

Stochastik für die Informatik

Wintersemester 2020/21

Anton Wakolbinger

Stofl-Webseite:

[https://www.math.uni-frankfurt.de/
~ismi/wakolbinger/teaching/Stofl2021/](https://www.math.uni-frankfurt.de/~ismi/wakolbinger/teaching/Stofl2021/)

Oder:

google: “Wakolbinger”

- Goethe-Universität – Homepage A. Wakolbinger
- Veranstaltungen WS 2020/21

Dort: Link zum

Moodle-Kurs “Stochastik für die Informatik”

Anmeldung zu einer **Übungsgruppe**:

elektronisch (← Stofl - Webseite)

bis Donnerstag 05.11.2020, 24 Uhr

nach dem first come - first serve Prinzip

Gruppenzeiten:

Di 14-16, Di 16-18,

Mi 12-14, 12-14

Do 12-14 , Do 12-14, Do 14-16

Fr 8-10, Fr 10-12, Fr 14-16

Seit gestern (02.11.) sind alle 10 Gruppen (à 40) voll.

Wenn Sie sich noch zu den Übungen anmelden wollen, dann schreiben Sie bitte eine E-Mail an den Übungs Koordinator: Herrn Florin Boenkost, boenkost@math.uni-frankfurt.de mit der Angabe von Ihnen möglichen Terminen.

Und bitte halten Sie Ihre Anmeldung nur dann aufrecht, wenn Sie wirklich an den Übungen teilnehmen wollen.

jeweils spätestens Dienstags:

Neues Übungsblatt im Moodle-Kurs

Tipps zu den Übungsaufgaben:

in den Tutorien

(und zu Blatt 1 auch schon in der Vorlesung diesen Freitag)

Termin für die **elektronische Abgabe**

der schriftlichen Lösungen der “S-Aufgaben”:

Freitags (10 Tage nach Ausgabe des Blattes)

Leitfaden zur elektronischen Abgabe: siehe Moodle-Kurs.

In der Woche nach der Abgabe
werden die **Lösungen in den Tutorien** besprochen.

Abschlussklausur: Dienstag, 16 Februar 2021

Zweitklausur: Donnerstag 08. April 2021.

Beide Termine sind derzeit noch unter Vorbehalt.

In der Klausur können 100 Klausurpunkte erreicht werden.

Die Note errechnet sich aus

der Summe der Anzahl der erreichten Klausurpunkte

plus der Anzahl der im Tutorium erreichten Bonuspunkte.

Abschlussklausur: Dienstag, 16 Februar 2021

Zweitklausur: Donnerstag 08. April 2021.

Beide Termine sind derzeit noch unter Vorbehalt.

In der Klausur können 100 Klausurpunkte erreicht werden.

Die Note errechnet sich aus

der Summe der Anzahl der erreichten Klausurpunkte

plus der Anzahl der im Tutorium erreichten Bonuspunkte.

Beträgt diese Summe mindestens 45,

gilt die Abschlussprüfung über die Veranstaltung
als bestanden.

Bonuspunkte (maximal 12):

durch aktive Beteiligung in den Tutorien.

Bonuspunkte bekommt man nur, wenn man
mindestes zweimal im Semester

Lösungen von Übungsaufgaben (oder Teile davon)
im Tutorium vorstellt,

und grundsätzlich nur für die Aufgaben, bei deren
Lösungsbesprechung man im Tutorium anwesend ist.

Lehrbuch:

Götz Kersting, Anton Wakolbinger

Elementare Stochastik, Birkhäuser, 2. Aufl. 2010,

Preis: 19,99 EUR

Semesterausleihe möglich aus der

Bibliothek des Mathematischen Seminars,

Robert-Mayer-Str. 8, 4. Stock

in der UB als E-Book vorhanden

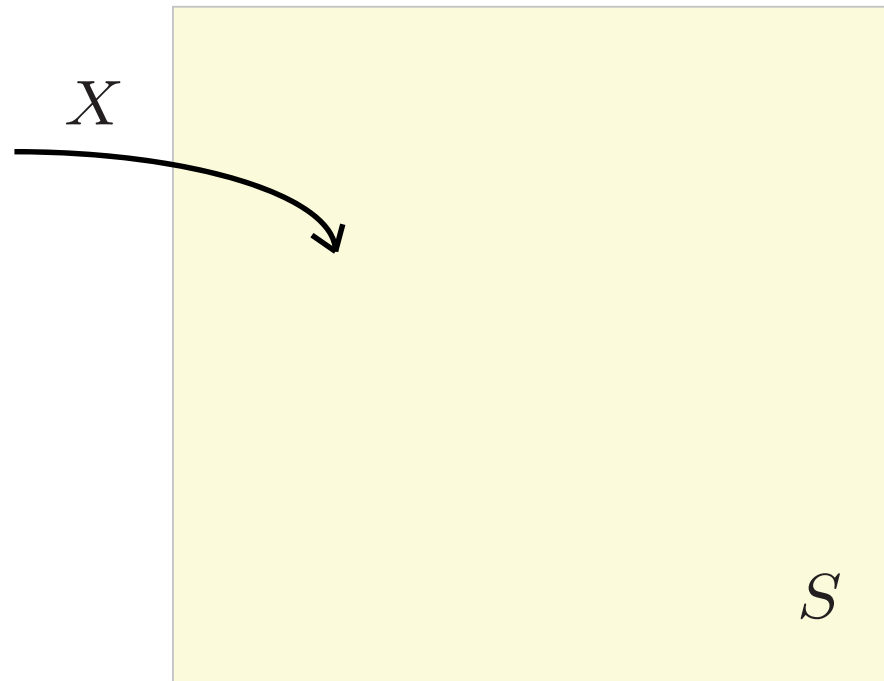
Vorlesung 1a

Zufallsvariable und Ereignisse, Wahrscheinlichkeiten und Verteilungen

Ein erster Blick

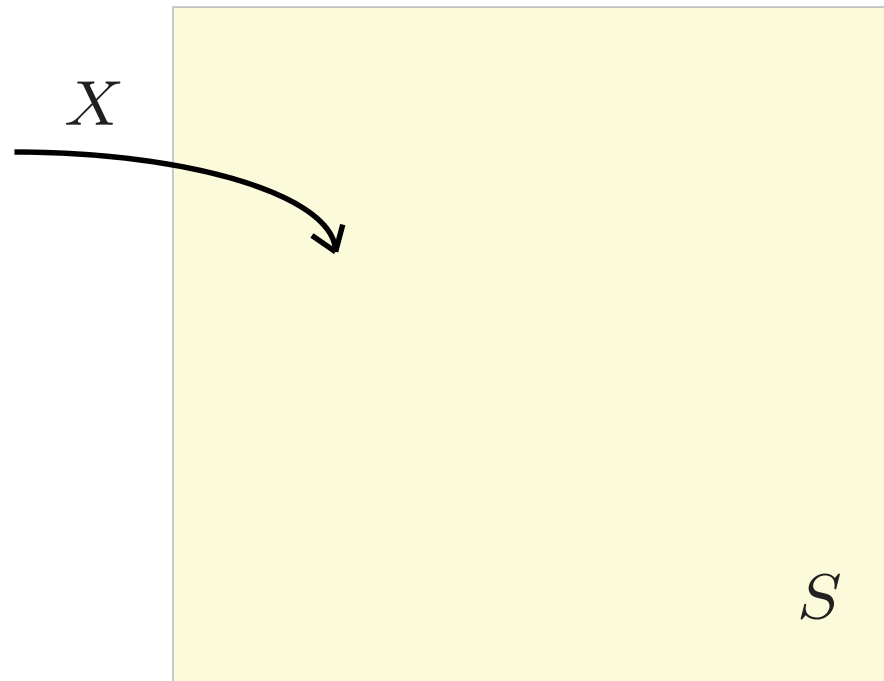
1. Rein zufällige Wahl aus einer endlichen Menge

Prototypisches Beispiel einer **Zufallsvariablen**:

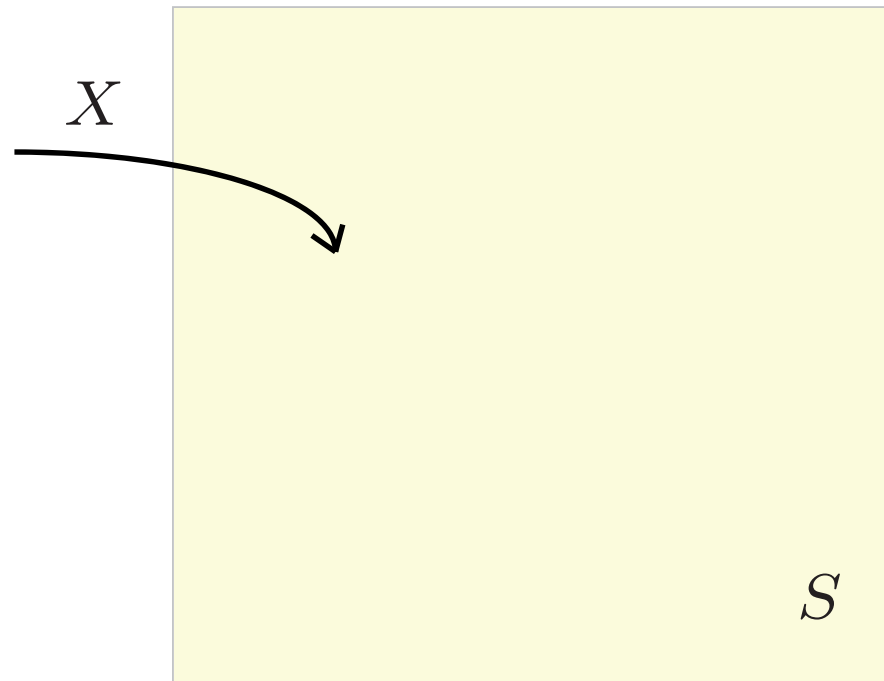


rein zufällige Wahl eines Punktes aus einer Fläche
(z. B. aus dem beigeen Quadrat S)

Prototypisches Beispiel einer **Zufallsvariablen**:

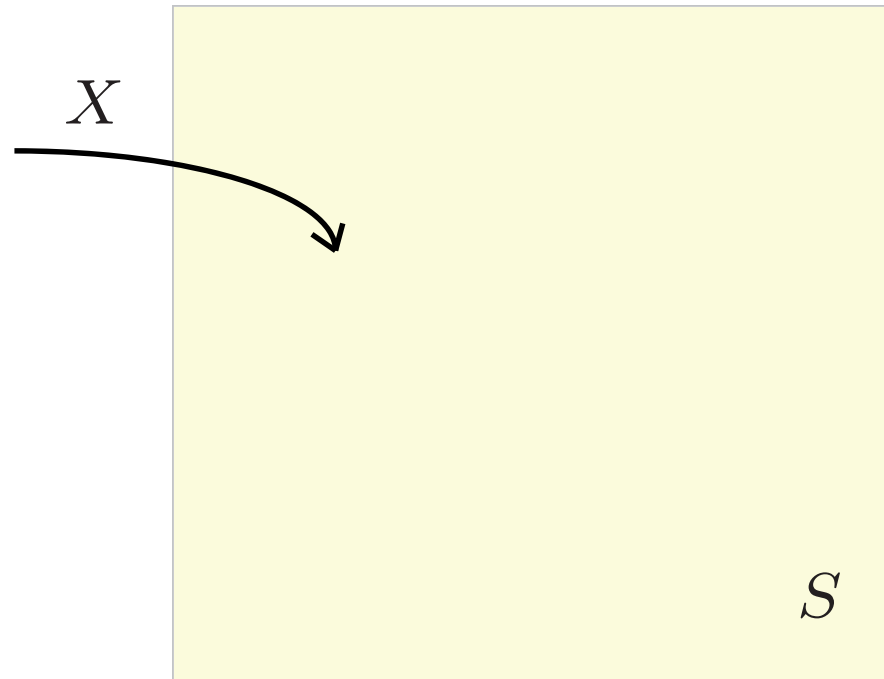


Was heißt “rein zufällige Wahl” eines Punktes aus S ?

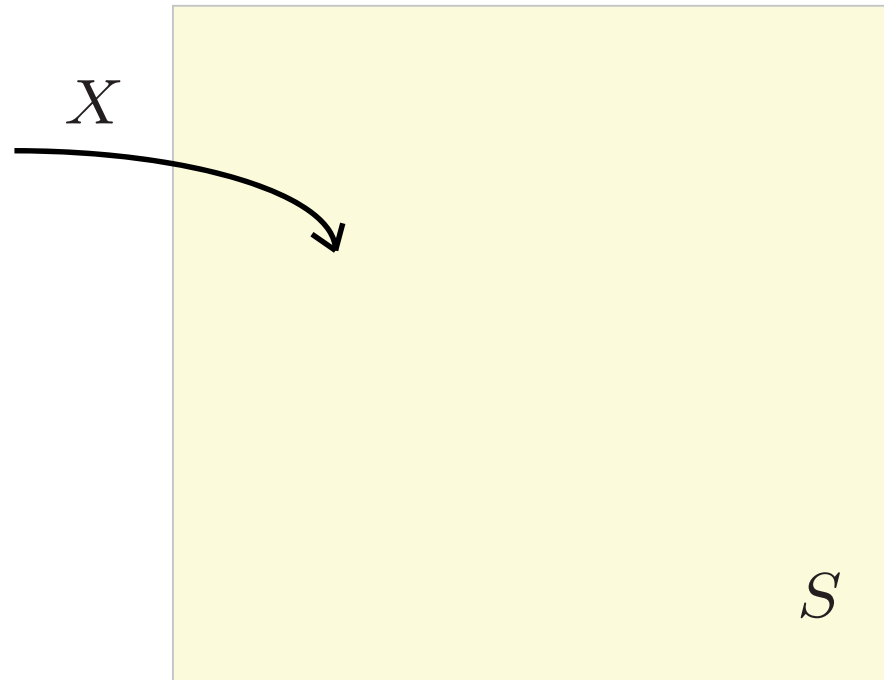


Fürs Erste sind endlich viele Pixel leichter vorstellbar
als ein Kontinuum aus unendlich vielen Punkten.

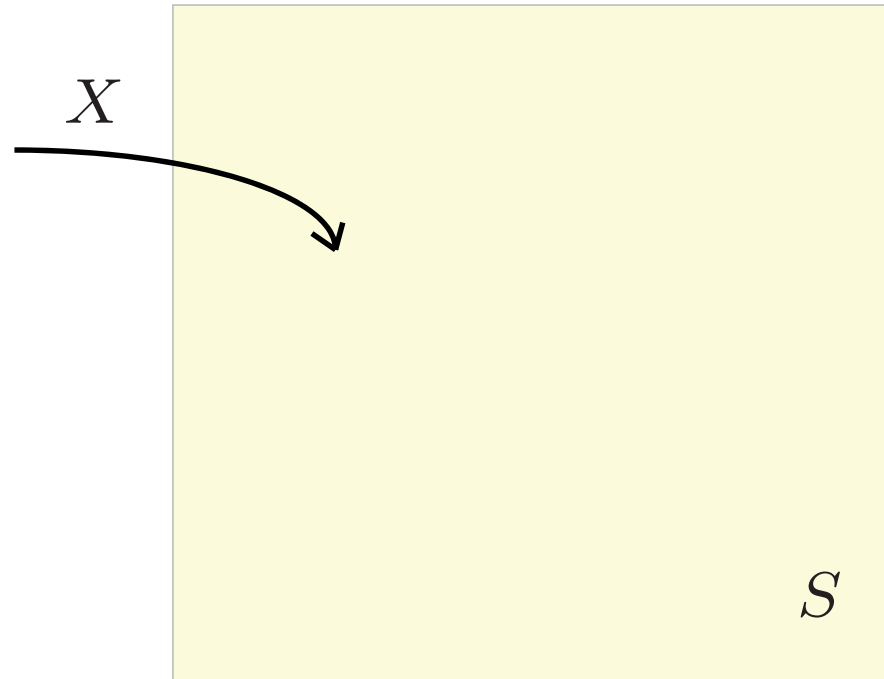
Stellen wir uns vor, S besteht aus 1000×1000 Pixeln



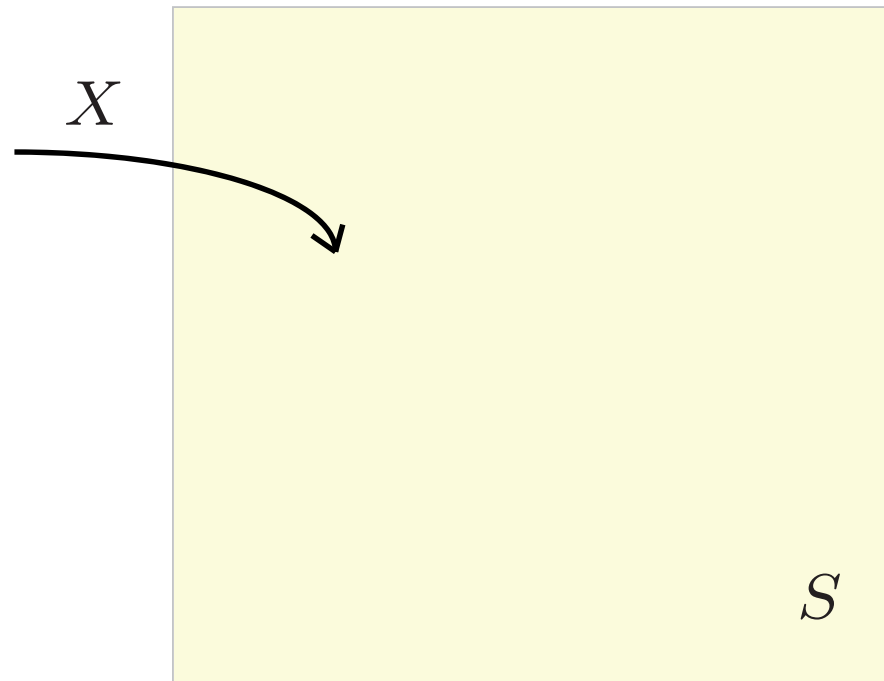
Stellen wir uns vor, S besteht aus g Pixeln, mit $g \in \mathbb{N}$



“Rein zufällige Wahl aus S ” soll heißen:



alle Pixel in S haben dieselbe Chance,
zum Zug zu kommen.



Man spricht dann von einer
uniform auf S verteilten Zufallsvariablen X

Analogie zum **fairen Würfeln**:

Die Menge der möglichen Ausgänge ist hier

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Auch hier wird der zufällige Ausgang beschrieben durch eine
uniform auf S verteilte Zufallsvariable X .

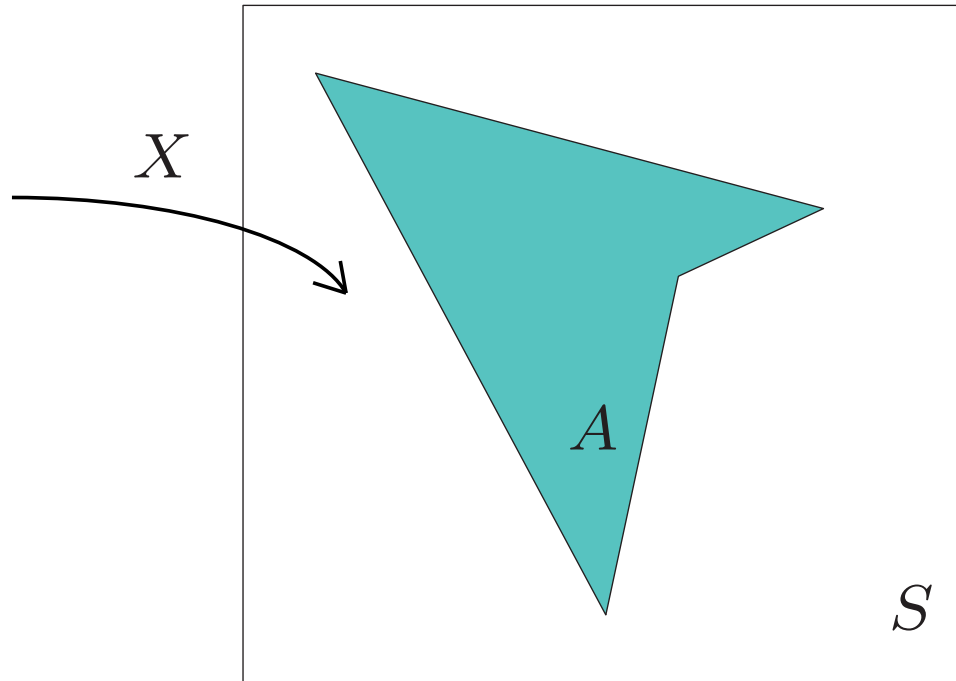
Wir halten fest:

Zufallsvariable mit Werten in einer endlichen Menge S
entsprechen einer zufälligen Wahl aus S .

Auf S **uniform** verteilte Zufallsvariable
entsprechen einer **rein** zufälligen Wahl aus S ,
bei der alle Elemente von S die gleiche Chance haben,
zum Zug zu kommen.

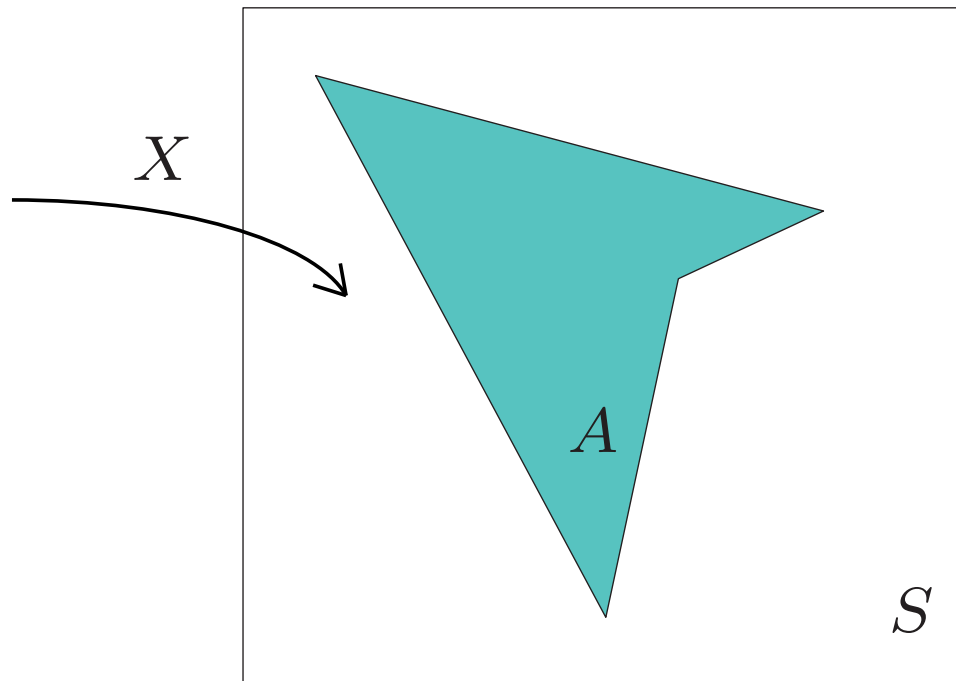
2. Ereignisse

Betrachten wir wieder unser Quadrat S



diesmal zusammen mit einer bestimmten **Teilmenge** A von S

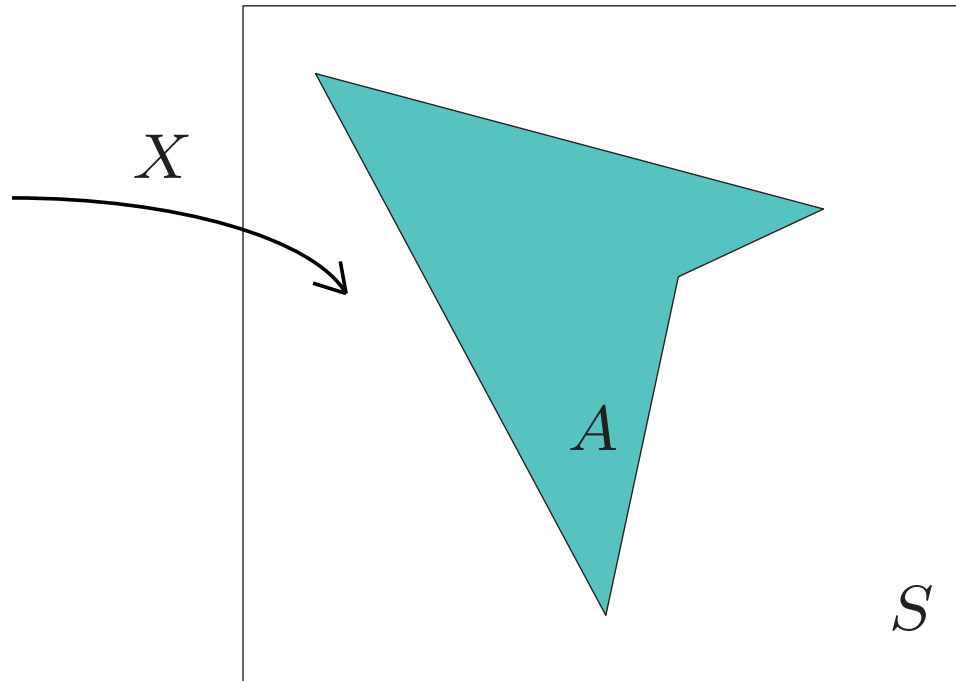
Bei der Wahl eines Pixels aus S



kann die Wahl auf A fallen

– oder auch nicht.

Bei der Wahl eines Pixels aus S



kann das Ereignis “ X fällt in A ” eintreten
– oder auch nicht.

Das Ereignis “ X fällt in A ”

notiert man als

$$\{X \in A\}.$$

Die Menge aller Ereignisse $\{X \in A\}$, $A \subset S$,

nennt man auch die

“von X erzeugte Kollektion von Ereignissen”.

Ereignisse kann man (aussagen-)logisch verknüpfen, z.B. gilt:

$$\{X \in A\} \text{ und } \{X \in B\} = \{X \in A \cap B\}.$$

Mehr dazu später!

3. Wahrscheinlichkeiten

Wie wahrscheinlich ist es,
dass bei einer rein zufälligen Wahl eines Pixels aus S
die Wahl auf A fällt?

Wie groß ist die Wahrscheinlichkeit
des Ereignisses $\{X \in A\}$?

.... eine Zahl zwischen 0 (= 0%) und 1 (= 100%)

Merke: Das Ereignis $\{X \in S\}$ hat Wahrscheinlichkeit 1.

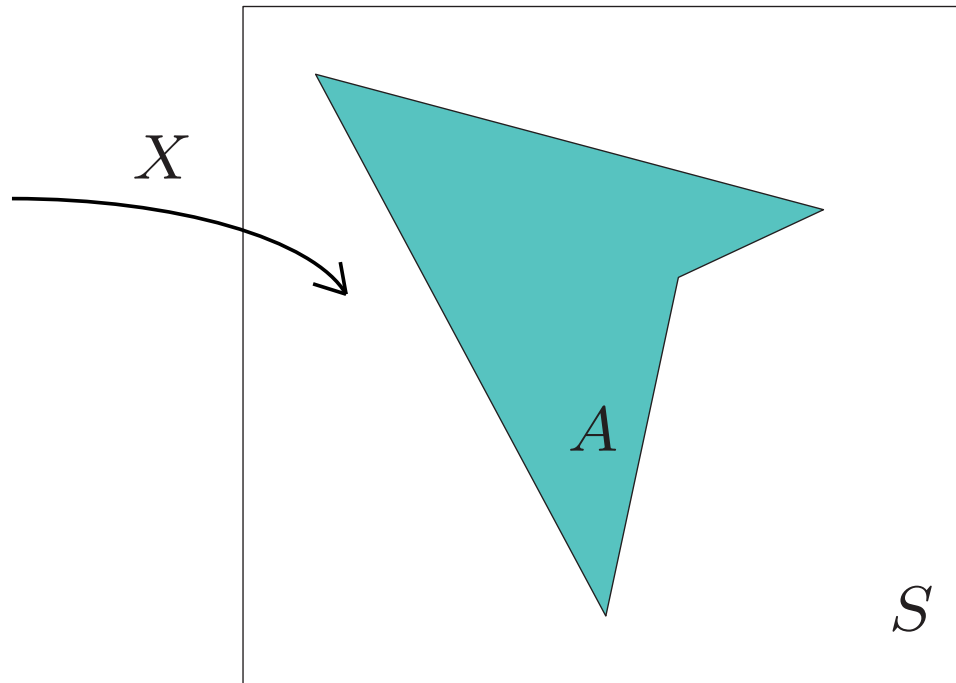
Für die Wahrscheinlichkeit des Ereignisses $\{X \in A\}$
schreiben wir

$$P(\{X \in A\})$$

oder kurz

$$P(X \in A).$$

P steht für **probabilitas** (= Wahrscheinlichkeit)



Bei der **rein zufälligen Wahl** eines Pixels aus S ,
beschrieben durch die Zufallsvariable X , ist
die Wahrscheinlichkeit des Ereignisses $\{X \in A\}$
proportional zur Anzahl der Pixel in A .

Zusammengefasst:

X ist rein zufälliger Pixel aus dem Quadrat S

bedeutet:

Für jede Teilmenge A von S ist

$$P(\{X \in A\}) = \frac{\text{Anzahl der Pixel in } A}{\text{Anzahl der Pixel in } S}.$$

Lies und merke:

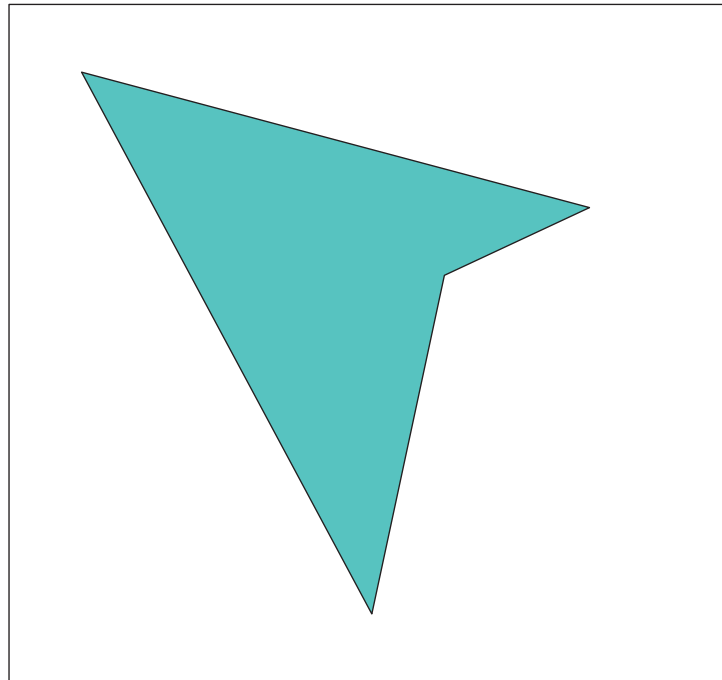
die Wahrscheinlichkeit, dass X in A fällt
ist der Anteil der Menge A an der Menge S .

4. Schätzung eines Flächenanteils

Eine Anwendung der “rein zufälligen Wahl”:

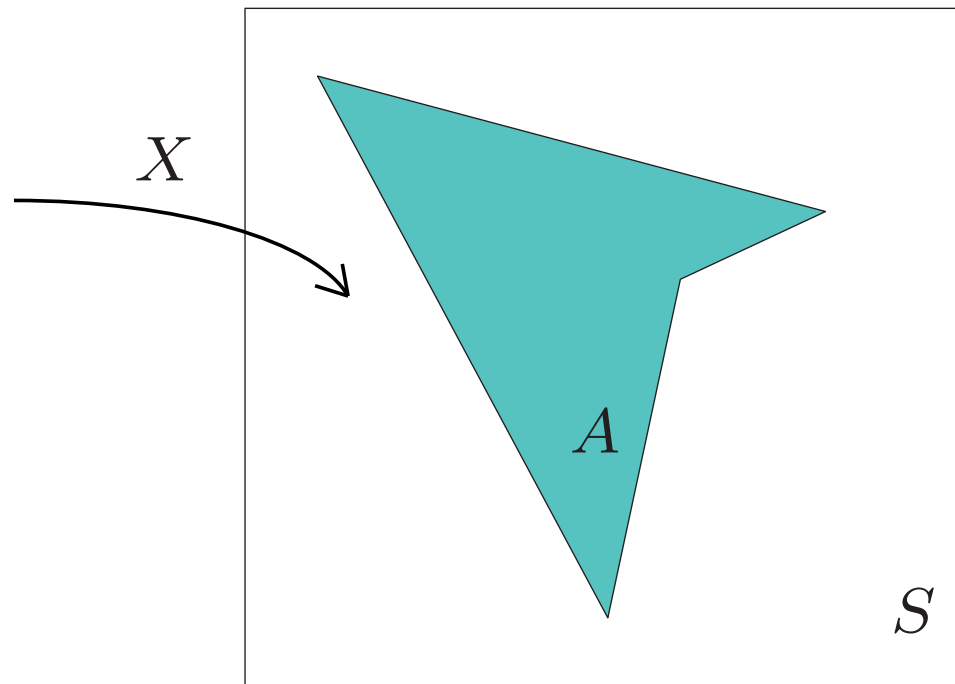
Monte-Carlo Schätzung eines Flächenanteils.

Wir fragen:



Wie groß ist der Anteil der blauen Fläche
an der Fläche des Quadrats?

und übersetzen in die Sprache der Stochastik:



Wie wahrscheinlich ist es, dass die
rein zufällige Wahl eines Pixels aus dem Quadrat S
in die blaue Fläche trifft?

Wie wir bald sehen werden

(und wie auch intuitiv klar ist)

gibt es einen engen Zusammenhang zwischen

Wahrscheinlichkeiten und Trefferquoten.

Angenommen wir haben ein Werkzeug, mit dem man einen **rein zufälligen Pixel** aus dem Quadrat wählen kann

– und das nicht nur einmal, sondern “immer wieder neu”.

Wir bekommen dann mit unserem Werkzeug
nicht nur *einen einzigen* rein zufälligen Pixel,

sondern sogar beliebig viele,
genauer:

eine *rein zufällige Folge* $(X_1, X_2, \dots)$

von Pixeln in S .

Sei $A \subset S$.

$$Z_i := \mathbf{1}_A(X_i), \quad i = 1, 2, \dots$$

zählt, ob X_i in A fällt.

Dabei ist

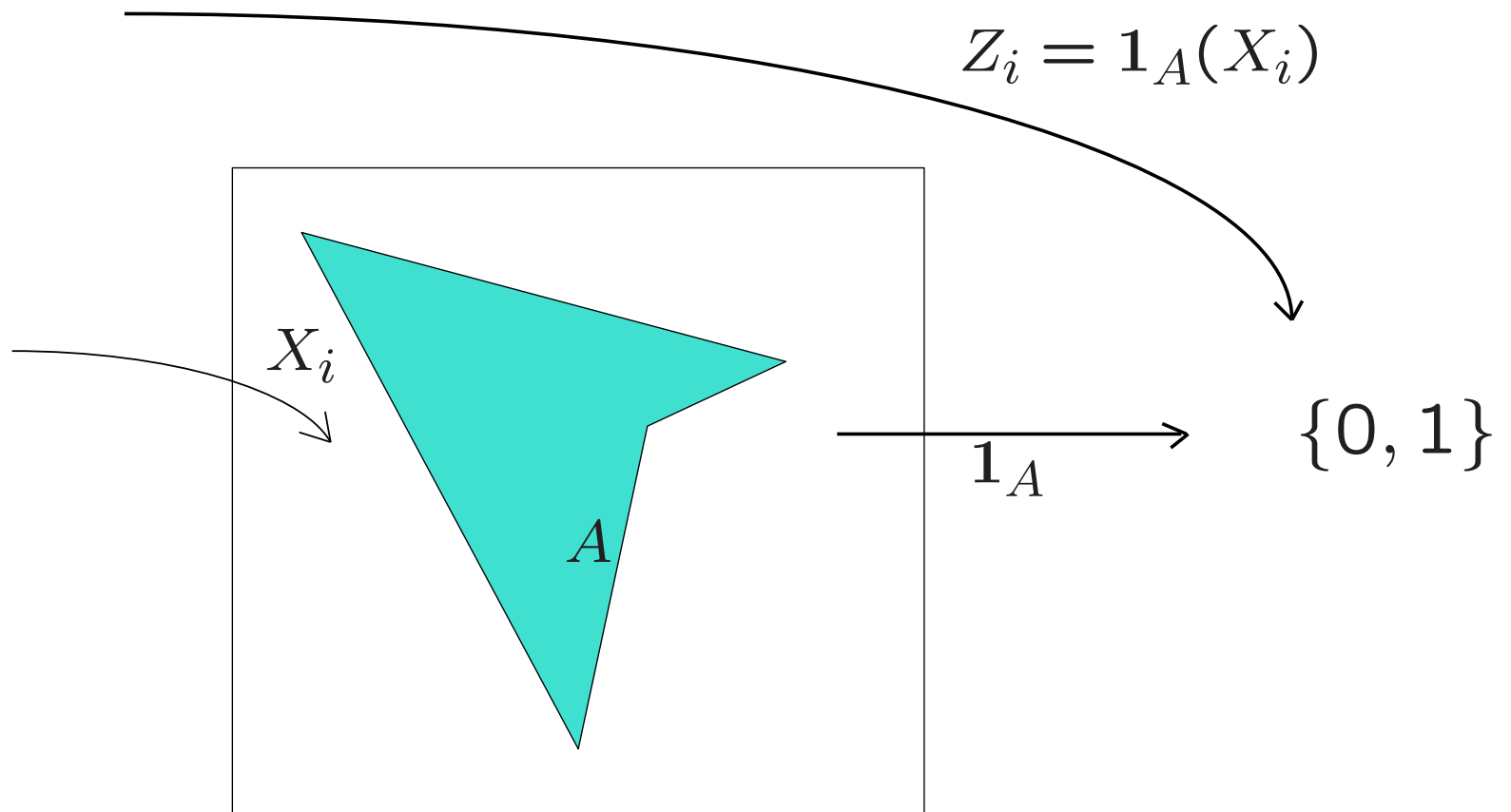
$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{für } x \in A \\ 0 & \text{für } x \in S \setminus A \end{cases}$$

die Indikatorfunktion der Menge A .

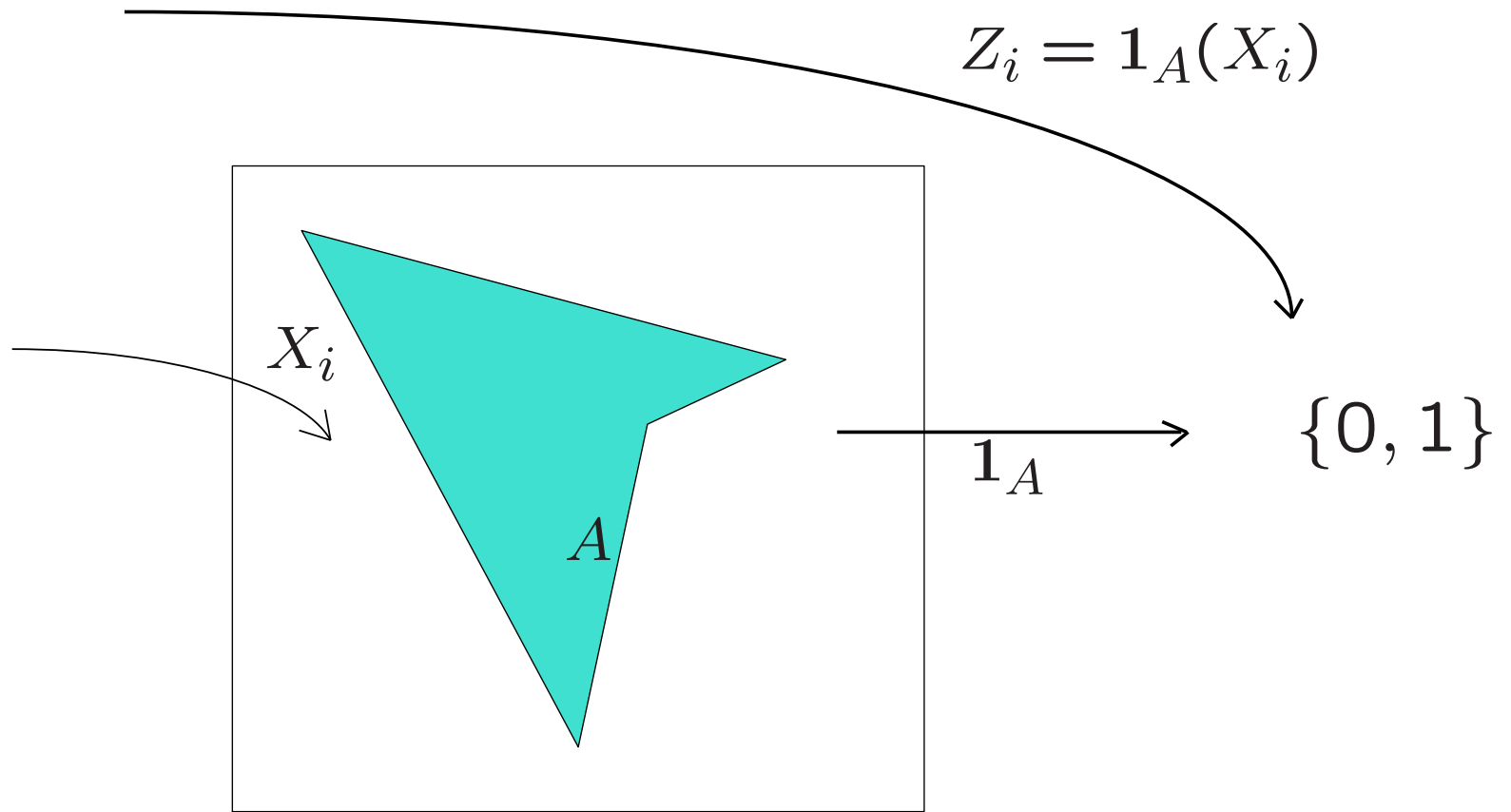
Mit dieser Abbildung von S nach $\{0, 1\}$

wird die S -wertige Zufallsvariable X_i

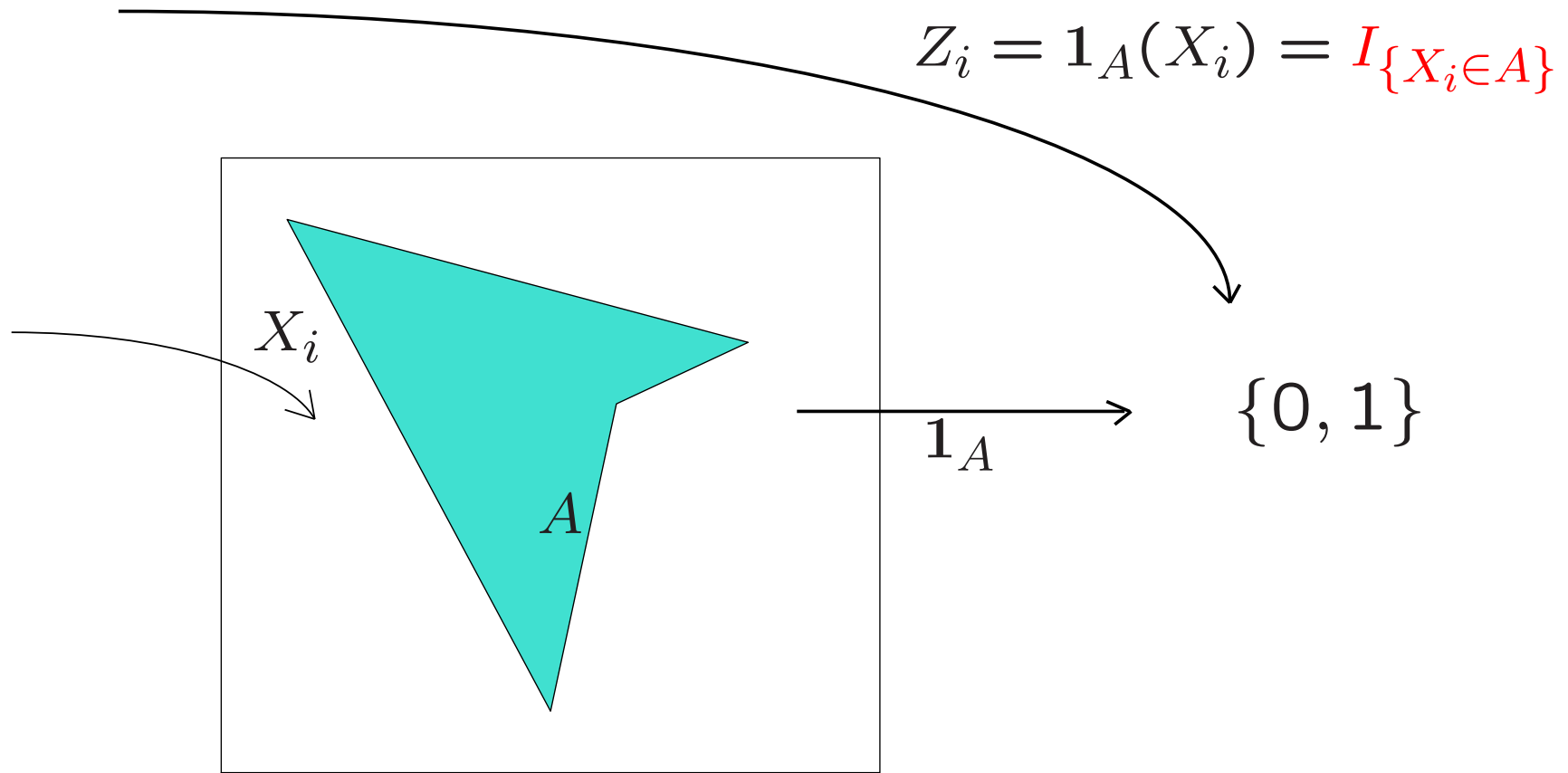
zur $\{0, 1\}$ -wertigen Zufallsvariablen Z_i “verarbeitet”.



