl der Möglichkeiten:

$$ {}_nC_r := \frac{n \cdot (n-1) \cdot \ldots \cdot (n-r+1)}{r \cdot (r-1) \cdot \ldots \cdot 1} = \frac{n!}{r!(n-r)!} = \binom{n}{r} $$

- Formale Darstellung aller Möglichkeiten:

$$ \{(m_1, \ldots, m_r) \in \mathbb{N}_n^r \mid m_1 < m_2 < \cdots < m_r\} $$

- gleichbedeutend: Verteilen von **nicht unterscheidbaren** Murmeln **ohne** Zulassung von **Mehrfachbelegungen**
- Häufig Anwendung bei *gleichzeitigem* Ziehen von r aus n Objekten bzw. simultane Auswahl von r aus n Plätzen; Anzahl der Möglichkeiten entspricht Anzahl r-elementiger Teilmengen aus n-elementiger Menge.

Variante IV

„ungeordnete Probe (Kombination) mit Wiederholung“

- Ziehen **mit Zurücklegen** und **ohne** Berücksichtigung der **Reihenfolge**.

  Verständnis schwieriger!
  Vorstellung: Erstelle „Strichliste“ mit r Strichen, verteilt auf n Felder (eines für jede Kugel) $\rightsquigarrow$ r Striche zwischen $n-1$ „Feldbegrenzungen“. Anzahl Möglichkeiten entspricht Anzahl der Möglichkeiten, die r Striche in der „Reihung“ der $n-1+r$ Striche&Feldbegrenzungen zu positionieren.
- Anzahl der Möglichkeiten:

$$ {}^{w}_{n}C_{r} := \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!} = \frac{(n+r-1)\cdot(n+r-2)\cdot\ldots\cdot n}{r\cdot(r-1)\cdot\ldots\cdot 1} $$

- Formale Darstellung aller Möglichkeiten:

$$ \{(m_1,\ldots,m_r)\in\mathbb{N}_n^r \mid m_1\le m_2\le\cdots\le m_r\} $$

- gleichbedeutend: Verteilen von **nicht unterscheidbaren** Murmeln **mit** Zulassung von **Mehrfachbelegungen**
- **Achtung:** Üblicherweise **nicht** geeignet als Ω in Laplaceschen Wahrscheinlichkeitsräumen, da die verschiedenen Möglichkeiten bei üblichen Ziehungsvorgängen nicht alle gleichwahrscheinlich sind!

Übersicht der Varianten I–IV

vgl. Ulrich Krengel, Einführung in die W.-Theorie und Statistik, 7. Aufl., Vieweg, Wiesbaden, 2003

r unterscheidbare Kugeln aus Urne mit n (unterscheidbaren) Kugeln	**mit** Zurücklegen	**ohne** Zurücklegen	
mit Berücksichtigung der Reihenfolge	Variante I $\quad {}^{w}_{n}V_r = n^r$	Variante II $\quad {}_{n}V_r = (n)_r$	**unterscheidbare** Murmeln
ohne Berücksichtigung der Reihenfolge	Variante IV $\quad {}^{w}_{n}C_r = \binom{n+r-1}{r}$	Variante III $\quad {}_{n}C_r = \binom{n}{r}$	**nicht unterscheidbare** Murmeln
	mit Mehrfachbesetzung	**ohne** Mehrfachbesetzung	r Murmeln in n unterscheidbare Schubladen

Bemerkungen

- Wird ohne Zurücklegen gezogen, muss natürlich $r \leq n$ gefordert werden!
  (Es können insgesamt nicht mehr Kugeln entnommen werden, als zu Beginn in der Urne enthalten waren.)
- Werden **alle** Kugeln ohne Zurücklegen unter Berücksichtigung der Reihenfolge entnommen, wird häufig folgende Verallgemeinerung betrachtet:
  - Nicht alle n Kugeln sind unterscheidbar (durchnummeriert).
  - Es gibt $m < n$ (unterscheidbare) Gruppen von Kugeln, die jeweils $n_1, n_2, \ldots, n_m$ **nicht** unterscheidbare Kugeln umfassen (mit $n = \sum_{i=1}^{m} n_i$).
  - Da es jeweils $n_1!, n_2!, \ldots, n_m!$ nicht unterscheidbare Anordnungen der Kugeln innerhalb der Gruppen gibt, reduziert sich die Anzahl der insgesamt vorhandenen Möglichkeiten von ${}_nP := {}_nV_n = n!$ auf

$$ {}_nP_{n_1,n_2,\ldots,n_m} := \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!} \ . $$

  - Typische Anwendung: Buchstabenanordnungen bei „Scrabble“
  - Nenner von ${}_nP_{n_1,n_2,\ldots,n_m}$ liefert Anzahl der Möglichkeiten für jeweils nicht unterscheidbare Anordnungen.

- Anwendung der Kombinatorik in Laplaceschen Wahrscheinlichkeitsräumen:
  - Auswahl eines geeigneten Ergebnisraums Ω mit **gleichwahrscheinlichen** Ausgängen des Zufallsexperiments.
  - Bestimmen von $|\Omega|$ mit kombinatorischen Hilfsmitteln.
  - Bestimmen von $|A|$ für interessierende Ereignisse $A \subseteq \Omega$ mit kombinatorischen Hilfsmitteln.
- Häufig gibt es nicht **die** richtige Lösung, sondern mehrere, da oft mehrere Modelle (mehr oder weniger) geeignet sind.
- Wird aber beispielsweise Ω unter Berücksichtigung der Ziehungsreihenfolge konstruiert, obwohl die Reihenfolge für das interessierende Ereignis A unwichtig ist, müssen unterschiedliche mögliche Anordnungen bei der Konstruktion von A ebenfalls berücksichtigt werden!
- Trotz vorhandener (und nützlicher) Modelle:

  Richtiges Zählen hat häufig „Knobelcharacter“, stures Einsetzen in Formeln selten ausreichend, Mitdenken erforderlich!

Beispiele

- **Lottospiel „6 aus 49“** (ohne Berücksichtigung einer Zusatzzahl)
  - Interessierendes Ereignis A: (Genau) 3 „Richtige“
  - $|\Omega| = \binom{49}{6} = 13983816$, $|A| = \binom{6}{3} \cdot \binom{43}{3} = 246820$

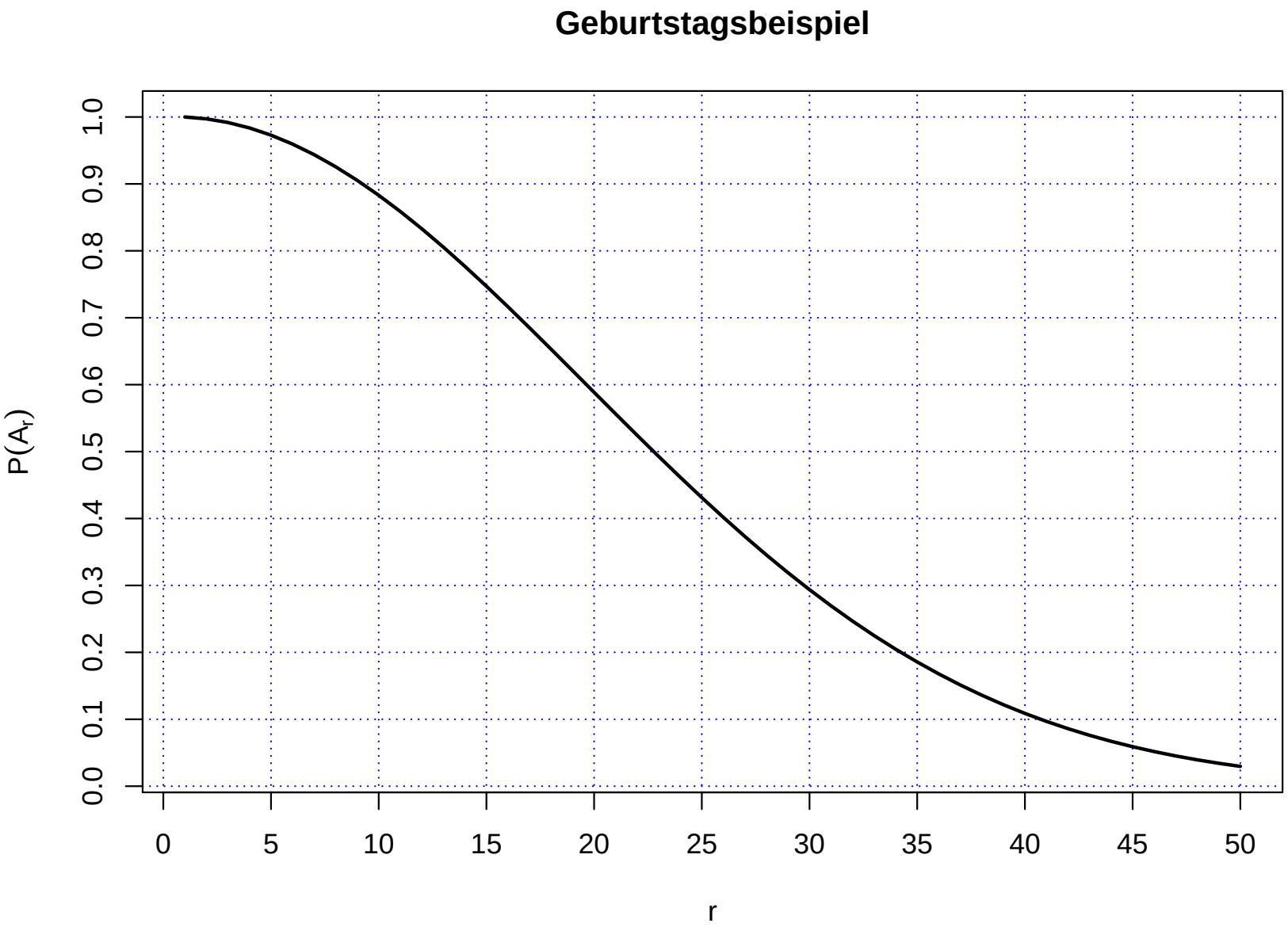
  $\Rightarrow P(A) = \dfrac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}} = 0.01765 = 1.765\%$
- Anzahl der Möglichkeiten bei **zweimaligem Würfelwurf**
  - wenn die Reihenfolge irrelevant (z.B. bei gleichzeitigem Werfen) ist:

    $|\Omega| = \binom{6+2-1}{2} = 21$

    Vorsicht: nicht alle Ergebnisse gleichwahrscheinlich!
  - wenn die Reihenfolge relevant ist:

    $|\Omega| = 6^2 = 36$
- **Geburtstage**
  Zusammensetzung der Geburtstage (ohne Jahreszahl; Vernachlässigung von Schaltjahren) bei r Personen (mit Reihenfolgeberücksichtigung)
  - Interessierendes Ereignis A_r: alle r Personen haben verschiedene Geburtstage
  - $|\Omega_r| = 365^r$, $|A_r| = (365)_r$ für $r \leq 365$, $|A_r| = 0$ sonst.

  $\Rightarrow P(A_r) = \frac{(365)_r}{365^r}$ für $r \leq 365$, $P(A_r) = 0$ sonst.


Geburtstagsbeispiel
P(A_r)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30
35
40
45
50
r

Allgemeine diskrete Wahrscheinlichkeitsräume I

- Verallgemeinerung von Laplaceschen Wahrscheinlichkeitsräumen: **Diskrete Wahrscheinlichkeitsräume**
- Ω endlich (mit $|\Omega| = n$) oder abzählbar unendlich.
- Nach wie vor: Jeder Teilmenge von Ω soll eine Wahrscheinlichkeit zugeordnet werden können, also $\mathcal{F} = \mathcal{P}(\Omega)$.
- Damit: Jedem Ergebnis kann Wahrscheinlichkeit zugeordnet werden.
- **Aber:**
  Ergebnisse (auch für endliches Ω) nicht mehr (zwingend) gleichwahrscheinlich.

Allgemeine diskrete Wahrscheinlichkeitsräume II

Definition 6.3 (Diskreter Wahrscheinlichkeitsraum)

Sei $\Omega \neq \emptyset$ endlich oder abzählbar unendlich und $\mathcal{F} = \mathcal{P}(\Omega)$. Sei $p : \Omega \to [0,1]$ eine Abbildung mit $p(\omega) \geq 0$ für alle $\omega \in \Omega$ und $\sum_{\omega \in \Omega} p(\omega) = 1$. Dann heißen das durch

$$P : \mathcal{F} \to \mathbb{R}; P(A) := \sum_{\omega \in A} p(\omega)$$

definierte Wahrscheinlichkeitsmaß sowie der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ **diskret**, p heißt auch **Wahrscheinlichkeitsfunktion**.

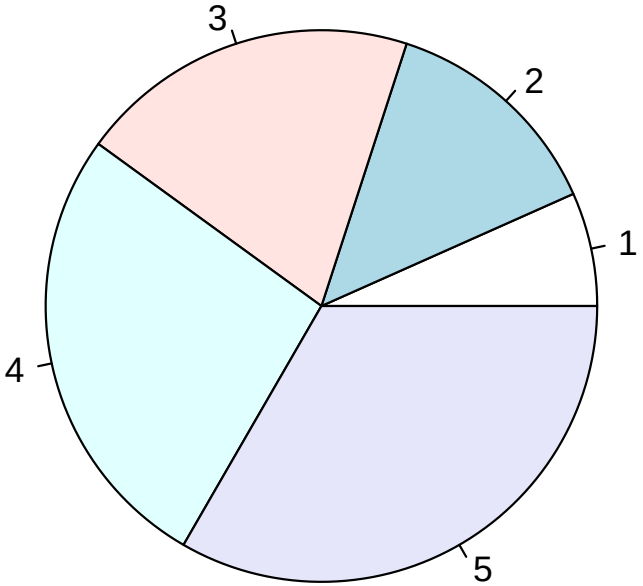
- Ein Laplacescher Wahrscheinlichkeitsraum ist also ein diskreter Wahrscheinlichkeitsraum mit $p : \Omega \to [0,1]; p(\omega) = \frac{1}{|\Omega|}$.

Beispiel I

- **„Glücksrad“** mit folgendem Aufbau:
  n Segmente, nummeriert von 1 bis n, deren Größe proportional zur Nummer ist (und die das Rad vollständig ausfüllen).
  - $\Omega = \{1, \ldots, n\}$
  - Mit $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ erhält man für die Wahrscheinlichkeitsfunktion

$$p : \Omega \to [0,1]; p(\omega) = \frac{\omega}{\frac{n(n+1)}{2}} = \frac{2\omega}{n(n+1)}$$

- Beispiel für $n = 5$:



Beispiel II

- **Münze** (mit „Wappen“ und „Zahl“) solange werfen, bis **zum ersten Mal „Zahl“** zu sehen ist:
  Mögliche Ergebnisse: {Z,WZ,WWZ,WWWZ,...}, *im Folgenden repräsentiert durch Anzahl der Würfe (insgesamt).*
  - $\Omega = \mathbb{N}$
  - Wahrscheinlichkeitsfunktion (bei „fairer“ Münze)

$$p : \Omega \to [0,1]; p(\omega) = \left(\frac{1}{2}\right)^{\omega} = \frac{1}{2^{\omega}}$$

- Wahrscheinlichkeit, höchstens n Würfe zu benötigen:

$$P(\{1,\ldots,n\}) = \sum_{\omega=1}^{n} p(\omega) = \sum_{\omega=1}^{n} \left(\frac{1}{2}\right)^{\omega} = \frac{\frac{1}{2}\cdot\left(1-\left(\frac{1}{2}\right)^{n}\right)}{1-\frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^{n}$$

Hier verwendet: Geometrische Summenformel $\sum_{i=1}^{n} q^i = \frac{q\cdot(1-q^n)}{1-q}$ *(für* $q \neq 1$*)*

Inhaltsverzeichnis

(Ausschnitt)

7 Bedingte Wahrscheinlichkeit und Unabhängigkeit
- Bedingte Wahrscheinlichkeiten
- Stochastische Unabhängigkeit

Bedingte Wahrscheinlichkeiten I

- Oft interessant: Eintrittswahrscheinlichkeit eines Ereignisses B, wenn schon bekannt ist, dass ein (anderes) Ereignis A, in diesem Zusammenhang auch **Bedingung** genannt, eingetreten ist.
- Beispiele:
  - Werfen eines Würfels:
    Wahrscheinlichkeit dafür, eine 2 gewürfelt zu haben, falls bekannt ist, dass eine ungerade Zahl gewüfelt wurde $\rightsquigarrow 0$
    Wahrscheinlichkeit dafür, eine 3 oder 4 gewürfelt zu haben, falls bekannt ist, dass eine Zahl größer als 3 gewürfelt wurde $\rightsquigarrow \frac{1}{3}$
  - Wahrscheinlichkeit, aus einer Urne mit 2 schwarzen und 2 weißen Kugeln bei Ziehung ohne Zurücklegen in der zweiten Ziehung eine weiße Kugel zu ziehen, wenn bekannt ist, dass in der ersten Ziehung eine schwarze Kugel gezogen wurde $\rightsquigarrow \frac{2}{3}$
  - Wahrscheinlichkeit, dass man beim Poker-Spiel (Texas Hold'em) zum Ende des Spiels einen Vierling als höchstes Blatt hat, wenn man bereits ein Paar in der Starthand hält $\rightsquigarrow 0.8424\%$

Bedingte Wahrscheinlichkeiten II

- Offensichtlich sind diese Wahrscheinlichkeiten nur dann interessant, wenn das Ereignis A auch mit positiver Wahrscheinlichkeit eintritt.
- *Rückblick:*
  In deskriptiver Statistik: Begriff der *bedingten relativen Häufigkeiten*
  - $r(a_i|Y = b_j) := \frac{h_{ij}}{h_{\cdot j}} = \frac{r_{ij}}{r_{\cdot j}}$
  - $r(b_j|X = a_i) := \frac{h_{ij}}{h_{i\cdot}} = \frac{r_{ij}}{r_{i\cdot}}$

  für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.
- Analog zur Einschränkung der statistischen Masse bei bedingten relativen Häufigkeiten: Einschränkung von Ω auf bedingendes Ereignis A.

Bedingte Wahrscheinlichkeiten III

- Zur Ermittlung der bedingten Wahrscheinlichkeit, eine 3 oder 4 gewürfelt zu haben (Ereignis $B = \{3, 4\}$), falls bekannt ist, dass eine Zahl größer als 3 gewürfelt wurde (Ereignis $A = \{4, 5, 6\}$):
  - Berechnung der Wahrscheinlichkeit des gemeinsamen Eintretens der Bedingung A und des interessierenden Ereignisses B: $P(A \cap B) = P(\{3, 4\} \cap \{4, 5, 6\}) = P(\{4\}) = \frac{1}{6}$
  - Berechnung der Wahrscheinlichkeit des Eintretens der Bedingung A: $P(A) = P(\{4, 5, 6\}) = \frac{3}{6}$
  - Analog zum Fall relativer bedingter Häufigkeiten: Berechnung des Verhältnisses $\frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$.

Bedingte Wahrscheinlichkeiten IV

Definition 7.1

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ mit $P(A) > 0$. Dann heißt

$$P(B|A) := \frac{P(B \cap A)}{P(A)}$$

die **bedingte Wahrscheinlichkeit** von B unter der Bedingung A.

Satz 7.2

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$ mit $P(A) > 0$. Dann ist die Abbildung

$$P(\,\cdot\,|A) : \mathcal{F} \to \mathbb{R}; B \mapsto P(B|A)$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$ gemäß Definition 5.4, also auch $(\Omega, \mathcal{F}, P(\,\cdot\,|A))$ ein Wahrscheinlichkeitsraum.

Bedingte Wahrscheinlichkeiten V

- **Wichtig:** Satz 7.2 gilt nur dann, wenn das bedingende Ereignis A festgehalten wird. Bei der Abbildung $P(A|\cdot) : \mathcal{F} \to \mathbb{R}$ handelt es sich im Allgemeinen **nicht** (und bei strenger Auslegung von Definition 7.1 sogar **nie**) um ein Wahrscheinlichkeitsmaß.
- Aus Definition 7.1 folgt unmittelbar

$$P(A \cap B) = P(A) \cdot P(B|A) \tag{1}$$

  für $A, B \in \mathcal{F}$ mit $P(A) > 0$.
- Obwohl $P(B|A)$ nur für $P(A) > 0$ definiert ist, bietet es sich aus technischen Gründen an, Schreibweisen wie in Gleichung (1) auch für $P(A) = 0$ zuzulassen. In diesem Fall sollen Produkte, in denen neben (eigentlich) nicht definierten bedingten Wahrscheinlichkeiten mindestens ein Faktor 0 auftritt, definitionsgemäß ebenfalls den Wert 0 annehmen.

Bedingte Wahrscheinlichkeiten VI

- Mit dieser Konvention für Bedingungen mit Eintrittswahrscheinlichkeit 0 lässt sich durch wiederholtes Einsetzen von Gleichung (1) leicht der **Multiplikationssatz**

  $$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

  für $A_1, \ldots, A_n \in \mathcal{F}$ herleiten.
- Mit dem Multiplikationssatz können insbesondere Wahrscheinlichkeiten in sogenannten Baumdiagrammen (für „mehrstufige“ Zufallsexperimente) ausgewertet werden.
- In vielen Anwendungen sind häufig vor allem bedingte Wahrscheinlichkeiten bekannt (bzw. werden als bekannt angenommen).
- Es ist nicht immer einfach, die Angabe bedingter Wahrscheinlichkeiten auch als solche zu erkennen!

Bedingte Wahrscheinlichkeiten VII

- Beispiel:
  In einer kleinen Druckerei stehen 3 Druckmaschinen unterschiedlicher Kapazität zur Verfügung:
  - Maschine 1, mit der 50% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.1% ein fehlerhaftes Ergebnis,
  - Maschine 2, mit der 30% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.25% ein fehlerhaftes Ergebnis,
  - Maschine 3, mit der 20% aller Druckjobs gedruckt werden, produziert mit einer Wahrscheinlichkeit von 0.5% ein fehlerhaftes Ergebnis.

  Sind die interessierenden Ereignisse gegeben durch
  - M_i: Maschine i wird zur Produktion des Druckjobs eingesetzt ($i \in \{1, 2, 3\}$),
  - F: Die Produktion des Druckjobs ist fehlerbehaftet,

  so sind insgesamt bekannt:

$$P(M_1) = 0.5 \qquad P(M_2) = 0.3 \qquad P(M_3) = 0.2$$
$$P(F|M_1) = 0.001 \qquad P(F|M_2) = 0.0025 \qquad P(F|M_3) = 0.005$$

Bedingte Wahrscheinlichkeiten VIII

- In der Situation des Druckmaschinen-Beispiels interessiert man sich häufig für die *unbedingte* Wahrscheinlichkeit, einen fehlerhaften Druckjob zu erhalten, also für $P(F)$.
- Diese Wahrscheinlichkeit kann mit einer Kombination von Gleichung (1) auf Folie 174 und der letzten Rechenregel von Folie 148 berechnet werden:

Satz 7.3 (Satz der totalen Wahrscheinlichkeit)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathcal{F}$ (für $n \in \mathbb{N}$) eine Zerlegung von Ω, es gelte also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{j=1}^{n} A_j = \Omega$. Für beliebige Ereignisse $B \in \mathcal{F}$ gilt dann

$$P(B) = \sum_{j=1}^{n} P(B|A_j) \cdot P(A_j) \ .$$

Bedingte Wahrscheinlichkeiten IX

- Im Druckmaschinen-Beispiel erhält man so:

$$\begin{aligned} P(F) &= P(F|M_1) \cdot P(M_1) + P(F|M_2) \cdot P(M_2) + P(F|M_3) \cdot P(M_3) \\ &= 0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2 \\ &= 0.00225 = 0.225\% \end{aligned}$$

- Die Wahrscheinlichkeit von $A \cap B$ kann mit Hilfe bedingter Wahrscheinlichkeiten in zwei Varianten berechnet werden:

$$P(B|A) \cdot P(A) = P(A \cap B) = P(A|B) \cdot P(B)$$

- Dies kann (bei Kenntnis der restlichen beteiligten Wahrscheinlichkeiten!) dazu benutzt werden, Bedingung und interessierendes Ereignis umzudrehen. Ist z.B. $P(B|A)$ (sowie $P(A)$ und $P(B)$) gegeben, so erhält man $P(A|B)$ durch

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \ .$$

Bedingte Wahrscheinlichkeiten X

- Wird dabei die unbedingte Wahrscheinlichkeit $P(B)$ mit dem Satz der totalen Wahrscheinlichkeit (Satz 7.3) berechnet, so erhält man:

Satz 7.4 (Formel von Bayes)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathcal{F}$ (für $n \in \mathbb{N}$) eine Zerlegung von Ω, es gelte also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{j=1}^n A_j = \Omega$. Sei $B \in \mathcal{F}$ ein weiteres Ereignis mit $P(B) > 0$. Dann gilt:

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{j=1}^n P(B|A_j) \cdot P(A_j)}, \qquad k \in \{1, \ldots, n\}$$

- Anwendung der Formel von Bayes im Druckmaschinen-Beispiel:

  Frage: Wenn eine Fehlproduktion aufgetreten ist, mit welchen Wahrscheinlichkeiten sind dann die verschiedenen Druckmaschinen für den Fehldruck „verantwortlich“?

Bedingte Wahrscheinlichkeiten XI

- Es interessiert nun also $P(M_i|F)$ für $i \in \{1, 2, 3\}$.
- Mit Formel von Bayes (ohne Verwendung des Zwischenergebnisses $P(F)$):

$$
\begin{aligned}
P(M_1|F) &= \frac{P(F|M_1) \cdot P(M_1)}{\sum_{i=1}^{3} P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.001 \cdot 0.5}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{2}{9} \\
P(M_2|F) &= \frac{P(F|M_2) \cdot P(M_2)}{\sum_{i=1}^{3} P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.0025 \cdot 0.3}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{3}{9} \\
P(M_3|F) &= \frac{P(F|M_3) \cdot P(M_3)}{\sum_{i=1}^{3} P(F|M_j) \cdot P(M_j)} \\
&= \frac{0.005 \cdot 0.2}{0.001 \cdot 0.5 + 0.0025 \cdot 0.3 + 0.005 \cdot 0.2} = \frac{4}{9}
\end{aligned}
$$

Beispiel: Fehler bei bedingten Wahrscheinlichkeiten I

(aus Walter Krämer: Denkste!, Piper, München, 2000)

- Häufig (auch in den Medien!) ist der Fehler zu beobachten, dass das bedingende und das eigentlich interessierende Ereignis vertauscht werden.
- Beispiel (Schlagzeile in einer Ausgabe der ADAC-Motorwelt):

  Der Tod fährt mit!
  Vier von zehn tödlich verunglückten Autofahrern trugen keinen Sicherheitsgurt!