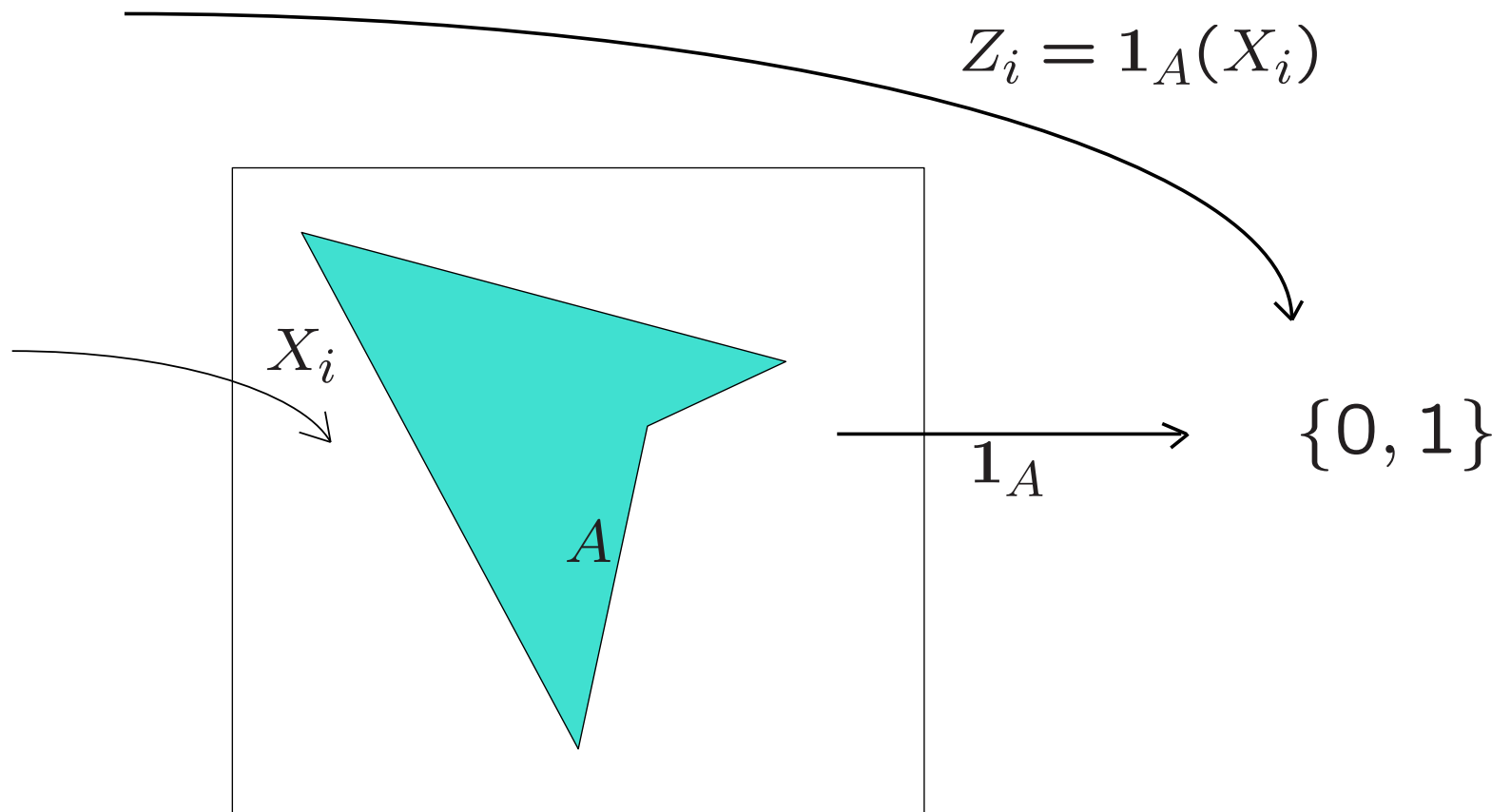
$Z_i = 1_A(X_i)$ ist der *der Indikator* (“der Zähler”) des Ereignisses “ X_i fällt in A ”.



Eine alternative Schreibweise für $1_A(X_i)$ ist $I_{\{X_i \in A\}}$



Eine alternative Schreibweise für $1_A(X_i)$ ist $I_{\{X_i \in A\}}$



$\{X_i \in A\}$ und $\{Z_i = 1\}$ sind zwei Schreibweisen für dasselbe Ereignis “ X_i fällt in A ”

5. Die Verteilung der zufälligen Trefferquote

Illustriert durch ein R-Programm
werden $n = 100$ Punkte
“rein zufällig” aus dem Quadrat gewählt.

$$Z_i = 1_A(X_i), \quad i = 1, 2, \dots$$

Die zufällige Zahl

$$M := \frac{1}{100}(Z_1 + \dots + Z_{100})$$

(die “Trefferquote”)

ist ein *Schätzer* für die Wahrscheinlichkeit

$$p := \mathbf{P}(X \in A)$$

(und damit für den gefragten Flächenanteil).

Ein Ergebnis (“eine Realisierung”) von $(X_1, \dots, X_{100})$
liefert eine Realisierung von $(Z_1, \dots, Z_{100})$
und damit eine Realisierung von M
(einen Schätzwert für p).

Wie “zuverlässig” ist dieser Schätzwert?

$$M := \frac{1}{100}(Z_1 + \cdots + Z_{100})$$

Es sei an dieser Stelle verraten:

Der Anteil der blauen Fläche am Quadrat
(den man in der Realität ja nicht kennt) ist in unserem Beispiel

$$p = 0.195$$

Damit hat M gar keine Chance, exakt auf p zu fallen, denn
der *Wertebereich* von M

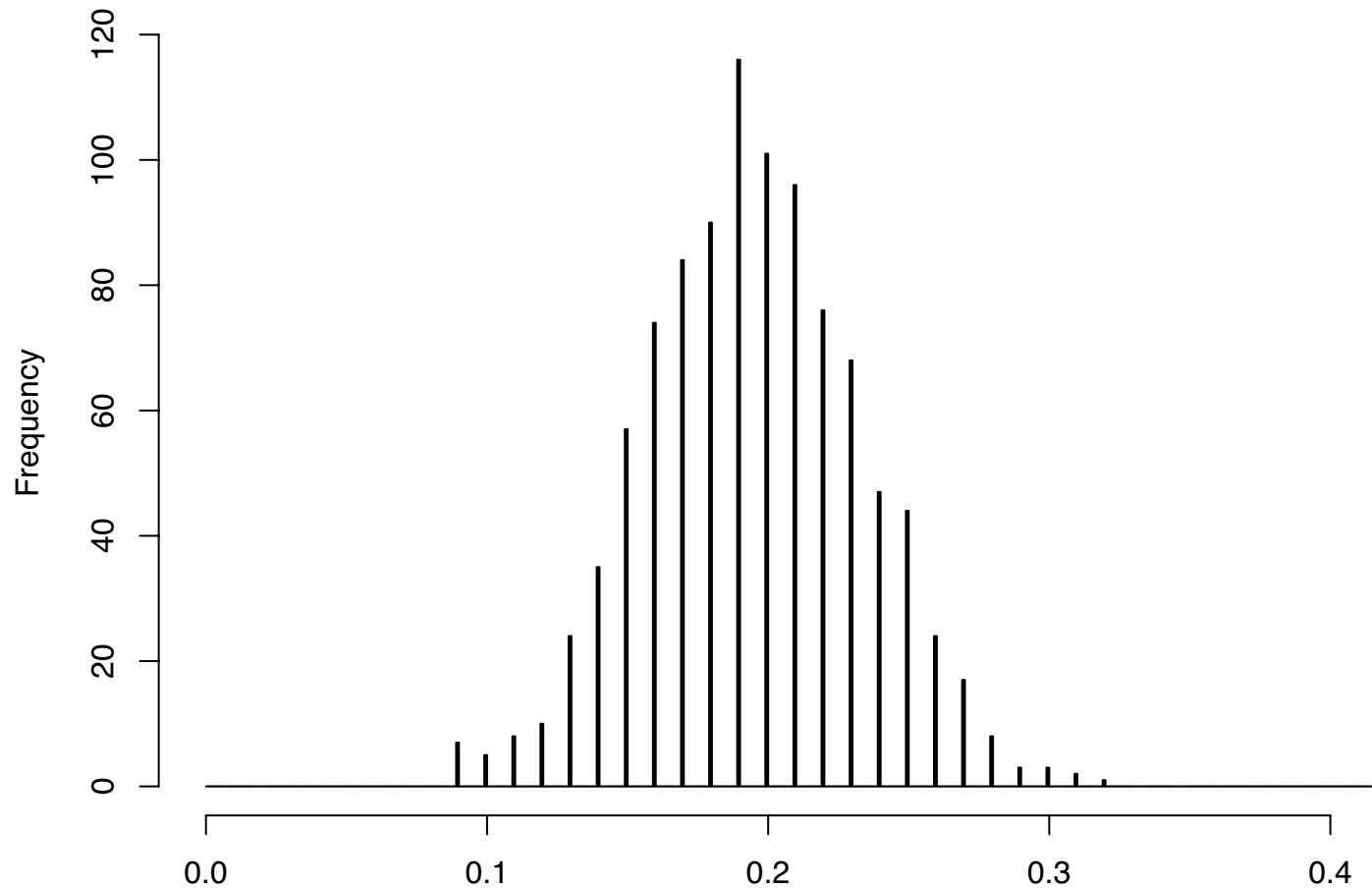
(d.h. die Menge der möglichen Ausgänge) ist

$$S' := \left\{ \frac{0}{100}, \frac{1}{100}, \cdots, \frac{99}{100}, \frac{100}{100} \right\}$$

Wie zuverlässig ist M als Schätzer für p ?

Davon machen wir uns ein Bild, indem wir viele
(z.B. 1000) “unabhängige Kopien” von M erzeugten
und in einem *Histogramm* darstellen,
wie oft welche Ausgänge realisiert wurden.

Verteilung von M (Näherung aus 1000 Wiederholungen):



Wie zuverlässig ist M als Schätzer für p ?

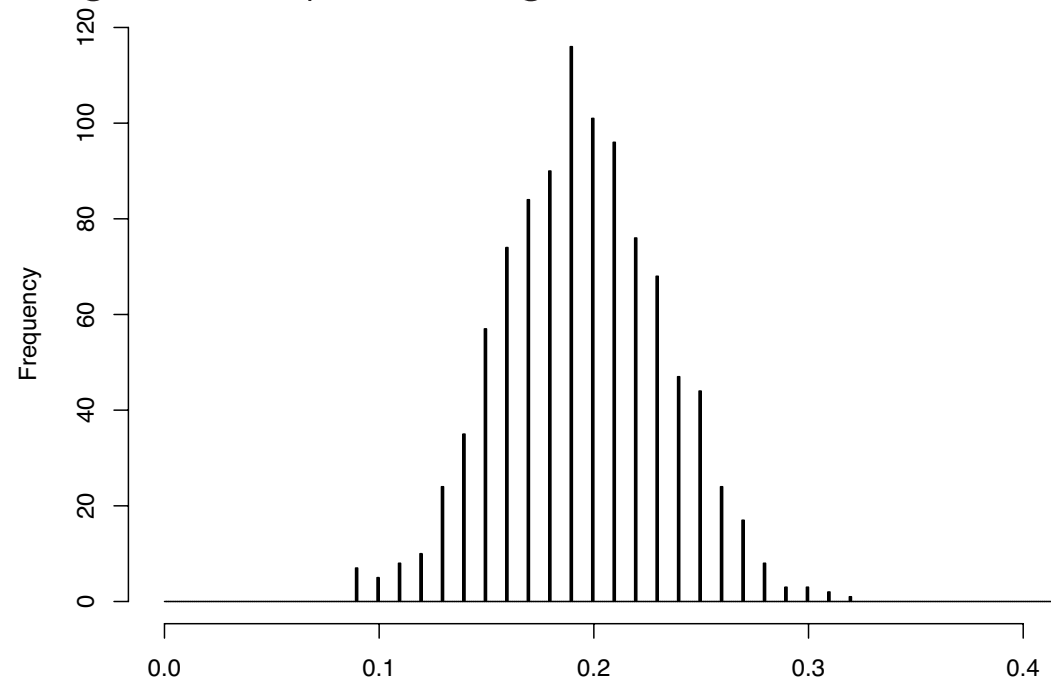
Davon machen wir uns ein Bild, indem wir viele (z.B. 1000) “unabhängige Kopien” von M erzeugen und in einem *Histogramm* darstellen, wie oft welche Ausgänge realisiert wurden.

So bekommen wir eine näherungsweise Darstellung der **Verteilung** von M .

Die **Verteilung** von M ist bestimmt durch ihre **Gewichte**

$$\rho(b) := \mathbf{P}(M = b), \quad b \in S'.$$

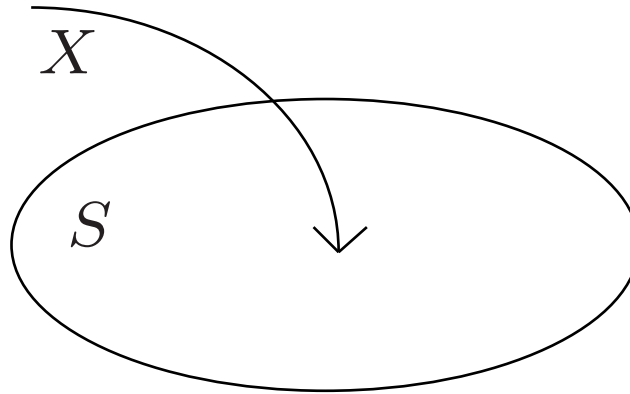
Verteilung von M (Näherung aus 1000 Wiederholungen):



Die 101 möglichen Ausgänge von M
sind (bei weitem) nicht gleich wahrscheinlich:
die Verteilung von M “ist um p konzentriert”.

6. Zusammenfassung der wichtigsten Begriffe der ersten Stunde

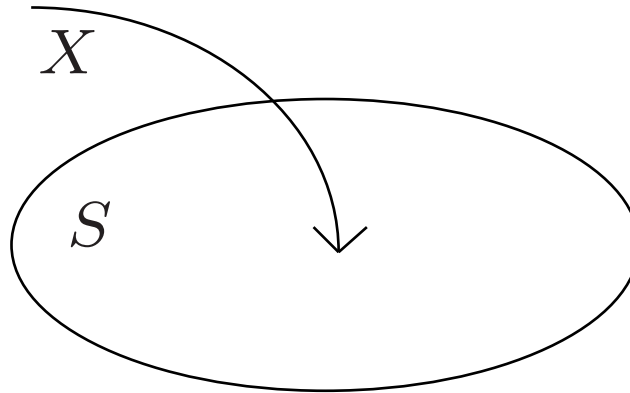
Ein Logo der Elementaren Stochastik:



X ... zufällige Wahl eines Elements aus S

S ... Menge von möglichen Ausgängen

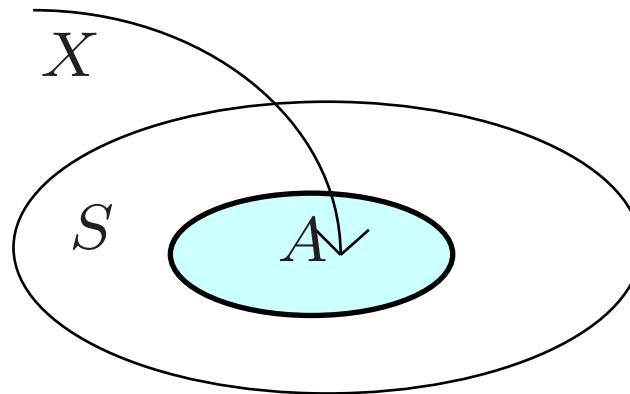
Ein Logo der Elementaren Stochastik:



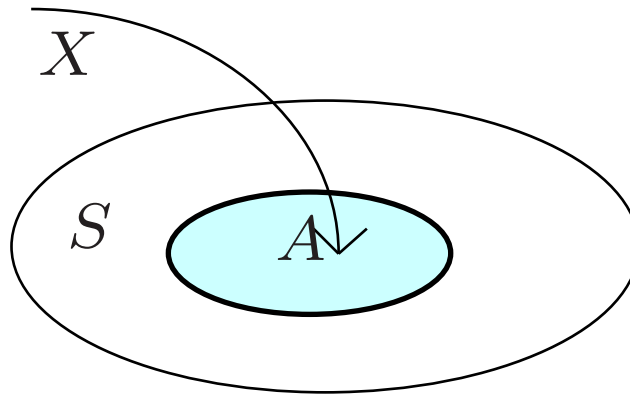
X ... **Zufallsvariable**

mit *Zielbereich (Wertebereich)* S

Wir interessieren uns für die *Wahrscheinlichkeit*
des *Ereignisses* “ X fällt in A ”



Dabei ist A eine bestimmte Teilmenge von S .



Ereignisse werden (wie Mengen)
in geschweiften Klammern notiert:

$$\{X \in A\}$$

Lies:

“ X fällt in A ”.

“ X ist *rein zufällig*”

heißt im Fall einer *endlichen* Menge S :

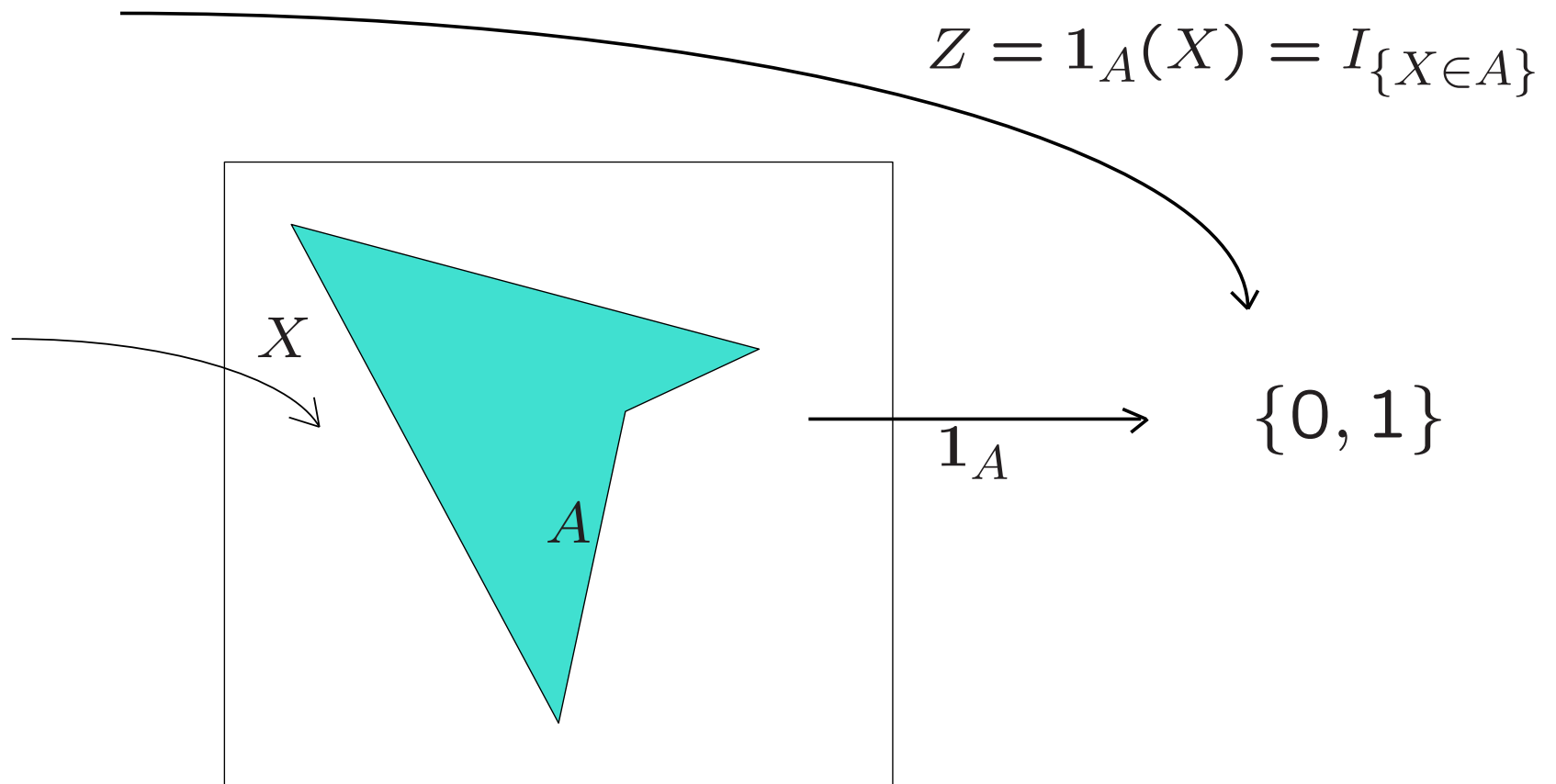
alle Elemente von S haben die gleiche W'keit
gewählt zu werden.

Dann ist die **Wahrscheinlichkeit** des Ereignisses “ X fällt in A ”:

$$\mathbf{P}(\{X \in A\}) = \frac{\#A}{\#S}.$$

Statt $\mathbf{P}(\{X \in A\})$ schreiben wir kurz:

$$\mathbf{P}(X \in A).$$



Die Ereignisse $\{X \in A\}$ und $\{Z = 1\}$ stimmen überein!

Ein Ausblick ins Kontinuum:

“ X ist *rein zufällige Wahl aus S* ”

heißt im Fall einer **kontinuierlichen** Menge $S \subset \mathbb{R}^d$:

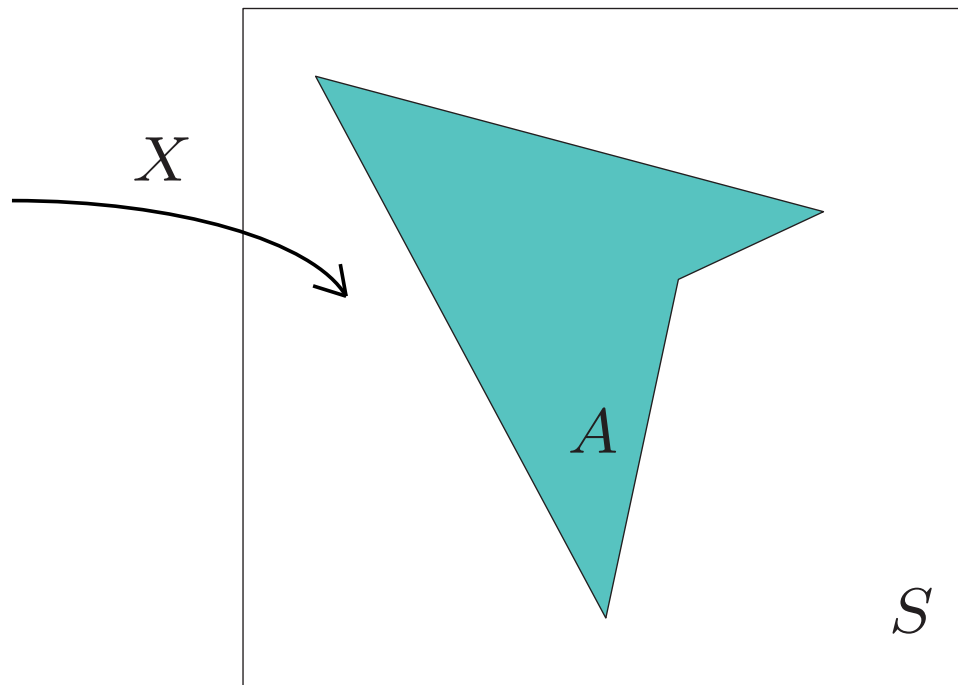
jede (messbare) Teilmenge A von S

kommt mit einer W'kt zum Zug,

die dem relativen Anteil ihres Volumens

am Volumen von S entspricht:

$$P(X \in A) = \frac{\text{Volumen von } A}{\text{Volumen von } S}$$



Vorlesung 1b

Wahrscheinlichkeit von Kollisionen

bei der wiederholten rein zufälligen Wahl

aus endlich vielen möglichen Ausgängen

(Buch S. 1-5)

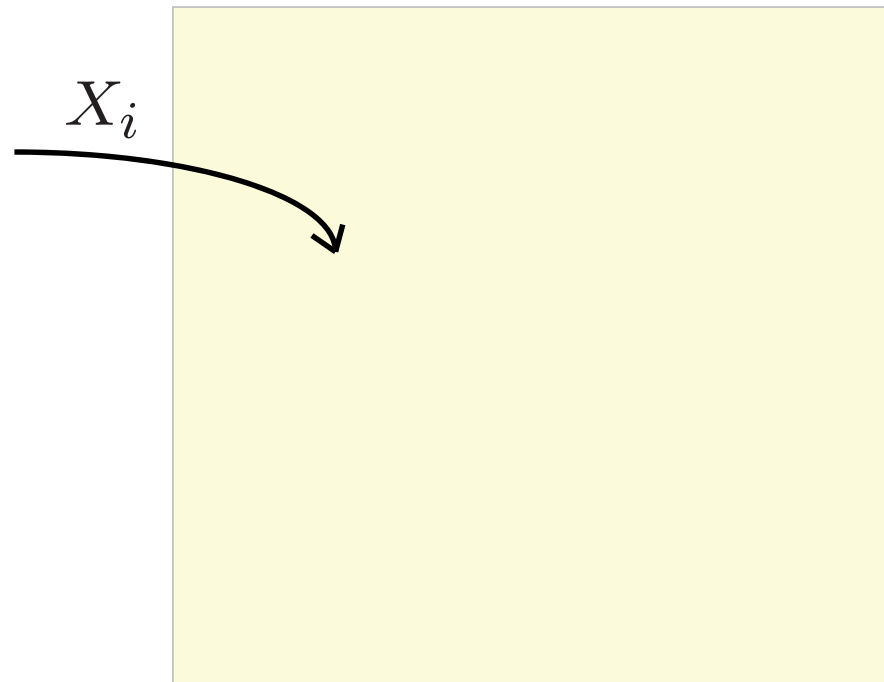
Teil Eins:

Vom Modell zur Formel

1. Die Fragestellung

– in verschiedenen Verpackungen

n -mal wiederholt wird rein zufällig
ein Pixel aus $g = 10^6$ Pixeln gewählt.



$$i = 1, \dots, n$$

n -mal wiederholt wird rein zufällig
ein Pixel aus $g = 10^6$ Pixeln gewählt.

Wie wahrscheinlich ist es, dass dabei
lauter verschiedene Pixel gewählt werden?

Ziemlich unwahrscheinlich
oder ziemlich hoch wahrscheinlich,
oder so “mittendrin”, für

$$n = 100 \quad ?$$

$$n = 1000 \quad ?$$

$$n = 10000 \quad ?$$

Was schätzen Sie?

Bei der “wiederholten rein zufälligen Pixelwahl”
handelt es sich um ein

Ziehen mit Zurücklegen.

Man kann auch denken an eine Urne
mit g Kugeln, nummeriert mit $1, 2, \dots, g$.

Es wird n -mal gezogen.

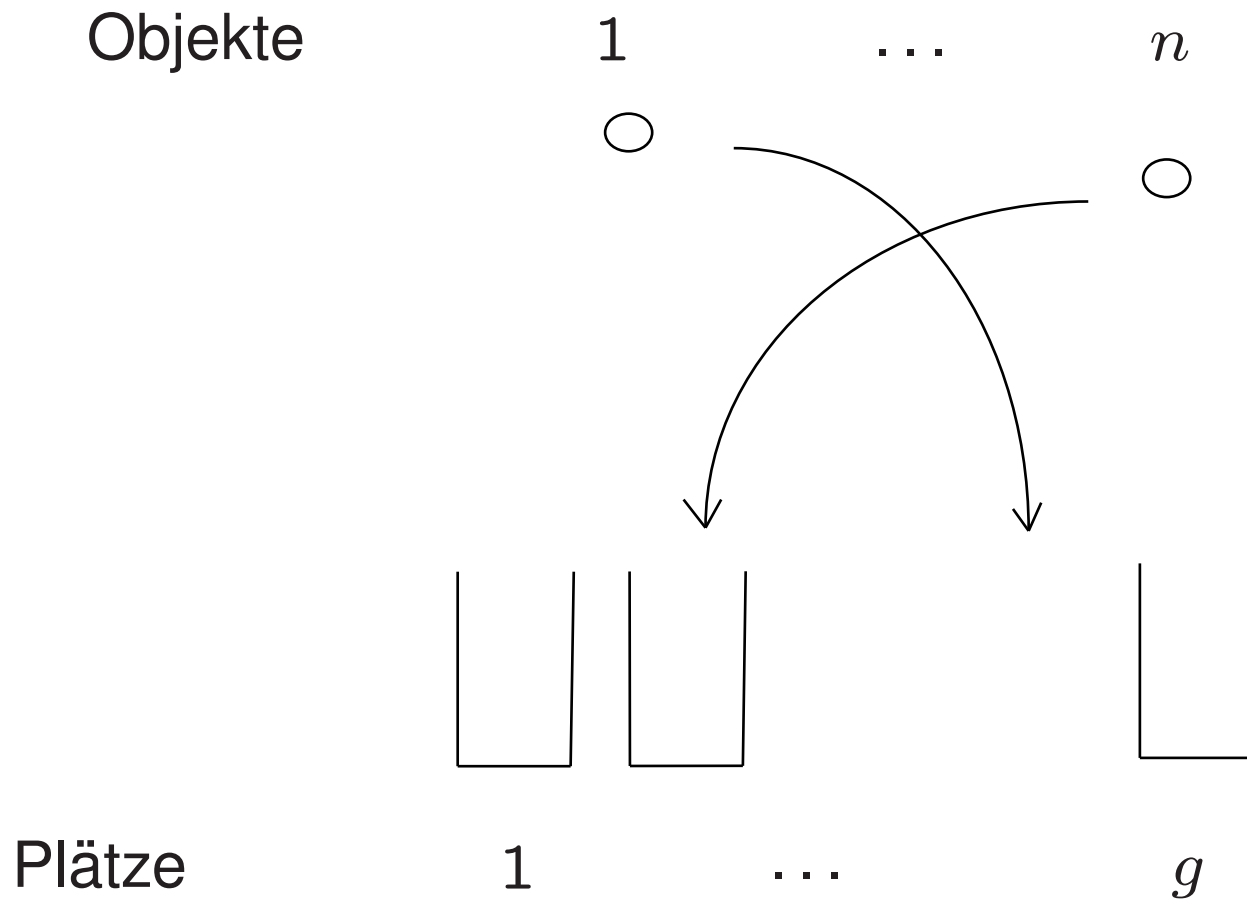
Nach jedem Zug wird die gezogene Kugel zurückgelegt,
vor jedem Zug wird “perfekt durchmischt”.

Tema con variazioni:

n Objekte,
 g Plätze.

Jedes Objekt wird auf einen
rein zufällig ausgewählten Platz gesetzt.
(Mehrfachbelegungen sind erlaubt!)

Wie wahrscheinlich ist es, dass dabei
keine Mehrfachbelegung ("Kollision") auftritt?



Anders als im Buch wird hier die Anzahl der Kennzeichen mit g (und nicht mit r) bezeichnet - wegen des besseren typographischen Unterschieds von g zu n und der Assoziation von g mit **G**esamtzahl der Plätze, **G**eburtstage,...

Populäre Version:

$n = 25$ Personen auf einer Party

Platz $\longleftrightarrow$ Geburtstag ($\in \{1, 2, \dots, 365\}$)

Wie wahrscheinlich ist es,
dass keine zwei Personen am selben Tag Geburtstag haben?

2. Beschreibung durch eine Zufallsvariable und ein Ereignis.

(Buch S. 2-3)

Die Objekte denken wir uns mit 1 bis n
und die Plätze mit 1 bis g nummeriert.

Ein **Ausgang** der Platzwahl (ein “**Wahlprotokoll**”)
lässt sich beschreiben durch das n -tupel

$$a = (a_1, \dots, a_n) ,$$

wobei a_i den für das i -te Objekt gewählte Platz bezeichnet
($1 \leq a_i \leq g$).

Die Menge der möglichen Wahlprotokolle ist

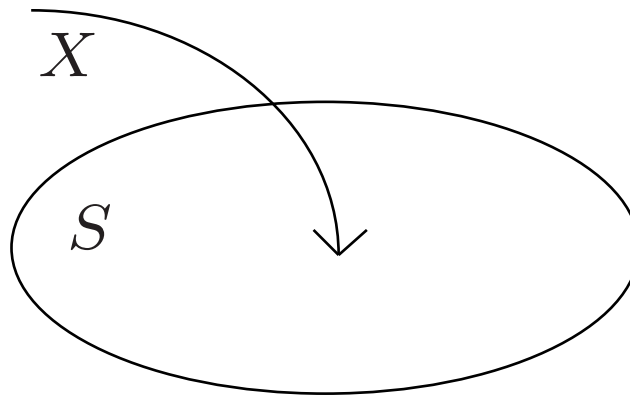
$$S := \{1, \dots, g\}^n ,$$

die Menge aller n -tupel $(a_1, \dots, a_n)$

mit Einträgen (Komponenten)

$$a_i \in \{1, \dots, g\}, \quad i = 1, \dots, n.$$

Die zufällige Platzwahl
beschreiben wir durch eine *S -wertige Zufallsvariable X* .



X kommt durch zufällige Wahl
eines Elementes aus S zustande.

Die Menge S heißt *Zielbereich* (oder *Wertebereich*)
der Zufallsvariable X .

Wie jedes Element $(a_1, \dots, a_n)$ unserer Menge S
besteht auch die Zufallsvariable X aus n Komponenten:

$$X = (X_1, \dots, X_n) .$$

Wir interessieren uns für das *Ereignis*,
dass **keine zwei Komponenten von X gleich** sind.

Dieses Ereignis schreiben wir als

$$\{X_i \neq X_j \text{ für alle } i \neq j\}$$

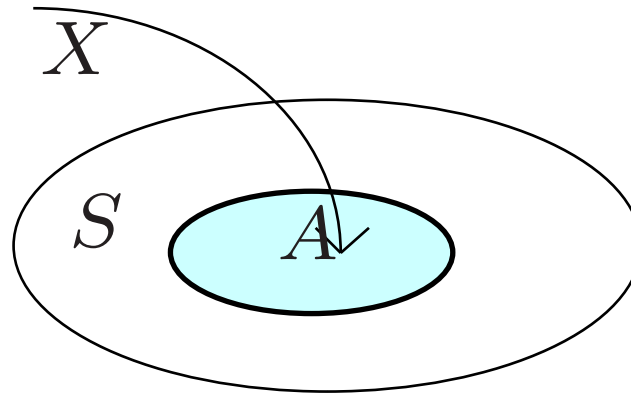
oder auch als

$$\{X \in A\}$$

mit der Teilmenge

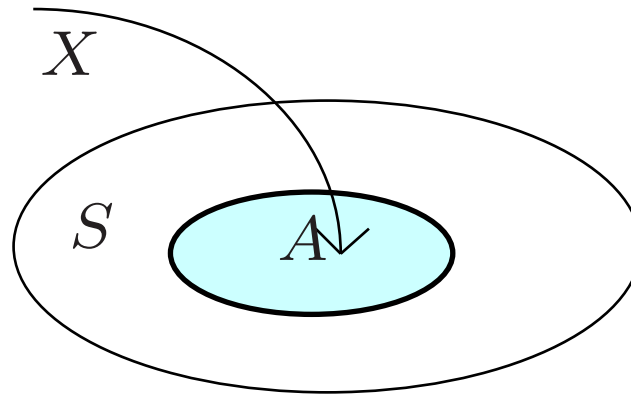
$$A := \{a \in S : a_i \neq a_j \text{ für alle } i \neq j\} .$$

Ein Logo für das Ereignis $\{X \in A\}$:



3. Die Wahrscheinlichkeit für Kollisionsfreiheit

(Buch S. 3)



Wie kommt man zur **Wahrscheinlichkeit**
des Ereignisses $\{X \in A\}$?

Zum Merken:

Wahrscheinlichkeiten gehören zu Ereignissen.

Die Wahrscheinlichkeit eines Ereignisses
misst dessen Chance einzutreten
mit einer Zahl zwischen 0 und 1:

$$\mathbf{P}(\{X \in A\})$$

Statt $\mathbf{P}(\{X \in A\})$

schreiben wir kurz

$$\mathbf{P}(X \in A)$$

Statt $\mathbf{P}(X \in \{a\})$

schreiben wir kurz

$$\mathbf{P}(X = a)$$

Eine einleuchtende Regel
für das Rechnen mit Wahrscheinlichkeiten ist die
Additivität:

Für disjunkte $A, A' \subset S$ gilt

$$\mathbf{P}(X \in A \cup A') = \mathbf{P}(X \in A) + \mathbf{P}(X \in A').$$

Daraus folgt im Fall endlich vieler möglicher Ausgänge,
d.h. für $\#S < \infty$:

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a)$$

Um die Wahrscheinlichkeit $P(X \in A)$ berechnen zu können,
muss man eine **Modellannahme** treffen.

Eine prominente Modellannahme ist die einer
rein zufälligen Wahl.

Damit ist gemeint, dass für je zwei $a, a' \in S$

$$\mathbf{P}(X = a) = \mathbf{P}(X = a').$$

Das heißt: kein Ausgang ist bevorzugt.

Gemäß der Additivität gilt:

$$\mathbf{P}(X \in S) = \sum_{a \in S} \mathbf{P}(X = a).$$

Und wir hatten schon vereinbart: $\mathbf{P}(X \in S) = 1$

Für eine rein zufällige Wahl folgt daraus sofort:

$$P(X = a) = \frac{1}{\#S}, \quad a \in S.$$

Und

$$\mathbf{P}(X \in A) = \frac{\#A}{\#S}.$$

Wir haben nun die Aufgabe des Abzählens der zwei Mengen

$$S := \{1, \dots, g\}^n = \{(a_1, \dots, a_n) : 1 \leq a_i \leq g\}$$

und

$$A := \{a \in S : a_i \neq a_j \text{ für alle } i \neq j\}$$

$$\#S = g^n$$

$$S = \{(a_1, \dots, a_n) : 1 \leq a_i \leq g\}$$

$$A := \{a \in S : a_i \neq a_j \text{ für alle } i \neq j\}$$

$$\#A = ?$$

Für a_1 gibt es g mögliche Werte, für a_2 dann noch $g - 1$,
usw. Also:

$$\#A = g(g - 1) \cdots (g - (n - 1))$$

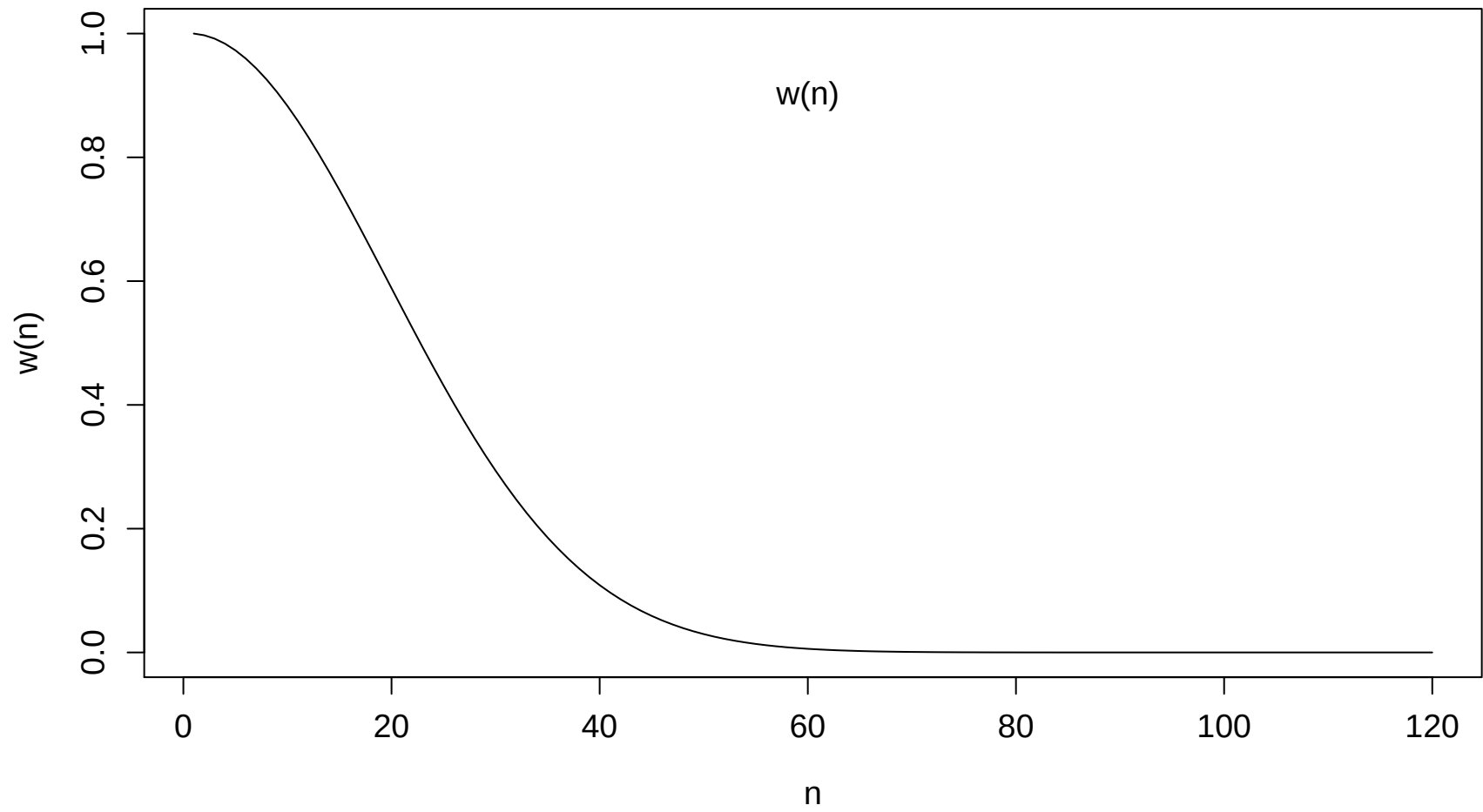
$$\mathbf{P}(X \in A) = \frac{g(g-1) \cdots (g-(n-1))}{g^n}$$

$$= \frac{g-1}{g} \frac{g-2}{g} \cdots \frac{g-(n-1)}{g}$$

$$\mathbf{P}(X \in A) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{g}\right) =: w(n, g)$$

ist die Wahrscheinlichkeit für Kollisionsfreiheit
 bei der n -maligen wiederholten rein zufälligen Platzwahl
 aus g Plätzen

Wahrscheinlichkeit fuer Kollisionsfreiheit bei $g=365$



4. Der Zeitpunkt der ersten Kollision und seine Verteilung

Ein dynamisches Bild:

Die Anzahl g der Plätze ist fest.

Jetzt wird ein Objekt nach dem anderen ($i = 1, 2, \dots, g + 1$)
auf einen (immer wieder neu)
rein zufällig gewählten Platz gesetzt.

Es sei X_i der von Objekt i gewählte Platz.

Eines ist sicher:

Spätestens bis $i = g + 1$ muss eine Kollision kommen
(warum?)

Das “zufällige Wahlprotokoll”

$$(X_1, X_2, \dots, X_{g+1})$$

ist somit eine Zufallsvariable mit Wertebereich

$$S :=$$

$$\{(a_1, \dots, a_{g+1}) : a_i \in \{1, \dots, g\}; \exists i \neq j \text{ mit } a_i = a_j\}$$

$$T := \min\{n \leq g + 1 : \exists i < n : X_i = X_n\}$$

ist der *Zeitpunkt der ersten Kollision*.

T ist eine aus $(X_1, \dots, X_{g+1})$ ablesbare Zufallvariable.