- Bezeichnet S das Ereignis „Sicherheitsgurt angelegt“ und T das Ereignis „Tödlicher Unfall“, so ist hier die Wahrscheinlichkeit $P(\overline{S}|T) = 0.4$, also die Wahrscheinlichkeit, keinen Sicherheitsgurt angelegt zu haben, falls man bei einen Unfall tödlich verunglückt ist, mit 40% angegeben.
- Was soll die Schlagzeile vermitteln?
  - Unwahrscheinlich: Anschnallen ist gefährlich, da 6 von 10 verunglückten Autofahren einen Sicherheitsgurt trugen.
  - Wohl eher: Anschnallen ist nützlich!
- **Aber:** Zitierte Wahrscheinlichkeit ist absolut ungeeignet, um Nutzen des Anschnallens zu untermauern!

Beispiel: Fehler bei bedingten Wahrscheinlichkeiten II

(aus Walter Krämer: Denkste!, Piper, München, 2000)

- Stattdessen interessant: $P(T|\overline{S})$ vs. $P(T|S)$
  (*Wie stehen Wahrscheinlichkeiten, bei nicht angelegtem bzw. angelegtem Sicherheitsgurt tödlich zu verunglücken, zueinander?*)
- Aus $P(\overline{S}|T)$ kann Verhältnis $\frac{P(T|\overline{S})}{P(T|S)}$ nur berechnet werden, falls die Wahrscheinlichkeit bekannt ist, mit der ein Autofahrer angeschnallt ist:

$$\begin{aligned} P(T|S) &= \frac{P(T \cap S)}{P(S)} = \frac{P(S|T) \cdot P(T)}{P(S)} \\ P(T|\overline{S}) &= \frac{P(T \cap \overline{S})}{P(\overline{S})} = \frac{P(\overline{S}|T) \cdot P(T)}{P(\overline{S})} \\ \Rightarrow \frac{P(T|\overline{S})}{P(T|S)} &= \frac{P(\overline{S}|T) \cdot P(S)}{P(\overline{S}) \cdot P(S|T)} \end{aligned}$$

- Für $P(S) = 0.9$: $\frac{P(T|\overline{S})}{P(T|S)} = 6$, also Risiko ohne Gurt 6 mal höher als mit Gurt.
- Für $P(S) = 0.2$: $\frac{P(T|\overline{S})}{P(T|S)} = \frac{1}{6}$, also Risiko ohne Gurt 6 mal niedriger als mit Gurt.

Stochastische Unabhängigkeit

- Analog zur Unabhängigkeit von Merkmalen in deskriptiver Statistik:

Definition 7.5 (Stochastische Unabhängigkeit)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. A und B heißen **stochastisch unabhängig** bezüglich P, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

- Ebenfalls analog zur deskriptiven Statistik sind Ereignisse genau dann unabhängig, wenn sich ihre unbedingten Wahrscheinlichkeiten nicht von den bedingten Wahrscheinlichkeiten — sofern sie definiert sind — unterscheiden:

Satz 7.6

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. Dann gilt

- *falls $P(A) > 0$:*
  A und B sind stochastisch unabhängig bzgl. $P \Leftrightarrow P(B|A) = P(B)$
- *falls $P(B) > 0$:*
  A und B sind stochastisch unabhängig bzgl. $P \Leftrightarrow P(A|B) = P(A)$

- Der Begriff „Stochastische Unabhängigkeit“ ist auch für Familien mit mehr als zwei Ereignissen von Bedeutung:

Definition 7.7

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $(A_i)_{i \in I}$ eine Familie von Ereignissen aus $\mathcal{F}$. Die Familie heißt **stochastisch unabhängig**, wenn für jede endliche Teilmenge $K \subseteq I$ gilt:

$$P(\bigcap_{i \in K} A_i) = \prod_{i \in K} P(A_i)$$

- Besteht die Familie $(A_i)_{i \in I}$ in Definition 7.7 nur aus den drei Ereignissen A_1, A_2, A_3, so ist die Familie also stochastisch unabhängig, falls
  1. $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$
  2. $P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$
  3. $P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$
  4. $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$

  gilt.
  Insbesondere genügt **nicht** die paarweise stochastische Unabhängigkeit der Ereignisse A_1, A_2, A_3!

Inhaltsverzeichnis

(Ausschnitt)

8 Messbarkeit und Bildwahrscheinlichkeit
- Messbare Abbildungen
- Bildwahrscheinlichkeit

Messbare Abbildungen I

- Häufig von Interesse:
  Nicht der Ausgang eines Zufallsexperiments selbst, sondern eine **Funktion** dieses Ausgangs, d.h. eine Abbildung $X : \Omega \to \Omega'$ vom Ergebnisraum Ω in eine andere Menge Ω'.
- Beispiele:
  - Augensumme beim gleichzeitigen Werfen von zwei Würfeln
  - Anzahl „Wappen“ bei mehrmaligem Münzwurf
  - Resultierender Gewinn zu gegebener Tippreihe beim Lottospiel
  - Anzahl der weißen Kugeln bei wiederholter Ziehung von Kugeln aus Urne mit schwarzen und weißen Kugeln (mit oder ohne Zurücklegen)
- Wahrscheinlichkeitsbegriff des ursprünglichen Zufallsexperiments $(\Omega, \mathcal{F}, P)$ soll in die Wertemenge/Bildmenge Ω' der Abbildung X „transportiert“ werden.

Messbare Abbildungen II

- Bestimmung der Wahrscheinlichkeit eines Ereignisses $A' \subseteq \Omega'$ wie folgt:
  - Feststellen, welche Elemente $\omega \in \Omega$ unter der Abbildung X auf Elemente $\omega' \in A'$ abgebildet werden.
  - Wahrscheinlichkeit von A' ergibt sich dann als Wahrscheinlichkeit der erhaltenen Teilmenge $\{\omega \in \Omega \,|\, X(\omega) \in A'\}$ von Ω, also als $P(\{\omega \in \Omega \,|\, X(\omega) \in A'\})$.

Definition 8.1 (Urbild)

Es seien $X : \Omega \to \Omega'$ eine Abbildung, $A' \subseteq \Omega'$. Dann heißt

$$X^{-1}(A') := \{\omega \in \Omega \,|\, X(\omega) \in A'\}$$

das **Urbild** von A' bzgl. X.

Messbare Abbildungen III

- Zur Funktionsfähigkeit des „Wahrscheinlichkeitstransports“ nötig: Geeignetes Mengensystem ($\rightsquigarrow$ σ-Algebra) $\mathcal{F}'$ über Ω', welches „kompatibel“ zur σ-Algebra $\mathcal{F}$ über Ω und zur Abbildung X ist.
- *„Kompatibilitätsanforderung“ ergibt sich nach Konstruktion:* Damit $P(\{\omega \in \Omega \,|\, X(\omega) \in A'\}) = P(X^{-1}(A'))$ für $A' \in \mathcal{F}'$ definiert ist, muss $X^{-1}(A') \in \mathcal{F}$ gelten für alle $A' \in \mathcal{F}'$.
- Beschriebene Eigenschaft der Kombination $\mathcal{F}$, $\mathcal{F}'$ und X heißt **Messbarkeit**:

Definition 8.2 (Messbarkeit, messbare Abbildung)

Es seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ zwei Messräume. Eine Abbildung $X : \Omega \to \Omega'$ heißt $\mathcal{F} - \mathcal{F}'-$**messbar**, wenn gilt:

$$X^{-1}(A') \in \mathcal{F} \text{ für alle } A' \in \mathcal{F}' \ .$$

Bildwahrscheinlichkeit

Definition 8.3 (Bildwahrscheinlichkeit)

Es seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{F}')$ ein Messraum, X eine $\mathcal{F} - \mathcal{F}'-$messbare Abbildung. Dann heißt das durch

$$P_X : \mathcal{F}' \to \mathbb{R}; P_X(A') := P(X^{-1}(A'))$$

definierte Wahrscheinlichkeitsmaß P_X **Bildwahrscheinlichkeit** von P bezüglich X. $(\Omega', \mathcal{F}', P_X)$ ist damit ebenfalls ein Wahrscheinlichkeitsraum.

- Gilt $\mathcal{F} = \mathcal{P}(\Omega)$, so ist offensichtlich jede Abbildung $X : \Omega \to \Omega'$ $\mathcal{F} - \mathcal{F}'-$messbar für beliebige σ-Algebren $\mathcal{F}'$ über Ω'.

Einbettung der deskriptiven Statistik in die Wahrscheinlichkeitsrechnung

- Ist Ω die (endliche) Menge von Merkmalsträgern einer deskriptiven statistischen Untersuchung, $\mathcal{F} = \mathcal{P}(\Omega)$ und P die Laplace-Wahrscheinlichkeit

$$P : \mathcal{P}(\Omega) \to \mathbb{R}; B \mapsto \frac{\#B}{\#\Omega} \ ,$$

  so kann jedes Merkmal X mit Merkmalsraum $A = \{a_1, \ldots, a_m\}$ als $\mathcal{P}(\Omega) - \mathcal{P}(A)-$messbare Abbildung $X : \Omega \to A$ verstanden werden.
- $(A, \mathcal{P}(A), P_X)$ ist dann ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $p(a_j) = r(a_j)$ bzw. — äquivalent — $P_X(\{a_j\}) = r(a_j)$ für $j \in \{1, \ldots, m\}$.
- Durch $(A, \mathcal{P}(A), P_X)$ wird damit die Erhebung des Merkmalswerts eines rein zufällig (gleichwahrscheinlich) ausgewählten Merkmalsträgers modelliert.

Inhaltsverzeichnis

(Ausschnitt)

9 Eindimensionale Zufallsvariablen
- Borelsche sigma-Algebra
- Wahrscheinlichkeitsverteilungen
- Verteilungsfunktionen
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- (Lineare) Abbildungen von Zufallsvariablen
- Momente von Zufallsvariablen
- Quantile von Zufallsvariablen
- Spezielle diskrete Verteilungen
- Spezielle stetige Verteilungen
- Verwendung spezieller Verteilungen

Borelsche σ-Algebra I

- Häufiger Wertebereich von Abbildungen aus Wahrscheinlichkeitsräumen: $\mathbb{R}$
- $\mathcal{P}(\mathbb{R})$ als σ-Algebra (aus technischen Gründen) aber ungeeignet!
- Alternative σ-Algebra über $\mathbb{R}$: „Borelsche“ σ-Algebra $\mathcal{B}$
- $\mathcal{B}$ ist die kleinste σ-Algebra über $\mathbb{R}$, die alle Intervalle

$$\begin{array}{llll} (a,b) & (a,b] & [a,b) & [a,b] \\ (-\infty,a) & (-\infty,a] & (a,\infty) & [a,\infty) \end{array}$$

für $a, b \in \mathbb{R}$ (mit $a < b$) enthält.

Borelsche σ-Algebra II

- Aufgrund der Eigenschaften von σ-Algebren sind auch alle
  - Einpunktmengen $\{x\}$ für $x \in \mathbb{R}$,
  - endliche Mengen $\{x_1, \ldots, x_m\} \subseteq \mathbb{R}$,
  - abzählbar unendliche Mengen sowie endliche und abzählbare Schnitte und Vereinigungen von Intervallen

  in $\mathcal{B}$ enthalten.
- Abbildungen $X : \Omega \to \mathbb{R}$ aus einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ heißen (eindimensionale) **Zufallsvariablen**, wenn sie $\mathcal{F} - \mathcal{B}-$messbar sind:

Eindimensionale Zufallsvariablen und deren Verteilung I

Definition 9.1 (Zufallsvariable, Verteilung, Realisation)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$ eine $\mathcal{F} - \mathcal{B}-$messbare Abbildung. Dann heißen X **(eindimensionale) Zufallsvariable** über $(\Omega, \mathcal{F}, P)$ und die gemäß Definition 8.3 gebildete Bildwahrscheinlichkeit

$$P_X : \mathcal{B} \to \mathbb{R}; B \mapsto P(X^{-1}(B))$$

Wahrscheinlichkeitsverteilung oder kürzer **Verteilung** von X. $(\mathbb{R}, \mathcal{B}, P_X)$ ist damit ebenfalls ein Wahrscheinlichkeitsraum.
Liegt nach Durchführung des Zufallsexperiments $(\Omega, \mathcal{F}, P)$ das Ergebnis $\omega \in \Omega$ vor, so heißt der zugehörige Wert $x = X(\omega)$ die **Realisierung** oder **Realisation** von X.

Eindimensionale Zufallsvariablen und deren Verteilung II

- $P_X(B)$ gibt nach Definition 8.3 für $B \in \mathcal{B}$ die Wahrscheinlichkeit an, als Ausgang des zugrundeliegenden Zufallsexperiments ein $\omega \in \Omega$ zu erhalten, das zu einer Realisation $X(\omega) \in B$ führt.
- Demzufolge sind folgende Kurzschreibweisen geläufig:
  - $P\{X \in B\} := P(\{X \in B\}) := P_X(B)$ für alle $B \in \mathcal{B}$,
  - $P\{X = x\} := P(\{X = x\}) := P_X(\{x\})$ für alle $x \in \mathbb{R}$,
  - $P\{X < x\} := P(\{X < x\}) := P_X((-\infty, x))$ für alle $x \in \mathbb{R}$, analog für $P\{X \leq x\}$, $P\{X > x\}$ und $P\{X \geq x\}$.
- In vielen Anwendungen interessiert man sich nur noch für die Bildwahrscheinlichkeit P_X der Zufallsvariablen X bzw. den (resultierenden) Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}, P_X)$. Häufig wird X dann nur noch durch die Angabe von P_X festgelegt und auf die Definition des zugrundeliegenden Wahrscheinlichkeitsraums $(\Omega, \mathcal{F}, P)$ verzichtet.
- Verteilungen P_X von Zufallsvariablen als Abbildungen von $\mathcal{B}$ nach $\mathbb{R}$ allerdings schlecht handhabbar.

Verteilungsfunktionen

- Man kann zeigen, dass Wahrscheinlichkeitsmaße P_X auf $\mathcal{B}$ bereits (z.B.) durch die Angabe aller Wahrscheinlichkeiten der Form $P_X((-\infty, x]) = P(\{X \leq x\})$ für $x \in \mathbb{R}$ eindeutig bestimmt sind!
- Daher überwiegend „Identifikation“ der Verteilung P_X mit Hilfe einer Abbildung $\mathbb{R} \to \mathbb{R}$ mit $x \mapsto P_X((-\infty, x])$ für $x \in \mathbb{R}$:

Definition 9.2 (Verteilungsfunktion)

Es seien X eine Zufallsvariable über dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und P_X die Wahrscheinlichkeitsverteilung von X. Dann heißt die Abbildung

$$F_X : \mathbb{R} \to \mathbb{R}; F_X(x) := P_X((-\infty, x]) = P(\{X \leq x\})$$

Verteilungsfunktion der Zufallsvariablen X.

Eigenschaften von Verteilungsfunktionen I

- Die Verteilungsfunktion F_X einer (eindimensionalen) Zufallsvariablen X hat folgende Eigenschaften:
  1. F_X ist monoton wachsend, d.h. für $x, y \in \mathbb{R}$ gilt:
  $$x < y \quad \Rightarrow \quad F_X(x) \leq F_X(y)$$
  2. F_X ist rechtsseitig stetig, d.h. für alle $x \in \mathbb{R}$ gilt:
  $$\lim_{\substack{h \to 0 \\ h > 0}} F_X(x+h) = F_X(x)$$
  3. $\lim\limits_{x \to \infty} F_X(x) =: F_X(\infty) = 1$
  4. $\lim\limits_{x \to -\infty} F_X(x) =: F_X(-\infty) = 0$
- Als abkürzende Schreibweise für die linksseitigen Grenzwerte verwendet man
$$F_X(x-0) := \lim_{\substack{h \to 0 \\ h > 0}} F_X(x-h) \qquad \text{für alle } x \in \mathbb{R}.$$

Eigenschaften von Verteilungsfunktionen II

- Analog zur empirischen Verteilungsfunktion (deskriptive Statistik, Folie 46) können Intervallwahrscheinlichkeiten leicht mit der Verteilungsfunktion F_X einer Zufallsvariablen X berechnet werden.
- Für $a, b \in \mathbb{R}$ mit $a < b$ gilt:
  - $P_X((-\infty, b]) = P(\{X \leq b\}) = F_X(b)$
  - $P_X((-\infty, b)) = P(\{X < b\}) = F_X(b - 0)$
  - $P_X([a, \infty)) = P(\{X \geq a\}) = 1 - F_X(a - 0)$
  - $P_X((a, \infty)) = P(\{X > a\}) = 1 - F_X(a)$
  - $P_X([a, b]) = P(\{a \leq X \leq b\}) = F_X(b) - F_X(a - 0)$
  - $P_X((a, b]) = P(\{a < X \leq b\}) = F_X(b) - F_X(a)$
  - $P_X([a, b)) = P(\{a \leq X < b\}) = F_X(b - 0) - F_X(a - 0)$
  - $P_X((a, b)) = P(\{a < X < b\}) = F_X(b - 0) - F_X(a)$
- Insbesondere gilt für $x \in \mathbb{R}$ auch:

$$P_X(\{x\}) = P(\{X = x\}) = F_X(x) - F_X(x - 0)$$

Diskrete Zufallsvariablen

- Einfacher, aber geläufiger Spezialfall für Zufallsvariable X über Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$: **Wertebereich**

$$X(\Omega) := \{x \in \mathbb{R} \,|\, x = X(\omega) \text{ für (mindestens) ein } \omega \in \Omega\}$$

  ist endlich oder abzählbar unendlich (bzw. etwas allgemeiner: es gibt eine endliche oder abzählbar unendliche Menge B mit $P(\{X \in B\}) = 1$).
- Analog zu Definition 6.3 heißen solche Zufallsvariablen „diskret“.

Definition 9.3 (Diskrete ZV, Wahrscheinlichkeitsfunktion, Träger)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, X eine Zufallsvariable über $(\Omega, \mathcal{F}, P)$ und $B \subseteq \mathbb{R}$ endlich oder abzählbar unendlich mit $P(\{X \in B\}) = 1$. Dann nennt man

- X eine **diskrete** Zufallsvariable,
- $p_X : \mathbb{R} \to [0, 1]; p_X(x) := P_X(\{x\})$ die **Wahrscheinlichkeitsfunktion** von X,
- $T(X) := \{x \in \mathbb{R} \,|\, p_X(x) > 0\}$ den **Träger** von X sowie alle Elemente $x \in T(X)$ **Trägerpunkte** von X und deren zugehörige Wahrscheinlichkeitsfunktionswerte $p_X(x)$ **Punktwahrscheinlichkeiten**.

Beispiel

Anzahl „Wappen“ bei dreimaligem Münzwurf

- Zufallsexperiment: Dreimaliger Münzwurf mit fairer Münze („**W**appen“ oder „**Z**ahl“)
  - $\Omega = \{WWW, WWZ, WZW, WZZ, ZWW, ZWZ, ZZW, ZZZ\}$
  - $\mathcal{F} = \mathcal{P}(\Omega)$
  - $P : \mathcal{F} \to \mathbb{R}; P(A) = \dfrac{|A|}{|\Omega|}$
- Zufallsvariable $X : \Omega \to \mathbb{R}$: Anzahl der auftretenden Wappen, also

$$X(WWW) = 3, \quad X(WWZ) = 2, \quad X(WZW) = 2, \quad X(WZZ) = 1,$$
$$X(ZWW) = 2, \quad X(ZWZ) = 1, \quad X(ZZW) = 1, \quad X(ZZZ) = 0.$$

- Für $X(\Omega) = \{0, 1, 2, 3\}$ gilt offensichtlich $P(X(\Omega)) = 1$, also X diskret.
- Konkreter ist $T(X) = \{0, 1, 2, 3\}$ und die Punktwahrscheinlichkeiten sind

$$p_X(0) = \frac{1}{8}, \qquad p_X(1) = \frac{3}{8}, \qquad p_X(2) = \frac{3}{8}, \qquad p_X(3) = \frac{1}{8}.$$

Diskrete Zufallsvariablen (Forts.) I

- $T(X)$ ist endlich oder abzählbar unendlich, die Elemente von $T(X)$ werden daher im Folgenden häufig mit x_i bezeichnet (für $i \in \{1, \ldots, n\}$ bzw. $i \in \mathbb{N}$), Summationen über Trägerpunkte mit dem Symbol $\sum_{x_i}$.
- Ist $p_X : \mathbb{R} \to \mathbb{R}$ die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen X, so gilt $P_X(A) = \sum_{x_i \in A \cap T(X)} p_X(x_i)$ für alle $A \in \mathcal{B}$.
- Spezieller gilt für die Verteilungsfunktion F_X einer diskreten Zufallsvariablen

$$F_X(x) = \sum_{\substack{x_i \in T(X) \\ x_i \leq x}} p_X(x_i) \qquad \text{für alle } x \in \mathbb{R}.$$

Verteilungsfunktionen diskreter Zufallsvariablen sind damit (vergleichbar mit empirischen Verteilungsfunktionen) Treppenfunktionen mit **Sprunghöhen** $p_X(x_i)$ an den **Sprungstellen** (=Trägerpunkten) $x_i \in T(X)$.

Diskrete Zufallsvariablen (Forts.) II

- *Im Münzwurf-Beispiel:*

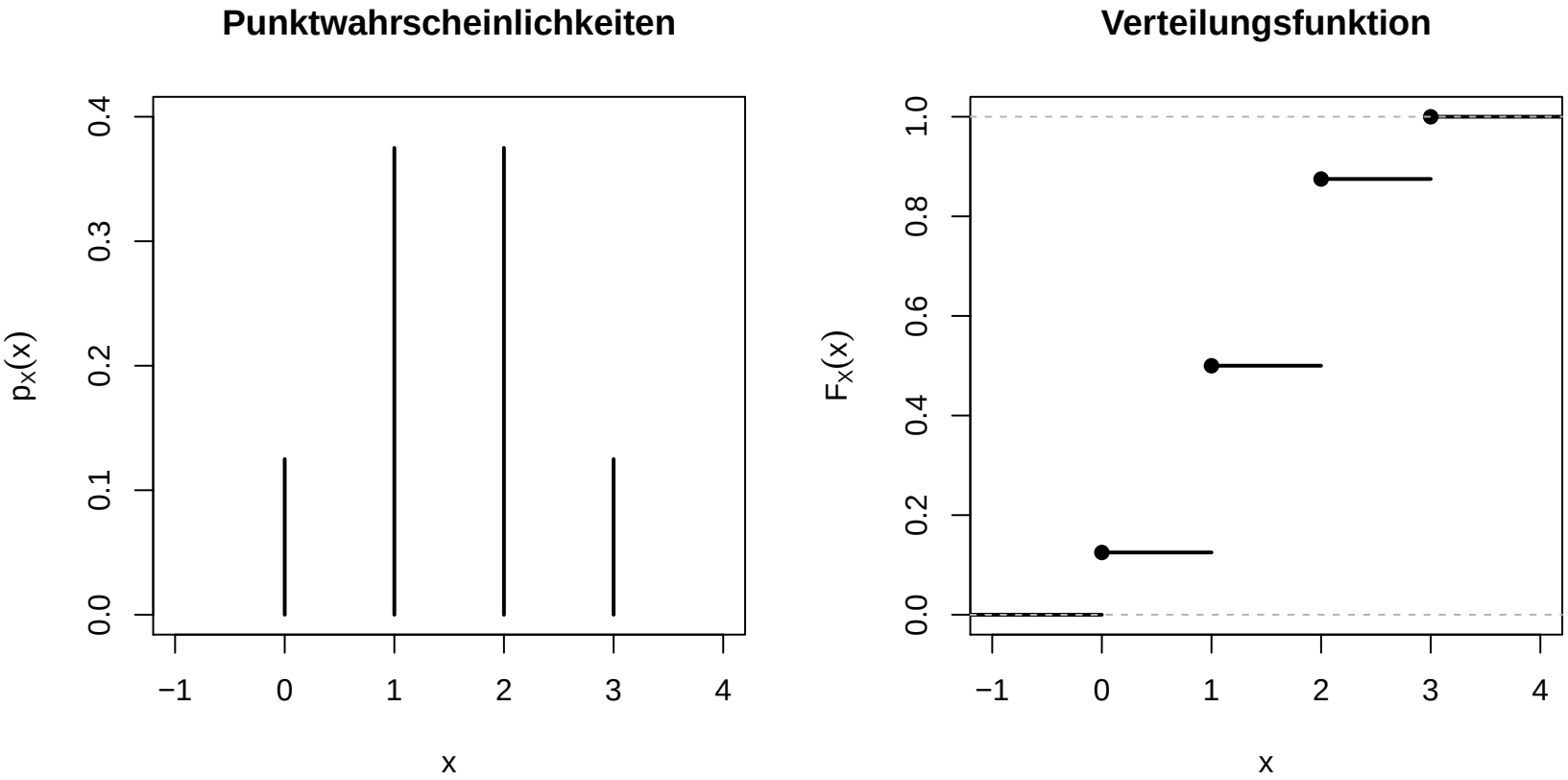
$$F_X(x) = \begin{cases} 0 & \text{für} \quad x < 0 \\ \frac{1}{8} & \text{für} \quad 0 \leq x < 1 \\ \frac{1}{2} & \text{für} \quad 1 \leq x < 2 \\ \frac{7}{8} & \text{für} \quad 2 \leq x < 3 \\ 1 & \text{für} \quad x \geq 3 \end{cases}$$

- Ist der Träger $T(X)$ endlich und die Anzahl der Elemente in $T(X)$ klein, so werden die Punktwahrscheinlichkeiten häufig in Tabellenform angegeben.
- *Im Münzwurf-Beispiel:*

x_i	0	1	2	3
$p_X(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Diskrete Zufallsvariablen (Forts.) III

- *Grafische Darstellung im Münzwurf-Beispiel:*



Stetige Zufallsvariablen I

- Weiterer wichtiger Spezialfall: **stetige** Zufallsvariablen
- Wertebereich stetiger Zufallsvariablen ist immer ein Kontinuum: es gibt **keine** endliche oder abzählbar unendliche Menge $B \subseteq \mathbb{R}$ mit $P_X(B) = 1$, stattdessen gilt sogar $P_X(B) = 0$ für alle endlichen oder abzählbar unendlichen Teilmengen $B \subseteq \mathbb{R}$.
- Verteilungsfunktionen stetiger Zufallsvariablen sind nicht (wie bei diskreten Zufallsvariablen) als Summe von Wahrscheinlichkeitsfunktionswerten darstellbar, sondern als Integral über eine sogenannte *Dichtefunktion*:

Stetige Zufallsvariablen II

Definition 9.4 (Stetige Zufallsvariable, Dichtefunktion)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable über $(\Omega, \mathcal{F}, P)$. Gibt es eine nichtnegative Abbildung $f_X : \mathbb{R} \to \mathbb{R}$ mit

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad \text{für alle } x \in \mathbb{R}, \tag{2}$$

so heißt die Zufallsvariable X **stetig**. Jede nichtnegative Abbildung f_X mit der Eigenschaft (2) heißt **Dichtefunktion** von X.