Für das Ereignis

$E_n :=$ “keine Kollision bis (einschließlich) n ” gilt:

$$E_n = \{T > n\}$$

also insbesondere auch

$$\mathbf{P}(E_n) = \mathbf{P}(T > n).$$

$$\mathbf{P}(E_n) = \mathbf{P}(T > n)$$

In Abschnitt 3 hatten wir berechnet:

$$\mathbf{P}(E_n) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{g}\right) =: w(n, g) =: w(n)$$

$$w(n) = \mathbf{P}(T > n)$$

Wie bekommen wir daraus die **Verteilungsgewichte**

$$\mathbf{P}(T = n), n = 1, 2, \dots$$

der $\mathbb{N}$ -wertigen Zufallsvariablen T ?

$$w(n) = \mathbf{P}(T > n)$$

Aus der Additivität der Wahrscheinlichkeit erhalten wir

$$\mathbf{P}(T > n - 1) = \mathbf{P}(T = n) + \mathbf{P}(T > n).$$

Also:

$$\mathbf{P}(T = n) = w(n - 1) - w(n).$$

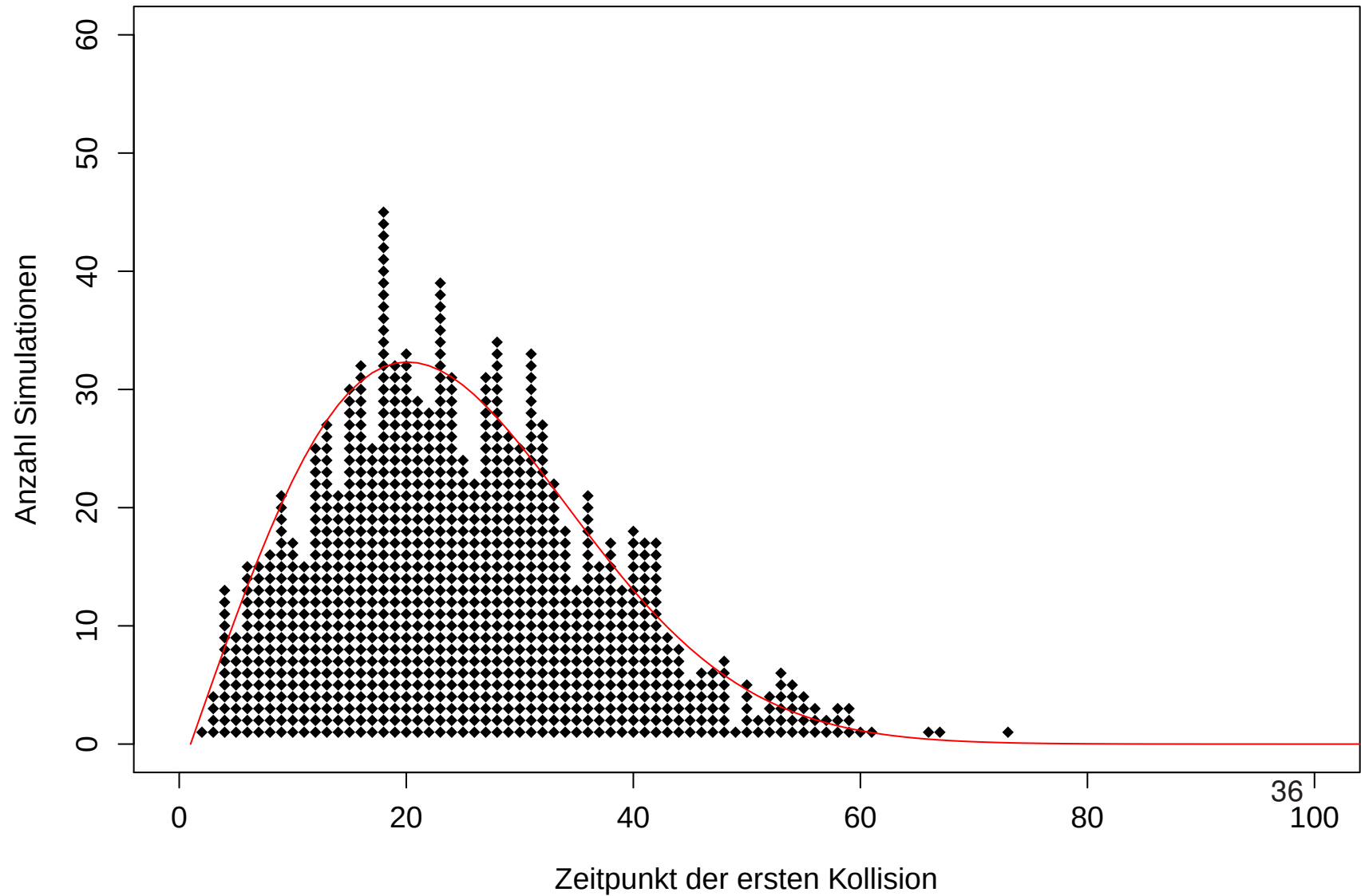
Die $\mathbb{N}$ -wertige Zufallsvariable T
(den Zeitpunkt der ersten Kollision) haben wir bekommen
als “Verarbeitung” des zufälligen Wahlprotokolls X :

$$T = h(X) \text{ mit passendem } h : S \rightarrow \mathbb{N}.$$

Aus der Gleichheit der Ereignisse
 $\{T = n\} = \{h(X) = n\} = \{X \in h^{-1}(n)\}$
ergibt sich die Gleichheit

$$\mathbf{P}(T = n) = \mathbf{P}(X \in h^{-1}(n)), \quad n \in \mathbb{N}.$$

Empirische Verteilung von T (1000 Simulationen)



Vorlesung 1b

Wahrscheinlichkeit von Kollisionen

bei der wiederholten rein zufälligen Wahl

aus endlich vielen möglichen Ausgängen

(Buch S. 1-5)

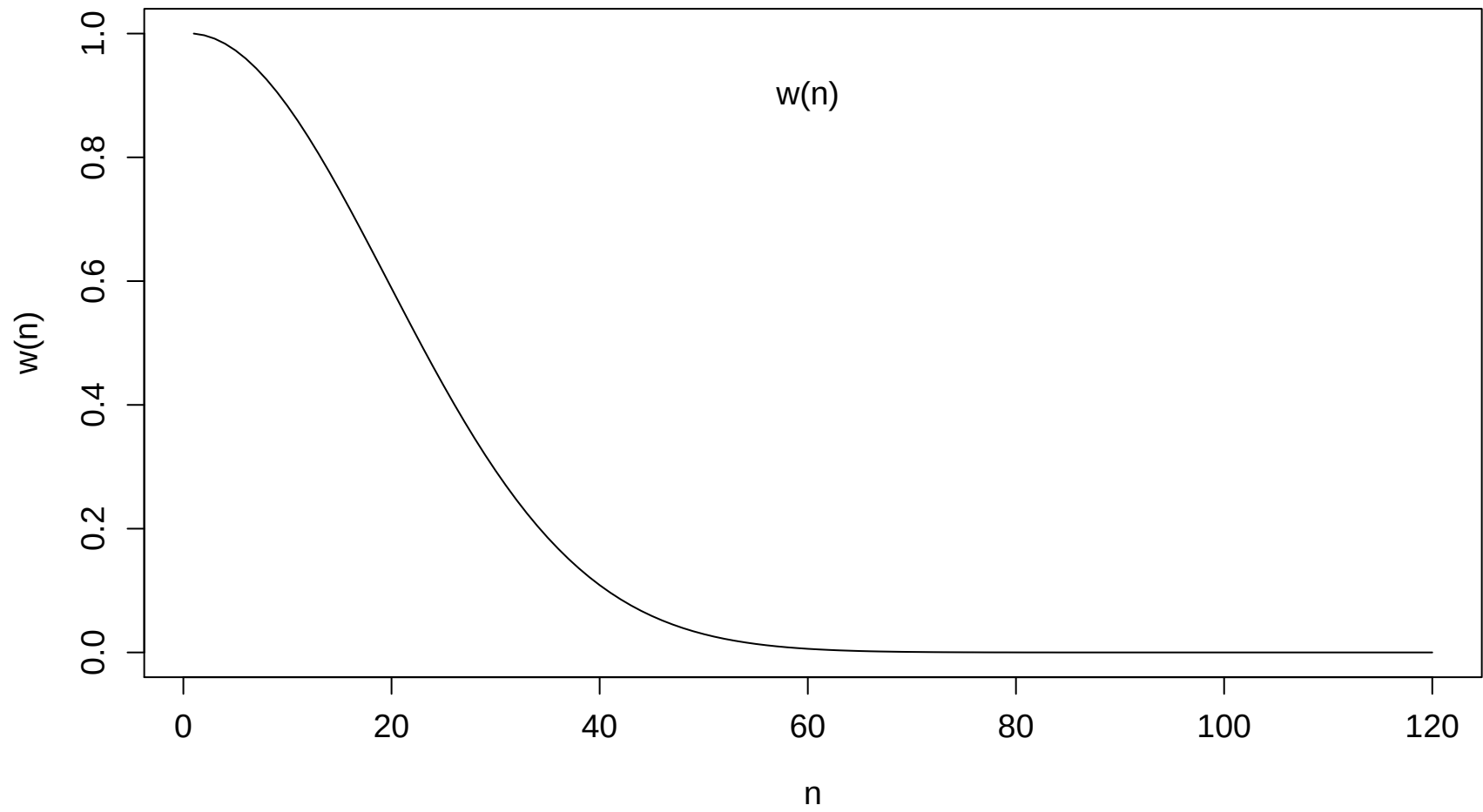
Teil Zwei:

Approximationen

der Wahrscheinlichkeit für Kollisionsfreiheit

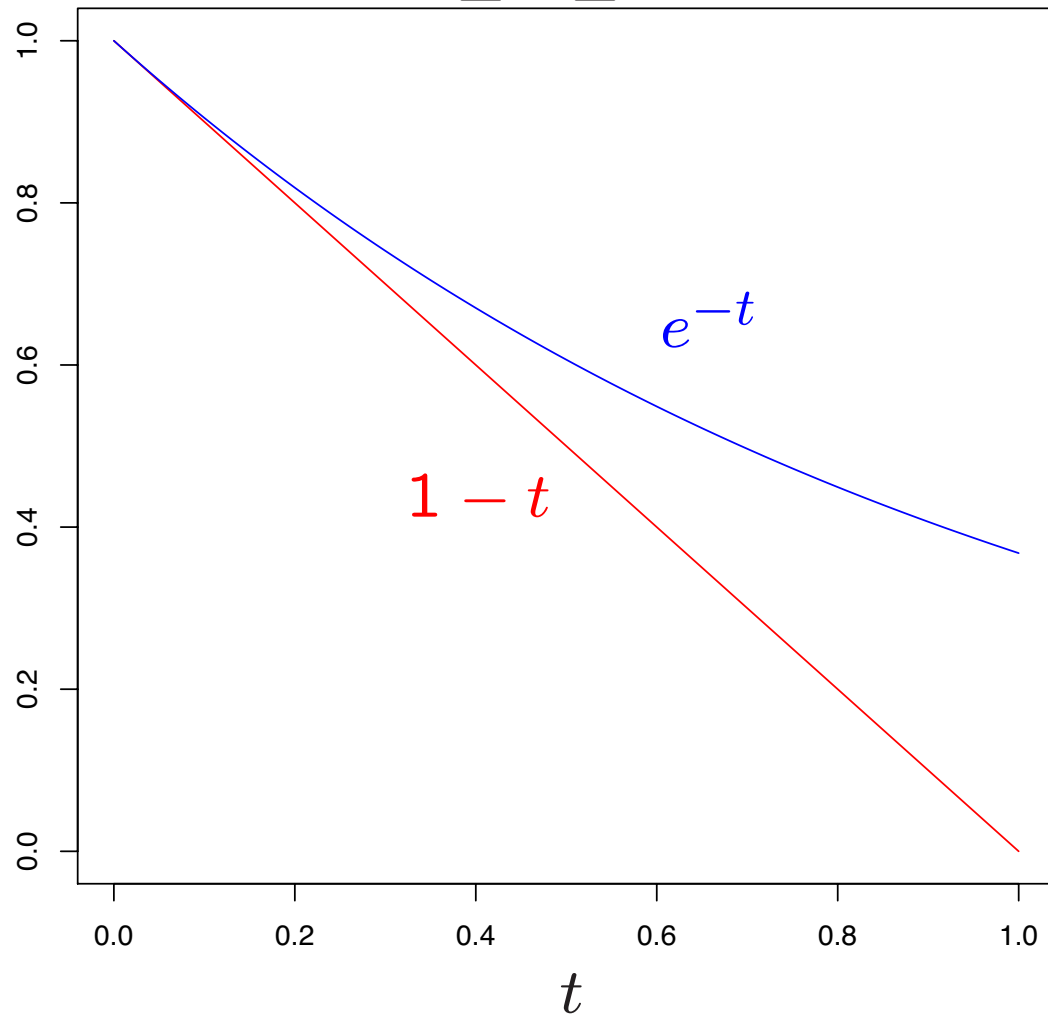
$$w(n, g) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{g}\right)$$

Wahrscheinlichkeit fuer Kollisionsfreiheit bei g=365

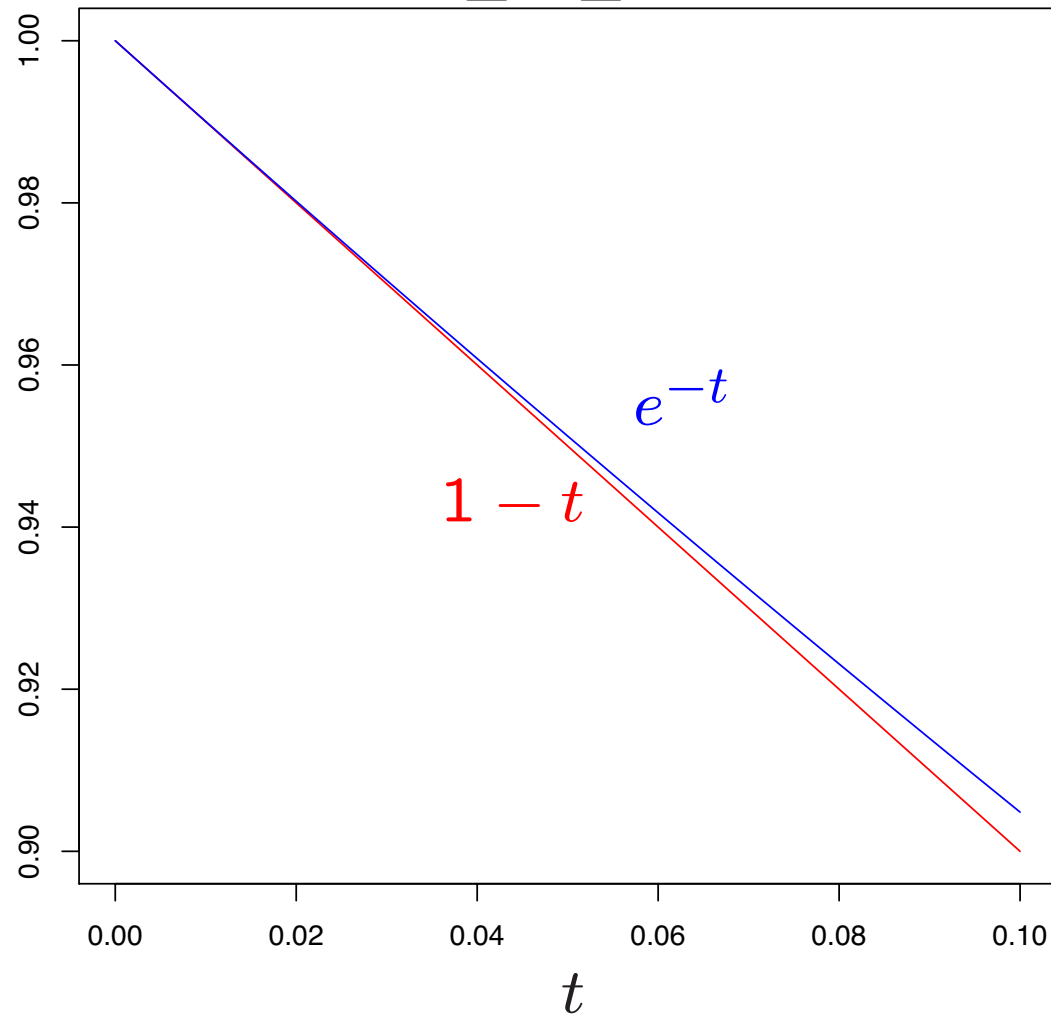


1. Approximation über die Linearisierung von $\exp$

für $0 \leq t \leq 1$:



für $0 \leq t \leq 0.1$:



Lokale lineare Approximation von $t \mapsto e^{-t}$ bei $t = 0$:

$$e^{-t} = 1 - t + o(t) \quad \text{für } t \rightarrow 0.$$

Dabei ist der Term $o(t)$ von kleinerer Ordnung als t , d.h. $\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0$.

Salopp geschrieben:

$$e^{-t} \approx 1 - t \quad \text{für kleine } t.$$

Damit bekommen wir für kleines $\frac{n}{g}$ die Approximation

$$\begin{aligned} w(n, g) &= \prod_{i=1}^{n-1} \left(1 - \frac{i}{g} \right) \approx \prod_{i=1}^{n-1} \exp \left(-\frac{i}{g} \right) \\ &= \exp \left(-\sum_{i=1}^{n-1} \frac{i}{g} \right) = \exp \left(-\frac{(n-1)n}{2g} \right). \end{aligned}$$

Also für kleines $\frac{n}{g}$, d.h. für $n \ll g$:

$$w(n, g) \approx \exp \left(-\frac{(n-1)n}{2g} \right) .$$

Gibt es auch eine Näherungsformel,
die brauchbar ist für alle $n < g$?

In der Tat!

2. Die Stirling-Approximation

(Buch S. 4-5)

$$\begin{aligned}
 w(n, g) &= \frac{g(g-1) \cdots (g-(n-1))}{g^n} \\
 &= \frac{g(g-1) \cdots (g-(n-1))}{g^n} \frac{(g-n)(g-n-1) \cdots 1}{(g-n)(g-n-1) \cdots 1}
 \end{aligned}$$

$$w(n, g) = \frac{g!}{g^n (g-n)!}$$

mit $k! := 1 \cdot 2 \cdots k$, lies: k -Fakultät

Die Stirling-Formel:

$$k! \approx \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$$

$$k! \approx \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$$

$$\begin{aligned} \frac{g!}{g^n(g-n)!} &\approx \frac{1}{g^n} \sqrt{\frac{g}{g-n}} \frac{g^g}{(g-n)^{g-n}} \frac{e^{g-n}}{e^g} \\ &= \sqrt{\frac{g}{g-n}} e^{-n} \left(\frac{g}{g-n}\right)^{g-n} \\ &= \frac{1}{\sqrt{1-\frac{n}{g}}} e^{-n} \left(1 - \frac{n}{g}\right)^{n-g} \end{aligned}$$

$$\frac{g!}{g^n(g-n)!} \approx \frac{1}{\sqrt{1 - \frac{n}{g}}} e^{-n} \left(1 - \frac{n}{g}\right)^{n-g}$$

Wir formen jetzt den Ausdruck

$$e^{-n} \left(1 - \frac{n}{g}\right)^{n-g}$$

weiter um. Ein Trick dabei ist es,

$$\left(1 - \frac{n}{g}\right)^{n-g} \text{ als } \exp\left((n-g) \ln\left(1 - \frac{n}{g}\right)\right)$$

zu schreiben

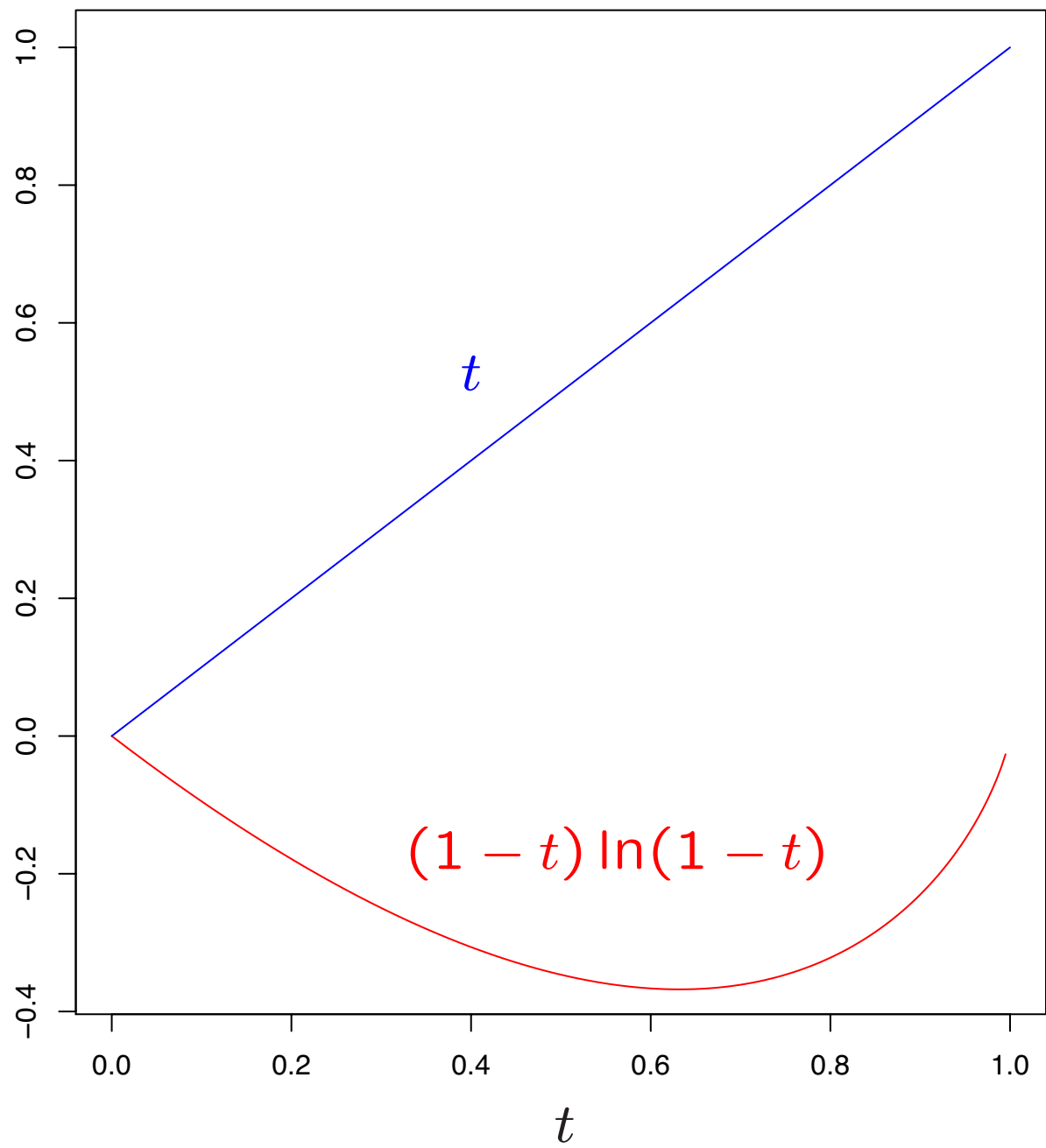
und dann die Beziehung $e^a e^b = e^{a+b}$ zu verwenden:

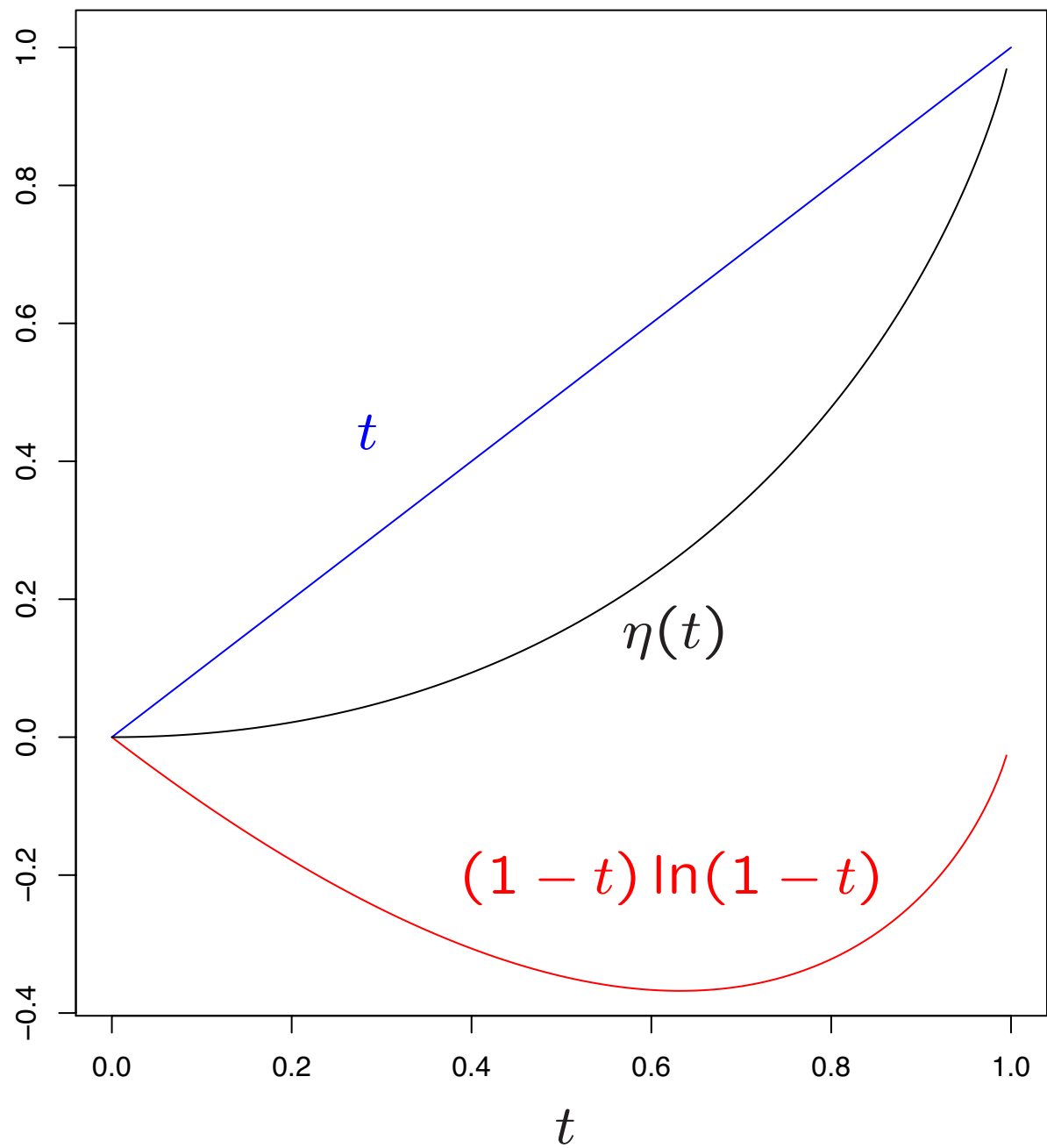
$$e^{-n} \left(1 - \frac{n}{g}\right)^{n-g} = \exp \left(-n + (n-g) \ln \left(1 - \frac{n}{g}\right) \right)$$

$$= \exp \left(-g \left(\frac{n}{g} + \left(1 - \frac{n}{g}\right) \ln \left(1 - \frac{n}{g}\right) \right) \right)$$

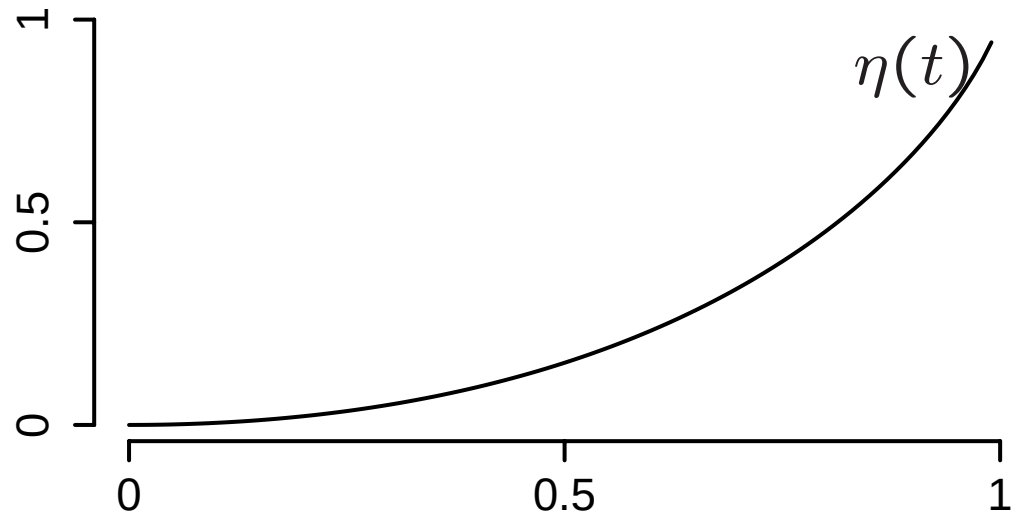
$$= \exp \left(-g \eta \left(\frac{n}{g} \right) \right)$$

mit $\eta(t) := t + (1-t) \ln(1-t), \quad 0 \leq t \leq 1.$





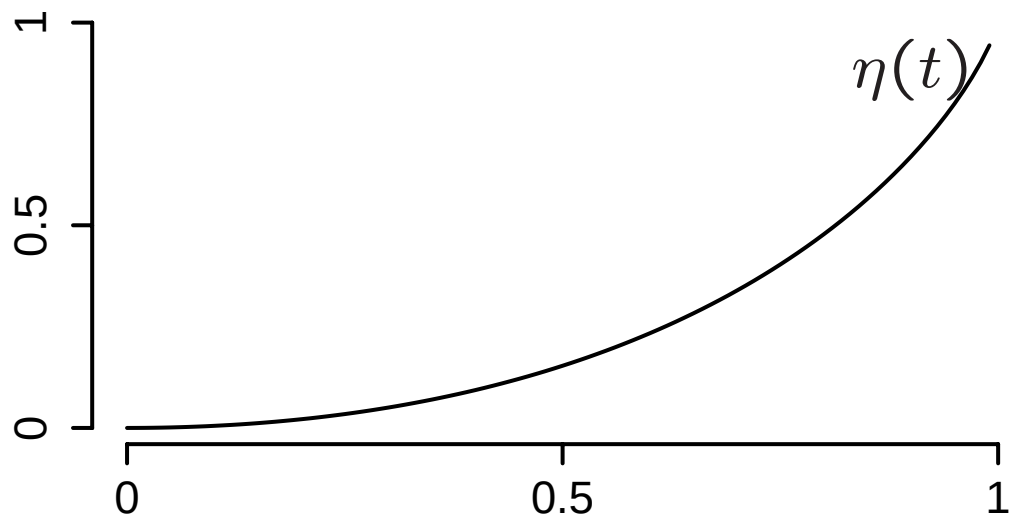
$$\eta(t) := t + (1 - t) \ln(1 - t), \quad 0 \leq t \leq 1.$$



$$\mathbf{P}(X \in A) \approx \frac{1}{\sqrt{1 - \frac{n}{g}}} \exp \left(-g \eta \left(\frac{n}{g} \right) \right)$$

Stirling-Approximation
der Wahrscheinlichkeit für Kollisionsfreiheit

$$\eta(t) := t + (1 - t) \ln(1 - t), \quad 0 \leq t \leq 1.$$



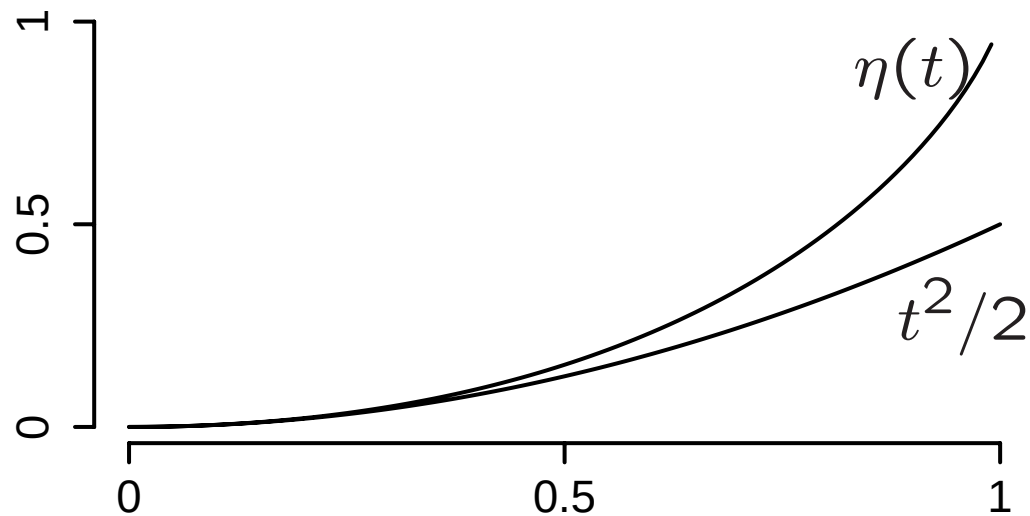
$$w(n, g) \approx \frac{1}{\sqrt{1 - \frac{n}{g}}} \exp \left(-g \eta \left(\frac{n}{g} \right) \right)$$

Die Folien 19 bis 23 wurden in der Vorlesung nicht besprochen. Ihre Lektüre ist für das Weitere nicht notwendig, kann aber (auch in der Remeinszenz an Mathe 1 und 2) genüsslich sein.

Für $n \ll g$, also $t := \frac{n}{g} \ll 1$,
können wir

- (i) $\eta(t)$ quadratisch um $t = 0$ approximieren
und
- (ii) den Vorfaktor $\frac{1}{\sqrt{1-\frac{n}{g}}}$ mittels $e^{-t} \approx 1 - t$ approximieren:

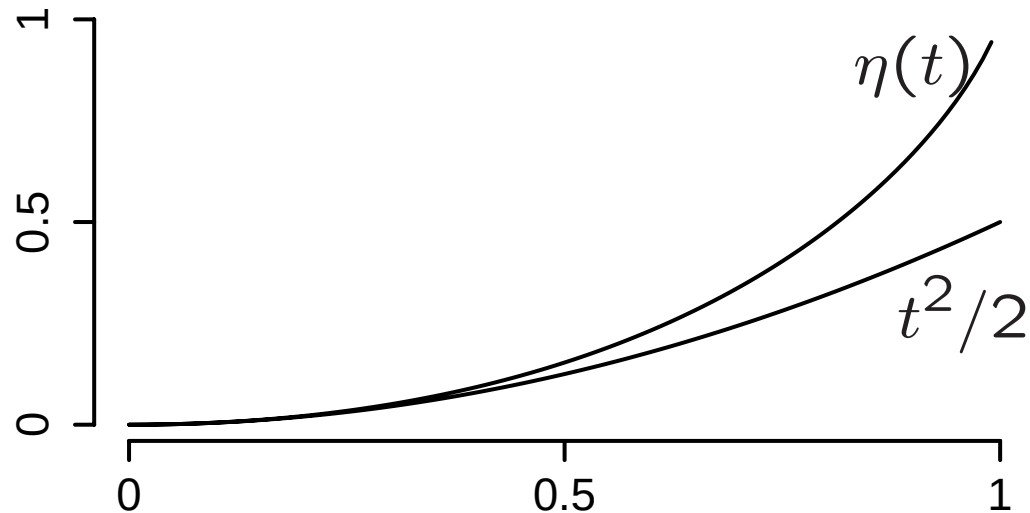
$$\eta(t) := t + (1 - t) \ln(1 - t), \quad 0 \leq t \leq 1.$$



$$\ln(1 - t) = -t - \frac{t^2}{2} - \frac{t^3}{3} - \dots$$

$$(1 - t) \ln(1 - t) = -t + \frac{t^2}{2} + o(t^2) \text{ für } t \rightarrow 0$$

$$\eta(t) = \frac{t^2}{2} + o(t^2) \text{ für } t \rightarrow 0$$



$\frac{1}{2}t^2$ ist die quadratische Approximation von $\eta(t)$ um $t = 0$.

Für $n \ll g$ (wie z. B. für $n = 25$, $g = 365$) ist also

$$\eta\left(\frac{n}{g}\right) \approx \frac{1}{2} \cdot \left(\frac{n}{g}\right)^2.$$

Jetzt zum Vorfaktor $\frac{1}{\sqrt{1 - \frac{n}{g}}}$:

$$\frac{1}{\sqrt{1 - t}} = (1 - t)^{-1/2} \approx (e^{-t})^{-1/2} = e^{t/2}.$$

Insgesamt ist also für kleine t

$$\frac{1}{\sqrt{1 - t}} e^{-g\eta(t)} \approx e^{t/2} e^{-gt^2/2}$$

Für $t := \frac{n}{g} \ll 1$ ergibt sich die **Stirling+Taylor-Approximation**:

$$w(n, g) \approx \exp\left(\frac{n}{2g}\right) \exp\left(-\frac{n^2}{2g}\right) = \exp\left(-\frac{n(n-1)}{2g}\right)$$

Man beachte:

Die Stirling+Taylor Approximation

$$\mathbf{P}(X \in A) \approx \exp \left(-\frac{n(n-1)}{2g} \right)$$

ist identisch mit der

Approximation über die Linearisierung von exp.

Im Buch haben wir hier den Faktor $\sqrt{1 - n/g}$
in der Stirling-Approximation vernachlässigt, d.h. gleich 1 gesetzt.
Dadurch erhielten wir im Buch die etwas weniger genaue Näherung

$$w(n, g) \approx \exp \left(-\frac{n^2}{2g} \right).$$

Fazit:

Für $n < g$ ist $w(n, g) \approx \frac{1}{\sqrt{1 - \frac{n}{g}}} \exp \left(-g \eta \left(\frac{n}{g} \right) \right)$

mit $\eta(t) = t + (1 - t) \ln(1 - t)$, $0 \leq t \leq 1$

(Stirling-Approximation).

Für $n \ll g$ ist $w(n, g) \approx \exp \left(-\frac{n(n-1)}{2g} \right)$

(Approximation über die Linearisierung von exp)

Für $n^2 \ll g$ ist $w(n, g) \approx 1$.

Beispiel: $n = 25, g = 365$:

$$\mathbf{P}(X \in A) = \frac{g(g-1) \cdots (g-n+1)}{g^n} = 0.431300$$

$$\frac{1}{\sqrt{1 - \frac{n}{g}}} \exp \left(-g \eta \left(\frac{n}{g} \right) \right) = 0.431308$$

$$\exp \left(-\frac{n(n-1)}{2g} \right) = 0.4396$$

Vorlesung 2a

Diskret uniform verteilte Zufallsvariable

(Buch S. 6-11)

Erinnerung und Auftakt

Sei S eine endliche Menge.

Eine Zufallsvariable X heißt *uniform verteilt auf S* ,
wenn

$$\mathbf{P}(X = a) = \frac{1}{\#S} \quad \text{für alle } a \in S.$$

Damit beschreibt X eine *rein zufällige Wahl* aus S .

Beispiel aus Vorlesung 1b:

$$S = \{1, 2, \dots, g\}^n$$

$X :=$ rein zufällige $1 \dots g$ - Folge der Länge n .

Eine auf einem endlichen Wertebereich
uniform verteilte Zufallsvariable nennt man auch
diskret uniform verteilt.

Heute lernen wir drei weitere Beispiele
von diskret uniform verteilten Zufallsvariablen kennen:

1. Rein zufällige Permutation
2. Rein zufällige k -elementige Teilmenge
3. Uniform verteilte Besetzung

Bei der Gelegenheit erarbeiten wir auch ein paar
Hilfen fürs Abzählen.

Teil 1

Rein zufällige Permutation

(Buch S. 6-8)

1. Elementares

Eine *Permutation* von $1, \dots, n$
ist eine bijektive Abbildung der Menge $\{1, \dots, n\}$ auf sich.

Z. B. mit $n = 7$

1	2	3	4	5	6	7
↓	↓	↓	↓	↓	↓	↓
5	2	7	3	1	4	6

Wie wahrscheinlich ist es,
dass eine rein zufällige Permutation
genau **so** ausfällt?

Wieviele Permutationen von $1, \dots, n$ gibt es?

n Möglichkeiten für das Bild von 1

mal $(n - 1)$ Möglichkeiten für das Bild von 2

mal $(n - 2)$ Möglichkeiten für das Bild von 3

$\dots$

$$= n(n - 1)(n - 2) \cdots 2 \cdot 1 = n!$$

Sei X eine rein zufällige Permutation von $1, \dots, n$,

d.h. eine Zufallsvariable, deren Zielbereich
 $S :=$ die Menge aller Permutationen von $1, \dots, n$
ist, und die auf S uniform verteilt ist.

Für alle Elemente $a \in S$ gilt also:

$$\mathbf{P}(X = a) = \frac{1}{n!}$$

2. Zufällige Permutation und zufälliges Ziehen

Wie kann man sich eine rein zufällige Permutation
entstanden denken?

Zum Beispiel: als Folge der gezogenen Nummern
beim n -maligen rein zufälligen *Ziehen ohne Zurücklegen*
aus $\{1, 2, \dots, n\}$.

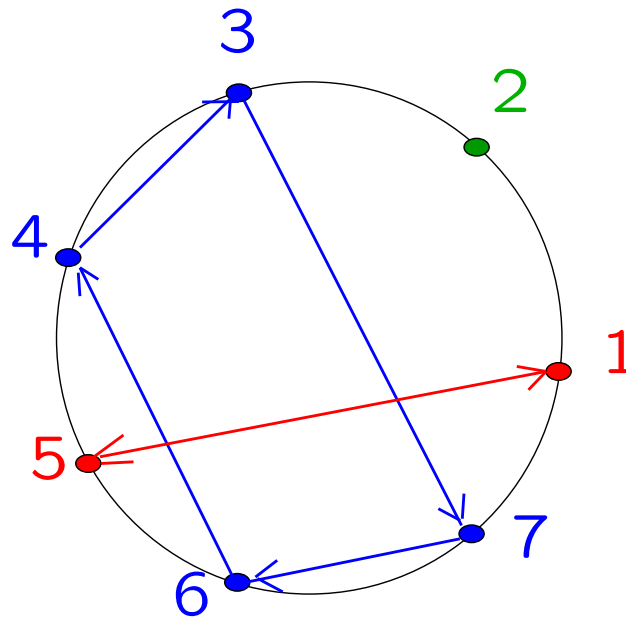
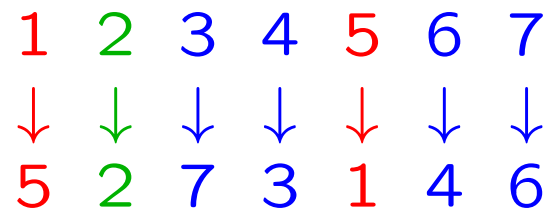
Szenario: eine stets ideal durchmischte Urne
mit anfangs n Kugeln, beschriftet mit den Nummern $1, \dots, n$.

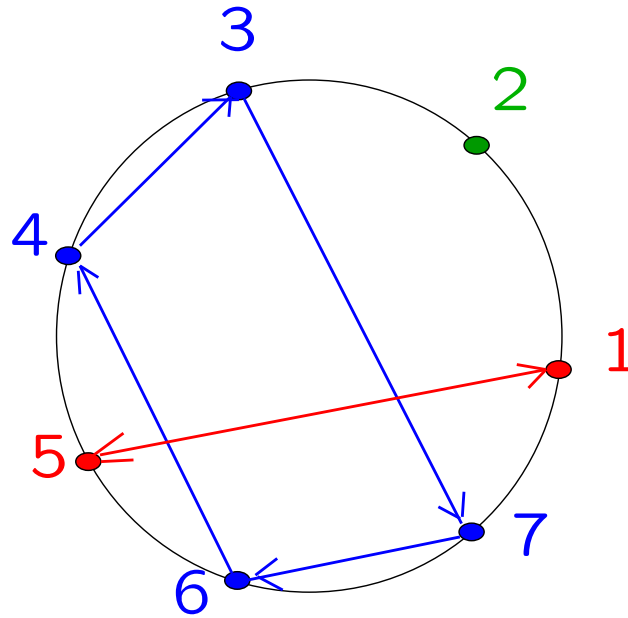
Ziehe sukzessive **ohne Zurücklegen** alle n Kugeln
und notiere die gezogenen Nummern in Reihenfolge.

3. Zyklendarstellung einer Permutation

Jede Permutation zerfällt in **Zyklen**

Beispiel:





Die Länge des Zyklus, der die Eins enthält, ist hier
zwei.

Sei X eine rein zufällige Permutation von $1, \dots, 7$,

Wie wahrscheinlich ist es, dass **der die 1 enthaltende Zykel**
genau die Länge 3 hat?

Wieviele Permutationen von $1, \dots, 7$ gibt es, bei denen
der die 1 enthaltende Zykel genau die Länge 3 hat?

Es gibt davon $6 \cdot 5 \cdot 1 \cdot 4!$ Stück (warum?)

Also ist die gefragte W'keit: $\frac{6!}{7!} = \frac{1}{7}$.

Jetzt allgemein:

Für eine Permutation $a \in S$ bezeichne

$$h(a)$$

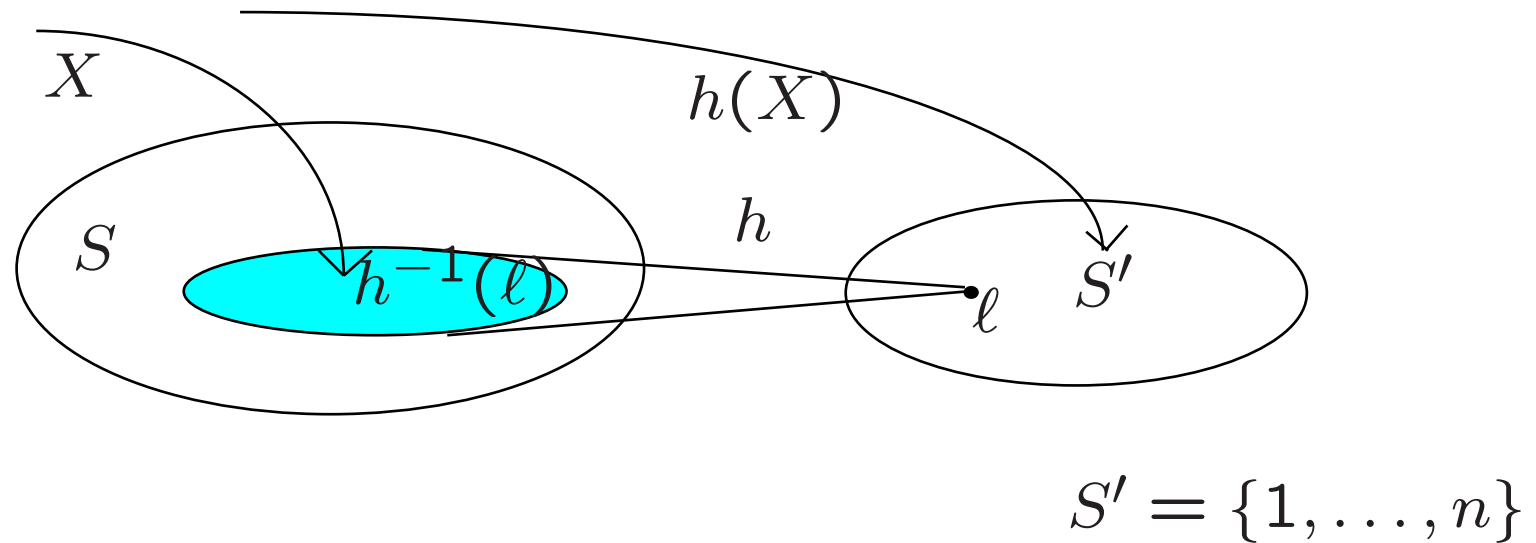
die Länge des Zyklus von a , der die Eins enthält.

Sei X eine rein zufällige Permutation von $\{1, \dots, n\}$,

also eine rein zufällige Wahl aus S ,

und sei $\ell \in \{1, 2, \dots, n\}$.

$$\mathbf{P}(h(X) = \ell) = ?$$



Wieviele Permutationen $a \in S$ gibt es mit $h(a) = \ell$?

$$A := \{a \in S : h(a) = \ell\}$$

$$\#A = ?$$

$$A = \{a \in S : a(1) \neq 1, a^2(1) \neq 1, \dots, \\ a^{\ell-1}(1) \neq 1, a^\ell(1) = 1\}$$

$$\begin{aligned} \#A &= (n-1)(n-2) \cdots (n-(\ell-1)) \cdot 1 \cdot (n-\ell) \cdots 1 \\ &= (n-1)! \end{aligned}$$

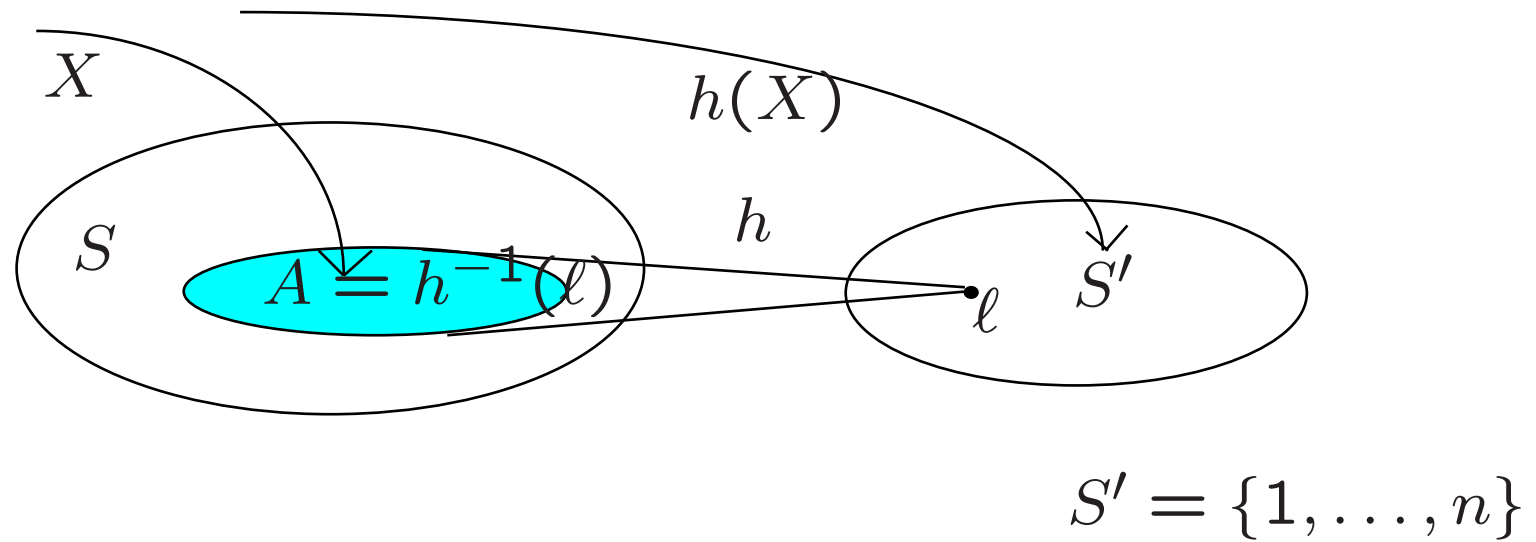
$$\#A = (n - 1)!, \quad \#S = n!$$

$$\mathbf{P}(X \in A) = \frac{\#A}{\#S}$$

$$= \frac{(n - 1)!}{n!}$$

$$= \frac{1}{n}$$

$$\mathbf{P}(X \in A) = \frac{1}{n}$$



$$A = \{a \in S : h(a) = \ell\}$$

$$\{X \in A\} = \{h(X) = \ell\}$$

$$\mathbf{P}(h(X) = \ell) = \frac{1}{n}$$

$$\mathbf{P}(h(X) = \ell) = \frac{1}{n}, \quad \ell = 1, \dots, n.$$

Fazit:

Die Länge desjenigen Zyklus
einer rein zufälligen Permutation von $1, \dots, n$,
der die Eins enthält,
ist uniform verteilt auf $\{1, \dots, n\}$.

Vorlesung 2a

Diskret uniform verteilte Zufallsvariable

Teil 2:
Rein zufällige Teilmenge
einer festen Größe

(Buch S. 9-10)

1. Die Anzahl der k -elementigen Teilmengen von $\{1, \dots, n\}$: **der Binomialkoeffizient.**

Sei $k \leq n$.

Jetzt sei S

die Menge aller k -elementigen Teilmengen von $\{1, \dots, n\}$.

Wieviele k -elementige Teilmengen von $\{1, \dots, n\}$ gibt es?

Wieviele Möglichkeiten gibt es, aus n Personen ein k -köpfiges Komitee ohne Reihung zu bilden?

Wird “nach der Reihe” ausgewählt, dann gibt es $n(n - 1) \cdots (n - (k - 1))$ mögliche Wahlprotokolle. Auf die Reihenfolge kommt es am Ende nicht an, somit führen jeweils $k!$ dieser Wahlprotokolle auf dieselbe k -elementige Teilmenge.

Also:

$$\#S = \frac{n(n-1) \cdots (n-k+1)}{k!}$$

$$= \frac{n!}{k!(n-k)!} =: \binom{n}{k}$$

Binomialkoeffizient „n über k“

die Anzahl der Möglichkeiten für „k aus n“

Beispiel: Binomischer Lehrsatz:

$$(x + y)^n = \underbrace{(x + y)(x + y) \cdots (x + y)}_{n\text{-mal}} = ?$$

Multipliziert man aus, dann ergeben sich 2^n Summanden
(entsprechend den 2^n x y -Folgen der Länge n).

Jeder dieser Summanden ist von der Form $x^k y^{n-k}$.

Für festes k gibt es $\binom{n}{k}$ Möglichkeiten,
 k aus den n Faktoren $(x + y)$ auszuwählen,
bei denen x zum Zug kommt.

$$\text{Also } (x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

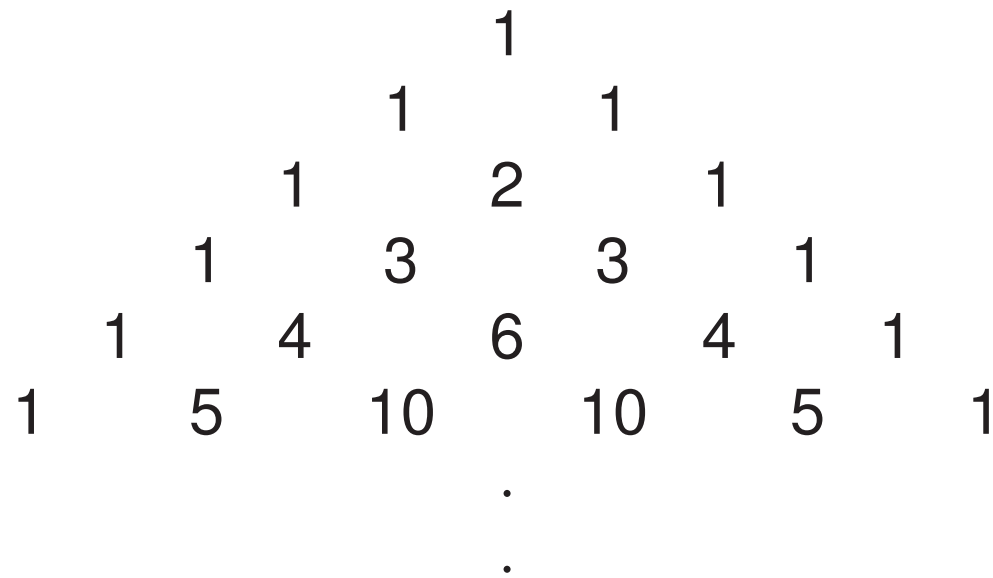
$\binom{n}{k}$... die Anzahl der Möglichkeiten für „ k aus n “

Rekursion:
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Rekursion:
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Interpretation: Anzahl der Möglichkeiten,
aus n **Hessen** und einem **Bayern**
ein $k+1$ köpfiges Komitee auszuwählen.
Entweder **der Bayer ist im Komitee...**
oder **er ist nicht im Komitee...**

Pascal'sches Dreieck



Rekursion:
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

	1	2	3	4	5	6	7	8	9	10
1	1	3	6	10	15	21	28	36	45	55
2	1	3	6	10	15	21	28	36	45	55
3	1	3	6	10	15	21	28	36	45	55
4	1	3	6	10	15	21	28	36	45	55
5	1	3	6	10	15	21	28	36	45	55
6	1	3	6	10	15	21	28	36	45	55
7	1	3	6	10	15	21	28	36	45	55
8	1	3	6	10	15	21	28	36	45	55
9	1	3	6	10	15	21	28	36	45	55
10	1	3	6	10	15	21	28	36	45	55

Rangs parallèles

Triangle Arithmétique

Rangs perpendiculaires

“La nombre de chaque cellule est égal a celui de la cellule
qui la précède dans son rang perpendiculaire,
plus à celui de la cellule qui la précède dans son rang parallèle.”
(Pascal 1654)

2. Rein zufällige Teilmengen

Sei $0 \leq k \leq n$

und sei Y eine rein zufällige k -elementige Teilmenge
von $\{1, \dots, n\}$.

Wie wahrscheinlich ist das Ereignis $\{Y = \{1, \dots, k\}\}$?

Der Zielbereich von Y ist

$$S := \{t : t \subset \{1, \dots, n\}, \#t = k\} ,$$

die Menge der k -elementigen Teilmengen von $\{1, \dots, n\}$.

Wir haben gesehen:

$$\#S = \binom{n}{k}$$

Fazit:

$$\mathbf{P}(Y = \{1, \dots, k\}) = \frac{1}{\binom{n}{k}} .$$

3. Ein Zusammenhang mit rein zufälligen Permutationen

Wie bekommt man eine rein zufällige k -elementige Teilmenge
aus einer rein zufälligen Permutation?

Fakt (für $k \leq n$):

Ist $X = (X_1, \dots, X_n)$

eine rein zufälligen Permutation von $1, \dots, n$,

dann ist $\{X_1, \dots, X_k\}$

eine rein zufällige k -elementige Teilmenge von $\{1, \dots, n\}$.

Anders gesagt:

Beim **Ziehen ohne Zurücklegen**
führen die ersten k Züge auf eine
rein zufällige Teilmenge des anfänglichen Reservoirs.

Noch eine Möglichkeit zum Erzeugen
einer rein zufälligen k -elementigen Teilmenge:

Ziehe sukzessive ohne Zurücklegen aus einer Urne
mit k roten und $n - k$ blauen Kugeln.

Notiere die **Nummern $X_1, \dots, X_k$ der Züge**,
bei denen eine rote Kugel gezogen wird.

Dann ist $\{X_1, \dots, X_k\}$
eine rein zufällige k -elementige Teilmenge von $\{1, \dots, n\}$.

Vorlesung 2a

Diskret uniform verteilte Zufallsvariable

Teil 3: Besetzungszahlen

(vgl. Buch S. 10-11)

1. Begriffsbildung

Sei $a = (a_1, \dots, a_n)$ eine $1, \dots, g$ - Folge der Länge n .

Vorstellung: Objekt Nr. i kommt auf Platz a_i .

Wie oft wird laut Protokoll a der Platz j besetzt?

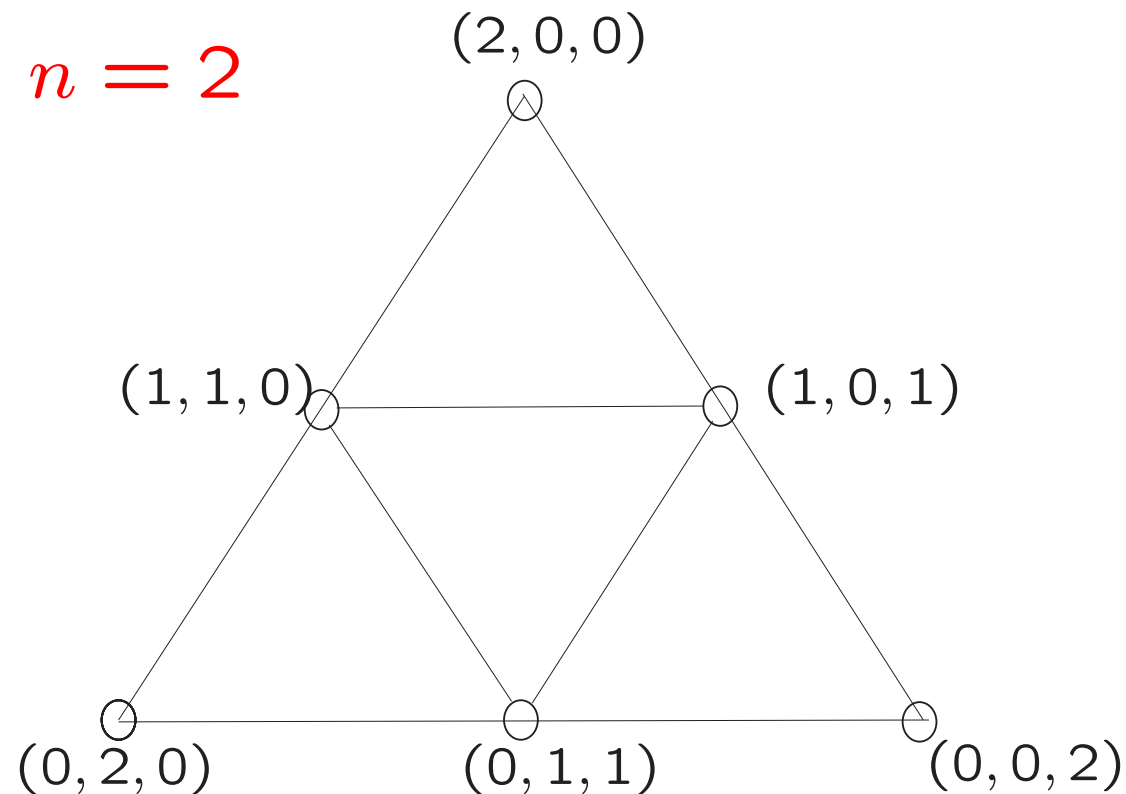
Anders gesagt: Für wieviele i ist $a_i = j$?

$$b_j(a) := \#\{i : a_i = j, 1 \leq i \leq n\}$$

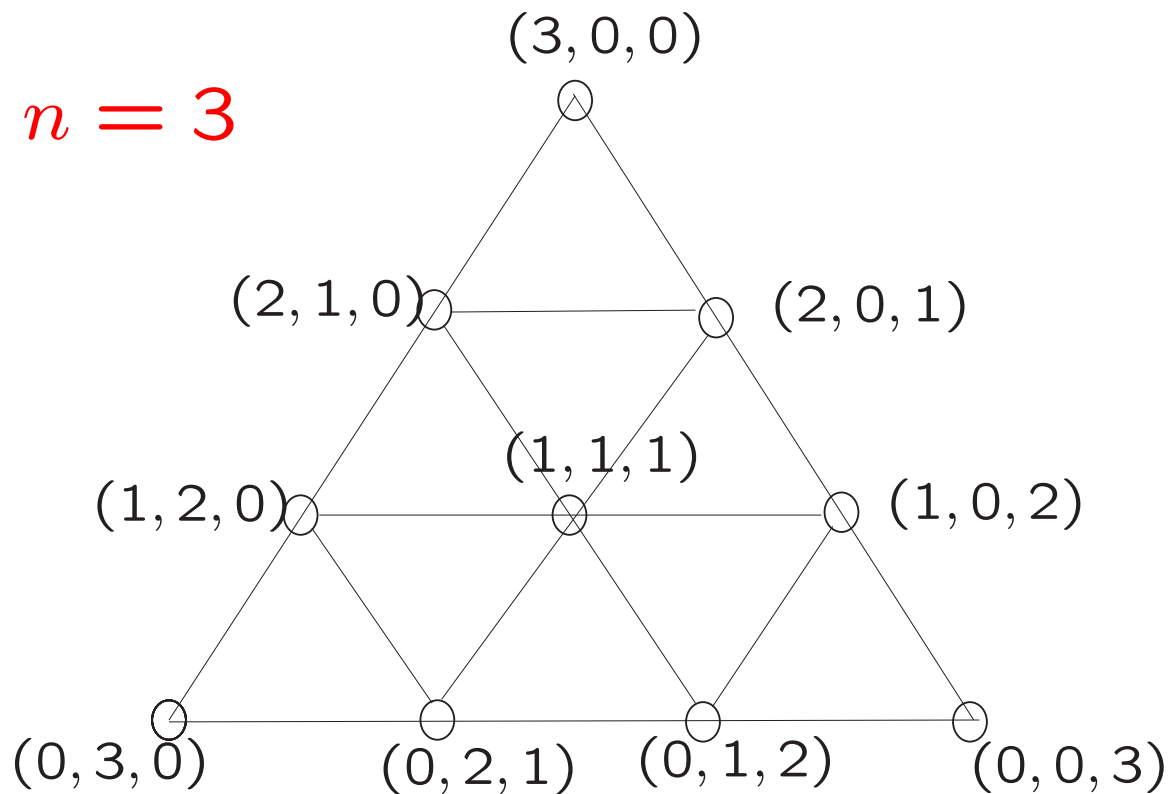
Das g -tupel der Besetzungszahlen $b_j(a)$ nennen wir kurz
“die durch a induzierte **Besetzung**”.

In der Vorstellung des
Setzens von n Objekten auf g mögliche Plätze
gibt sie an, wieviele Objekte auf welchem Platz landen
(und unterscheidet nicht, welche Objekte das sind).

Im Fall $g = 3$ gibt es eine nette Darstellung der Menge der Besetzungen mittels des sogenannten de Finetti-Dreiecks:



Im Fall $g = 3$ gibt es eine nette Darstellung der Menge der Besetzungen mittels des sogenannten de Finetti-Dreiecks:



2. Besetzungszahlen bei wiederholter rein zufälliger Platzwahl

Machen wir uns ein Bild von der zufälligen Besetzung,
die aus einem auf $\{1, \dots, g\}^n$ uniform verteilten X entsteht.

Das war das Szenario aus Vorlesung 1b:

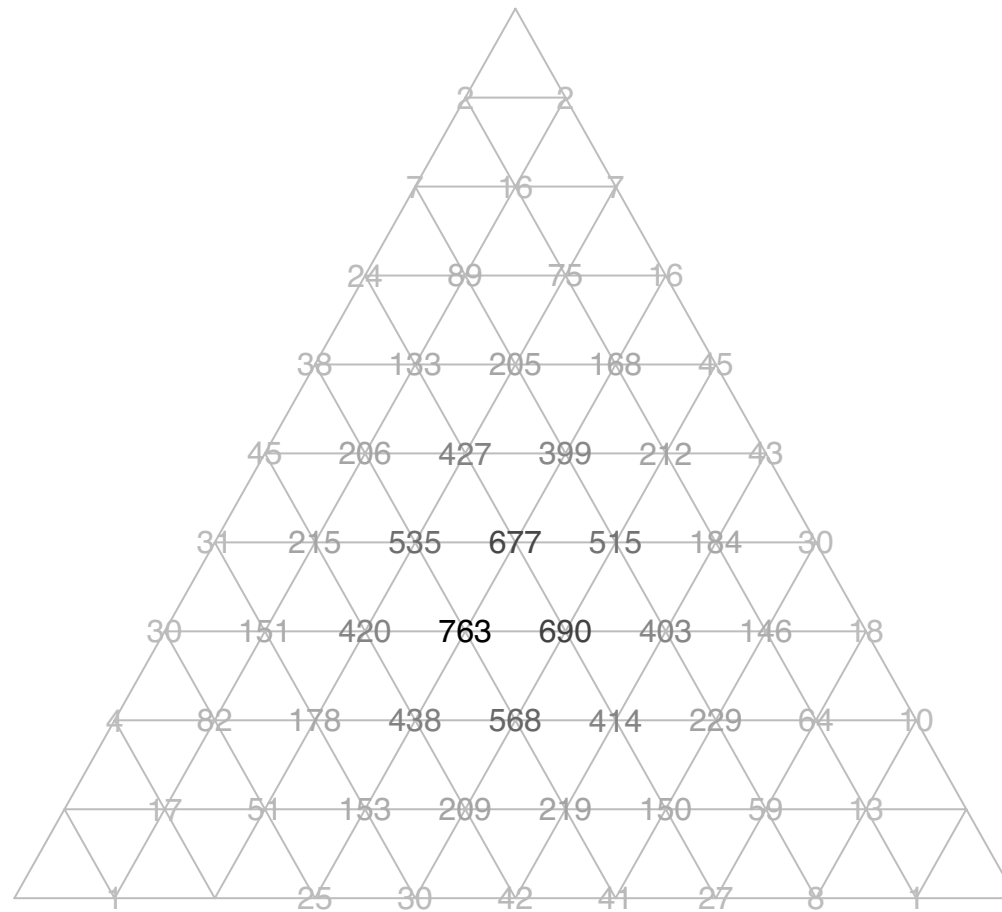
n -malige rein zufällige Wahl
aus g möglichen Plätzen.

Es folgt das Ergebnis einer Simulation für $n = 10$ und $g = 3$.

Die Ecken des *de Finettii-Dreiecks* auf der nächsten Folie
entsprechen den Besetzungen

$(10, 0, 0)$ (oben), $(0, 10, 0)$ (links) und $(0, 0, 10)$ (rechts).

Häufigkeiten der Besetzungen bei 10000 Wiederholungen



Wir sehen aus der Simulation:

Die Verteilung dieser zufälligen Besetzung ist
(bei weitem) nicht uniform.

Wir kommen auf diese Verteilung
in der nächsten Vorlesung zurück.

Zum Kontrast betrachten wir jetzt die

3. Uniform verteilte Besetzung

von g Plätzen mit n Objekten:

Der Zielbereich ist

$$S_{n,g} := \{b = (b_1, \dots, b_g) : b_j \in \mathbb{N}_0, b_1 + \dots + b_g = n\}$$

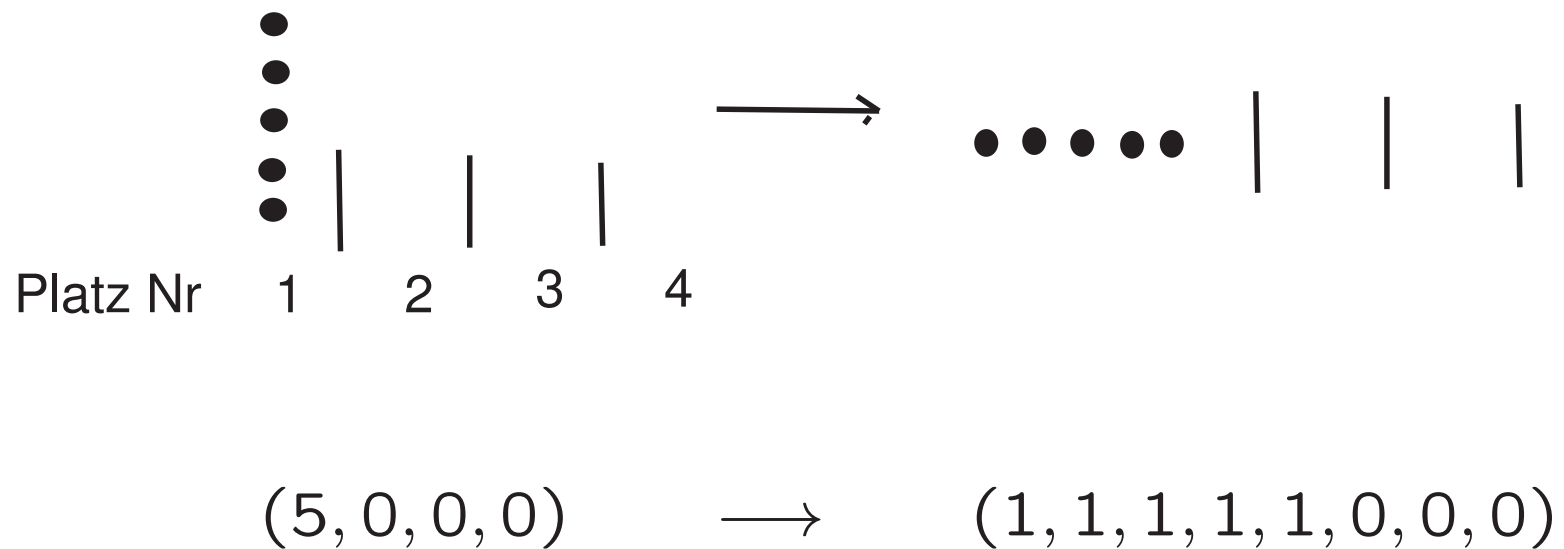
b ist ein g -tupel von Besetzungszahlen, kurz: eine Besetzung.

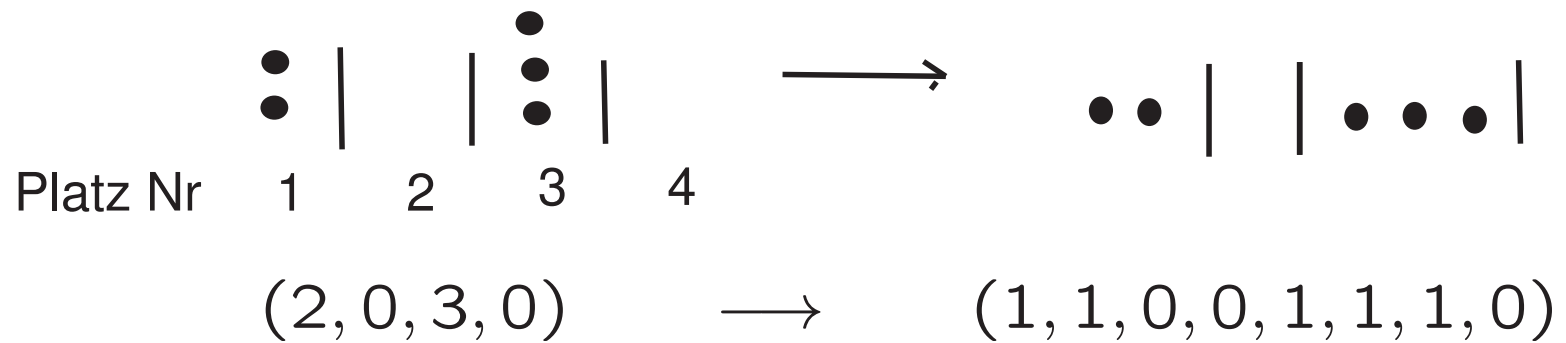
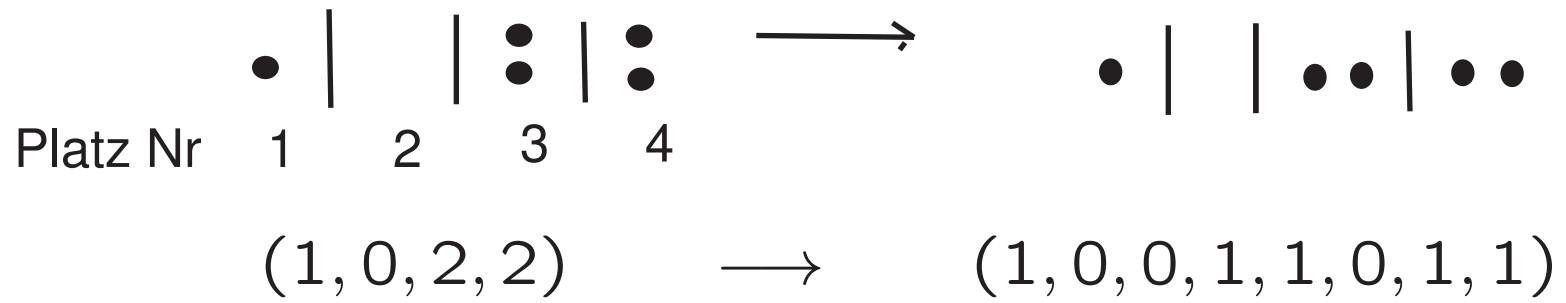
$$S_{n,g} := \{b = (b_1, \dots, b_g) : b_j \in \mathbb{N}_0, b_1 + \dots + b_g = n\}$$

$$\#S_{n,g} = ?$$

Hier hilft ein hübscher Trick, $S_{n,g}$ anders darzustellen.

Beispiele: $g = 4$ Plätze, $n = 5$ Objekte





Fakt:

$$h(b_1, b_2, \dots, b_g) := \underbrace{1 \dots 1}_{b_1\text{-mal}} 0 \underbrace{1 \dots 1}_{b_2\text{-mal}} 0 \dots 0 \underbrace{1 \dots 1}_{b_g\text{-mal}}$$

ist eine bijektive Abbildung von $S_{n,g}$ nach

$S :=$ Menge der 01-Folgen der Länge $n + g - 1$
mit genau n Einsen

$$h(b_1, b_2, \dots, b_g) := \underbrace{1 \dots 1}_{b_1\text{-mal}} 0 \underbrace{1 \dots 1}_{b_2\text{-mal}} 0 \dots 0 \underbrace{1 \dots 1}_{b_g\text{-mal}}$$

Voriges Beispiel: $n = 5$, $g = 4$:

$$(b_1, \dots, b_4) = (2, 0, 3, 0)$$

$$h(2, 0, 3, 0) = 11001110$$

Der zweite und der vierte Block aus Einsen sind hier leer.

$$h(b_1, b_2, \dots, b_g) := \underbrace{1 \dots 1}_{b_1\text{-mal}} 0 \underbrace{1 \dots 1}_{b_2\text{-mal}} 0 \dots 0 \underbrace{1 \dots 1}_{b_g\text{-mal}}$$

Die Länge des j -ten Blocks aus Einsen steht für die Anzahl der Objekte auf Platz j .

Die Blöcke aus Einsen sind durch Nullen getrennt. Die Nullen fungieren als “Trennwände” zwischen den g Plätzen, insgesamt gibt es $g - 1$ solche Trennwände.

$S :=$ Menge der 01-Folgen der Länge $n + g - 1$
mit genau n Einsen

$$\#S = ?$$

$$\#S = \binom{n + g - 1}{n}$$

Also (wegen der Bijektion mittels h) auch:

$$\#S_{n,g} = \binom{n + g - 1}{n}$$

Eine Möglichkeit zum Erzeugen
einer uniform verteilten Besetzung:

Ziehe aus einer Urne mit n weißen und $g - 1$ schwarzen
Kugeln sukzessive ohne Zurücklegen.

Notiere 0 beim Zug einer schwarzen
und 1 beim Zug einer weißen Kugel.

Erzeuge so ein rein zufälliges Element aus S .
Übersetze dieses (mit der Umkehrung von h)
in eine rein zufällige Besetzung.

Hier ist noch ein weiteres Zufallsexperiment, das auf eine **uniform verteilte Besetzung** $(Z_1, \dots, Z_g)$ führt – dass das tatsächlich so ist, werden wir im Kapitel über mehrstufige Zufallsexperimente unter der Überschrift *Pólya-Urne* sehen.

Das Experiment (unter dem Stichwort *Wo Tauben sind, fliegen Tauben zu*) läuft wie folgt:

Anfangs sind g Personen auf der Bühne, nummeriert mit $1, \dots, g$.

Sukzessive kommen Ankömmlinge dazu.

Ankömmling 1 wählt rein zufällig eine der g Personen und stellt sich zu ihr, damit steigt die Zahl der Personen auf der Bühne auf $g + 1$.

Ankömmling 2 wählt rein zufällig eine der $g + 1$ Personen und stellt sich zu ihr, u.s.w.

Betrachtet wird dann (bei insgesamt n Ankömmlingen) das g -Tupel $(Z_1, \dots, Z_g)$, wobei Z_j die Anzahl der Ankömmlinge ist, die bei der Person Nr. j stehen.

Vorlesung 2b

Diskrete Zufallsvariable und ihre Verteilungen

mit den Beispielen

Anzahl der Erfolge beim n -fachen p -**Münzwurf**

und

Besetzungen beim n -fachen $(p_1, \dots, p_g)$ -**Würfeln**

Teil 1: Grundbegriffe

(Buch S. 20-21)

Abzählbare Additivität von Wahrscheinlichkeiten

Verteilung und Verteilungsgewichte

Zufällige Paare: gemeinsame Verteilung und Projektionen

Weiterverarbeitung von Zufallsvariablen und

Transport von Verteilungen

Bisher hatten wir uns mit Zufallsvariablen beschäftigt, deren Wertebereich S endlich war.

Die (schon in Vorlesung 1b formulierten)

zwei Grundregeln

für Wahrscheinlichkeiten lauteten für diesen Fall:

Normiertheit auf Eins:

$$\mathbf{P}(X \in S) = 1 .$$

(Folgerung aus der) Additivität:

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a), \quad A \subset S$$

Diese beiden Regeln behalten ihren Sinn, wenn der Wertebereich nicht endlich, sondern abzählbar unendlich ist.

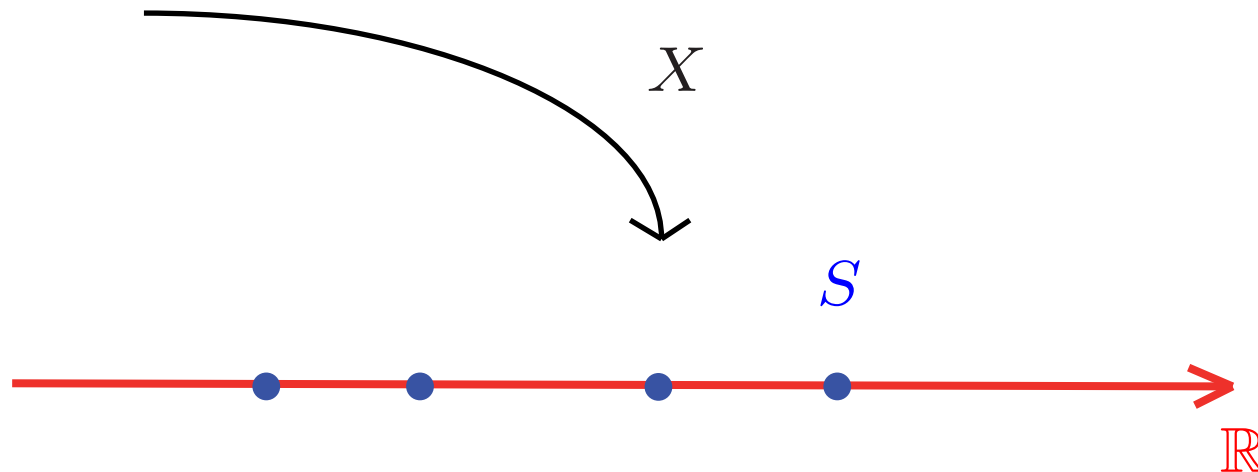
Beispiel: $S = \mathbb{N}$

$$\mathbf{P}(X = 1) = \frac{1}{2}, \quad \mathbf{P}(X = 2) = \frac{1}{4}, \quad \mathbf{P}(X = 3) = \frac{1}{8}, \dots$$

$$\mathbf{P}(X = a) = 1/2^a, \quad a \in \mathbb{N}.$$

Auch wenn der Wertebereich von X
eine überabzählbare Menge ist
(wie z.B. die Menge der reellen Zahlen $\mathbb{R}$
oder das “Einheitsintervall” $[0, 1]$
oder das “Einheitsquadrat” $[0, 1] \times [0, 1]$),
behalten beide Regeln ihren Sinn, wenn man fordert,
dass der Wertebereich
eine endliche oder abzählbar unendliche Menge S enthält
mit $P(X \in S) = 1$.
Eine Zufallsvariable X mit dieser Eigenschaft nennen wir
diskret.

Beispiel: Wertebereich $\mathbb{R}$



$S \subset \mathbb{R}$ endlich oder abzählbar unendlich mit $P(X \in S) = 1$

(z.B. $S = \{1, 2, \dots, g\}$, oder $S = \mathbb{N}$)

Warum ist das interessant?

Wie wir sehen werden (und wie jetzt schon intuitiv klar ist),
kann man mit reellwertigen Zufallsvariablen **rechnen**.

Man kann z.B. eine reellwertige Zufallsvariable X
durch 2 teilen,
und wenn die Zufallsvariable X diskret ist,
ist auch die Zufallsvariable $X/2$ diskret.

Nochmal die Definition:

Eine Zufallsvariable X heißt **diskret**,
falls ihr Wertebereich (nennen wir ihn W)
eine **diskrete** (d.h. endliche oder abzählbar unendliche)

Menge S enthält mit

$$P(X \in S) = 1.$$

Für diskretes X mit Wertebereich W
und diskretem S mit $\mathbf{P}(X \in S) = 1$ gilt:

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a), \quad A \subset S$$

(Wahrscheinlichkeiten sind abzählbar additiv)

$$\mathbf{P}(X \in W \setminus S) = \mathbf{P}(X \in W) - \mathbf{P}(X \in S) = 1 - 1 = 0$$

(X fällt mit Wahrscheinlichkeit 0 außerhalb von S .)

Die Abbildung $A \mapsto \rho(A) := \mathbf{P}(X \in A)$, $A \subset S$,

heißt die **Verteilung** von X .

Die Zahlen $\rho(a) := \mathbf{P}(X = a)$, $a \in S$,

sind die (atomaren) **Gewichte** der Verteilung ρ .

Wir nennen sie auch die **Verteilungsgewichte** von X .

Zufällige Paare und ihre Komponenten

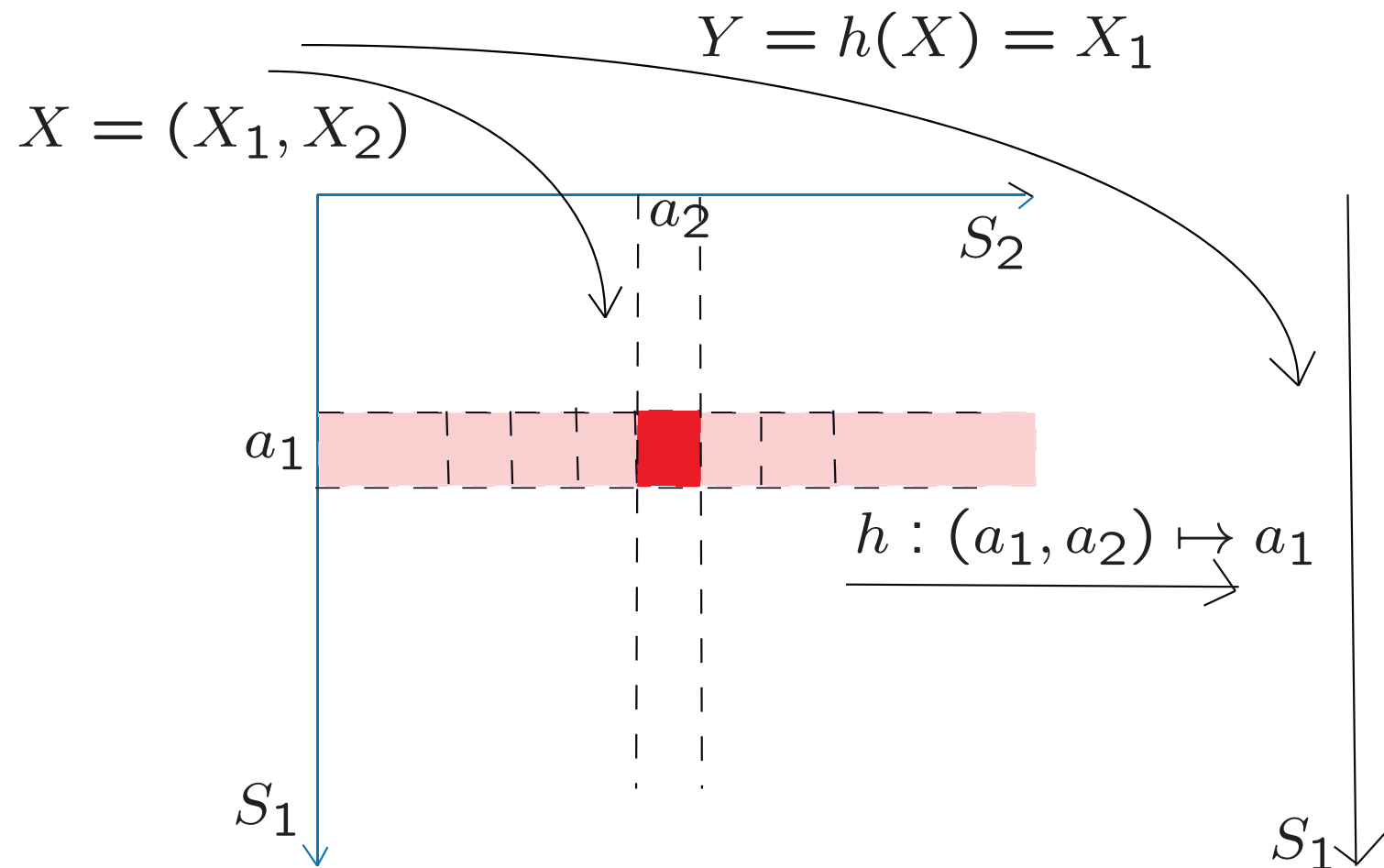
(Buch S. 21)

$X = (X_1, X_2)$ sei eine Zufallsvariable
(ein “zufälliges Paar”) mit diskreten Komponenten X_1, X_2 .

Dann ist auch X diskret, denn die Diskretheit von S_1 und S_2
vererbt sich auf $S_1 \times S_2$, und $\mathbf{P}(X_i \in S_i) = 1, i = 1, 2$,
impliziert $\mathbf{P}(X \in S_1 \times S_2) = 1$.

Die Ereignisse $\{(X_1, X_2) = (a_1, a_2)\}$ notieren wir auch als

$$\{X_1 = a_1, X_2 = a_2\}.$$



Die Verteilungsgewichte von X (man sagt auch:
die gemeinsamen Verteilungsgewichte von X_1 und X_2)
schreiben wir als

$$\begin{aligned}\rho(a_1, a_2) &= \mathbf{P}((X_1, X_2) = (a_1, a_2)) \\ &= \mathbf{P}(X_1 = a_1, X_2 = a_2) ,\end{aligned}$$

Die Verteilungsgewichte von X_1 erhält man
durch Summation über die a_2 :

$$\rho_1(a_1) = \sum_{a_2 \in S_2} \rho(a_1, a_2) .$$

3. Weiterverarbeitung von Zufallsvariablen und Transport von Verteilungen

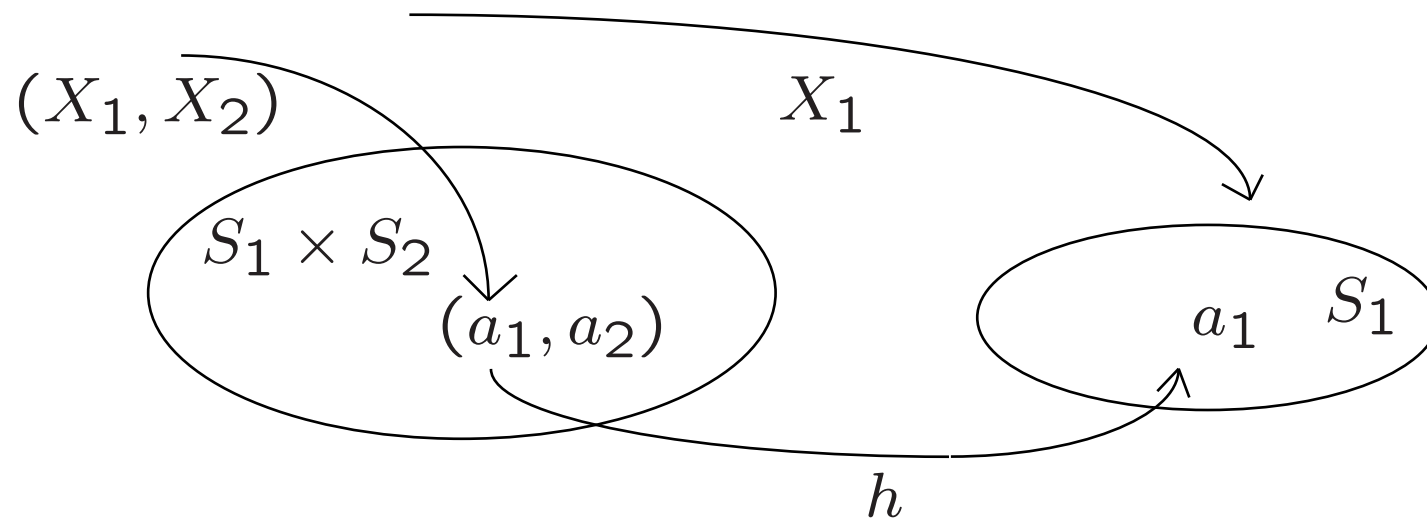
(Buch S. 21-22)

Der Übergang von $X = (X_1, X_2)$
zur Komponente X_1
ist ein Beispiel einer
Weiterverarbeitung einer Zufallsvariablen:

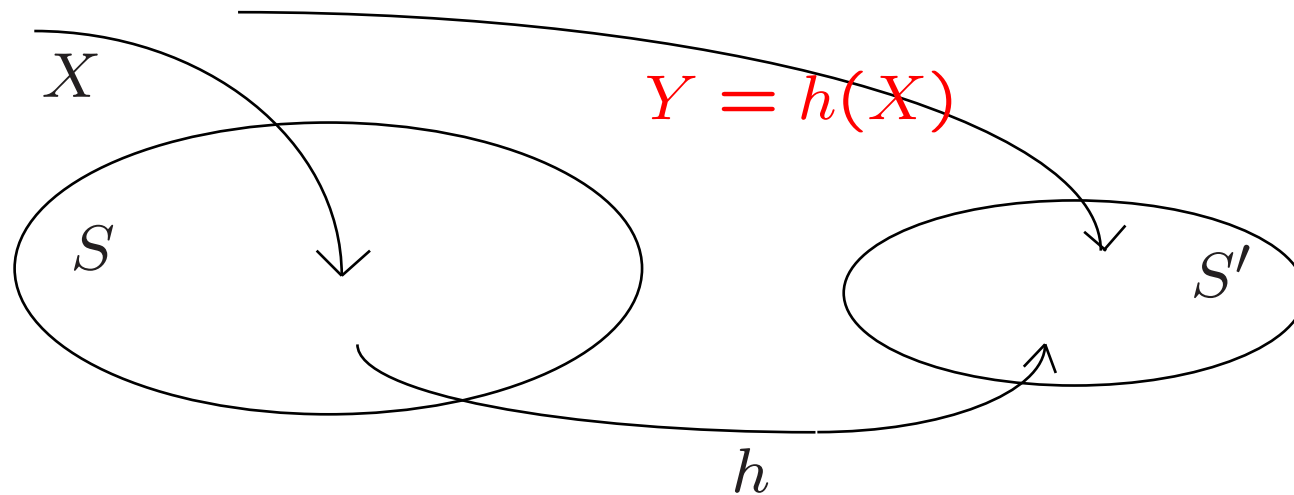
$$X_1 = h(X)$$

mit $h((a_1, a_2)) := a_1$,

also die Projektion des Paares (a_1, a_2)
auf seine erste Komponente.



Sind S und S' zwei Mengen,
 X eine Zufallsvariable mit Zielbereich S ,
 h eine Abbildung von S nach S' ,
und nimmt man X als zufällige Eingabe von h ,
dann bekommt man eine Zufallsvariable Y mit Zielbereich S' :

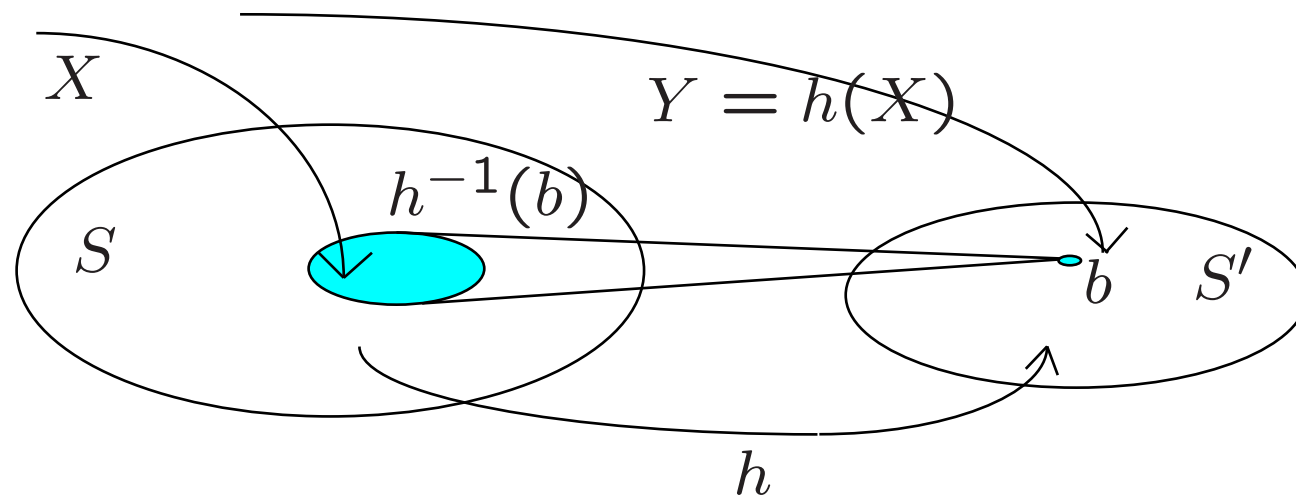


Für jedes $b \in S'$ gilt:

$$\{h(X) = b\} = \{X \in h^{-1}(b)\}.$$

Für die Verteilungsgewichte von $Y = h(X)$ ergibt sich:

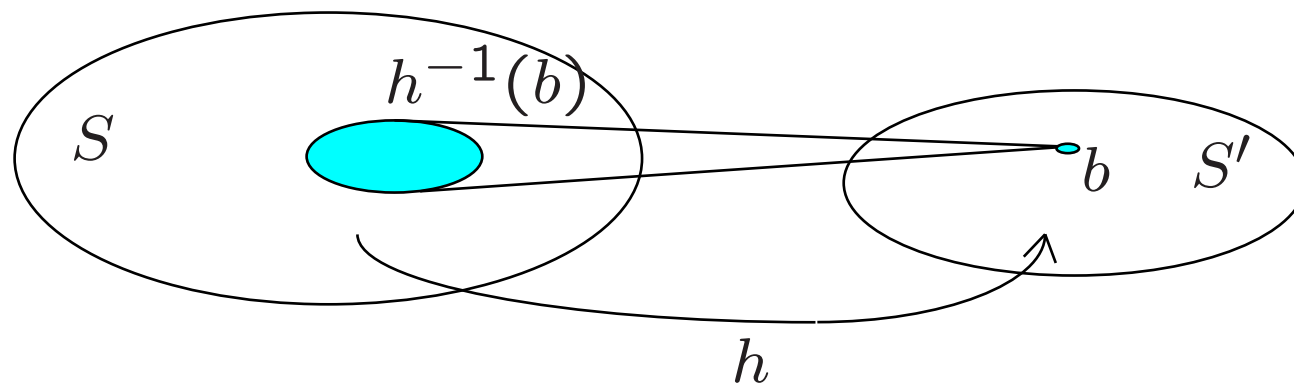
$$\mathbf{P}(Y = b) = \mathbf{P}(X \in h^{-1}(b)) = \sum_{a \in h^{-1}(b)} \mathbf{P}(X = a).$$



Bezeichnet ρ die Verteilung von X und ρ' die von Y ,
dann ist

$$\rho'(b) = \sum_{a \in h^{-1}(b)} \rho(a).$$

Man sagt: Die Verteilung ρ wird durch die Abbildung h
in die Verteilung ρ' transportiert.



Diese Situation haben wir schon mehrmals angetroffen:

in Vorlesung 1b:

$X :=$ rein zufällige $1, \dots, g$ -Folge der Länge $g + 1$

$T = h(X) :=$ Zeitpunkt der ersten Kollision.

in Vorlesung 2a:

$X :=$ rein zufällige Permutation von $1, \dots, n$

$h(X) :=$ Länge des Zyklus von X , der die Eins enthält.