Stetige Zufallsvariablen III

- Aus Definition 9.4 lassen sich weitere Eigenschaften von stetigen Zufallsvariablen bzw. Dichtefunktionen ableiten, zum Beispiel:
  - $P_X(\{x\}) = 0$ für alle $x \in \mathbb{R}$.
  - $F_X(x-0) = F_X(x)$ für alle $x \in \mathbb{R}$, F_X ist also stetig auf $\mathbb{R}$.
  - $P(\{a \leq X \leq b\}) = P(\{a < X \leq b\}) = P(\{a \leq X < b\}) = P(\{a < X < b\})$ $= F_X(b) - F_X(a) = \int_a^b f_X(t)dt$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$.
  - (Mindestens) in allen Stetigkeitsstellen $x \in \mathbb{R}$ von f_X ist F_X differenzierbar und es gilt $F_X'(x) = f_X(x)$.
- Wahrscheinlichkeit von Intervallen stimmt mit Fläche zwischen Intervall und Dichtefunktion (analog zu Histogrammen bei deskriptiver Statistik mit klassierten Daten) überein.

Stetige Zufallsvariablen IV

- Dichtefunktion f_X zu einer Verteilungsfunktion F_X ist nicht eindeutig bestimmt; Abänderungen von f_X an endlich oder abzählbar unendlich vielen Stellen sind beliebig möglich!
- **Aber:** Ist $f_X : \mathbb{R} \to \mathbb{R}$ nichtnegative uneigentlich integrierbare Abbildung mit $\int_{-\infty}^{+\infty} f_X(x)dx = 1$, so gibt es genau eine Verteilungsfunktion $F_X : \mathbb{R} \to \mathbb{R}$, zu der f_X Dichtefunktion ist.
- Neben diskreten und stetigen Zufallsvariablen sind auch Mischformen (mit einem diskreten und stetigen Anteil) möglich, auf die hier aber nicht näher eingegangen wird!

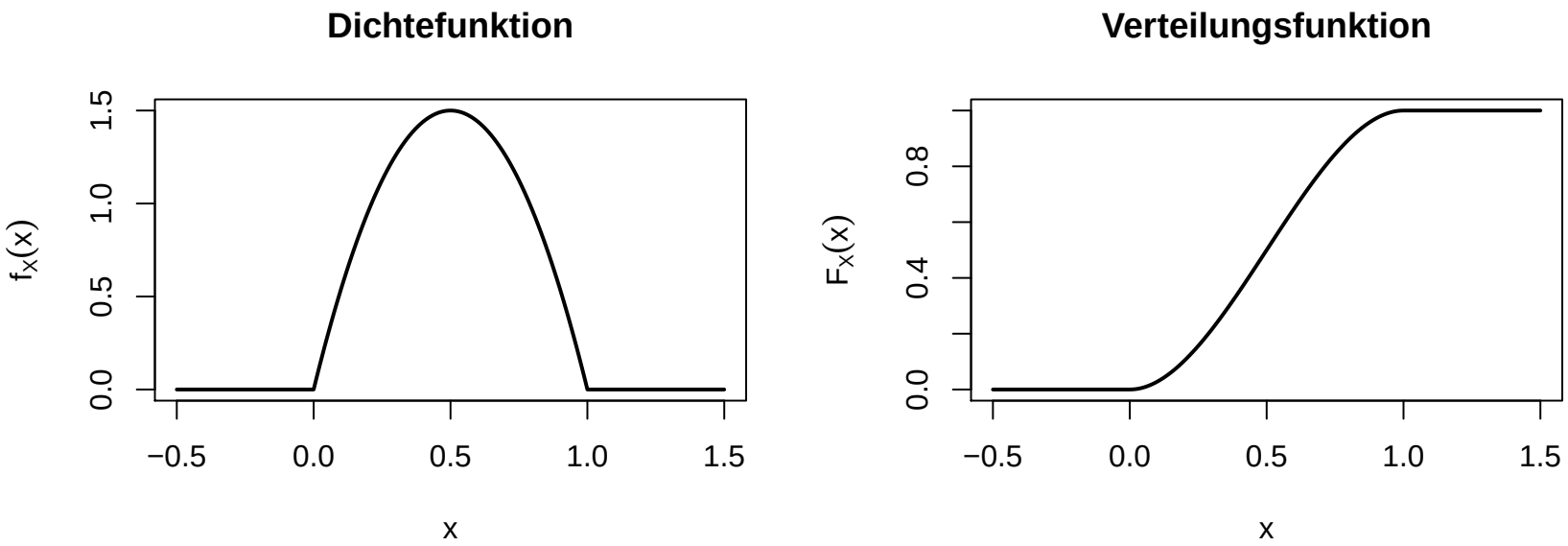
- **Beispiel:** Die Abbildung

$$f_X : \mathbb{R} \to \mathbb{R}; f_X(x) := \begin{cases} 6(x - x^2) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

ist nichtnegativ und uneigentlich integrierbar mit $\int_{-\infty}^{+\infty} f_X(x)dx = 1$, also eine Dichtefunktion.

- Die Verteilungsfunktion F_X zu f_X erhält man als

$$F_X : \mathbb{R} \to \mathbb{R}; F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{für } x < 0 \\ 3x^2 - 2x^3 & \text{für } 0 \leq x \leq 1 \\ 1 & \text{für } x > 1 \end{cases} .$$



(Lineare) Abbildungen von Zufallsvariablen I

- Genauso, wie man mit Hilfe einer $\mathcal{F}-\mathcal{B}-$messbaren Abbildung $X : \Omega \to \mathbb{R}$ als Zufallsvariable über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ einen neuen Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}, P_X)$ erhält, kann aus diesem mit einer „nachgeschalteten" $\mathcal{B}-\mathcal{B}-$messbaren Abbildung $G : \mathbb{R} \to \mathbb{R}$ ein weiterer Wahrscheinlichkeitsraum gewonnen werden!
- Mehrere „Auffassungen" möglich:
  1. G als Zufallsvariable über $(\mathbb{R}, \mathcal{B}, P_X)$ durch
  $$(P_X)_G(B) = P_X(G^{-1}(B)) \text{ für alle } B \in \mathcal{B}.$$
  2. $G(X) := G \circ X$ als Zufallsvariable über $(\Omega, \mathcal{F}, P)$ durch
  $$P_{G(X)}(B) = P((G \circ X)^{-1}(B)) = P(X^{-1}(G^{-1}(B))) \text{ für alle } B \in \mathcal{B}.$$
- Man erkennt leicht, dass beide Auffassungen miteinander vereinbar sind (es gilt $(P_X)_G = P_{G(X)}$), daher Schreibweise $G(X)$ auch geläufig, wenn $(\Omega, \mathcal{F}, P)$ nicht im Vordergrund steht.

(Lineare) Abbildungen von Zufallsvariablen II

- Im Folgenden: Betrachtung besonders einfacher (linearer) Abbildungen G, also Abbildungen der Form $G(x) = a \cdot x + b$ für $a, b \in \mathbb{R}$ mit $a \neq 0$.
- Man kann zeigen, dass G (als stetige Funktion) stets $\mathcal{B} - \mathcal{B}-$messbar ist.
- **Problemstellung:** Wie kann die Verteilung von $Y := G(X)$ (möglichst leicht!) aus der Verteilung von X gewonnen werden?
- **Idee:** Abbildung G ist insbesondere bijektiv, es existiert also die **Umkehrfunktion** $G^{-1}(y) = \dfrac{y-b}{a}$.
- Für diskrete Zufallsvariablen X mit Trägerpunkten x_i und zugehöriger Wahrscheinlichkeitsfunktion p_X ist offensichtlich auch Y diskret mit
  - Trägerpunkten $y_i = a \cdot x_i + b$ und
  - Wahrscheinlichkeitsfunktion $p_Y(y) = P_Y(\{y\}) = P_X(\{\frac{y-b}{a}\}) = p_X(\frac{y-b}{a})$.

  (Es gilt also $p_i = p_X(x_i) = p_Y(a \cdot x_i + b) = p_Y(y_i)$)
- Ähnlich lassen sich zu $Y = G(X)$ auch Dichtefunktionen f_Y (bei stetigen Zufallsvariablen X) aus f_X sowie (allgemein) Verteilungsfunktionen F_Y aus F_X bestimmen:

Satz 9.5

Es seien X eine (eindimensionale) Zufallsvariable, $a, b \in \mathbb{R}$ mit $a \neq 0$ und $G : \mathbb{R} \to \mathbb{R}$ die Abbildung mit $G(x) = ax + b$ für alle $x \in \mathbb{R}$. Dann ist $Y := G(X) = aX + b$ ebenfalls eine Zufallsvariable und es gilt

$$F_Y(y) = \left\{ \begin{array}{cc} F_X(\frac{y-b}{a}) & \text{für } a > 0 \\ 1 - F_X(\frac{y-b}{a} - 0) & \text{für } a < 0 \end{array} \right\} \quad \text{für alle } y \in \mathbb{R}.$$

- *Ist X diskret und p_X die Wahrscheinlichkeitsfunktion von X, so ist auch Y diskret und die Wahrscheinlichkeitsfunktion p_Y von Y gegeben durch:*

$$p_Y : \mathbb{R} \to \mathbb{R}; p_Y(y) = p_X\left(\frac{y-b}{a}\right)$$

- *Ist X stetig und f_X eine Dichtefunktion von X, so ist auch Y stetig und*

$$f_Y : \mathbb{R} \to \mathbb{R}; f_Y(y) := \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$$

eine Dichtefunktion zu Y.

Erwartungswert von Zufallsvariablen I

- Analog zur Lage- und Streuungsmaßen in deskriptiver Statistik:
  Oft Verdichtung der Information aus Verteilungen von Zufallsvariablen auf eine oder wenige Kennzahlen erforderlich.
- Kennzahl für **Lage** der Verteilung: **Erwartungswert**
- Zur allgemeinen Definition „zuviel“ Maß- und Integrationstheorie erforderlich, daher Beschränkung auf diskrete und stetige Zufallsvariablen (separat).
- In deskriptiver Statistik: Arithmetischer Mittelwert eines Merkmals als (mit den relativen Häufigkeiten) gewichtetes Mittel der aufgetretenen Merkmalswerte.
- Bei diskreten Zufallsvariablen analog: Erwartungswert als (mit den Punktwahrscheinlichkeiten) gewichtetes Mittel der Trägerpunkte.

Erwartungswert von Zufallsvariablen II

Definition 9.6 (Erwartungswerte diskreter Zufallsvariablen)

Es sei X eine diskrete Zufallsvariable mit Trägerpunkten x_i und Wahrscheinlichkeitsfunktion p_X. Gilt

$$\sum_{x_i} |x_i| \cdot p_X(x_i) < \infty, \tag{3}$$

so heißt

$$\mu_X := \mathrm{E}\, X := \mathrm{E}(X) := \sum_{x_i} x_i \cdot p_X(x_i)$$

der **Erwartungswert (Mittelwert)** der Zufallsvariablen X.
Gilt (3) nicht, so sagt man, der Erwartungswert von X existiere nicht.

Erwartungswert von Zufallsvariablen III

- Anders als in deskriptiver Statistik: Erwartungswert einer Zufallsvariablen existiert möglicherweise nicht!
- Existenzbedingungen sind zwar (auch in anderen Definitionen) angeführt, bei den hier betrachteten Zufallsvariablen sind diese aber stets erfüllt.
- Ist der Träger $T(X)$ einer diskreten Zufallsvariablen X endlich, so existiert $\mathrm{E}(X)$ stets (Summe endlich).
- Gilt spezieller $T(X) = \{a\}$ für ein $a \in \mathbb{R}$, d.h. gilt $p_X(a) = 1$ und $p_X(x) = 0$ für alle $x \in \mathbb{R}$ mit $x \neq a$, dann nennt man die Verteilung von X eine **Einpunktverteilung** und es gilt offensichtlich $\mathrm{E}(X) = a$.
- Für stetige Zufallsvariablen ist die Summation durch ein Integral, die Wahrscheinlichkeitsfunktion durch eine Dichtefunktion und die Trägerpunkte durch die Integrationsvariable zu ersetzen:

Erwartungswert von Zufallsvariablen IV

Definition 9.7 (Erwartungswerte stetiger Zufallsvariablen)

Es sei X eine stetige Zufallsvariable und f_X eine Dichtefunktion von X. Gilt

$$\int_{-\infty}^{+\infty} |x| \cdot f_X(x)dx < \infty, \tag{4}$$

so heißt

$$\mu_X := \mathrm{E}\,X := \mathrm{E}(X) := \int_{-\infty}^{+\infty} x \cdot f_X(x)dx$$

der **Erwartungswert (Mittelwert)** der Zufallsvariablen X.
Gilt (4) nicht, so sagt man, der Erwartungswert von X existiere nicht.

Symmetrie

Definition 9.8 (Symmetrische Zufallsvariablen)

Eine Zufallsvariable X über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ heißt **symmetrisch** um $a \in \mathbb{R}$, falls die Verteilungen von $X - a$ und $a - X$ übereinstimmen.

- Alternative (äquivalente) Bedingungen
  - für beliebige Zufallsvariablen X:

    $$P_X(\{X \leq a + x\}) = P_X(\{X \geq a - x\}) \quad \text{bzw.} \quad F_X(a + x) = 1 - F_X(a - x - 0)$$

    für alle $x \in \mathbb{R}$.
  - für diskrete Zufallsvariablen X:
    Für alle $x_i \in T(X)$ gilt $2a - x_i \in T(X)$ und spezieller $p_X(x_i) = p_X(2a - x_i)$.
  - für stetige Zufallsvariablen X:
    Es existiert eine Dichte f_X von X mit

    $$f_X(a + x) = f_X(a - x) \quad \text{bzw.} \quad f_X(x) = f_X(2a - x)$$

    für alle $x \in \mathbb{R}$.
- Existiert der Erwartungswert $\mathrm{E}(X)$ von X und ist X symmetrisch um $a \in \mathbb{R}$, dann gilt stets $a = \mathrm{E}(X)$.

Erwartungswert von $G(X)$

- Zur Einführung weiterer Kennzahlen für Verteilung einer Zufallsvariablen X nötig: Erwartungswerte von Transformationen $G(X)$ für verschiedene (nicht nur lineare) $\mathcal{B}-\mathcal{B}-$messbare Abbildungen $G : \mathbb{R} \to \mathbb{R}$.

Definition 9.9

Es seien X eine Zufallsvariable und $G : \mathbb{R} \to \mathbb{R}$ eine $\mathcal{B}-\mathcal{B}-$messbare Abbildung.

- Ist X diskrete Zufallsvariable, x_i die Trägerpunkte sowie p_X die Wahrscheinlichkeitsfunktion von X und gilt $\sum_{x_i} |G(x_i)| \cdot p_X(x_i) < \infty$, dann existiert der Erwartungswert $\mathrm{E}(G(X))$ und es gilt

$$\mathrm{E}(G(X)) = \sum_{x_i} G(x_i) \cdot p_X(x_i).$$

- Ist X stetige Zufallsvariable mit Dichtefunktion f_X und gilt $\int_{-\infty}^{+\infty} |G(x)| \cdot f_X(x)dx < \infty$, dann existiert der Erwartungswert $\mathrm{E}(G(X))$ und es gilt

$$\mathrm{E}(G(X)) = \int_{-\infty}^{+\infty} G(x) \cdot f_X(x)dx.$$

Höhere Momente von Zufallsvariablen

Definition 9.10 (k-te Momente, Varianz, Standardabweichung)

Es seien X eine (eindimensionale) Zufallsvariable, $k \in \mathbb{N}$.

- Man bezeichnet den Erwartungswert $\mathrm{E}(X^k)$ (falls er existiert) als das **Moment k-ter Ordnung (um Null)** von X.
- Existiert $\mathrm{E}(X)$, so bezeichnet man den Erwartungswert $\mathrm{E}[(X - \mathrm{E}(X))^k]$ (falls er existiert) als das **zentrale Moment k-ter Ordnung** von X.
- Das zweite zentrale Moment heißt auch **Varianz** von X, und man schreibt $\sigma_X^2 := \mathrm{Var}(X) := \mathrm{E}[(X - \mathrm{E}(X))^2] = \mathrm{E}[(X - \mu_X)^2]$.
- Die positive Wurzel der Varianz von X heißt auch **Standardabweichung** von X, und man schreibt $\sigma_X := \mathrm{Sd}(X) := +\sqrt{\sigma_X^2}$.

Rechenregeln für Erwartungswerte und Varianz I

Satz 9.11

Es seien X eine (eindimensionale) Zufallsvariable, $a, b \in \mathbb{R}$.

- *Existiert $\mathrm{E}(X)$, so existiert auch der Erwartungswert von $aX + b$ und es gilt:*

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$$

- *Existiert $\mathrm{Var}(X)$, so existiert auch die Varianz von $aX + b$ und es gilt:*

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

Für die Standardabweichung gilt dann $\mathrm{Sd}(aX + b) = |a|\,\mathrm{Sd}(X)$.

Rechenregeln für Erwartungswerte und Varianz II

Satz 9.12 (Varianzzerlegungssatz)

Es sei X eine (eindimensionale) Zufallsvariable. Dann existiert die Varianz $\mathrm{Var}(X)$ *genau dann, wenn* $\mathrm{E}(X^2)$ *(und* $\mathrm{E}(X)$*) existiert, und in diesem Fall gilt:*

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

- Die Eigenschaft aus Satz 9.11, den „Erwartungswertoperator“ $\mathrm{E}(\,\cdot\,)$ mit linearen Abbildungen vertauschen zu dürfen, lässt sich zum Beispiel wie folgt verallgemeinern:

  Es seien X eine (eindimensionale) Zufallsvariable, $a, b \in \mathbb{R}$ sowie $G : \mathbb{R} \to \mathbb{R}$ und $H : \mathbb{R} \to \mathbb{R}$ zwei $\mathcal{B} - \mathcal{B}-$messbare Abbildungen.
  Existieren $\mathrm{E}(G(X))$ *und* $\mathrm{E}(H(X))$*, dann gilt:*

$$\mathrm{E}(aG(X) + bH(X)) = a\,\mathrm{E}(G(X)) + b\,\mathrm{E}(H(X))$$

Rechenregeln für Erwartungswerte und Varianz III

- Mit Satz 9.11 folgt direkt:

  Ist X eine Zufallsvariable mit existierendem Erwartungswert $\mathrm{E}(X)$ und existierender Varianz $\mathrm{Var}(X)$, so erhält man mit

$$Y := \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} = \frac{X - \mathrm{E}(X)}{\mathrm{Sd}(X)} = \frac{X - \mu_X}{\sigma_X}$$

  eine neue Zufallsvariable mit $\mathrm{E}(Y) = 0$ und $\mathrm{Var}(Y) = \mathrm{Sd}(Y) = 1$.
  Man nennt Y dann eine **standardisierte Zufallsvariable**.

Schiefe und Wölbung von Zufallsvariablen I

Definition 9.13 (Schiefe (Skewness), Wölbung (Kurtosis))

Sei X eine (eindimensionale) Zufallsvariable mit existierender Varianz $\sigma_X^2 > 0$.

1. Existiert das zentrale Moment 3. Ordnung, so nennt man

$$\gamma_X := \gamma(X) := \frac{\mathrm{E}[(X - \mathrm{E}(X))^3]}{\sigma_X^3} = \mathrm{E}\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right]$$

   die **Schiefe (Skewness)** von X.

2. Existiert das zentrale Moment 4. Ordnung, so nennt man

$$\kappa_X := \kappa(X) := \frac{\mathrm{E}[(X - \mathrm{E}(X))^4]}{\sigma_X^4} = \mathrm{E}\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^4\right]$$

   die **Wölbung (Kurtosis)** von X.
   Die um 3 verminderte Kurtosis $\kappa_X - 3$ wird auch **Exzess-Kurtosis** genannt.

Schiefe und Wölbung von Zufallsvariablen II

- Ist X symmetrisch, so ist die Schiefe von X (falls sie existiert) gleich Null. Die Umkehrung der Aussage gilt **nicht**.
- Existieren Schiefe bzw. Kurtosis einer Zufallsvariablen X, so existieren auch die Schiefe bzw. Kurtosis von $Y := aX + b$ für $a, b \in \mathbb{R}$ mit $a \neq 0$ und es gilt:
  - $\gamma_X = \gamma_Y$ sowie $\kappa_X = \kappa_Y$, falls $a > 0$,
  - $\gamma_X = -\gamma_Y$ sowie $\kappa_X = \kappa_Y$, falls $a < 0$.
- In Abhängigkeit von γ_X heißt die Verteilung von X
  - **linkssteil** oder **rechtsschief**, falls $\gamma_X > 0$ gilt und
  - **rechtssteil** oder **linksschief**, falls $\gamma_X < 0$ gilt.

Schiefe und Wölbung von Zufallsvariablen III

- In Abhängigkeit von κ_X heißt die Verteilung von X
  - **platykurtisch** oder **flachgipflig**, falls $\kappa_X < 3$ gilt,
  - **mesokurtisch** oder **normalgipflig**, falls $\kappa_X = 3$ gilt und
  - **leptokurtisch** oder **steilgipflig**, falls $\kappa_X > 3$ gilt.
- γ_X und κ_X können (auch wenn sie existieren) nicht beliebige Werte annehmen. Es gilt stets $\kappa_X \geq 1$ und $\gamma_X^2 \leq \kappa_X - 1$.
- Man kann (zur einfacheren Berechnung von γ_X und κ_X) leicht zeigen:
  - $$\begin{aligned} \mathrm{E}[(X - \mathrm{E}(X))^3] &= \mathrm{E}(X^3) - 3\,\mathrm{E}(X^2)\,\mathrm{E}(X) + 2[\mathrm{E}(X)]^3 \\ &= \mathrm{E}(X^3) - 3\mu_X\sigma_X^2 - \mu_X^3 \end{aligned}$$
  - $$\begin{aligned} \mathrm{E}[(X - \mathrm{E}(X))^4] &= \mathrm{E}(X^4) - 4\,\mathrm{E}(X^3)\,\mathrm{E}(X) + 6\,\mathrm{E}(X^2)[\mathrm{E}(X)]^2 - 3[\mathrm{E}(X)]^4 \\ &= \mathrm{E}(X^4) - 4\,\mathrm{E}(X^3)\mu_X + 6\mu_X^2\sigma_X^2 + 3\mu_X^4 \end{aligned}$$

Beispiel

zu stetiger Zufallsvariable X aus Folie 208 mit Dichte $f_X(x) = \begin{cases} 6(x-x^2) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$.

- $\mathrm{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx = \int_0^1 (6x^2 - 6x^3)dx = \left[2x^3 - \frac{3}{2}x^4\right]_0^1 = \frac{1}{2}$
- $\mathrm{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x)dx = \int_0^1 (6x^3 - 6x^4)dx = \left[\frac{3}{2}x^4 - \frac{6}{5}x^5\right]_0^1 = \frac{3}{10}$
  $\Rightarrow \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20}$
- $\mathrm{E}(X^3) = \int_{-\infty}^{+\infty} x^3 \cdot f_X(x)dx = \int_0^1 (6x^4 - 6x^5)dx = \left[\frac{6}{5}x^5 - x^6\right]_0^1 = \frac{1}{5}$
  $\Rightarrow \gamma(X) = \frac{\frac{1}{5} - 3 \cdot \frac{3}{10} \cdot \frac{1}{2} + 2 \cdot \left(\frac{1}{2}\right)^3}{(1/20)^{3/2}} = \frac{\frac{4-9+5}{20}}{(1/20)^{3/2}} = 0$
- $\mathrm{E}(X^4) = \int_{-\infty}^{+\infty} x^4 \cdot f_X(x)dx = \int_0^1 (6x^5 - 6x^6)dx = \left[x^6 - \frac{6}{7}x^7\right]_0^1 = \frac{1}{7}$
  $\Rightarrow \kappa(X) = \frac{\frac{1}{7} - 4 \cdot \frac{1}{5} \cdot \frac{1}{2} + 6 \cdot \frac{3}{10} \cdot \left(\frac{1}{2}\right)^2 - 3 \cdot \left(\frac{1}{2}\right)^4}{(1/20)^2} = \frac{3/560}{1/400} = \frac{15}{7}$

Quantile von Zufallsvariablen I

Definition 9.14 (p-Quantil)

Seien X eine eindimensionale Zufallsvariable, $p \in (0,1)$. Jeder Wert $x_p \in \mathbb{R}$ mit

$$P\{X \leq x_p\} \geq p \qquad \text{und} \qquad P\{X \geq x_p\} \geq 1 - p$$

heißt p-**Quantil** (auch p-**Perzentil**) von X. Man nennt Werte x_p mit dieser Eigenschaft spezieller

- **Median** von X für $p = 0.5$,
- **unteres Quartil** von X für $p = 0.25$ sowie
- **oberes Quartil** von X für $p = 0.75$.

Ist F_X die Verteilungsfunktion von X, so ist x_p also genau dann p-Quantil von X, wenn

$$F_X(x_p - 0) \leq p \leq F_X(x_p)$$

gilt, für stetige Zufallsvariablen X also genau dann, wenn $F_X(x_p) = p$ gilt.

Quantile von Zufallsvariablen II

- p-Quantile sind nach Definition 9.14 eindeutig bestimmt, wenn die Verteilungsfunktion F_X der Zufallsvariablen X (dort, wo sie Werte in $(0,1)$ annimmt) invertierbar ist, also insbesondere stetig und streng monoton wachsend.
  Bezeichnet F_X^{-1} die Umkehrfunktion von F_X, so gilt dann

  $$x_p = F_X^{-1}(p) \qquad \text{für alle } p \in (0,1) \ .$$

  F_X^{-1} wird in diesem Fall auch **Quantilsfunktion** genannt.
- Der Abstand $x_{0.75} - x_{0.25}$ zwischen unterem und oberen Quartil wird (wie auch bei empirischen Quartilen) auch **Interquartilsabstand (IQA)** genannt.