Heutiges Programm:
Zusätzliche Beispiele für
“Weiterverarbeitungen von zufälligen Folgen”

Daraus bekommen wir wichtige Beispiele
diskreter Zufallsvariabler und diskreter Verteilungen.

Vorlesung 2b

Diskrete Zufallsvariable und ihre Verteilungen

Teil 2:

Die Anzahl der Erfolge:
Vom Münzwurf zur Binomialverteilung
(vgl. Buch S. 22)

1. Vom n -fachen fairen Münzwurf
zur Binomial($n, \frac{1}{2}$)-Verteilung:

$$S := \{0, 1\}^n$$

die Menge der 01-Folgen der Länge n

X sei uniform verteilt auf S ,

jeder Ausgang hat somit das Gewicht

$$\frac{1}{2^n} = \frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}.$$

Man sagt dann auch:

X ist ein n -facher “fairer Münzwurf”.

$Y :=$ die Anzahl der Einsen in X .

Was ergibt sich für die Verteilungsgewichte von Y ?

Jede einzelne 01-Folge a der Länge n mit genau k Einsen
hat Gewicht $\frac{1}{2^n}$

Es gibt $\binom{n}{k}$ derartiger Folgen a .

Also:

$$\mathbf{P}(Y = k) = \binom{n}{k} \frac{1}{2^n}, \quad k = 0, \dots, n.$$

Das sind die Gewichte der Binomial($n, \frac{1}{2}$)-Verteilung.

2. Die Anzahl der Sechsen beim fairen Würfeln:

Vom n -fachen Würfeln zur Binomial($n, \frac{1}{6}$)-Verteilung:

(vgl. Buch S. 28)

$X = (X_1, \dots, X_n)$ uniform verteilt auf
 $S := \{1, \dots, 6\}^n.$

$$Z := (Z_1, \dots, Z_n), \quad \text{mit}$$
$$Z_i := \mathbf{1}_{\{6\}}(X_i)$$

Z ist also eine zufällige 01-Folge, mit
 $Z_i = 1$ falls der i -te Wurf eine Sechs ergibt
und $Z_i = 0$ sonst.
Wie ist Z verteilt?

$$\begin{aligned}
& \mathbf{P}(Z_1 = 1, \dots, Z_k = 1, Z_{k+1} = 0, \dots, Z_n = 0) \\
&= \mathbf{P}(X_1 = 6, \dots, X_k = 6, X_{k+1} \neq 6, \dots, X_n \neq 6) \\
&= \frac{1^k \cdot 5^{n-k}}{6^n} \\
&= p^k q^{n-k}, \quad \text{mit } p := \frac{1}{6} \text{ und } q := \frac{5}{6}.
\end{aligned}$$

Auch für jede andere Platzierung von **genau k “Sechsen”**
in den n Würfeln ergibt sich diese W'keit.

Es gibt Es gibt $\binom{n}{k}$ derartige Platzierungen.

Also hat die Verteilung von

$$Y := Z_1 + \cdots + Z_n$$

die Gewichte

$$\mathbf{P}(Y = k) = \binom{n}{k} p^k q^{n-k}.$$

Das sind die Gewichte der Binomial($n, \frac{1}{6}$)-Verteilung.

3. Vom p -Münzwurf zur Binomialverteilung

(“Was $1/6$ recht ist, soll p billig sein...”)

(Buch S. 22)

Definition (p -Münzwurf):

Sei $p \in [0, 1]$, $q := 1 - p$.

Eine Zufallsvariable Z mit Zielbereich

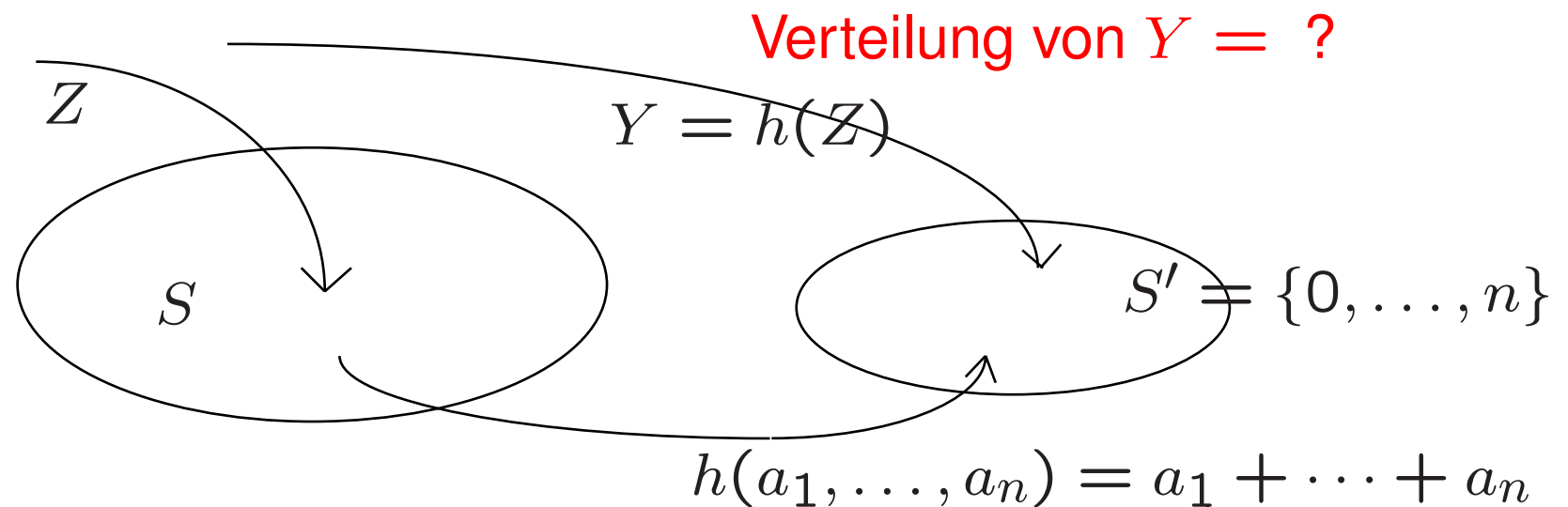
$$S = \{0, 1\}^n = \{a = (a_1, \dots, a_n) : a_i \in \{0, 1\}\}$$

heißt **n -facher p -Münzwurf**,

wenn für alle $a \in S$ mit k Einsen und $n - k$ Nullen gilt:

$$\mathbf{P}(Z = a) = p^k q^{n-k}.$$

Sei $Z = (Z_1, \dots, Z_n)$ ein n -facher p -Münzwurf
und $Y = Z_1 + \dots + Z_n =: h(Z)$ die *Anzahl der Erfolge*
(die Anzahl der Einsen in der zufälligen 0-1 Folge Z)

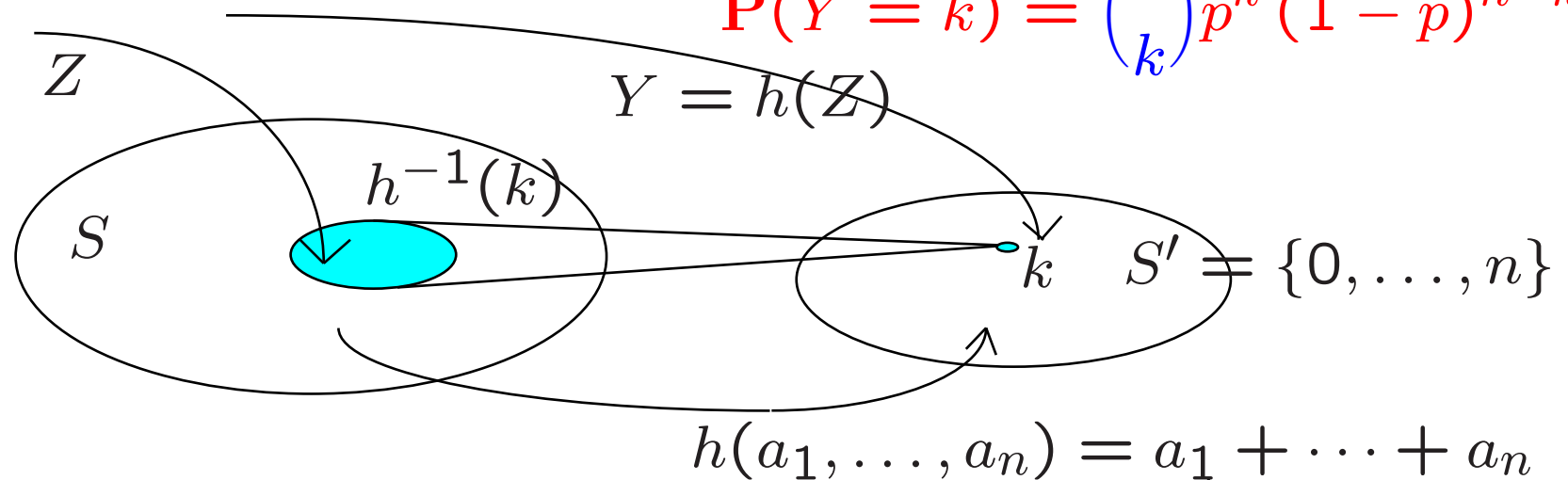


Jedes $a \in S$ mit $h(a) = k$
 (d.h. mit k Einsen und $n - k$ Nullen)

hat Gewicht $p^k (1 - p)^{n-k}$.

Es gibt $\binom{n}{k}$ solche a .

$$\mathbf{P}(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$



Definition:

Eine Zufallsvariable Y mit Zielbereich $\{0, 1, \dots, n\}$

heißt *binomialverteilt* mit Parametern n und p ,

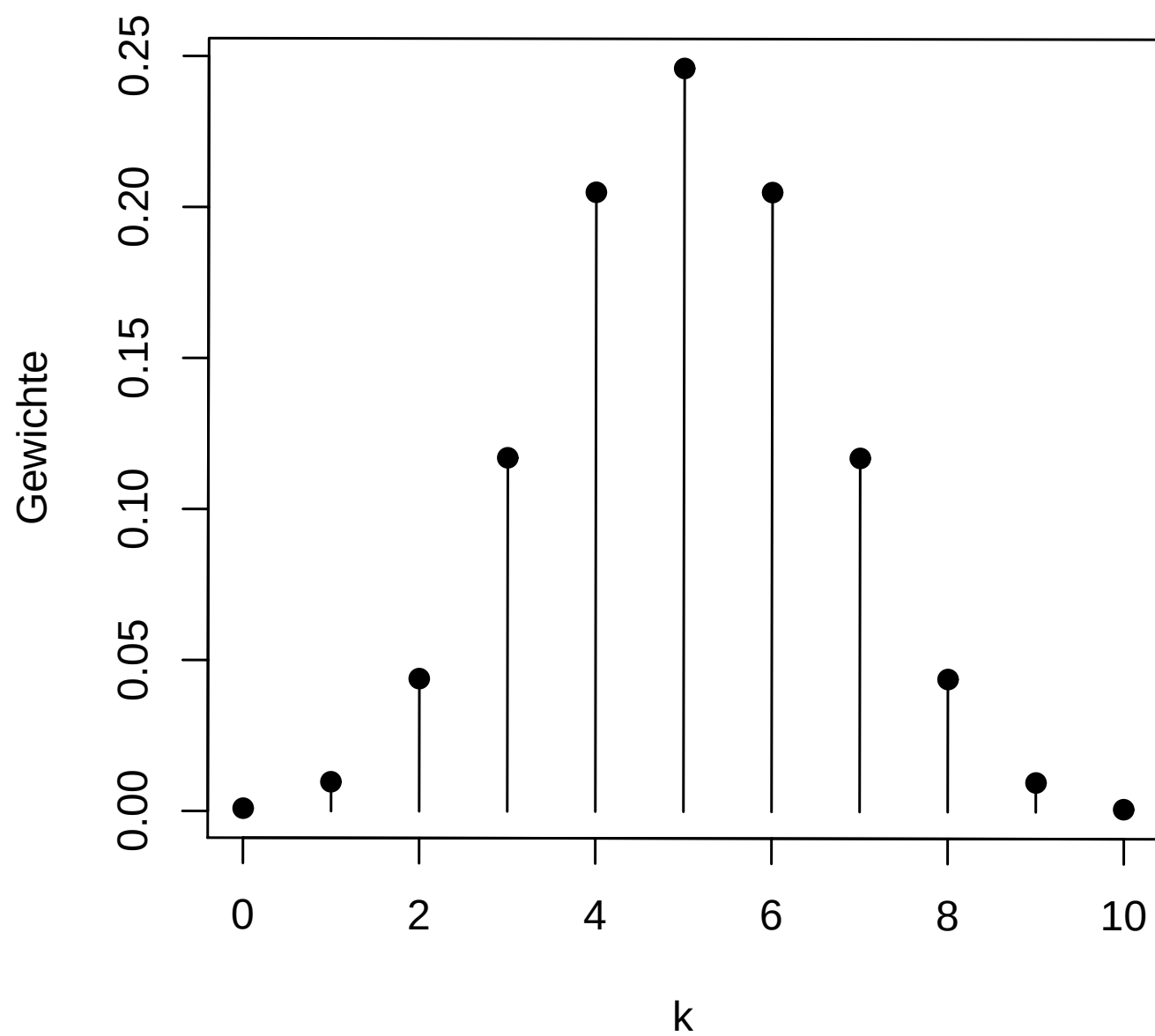
kurz

$\text{Bin}(n, p)$ -verteilt,

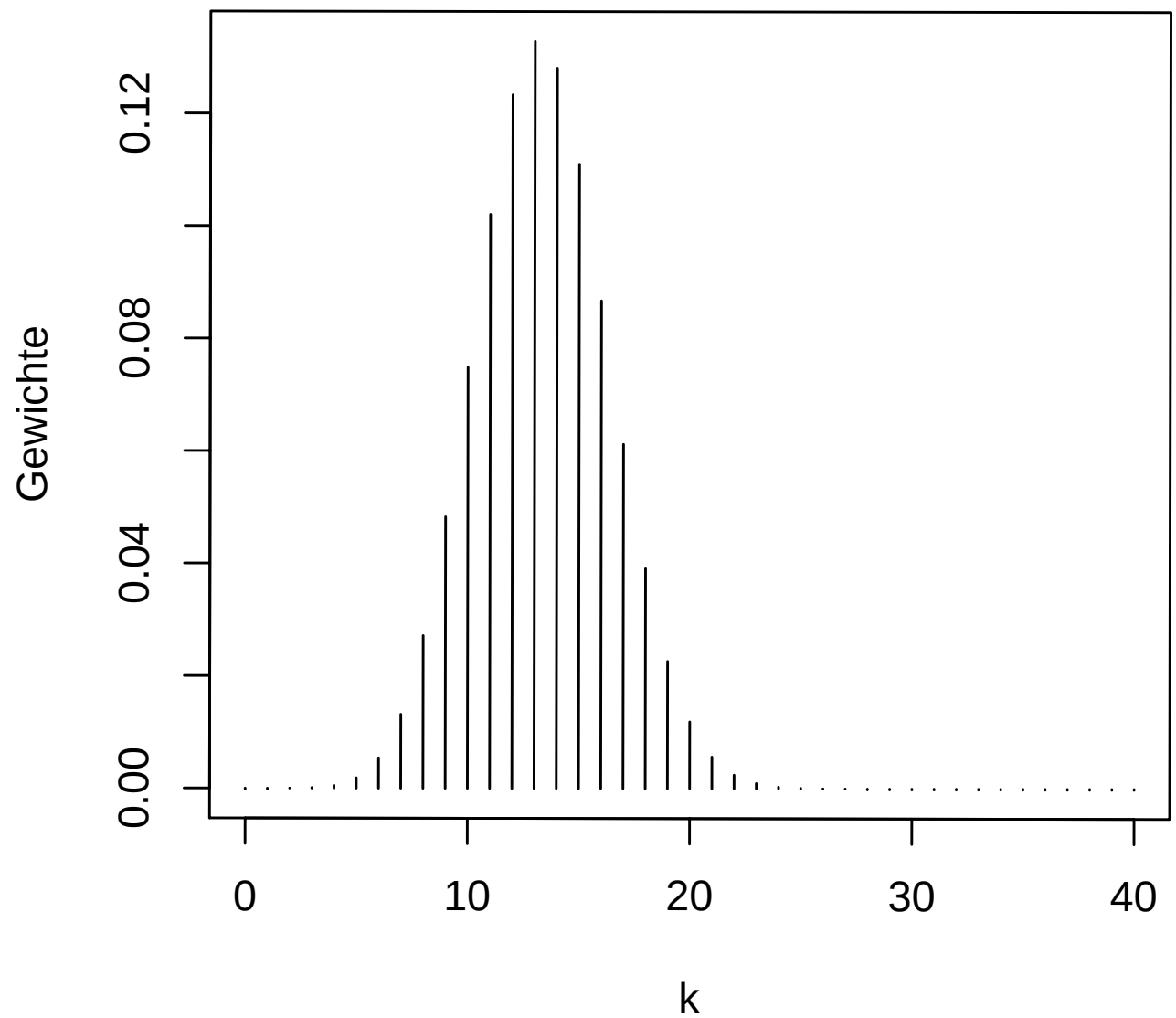
wenn

$$\mathbf{P}(Y = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n,$$

mit $q = 1 - p$.



Gewichte der Bin(10, 1/2) Verteilung



Gewichte der $\text{Bin}(40, 1/3)$ Verteilung

4. Vom Ziehen mit Zurücklegen zum p -Münzwurf

(vgl. Buch S. 33)

n -maliges Ziehen *mit Zurücklegen*
aus einer ideal durchmischten Urne.

Ein Anteil p der Kugeln ist **rot**,
der restliche Anteil $q = 1 - p$ ist **blau**.

Zufällige 0-1 Folge $Z = (Z_1, \dots, Z_n)$:

$Z_i = 1$ wenn beim i -ten Zug eine rote Kugel kommt,
und $Z_i = 0$ wenn beim i -ten Zug eine blaue Kugel kommt.

Sei a eine vorgegebene 0-1 Folge der Länge n mit k Einsen,

$$\text{z. B.: } a := (\underbrace{1, \dots, 1}_{k\text{-mal}}, \underbrace{0, \dots, 0}_{(n-k)\text{-mal}})$$

$$P(Z = a) = ?$$

Sei g die Gesamtanzahl der Kugeln in der Urne.

Die Anzahl der roten Kugeln ist pg ,

die der blauen Kugeln ist qg . Für obiges a gilt:

$$P(Z = a) = \frac{(pg)^k (qg)^{n-k}}{g^n} = p^k q^{n-k}$$

Das ist so für jede 0-1 Folge a mit k Einsen und $n - k$ Nullen.

Zur Erinnerung:

Definition (p -Münzwurf):

Sei $p \in [0, 1]$, $q := 1 - p$.

Eine Zufallsvariable Z mit Zielbereich

$$S = \{0, 1\}^n = \{a = (a_1, \dots, a_n) : a_i \in \{0, 1\}\}$$

heißt **n -facher p -Münzwurf**,

wenn für alle $a \in S$ mit k Einsen und $n - k$ Nullen gilt:

$$\mathbf{P}(Z = a) = p^k q^{n-k}.$$

Vorlesung 2b

Diskrete Zufallsvariable und ihre Verteilungen

Teil 3:

Zufällige Besetzungen:

Vom Würfeln zur Multinomialverteilung

1. Das $(p_1, \dots, p_g)$ -Würfeln als Verallgemeinerung des p -Münzwurfs

Oder: Was 2 recht ist, soll g billig sein!

(vgl. Buch S. 28)

Definition (“ n -faches $(p_1, \dots, p_g)$ -Würfeln”):

Seien $g \in \mathbb{N}$ und $p_1, \dots, p_g \geq 0$ mit $p_1 + \dots + p_g = 1$.

Wir definieren **Verteilungsgewichte** auf

$$S := \{a = (a_1, \dots, a_n) : a_i \in \{1, \dots, g\}\}$$

durch

$$\rho(a_1, \dots, a_n) := p_{a_1} \cdot p_{a_2} \cdots p_{a_n}.$$

Eine Zufallsvariable X mit diesem Zielbereich S
und diesen Verteilungsgewichten ρ nennen wir

n -faches $(p_1, \dots, p_g)$ -Würfeln.

Für *jedes* $a \in S$ mit
 b_1 Komponenten gleich 1,
 b_2 Komponenten gleich 2,
...
 b_g Komponenten gleich g

ist dann

$$\mathbf{P}(X = a) = p_1^{b_1} p_2^{b_2} \cdots p_g^{b_g}$$

2. Vom $(p_1, \dots, p_g)$ -Würfeln zur Multinomialverteilung

$X = (X_1, \dots, X_n)$ sei ein n -faches $(p_1, \dots, p_g)$ -Würfeln

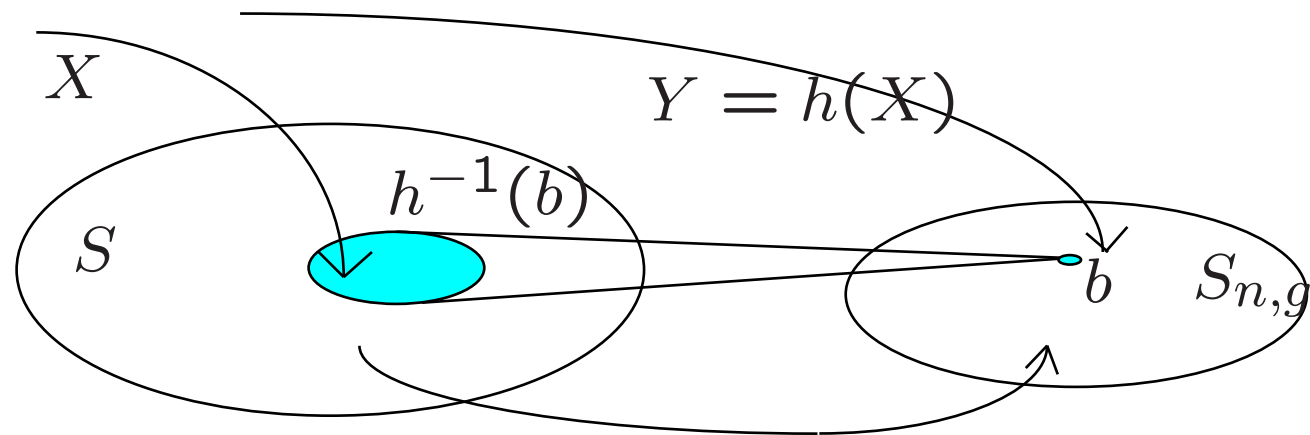
$$Y_j := \#\{i : X_i = j\}$$

(die Anzahl der Würfe mit Ergebnis j).

$Y := (Y_1, \dots, Y_g)$ hat dann den Zielbereich

$$S_{n,g} = \{(b_1, \dots, b_g) : b_j \in \mathbb{N}_0, b_1 + \dots + b_g = n\}.$$

Verteilung von Y = ?



$$h(a_1, \dots, a_n) = (b_1, \dots, b_g) =: b$$

mit $b_j := \#\{i : a_i = j\}$, $j = 1, \dots, g$

Jedes $a \in S$ mit $h(a) = (b_1, \dots, b_g)$

hat Gewicht $p_1^{b_1} \dots p_g^{b_g}$

Wieviele solche a gibt es?

Dazu überlegen wir:

Auf wieviele Arten kann man
 n Objekte so auf g Fächer verteilen,
dass das j -te Fach jeweils genau b_j Objekte enthält?

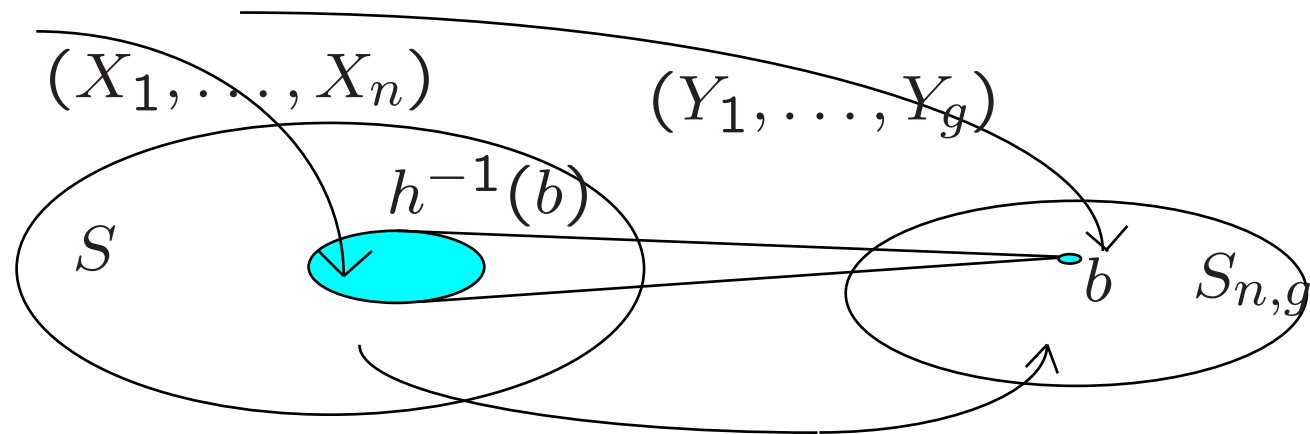
Dabei ist $b_1 + \dots + b_g = n$.

Die Antwort ist:

$$\binom{n}{b_1} \cdot \binom{n - b_1}{b_2} \dots \binom{n - b_1 - \dots - b_{g-1}}{b_g}$$

$$= \frac{n!}{b_1! b_2! \dots b_g!} =: \binom{n}{b_1, \dots, b_g}$$

Multinomialkoeffizient, lies: n über $b_1, \dots, b_g$



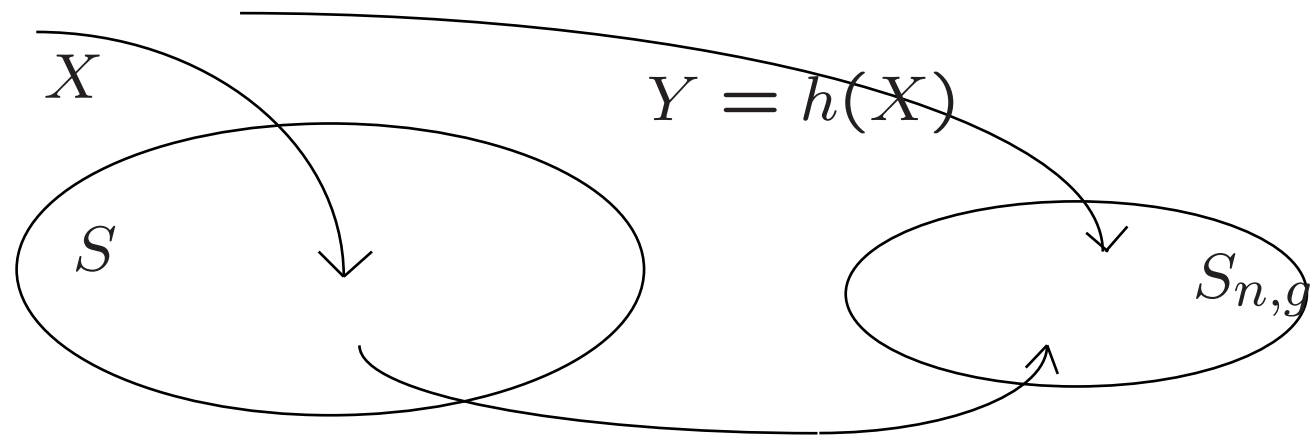
$$h(a_1, \dots, a_n) = (b_1, \dots, b_g)$$

Jedes $a \in S$ mit $h(a) = (b_1, \dots, b_g)$

hat Gewicht $p_1^{b_1} \dots p_g^{b_g}$

Es gibt $\binom{n}{b_1, \dots, b_g}$ solche a .

$$\mathbf{P}(Y_1 = b_1, \dots, Y_g = b_g) = \binom{n}{b_1, \dots, b_g} p_1^{b_1} \dots p_g^{b_g}$$



$$h(a_1, \dots, a_n) = (b_1, \dots, b_g) =: k$$

mit $b_j := \#\{i : a_i = j\}$, $j = 1, \dots, g$

$$\mathbf{P}(Y_1 = b_1, \dots, Y_g = b_g) = \binom{n}{b_1, \dots, b_g} p_1^{b_1} \dots p_g^{b_g}$$

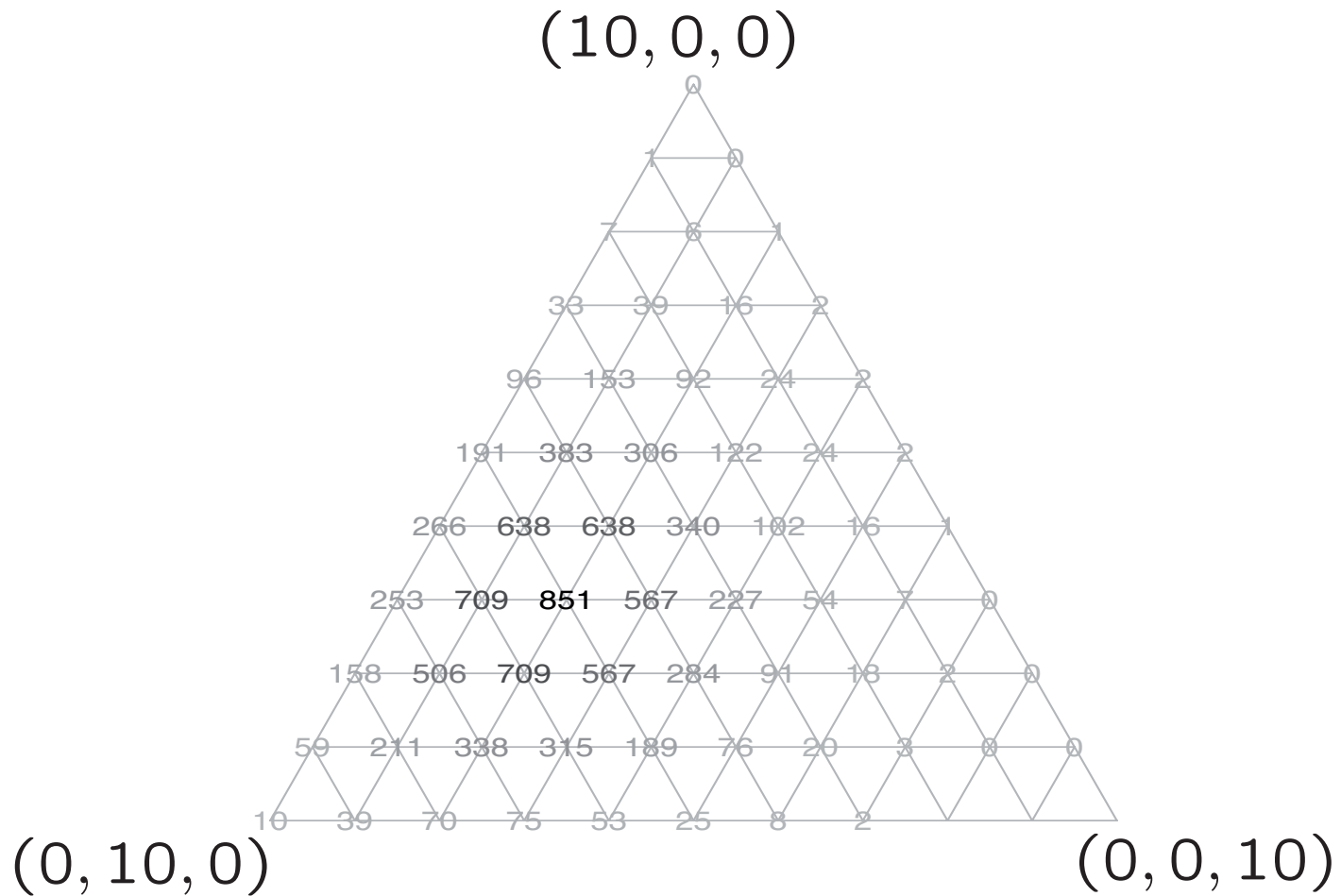
Definition:

Eine Zufallsvariable X mit Zielbereich $S_{n,g}$
heißt *multinomialverteilt* mit Parametern $n; p_1, \dots, p_g$,

wenn

$$\mathbf{P}(X = (b_1, \dots, b_g)) = \binom{n}{b_1, \dots, b_g} p_1^{b_1} \dots p_g^{b_g},$$

$$(b_1, \dots, b_g) \in S_{n,g} .$$



Gewichte der Multinomialverteilung, notiert in Vielfachen von $\frac{1}{10000}$,
für $n = 10$, $g = 3$, $p_1 = 0.3$, $p_2 = 0.5$, $p_3 = 0.2$

Resumé

$$g \in \mathbb{N}$$

Würfeln

Besetzung der Ausgänge

Multinomialverteilung

$$g = 2$$

Münzwurf

Anzahl der Erfolge

Binomialverteilung

Vorlesung 3a

Der Erwartungswert

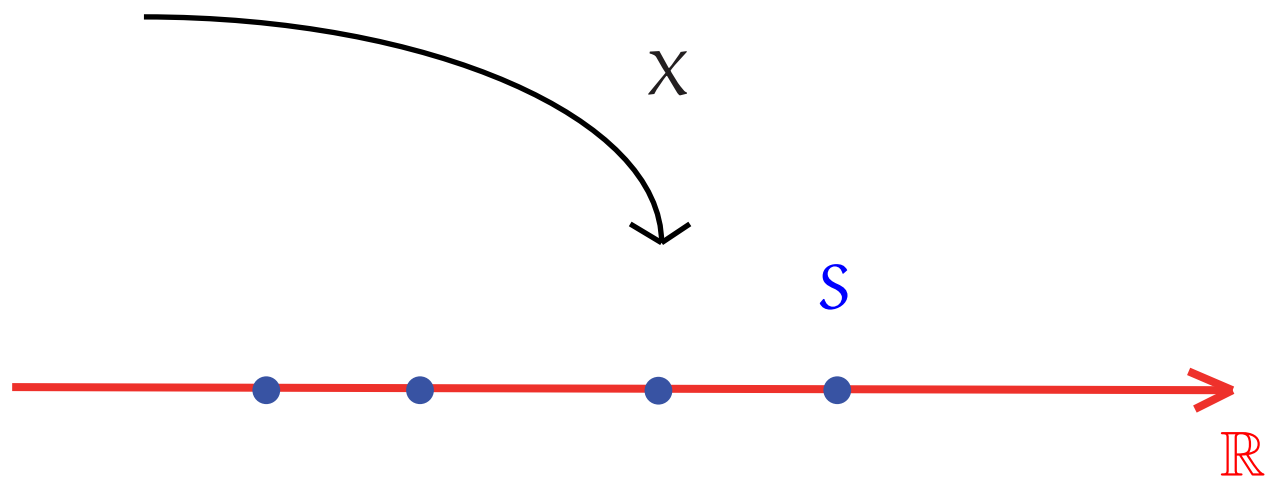
von diskreten reellwertigen Zufallsvariablen

Teil 1

Der Erwartungswert als gewichtetes Mittel

(Buch S. 23)

Unter einer **diskreten reellwertigen** Zufallsvariablen
verstehen wir eine ZV'e, deren Wertebereich
die Menge $\mathbb{R}$ der reellen Zahlen
(oder einer Teilmenge davon) ist,
und für die eine
endliche oder abzählbar unendliche Menge S ($\subset \mathbb{R}$) existiert
mit **$\mathbf{P}(X \in S) = 1$** .

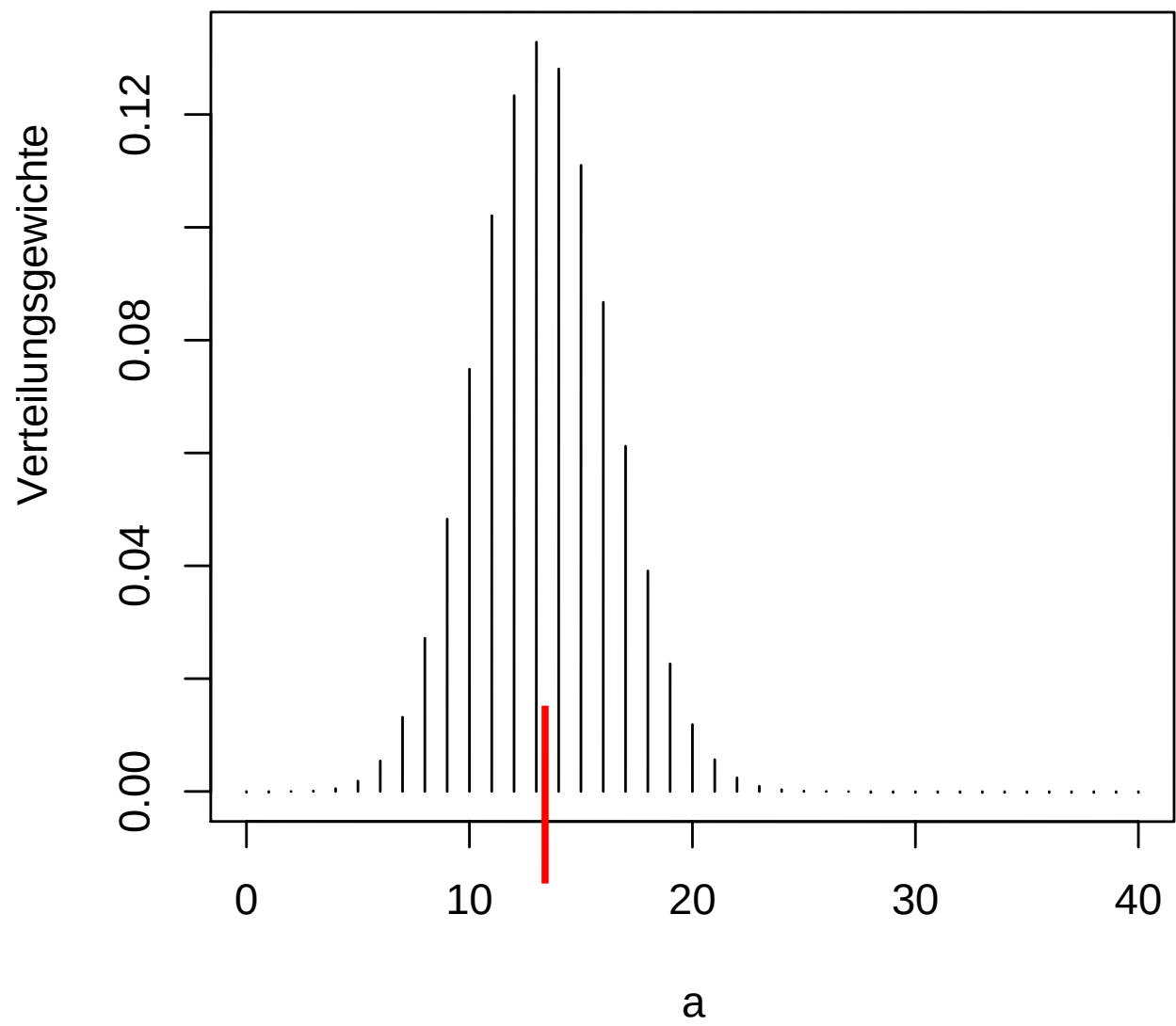


Eine einprägsame Kenngröße
für die *Lage* der Verteilung von X

ist das mit den Wahrscheinlichkeiten gewichtete Mittel
der möglichen Werte von X :

$$\mathbf{E}[X] := \sum_{a \in S} a \mathbf{P}(X = a) .$$

Man spricht vom *Erwartungswert von X* .
(Wir bezeichnen ihn auch mit μ oder μ_X .)



Das elementarste Beispiel:

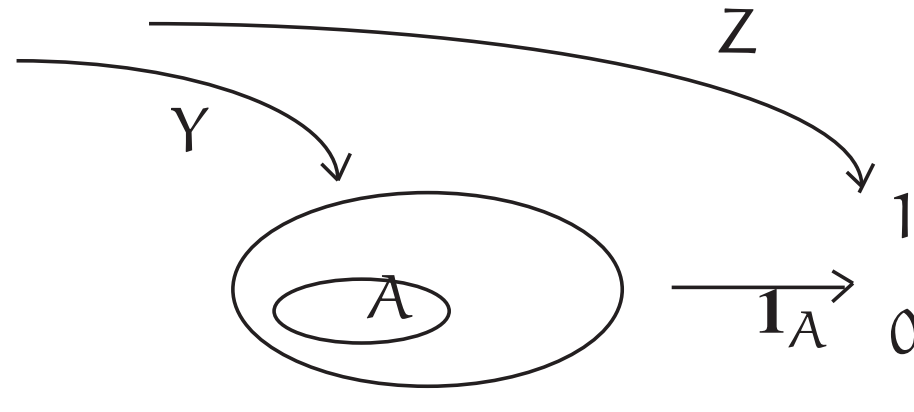
$$Z = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p \\ 0 & \text{mit Wahrscheinlichkeit } q = 1 - p \end{cases}$$

Z beschreibt also einen einfachen p -Münzwurf.

$$\mathbf{E}[Z] = p \cdot 1 + q \cdot 0 = p$$

$$\mathbf{E}[Z] = \mathbf{P}(Z = 1).$$

Indikatorvariable von Ereignissen



$$Z = \mathbf{1}_A(Y) =: I_{\{Y \in A\}}$$

.... die *Indikatorvariable* des Ereignisses $\{Y \in A\}$

$$\mathbf{E}[Z] = \mathbf{P}(Z = 1)$$

$$\{Z = 1\} = \{Y \in A\}$$

$$\mathbf{E}[I_{\{Y \in A\}}] = \mathbf{P}(Y \in A).$$

$$\mathbf{E}[I_{\{Y \in A\}}] = \mathbf{P}(Y \in A).$$

Die Wahrscheinlichkeit eines Ereignisses ist gleich dem Erwartungswert der Indikatorvariable des Ereignisses.

Für eine diskrete reellwertige ZV'e X hatten wir

$$\begin{aligned}\mathbf{E}[X] &= \sum_{a \in S} a \mathbf{P}\{X = a\} \\ &= \sum_{a \in S} a \rho(a).\end{aligned}$$

Dabei sind die Zahlen $\rho(a)$ die Verteilungsgewichte von X .

Merke:

Der Erwartungswert der Zufallsvariablen X
hängt nur von ihrer Verteilung ρ ab.

Synonym sprechen wir daher auch manchmal vom
Erwartungswert der Verteilung ρ .

Gilt $S \subset \mathbb{R}_+$ oder $\sum_{a \in S} |a| \rho(a) < \infty$,

dann hängt der Summenwert $\sum_{a \in S} a \rho(a)$
nicht von der Reihenfolge der Summanden ab.

Man sagt dann:

Der Erwartungswert der Zufallsvariablen X mit Verteilung ρ

ist wohldefiniert

oder kurz: Der Erwartungswert von X *existiert*.

Merke:

X

ist eine Zufallsgröße;

$E[X]$

ist eine Zahl.

Vorlesung 3a

Der Erwartungswert

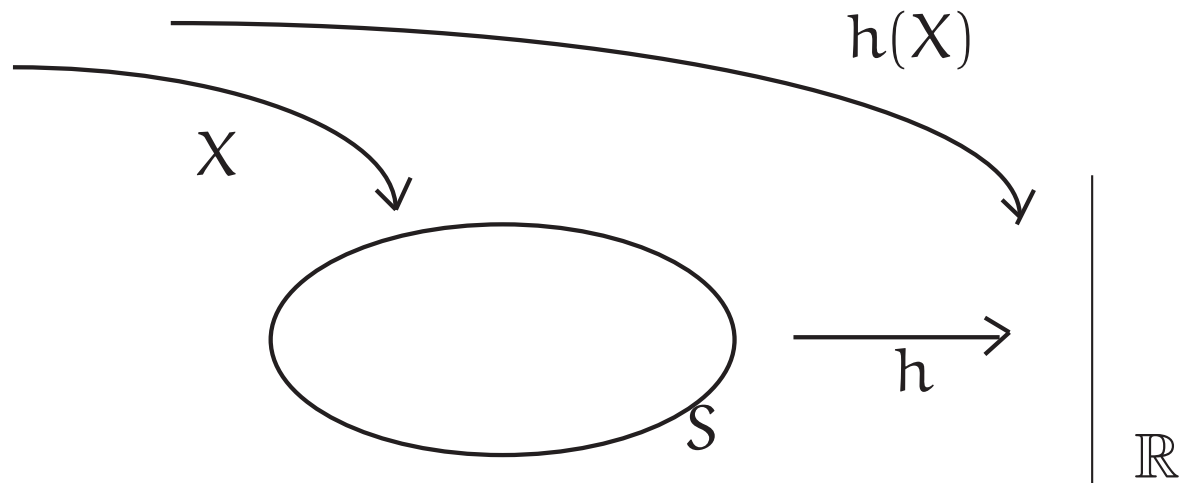
von diskreten reellwertigen Zufallsvariablen

Teil 2

Die Transformationsformel für Erwartungswerte

(vgl. Buch S. 23)

Sei X diskrete Zufallsvariable mit $\mathbf{P}(X \in S) = 1$
und h eine Abbildung von S nach $\mathbb{R}$



Sei X diskrete Zufallsvariable mit $\mathbf{P}(X \in S) = 1$
und h eine Abbildung von S nach $\mathbb{R}$

$$\text{mit } \sum_{a \in S} |h(a)| \mathbf{P}(X = a) < \infty.$$

Dann ist

$$\mathbf{E}[h(X)] = \sum_{a \in S} h(a) \mathbf{P}(X = a) .$$

Diese Formel ist oft hilfreich
bei der Berechnung von Erwartungswerten.