Quantile von Zufallsvariablen III

- Auch ohne die Invertierbarkeit von F_X kann Eindeutigkeit von Quantilen zum Beispiel durch die Festsetzung

$$x_p := \min\{x \mid P\{X \leq x\} \geq p\} = \min\{x \mid F_X(x) \geq p\}$$

  erreicht werden.
  Man nennt die Abbildung

$$(0,1) \to \mathbb{R}; p \mapsto x_p = \min\{x \mid F_X(x) \geq p\}$$

  häufig auch *verallgemeinerte Inverse* von F_X und verwendet hierfür dann ebenfalls das Symbol F_X^{-1} sowie die Bezeichnung Quantilsfunktion.
- Diese Eindeutigkeitsfestlegung **unterscheidet** sich von der vergleichbaren Konvention aus der deskriptiven Statistik für empirische Quantile!

Spezielle diskrete Verteilungen

- Im Folgenden: Vorstellung spezieller (parametrischer) **Verteilungsfamilien**, die häufig Verwendung finden.
- Häufige Verwendung ist dadurch begründet, dass diese Verteilungen in vielen verschiedenen Anwendungen anzutreffen sind bzw. die Zufallsabhängigkeit interessanter Größen geeignet modellieren.
- Parametrische Verteilungsfamilien sind Mengen von (ähnlichen) Verteilungen Q_θ, deren Elemente sich nur durch die Ausprägung eines oder mehrerer **Verteilungsparameter** unterscheiden, d.h. die spezielle Verteilung hängt von einem Parameter oder einem Parametervektor θ ab, und zu jedem Parameter(vektor) gehört jeweils eine eigene Verteilung Q_θ.
- Die Menge aller möglichen Parameter(vektoren) θ, auch **Parameterraum** genannt, wird meist mit Θ bezeichnet. Die Verteilungsfamilie ist damit die Menge $\{Q_\theta \mid \theta \in \Theta\}$.
- Besitzt eine Zufallsvariable X die Verteilung Q_θ, so schreibt man auch kurz: $X \sim Q_\theta$.
- *Zunächst:* Vorstellung spezieller *diskreter* Verteilungen.

Bernoulli-/Alternativverteilung

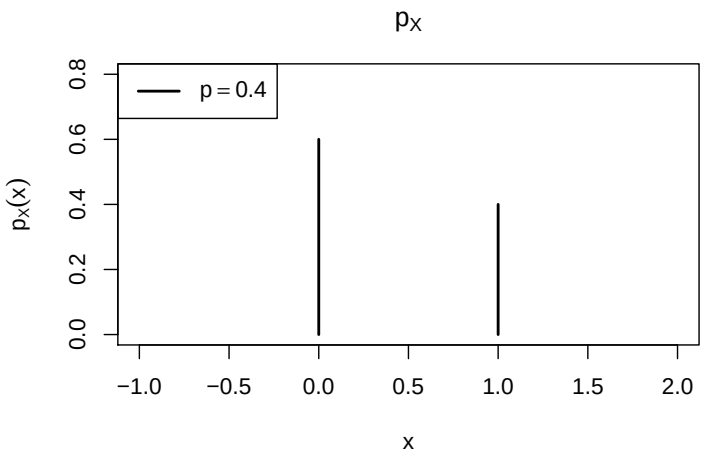
- Verwendung:
  - Modellierung eines Zufallsexperiments $(\Omega, \mathcal{F}, P)$, in dem nur das Eintreten bzw. Nichteintreten eines einzigen Ereignisses A von Interesse ist.
  - Eintreten des Ereignisses A wird oft als „Erfolg“ interpretiert, Nichteintreten (bzw. Eintreten von $\overline{A}$) als „Misserfolg“.
  - Zufallsvariable soll im Erfolgsfall Wert 1 annehmen, im Misserfolgsfall Wert 0, es sei also
$$X(\omega) := \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \in \overline{A} \end{cases}$$
  - Beispiel: Werfen eines fairen Würfels, Ereignis A: „6 gewürfelt“ mit $P(A) = \frac{1}{6}$.
- Verteilung von X hängt damit *nur* von „Erfolgswahrscheinlichkeit“ $p := P(A)$ ab; p ist also einziger Parameter der Verteilungsfamilie.
- Um triviale Fälle auszuschließen, betrachtet man nur Ereignisse mit $p \in (0, 1)$
- Der Träger der Verteilung ist dann $T(X) = \{0, 1\}$, die Punktwahrscheinlichkeiten sind $p_X(0) = 1 - p$ und $p_X(1) = p$.
- Symbolschreibweise für Bernoulli-Verteilung mit Parameter p: $B(1, p)$
- Ist X also Bernoulli-verteilt mit Parameter p, so schreibt man $X \sim B(1, p)$.

Bernoulli-/Alternativverteilung
$B(1,p)$

Parameter:
$p \in (0,1)$

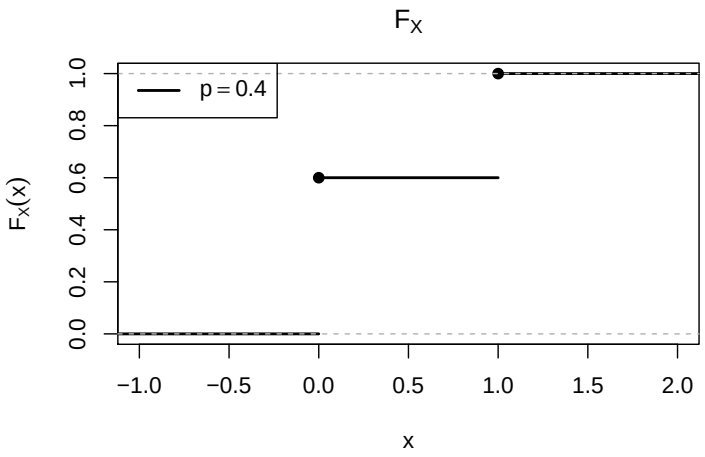
Träger: $T(X) = \{0,1\}$
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} 1-p & \text{für } x = 0 \\ p & \text{für } x = 1 \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1-p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$$



Momente:

$$E(X) = p \qquad \mathrm{Var}(X) = p \cdot (1-p)$$

$$\gamma(X) = \frac{1-2p}{\sqrt{p(1-p)}} \qquad \kappa(X) = \frac{1-3p(1-p)}{p(1-p)}$$

Binomialverteilung

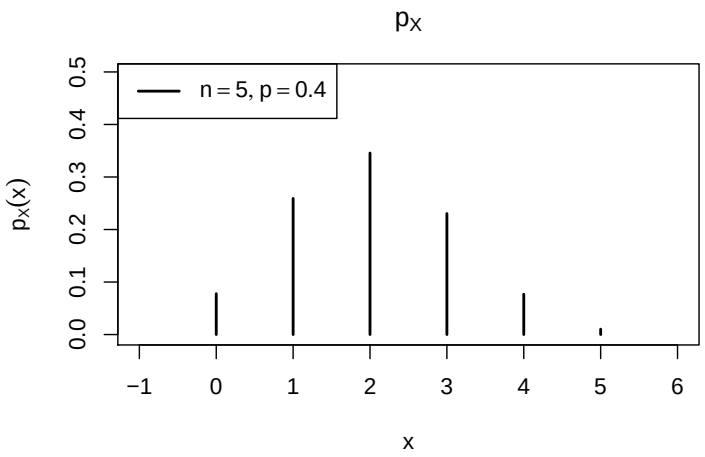
- Verallgemeinerung der Bernoulli-Verteilung
- Verwendung:
  - Modellierung der **unabhängigen, wiederholten** Durchführung eines Zufallsexperiments, in dem nur die **Häufigkeit** des Eintretens bzw. Nichteintretens eines Ereignisses A interessiert („Bernoulli-Experiment“).
  - Eintreten des Ereignisses A wird auch hier oft als „Erfolg“ interpretiert, Nichteintreten (bzw. Eintreten von $\overline{A}$) als „Misserfolg“.
  - Zufallsvariable X soll die **Anzahl der Erfolge** bei einer vorgegebenen Anzahl von n Wiederholungen des Experiments zählen.
  - Nimmt X_i für $i \in \{1, \ldots, n\}$ im Erfolgsfall (für Durchführung i) den Wert 1 an, im Misserfolgsfall den Wert 0, dann gilt also $X = \sum_{i=1}^{n} X_i$.
  - Beispiel: 5-faches Werfen eines fairen Würfels, Anzahl der Zahlen kleiner 3. $\rightsquigarrow n = 5, p = 1/3$.
- Verteilung von X hängt damit *nur* von „Erfolgswahrscheinlichkeit“ $p := P(A)$ sowie der Anzahl der Durchführungen n des Experiments ab.
- Um triviale Fälle auszuschließen, betrachtet man nur die Fälle $n \in \mathbb{N}$ und $p \in (0, 1)$. Träger der Verteilung ist dann $T(X) = \{0, 1, \ldots, n\}$.
- Symbolschreibweise für Binomialverteilung mit Parameter n und p: $B(n, p)$
- Übereinstimmung mit Bernoulli-Verteilung (mit Parameter p) für $n = 1$.

Binomialverteilung
$B(n,p)$

Parameter:
$n \in \mathbb{N}, p \in (0,1)$

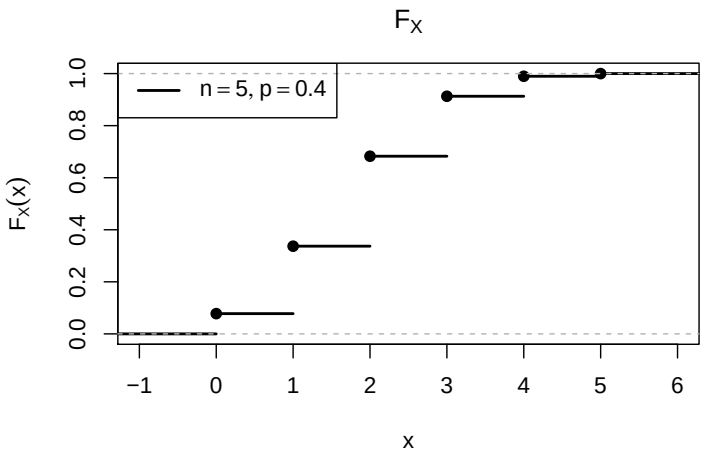
Träger: $T(X) = \{0, 1, \ldots, n\}$
Wahrscheinlichkeitsfunktion: $p_X(x)$

$$= \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \sum_{\substack{x_i \in T(X) \\ x_i \leq x}} p_X(x_i)$$



Momente: $E(X) = n \cdot p$ $\quad$ $\mathrm{Var}(X) = n \cdot p \cdot (1-p)$

$\gamma(X) = \frac{1-2p}{\sqrt{np(1-p)}}$ $\quad$ $\kappa(X) = \frac{1+(3n-6)p(1-p)}{np(1-p)}$

Geometrische Verteilung

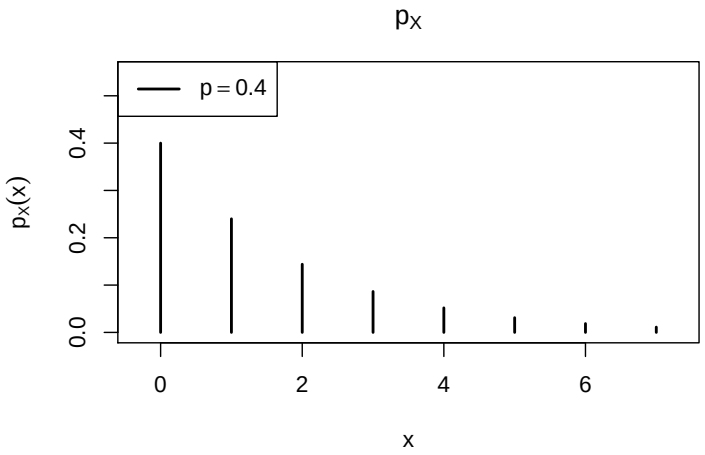
- Verwendung:
  - Modellierung der **unabhängigen, wiederholten** Durchführung eines Bernoulli-Experiments (nur das Eintreten bzw. Nichteintreten eines einzigen Ereignisses A ist von Interesse), bis das Ereignis A **zum ersten Mal** eintritt.
  - Zufallsvariable X zählt **Anzahl der Misserfolge**, ausschließlich des (letzten) „erfolgreichen" Versuchs, bei dem Ereignis A zum ersten Mal eintritt.
  - X kann also nur Werte $x \in \mathbb{N}_0$ annehmen, man erhält die Realisation x von X, wenn nach genau x Misserfolgen (Nicht-Eintreten von A) in der $(x+1)$-ten Durchführung ein Erfolg (Eintreten von A) zu verzeichnen ist.
  - Ist $p := P(A)$ die „Erfolgswahrscheinlichkeit" des Bernoulli-Experiments, so gilt offensichtlich $P\{X = x\} = (1-p)^x \cdot p$ für alle $x \in \mathbb{N}_0$.
  - Beispiel (vgl. Folie 168): Anzahl des Auftretens von „Zahl" beim Werfen einer Münze („Wappen" oder „Zahl"), bis zum ersten Mal „Wappen" erscheint $\rightsquigarrow p = 1/2$ (bei fairer Münze).
- Verteilung von X hängt damit *nur* von Erfolgswahrscheinlichkeit p ab.
- Um triviale Fälle auszuschließen, betrachtet man nur den Fall $p \in (0,1)$. Träger der Verteilung ist dann $T(X) = \mathbb{N}_0 = \{0, 1, \ldots\}$.
- Symbolschreibweise für geometrische Verteilung mit Parameter p: $\mathrm{Geom}(p)$

Geometrische Verteilung
Geom(p)

Parameter:
$p \in (0,1)$

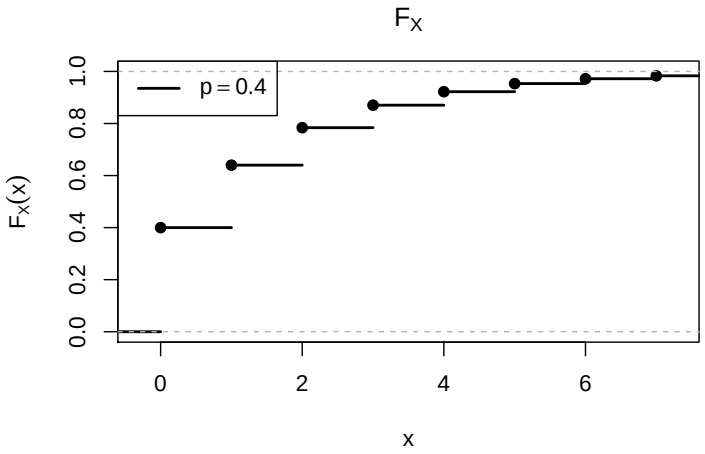
Träger: $T(X) = \mathbb{N}_0 = \{0, 1, \ldots\}$
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} (1-p)^x \cdot p & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & \text{für } x \geq 0 \end{cases}$$



Momente: $E(X) = \frac{1-p}{p}$ $\quad$ $\mathrm{Var}(X) = \frac{1-p}{p^2}$

$\gamma(X) = \frac{2-p}{\sqrt{1-p}}$ $\quad$ $\kappa(X) = \frac{p^2-9p+9}{1-p}$

Poisson-Verteilung

- „Grenzverteilung“ der Binomialverteilung
- Verwendung:
  - Approximation einer $B(n, p)$-Verteilung, wenn n (sehr) groß und p (sehr) klein ist.
  - „Faustregeln“ zur Anwendung der Approximation:

$$n \geq 50, \qquad p \leq 0.1, \qquad n \cdot p \leq 10$$

  - Poisson-Verteilung hat einzigen Parameter $\lambda > 0$, der zur Approximation einer $B(n, p)$-Verteilung auf $\lambda = n \cdot p$ gesetzt wird.
- Träger von Poisson-verteilten Zufallsvariablen X: $T(X) = \mathbb{N}_0 = \{0, 1, \ldots\}$
- Wahrscheinlichkeitsfunktion für $x \in T(X)$: $p_X(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$, wobei $e = \exp(1)$ die Eulersche Zahl ist, also $e \approx 2.71828$.
- Gültigkeit der Approximation beruht auf Konvergenz der Punktwahrscheinlichkeiten. Es gilt nämlich für alle $x \in \mathbb{N}_0$:

$$\lim_{\substack{n \to \infty \\ p \to 0 \\ n \cdot p \to \lambda}} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

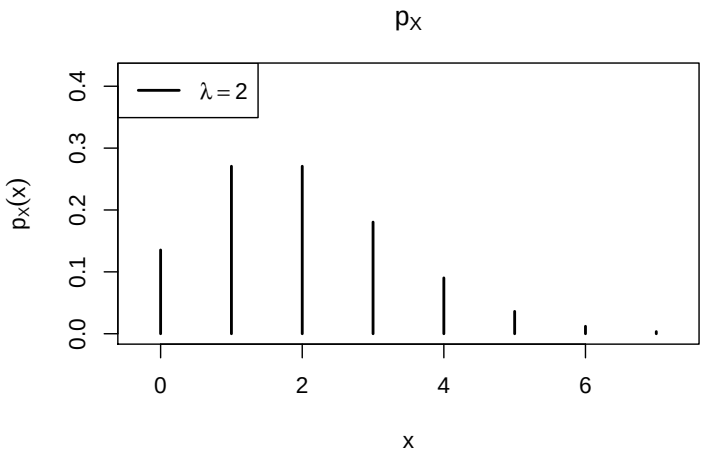
- Symbolschreibweise für Poisson-Verteilung mit Parameter λ: $\mathrm{Pois}(\lambda)$

Poisson-Verteilung
Pois(λ)

Parameter:
$\lambda > 0$

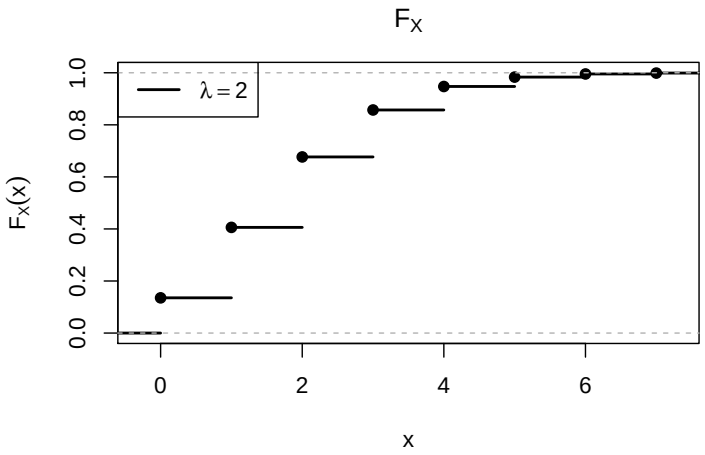
Träger: $T(X) = \mathbb{N}_0 = \{0, 1, \ldots\}$
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \sum_{\substack{x_i \in T(X) \\ x_i \leq x}} p_X(x_i)$$



Momente:

$$E(X) = \lambda \qquad \mathrm{Var}(X) = \lambda$$
$$\gamma(X) = \frac{1}{\sqrt{\lambda}} \qquad \kappa(X) = 3 + \frac{1}{\lambda}$$

Spezielle stetige Verteilungen

- Nun: Vorstellung spezieller parametrischer Verteilungsfamilien von *stetigen* Verteilungen
- In Verallgemeinerung des **Trägers** diskreter Verteilungen:
  Träger $T(X)$ einer stetigen Verteilung als **„Bereich positiver Dichte“**.
- Wegen Möglichkeit, Dichtefunktionen abzuändern, etwas genauer:

$$\begin{aligned} T(X) := \{x \in \mathbb{R} \,|\, & \text{es gibt eine Dichtefunktion } f_X \text{ von } X \text{ und ein } \epsilon > 0 \\ & \text{mit } (f_X(t) > 0 \text{ für alle } t \in [x-\epsilon, x]) \\ & \text{oder } (f_X(t) > 0 \text{ für alle } t \in [x, x+\epsilon])\} \end{aligned}$$

Stetige Gleichverteilung

- Einfachste stetige Verteilungsfamilie:
  Stetige Gleichverteilung auf Intervall $[a, b]$
- Modellierung einer stetigen Verteilung, in der alle Realisationen in einem Intervall $[a, b]$ als „gleichwahrscheinlich“ angenommen werden.
- Verteilung hängt von den beiden Parametern $a, b \in \mathbb{R}$ mit $a < b$ ab.
- Dichtefunktion f_X einer gleichverteilten Zufallsvariablen X kann auf Intervall $[a, b]$ konstant zu $\frac{1}{b-a}$ gewählt werden.
- Träger der Verteilung: $T(X) = [a, b]$
- Symbolschreibweise für stetige Gleichverteilung auf $[a, b]$: $X \sim \text{Unif}(a, b)$

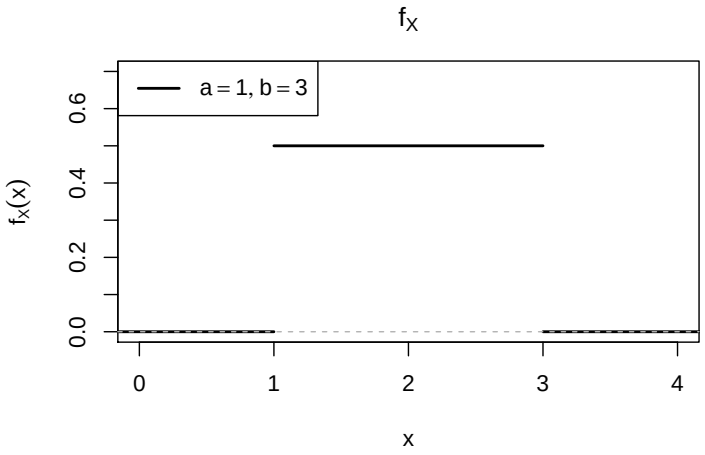
Stetige Gleichverteilung

Unif(a, b)

Parameter:
$a, b \in \mathbb{R}$ mit $a < b$

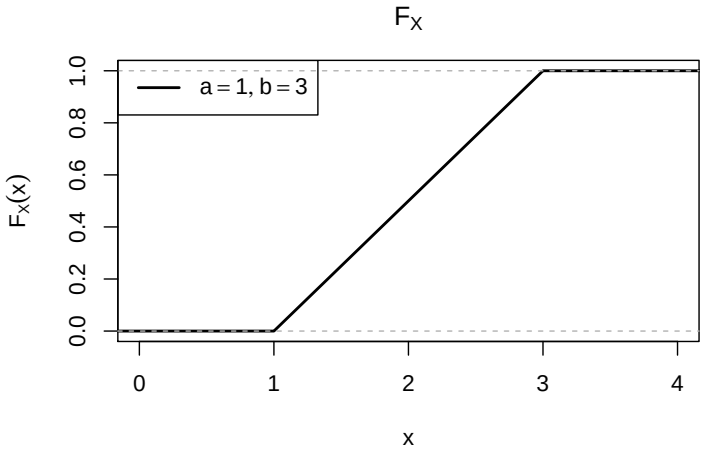
Träger: $T(X) = [a, b]$
Dichtefunktion: $f_X : \mathbb{R} \to \mathbb{R}$;

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion: $F_X : \mathbb{R} \to \mathbb{R}$;

$$F_X(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$



Momente:

$$E(X) = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

$$\gamma(X) = 0 \qquad \kappa(X) = \frac{9}{5}$$

Normalverteilung

- Verteilung entsteht als Grenzverteilung bei Durchschnittsbildung vieler (unabhängiger) Zufallsvariablen (später mehr!) $\rightsquigarrow$ Einsatz für Näherungen
- Familie der Normalverteilungen hat Lageparameter $\mu \in \mathbb{R}$, der mit Erwartungswert übereinstimmt, und Streuungsparameter $\sigma^2 > 0$, der mit Varianz übereinstimmt, Standardabweichung ist dann $\sigma := +\sqrt{\sigma^2}$.
- Verteilungsfunktion von Normalverteilungen schwierig zu handhaben, Berechnung muss i.d.R. mit Software/Tabellen erfolgen.
- Wichtige Eigenschaft der Normalverteilungsfamilie:

  Ist X normalverteilt mit Parameter $\mu = 0$ und $\sigma^2 = 1$, dann ist $aX + b$ für $a, b \in \mathbb{R}$ normalverteilt mit Parameter $\mu = b$ und $\sigma^2 = a^2$.

$\rightsquigarrow$ Zurückführung allgemeiner Normalverteilungen auf den Fall der **Standardnormalverteilung (Gauß-Verteilung)** mit Parameter $\mu = 0$ und $\sigma^2 = 1$, Tabellen/Algorithmen für Standardnormalverteilung damit einsetzbar.

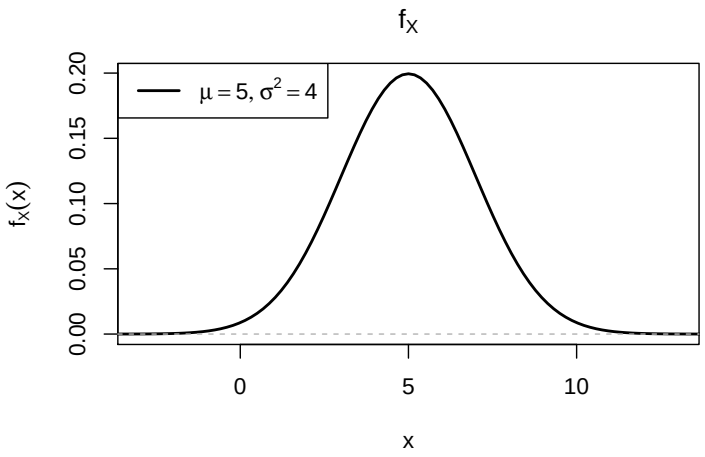
- Dichtefunktion der Standardnormalverteilung: φ, Verteilungsfunktion: Φ.
- Träger aller Normalverteilungen ist $T(X) = \mathbb{R}$.
- Symbolschreibweise für Normalverteilung mit Parameter μ, σ^2: $X \sim \mathrm{N}(\mu, \sigma^2)$

Normalverteilung
$\mathrm{N}(\mu, \sigma^2)$

Parameter:
$\mu \in \mathbb{R}, \sigma^2 > 0$

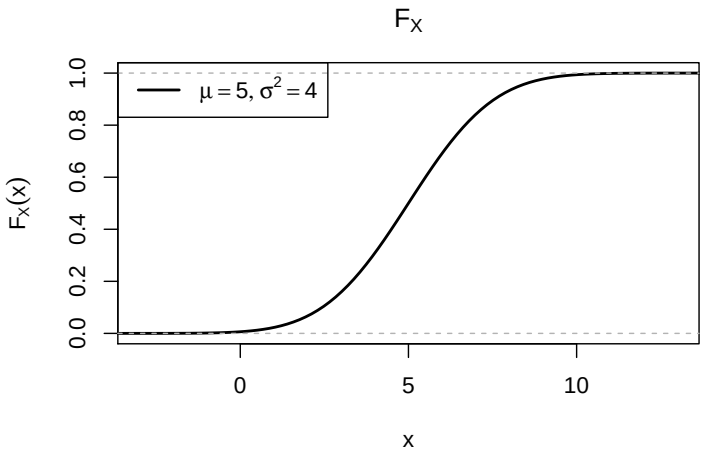
Träger: $T(X) = \mathbb{R}$
Dichtefunktion: $f_X : \mathbb{R} \to \mathbb{R}$;

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



Verteilungsfunktion:

$$F_X : \mathbb{R} \to \mathbb{R}; F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$



Momente: $E(X) = \mu$ $\quad$ $\mathrm{Var}(X) = \sigma^2$
$\gamma(X) = 0$ $\quad$ $\kappa(X) = 3$

Exponentialverteilung

- Beliebte Verteilungsfamilie zur Modellierung von **Wartezeiten**.
- Verteilung entsteht als Grenzverteilung der geometrischen Verteilung (Anzahl Fehlversuche vor erstem Erfolg bei wiederholter, unabhängiger Ausführung eines Bernoulli-Experiments) bei Erfolgswahrscheinlichkeit $p \to 0$.
- Da die Anzahl X der benötigten Versuche für $p \to 0$ offensichtlich immer größere Werte annehmen wird, wird statt der *Anzahl* der benötigten Versuche die *Zeit* zur Durchführung der benötigten Versuche modelliert, und mit $p \to 0$ zugleich die *pro Zeiteinheit* durchgeführten Versuche n des Bernoulli-Experiments so erhöht, dass $p \cdot n =: \lambda$ konstant bleibt.
- Einziger Parameter der resultierenden Exponentialverteilung ist damit die als „erwartete Anzahl von Erfolgen pro Zeiteinheit“ interpretierbare Größe $\lambda > 0$.
- Ist X exponentialverteilt mit Parameter λ, so erhält man $F_X(x)$ aus der Verteilungsfunktion der geometrischen Verteilung für $x \geq 0$ gemäß

$$F_X(x) = \lim_{n\to\infty} 1 - \left(1 - \frac{\lambda}{n}\right)^{n\cdot x} = \lim_{n\to\infty} 1 - \left(1 + \frac{-\lambda \cdot x}{n \cdot x}\right)^{n\cdot x} = 1 - e^{-\lambda x} \ .$$

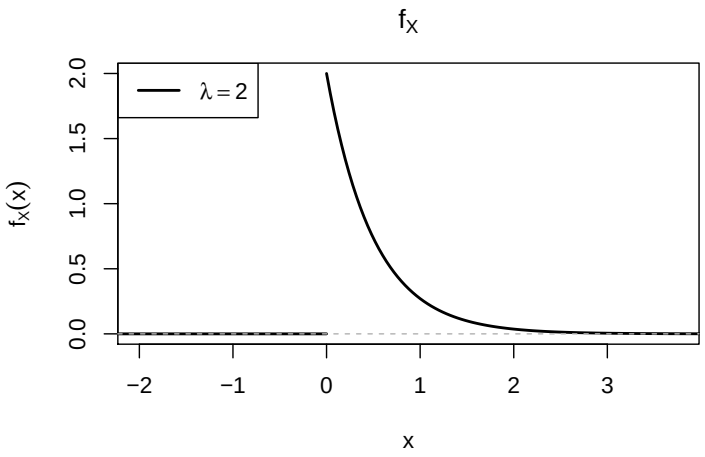
- Träger der Exponentialverteilungsfamilie ist $\mathbb{R}_+ := \{x \in \mathbb{R} \,|\, x \geq 0\}$.
- Symbolschreibweise für Exponentialverteilung mit Parameter λ: $X \sim \mathrm{Exp}(\lambda)$

Exponentialverteilung
$\mathrm{Exp}(\lambda)$

Parameter:
$\lambda > 0$

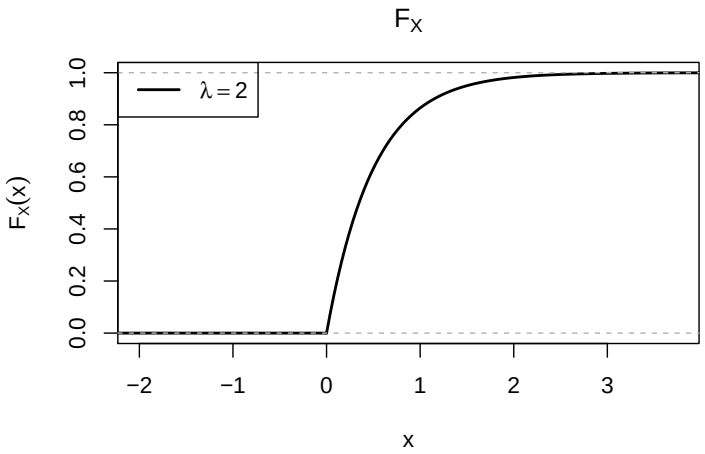
Träger: $T(X) = \mathbb{R}_+$
Dichtefunktion: $f_X : \mathbb{R} \to \mathbb{R}$;

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



Verteilungsfunktion: $F_X : \mathbb{R} \to \mathbb{R}$;

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } x \geq 0 \end{cases}$$



Momente: $E(X) = \frac{1}{\lambda}$ $\quad$ $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
$\gamma(X) = 2$ $\quad$ $\kappa(X) = 9$

Verwendung spezieller Verteilungen

- Übliche Vorgehensweise zur Berechnung von (Intervall-)Wahrscheinlichkeiten für Zufallsvariablen X: Verwendung der Verteilungsfunktion F_X
- Problem bei einigen der vorgestellten Verteilungen:
  Verteilungsfunktion F_X schlecht handhabbar bzw. nicht leicht auszuwerten!
- Traditionelle Lösung des Problems: *Vertafelung* bzw. *Tabellierung* der Verteilungsfunktionswerte, Ablesen der Werte dann aus Tabellenwerken.
- Lösung nicht mehr zeitgemäß: (kostenlose) PC-Software für alle benötigten Verteilungsfunktionen verfügbar, zum Beispiel Statistik-Software **R** (`http://www.r-project.org`)
- **Aber:** In Klausur keine PCs verfügbar, daher dort Rückgriff auf Tabellen.
- Problematische Verteilungsfunktionen (bisher) sind die der Standardnormalverteilung, Binomialverteilung sowie Poisson-Verteilung.
- Tabellen oder Tabellenausschnitte zu diesen Verteilungen werden in Klausur (sofern benötigt) zur Verfügung gestellt!
- Auch das Bestimmen von Quantilen ist für diese Verteilungen nicht ohne Hilfsmittel möglich und muss mit Hilfe weiterer Tabellen oder auf Grundlage der tabellierten Verteilungsfunktionswerte erfolgen.

Ausschnitt aus Tabelle für $\Phi(x)$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

Ausschnitt aus Tabelle für $F_{B(n,p)}(x)$

n	x	$p = 0.05$	$p = 0.10$	$p = 0.15$	$p = 0.20$	$p = 0.25$	$p = 0.30$	$p = 0.35$	$p = 0.40$
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000
1	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600
2	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400
2	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160
3	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480
3	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360
3	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296
4	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752
4	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208
4	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744
4	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778
5	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370
5	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826
5	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130
5	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898
5	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467
6	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333
6	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443
6	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208
6	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590
6	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959
6	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

R-Befehle für spezielle Verteilungen

- Verteilungsfunktionen können sofort nach dem Start von **R** mit den folgenden Befehlen ausgewertet werden:

Verteilung von X	Parameter	F_X an Stelle `x` mit **R**
$B(n,p)$	`size`$=n$, `prob`$=p$	`pbinom(x,size,prob)`
$\mathrm{Geom}(p)$	`prob`$=p$	`pgeom(x,prob)`
$\mathrm{Pois}(\lambda)$	`lambda`$=\lambda$	`ppois(x,lambda)`
$\mathrm{Unif}(a,b)$	`min`$=a$, `max`$=b$	`punif(x,min,max)`
$N(\mu,\sigma^2)$	`mean`$=\mu$, `sd`$=\sqrt{\sigma^2}$	`pnorm(x,mean,sd)`
$\mathrm{Exp}(\lambda)$	`rate`$=\lambda$	`pexp(x,rate)`

- Ersetzt man in den Befehlen den ersten Buchstaben `p` durch `d` (z.B. `dnorm`), so erhält man den Wert der Dichtefunktion bzw. Wahrscheinlichkeitsfunktion an der Stelle `x`.
- Ersetzt man in den Befehlen den ersten Buchstaben `p` durch `q` (z.B. `qnorm`) und `x` durch `p`, so erhält man das (bzw. ein) `p`-Quantil der zugehörigen Verteilung.
- Ersetzt man schließlich in den Befehlen den ersten Buchstaben `p` durch `r` (z.B. `rnorm`) und `x` durch `n`$\in \mathbb{N}$, so erhält man `n` (Pseudo-)Zufallszahlen zur zugehörigen Verteilung.

Hinweise zur Tabellennutzung

- Bezeichnet $F_{B(n,p)}$ für $n \in \mathbb{N}$ und $p \in (0,1)$ die Verteilungsfunktion der $B(n,p)$-Verteilung, so gilt (!)

$$F_{B(n,1-p)}(x) = 1 - F_{B(n,p)}(n-x-1)$$

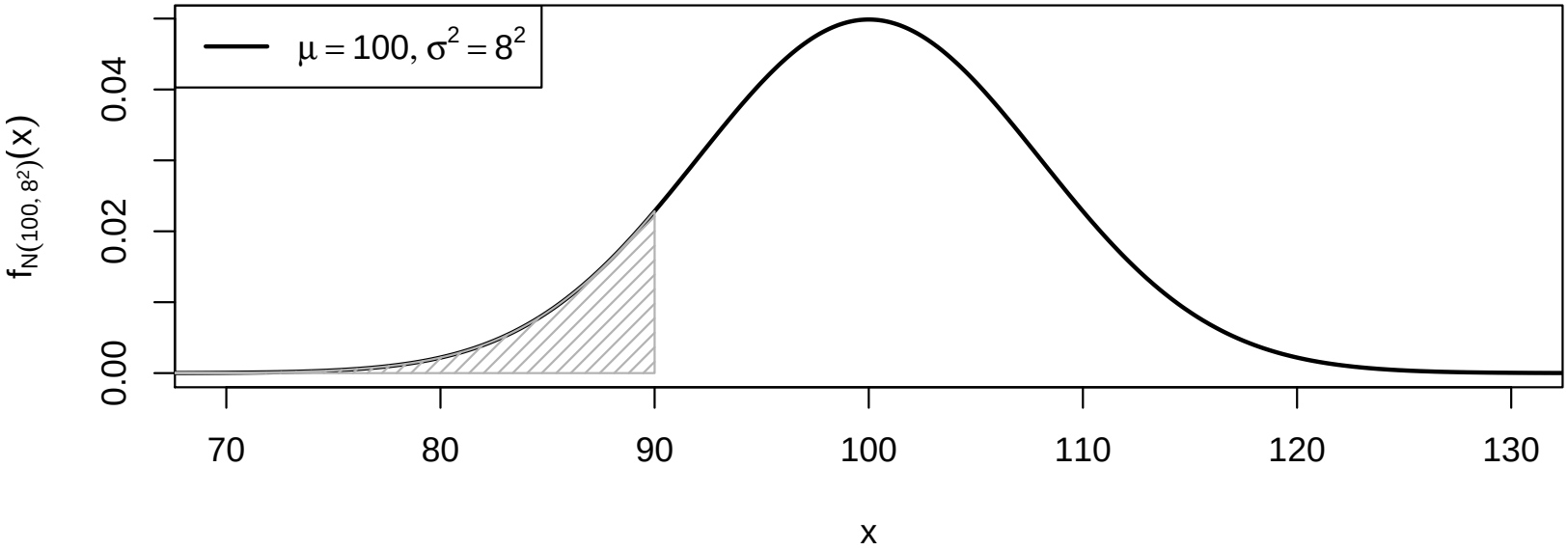
  für alle $n \in \mathbb{N}$, $p \in (0,1)$, $x \in \{0,\ldots,n-1\}$. Daher werden Tabellen zur Binomialverteilung nur für $p \in (0,0.5]$ erstellt, und die benötigten Werte für $p \in [0.5,1)$ mit obiger Formel aus den Werten für $p \in (0,0.5]$ gewonnen.
- Wegen der Symmetrie der Standardnormalverteilung um 0 gilt nicht nur $\varphi(x) = \varphi(-x)$ für alle $x \in \mathbb{R}$, sondern auch (vgl. Folie 216)

$$\Phi(x) = 1 - \Phi(-x) \qquad \text{für alle } x \in \mathbb{R}\ .$$

  Daher werden Tabellen für $\Phi(x)$ in der Regel nur für $x \in \mathbb{R}_+$ erstellt.
- Zur Bestimmung von Quantilen darf in der Klausur ein beliebiger Wert des Intervalls, in dem das Quantil laut Tabelle liegen muss, eingesetzt werden; eine lineare Interpolation ist zwar sinnvoll, aber nicht nötig!
- Generell gilt: Ist ein Wert nicht tabelliert, wird stattdessen ein „naheliegender“ Wert aus der Tabelle eingesetzt.
  Beispiel: Für fehlenden Wert $F_{B(4,0.28)}(2)$ wird $F_{B(4,0.3)}(2)$ eingesetzt.

Beispiel: Arbeiten mit Normalverteilungstabelle

- **Frage:** Mit welcher Wahrscheinlichkeit nimmt eine $N(100, 8^2)$-verteilte Zufallsvariable Werte kleiner als 90 an? (Wie groß ist die schraffierte Fläche?)

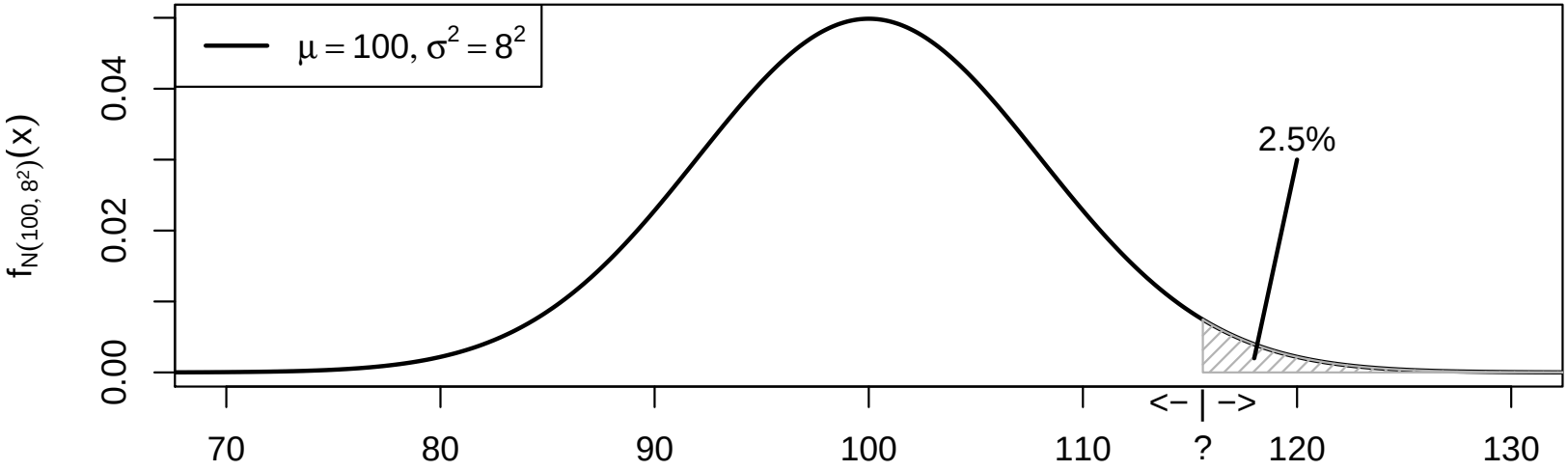


- **Antwort:** Ist $X \sim N(100, 8^2)$, so gilt:

$$\begin{aligned} P\{X < 90\} &= F_{N(100,8^2)}(90) = \Phi\left(\frac{90-100}{8}\right) \\ &= \Phi(-1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056 \end{aligned}$$

$\rightsquigarrow$ Die gesuchte Wahrscheinlichkeit ist $0.1056 = 10.56\%$.

- **Frage:** Welchen Wert x überschreitet eine $N(100, 8^2)$-verteilte Zufallsvariable nur mit 2.5% Wahrscheinlichkeit? (Welche linke Grenze x führt bei der schraffierten Fläche zu einem Flächeninhalt von 0.025?)



- **Antwort:** Ist $X \sim N(100, 8^2)$, so ist das 97.5%- bzw. 0.975-Quantil von X gesucht. Mit

$$F_X(x) = F_{N(100,8^2)}(x) = \Phi\left(\frac{x-100}{8}\right)$$

erhält man

$$\Phi\left(\frac{x-100}{8}\right) \stackrel{!}{=} 0.975 \quad \Leftrightarrow \quad \frac{x-100}{8} = \Phi^{-1}(0.975) = 1.96$$

$$\Rightarrow \quad x = 8 \cdot 1.96 + 100 = 115.68$$

Beispiel: Arbeiten mit Statistik-Software **R**

- Beantwortung der Fragen (noch) einfacher mit Statistik-Software **R**:
- **Frage:** Mit welcher Wahrscheinlichkeit nimmt eine $N(100, 8^2)$-verteilte Zufallsvariable Werte kleiner als 90 an?
- **Antwort:**

```
> pnorm(90,mean=100,sd=8)
[1] 0.1056498
```

- **Frage:** Welchen Wert x überschreitet eine $N(100, 8^2)$-verteilte Zufallsvariable nur mit 2.5% Wahrscheinlichkeit?
- **Antwort:**

```
> qnorm(0.975,mean=100,sd=8)
[1] 115.6797
```

oder alternativ

```
> qnorm(0.025,mean=100,sd=8,lower.tail=FALSE)
[1] 115.6797
```

Inhaltsverzeichnis

(Ausschnitt)

10 Mehrdimensionale Zufallsvariablen
- Borelsche sigma-Algebra
- Diskrete Zufallsvektoren
- Stetige Zufallsvektoren
- Randverteilungen
- (Stochastische) Unabhängigkeit
- Bedingte Verteilungen
- Momente zweidimensionaler Zufallsvektoren
- Momente höherdimensionaler Zufallsvektoren

Mehrdimensionale Zufallsvariablen/Zufallsvektoren I

- Im Folgenden: *Simultane* Betrachtung *mehrerer* (endlich vieler) Zufallsvariablen über *demselben* Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.
- Ist $n \in \mathbb{N}$ die Anzahl der betrachteten Zufallsvariablen, so fasst man die n Zufallsvariablen $X_1, \ldots, X_n$ in einem n-dimensionalen Vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ zusammen.
- Damit ist $\mathbb{R}^n$ der Wertebereich der Abbildung $\mathbf{X} : \Omega \to \mathbb{R}^n$, als σ-Algebra über $\mathbb{R}^n$ wählt man die n-dimensionale Borelsche σ-Algebra $\mathcal{B}^n$, in der alle karthesischen Produkte von n Elementen aus $\mathcal{B}$ enthalten sind.
- Insbesondere enthält $\mathcal{B}^n$ alle endlichen und abzählbar unendlichen Teilmengen von $\mathbb{R}^n$ sowie alle karthesischen Produkte von n Intervallen aus $\mathbb{R}$.
- Damit lassen sich die meisten bekannten Konzepte eindimensionaler Zufallsvariablen leicht übertragen.
- Ähnlich zur Situation bei mehrdimensionalen Merkmalen in der deskriptiven Statistik werden viele Darstellungen im Fall $n > 2$ allerdings schwierig.

Mehrdimensionale Zufallsvariablen/Zufallsvektoren II

Definition 10.1 (Zufallsvektor, Mehrdimensionale Zufallsvariable)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $n \in \mathbb{N}$,

$$\mathbf{X} : \Omega \to \mathbb{R}^n : \mathbf{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega))'$$

eine $\mathcal{F} - \mathcal{B}^n-$messbare Abbildung. Dann heißen $\mathbf{X}$ n-**dimensionale Zufallsvariable** bzw. n-**dimensionaler Zufallsvektor** über $(\Omega, \mathcal{F}, P)$ und die gemäß Definition 8.3 gebildete Bildwahrscheinlichkeit

$$P_{\mathbf{X}} : \mathcal{B}^n \to \mathbb{R}; \mathbf{B} \mapsto P(\mathbf{X}^{-1}(\mathbf{B}))$$

(gemeinsame) Wahrscheinlichkeitsverteilung oder kürzer **(gemeinsame) Verteilung** von $\mathbf{X}$. $(\mathbb{R}^n, \mathcal{B}^n, P_{\mathbf{X}})$ ist damit ebenfalls ein Wahrscheinlichkeitsraum. Liegt nach Durchführung des Zufallsexperiments $(\Omega, \mathcal{F}, P)$ das Ergebnis $\omega \in \Omega$ vor, so heißt der zugehörige Wert $\mathbf{x} = \mathbf{X}(\omega)$ die **Realisierung** oder **Realisation** von $\mathbf{X}$.

Mehrdimensionale Zufallsvariablen/Zufallsvektoren III

- Wie im eindimensionalen Fall sind Kurzschreibweisen (zum Beispiel) der Form

$$P\{X_1 \leq x_1, \ldots, X_n \leq x_n\} := P_{\mathbf{X}}((-\infty, x_1] \times \cdots \times (-\infty, x_n])$$

  für $x_1, \ldots, x_n \in \mathbb{R}$ geläufig.
- Auch hier legen die Wahrscheinlichkeiten solcher Ereignisse die Verteilung des n-dimensionalen Zufallsvektors bereits eindeutig fest, und man definiert analog zum eindimensionalen Fall die gemeinsame Verteilungsfunktion

$$F_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}; F_{\mathbf{X}}(x_1, \ldots, x_n) := P_{\mathbf{X}}((-\infty, x_1] \times \ldots \times (-\infty, x_n]) \ .$$

- Gemeinsame Verteilungsfunktionen mehrdimensionaler Zufallsvariablen sind allerdings für den praktischen Einsatz im Vergleich zur eindimensionalen Variante relativ unbedeutend und werden daher hier nicht weiter besprochen.

Diskrete Zufallsvektoren I

- Ist analog zum eindimensionalen Fall

$$\mathbf{X}(\Omega) := \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \,|\, \mathbf{x} = \mathbf{X}(\omega) \text{ für (mindestens) ein } \omega \in \Omega\}$$

  endlich oder abzählbar unendlich bzw. existiert (wiederum etwas allgemeiner) eine endliche oder abzählbar unendliche Menge $\mathbf{B} \subseteq \mathbb{R}^n$ mit $P(\{\mathbf{X} \in \mathbf{B}\}) = 1$, so nennt man auch solche Zufallsvektoren „diskret“.
- Mit Hilfe einer (mehrdimensionalen) Wahrscheinlichkeitsfunktion $p_{\mathbf{X}}$ mit $p_{\mathbf{X}}(\mathbf{x}) := P_{\mathbf{X}}(\{\mathbf{x}\})$ für $\mathbf{x} \in \mathbb{R}^n$ können Wahrscheinlichkeiten $P\{\mathbf{X} \in \mathbf{A}\}$ für Ereignisse $\mathbf{A} \in \mathcal{B}^n$ wiederum durch Aufsummieren der Punktwahrscheinlichkeiten aller Trägerpunkte $\mathbf{x}_i$ mit $\mathbf{x}_i \in \mathbf{A}$ berechnet werden, also durch:

$$P\{\mathbf{X} \in \mathbf{A}\} = \sum_{\mathbf{x}_i \in \mathbf{A} \cap T(\mathbf{X})} p_{\mathbf{X}}(\mathbf{x}_i) \qquad \text{für alle } \mathbf{A} \in \mathcal{B}^n$$

Diskrete Zufallsvektoren II

Definition 10.2 (Diskreter Zufallsvektor)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $n \in \mathbb{N}$, $\mathbf{X}$ ein n-dimensionaler Zufallsvektor über $(\Omega, \mathcal{F}, P)$ und $\mathbf{B} \subseteq \mathbb{R}^n$ endlich oder abzählbar unendlich mit $P(\{\mathbf{X} \in \mathbf{B}\}) = 1$. Dann nennt man

- $\mathbf{X}$ einen **diskreten** Zufallsvektor,
- $p_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]; p_{\mathbf{X}}(\mathbf{x}) := P_{\mathbf{X}}(\{\mathbf{x}\})$ die **(gemeinsame) Wahrscheinlichkeitsfunktion** von $\mathbf{X}$,
- $T(\mathbf{X}) := \{\mathbf{x} \in \mathbb{R}^n \mid p_{\mathbf{X}}(\mathbf{x}) > 0\}$ den **Träger** von $\mathbf{X}$ sowie alle Elemente $\mathbf{x} \in T(\mathbf{X})$ **Trägerpunkte** von $\mathbf{X}$ und deren zugehörige Wahrscheinlichkeitsfunktionswerte $p_{\mathbf{X}}(\mathbf{x})$ **Punktwahrscheinlichkeiten**.

Stetige Zufallsvektoren I

- Zweiter wichtiger Spezialfall (wie im eindimensionalen Fall): **stetige** n-dimensionale Zufallsvektoren $\mathbf{X}$
- Wiederum gilt $P_{\mathbf{X}}(\mathbf{B}) = 0$ *insbesondere* für alle endlichen oder abzählbar unendlichen Teilmengen $\mathbf{B} \subseteq \mathbb{R}^n$.
- Auch hier ist die definierende Eigenschaft die Möglichkeit zur Berechnung spezieller Wahrscheinlichkeiten als Integral über eine (nun mehrdimensionale) Dichtefunktion.
- In Verallgemeinerung der Berechnung von *Intervallwahrscheinlichkeiten* im eindimensionalen Fall müssen nun *Wahrscheinlichkeiten von Quadern* als (Mehrfach-)Integral über eine Dichtefunktion berechnet werden können.

Stetige Zufallsvektoren II

Definition 10.3 (Stetiger Zufallsvektor, (gemeinsame) Dichtefunktion)

Seien $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $n \in \mathbb{N}$ und $\mathbf{X}$ ein n-dimensionaler Zufallsvektor über $(\Omega, \mathcal{F}, P)$. Gibt es eine nichtnegative Abbildung $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ mit

$$P_{\mathbf{X}}(\mathbf{A}) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\mathbf{X}}(t_1, \ldots, t_n) dt_n \cdots dt_1 \tag{5}$$

für alle Quader $\mathbf{A} = (a_1, b_1] \times \cdots \times (a_n, b_n] \subset \mathbb{R}^n$ mit $a_1 \leq b_1, \ldots, a_n \leq b_n$, so heißt der Zufallsvektor $\mathbf{X}$ **stetig**. Jede nichtnegative Abbildung $f_{\mathbf{X}}$ mit der Eigenschaft (5) heißt **(gemeinsame) Dichtefunktion** von $\mathbf{X}$.