Sie erinnert an die Einsetzungsregel(Substitutionsregel)
zum Berechnen von Summen und Integralen,

und wird uns im nächsten Teil helfen,
die Linearität des Erwartungswertes herzuleiten.

$$\mathbf{E}[h(X)] = \sum_{a \in S} h(a) \mathbf{P}(X = a)$$

Die Idee ist einfach: anstatt die Werte $b = h(a)$, $a \in S$,
mit deren Gewichten zu mitteln,
“zerlegt man nach dem Urbild”
und mittelt dann mit den Gewichten der Werte a .

$$\mathbf{E}[h(X)] = \sum_{a \in S} h(a) \mathbf{P}(X = a)$$

Denn:

$$\begin{aligned} & \sum_{b \in h(S)} b \mathbf{P}(h(X) = b) \\ &= \sum_{b \in h(S)} b \sum_{a \in h^{-1}(b)} \mathbf{P}(X = a) \\ &= \sum_{b \in h(S)} \sum_{a \in h^{-1}(b)} h(a) \mathbf{P}(X = a) \\ &= \sum_{a \in S} h(a) \mathbf{P}(X = a) . \quad \square \end{aligned}$$

Vorlesung 3a

Der Erwartungswert

von diskreten reellwertigen Zufallsvariablen

Teil 3

Linearität des Erwartungswertes

(vgl. (Buch S. 52))

Satz 1. (Homogenität des EW)

Der Erwartungswert des c -fachen
einer diskreten reellwertigen Zufallsvariablen X
ist das c -fache des Erwartungswertes:

$$\mathbf{E}[cX] = c\mathbf{E}[X]$$

Satz 1. (Homogenität des EW)

Der Erwartungswert des c -fachen
einer diskreten reellwertigen Zufallsvariablen X
ist das c -fache des Erwartungswertes:

$$h(a) := ca, \quad Y := h(X) = cX.$$

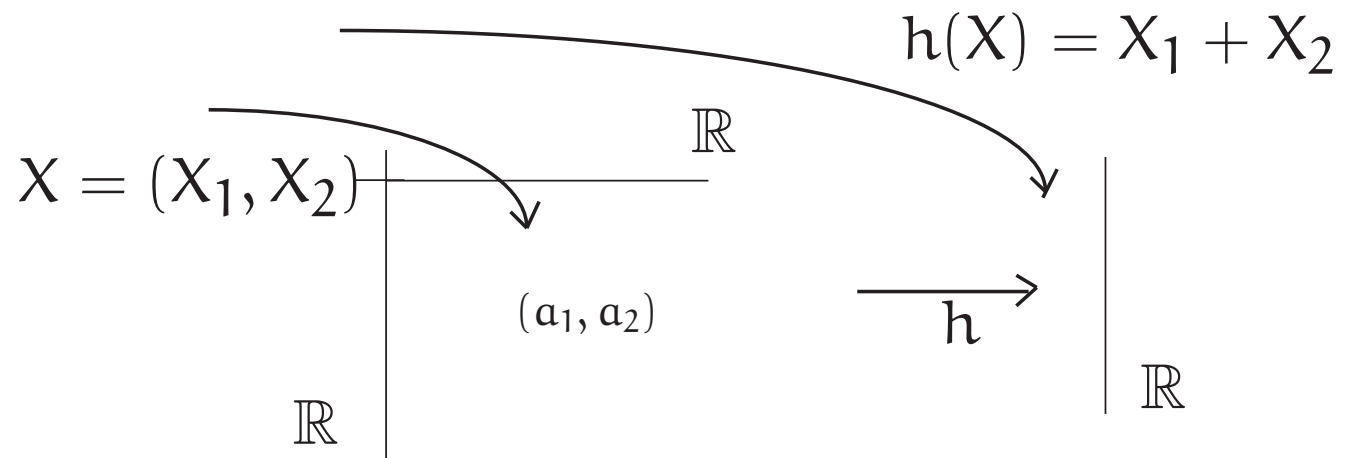
$$\begin{aligned} \mathbf{E}[cX] &= \mathbf{E}[h(X)] = \sum_{a \in S} h(a) \mathbf{P}(X = a) \\ &= \sum_{a \in S} ca \mathbf{P}(X = a) = c \sum_{a \in S} a \mathbf{P}(X = a) = c\mathbf{E}[X]. \end{aligned}$$

Satz 2. (Additivität des Erwartungswertes)

Sei (X_1, X_2) ein zufälliges Paar reellwertiger ZV'er,
für die $\mathbf{E}[X_1]$ und $\mathbf{E}[X_2]$ existieren und endlich sind.

Dann gilt

$$\mathbf{E}[X_1 + X_2] = \mathbf{E}[X_1] + \mathbf{E}[X_2]$$



$$h(a_1, a_2) = a_1 + a_2$$

Beweis von Satz 2

(hier nur für **diskrete** Zufallsvariable):

Seien $S_1, S_2 \subset \mathbb{R}$ abzählbar mit

$$\mathbf{P}(X_1 \in S_1) = \mathbf{P}(X_2 \in S_2) = 1.$$

Aus der Transformationsformel folgt mit

$$h(a_1, a_2) := a_1 + a_2:$$

$$\mathbf{E}[X_1 + X_2] = \sum_{(a_1, a_2) \in S_1 \times S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

In der nächsten Umformung werden wir
die in V2b1 besprochene Beziehung

$$\sum_{a_2 \in S_2} \mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 \in a_1)$$

verwenden.

$$\mathbf{E}[X_1 + X_2]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1 \in S_1} a_1 \sum_{a_2 \in S_2} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$+ \sum_{a_2 \in S_2} a_2 \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$\mathbf{E}[X_1 + X_2]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1 \in S_1} a_1 \mathbf{P}(X_1 = a_1)$$

$$+ \sum_{a_2 \in S_2} a_2 \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$\mathbf{E}[X_1 + X_2]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1 \in S_1} a_1 \mathbf{P}(X_1 = a_1)$$

$$+ \sum_{a_2 \in S_2} a_2 \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$\mathbf{E}[X_1 + X_2]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \mathbf{E}[X_1]$$

$$+ \sum_{a_2 \in S_2} a_2 \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$\mathbf{E}[X_1 + X_2]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} (a_1 + a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \mathbf{E}[X_1]$$

$$+ \mathbf{E}[X_2]$$



Korollar aus Satz 1 und Satz 2
(Linearität des Erwartungswertes)
(Buch S. 52)

Für reellwertige Zufallsvariable X_1, X_2
für die $\mathbf{E}[X_1]$ und $\mathbf{E}[X_2]$ existieren und endlich sind,
gilt

$$\mathbf{E}[c_1 X_1 + c_2 X_2] = c_1 \mathbf{E}[X_1] + c_2 \mathbf{E}[X_2] , \quad c_1, c_2 \in \mathbb{R} .$$

Vorlesung 3a

Der Erwartungswert

von diskreten reellwertigen Zufallsvariablen

Teil 4

Beispiele

1. Der Erwartungswert der Binomialverteilung

(als Erwartungswert der Anzahl der Erfolge
beim n -fachen p -Münzwurf)

(Buch S. 49)

X sei $\text{Bin}(n, p)$ -verteilt.

$$\mathbf{E}[X] = ?$$

$$\sum_{k=0}^n k \mathbf{P}(X = k) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \dots$$

Es GEHT so (vgl Buch Seite 23-24)

Aber es geht auch einfacher (vgl. Buch S. 49):

Sei $Z = (Z_1, \dots, Z_n)$ ein n -facher p -Münzwurf.

Dann ist $(Z_1 + \dots + Z_n)$ $\text{Bin}(n, p)$ -verteilt.

$$\mathbf{E}[Z_1 + \dots + Z_n] = \mathbf{E}[Z_1] + \dots + \mathbf{E}[Z_n]$$

$$\mathbf{E}[Z_i] = 1 \cdot p + 0 \cdot q = p$$

Fazit:

Der Erwartungswert einer $\text{Bin}(n, p)$ verteilten ZV ist

np .

2. Der Erwartungswert der hypergeometrischen Verteilung

(als Erwartungswert der Anzahl der “Erfolge”
beim n -fachen Ziehen ohne Zurücklegen)

(Buch S. 50 und S. 28)

BEISPIEL

Ziehen ohne Zurücklegen

Eine Urne enthält r rote und b blaue Kugeln.

ooooooooooooo $r = 8$ $b = 5$

Aus der Urne werden ohne Zurücklegen n Kugeln gezogen.

oooooooooooooo $n = 9$

$R :=$ Anzahl der gezogenen roten Kugeln

$$\mathbf{E}[R] = ?$$

$$R = Z_1 + Z_2 + \dots + Z_n$$

$Z_i = 1$ falls i-te gezogene Kugel rot

$Z_i = 0$ falls i-te gezogene Kugel blau

ooooooooooooo $r = 8$ $b = 5$

$$\mathbf{P}(Z_i = 1) = \frac{r}{r + b}$$

Man stelle sich vor, die Nummern der Züge werden
als rein zufällige Permutation an die $r + b$ Kugeln vergeben.

$\{Z_i = 1\}$ ist das Ereignis “die i-te gezogene Kugel ist rot”.

$$R = Z_1 + Z_2 + \dots + Z_n$$

$Z_i = 1$ falls i-te gezogene Kugel rot

$Z_i = 0$ falls i-te gezogene Kugel blau

ooooooooooooo $r = 8$ $b = 5$

$$\mathbf{P}(Z_i = 1) = \frac{r}{r + b}$$

$$\mathbf{E}[Z_i] = \frac{r}{r + b}$$

$$\mathbf{E}[R] = \mathbf{E}[Z_1] + \mathbf{E}[Z_2] + \dots + \mathbf{E}[Z_n]$$

$$\mathbf{E}[R] = n \frac{r}{r + b}$$

BEISPIEL

Ziehen ohne Zurücklegen

Eine Urne enthält r rote und b blaue Kugeln.

ooooooooooooo $r = 8$ $b = 5$

Aus der Urne werden ohne Zurücklegen n Kugeln gezogen.

oooooooooooooo $n = 9$

$R :=$ Anzahl der gezogenen roten Kugeln

Verteilung von R ?

Verteilungsgewichte von R ?

$$\mathbf{P}(R = k) = ?$$

$$\mathbf{P}(R = k) = \binom{r}{k} \binom{b}{n-k} / \binom{r+b}{n}$$
$$(k = 0, \dots, n)$$

Eine ZV mit diesen Verteilungsgewichten
heißt

hypergeometrisch verteilt zu den Parametern $(n, r+b, r)$.

(vgl. Buch Seite 28)

$$\mathbf{P}(R = k) = \binom{r}{k} \binom{b}{n-k} / \binom{r+b}{n}$$

$$\mathbf{E}(R) = ?$$

$$\mathbf{E}[R] = \sum_{k=0}^n k \binom{r}{k} \binom{b}{n-k} / \binom{r+b}{n} = \dots$$

Es GEHT so (vgl. Buch Seite 32)

Aber wie wir eben gesehen haben,

(über die Darstellung von R als Summe von Zählern)

geht's auch einfacher (vgl. Buch S. 50/51).

3. Der Erwartungswert einer Anzahl von Runs

Runs beim fairen Münzwurf

$Z := (Z_1, Z_2, \dots, Z_n)$ n-facher fairer Münzwurf

$$P\{Z_i = 1\} = \frac{1}{2} \quad P\{Z_i = 0\} = \frac{1}{2}$$

Run: ein Block von Nullen (Einsen),
der nicht echt in einem größeren Block enthalten ist

$R :=$ Anzahl Runs in Z

00000000 $R = 1$

11100011 $R = 3$

10101010 $R = 8$

$$\mathbf{E}[R] = ?$$

Dazu schreiben wir R als Summe von Zählern.

Bei jedem Wurf zählen wir eins dazu,
wenn bei diesem Wurf ein Run beginnt:

$Y_i := 1$ falls bei i ein Run beginnt, $Y_i := 0$ sonst

$$R = Y_1 + Y_2 + \dots + Y_n$$

$$Y_1 \equiv 1$$

$$\{Y_i = 1\} = \{(Z_{i-1}, Z_i) = (0, 1) \text{ oder } (1, 0)\} \quad (i > 1)$$

$$P(Y_i = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \quad (i > 1)$$

$$E[Y_i] = \frac{1}{2} \quad (i > 1)$$

$$E[R] = E[Y_1] + E[Y_2] + E[Y_3] + \dots + E[Y_n]$$

$$E[R] = 1 + \frac{1}{2}(n - 1)$$

Vorlesung 3a

Der Erwartungswert

von diskreten reellwertigen Zufallsvariablen

Teil 5

Wie erlebt man den Erwartungswert?

X

eine Zufallsgröße;

$E[X]$

eine Zahl.

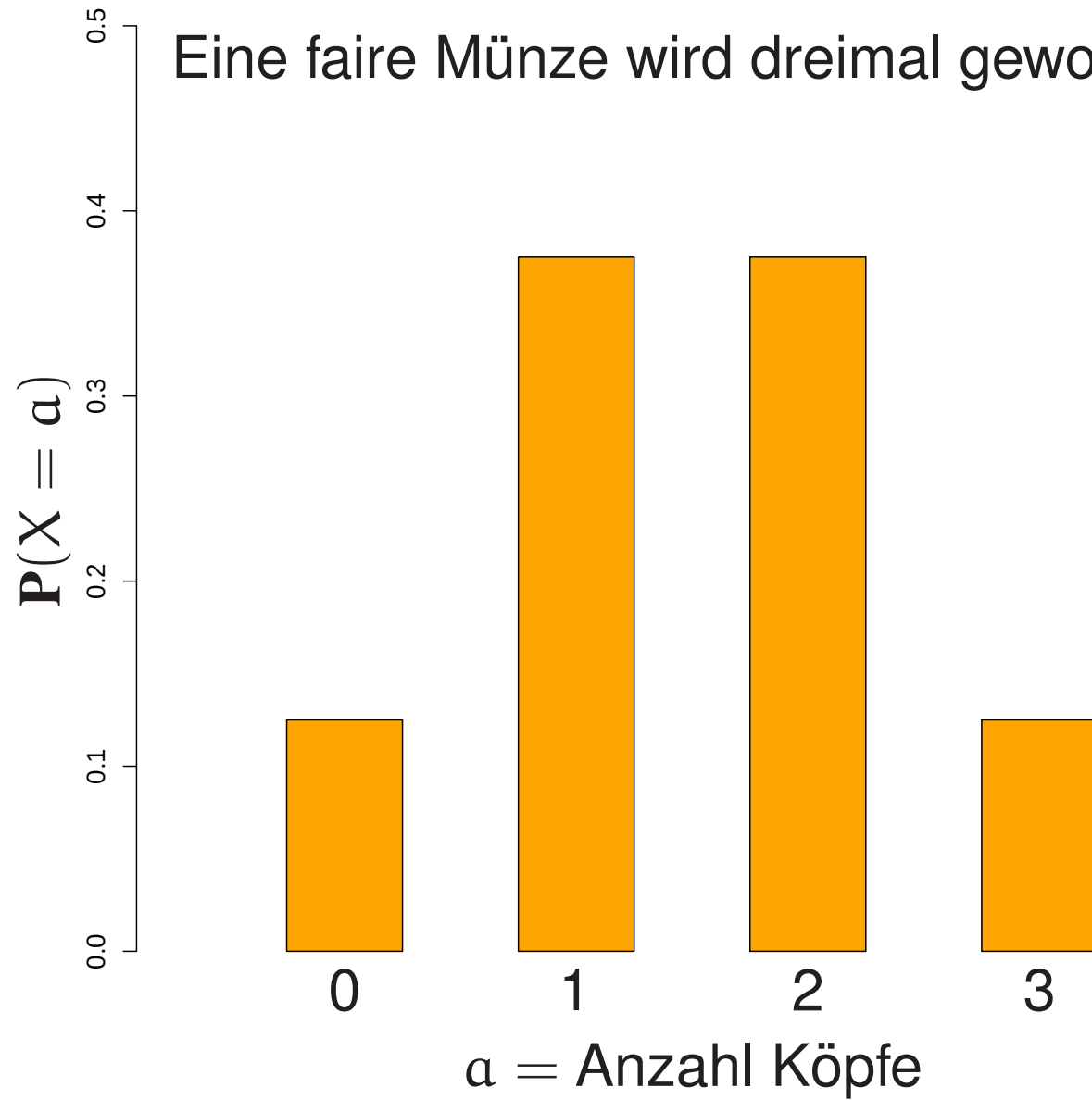
Ein Beispiel: Der Erwartungswert der Anzahl der Erfolge beim dreifachen Münzwurf

Eine faire Münze wird dreimal geworfen.

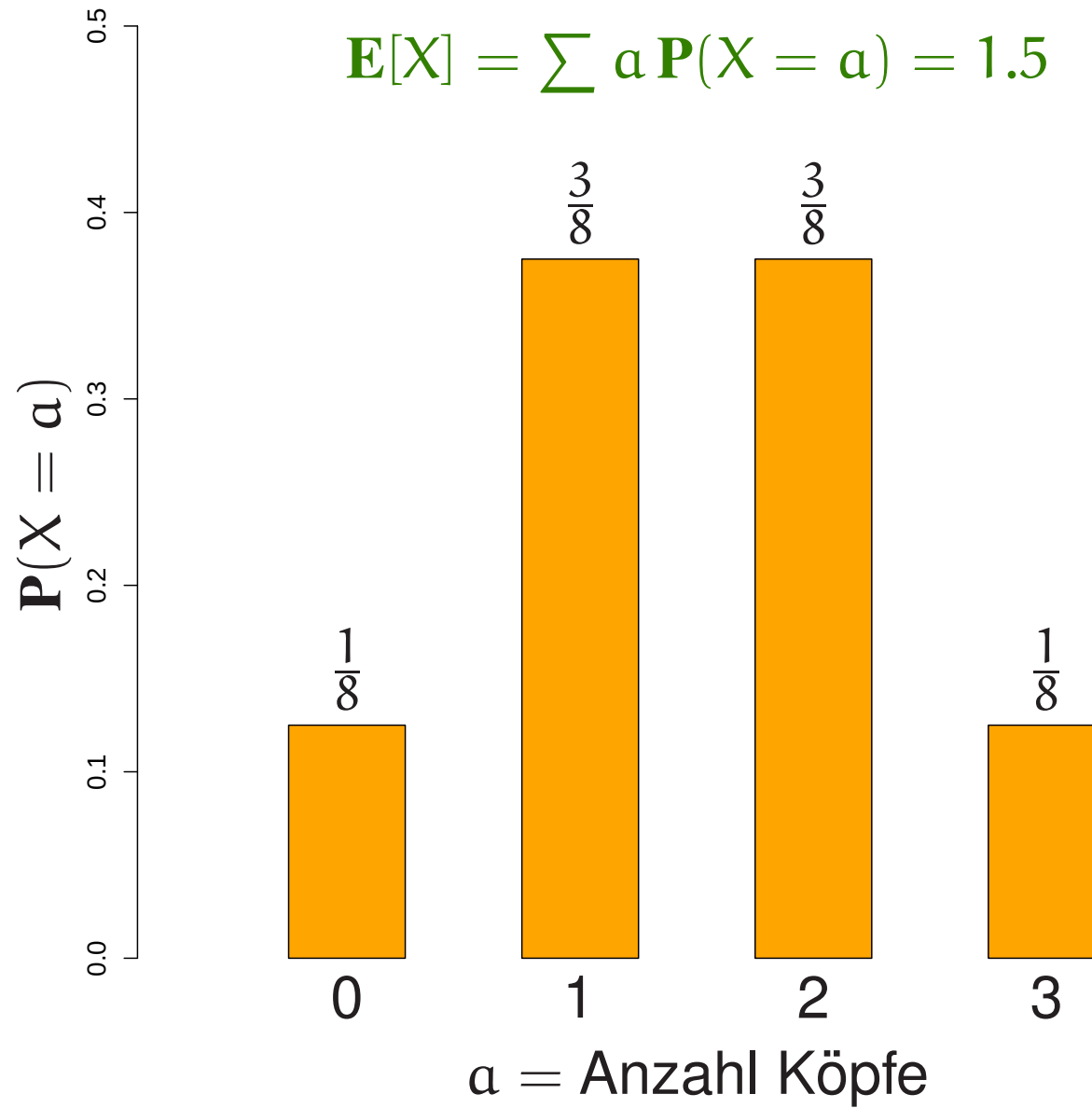
$X :=$ Anzahl der geworfenen Köpfe.

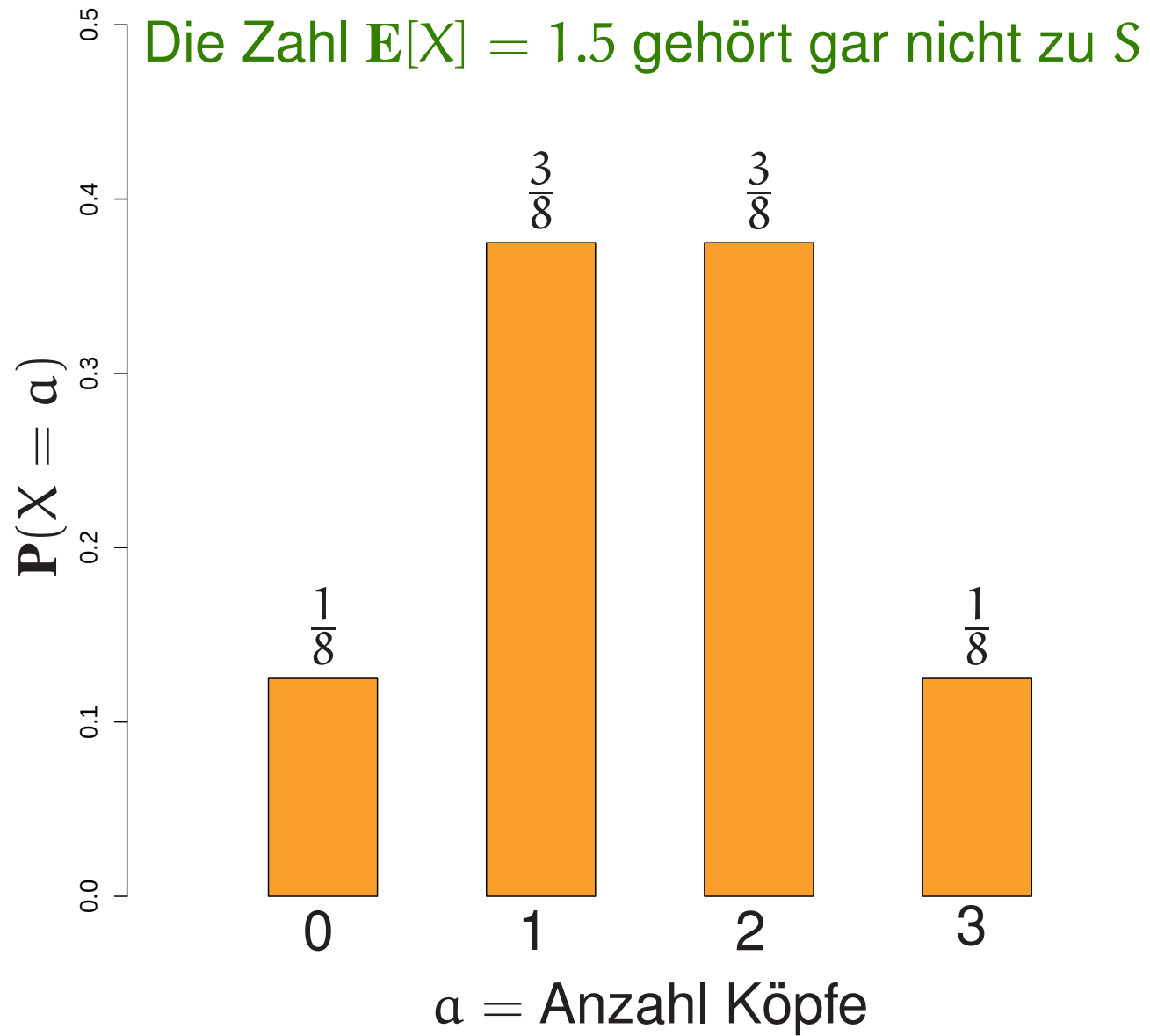
$S := \{0, 1, 2, 3\}.$

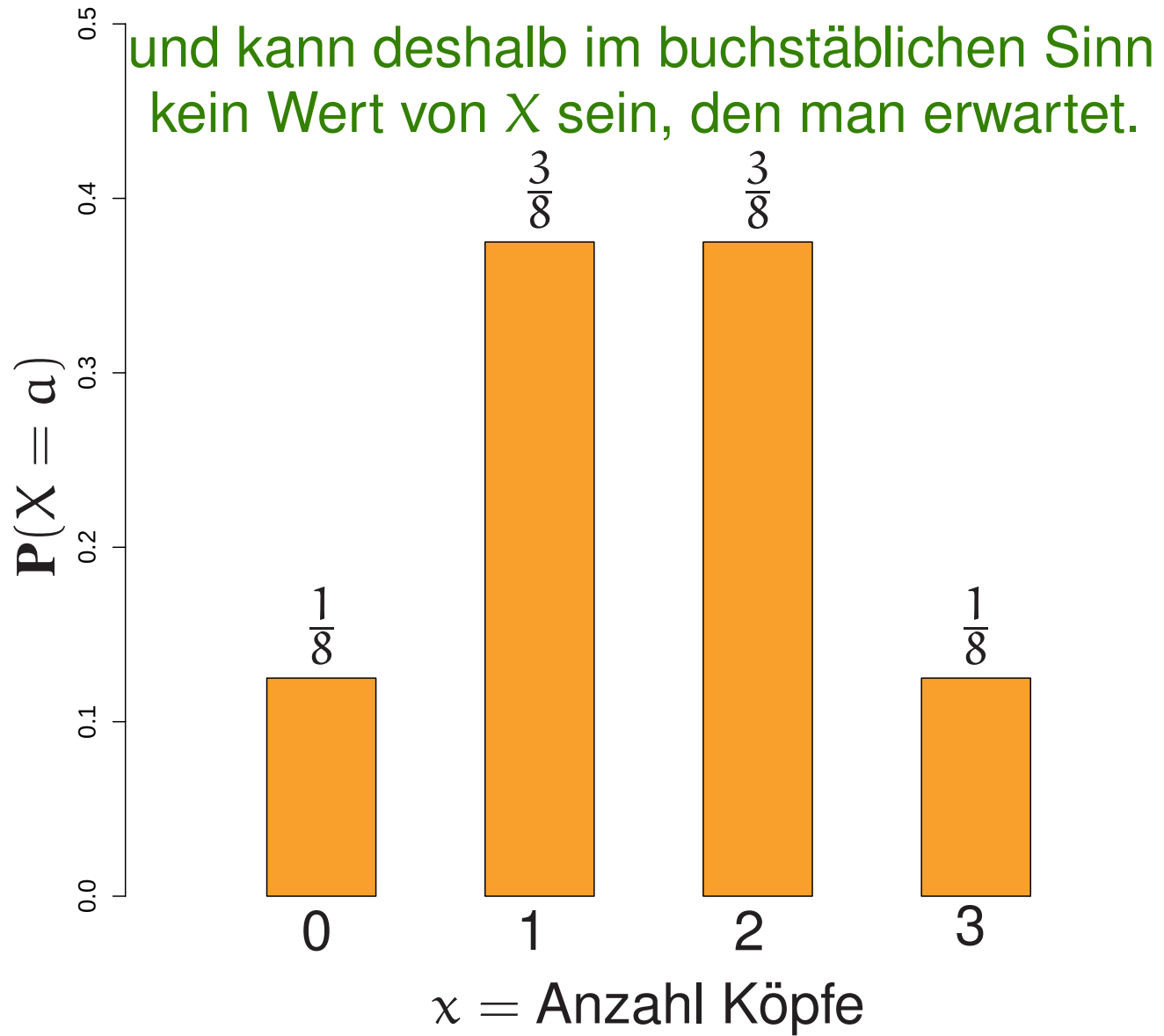
Eine faire Münze wird dreimal geworfen.



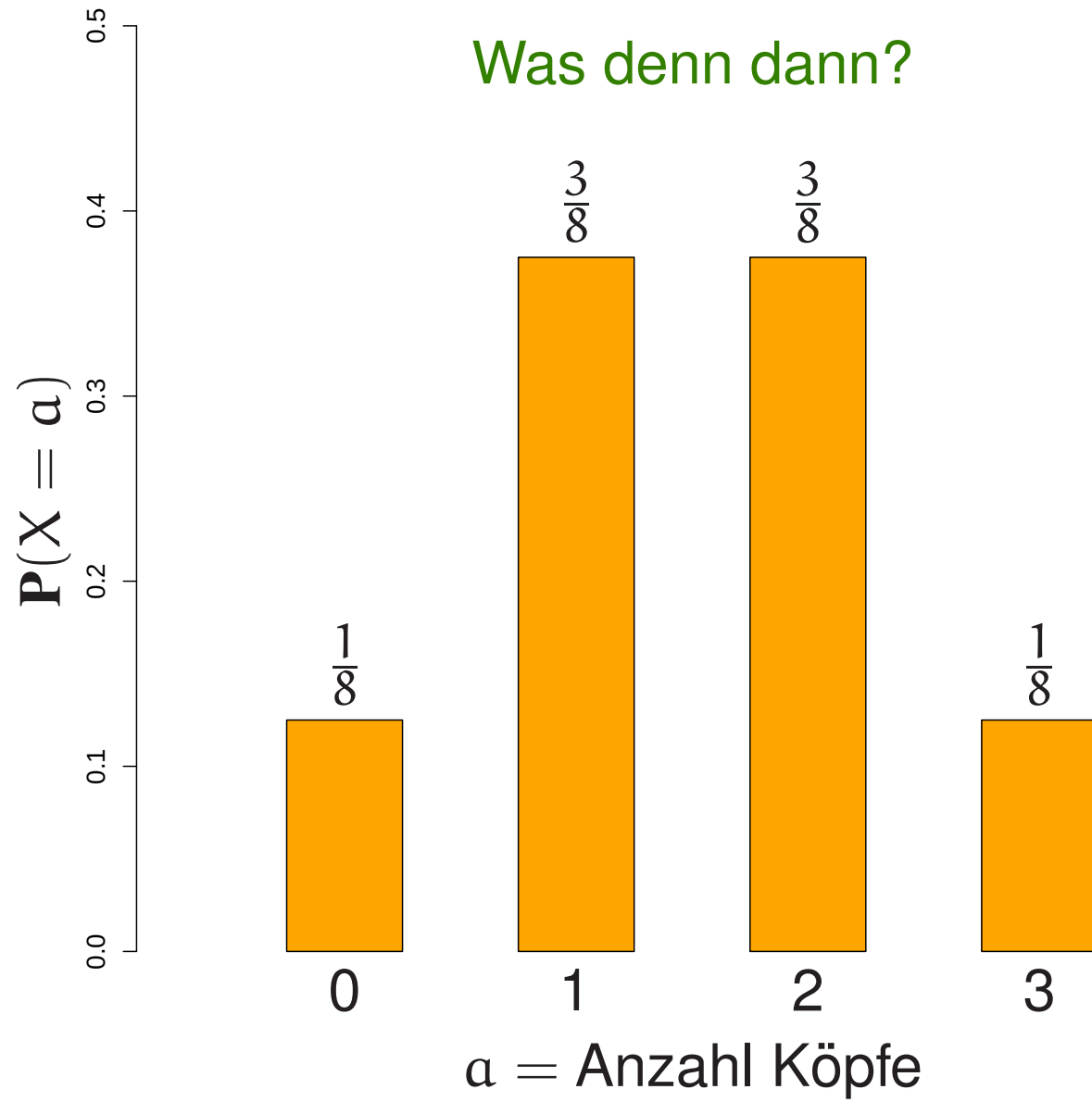
$$\mathbf{E}[X] = \sum a \mathbf{P}(X = a) = 1.5$$



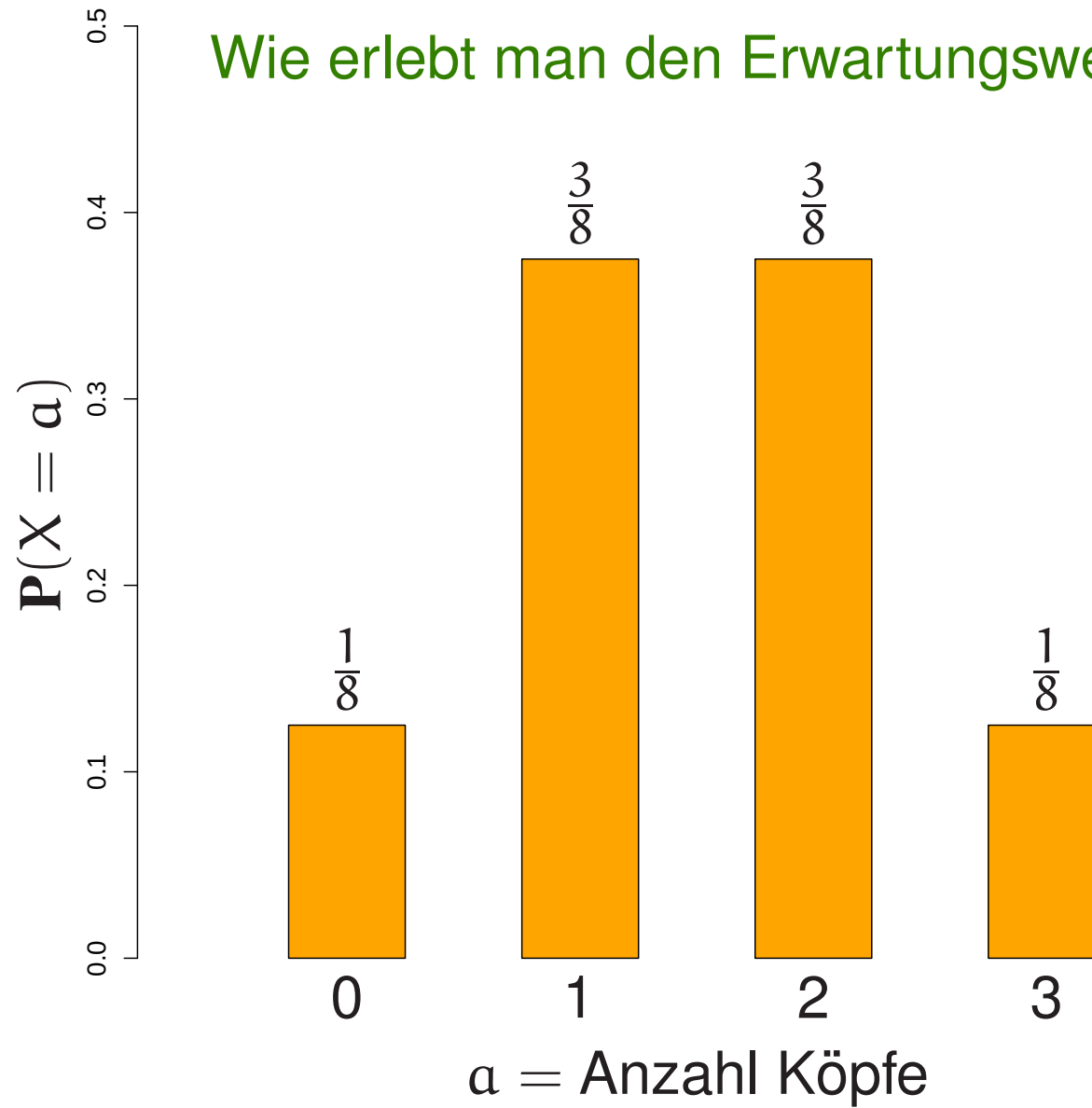




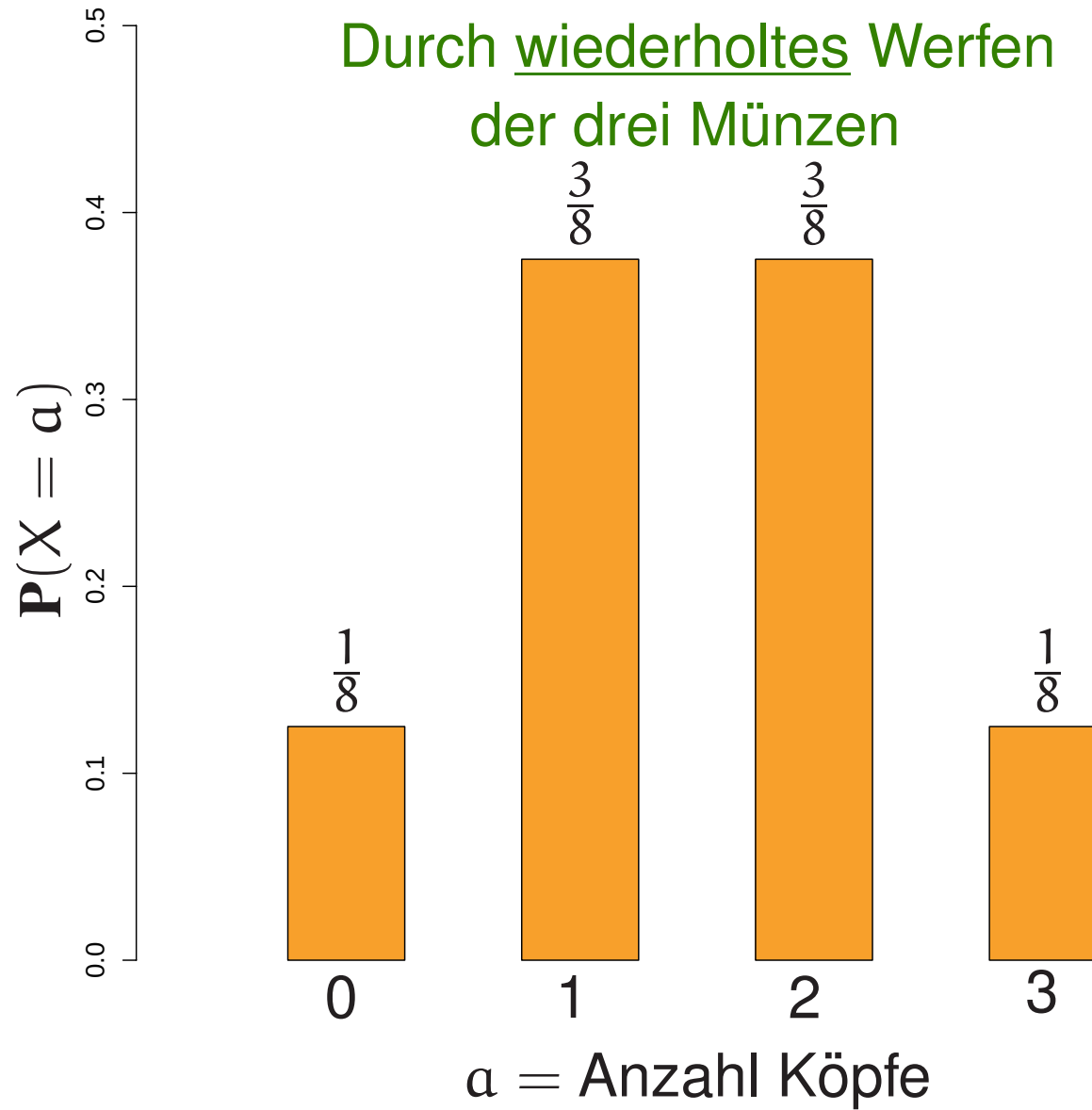
Was denn dann?



Wie erlebt man den Erwartungswert?



Durch wiederholtes Werfen
der drei Münzen

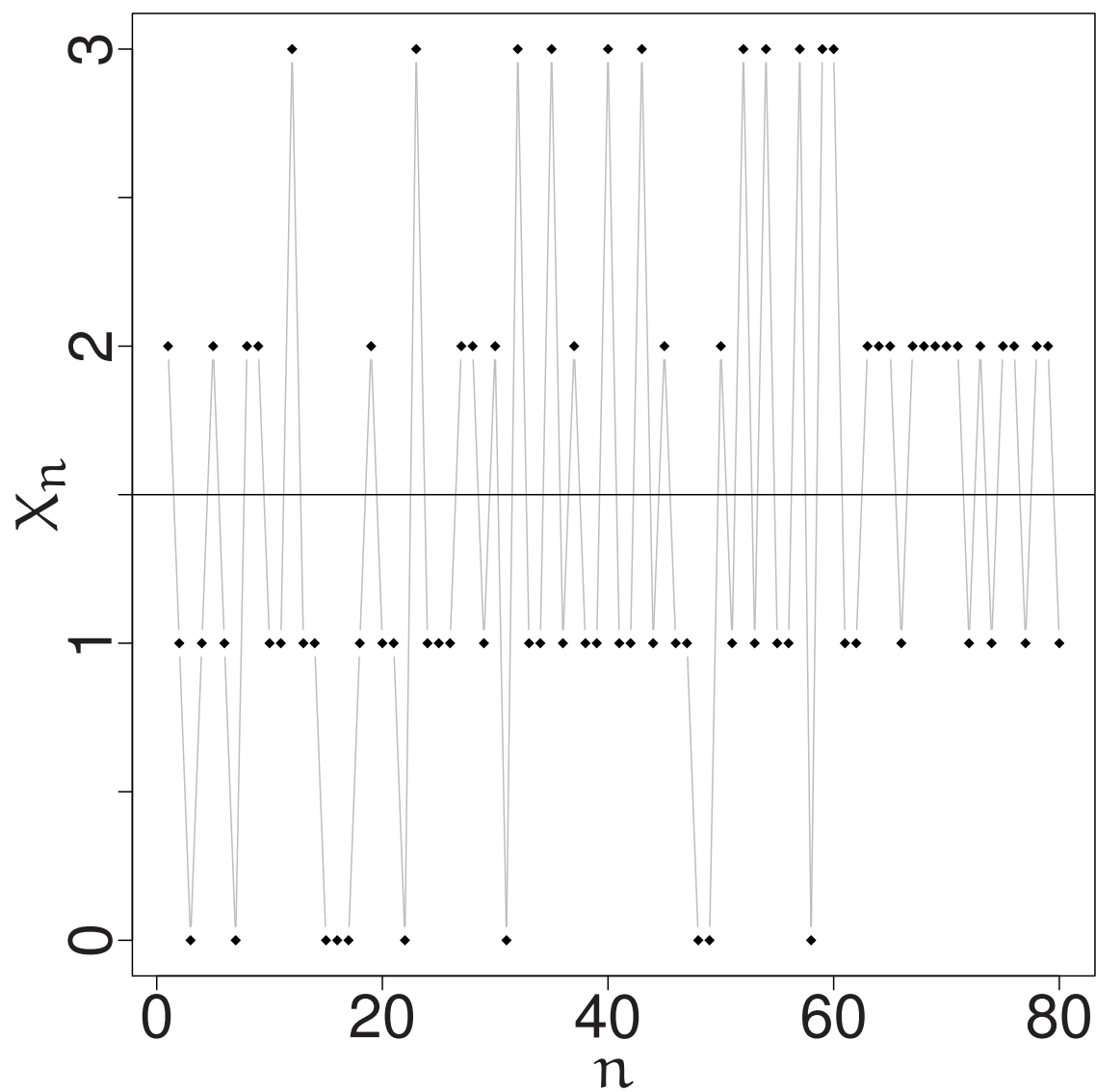


Der Erwartungswert als Langzeitmittel

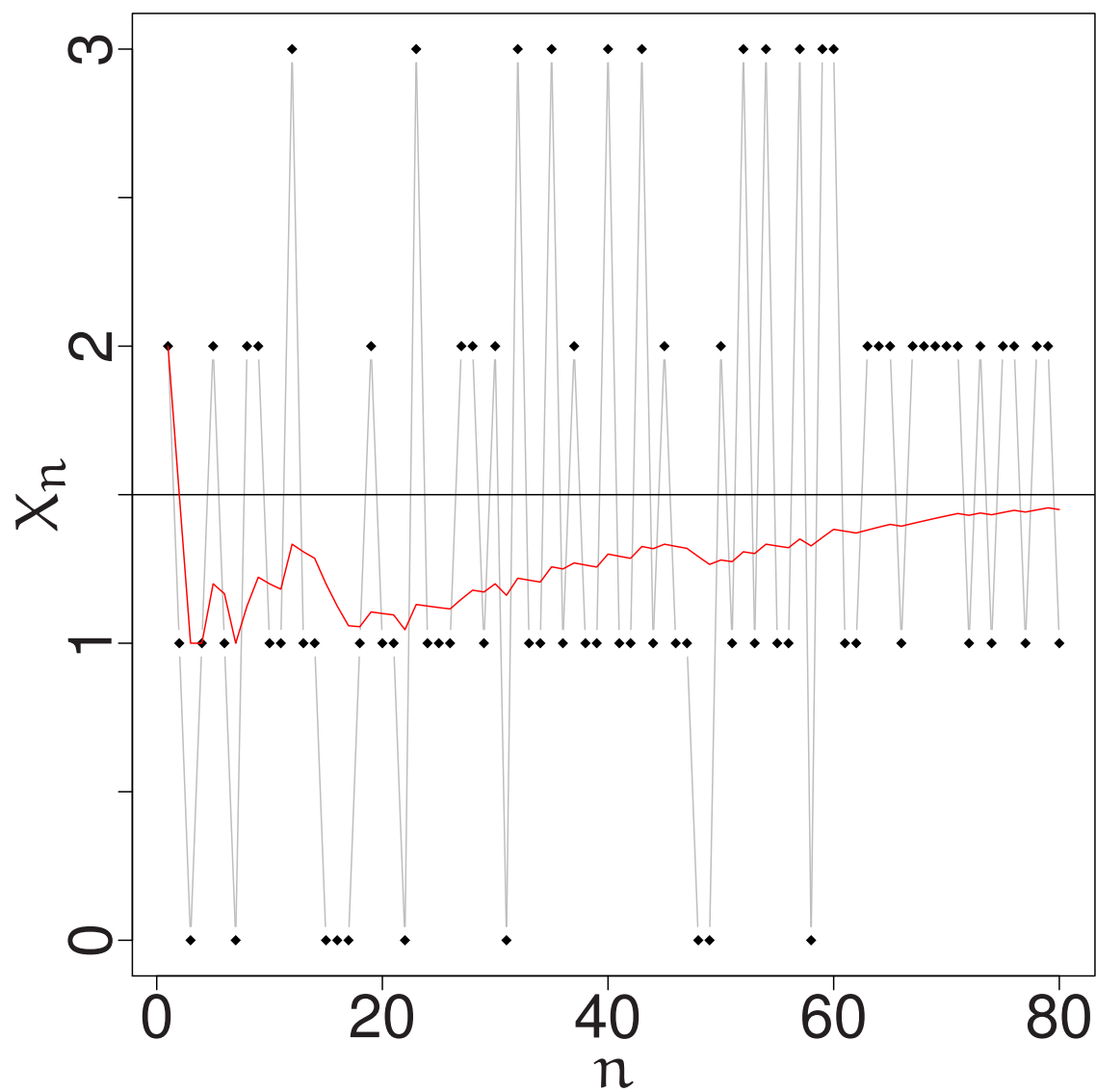
Beispiel:

X ... Anzahl Köpfe beim dreimaligen fairen Münzwurf

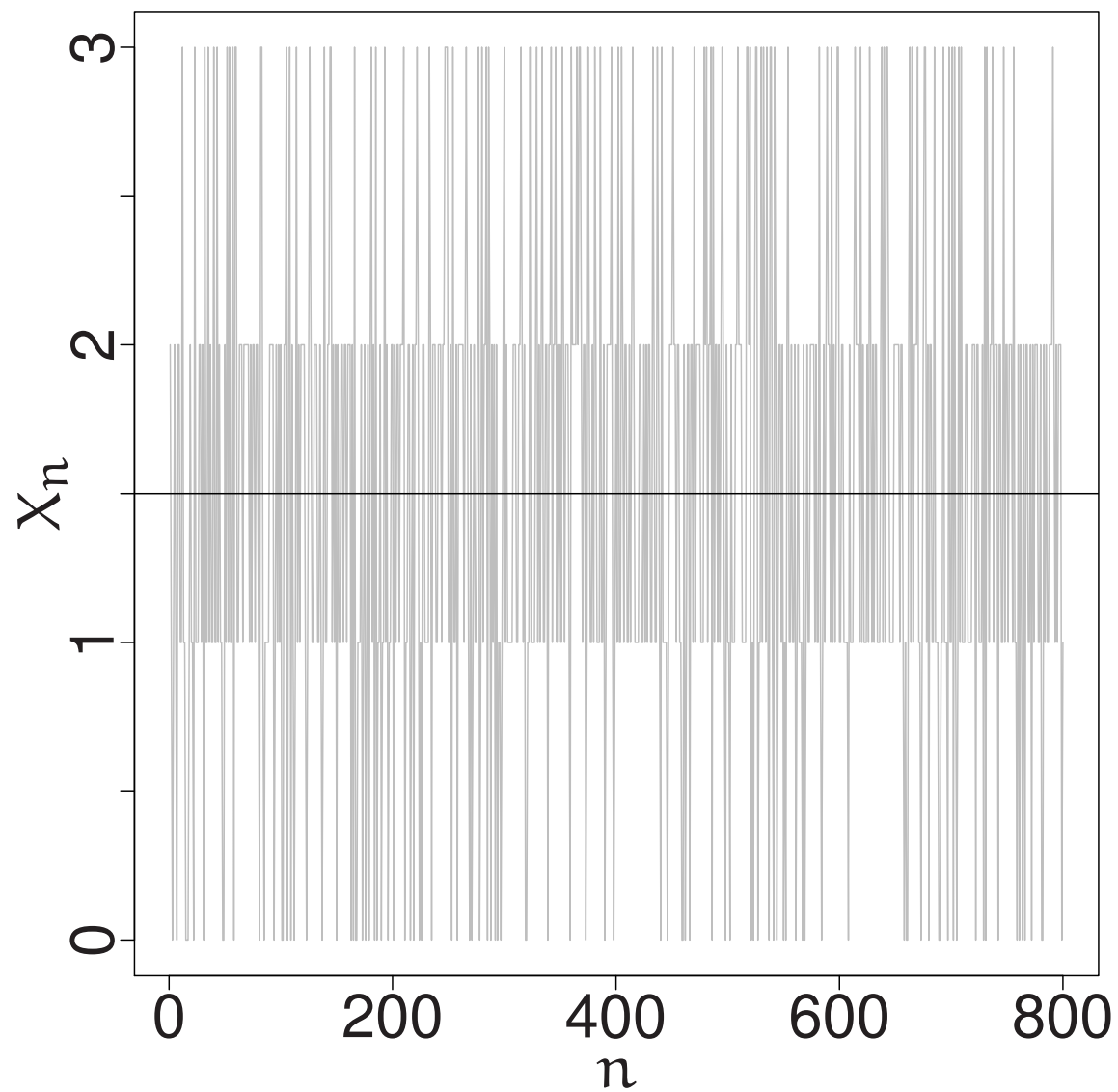
80 Wiederholungen: $X_1, X_2, \dots, X_{80}$