Notationen im Spezialfall $n = 2$

- Im Folgenden wird (auch für weitere Anwendungen) regelmäßig der Spezialfall $n = 2$ betrachtet.
- Zur Vereinfachung der Darstellung (insbesondere zur Vermeidung doppelter Indizes) sei der betrachtete Zufallsvektor $\mathbf{X}$ dann mit $\mathbf{X} = (X, Y)'$ oder $\mathbf{X} = (X, Y)$ statt $\mathbf{X} = (X_1, X_2)'$ bezeichnet.
- Stetige 2-dimensionale Zufallsvektoren $\mathbf{X} = (X, Y)$ werden in der Regel durch die Angabe einer gemeinsamen Dichtefunktion

$$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R};\ (x, y) \mapsto f_{X,Y}(x, y)$$

  spezifizert.
- Ist $\mathbf{X} = (X, Y)$ ein zweidimensionaler diskreter Zufallsvektor mit „wenigen“ Trägerpunkten, stellt man die gemeinsame Verteilung — analog zu den Kontingenztabellen der deskriptiven Statistik — gerne in Tabellenform dar.

Beispiel: (Gemeinsame) Wahrscheinlichkeitsfunktion

bei zweidimensionaler diskreter Zufallsvariable

- Ist $\mathbf{X} = (X, Y)$ zweidimensionale diskrete Zufallsvariable mit endlichem Träger, $A := T(X) = \{x_1, \ldots, x_k\}$ der Träger von X und $B := T(Y) = \{y_1, \ldots, y_l\}$ der Träger von Y, so werden die Werte der Wahrscheinlichkeitsfunktion $p_{\mathbf{X}}$ auch mit

$$p_{ij} := p_{(X,Y)}(x_i, y_j) \qquad \text{für } i \in \{1, \ldots, k\} \text{ und } j \in \{1, \ldots, l\}$$

bezeichnet und wie folgt tabellarisch dargestellt:

$X \setminus Y$	y_1	y_2	$\cdots$	y_l
x_1	p_{11}	p_{12}	$\cdots$	p_{1l}
x_2	p_{21}	p_{22}	$\cdots$	p_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_k	p_{k1}	p_{k2}	$\cdots$	p_{kl}

Randverteilungen I

- Wie in der deskriptiven Statistik lassen sich die Verteilungen der einzelnen Zufallsvariablen eines n-dimensionalen Zufallsvektors auch aus der gemeinsamen Verteilung gewinnen.
- Analog zu den „Randhäufigkeiten“ erhält man so die **Randverteilungen** der einzelnen Komponenten des Zufallsvektors.
- Ist $\mathbf{X}$ diskreter n-dimensionaler Zufallsvektor mit Wahrscheinlichkeitsfunktion $p_{\mathbf{X}}$, so erhält man für $j \in \{1, \ldots, n\}$ die Wahrscheinlichkeitsfunktion p_{X_j} zur j-ten Komponente X_j durch:

$$p_{X_j}(x) = \sum_{\substack{\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n}) \\ x_{i,j} = x}} p_{\mathbf{X}}(\mathbf{x}_i)$$

Randverteilungen II

- Ist $\mathbf{X}$ stetiger n-dimensionaler Zufallsvektor mit gemeinsamer Dichtefunktion $f_{\mathbf{X}}$, so erhält man für $j \in \{1, \ldots, n\}$ eine Dichtefunktion $f_{X_j} : \mathbb{R} \to \mathbb{R}$ zur j-ten Komponente X_j durch:

$$f_{X_j}(x) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1)\text{-mal}} f_{\mathbf{X}}(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n) dx_n \cdots dx_{j+1} dx_{j-1} \cdots dx_1$$

- Für $\mathbf{X} = (X, Y)$ erhält man also eine Randdichtefunktion f_X zu X durch

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

sowie eine Randdichtefunktion f_Y zu Y durch

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \ .$$

Fortsetzung Beispiel (zweidimensional, diskret)

Ergänzung der gemeinsamen Wahrscheinlichkeitstabelle um Randverteilungen

- Ist $A = T(X) = \{x_1, \ldots, x_k\}$ der Träger von X und $B = T(Y) = \{y_1, \ldots, y_l\}$ der Träger von Y, so erhält man für $i \in \{1, \ldots, k\}$ als Zeilensummen

$$p_{i\cdot} := p_X(x_i) = \sum_{j=1}^{l} p_{(X,Y)}(x_i, y_j) = \sum_{j=1}^{l} p_{ij}$$

sowie für $j \in \{1, \ldots, l\}$ als Spaltensummen

$$p_{\cdot j} := p_Y(y_j) = \sum_{i=1}^{k} p_{(X,Y)}(x_i, y_j) = \sum_{i=1}^{k} p_{ij}$$

und damit insgesamt die folgende ergänzte Tabelle:

$X \setminus Y$	y_1	y_2	$\cdots$	y_l	$p_{i\cdot}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1l}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2l}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	p_{k1}	p_{k2}	$\cdots$	p_{kl}	$p_{k\cdot}$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot l}$	1

(Stochastische) Unabhängigkeit von Zufallsvariablen I

- Die Komponenten $X_1, \ldots, X_n$ eines n-dimensionalen Zufallsvektors werden genau dann **stochastisch unabhängig** genannt, wenn alle Ereignisse der Form

$$\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\} \qquad \text{für } B_1, \ldots, B_n \in \mathcal{B}$$

stochastisch unabhängig sind:

Definition 10.4

Seien $n \in \mathbb{N}$ und $X_1, \ldots, X_n$ Zufallsvariablen über demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Dann heißen die Zufallsvariablen $X_1, \ldots, X_n$ **(stochastisch) unabhängig**, wenn für alle $B_1, \ldots, B_n \in \mathcal{B}$ gilt:

$$P\{X_1 \in B_1, \ldots, X_n \in B_n\} = \prod_{i=1}^{n} P\{X_i \in B_i\} = P\{X_1 \in B_1\} \cdot \ldots \cdot P\{X_n \in B_n\}$$

(Stochastische) Unabhängigkeit von Zufallsvariablen II

- Man kann weiter zeigen, dass n diskrete Zufallsvariablen $X_1, \ldots, X_n$ über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ genau dann stochastisch unabhängig sind, wenn für den (in diesem Fall ebenfalls diskreten) Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)'$ bzw. die zugehörigen Wahrscheinlichkeitsfunktionen für alle $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt:

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n)$$

- Insbesondere sind X und Y im Fall $n = 2$ mit $\mathbf{X} = (X, Y)$ bei endlichen Trägern $A = T(X) = \{x_1, \ldots, x_k\}$ von X und $B = T(Y) = \{y_1, \ldots, y_l\}$ von Y unabhängig, falls $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ gilt für alle $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.

(Stochastische) Unabhängigkeit von Zufallsvariablen III

- Weiterhin sind n stetige Zufallsvariablen $X_1, \ldots, X_n$ über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ stochastisch unabhängig, wenn der Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)'$ stetig ist und eine gemeinsame Dichte $f_{\mathbf{X}}$ bzw. Randdichten $f_{X_1}, \ldots, f_{X_n}$ existieren, so dass für alle $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

- Insbesondere sind stetige Zufallsvariablen X und Y im Fall $n = 2$ mit $\mathbf{X} = (X, Y)$ genau dann unabhängig, wenn es Dichtefunktionen f_X von X, f_Y von Y sowie $f_{X,Y}$ von (X, Y) gibt mit

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{für alle } x, y \in \mathbb{R} \ .$$

Bedingte Verteilungen 2-dim. Zufallsvektoren I

Definition 10.5 (Bedingte Wahrscheinlichkeitsverteilung (diskret))

Seien $\mathbf{X} = (X, Y)$ ein diskreter zweidimensionaler Zufallsvektor mit Wahrscheinlichkeitsfunktion $p_{(X,Y)}$, p_Y die Wahrscheinlichkeitsfunktion zur Randverteilung von Y, $y \in T(Y)$. Dann heißt die durch die Wahrscheinlichkeitsfunktion

$$p_{X|Y=y} : \mathbb{R} \to [0,1]; p_{X|Y=y}(x) = \frac{p_{(X,Y)}(x,y)}{p_Y(y)}$$

definierte diskrete Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von X unter der Bedingung $Y = y$**. Analog nennt man für $x \in T(X)$ die durch die Wahrscheinlichkeitsfunktion

$$p_{Y|X=x} : \mathbb{R} \to [0,1]; p_{Y|X=x}(y) = \frac{p_{(X,Y)}(x,y)}{p_X(x)}$$

definierte diskrete Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von Y unter der Bedingung $X = x$**.

Bedingte Verteilungen 2-dim. Zufallsvektoren II

Definition 10.6 (Bedingte Wahrscheinlichkeitsverteilung (stetig))

Seien $\mathbf{X} = (X, Y)$ ein stetiger zweidimensionaler Zufallsvektor, $f_{(X,Y)}$ eine gemeinsame Dichtefunktion von (X, Y), f_Y eine Dichtefunktion zur Randverteilung von Y, $y \in \mathbb{R}$ eine Stetigkeitsstelle von f_Y mit $f_Y(y) > 0$. Dann heißt die durch die Dichtefunktion

$$f_{X|Y=y} : \mathbb{R} \to \mathbb{R}; f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

definierte stetige Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von X unter der Bedingung $Y = y$.**
Analog nennt man für Stetigkeitsstellen $x \in \mathbb{R}$ von f_X mit $f_X(x) > 0$ die durch die Dichtefunktion

$$f_{Y|X=x} : \mathbb{R} \to \mathbb{R}; f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}$$

definierte stetige Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von Y unter der Bedingung $X = x$.**

Bemerkungen I

- Mit den üblichen Bezeichnungen und Abkürzungen für zweidimensionale Zufallsvektoren mit endlichem Träger erhält man mit Definition 10.5 analog zu den bedingten Häufigkeiten bei zweidimensionalen Merkmalen

$$p_{X|Y=y_j}(x_i) = \frac{p_{ij}}{p_{\cdot j}} \qquad \text{und} \qquad p_{Y|X=x_i}(y_j) = \frac{p_{ij}}{p_{i\cdot}}$$

  für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.
- Ebenfalls analog zur deskriptiven Statistik sind zwei Zufallsvariablen also genau dann stochastisch unabhängig, wenn die bedingten Verteilungen jeweils mit den zugehörigen Randverteilungen übereinstimmen, also wenn
  - im diskreten Fall $p_{X|Y=y}(x) = p_X(x)$ für alle $x \in \mathbb{R}$ und alle $y \in T(Y)$ sowie $p_{Y|X=x}(y) = p_Y(y)$ für alle $y \in \mathbb{R}$ und alle $x \in T(X)$ gilt.
  - im stetigen Fall (bedingte) Dichtefunktionen f_X, f_Y, $f_{Y|X=x}$, $f_{X|Y=y}$ existieren mit $f_{X|Y=y}(x) = f_X(x)$ für alle $x \in \mathbb{R}$ und alle Stetigkeitsstellen $y \in \mathbb{R}$ von f_Y mit $f_Y(y) > 0$ sowie $f_{Y|X=x}(y) = f_Y(y)$ für alle $y \in \mathbb{R}$ und alle Stetigkeitsstellen $x \in \mathbb{R}$ von f_X mit $f_X(x) > 0$.

Bemerkungen II

- Durch bedingte Dichtefunktionen $f_{X|Y=y}$ bzw. $f_{Y|X=x}$ oder bedingte Wahrscheinlichkeitsfunktionen $p_{X|Y=y}$ bzw. $p_{Y|X=x}$ definierte bedingte Verteilungen können *im weiteren Sinne* als Verteilungen eindimensionaler Zufallsvariablen $X|Y=y$ bzw. $Y|X=x$ aufgefasst werden.
- Damit lassen sich viele für eindimensionale Zufallsvariablen bekannte Methoden/Konzepte in natürlicher Weise übertragen, insbesondere auf
  - bedingte Verteilungsfunktionen $F_{X|Y=y}$ bzw. $F_{Y|X=x}$,
  - bedingte Wahrscheinlichkeitsmaße $P_{X|Y=y}$ bzw. $P_{Y|X=x}$,
  - bedingte Träger $T(X|Y=y)$ bzw. $T(Y|X=x)$,
  - bedingte Erwartungswerte $\mathrm{E}(X|Y=y)$ bzw. $\mathrm{E}(Y|X=x)$,
  - bedingte Varianzen $\mathrm{Var}(X|Y=y)$ bzw. $\mathrm{Var}(Y|X=x)$,
- Den bedingten Erwartungswert $\mathrm{E}(Y|X=x)$ erhält man zum Beispiel
  - im diskreten Fall durch: $\mathrm{E}(Y|X=x) = \sum_{y_i} y_i \cdot p_{Y|X=x}(y_i)$
  - im stetigen Fall durch: $\mathrm{E}(Y|X=x) = \int_{-\infty}^{+\infty} y \cdot f_{Y|X=x}(y)dy$
- Gültig bleibt auch der VZS: $\mathrm{Var}(X|Y=y) = \mathrm{E}(X^2|Y=y) - [\mathrm{E}(X|Y=y)]^2$

Beispiel: Zweidimensionale Normalverteilung I

- Wichtige mehrdimensionale stetige Verteilung: **mehrdimensionale (multivariate) Normalverteilung**
- Spezifikation am Beispiel der zweidimensionalen (bivariaten) Normalverteilung durch Angabe einer Dichtefunktion

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}$$

  abhängig von den Parametern $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, $\rho \in (-1, 1)$.
- Man kann zeigen, dass die Randverteilungen von (X, Y) dann wieder (eindimensionale) Normalverteilungen sind, genauer gilt $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$

Beispiel: Zweidimensionale Normalverteilung II

- Sind f_X bzw. f_Y die wie auf Folie 242 definierten Dichtefunktionen zur $N(\mu_X, \sigma_X^2)$- bzw. $N(\mu_Y, \sigma_Y^2)$-Verteilung, so gilt (genau) im Fall $\rho = 0$

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{für alle } x, y \in \mathbb{R} \ ,$$

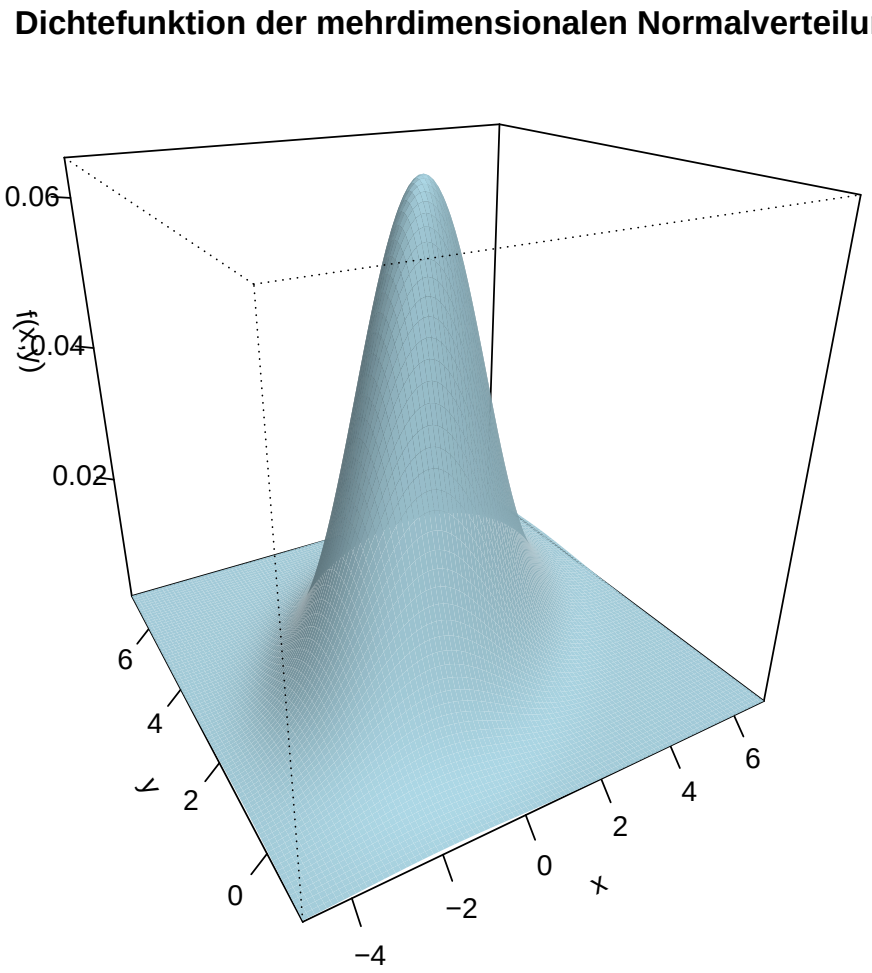
  also sind X und Y (genau) für $\rho = 0$ stochastisch unabhängig.
- Auch für $\rho \neq 0$ sind die bedingten Verteilungen von $X|Y = y$ und $Y|X = x$ wieder Normalverteilungen, es gilt genauer:

$$X|Y = y \quad \sim \quad N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right)$$

  bzw.

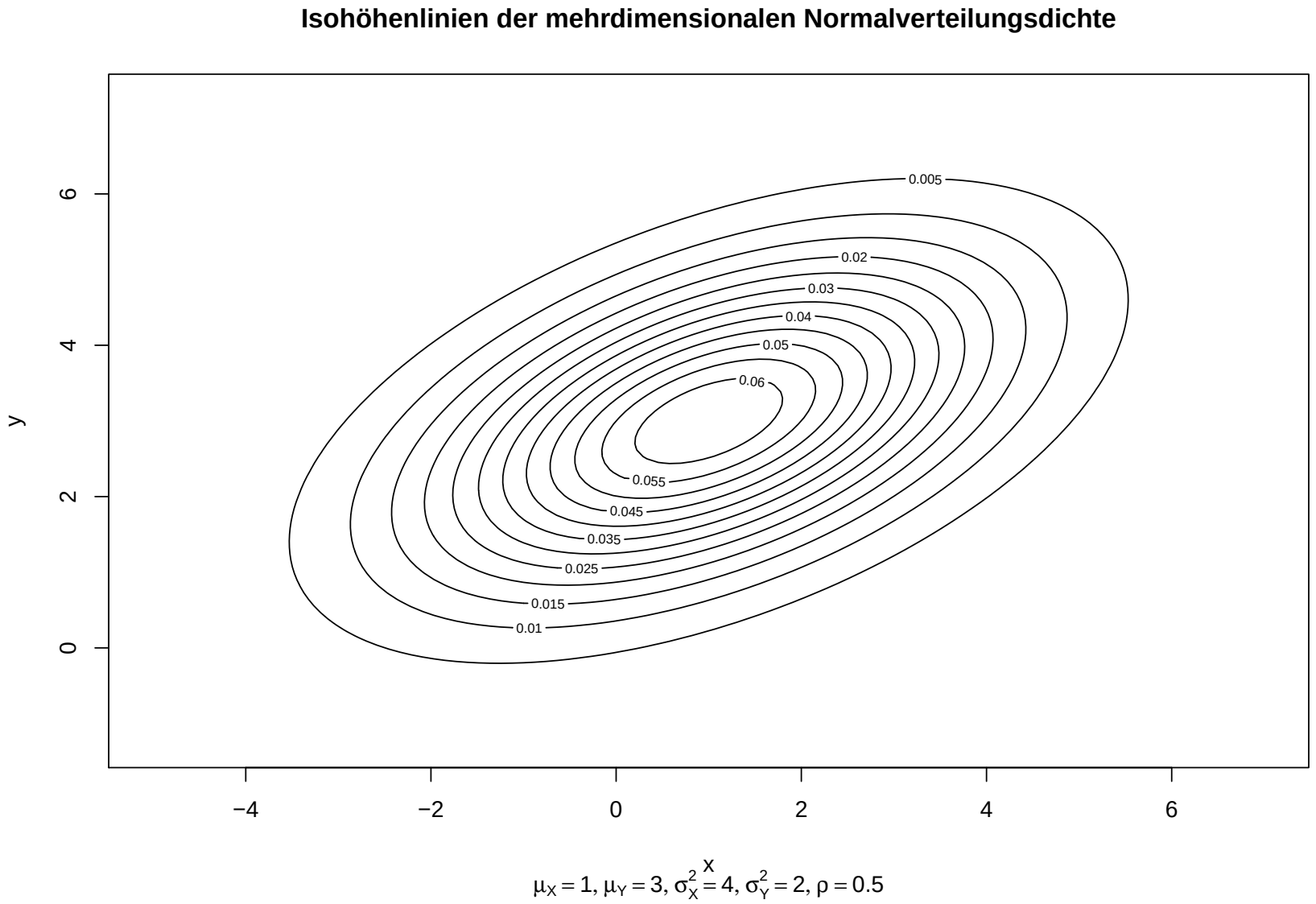
$$Y|X = x \quad \sim \quad N\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right)$$

Beispiel: Zweidimensionale Normalverteilung III

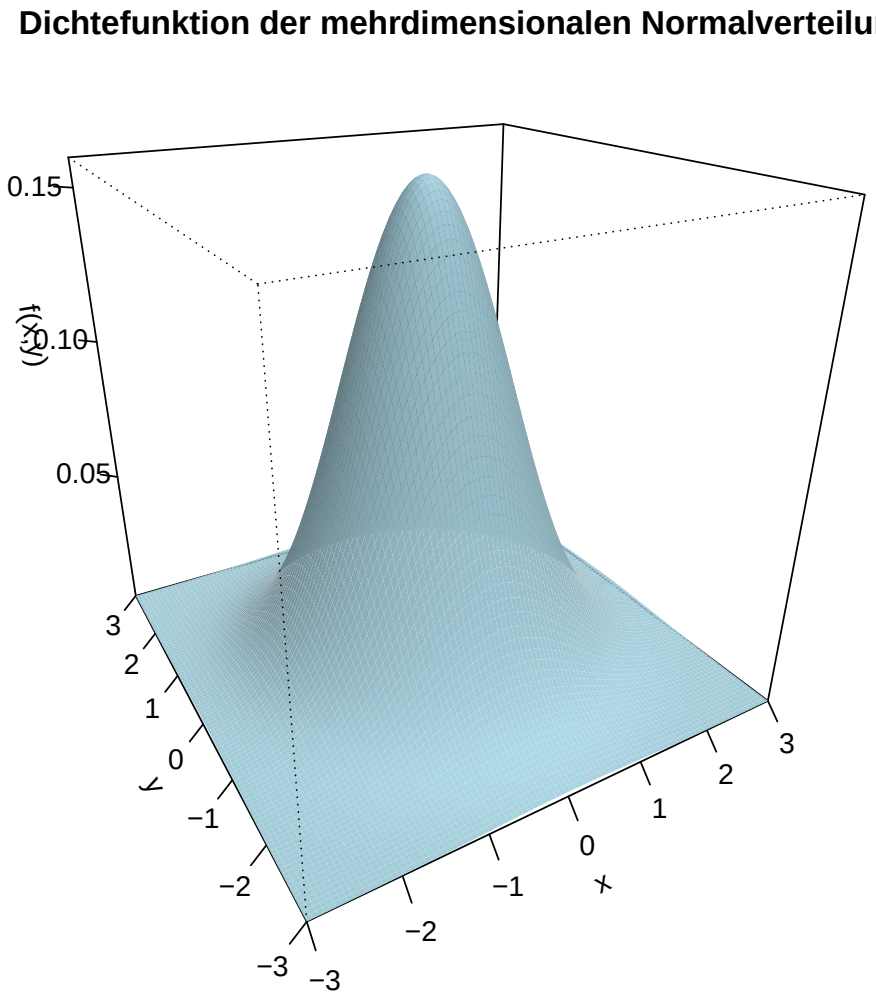


$\mu_X = 1$, $\mu_Y = 3$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 2$, $\rho = 0.5$

Beispiel: Zweidimensionale Normalverteilung IV

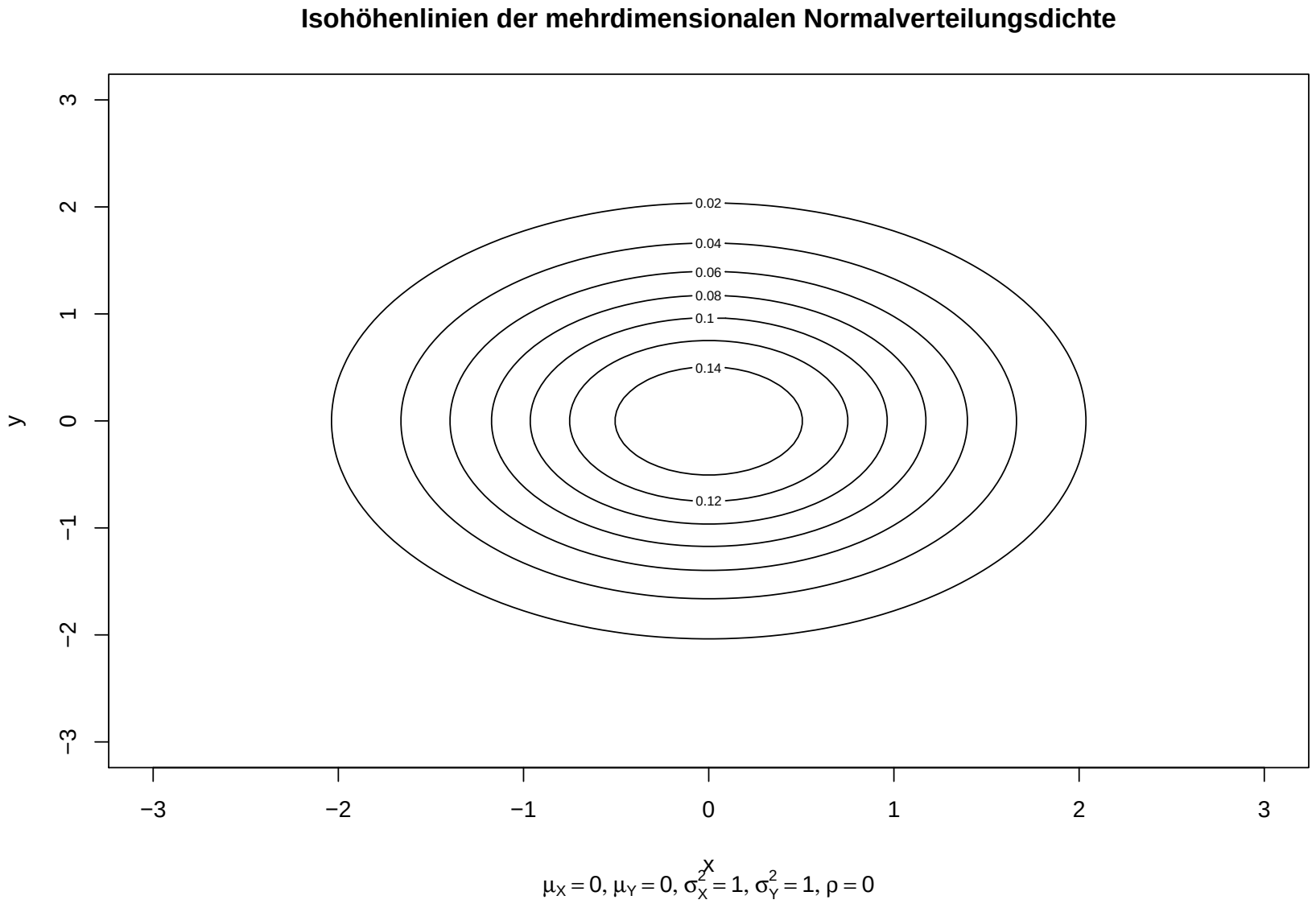


Beispiel: Zweidimensionale Normalverteilung V

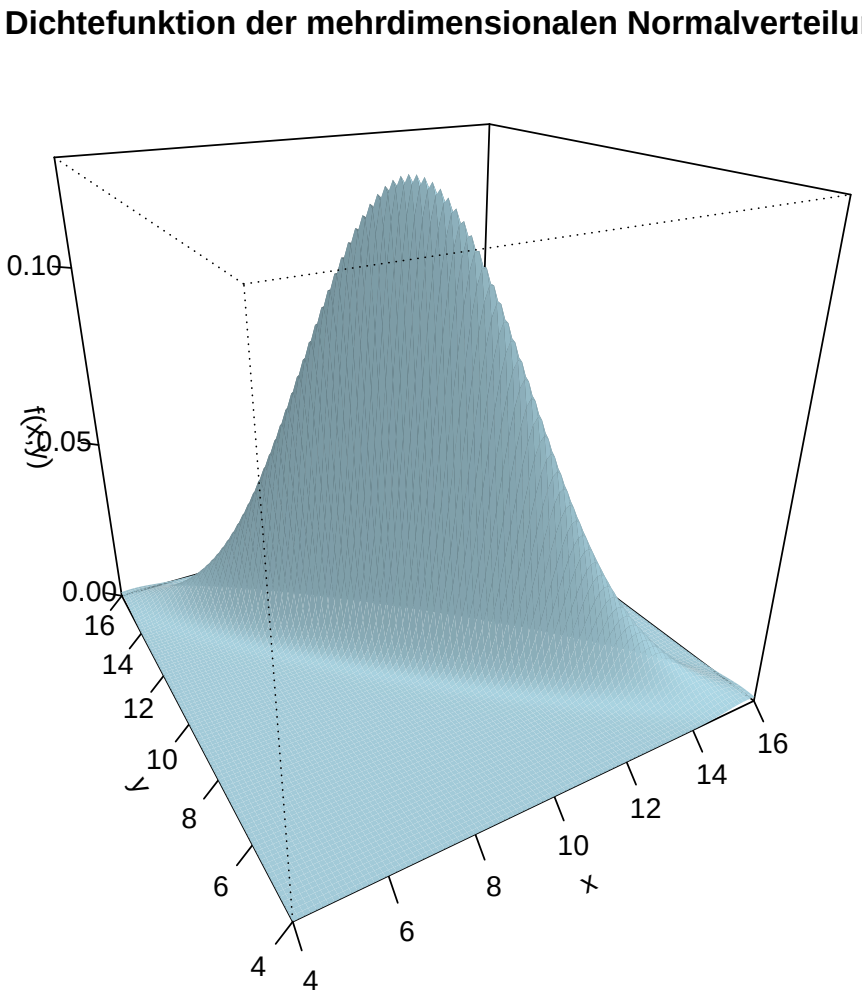


$\mu_X = 0$, $\mu_Y = 0$, $\sigma_X^2 = 1$, $\sigma_Y^2 = 1$, $\rho = 0$

Beispiel: Zweidimensionale Normalverteilung VI

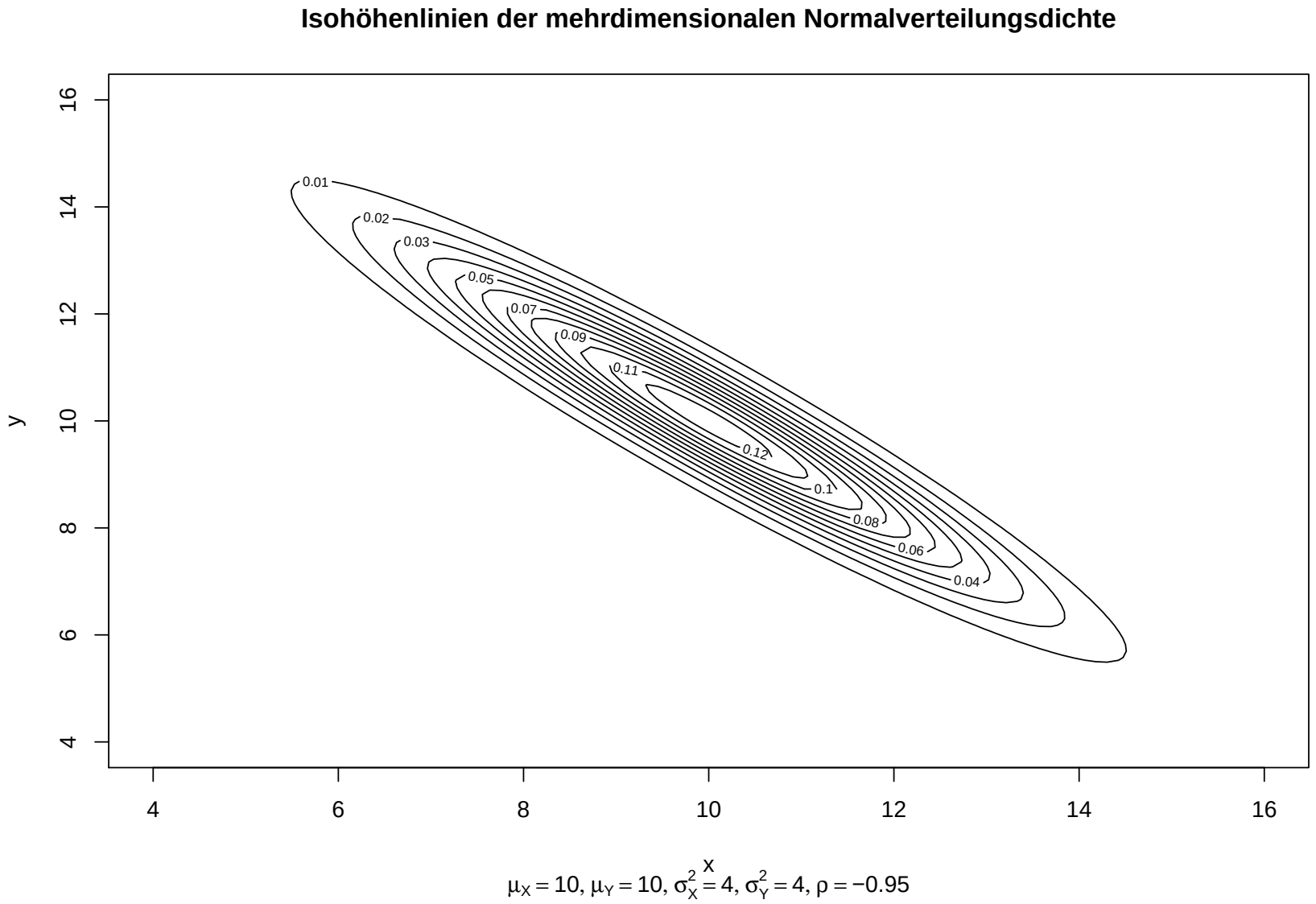


Beispiel: Zweidimensionale Normalverteilung VII



$\mu_X = 10$, $\mu_Y = 10$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 4$, $\rho = -0.95$

Beispiel: Zweidimensionale Normalverteilung VIII



Erwartungswert von $G(X, Y)$

Definition 10.7

Es seien (X, Y) eine zweidimensionale Zufallsvariable und $G : \mathbb{R}^2 \to \mathbb{R}$ eine $\mathcal{B}^2 - \mathcal{B}-$messbare Abbildung.

- Ist (X, Y) diskreter Zufallsvektor, sind (x_i, y_j) die Trägerpunkte sowie $p_{(X,Y)}$ die Wahrscheinlichkeitsfunktion von (X, Y) und gilt $\sum_{x_i} \sum_{y_j} |G(x_i, y_j)| \cdot p_{(X,Y)}(x_i, y_j) < \infty$, dann existiert der Erwartungswert $\mathrm{E}(G(X, Y))$ und es gilt

$$\mathrm{E}(G(X, Y)) = \sum_{x_i} \sum_{y_j} G(x_i, y_j) \cdot p_{(X,Y)}(x_i, y_j).$$

- Ist (X, Y) stetiger Zufallsvektor mit Dichtefunktion $f_{(X,Y)}$ und gilt $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |G(x, y)| \cdot f_{(X,Y)}(x, y) dy dx < \infty$, dann existiert der Erwartungswert $\mathrm{E}(G(X, Y))$ und es gilt

$$\mathrm{E}(G(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x, y) \cdot f_{(X,Y)}(x, y) dy dx.$$

- *Beispiel 1:* $G(X,Y) = a \cdot X + b \cdot Y + c$ für $a, b, c \in \mathbb{R}$
  Wegen der Linearität von Summenbildung und Integration gilt stets:

$$\mathrm{E}(a \cdot X + b \cdot Y + c) = a \cdot \mathrm{E}(X) + b \cdot \mathrm{E}(Y) + c$$

- *Beispiel 2:* $G(X,Y) = X \cdot Y$
  Hier gilt im allgemeinen **nicht** $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$!
  Sind X und Y allerdings **stochastisch unabhängig**, so gilt
  - im diskreten Fall wegen $p_{(X,Y)}(x,y) = p_X(x) \cdot p_Y(y)$ insgesamt

$$\begin{aligned} \mathrm{E}(X \cdot Y) &= \sum_{x_i} \sum_{y_j} x_i \cdot y_j \cdot p_X(x_i) \cdot p_Y(y_j) \\ &= \sum_{x_i} x_i \cdot p_X(x_i) \cdot \sum_{y_j} y_j \cdot p_Y(y_j) \end{aligned}$$

  - und im stetigen Fall wegen $f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y)$ insgesamt

$$\begin{aligned} \mathrm{E}(X \cdot Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f_X(x) \cdot f_Y(y) dy dx \\ &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \cdot \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy \ , \end{aligned}$$

  womit man **in diesem Fall** speziell $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$ erhält.

Abhängigkeitsmaße

- Analog zur deskriptiven Statistik ist man an Maßzahlen für die Abhängigkeit von zwei Zufallsvariablen X und Y über demselben Wahrscheinlichkeitsraum interessiert.
- Bei stochastischer Unabhängigkeit von X und Y sollten diese Maßzahlen naheliegenderweise den Wert 0 annehmen.
- Wie in deskriptiver Statistik: Maß für **lineare** Abhängigkeit von X und Y vordergründig:

Definition 10.8 (Kovarianz)

Es sei (X, Y) ein zweidimensionaler Zufallsvektor. Den Erwartungswert

$$\sigma_{XY} := \mathrm{Cov}(X, Y) := \mathrm{E}\left[(X - \mathrm{E}(X)) \cdot (Y - \mathrm{E}(Y))\right]$$

nennt man (im Falle der Existenz) **Kovarianz** zwischen X und Y.

- Eine dem Varianzzerlegungssatz ähnliche Rechenvorschrift zeigt man auch leicht für die Berechnung der Kovarianz:

Rechenregeln für Kovarianzen

Satz 10.9 (Kovarianzzerlegungssatz)

Ist (X, Y) ein zweidimensionaler Zufallsvektor, so gilt (im Fall der Existenz der beteiligten Erwartungswerte)

$$\mathrm{Cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y)$$

Sind X, Y und Z Zufallsvariablen (über demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) und $a, b \in \mathbb{R}$, so gelten außerdem die folgenden Rechenregeln:

1. $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y)$
2. $\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X, Y)$
   (Translationsinvarianz)
3. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
   (Symmetrie)
4. $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(Z, Y)$
5. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
6. X, Y stochastisch unabhängig $\Rightarrow \mathrm{Cov}(X, Y) = 0$

- Die erreichbaren Werte der Größe $\mathrm{Cov}(X, Y)$ hängen nicht nur von der Stärke der linearen Abhängigkeit ab, sondern (wie insbesondere aus Rechenregel 1 von Folie 284 ersichtlich) auch von der Streuung von X bzw. Y.
- Wie in deskriptiver Statistik: Alternatives Abhängigkeitsmaß mit normiertem „Wertebereich“, welches invariant gegenüber Skalierung von X bzw. Y ist.
- Hierzu Standardisierung der Kovarianz über Division durch Standardabweichungen von X und Y (falls möglich!):

Definition 10.10

Es sei (X, Y) ein zweidimensionaler Zufallsvektor mit $\mathrm{Var}(X) = \sigma_X^2 > 0$ und $\mathrm{Var}(Y) = \sigma_Y^2 > 0$. Man nennt

$$\rho_{XY} := \mathrm{Korr}(X, Y) := \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{Cov}(X, Y)}{+\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

den **Korrelationskoeffizienten** (nach Bravais-Pearson) zwischen X und Y.

Rechenregeln für Korrelationskoeffizienten

Sind X und Y Zufallsvariablen (über demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) mit $\mathrm{Var}(X) > 0, \mathrm{Var}(Y) > 0$ und $a, b \in \mathbb{R}$, so gilt:

1. $\mathrm{Korr}(aX, bY) = \begin{cases} \mathrm{Korr}(X, Y) & \text{falls } a \cdot b > 0 \\ -\mathrm{Korr}(X, Y) & \text{falls } a \cdot b < 0 \end{cases}$
2. $\mathrm{Korr}(X + a, Y + b) = \mathrm{Korr}(X, Y)$
   (Translationsinvarianz)
3. $\mathrm{Korr}(X, Y) = \mathrm{Korr}(Y, X)$
   (Symmetrie)
4. $-1 \leq \mathrm{Korr}(X, Y) \leq 1$
5. $\mathrm{Korr}(X, X) = 1$
6. $\left.\begin{array}{rl} \mathrm{Korr}(X, Y) = & 1 \\ \mathrm{Korr}(X, Y) = & -1 \end{array}\right\}$ genau dann, wenn $Y = aX + b$ mit $\left\{\begin{array}{l} a > 0 \\ a < 0 \end{array}\right.$
7. X, Y stochastisch unabhängig $\Rightarrow \mathrm{Korr}(X, Y) = 0$

Zufallsvariablen X, Y mit $\mathrm{Cov}(X, Y) = 0$ (!) heißen **unkorreliert**.

Zusammenhang Unkorreliertheit und Unabhängigkeit

- Offensichtlich gilt stets (falls die Kovarianz existiert):

$$X, Y \text{ stochastisch unabhängig} \quad \Rightarrow \quad X, Y \text{ unkorreliert}$$

- Die Umkehrung ist allerdings *im allgemeinen* falsch, es gilt **außer in speziellen Ausnahmefällen**:

$$X, Y \text{ unkorreliert} \quad \not\Rightarrow \quad X, Y \text{ stochastisch unabhängig}$$

- Einer dieser Ausnahmefälle ist die bivariate Normalverteilung:
  Ist der Zufallsvektor (X, Y) zweidimensional normalverteilt, so gilt **in dieser Situation** nämlich $\text{Korr}(X, Y) = \rho$ und damit (siehe Folie 274):

$$X, Y \text{ unkorreliert} \quad \Leftrightarrow \quad X, Y \text{ stochastisch unabhängig}$$

Varianzen von Summen zweier Zufallsvariablen

- Durch Verknüpfung verschiedener Rechenregeln aus Folie 284 lässt sich leicht zeigen, dass stets

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

  bzw. für $a, b, c \in \mathbb{R}$ allgemeiner stets

$$\operatorname{Var}(aX+bY+c) = a^2\operatorname{Var}(X) + 2ab\operatorname{Cov}(X,Y) + b^2\operatorname{Var}(Y)$$

  gilt.
- **Nur für unkorrelierte** (also insbesondere auch für stochastisch unabhängige) Zufallsvariablen X und Y gilt offensichtlich spezieller

$$\operatorname{Var}(aX+bY+c) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y)\ .$$

  Dies kann für mehr als zwei Zufallsvariablen weiter verallgemeinert werden.

Momente höherdimensionaler Zufallsvektoren

Definition 10.11

Seien $n \in \mathbb{N}$ und $\mathbf{X} = (X_1, \ldots, X_n)'$ ein n-dimensionaler Zufallsvektor. Dann heißt der n-dimensionale Vektor

$$\mathrm{E}(\mathbf{X}) := [\mathrm{E}(X_1), \ldots, \mathrm{E}(X_n)]' = \begin{bmatrix} \mathrm{E}(X_1) \\ \vdots \\ \mathrm{E}(X_n) \end{bmatrix}$$

Erwartungswertvektor von $\mathbf{X}$ und die $n \times n$-Matrix

$$\begin{aligned} \mathrm{V}(\mathbf{X}) &:= \mathrm{E}\left[(\mathbf{X} - \mathrm{E}(\mathbf{X})) \cdot (\mathbf{X} - \mathrm{E}(\mathbf{X}))'\right] \\ &:= \begin{bmatrix} \mathrm{E}[(X_1 - \mathrm{E}(X_1)) \cdot (X_1 - \mathrm{E}(X_1))] & \cdots & \mathrm{E}[(X_1 - \mathrm{E}(X_1)) \cdot (X_n - \mathrm{E}(X_n))] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mathrm{E}(X_n)) \cdot (X_1 - \mathrm{E}(X_1))] & \cdots & \mathrm{E}[(X_n - \mathrm{E}(X_n)) \cdot (X_n - \mathrm{E}(X_n))] \end{bmatrix} \end{aligned}$$

(Varianz-)Kovarianzmatrix von $\mathbf{X}$.

Es gilt $\mathrm{V}(\mathbf{X})$

$$= \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_{n-1}) & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \cdots & \mathrm{Cov}(X_2, X_{n-1}) & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{Cov}(X_{n-1}, X_1) & \mathrm{Cov}(X_{n-1}, X_2) & \cdots & \mathrm{Cov}(X_{n-1}, X_{n-1}) & \mathrm{Cov}(X_{n-1}, X_n) \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Cov}(X_n, X_{n-1}) & \mathrm{Cov}(X_n, X_n) \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_{n-1}) & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_{n-1}) & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{Cov}(X_{n-1}, X_1) & \mathrm{Cov}(X_{n-1}, X_2) & \cdots & \mathrm{Var}(X_{n-1}) & \mathrm{Cov}(X_{n-1}, X_n) \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Cov}(X_n, X_{n-1}) & \mathrm{Var}(X_n) \end{bmatrix}$$

- Der Eintrag v_{ij} der i-ten Zeile und j-ten Spalte von $\mathrm{V}(\mathbf{X})$ ist also gegeben durch $\mathrm{Cov}(X_i, X_j)$.
- $\mathrm{V}(\mathbf{X})$ ist wegen $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$ offensichtlich symmetrisch.
- Für $n = 1$ erhält man die „klassische“ Definition der Varianz, das heißt die Matrix $\mathrm{V}(\mathbf{X})$ wird als 1×1-Matrix skalar und stimmt mit $\mathrm{Var}(X_1)$ überein.

- Ein n-dimensionaler Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)'$ bzw. die n eindimensionalen Zufallsvariablen $X_1, \ldots, X_n$ heißen **unkorreliert**, wenn $\mathrm{V}(\mathbf{X})$ eine Diagonalmatrix ist, also höchstens die Diagonaleinträge von 0 verschieden sind.
- Es gilt in Verallgemeinerung von Folie 287 (bei Existenz der Momente):

$$X_1, \ldots, X_n \text{ stochastisch unabhängig} \quad \Rightarrow \quad X_1, \ldots, X_n \text{ unkorreliert},$$

  die Umkehrung gilt *im allgemeinen* nicht!
- Dass stochastische Unabhängigkeit eine stärkere Eigenschaft ist als Unkorreliertheit, zeigt auch der folgende Satz:

Satz 10.12

Seien für $n \in \mathbb{N}$ stochastisch unabhängige Zufallsvariablen $X_1, \ldots, X_n$ über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ sowie $\mathcal{B} - \mathcal{B}-$messbare Funktionen $G_i : \mathbb{R} \to \mathbb{R}$ für $i \in \{1, \ldots, n\}$ gegeben.
Dann sind auch die Zufallsvariablen $G_1(X_1), \ldots, G_n(X_n)$ stochastisch unabhängig.

Momente von Summen von Zufallsvariablen

- Regelmäßig ist man für $n \in \mathbb{N}$ an der Verteilung bzw. an Maßzahlen der Summe $\sum_{i=1}^{n} X_i = X_1 + \cdots + X_n$ oder einer gewichteten Summe

$$\sum_{i=1}^{n} w_i \cdot X_i = w_1 \cdot X_1 + \cdots + w_n \cdot X_n \qquad (w_1, \ldots, w_n \in \mathbb{R})$$

  von n Zufallsvariablen $X_1, \ldots, X_n$ interessiert.
- In Verallgemeinerung von Folie 282 zeigt man für den Erwartungswert leicht

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) \qquad \text{bzw.} \qquad \mathrm{E}\left(\sum_{i=1}^{n} w_i \cdot X_i\right) = \sum_{i=1}^{n} w_i \cdot \mathrm{E}(X_i)\ .$$

- Insbesondere gilt für das (arithmetische) Mittel aus $X_1, \ldots, X_n$

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i)\ .$$

- In Verallgemeinerung von Folie 288 erhält man für die Varianz weiter

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Cov}(X_i, X_j) &= \sum_{i=1}^{n}\operatorname{Var}(X_i) + \sum_{\substack{i,j\in\{1,\ldots,n\}\\ i\neq j}}\operatorname{Cov}(X_i, X_j)\\ &= \sum_{i=1}^{n}\operatorname{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{Cov}(X_i, X_j)\end{aligned}$$

bzw.

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{n} w_i \cdot X_i\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i \cdot w_j \cdot \operatorname{Cov}(X_i, X_j)\\ &= \sum_{i=1}^{n} w_i^2 \cdot \operatorname{Var}(X_i) + \sum_{\substack{i,j\in\{1,\ldots,n\}\\ i\neq j}} w_i \cdot w_j \cdot \operatorname{Cov}(X_i, X_j)\\ &= \sum_{i=1}^{n} w_i^2 \cdot \operatorname{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_i \cdot w_j \cdot \operatorname{Cov}(X_i, X_j)\ .\end{aligned}$$

- Sind $X_1, \ldots, X_n$ unkorreliert, so vereinfacht sich die Varianz der Summe wegen $\mathrm{Cov}(X_i, X_j) = 0$ für $i \neq j$ offensichtlich zu

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

bzw.

$$\mathrm{Var}\left(\sum_{i=1}^{n} w_i \cdot X_i\right) = \sum_{i=1}^{n} w_i^2 \cdot \mathrm{Var}(X_i) \ .$$

- Fasst man die Zufallsvariablen $X_1, \ldots, X_n$ im n-dimensionalen Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)$ und die Gewichte $w_1, \ldots, w_n$ im Vektor $\mathbf{w} = (w_1, \ldots, w_n)' \in \mathbb{R}^n$ zusammen, so lassen sich die Ergebnisse kürzer darstellen als

$$\mathrm{E}\left(\sum_{i=1}^{n} w_i \cdot X_i\right) = \mathbf{w}' \, \mathrm{E}(\mathbf{X}) \qquad \text{bzw.} \qquad \mathrm{Var}\left(\sum_{i=1}^{n} w_i \cdot X_i\right) = \mathbf{w}' \, \mathrm{V}(\mathbf{X}) \mathbf{w} \ .$$

Summen von Zufallsvariablen spezieller Verteilungen

- Sind die Zufallsvariablen $X_1, \ldots, X_n$ nicht nur unkorreliert, sondern sogar unabhängig, dann sind einige der erläuterten Verteilungsfamilien „abgeschlossen“ gegenüber Summenbildungen.
- Besitzen darüberhinaus alle n Zufallsvariablen $X_1, \ldots, X_n$ exakt dieselbe Verteilung, spricht man von unabhängig identisch verteilten Zufallsvariablen gemäß folgender Definition:

Definition 11.1 (i.i.d. Zufallsvariablen)

Seien für $n \in \mathbb{N}$ Zufallsvariablen $X_1, \ldots, X_n$ über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ gegeben mit

- $Q_X := P_{X_1} = P_{X_2} = \cdots = P_{X_n}$,
- $X_1, \ldots, X_n$ stochastisch unabhängig.

Dann heißen die Zufallsvariablen $X_1, \ldots, X_n$ **unabhängig identisch verteilt (independent and identically distributed)** gemäß Q_X, in Zeichen: $X_i \overset{i.i.d.}{\sim} Q_X$ oder $X_i \overset{u.i.v.}{\sim} Q_X$ für $i \in \{1, \ldots, n\}$.

Satz 11.2 (Summen spezieller Zufallsvariablen)

Seien $n \in \mathbb{N}$, $X_1, \ldots, X_n$ *stochastisch unabhängige Zufallsvariablen und*

$$Y := X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i \ .$$

Gilt für $i \in \{1, \ldots, n\}$ *weiterhin*

1. $X_i \sim B(1,p)$ *für ein* $p \in (0,1)$, *also insgesamt* $X_i \overset{i.i.d.}{\sim} B(1,p)$, *so gilt* $Y \sim B(n,p)$ *(vgl. Folie 232),*
2. $X_i \sim B(n_i,p)$ *für* $n_i \in \mathbb{N}$ *und ein* $p \in (0,1)$, *so gilt* $Y \sim B(N,p)$ *mit* $N := n_1 + \cdots + n_n = \sum_{i=1}^{n} n_i$,
3. $X_i \sim \mathrm{Pois}(\lambda_i)$ *für* $\lambda_i \in \mathbb{R}_+$, *so gilt* $Y \sim \mathrm{Pois}(\lambda)$ *mit* $\lambda := \lambda_1 + \cdots + \lambda_n = \sum_{i=1}^{n} \lambda_i$,
4. $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ *für* $\mu_i \in \mathbb{R}$ *und* $\sigma_i^2 > 0$, *so gilt* $Y \sim \mathrm{N}(\mu, \sigma^2)$ *mit* $\mu = \mu_1 + \cdots + \mu_n = \sum_{i=1}^{n} \mu_i$ *und* $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 = \sum_{i=1}^{n} \sigma_i^2$,
5. $X_i \sim \mathrm{N}(\mu, \sigma^2)$ *für ein* $\mu \in \mathbb{R}$ *und ein* $\sigma^2 > 0$, *also insgesamt* $X_i \overset{i.i.d.}{\sim} \mathrm{N}(\mu, \sigma^2)$, *so gilt für* $\overline{X} := \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ *insbesondere* $\overline{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$.