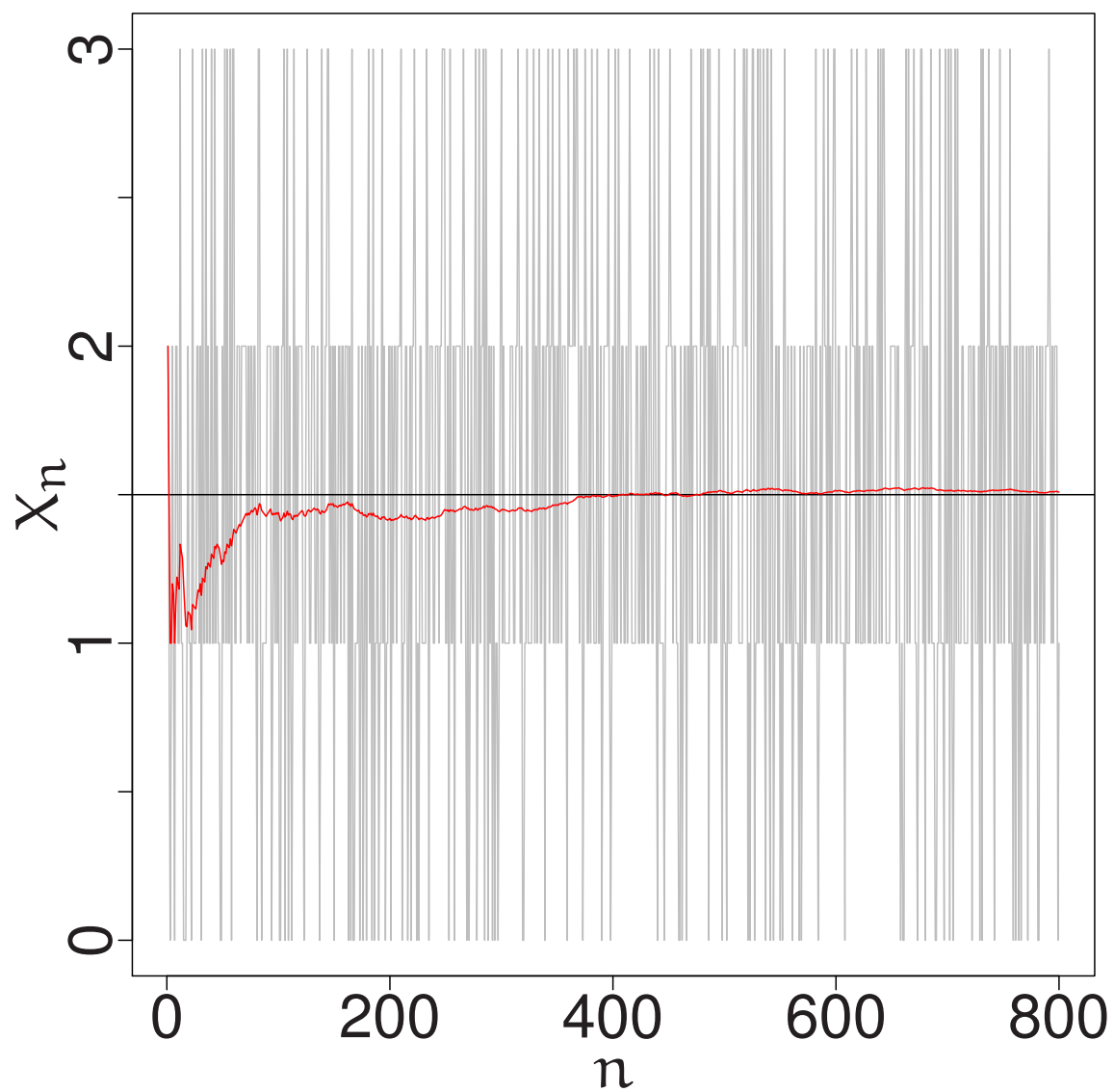
$$M_n := (X_1 + X_2 + \dots + X_n) / n$$



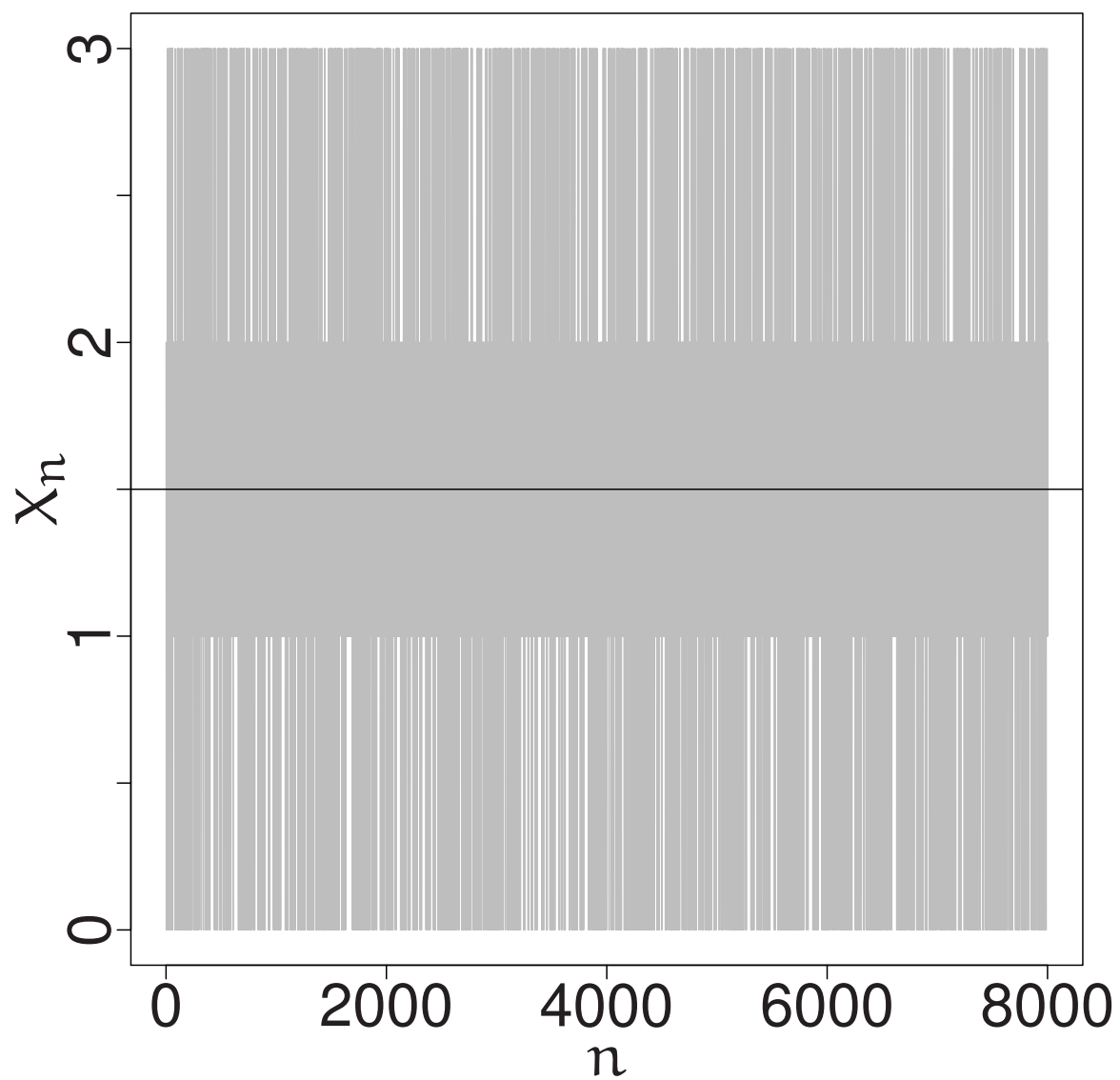
800 Wiederholungen: $X_1, X_2, \dots, X_{800}$



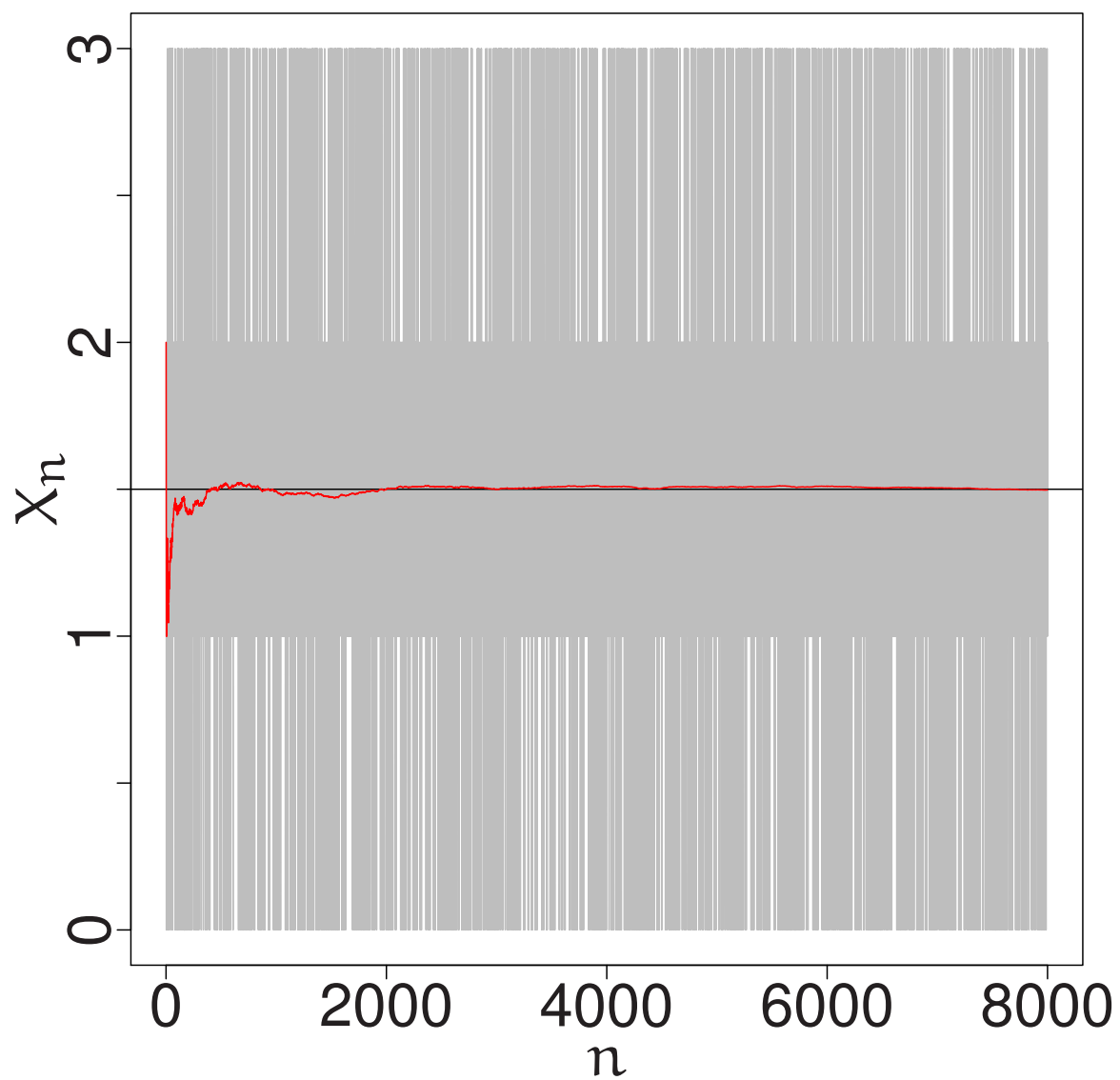
$$M_n := (X_1 + X_2 + \dots + X_n) / n$$



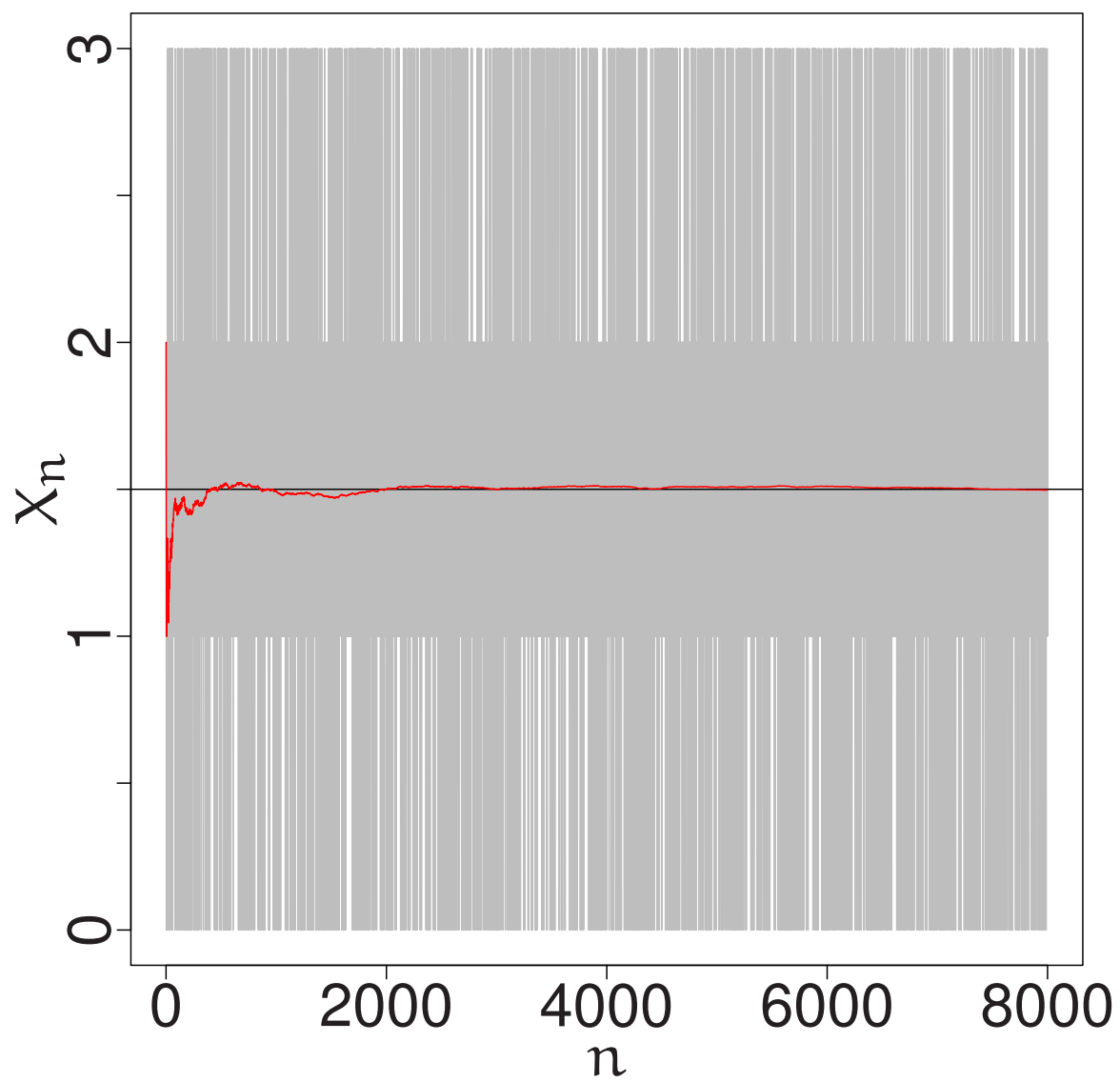
8000 Wiederholungen: $X_1, X_2, \dots, X_{8000}$



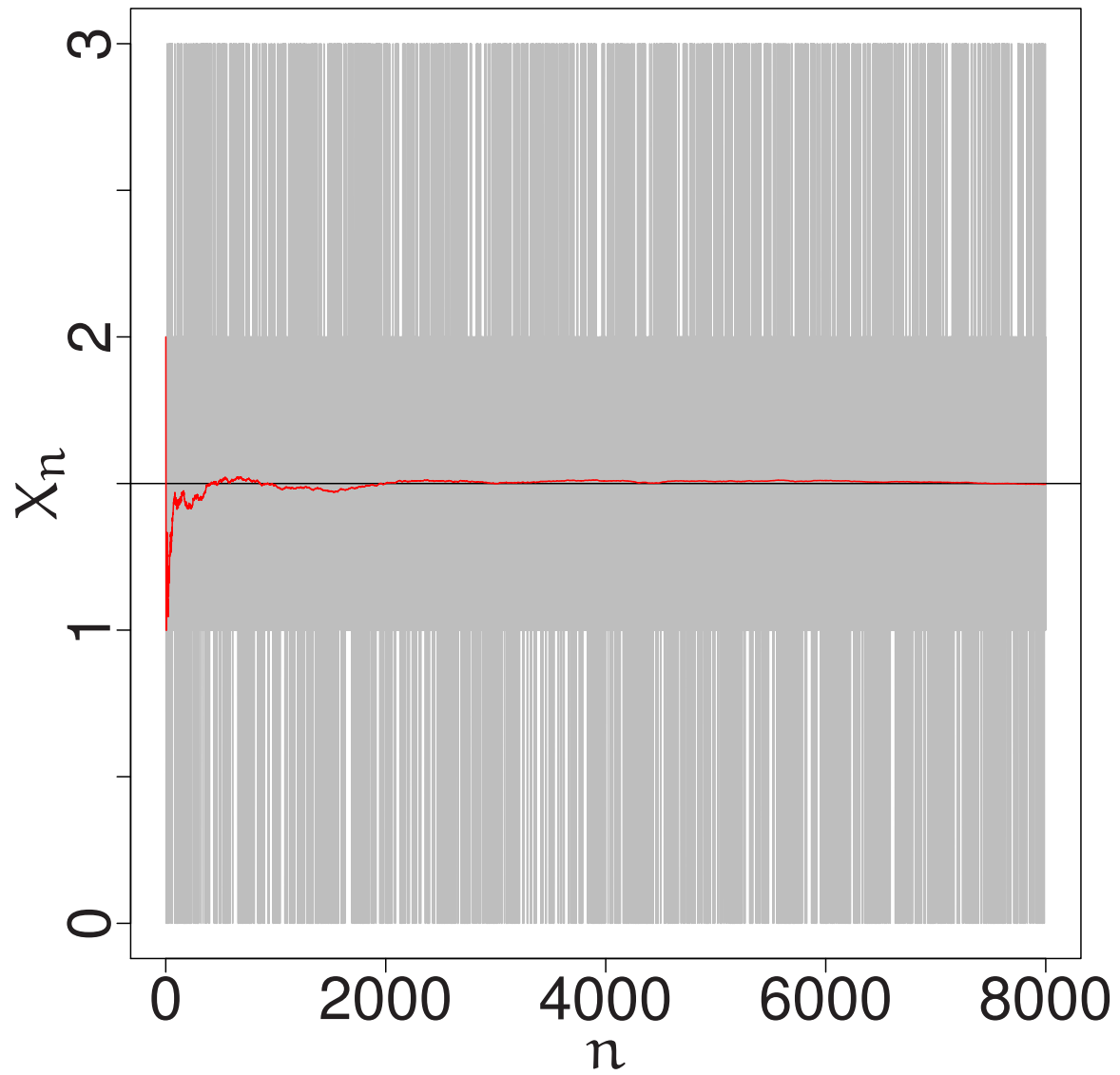
$$M_n := (X_1 + X_2 + \dots + X_n) / n$$



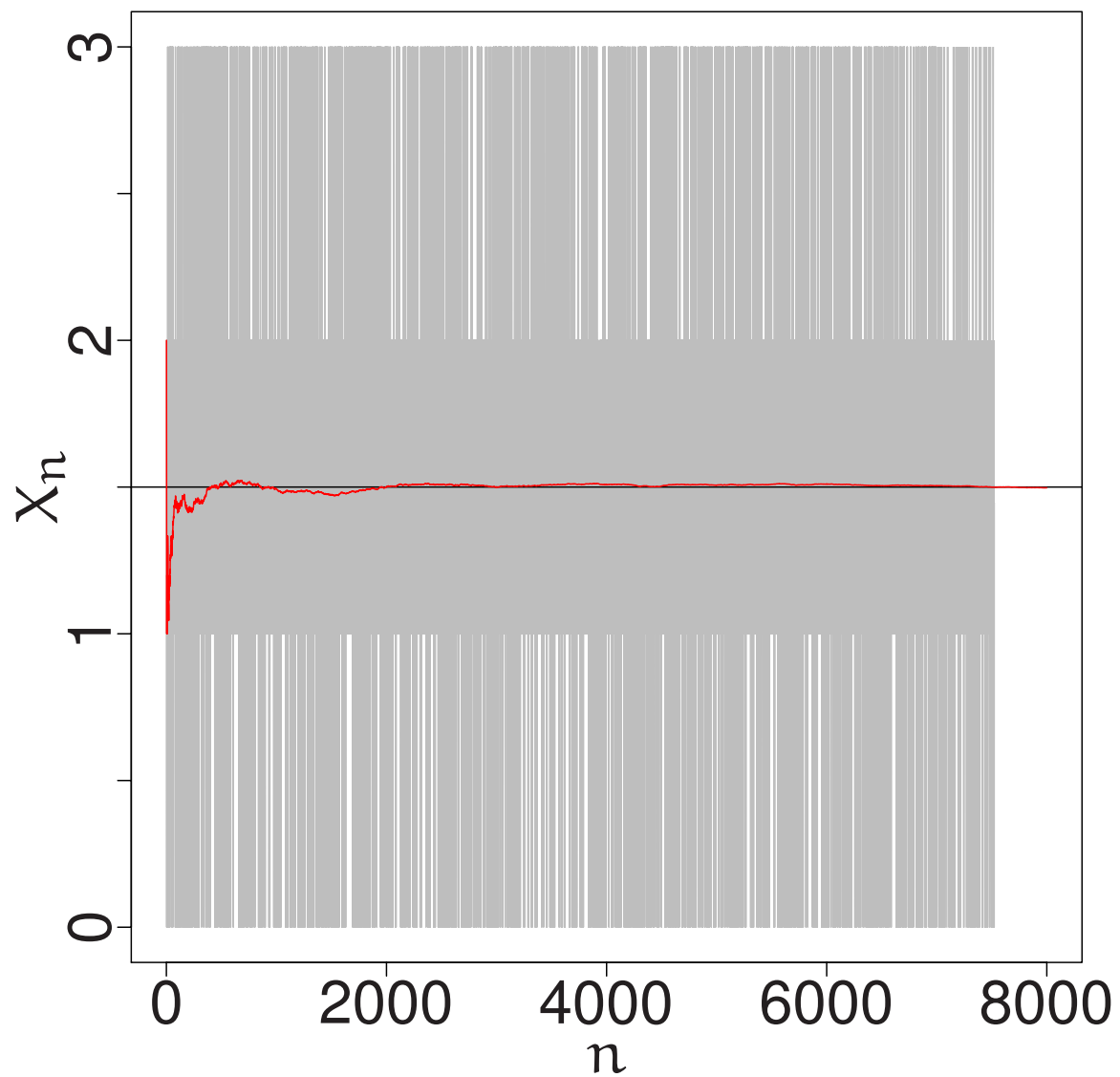
$$M_n \rightarrow \mathbf{E}[X]$$



Warum?



$$M_n = (X_1 + X_2 + \dots + X_n) / n$$



Die Summe von n Zahlen $x_1, \dots, x_n \in \{0, 1, 2, 3\}$
kann man auch so berechnen:

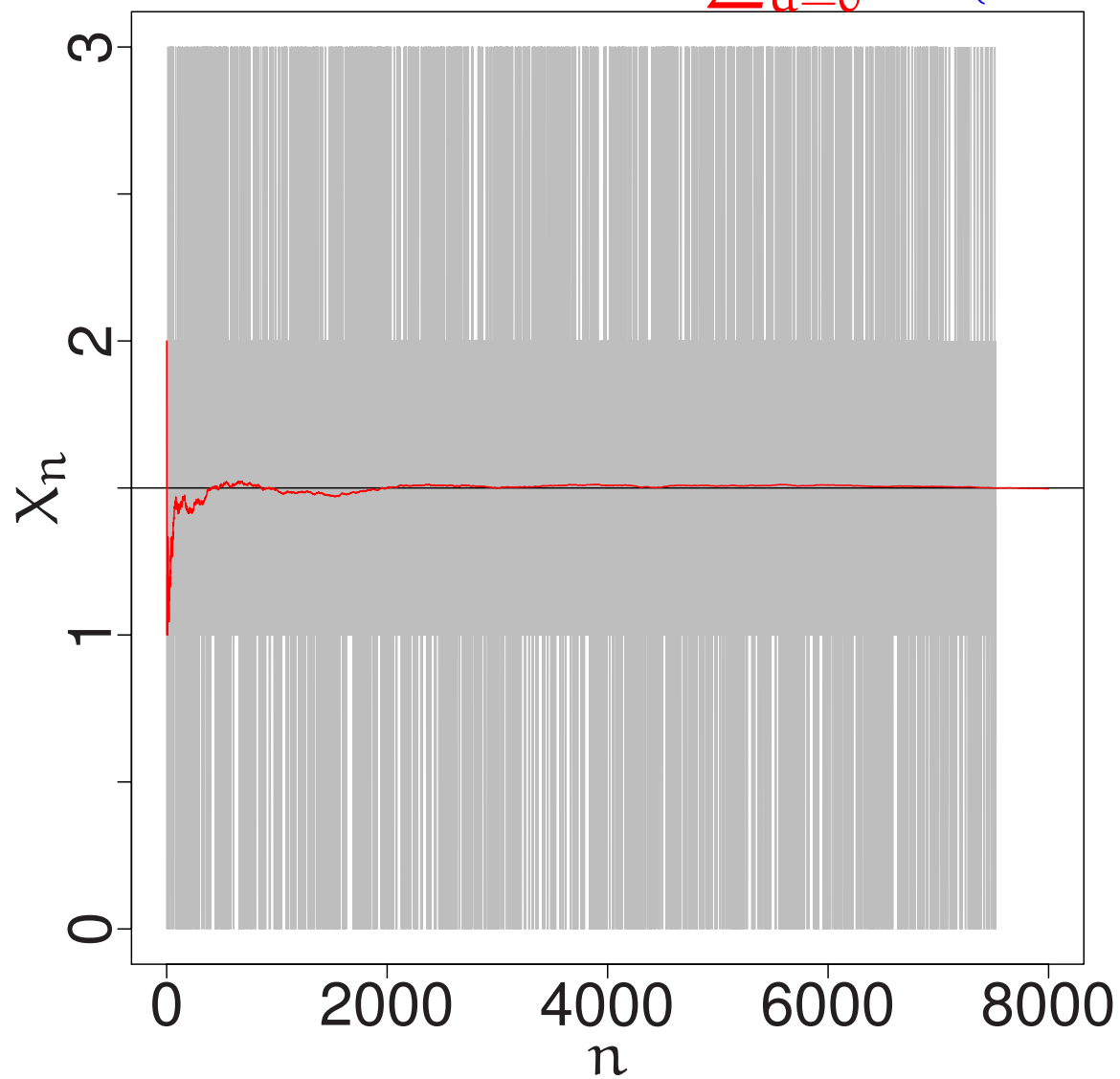
Man zählt für $a = 1, 2, 3$, wieviele der x_i gleich a sind

und bekommt

$$x_1 + \dots + x_n = \sum_{a=0}^3 a \cdot \#\{i \leq n : x_i = a\}.$$

$$M_n = \sum_{a=0}^3 a \#\{i \leq n : X_i = a\}/n$$

$$\rightarrow \sum_{a=0}^3 a \mathbf{P}(X = a)$$



Dazu später mehr.

Für den Moment nur als kurzer Ausblick:

DAS GESETZ DER GROSSEN ZAHLEN

Sei X eine Zufallsgröße mit Erwartungswert $\mathbf{E}[X]$.

Seien $X_1, X_2, \dots$ unabhängige Kopien von X .

Dann gilt

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mathbf{E}[X]$$

Zu klären

1. Was heißt „unabhängig“?

2. Was heißt „ $\rightarrow$ “?

Diese Klärung wird in der Vorlesung
in wenigen Wochen erfolgen.

Jetzt halten wir erst einmal fest:

Zwei Vorstellungen von $\mathbf{E}[X]$

1. Gewichtetes Mittel

der möglichen Werte:

$$\mathbf{E}[X] := \sum a \mathbf{P}(X = a)$$

2. Langzeitmittelwert

bei “unabhängigen” Wiederholungen:

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mathbf{E}[X]$$

Zusammenfassung

des Wichtigsten

A.

Was ist der Erwartungswert?

$$\mathbf{E}[X] = \sum a \mathbf{P}(X = a)$$

und

$$\mathbf{E}[X] = \lim \frac{X_1 + \dots + X_n}{n}$$

für “unabhängige Wiederholungen” $X_1, X_2, \dots$

B.

Was ist die wichtigste Eigenschaft des Erwartungswertes?

Die Linearität:

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y]$$

C.

Wie berechnet man $\mathbf{E}[X]$ am besten?

Oft dadurch,
dass man X als Summe schreibt:

$$X = Z_1 + \dots + Z_n$$

$$\mathbf{E}[X] = \mathbf{E}[Z_1] + \dots + \mathbf{E}[Z_n]$$

Vorlesung 3b

Indikatorvariable

Rechnen mit Ereignissen und
Wahrscheinlichkeiten.

Teil 1

Ereignisse und ihre Indikatorvariablen

(Buch S. 36-37)

Wir wissen schon:

**Ein- und dasselbe Ereigniss kann man
auf verschiedene Weisen darstellen:**

Beispiel:

Sei $X = (X_1, X_2)$ das Paar der Augenzahlen
beim zweimaligen (gewöhnlichen) Würfeln. Dann gilt:

$$\{X_1 = 3\} = \{X_1 = 3, X_2 \in \{1, 2, 3, 4, 5, 6\}\}$$

Allgemeiner:

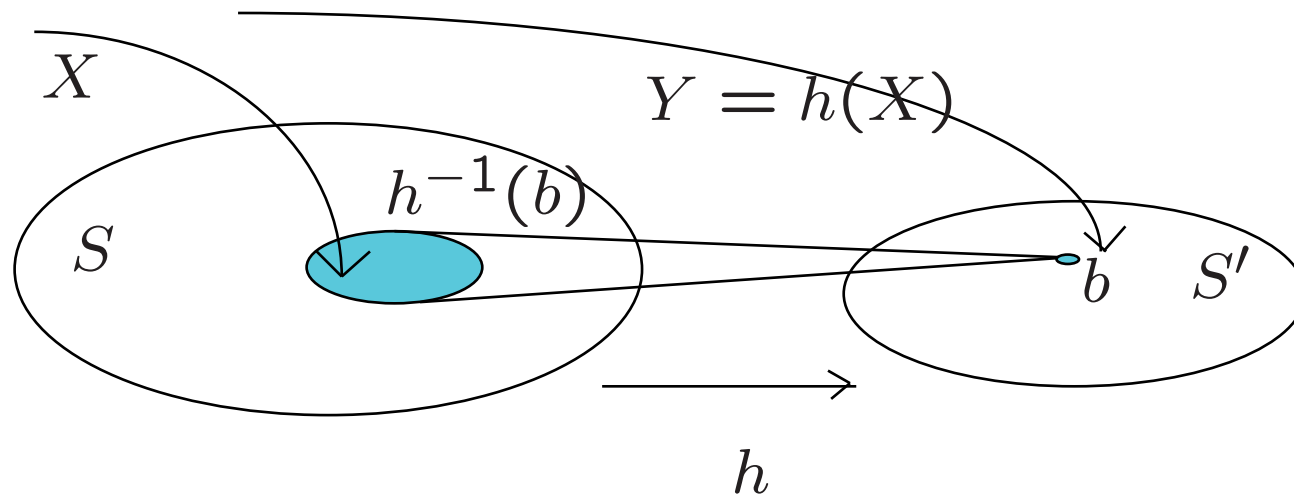
Sei $X = (X_1, X_2)$ ein zufälliges Paar
mit Wertebereich $S_1 \times S_2$.

Dann ist für $a_1 \in S_1$

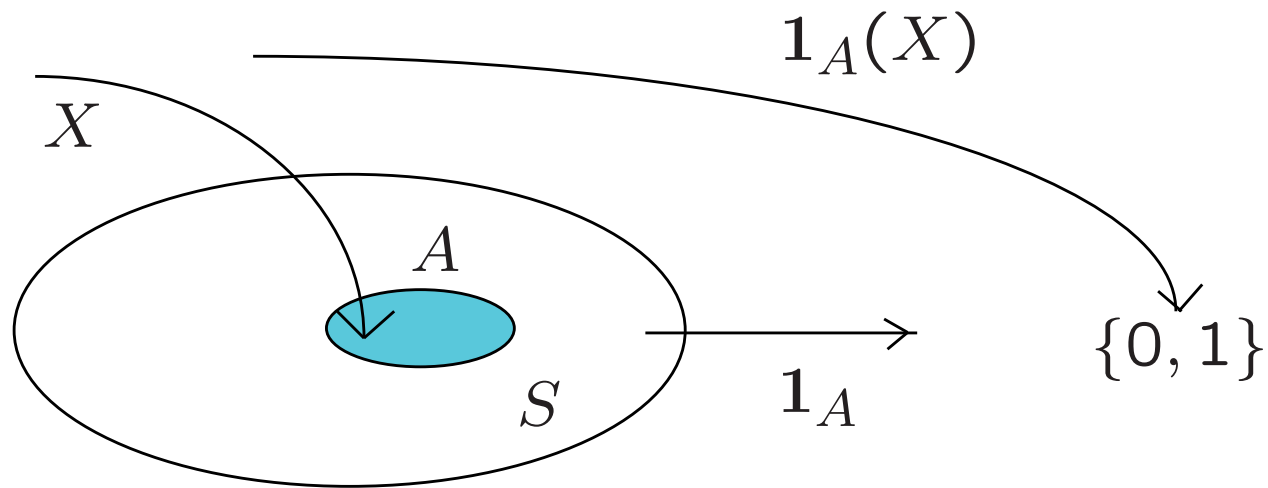
$$\{X_1 = a_1\} = \{X_1 \in a_1, X_2 \in S_2\}.$$

Noch allgemeiner gilt
für die “Verarbeitung” $Y = h(X)$ einer Zufallsvariablen X :

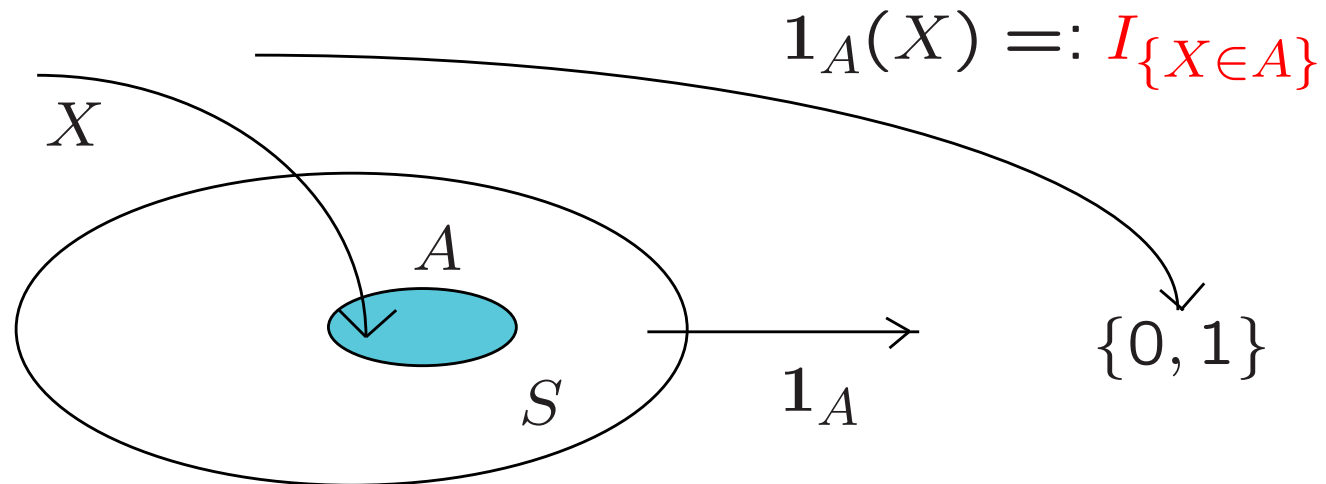
$$\{Y = b\} = \{X \in h^{-1}(b)\}.$$



Ein wichtiger Spezialfall hiervon ist die Verarbeitung von X mittels der Indikatorfunktion $\mathbf{1}_A$ einer Teilmenge A von S :

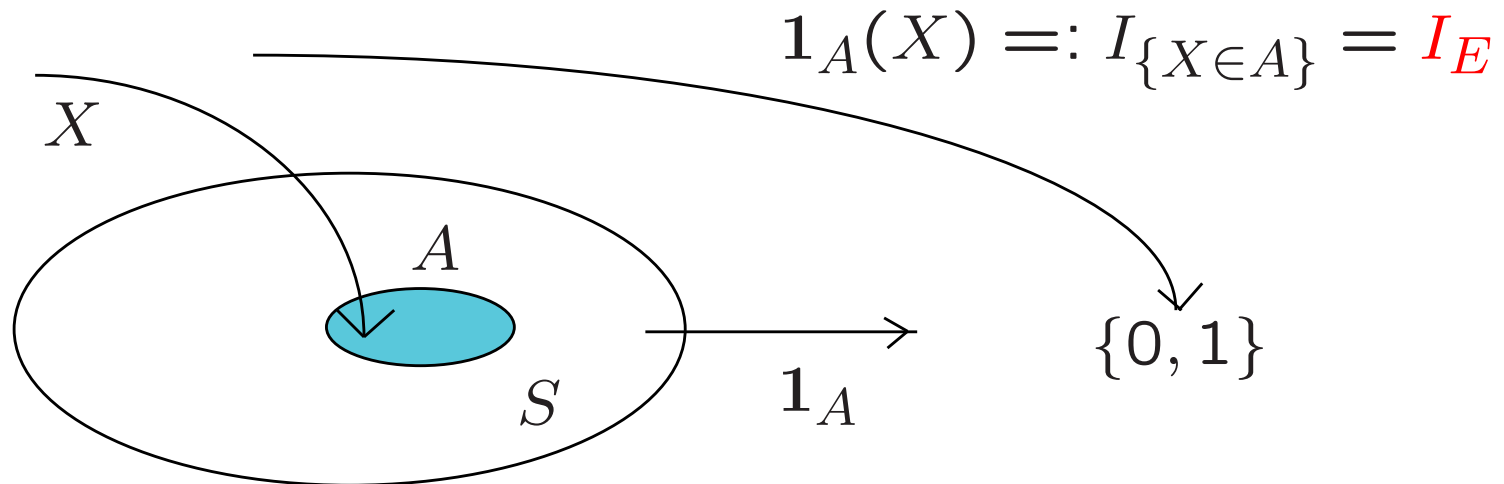


$$\{X \in A\} = \{\mathbf{1}_A(X) = 1\}$$



$$\{X \in A\} = \{1_A(X) = 1\}$$

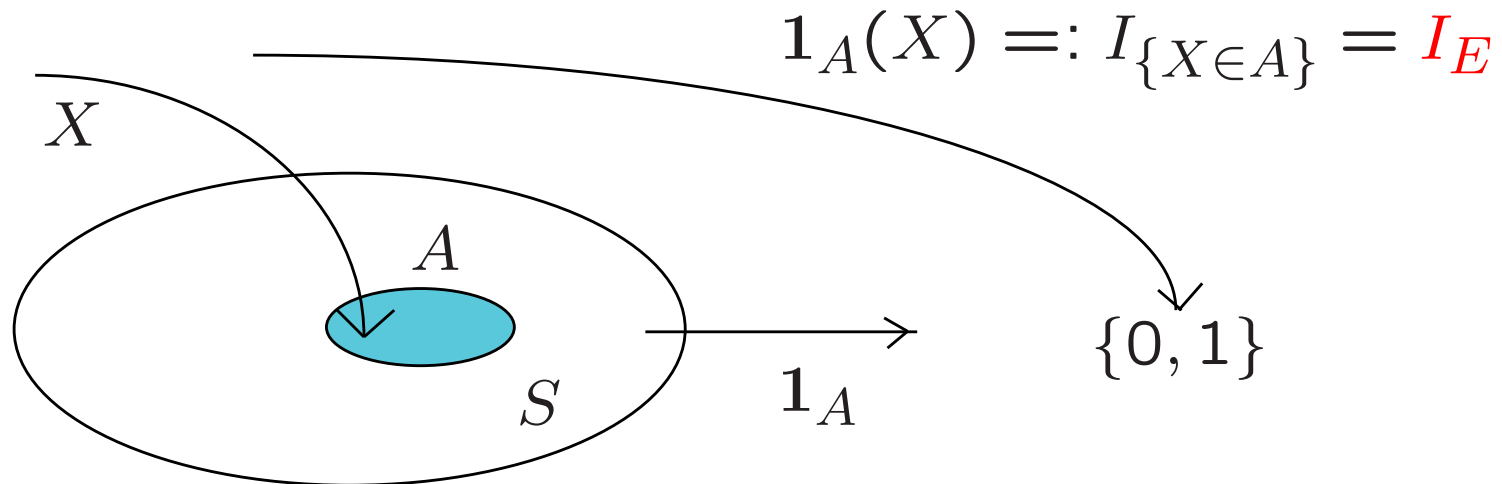
$I_{\{X \in A\}}$ heißt **Indikatorvariable** des Ereignisses $\{X \in A\}$



$$E := \{X \in A\} = \{1_A(X) = 1\}$$

$$E = \{I_E = 1\}$$

I_E fällt auf den Wert 1 genau dann,
wenn das Ereignis E eintritt.



$$E := \{X \in A\} = \{1_A(X) = 1\}$$

$$E = \{I_E = 1\}$$

I_E fällt auf den Wert 0 genau dann,
wenn das Ereignis E nicht eintritt.

$$E = \{I_E = 1\}.$$

Ereignisse sind gleich,
wenn ihre Indikatorvariablen gleich sind
(in dem Sinn, dass beide stets gemeinsam auf 0
oder gemeinsam auf 1 fallen)

Wann “sind zwei Zufallsvariablen gleich”?

Intuitiv dann, wenn stets
die eine auf denselben Wert fällt wie die andere.

Das präzisieren wir im Folgenden.

Das sichere Ereignis

Für jede Zufallsvariable X mit Wertebereich S gilt:

$$I_{\{X \in S\}} = 1_S(X)$$

ist eine Zufallsvariable, die “sicher” auf den Ausgang 1 fällt.

Das *sichere Ereignis* E_S ist dadurch charakterisiert, dass seine Indikatorvariable stets auf den Wert 1 fällt:

$$I_{E_S} = 1.$$

Die Aussage $X = Y$ und das Ereignis $\{X = Y\}$

Seien X, Y Zufallsvariable mit demselben Wertebereich S .

$D := \{(x, y) \in S^2 : x = y\}$, die „Diagonale“ in S^2 .

Wir definieren **das Ereignis „ X und Y fallen gleich aus“**

als

$$\{X = Y\} := \{(X, Y) \in D\}$$

Ist dieses Ereignis gleich dem sicheren Ereignis,
so schreiben wir dafür kurz:

$$X = Y$$

Vorlesung 3b

Indikatorvariable

Rechnen mit Ereignissen und
Wahrscheinlichkeiten.

Teil 2

Rechnen mit Ereignissen

(Buch S. 37-38)

Das *sichere Ereignis* E_S hatten wir dadurch charakterisiert, dass seine Indikatorvariable stets den Wert 1 annimmt:

$$I_{E_S} = 1$$

Das *unmögliche Ereignis* E_U ist dadurch charakterisiert, dass seine Indikatorvariable stets auf den Wert 0 fällt:

$$I_{E_U} = 0.$$

Für zwei Ereignisse E_1, E_2
hat deren “Oder-Ereignis” die Indikatorvariable

$$I_{E_1 \cup E_2} := \max(I_{E_1}, I_{E_2})$$

und deren “Und-Ereignis” die Indikatorvariable

$$I_{E_1 \cap E_2} := \min(I_{E_1}, I_{E_2}).$$

Man nennt $E_1 \cup E_2$ bzw. $E_1 \cap E_2$ auch
die “Vereinigung” bzw. den “Durchschnitt”
der Ereignisse E_1 und E_2 .

Für $b_1, b_2 \in \{0, 1\}$ gilt die Identität:

$$b_1 + b_2 = \max(b_1, b_2) + \min(b_1, b_2).$$

Dies überträgt sich auf eine Gleichheit von Zufallsvariablen:

$$I_{E_1} + I_{E_2} = I_{E_1 \cup E_2} + I_{E_1 \cap E_2}$$

Falls

$$E_1 \cap E_2 = E_u ,$$

so heißen E_1 und E_2

disjunkte oder sich ausschließende Ereignisse.

Gilt $E_1 = E_1 \cap E_2$, so schreiben wir

$$E_1 \subset E_2 .$$

und sagen:

“Mit E_1 tritt sicher auch E_2 ein”

oder auch

“Das Ereignis E_1 zieht das Ereignis E_2 nach sich.”

Für jedes Ereignis E ist sein *Komplementärereignis*

$$E^c$$

definiert durch

$$I_{E^c} := 1 - I_E \quad \text{bzw.} \quad E^c := \{I_E = 0\} .$$

Wegen $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$ gilt

$$\{X \in A\}^c = \{X \in A^c\} .$$

Vorlesung 3b

Indikatorvariable

Rechnen mit Ereignissen und
Wahrscheinlichkeiten.

Teil 3

Rechnen mit Wahrscheinlichkeiten.

(Buch S. 57)

Für Ereignisse und ihre Indikatorvariable
gelten die Beziehungen

$$E = \{I_E = 1\},$$

$$\mathbf{E}[I_E] = \mathbf{P}(I_E = 1) = \mathbf{P}(E).$$

Aus dem Rechnen mit Indikatorvariablen
und aus der Linearität des Erwartungswertes
ergeben sich die Regeln
für das Rechnen mit Wahrscheinlichkeiten:

$$(i) \quad \mathbf{P}(E_S) = 1, \quad \mathbf{P}(E_U) = 0,$$

denn

eine $\{0, 1\}$ -wertige ZV'e, die stets den Wert 1 annimmt,
hat Erwartungswert 1.