Der zentrale Grenzwertsatz

- Die Verteilung von Summen (insbesondere) unabhängig identisch verteilter Zufallsvariablen weist ein spezielles Grenzverhalten auf, wenn die Anzahl der summierten Zufallsvariablen wächst.
- Dieses Grenzverhalten ist wesentlicher Grund für großes Anwendungspotenzial der (Standard-)Normalverteilung.
- Betrachte im Folgenden für $i \in \mathbb{N}$ unabhängig identisch verteilte Zufallsvariablen X_i mit $\mathrm{E}(X_i) = \mu_X \in \mathbb{R}$ und $\mathrm{Var}(X_i) = \sigma_X^2 > 0$.
- Gilt $n \to \infty$ für die Anzahl n der summierten Zufallsvariablen X_i, gilt für die Summen $Y_n := \sum_{i=1}^n X_i$ offensichtlich $\mathrm{Var}(Y_n) = n \cdot \sigma_X^2 \to \infty$ und für $\mu_X \neq 0$ auch $\mathrm{E}(Y_n) = n \cdot \mu_X \to \pm\infty$, eine Untersuchung der Verteilung von Y_n für $n \to \infty$ ist also schwierig.
- Stattdessen: Betrachtung der **standardisierten** Zufallsvariablen

$$Z_n := \frac{Y_n - \mathrm{E}(Y_n)}{\sqrt{\mathrm{Var}(Y_n)}} = \frac{\left(\sum_{i=1}^n X_i\right) - n\mu_X}{\sigma_X \sqrt{n}} = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \mu_X}{\sigma_X}\sqrt{n}$$

  mit $\mathrm{E}(Z_n) = 0$ und $\mathrm{Var}(Z_n) = 1$ für alle $n \in \mathbb{N}$.
- Gemäß folgendem Satz „konvergiert“ die Verteilung der Z_n für $n \to \infty$ gegen eine Standardnormalverteilung:

Satz 11.3 (Zentraler Grenzwertsatz)

Es seien X_i für $i \in \mathbb{N}$ unabhängig identisch verteilte Zufallsvariablen mit $\mathrm{E}(X_i) = \mu_X \in \mathbb{R}$ und $\mathrm{Var}(X_i) = \sigma_X^2 > 0$. Für $n \in \mathbb{N}$ seien die Summen $Y_n := \sum_{i=1}^n X_i$ bzw. die standardisierten Summen bzw. Mittelwerte

$$Z_n := \frac{Y_n - \mathrm{E}(Y_n)}{\sqrt{\mathrm{Var}(Y_n)}} = \frac{\left(\sum_{i=1}^n X_i\right) - n\mu_X}{\sigma_X \sqrt{n}} = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \mu_X}{\sigma_X}\sqrt{n}$$

von $X_1, \ldots, X_n$ definiert.
Dann gilt für die Verteilungsfunktionen F_{Z_n} der Zufallsvariablen Z_n

$$\lim_{n\to\infty} F_{Z_n}(z) = F_{\mathrm{N}(0,1)}(z) = \Phi(z) \quad \text{für alle } z \in \mathbb{R}\ ,$$

man sagt auch, die Folge Z_n von Zufallsvariablen konvergiere in Verteilung *gegen die Standardnormalverteilung, in Zeichen $Z_n \overset{n\to\infty}{\longrightarrow} \mathrm{N}(0,1)$.*

- Mit Hilfe des zentralen Grenzwertsatzes lassen sich insbesondere
  - weitere (schwächere oder speziellere) Grenzwertsätze zeigen,
  - Wahrscheinlichkeiten von Summen von i.i.d. Zufallsvariablen näherungsweise mit Hilfe der Standardnormalverteilung auswerten.

Das schwache Gesetz der großen Zahlen

- Eine direkte Folge von Satz 11.3 ist zum Beispiel das folgende Resultat:

Satz 11.4 (Schwaches Gesetz der großen Zahlen)

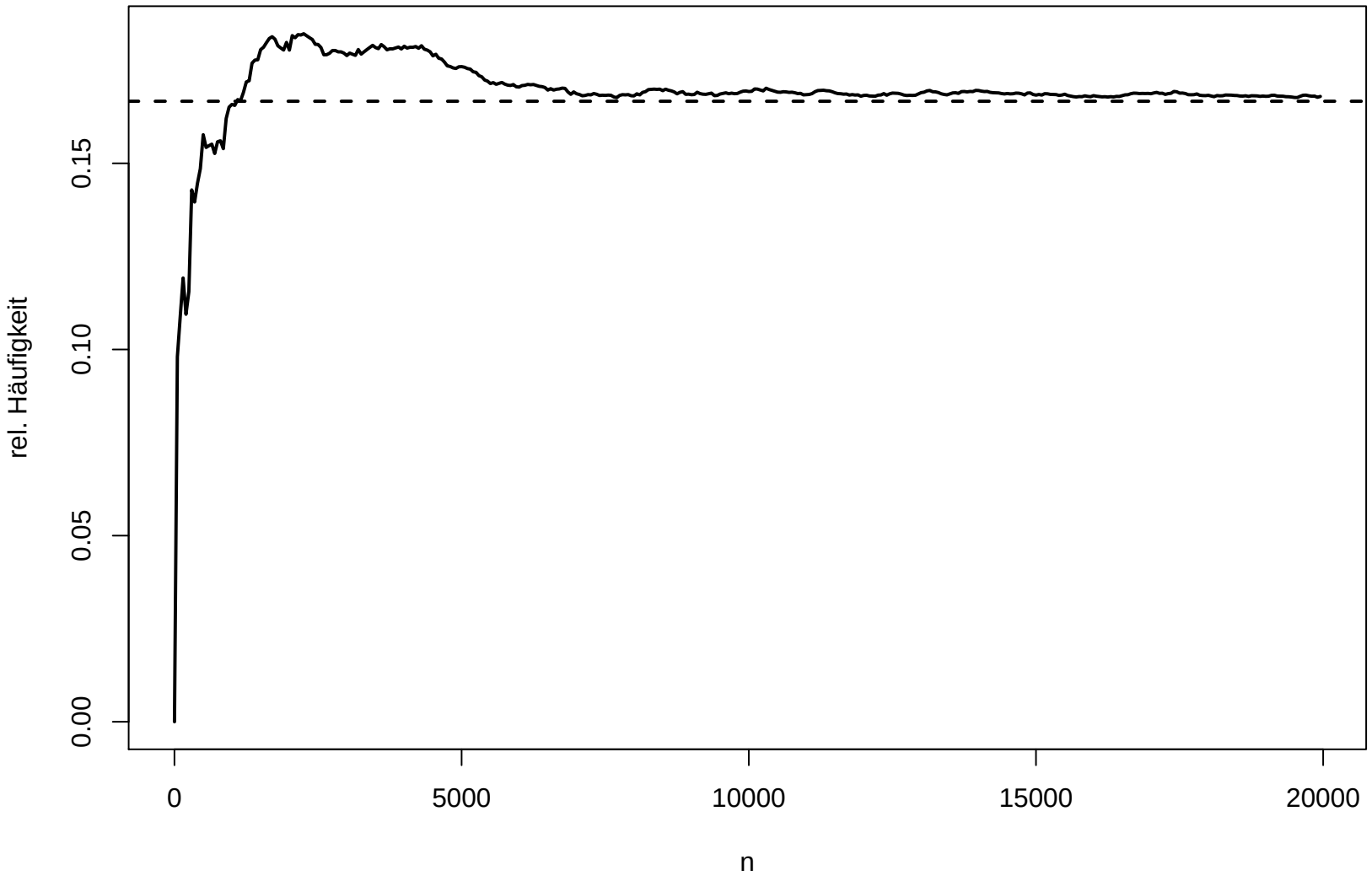
Es seien X_i für $i \in \mathbb{N}$ unabhängig identisch verteilte Zufallsvariablen mit $\mathrm{E}(X_i) = \mu_X \in \mathbb{R}$. Dann gilt:

$$\lim_{n\to\infty} P\left\{\left|\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu_X\right| \geq \varepsilon\right\} = 0 \qquad \textit{für alle } \varepsilon > 0$$

- Das schwache Gesetz der großen Zahlen besagt also, dass die Wahrscheinlichkeit, mit der ein Mittelwert von n i.i.d. Zufallsvariablen betragsmäßig um mehr als eine vorgegebene (kleine) Konstante $\varepsilon > 0$ vom Erwartungswert der Zufallsvariablen abweicht, für $n \to \infty$ gegen Null geht.
- Insbesondere „konvergiert“ also die Folge der beobachteten relativen Häufigkeiten der Erfolge bei unabhängiger wiederholter Durchführung eines Bernoulli-Experiments gegen die Erfolgswahrscheinlichkeit p.
- Letztendlich ist dies auch eine Rechtfertigung für den häufigkeitsbasierten Wahrscheinlichkeitsbegriff!

Veranschaulichung „Schwaches Gesetz der großen Zahlen“

(relative Häufigkeit des Auftretens der Zahl 6 beim Würfeln)



Grenzwertsatz von de Moivre-Laplace

- *Erinnerung: (siehe Satz 11.2, Folie 296)*
  Summen i.i.d. bernoulliverteilter Zufallsvariablen sind binomialverteilt.
- Die Anwendung des zentralen Grenzwertsatzes 11.3 auf binomialverteilte Zufallsvariablen als Summen i.i.d. bernoulliverteilter Zufallsvariablen führt zur (zumindest historisch wichtigen und beliebten) Näherung von Binomialverteilungen durch Normalverteilungen.
- Resultat ist auch als eigenständiger Grenzwertsatz bekannt:

Satz 11.5 (Grenzwertsatz von de Moivre-Laplace)

Für $n \in \mathbb{N}$ und $p \in (0,1)$ sei $Y_n \sim B(n,p)$. Dann konvergieren die standardisierten Zufallsvariablen

$$Z_n := \frac{Y_n - np}{\sqrt{np(1-p)}}$$

in Verteilung gegen die Standardnormalverteilung, es gilt also

$$\lim_{n\to\infty} F_{Z_n}(z) = F_{\mathrm{N}(0,1)}(z) = \Phi(z) \quad \text{für alle } z \in \mathbb{R}\ .$$

Anwendung der Grenzwertsätze für Näherungen

- Grenzwertsätze treffen nur Aussagen für $n \to \infty$.
- In praktischen Anwendungen wird verwendet, dass diese Aussagen für endliche, aber hinreichend große n, schon „näherungsweise“ gelten.
- Eine häufige verwendete „Faustregel“ zur Anwendung des Grenzwertsatzes von de Moivre-Laplace ist zum Beispiel $np(1-p) \geq 9$.
- (Äquivalente!) Anwendungsmöglichkeit der Grenzwertsätze für endliches n: Verwendung
  - der $N(n\mu_X, n\sigma_X^2)$-Verteilung für $Y_n := \sum_{i=1}^{n} X_i$ oder
  - der $N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$-Verteilung für $\overline{X}_n := \frac{1}{n} Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ oder
  - der Standardnormalverteilung für $Z_n := \frac{Y_n - n\mu_X}{\sigma_X \sqrt{n}} = \frac{\overline{X}_n - \mu_X}{\sigma_X} \sqrt{n}$

  mit $\mu_X = \mathrm{E}(X_i)$ und $\sigma_X^2 = \mathrm{Var}(X_i)$ statt der jeweiligen exakten Verteilung.
- Gilt $X_i \overset{i.i.d.}{\sim} N(\mu_X, \sigma_X^2)$, stimmen die „Näherungen“ sogar mit den exakten Verteilungen überein (siehe Folie 296)!

Beispiele I

zur Anwendung des zentralen Grenzwertsatzes

Für $Y_n \sim B(n,p)$ erhält man beispielsweise die Näherung

$$P\{Y_n \leq z\} = F_{Y_n}(z) \approx F_{N(np,np(1-p))}(z) = \Phi\left(\frac{z-np}{\sqrt{np(1-p)}}\right)$$

oder gleichbedeutend

$$\begin{aligned} P\{Y_n \leq z\} &= P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq \frac{z-np}{\sqrt{np(1-p)}}\right\} = P\left\{Z_n \leq \frac{z-np}{\sqrt{np(1-p)}}\right\} \\ &\approx F_{N(0,1)}\left(\frac{z-np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{z-np}{\sqrt{np(1-p)}}\right) . \end{aligned}$$

Beispiele II

zur Anwendung des zentralen Grenzwertsatzes

Gilt $Y \sim B(5000, 0.2)$, so erhält man für die Wahrscheinlichkeit, dass Y Werte > 1000 und ≤ 1050 annimmt, näherungsweise

$$\begin{aligned} P\{1000 < Y \leq 1050\} &= F_Y(1050) - F_Y(1000) \\ &\approx \Phi\left(\frac{1050-5000\cdot 0.2}{\sqrt{5000\cdot 0.2\cdot(1-0.2)}}\right) - \Phi\left(\frac{1000-5000\cdot 0.2}{\sqrt{5000\cdot 0.2\cdot(1-0.2)}}\right) \\ &= \Phi(1.77) - \Phi(0) = 0.9616 - 0.5 = 0.4616 \end{aligned}$$

(Exakte Wahrscheinlichkeit: 0.4538*)*

Beispiele III

zur Anwendung des zentralen Grenzwertsatzes

Seien $X_i \overset{i.i.d}{\sim} \text{Pois}(5)$ für $i \in \{1, \ldots, 100\}$ (es gilt also $\text{E}(X_i) = 5$ und $\text{Var}(X_i) = 5$), sei $Y := \sum_{i=1}^{100} X_i$. Dann gilt für das untere Quartil $y_{0.25}$ von Y

$$F_Y(y_{0.25}) \approx F_{N(100\cdot 5, 100\cdot 5)}(y_{0.25}) = \Phi\left(\frac{y_{0.25}-500}{\sqrt{500}}\right) \overset{!}{=} 0.25$$

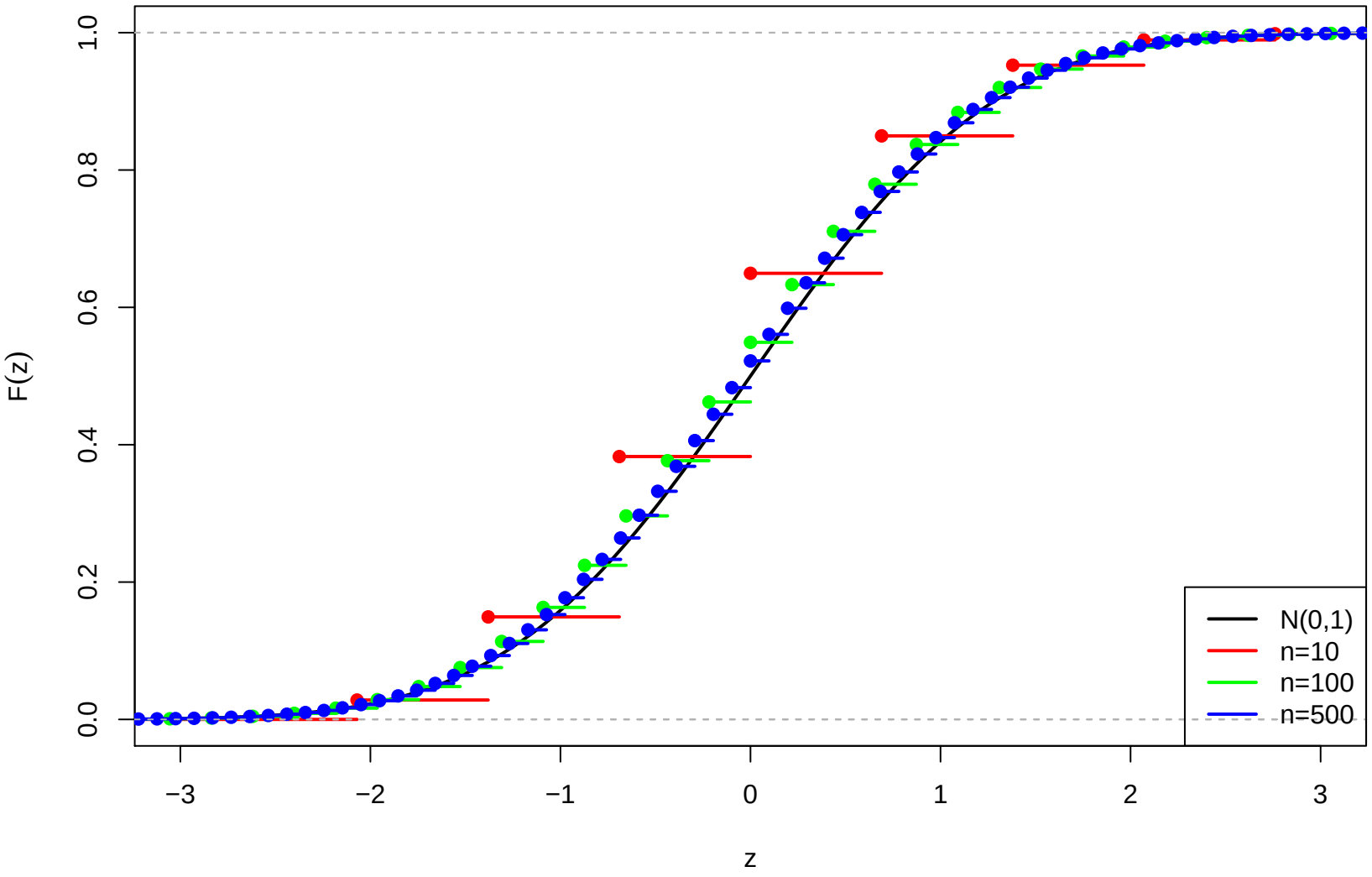
und unter Verwendung von $\Phi(z) = 1 - \Phi(-z)$ bzw. $\Phi(-z) = 1 - \Phi(z)$ weiter

$$\begin{aligned} \Phi\left(\frac{500-y_{0.25}}{\sqrt{500}}\right) &\overset{!}{=} 1 - 0.25 = 0.75 \quad \Rightarrow \quad \frac{500-y_{0.25}}{\sqrt{500}} = \Phi^{-1}(0.75) \approx 0.675 \\ \Rightarrow \quad y_{0.25} &\approx 500 - 0.675 \cdot \sqrt{500} = 484.9065 \end{aligned}$$

(exaktes unteres Quartil unter Verwendung von $Y \sim \text{Pois}(500)$: $y_{0.25} = 485$)

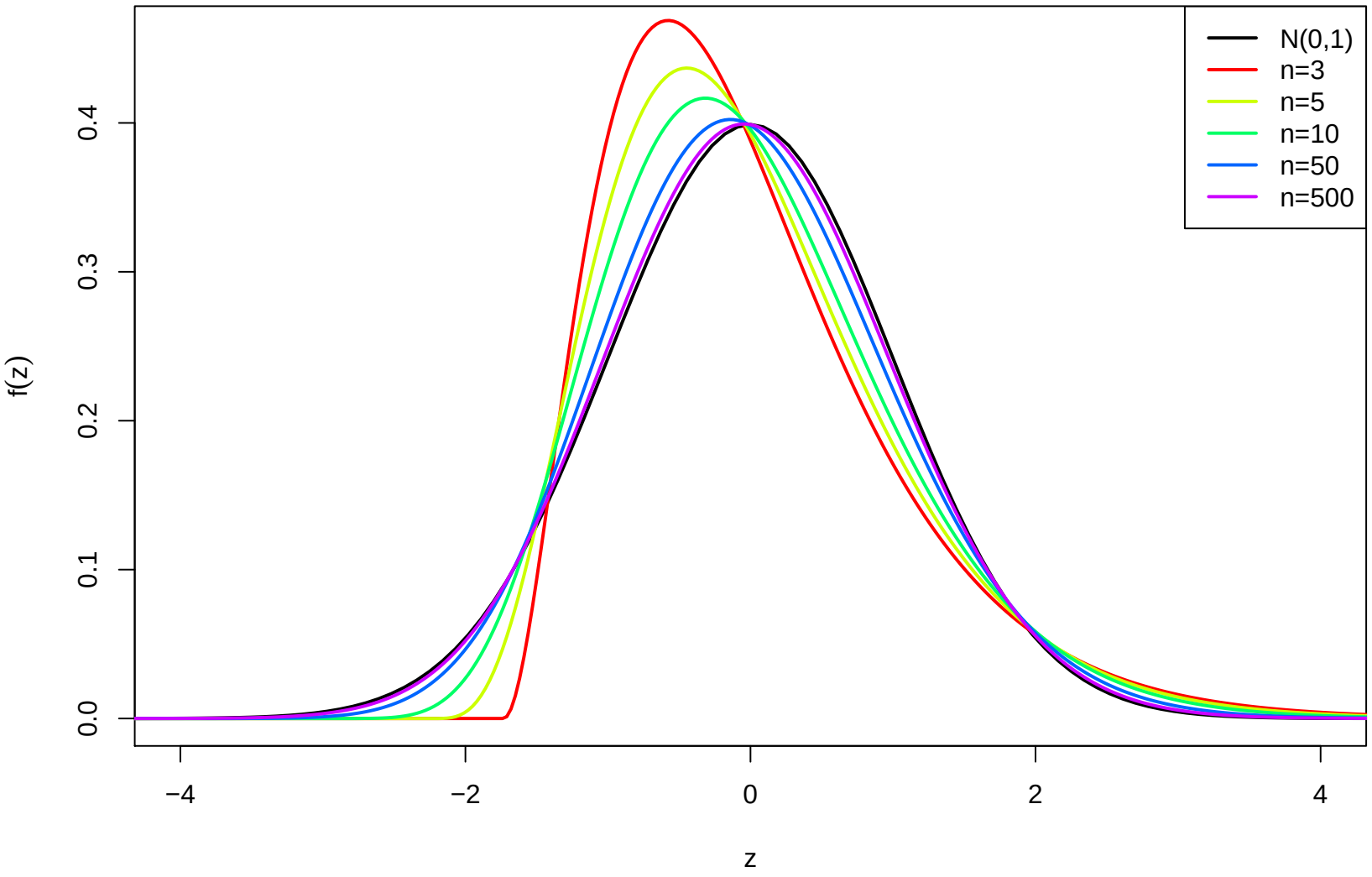
Veranschaulichung des zentralen Grenzwertsatzes

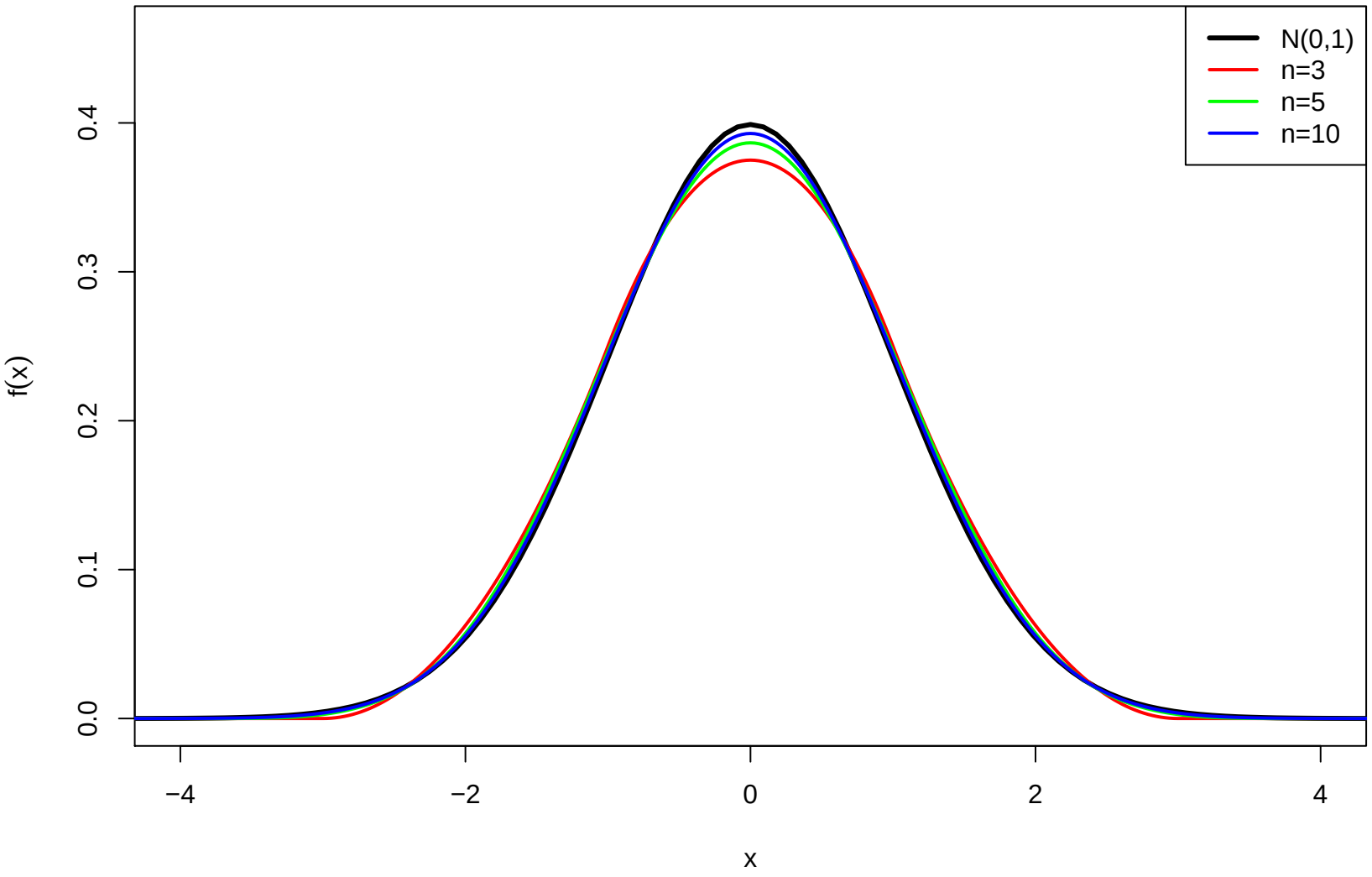
an der Verteilungsfunktion standardisierter Binomialverteilungen $B(n, p)$ mit $p = 0.3$



Veranschaulichung des zentralen Grenzwertsatzes

an der Dichtefunktion standardisierter Summen von Exponentialverteilungen mit $\lambda = 2$



Veranschaulichung des zentralen Grenzwertsatzes

an der Dichtefunktion standardisierter Summen von Unif$(20, 50)$-Verteilungen