Das passt auch zu unserer Vereinbarung der ersten Stunde:

Für eine S -wertige Zufallsvariable X ist

$$\mathbf{P}(X \in S) = 1.$$

$$(ii) \quad \mathbf{P}(E_1) + \mathbf{P}(E_2) = \mathbf{P}(E_1 \cup E_2) + \mathbf{P}(E_1 \cap E_2)$$

Das sieht man aus der Identität

$$b_1 + b_2 = \max(b_1, b_2) + \min(b_1, b_2), \quad b_1, b_2 \in \{0, 1\}$$

mit ihrer Folgerung

$$I_{E_1} + I_{E_2} = I_{E_1 \cup E_2} + I_{E_1 \cap E_2}$$

zusammen mit der Linearität des Erwartungswertes.

Aus (ii) folgt sofort die **Subadditivität**:

$$\mathbf{P}(E_1 \cup E_2) \leq \mathbf{P}(E_1) + \mathbf{P}(E_2)$$

Eine weitere unmittelbare Konsequenz aus (ii) ist

$$(iii) \quad \mathbf{P}(E_1 \cup E_2) = \mathbf{P}(E_1) + \mathbf{P}(E_2)$$

falls E_1 und E_2 disjunkt.

“(endliche) Additivität”

In der Wahrscheinlichkeitstheorie arbeitet man mit einer stärkeren Form dieser Eigenschaft, der σ -Additivität. Sie besagt:

Für unendliche Folgen $E_1, E_2, \dots$ von paarweise disjunkten Ereignissen gilt $\mathbf{P}(E_1 \cup E_2 \cup \dots) = \mathbf{P}(E_1) + \mathbf{P}(E_2) + \dots$.

Diese Formel entspricht zusammen mit der Eigenschaft (i) den Axiomen von Kolmogorov für Wahrscheinlichkeitsmaße.

Eine unmittelbare Konsequenz aus
der Additivität des Erwartungswertes ist

$$(iv) \quad \mathbf{P}(E^c) = 1 - \mathbf{P}(E)$$

(Regel von der Gegenwahrscheinlichkeit)

$$\text{denn: } I_{E^c} = 1 - I_E.$$

Für je zwei Ereignisse E_1, E_2 gilt die folgende
“Zerlegung von E_2 ”:

$$E_2 = (E_1 \cap E_2) \cup (E_1^c \cap E_2),$$

dabei sind $E_1 \cap E_2$ und $E_1^c \cap E_2$ disjunkt.

Also folgt aus der Additivität:

$$\mathbf{P}(E_1 \cap E_2) \leq \mathbf{P}(E_2).$$

Daraus ergibt sich

$$(v) \quad \mathbf{P}(E_1) \leq \mathbf{P}(E_2)$$

falls das Ereignis E_1 das Ereignis E_2 nach sich zieht

(d.h. falls gilt: $E_1 = E_1 \cap E_2$)

Wir fassen zusammen:

$$(i) \quad \mathbf{P}(E_s) = 1, \quad \mathbf{P}(E_u) = 0.$$

$$(ii) \quad \mathbf{P}(E_1) + \mathbf{P}(E_2) = \mathbf{P}(E_1 \cup E_2) + \mathbf{P}(E_1 \cap E_2),$$

insbesondere $\mathbf{P}(E_1 \cup E_2) \leq \mathbf{P}(E_1) + \mathbf{P}(E_2)$.

$$(iii) \quad \mathbf{P}(E_1 \cup E_2) = \mathbf{P}(E_1) + \mathbf{P}(E_2)$$

falls E_1 und E_2 disjunkt.

$$(iv) \quad \mathbf{P}(E^c) = 1 - \mathbf{P}(E).$$

$$(v) \quad \mathbf{P}(E_1) \leq \mathbf{P}(E_2), \text{ falls } E_1 \subset E_2.$$

Vorlesung 3b

Indikatorvariable

Rechnen mit Ereignissen und
Wahrscheinlichkeiten.

Teil 4

Die Einschluss-Ausschlussformel

(Buch S. 57-58)

Die im vorigen Teil diskutierte Eigenschaft

$$(ii) \quad P(E_1) + P(E_2) = P(E_1 \cup E_2) + P(E_1 \cap E_2)$$

bzw. äquivalent dazu

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

wird verallgemeinert durch die

Einschluss-Ausschluss-Formel:

$$\begin{aligned} &P(E_1 \cup \dots \cup E_n) \\ &= \sum_i P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \dots - \dots \pm P(E_1 \cap \dots \cap E_n) . \end{aligned}$$

Beweis der Einschluss-Ausschluss-Formel:

$$1 - I_{E_1 \cup \dots \cup E_n}$$

fällt genau dann auf 0,

wenn mindestens eines der I_{E_i} auf 1 fällt,

ist also gleich dem Produkt

$$(1 - I_{E_1}) \cdots (1 - I_{E_n})$$

Ausmultiplizieren ergibt

$$1 - \sum_i I_{E_i} + \sum_{i < j} I_{E_i \cap E_j} - \cdots$$

Gehe dann über zum Erwartungswert. $\square$

Beispiel (vgl Buch S. 58). $X = (X_1, \dots, X_n)$ sei eine rein zufällige Permutation von $1, \dots, n$.

Was ist die W'keit, dass X mindestens einen Fixpunkt hat?

Sei $E_i := \{X_i = i\}$. Wir arbeiten mit der E-A-Formel.

Offenbar gilt für $i_1 < \dots < i_k$:

$$\mathbf{P}(E_{i_1} \cap \dots \cap E_{i_k}) = \frac{(n - k)!}{n!}$$

(denn es gibt $(n - k)!$ Permutationen von $1, \dots, n$
mit Fixpunkten bei $i_1, \dots, i_k$)

Für jedes k gibt es $\binom{n}{k}$ Mögl'k'ten, $i_1 < \dots < i_k$ zu wählen.

Also:

$$\sum_{i_1 < i_2 < \dots < i_k} \mathbf{P}(E_{i_1} \cap \dots \cap E_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Mit der E-A-Formel folgt für die gefragte W'keit

$$\mathbf{P}(E_1 \cup \dots \cup E_n) = 1 - \frac{1}{2!} + \dots \pm \frac{1}{n!}. \quad \square$$

(Für $n \rightarrow \infty$ konvergiert das übrigens gegen $1 - e^{-1}$,
denn aus der Mathe 1 ist bekannt:

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!} = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \dots,$$

$$e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots$$

Vorlesung 3b

Indikatorvariable

Rechnen mit Ereignissen und
Wahrscheinlichkeiten.

Teil 5

Positivität und Monotonie des Erwartungswertes

(Buch S. 55)

Wir beweisen jetzt
(hier nur für *diskrete* reellwertige Zufallsvariable)

zwei weitere fundamentale
Eigenschaften des Erwartungswerts:
die Positivität und die Monotonie.

Als Vorbereitung dazu ist hier ein
Nachtrag zu Teil 1 der heutigen Vorlesung.

Die Aussage “ $X \geq 0$ ” und das Ereignis $\{X \geq 0\}$:

X sei eine reellwertige Zufallsvariable.

Die Aussage “ $X \geq 0$ ”
definieren wir als gleichbedeutend damit,
dass $\{X \geq 0\}$ das sichere Ereignis ist.

Ist $\{X \geq 0\} = E_S$, dann können wir wahlweise $[0, \infty)$
oder $\mathbb{R}$ (oder jede andere Obermenge von $[0, \infty)$)
als Wertebereich von X verwenden.

Die Aussage $X \leq Y$ und das Ereignis $\{X \leq Y\}$:

Es sei $H := \{(x, y) \in \mathbb{R}^2 : x \leq y\}$,
der Halbraum über (und einschließlich) der Diagonalen.

Für reellwertige Zufallsvariable X, Y setzen wir

$$\{X \leq Y\} := \{(X, Y) \in H\}.$$

Die Aussage “ $X \leq Y$ ”
definieren wir als gleichbedeutend damit,
dass $\{X \leq Y\}$ das sichere Ereignis ist.

Positivität des Erwartungswertes

Für die reellwertige Zufallsvariable X gelte $X \geq 0$. Dann gilt

$$(i) \mathbf{E}[X] \geq 0,$$

$$(ii) \mathbf{E}[X] = 0 \text{ genau dann, wenn } \mathbf{P}(X = 0) = 1.$$

Monotonie des Erwartungswertes

Für reellwertige Zufallsvariable $X_1 \leq X_2$
mit wohldefinierten Erwartungswerten gilt

$$\mathbf{E}[X_1] \leq \mathbf{E}[X_2].$$

Positivität

Für die reellwertige Zufallsvariable X gelte $X \geq 0$. Dann gilt

(i) $E[X] \geq 0$,

(ii) $E[X] = 0$ genau dann, wenn $P(X = 0) = 1$.

Wir geben hier einen Beweis nur im diskreten Fall:

Nach Voraussetzung (siehe die “Vorbereitung” am Beginn dieses Teils) können wir $[0, \infty)$ als Wertebereich ansehen.

Weil X als diskret vorausgesetzt war, existiert dann eine abzählbare Teilmenge $\tilde{S} \subset [0, \infty)$ mit $P(X \in \tilde{S}) = 1$.

Dann gilt auch $P(X \in S) = 1$ mit $S := \tilde{S} \cup \{0\}$.

Nach der Definition des Erwartungswerts aus V3b1 gilt:

$$\mathbf{E}[X] = \sum_{a \in S} a \mathbf{P}(X = a) = \sum_{a \in S: a > 0} a \mathbf{P}(X = a) \quad (1)$$

Aus $\mathbf{P}(X = S) = 1$ folgt

$$1 = \mathbf{P}(X = 0) + \sum_{a \in S: a > 0} \mathbf{P}(X = a) \quad (2)$$

Aus (1) folgt sofort: $\mathbf{E}[X] \geq 0$ also die Aussage (i). Weiter gilt:

$$\mathbf{P}(X = 0) = 1$$

$$\stackrel{(2)}{\iff}$$

$$\mathbf{P}(X = a) = 0 \text{ für alle strikt positiven } a \in S$$

$$\stackrel{(1)}{\iff}$$

$$\mathbf{E}[X] = 0 \quad \square.$$

Monotonie

Für reellwertige Zufallsvariable $X_1 \leq X_2$
mit wohldefiniertem Erwartungswert gilt

$$\mathbf{E}[X_1] \leq \mathbf{E}[X_2].$$

Beweis:

$X_1 \leq X_2$ ist gleichbedeutend mit $X_2 - X_1 \geq 0$.

Aus der Positivität und der Linearität des Erwartungswertes

folgt $\mathbf{E}[X_2] - \mathbf{E}[X_1] \geq 0$. $\square$

Die Ungleichung von Markov

X reellwertige Zufallsvariable mit $X \geq 0$, $c > 0$. Dann gilt

$$\mathbf{P}(X \geq c) \leq \frac{1}{c} \mathbf{E}[X]$$

Beweis:

Wegen $c1_{[c,\infty)}(a) \leq a$, $a \geq 0$,

gilt

$$cI_{\{X \geq c\}} \leq X.$$

Aus Linearität und Monotonie des Erwartungswertes folgt:

$$c\mathbf{E}[I_{\{X \geq c\}}] \leq \mathbf{E}[X]. \quad \square$$

Vorlesung 4a

Versuche, Erfolge, Wartezeiten:

Die Welt des p -Münzwurfs -
von Bernoulli zu Poisson

Teil 1

Der fortgesetzte p -Münzwurf

(Buch S. 19-20)

Zur Erinnerung:

Der n -fache p -Münzwurf ...

... ist eine $\{0, 1\}^n$ -wertige ZV'e $(Z_1, \dots, Z_n)$

mit

$$(*) \quad \mathbb{P}(Z_1 = a_1, \dots, Z_n = a_n) = p^k (1-p)^{n-k}$$

falls $a_1 + \dots + a_n = k$.

Ereignisse kann man oft auf verschiedene Weise darstellen.

Für einen $(n + 1)$ -fachen p -Münzwurf gilt z.B.

$$\begin{aligned} &\{Z_1 = a_1, \dots, Z_n = a_n\} \\ &= \{Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 1\} \\ &\quad \cup \{Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 0\} \end{aligned}$$

Weil rechts zwei disjunkte Ereignisse stehen, muss gelten:

$$\begin{aligned} &\mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) \\ &= \mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 1) \\ &\quad + \mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 0) \end{aligned}$$

Die Definition (*) auf der vorigen Folie ist damit verträglich:

$$\begin{aligned}
& \mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) \\
&= \mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 1) \\
&\quad + \mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n, Z_{n+1} = 0)
\end{aligned}$$

$$\begin{aligned}
& p^k(1-p)^{n-k} \\
&= p^{k+1}(1-p)^{n-k} \\
&\quad + p^k(1-p)^{n-k+1}
\end{aligned}$$

Dies zeigt, dass auch die folgende Definition
konsistent über n ist.

Definition: Sei $p \in (0, 1)$, $q := 1 - p$.

Eine *Bernoulli-Folge zum Parameter p*

(man sagt manchmal auch: ein *fortgesetzter p -Münzwurf*)

ist eine zufällige 01-Folge $(Z_1, Z_2, \dots)$, deren Verteilung die folgende Eigenschaft hat:

Für jedes $n \in \mathbb{N}$ und jede endliche 01-Folge $(a_1, \dots, a_n)$
mit k Einsen und $n - k$ Nullen ist

$$\mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) = p^k q^{n-k}.$$

(d.h. für jedes n ist $(Z_1, \dots, Z_n)$ ein n -facher p -Münzwurf)

Wir wissen schon:

Für jedes n ist dann

die Anzahl der Einsen in $(Z_1, \dots, Z_n)$

(die “Anzahl der Erfolge in n Versuchen”)

binomial(n, p)-verteilt:

$$\mathbf{P}(Z_1 + \dots + Z_n = k) = \binom{n}{k} p^k q^{n-k}.$$

Vorlesung 4a

Versuche, Erfolge, Wartezeiten:

Teil 2

Der Zeitpunkt des ersten Erfolgs
und die geometrische Verteilung

(Buch S. 34-35)

$(Z_1, Z_2, \dots)$ sei ein fortgesetzter p -Münzwurf.

$$T := \min\{i : i \in \mathbb{N}, Z_i = 1\}$$

ist der *Zeitpunkt des ersten Erfolges*.

Wie sieht die Verteilung von T aus?

$$\mathbf{P}(T = n) = ?$$

$$\{T = n\} = \{Z_1 = 0, \dots, Z_{n-1} = 0, Z_n = 1\}$$

Also:

$$\mathbf{P}(T = n) = \mathbf{P}(Z_1 = 0, \dots, Z_{n-1} = 0, Z_n = 1)$$

$$= q^{n-1} p.$$

Alternativ:

$$\{T > n\} = \{Z_1 = 0, \dots, Z_n = 0\}$$

Also

$$\mathbf{P}(T > n) = q^n.$$

$$\mathbf{P}(T = n) = q^{n-1} p$$

$$\mathbf{P}(T > n) = q^n$$

Das passt zusammen.

Denn das Ereignis $\{T > n - 1\}$ ist die disjunkte Vereinigung der beiden Ereignisse $\{T = n\}$ und $\{T > n\}$.

Dementsprechend:

$$q^{n-1} = q^{n-1} p + q^n.$$

Definition

Sei $p \in (0, 1)$. Eine Zufallsvariable T mit Zielbereich $\mathbb{N}$ heißt

geometrisch verteilt mit Parameter p ,

kurz **Geom(p)-verteilt**,

wenn

$$\mathbf{P}(T > a) = q^a, \quad a = 0, 1, 2, \dots,$$

mit $q := 1 - p$.

$$\mathbf{E}[T] = ?$$

Anschaulich ist klar:

Beim gewöhnlichen Würfeln kommt im Mittel
jedes 6-te Mal eine Sechs.

Beim Münzwurf mit Erfolgswahrscheinlichkeit p
kommt im Mittel jedes $(1/p)$ -te Mal ein Erfolg.

Also wird gelten:

$$\mathbf{E}[T] = \frac{1}{p}.$$

Das beweist man auch schnell
mit dem folgenden

Lemma zum Erwartungswert $\mathbb{N}$ -wertiger ZV'er:
(Buch S. 34)

Ist X eine Zufallsvariable mit Zielbereich $\mathbb{N}$ oder $\mathbb{N}_0$, dann ist

$$\mathbf{E}[X] = \sum_{i \geq 0} \mathbf{P}(X > i)$$

Folgerung aus dem Lemma:

Für eine $\text{Geom}(p)$ -verteilte Zufallsvariable T ist

$$\mathbf{E}[T] = \sum_{i \geq 0} \mathbf{P}(T > i) = \sum_{i \geq 0} q^i = \frac{1}{1 - q} = \frac{1}{p}.$$

$$\boxed{\mathbf{E}[T] = \frac{1}{p}}$$

Lemma

Ist X eine Zufallsvariable mit Zielbereich $\mathbb{N}$ oder $\mathbb{N}_0$, dann ist

$$\mathbf{E}[X] = \sum_{i \geq 0} \mathbf{P}(X > i)$$

Beweis.

$\rho(j)$ seien die Verteilungsgewichte von X .

$$\mathbf{E}[X] = \sum_{j \geq 1} j \rho(j)$$

$$\sum_{i \geq 0} \mathbf{P}(X > i) = \sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j)$$

Warum sind die beiden rechten Seiten gleich?

Beweis.

$\rho(j)$ seien die Verteilungsgewichte von X .

$$\mathbf{E}[X] = \sum_{j \geq 1} j \rho(j)$$

$$\sum_{i \geq 0} \mathbf{P}(X > i) = \sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j)$$

Warum sind die beiden rechten Seiten gleich?

Beweis.

$\rho(j)$ seien die Verteilungsgewichte von X .

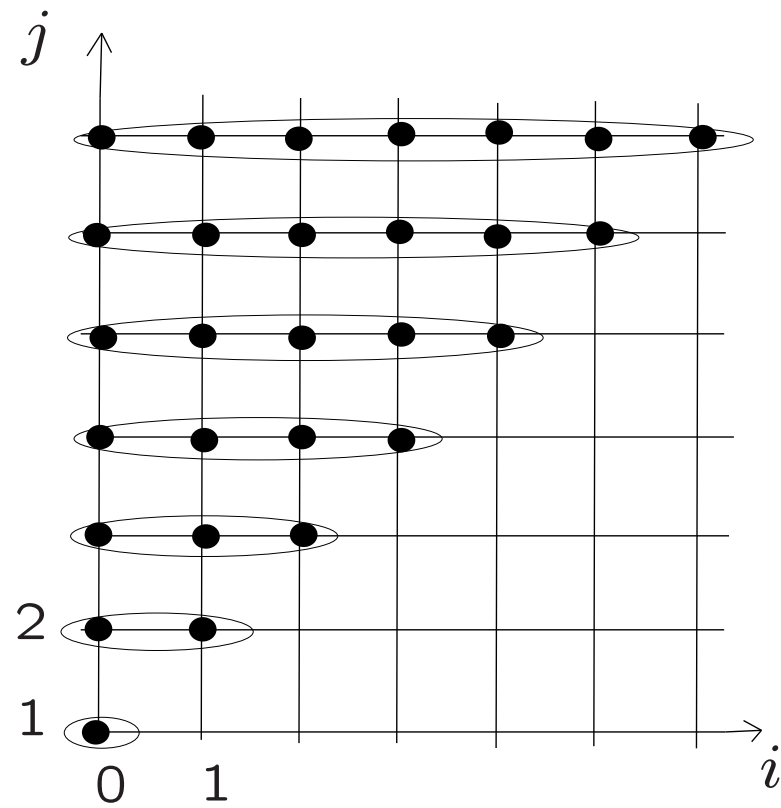
$$\mathbf{E}[X] = \sum_{j \geq 1} j \rho(j) = \sum_{j \geq 1} \sum_{i=0}^{j-1} \rho(j)$$

$$\sum_{i \geq 0} \mathbf{P}(X > i) = \sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j)$$

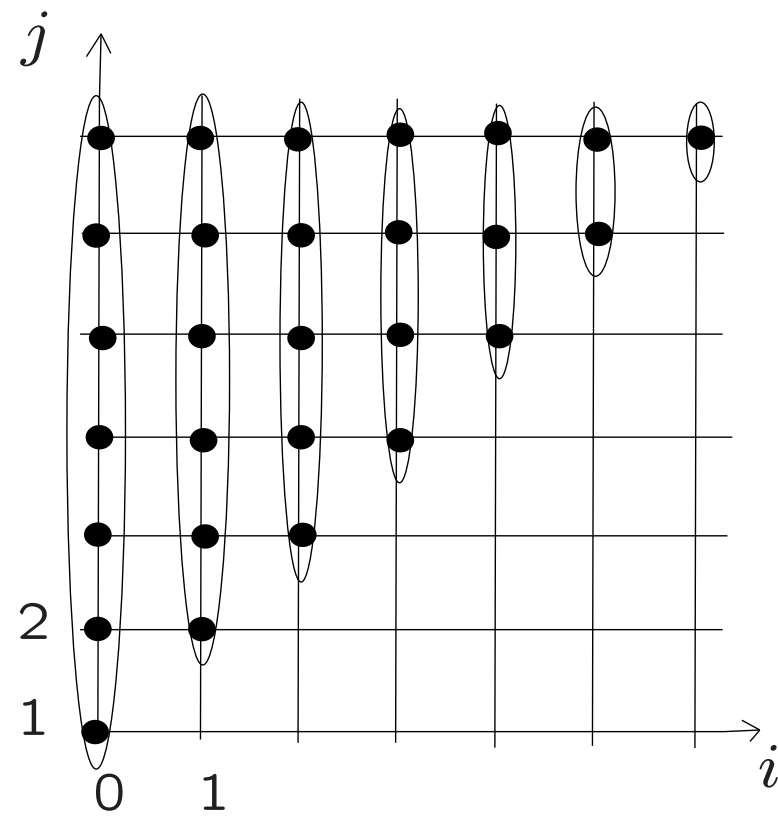
Warum sind die beiden rechten Seiten gleich?

Wie sieht man die Gleichheit

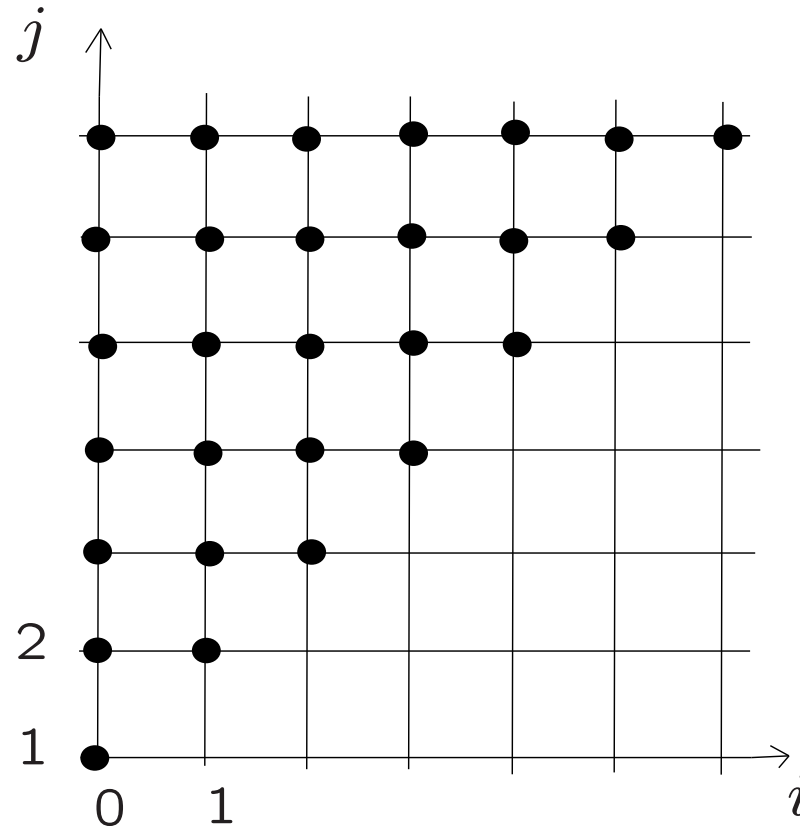
$$\sum_{j \geq 1} \sum_{i=0}^{j-1} \rho(j) = \sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j) \quad ?$$



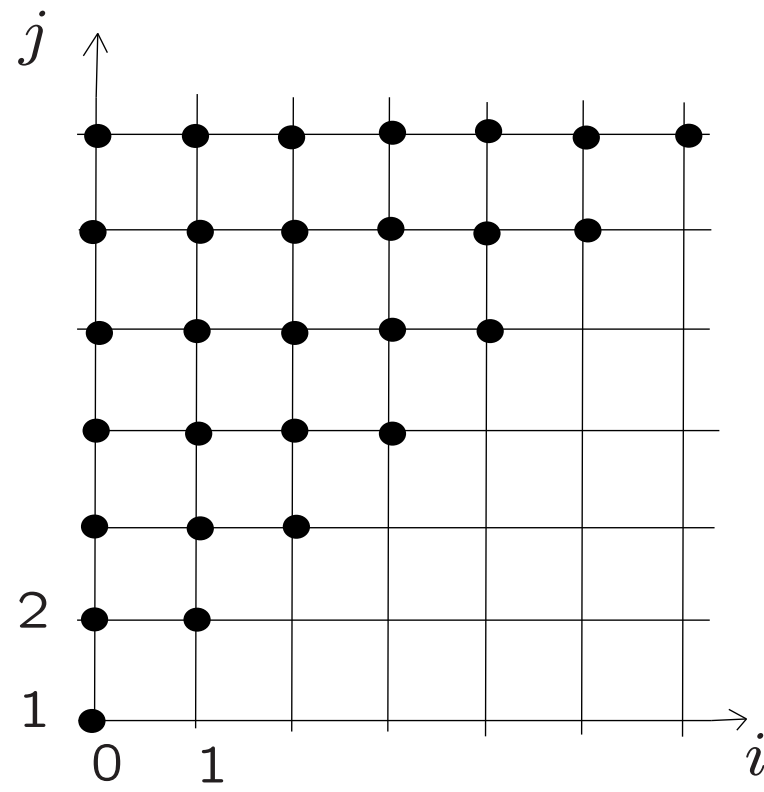
$$\sum_{j \geq 1} \sum_{i=0}^{j-1} \rho(j)$$



$$\sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j)$$



Es kommt bei nichtnegativen Summanden
nicht auf die Reihenfolge der Summation an!



$$\sum_{j \geq 1} \sum_{i=0}^{j-1} \rho(j) = \sum_{i \geq 0} \sum_{j=i+1}^{\infty} \rho(j) \quad \square$$

Vorlesung 4a

Versuche, Erfolge, Wartezeiten:

Teil 3

Die Exponentialapproximation.

Oder:

Münzwurf mit kleiner Erfolgswahrscheinlichkeit:

Wie lange dauert es bis zum ersten Erfolg?

(Buch S. 42)

Wieder sei

T

der zufällige Zeitpunkt des ersten Erfolgs
in einem fortgesetzten p -Münzwurf.

Beispiel:

$$p = \frac{1}{1000}$$

$$\mathbf{P}(T > 2000) = q^{2000} = \left(1 - \frac{1}{1000}\right)^{2000} \approx e^{-2}$$

$$\mathbf{P}(T > 2 \cdot \mathbf{E}[T]) \approx e^{-2}$$

$$\boxed{\mathbf{P}\left(\frac{T}{\mathbf{E}[T]} > 2\right) \approx e^{-2}}$$

Betrachten wir T auf der Skala seines Erwartungswertes:

$$\widetilde{T} := \frac{T}{\mathbf{E}[T]} = pT.$$

Für $t \in \mathbb{R}_+$ ist

$$\begin{aligned} \mathbf{P}\{\widetilde{T} > t\} &= \mathbf{P}\left(T > \frac{t}{p}\right) = \mathbf{P}\left(T > \left\lfloor \frac{t}{p} \right\rfloor\right) \\ &= (1 - p)^{\left\lfloor \frac{t}{p} \right\rfloor} \\ &= (1 - p)^{\frac{1}{p}} p^{\left\lfloor \frac{t}{p} \right\rfloor} \end{aligned}$$

Für $p \rightarrow 0$ konvergiert dies gegen

$$(e^{-1})^t = e^{-t}.$$

Diese Tatsache formulieren wir als einen *Grenzwertsatz*:

(vgl. Buch S. 42)

Satz Sei $T_1, T_2, \dots$ eine Folge von geometrisch verteilten Zufallsvariablen mit der Eigenschaft

$$\mathbf{E}[T_m] \xrightarrow{m \rightarrow \infty} \infty.$$

Dann gilt für jedes $t \geq 0$:

$$\mathbf{P} \left(\frac{T_m}{\mathbf{E}[T_m]} > t \right) \xrightarrow{m \rightarrow \infty} e^{-t}$$

Vorlesung 4a

Versuche, Erfolge, Wartezeiten:

Teil 4

Die Poissonapproximation.

Oder:

Münzwurf mit kleiner Erfolgswahrscheinlichkeit:

Wie ist die Anzahl der Erfolge verteilt
bei einer großen Zahl von Versuchen?

(Buch S. 29-30)

p klein, n groß

$$X := Z_1 + Z_2 + \dots + Z_n$$

$$\mathbf{P}(X = k) \approx ?$$

Beispiel:

$$p = \frac{1}{1000}, \quad n = 3000$$

$$\mathbf{P}(X = 0) = q^n = \left(1 - \frac{1}{1000}\right)^{3000} \\ \approx e^{-3}$$

$$\mathbf{P}(X = 1) = npq^{n-1} \approx 3e^{-3}$$

$$\mathbf{P}(X = 2) = \binom{n}{2} p^2 q^{n-2} \approx \frac{1}{2} (np)^2 q^n \approx \frac{1}{2} 3^2 e^{-3}$$

Clou:

p klein, n groß:

$$q^n = (1 - p)^n \approx e^{-np}$$

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{k!} n^k p^k q^n \approx \frac{1}{k!} (np)^k e^{-np}$$

Fazit

Sei p eine kleine positive Zahl,
 n eine große natürliche Zahl
und X eine $\text{Bin}(n, p)$ -verteilte Zufallsvariable.

Man kann dann die Verteilungsgewichte von X
approximativ als Funktion von $\mathbb{E}[X] = np$ ausdrücken.

Rigoros fasst man diese Behauptung im folgenden

Grenzwertsatz:

Satz (Poissons Gesetz der seltenen Ereignisse)

(vgl. Buch S. 30)

Sei $\lambda > 0$ und sei $X_n, n = 1, 2, \dots$,
eine Folge von $\text{Bin}(n, p_n)$ -verteilten Zufallsvariablen,

so dass für $n \rightarrow \infty$

$$\mathbf{E}[X_n] \rightarrow \lambda, \quad \text{d. h. } p_n \sim \frac{\lambda}{n}.$$

Dann gilt für jedes $k = 0, 1, 2, \dots$

$$\mathbf{P}(X_n = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Beweis:

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} =$$

$$\frac{1}{k!} \underbrace{\frac{n(n-1) \cdots (n-k+1)}{n^k}}_{\rightarrow 1} \underbrace{(np_n)^k}_{\rightarrow \lambda^k} \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\rightarrow e^{-\lambda}} \underbrace{(1 - p_n)^{-k}}_{\rightarrow 1}$$

$$\rightarrow \frac{1}{k!} \lambda^k e^{-\lambda} . \quad \square$$

Definition (Poissonverteilung)

(Buch S. 29)

Sei $\lambda \in \mathbb{R}_+$.

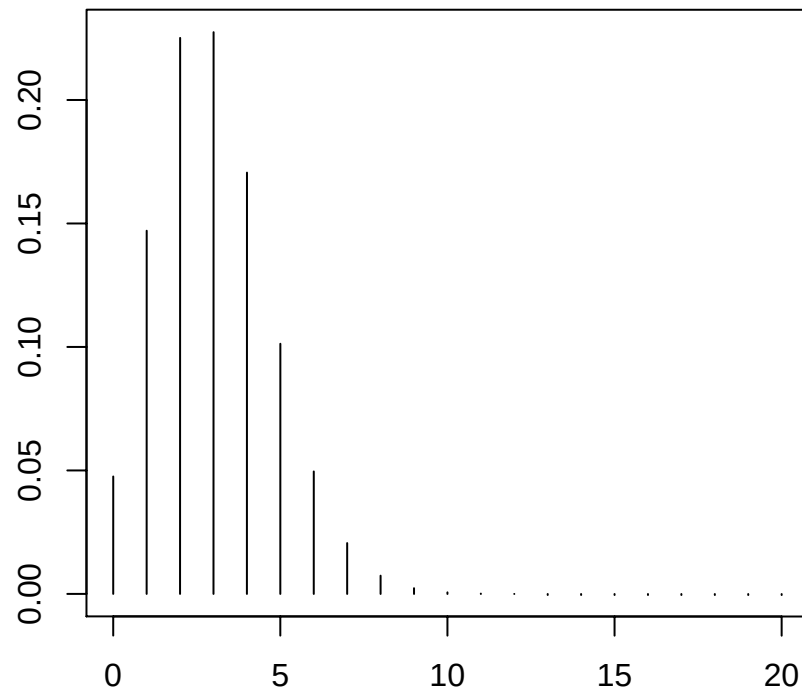
Eine Zufallsvariable X mit Zielbereich $\mathbb{N}_0$ heißt

Poissonverteilt mit Parameter λ ,

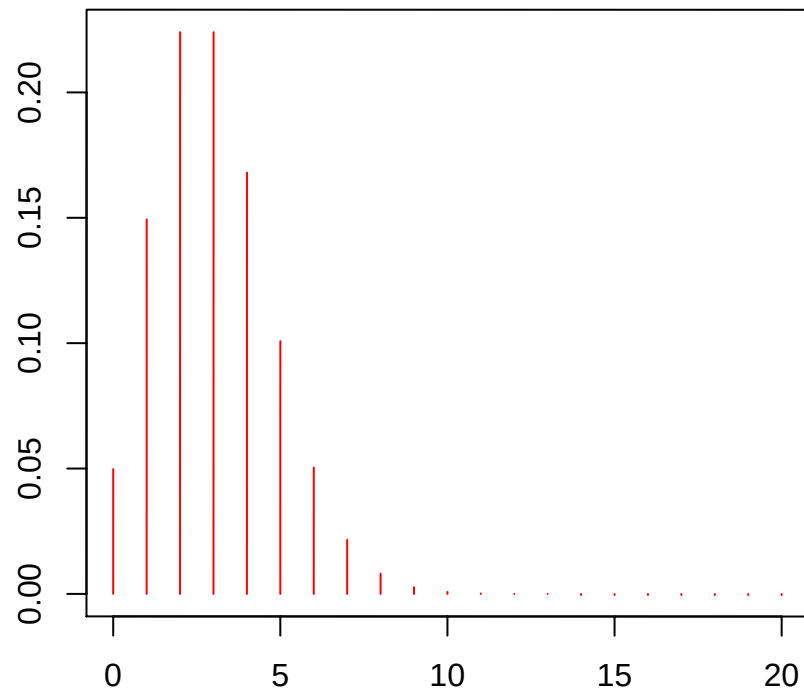
kurz $\text{Pois}(\lambda)$ -verteilt,

wenn

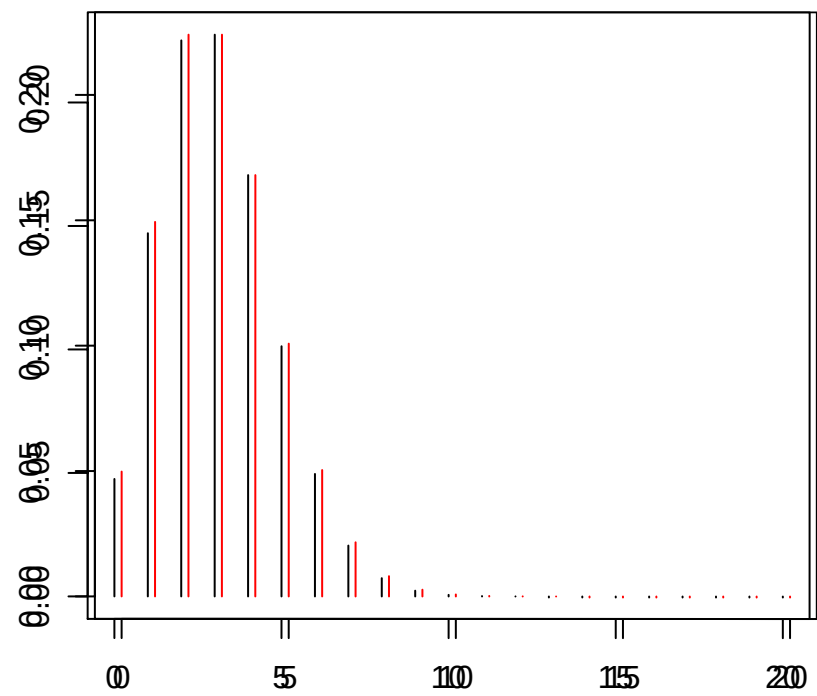
$$\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$



Binomialgewichte zu $n = 100$ und $p = 0.03$



Poissongewichte zum Parameter $\lambda = 3$



Satz.

Der Erwartungswert
einer $\text{Pois}(\lambda)$ -verteilten Zufallsvariablen X ist

$$\mathbf{E}[X] = \lambda.$$

Beweis:

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \cdot 1 \quad \square$$

Vorlesung 4b

Zufallsvariable mit Dichten

Teil 1

Kontinuierlich uniform verteilte Zufallsvariable

Kontinuierlich anstelle von diskret:

Bisher hatten wir im Fokus der Vorlesung:

Diskrete Zufallsvariable.

Sie fallen mit W'keit 1 in eine diskrete
(d.h. endliche oder abzählbar unendliche)

Menge S .

Jetzt wenden wir uns Zufallsvariablen zu,

die **kontinuierlich verteilt** sind.

Dann ist der Wertebereich *überabzählbar*.

Ein prominentes Beispiel ist der (faire) Münzwurf.

Man kann ihn auffassen als
rein zufällige Wahl eines Elementes aus $\{0, 1\}^{\mathbb{N}}$.

Ein weiteres einprägsames Beispiel ist die rein zufällige Wahl
eines Punktes aus dem Einheitsintervall
oder aus dem Einheitsquadrat.

Idee bei der rein zufälligen Wahl aus einem Kontinuum:

$P(X \in A)$ ist gegeben durch den Anteil von A an S

Uniforme Verteilung auf dem Einheitsintervall

Eine Zufallsvariable X mit Wertebereich $S = [0, 1]$ heißt

uniform verteilt auf S ,

wenn für alle $A \subset S$ mit wohldefiniertem Längenmaß $\lambda(A)$
gilt:

$$\mathbf{P}(X \in A) = \lambda(A)$$

(denn hier ist ja $\lambda(S) = 1$).

Beispiel 1:

$$A := [b, c] \quad \text{mit } 0 \leq b \leq c \leq 1$$



$$\mathbf{P}(X \in A) = c - b.$$

Beispiel 2:

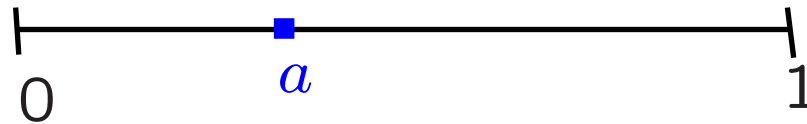
$$A := [b, c] \cup [\tilde{b}, \tilde{c}] \quad \text{mit } 0 \leq b \leq c < \tilde{b} \leq \tilde{c} \leq 1$$



$$\mathbf{P}(X \in A) = (c - b) + (\tilde{c} - \tilde{b}).$$

Beispiel 3:

$$A := \{a\} \quad \text{mit } 0 \leq a \leq 1$$



$$\mathbf{P}(X = a) = a - a = 0.$$

Uniforme Verteilung auf einem Rechteck in $\mathbb{R}^2$:

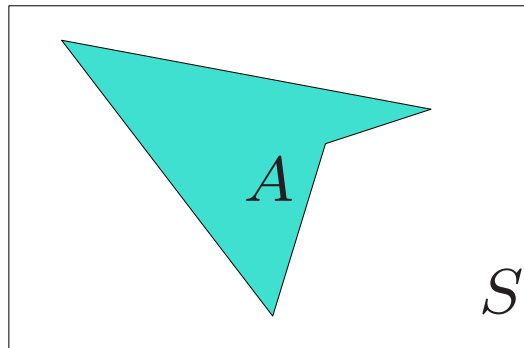
Eine Zufallsvariable X mit Zielbereich

$$S := [0, \ell] \times [0, b] \subset \mathbb{R}^2$$

heißt *uniform verteilt auf S* , wenn

für alle $A \subset S$ mit wohldefiniertem Flächenmaß $\lambda^2(A)$ gilt:

$$\mathbf{P}(X \in A) = \frac{\lambda^2(A)}{\lambda^2(S)} = \frac{\lambda^2(A)}{\ell \cdot b}.$$



Uniforme Verteilung auf einer Teilmenge des $\mathbb{R}^d$

Definition (Buch S. 12)

Sei S eine Teilmenge des $\mathbb{R}^d$ mit endlichem Inhalt $\lambda^d(S) > 0$.

Eine Zufallsvariable X mit Wertebereich S heißt

uniform verteilt auf S ,

wenn für alle $A \subset S$ mit wohldefiniertem (Volums-)Inhalt

$v(A) := \lambda^d(A)$ gilt:

$$\mathbf{P}(X \in A) = \frac{v(A)}{v(S)}.$$

$$\mathbf{P}(X \in A) = \frac{v(A)}{v(S)}.$$

Man beachte die Analogie zu
“Anzahl günstige durch Anzahl mögliche Fälle”:

Der zahlenmäßige Anteil von A
an einem endlichen Wertebereich S

wird jetzt ersetzt durch den volumsmäßigen Anteil von A
am (überabzählbar) unendlichen Wertebereich S .

Wie im Diskreten werden wir uns nicht nur mit *rein* zufälliger Wahl begnügen.

Vorlesung 4b

Zufallsvariable mit Dichten

Teil 2

Dichten auf $\mathbb{R}$ bzgl des Längenmaßes

Beginnen wir mit einem **Beispiel**:

X sei eine $\mathbb{R}_+$ -wertige Zufallsvariable mit

$$\mathbf{P}(X > b) = e^{-b} \quad \text{für alle } b \geq 0.$$

$$\mathbf{P}(X \geq b) = ?$$

$$\mathbf{P}(X = b) = ?$$

Für $b = 0$ ist

$$\mathbf{P}(X \geq 0) = 1 = e^{-0} = \mathbf{P}(X > 0).$$

Für $b = 0$ ist

$$\mathbf{P}(X \geq 0) = 1 = e^{-0} = \mathbf{P}(X > 0).$$

Für $b > 0$ und $0 < \varepsilon < b$ ist

$$\begin{aligned} e^{-b} &= \mathbf{P}(X > 0) \leq \mathbf{P}(b \leq X) \\ &\leq \mathbf{P}(b - \varepsilon < X) = e^{-b+\varepsilon} \end{aligned}$$

$$\implies \mathbf{P}(X \geq b) = e^{-b} = \mathbf{P}(X > b).$$

$$\mathbf{P}(X = b) = \mathbf{P}(X \geq b) - \mathbf{P}(X > b) = 0.$$

Für alle $0 \leq b \leq c$ folgt:

$$\begin{aligned} \mathbf{P}(b \leq X \leq c) &= \mathbf{P}(b \leq X) - \mathbf{P}(c < X) \\ &= e^{-b} - e^{-c} = \int_b^c e^{-a} da. \end{aligned}$$

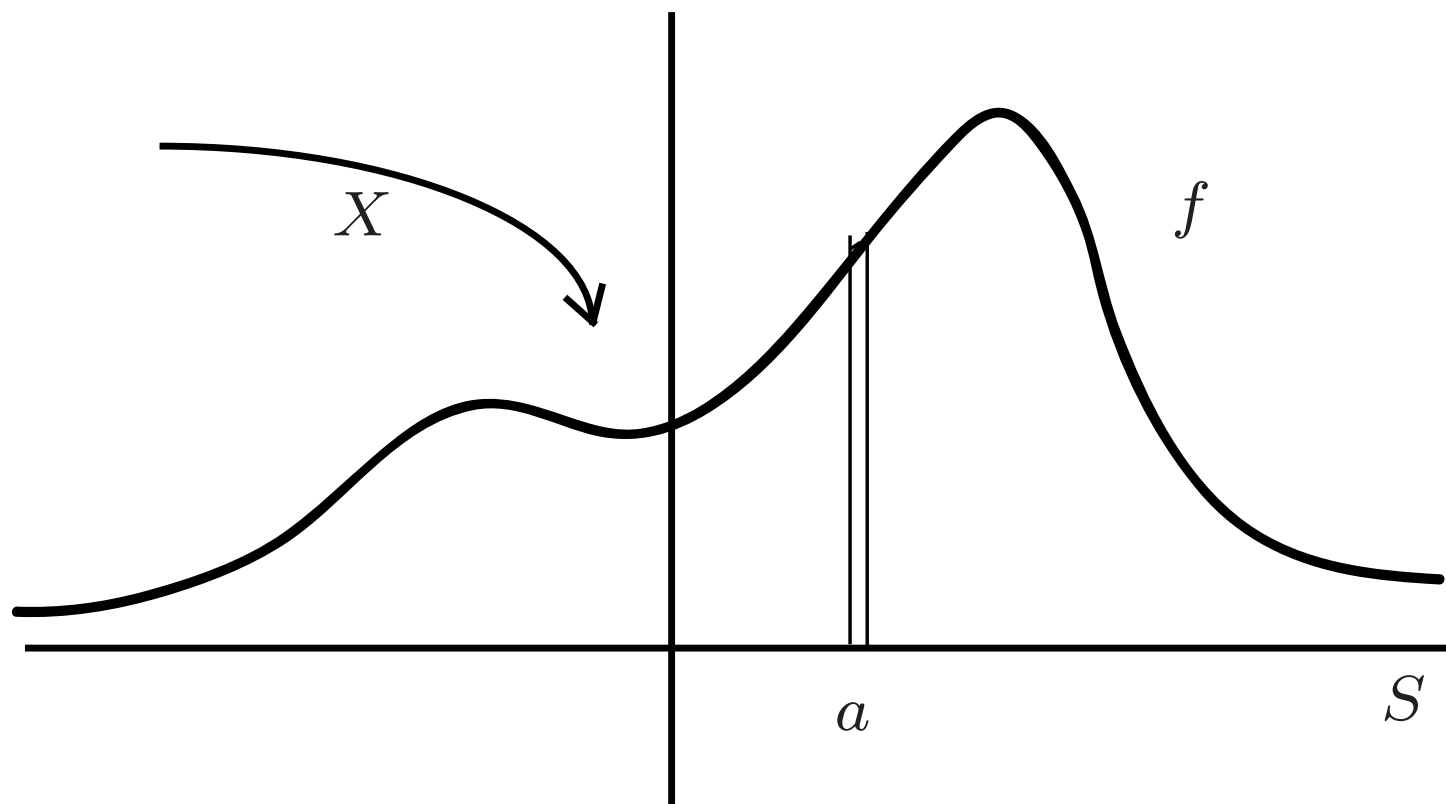
Dieses Beispiel wird im Folgenden verallgemeinert:

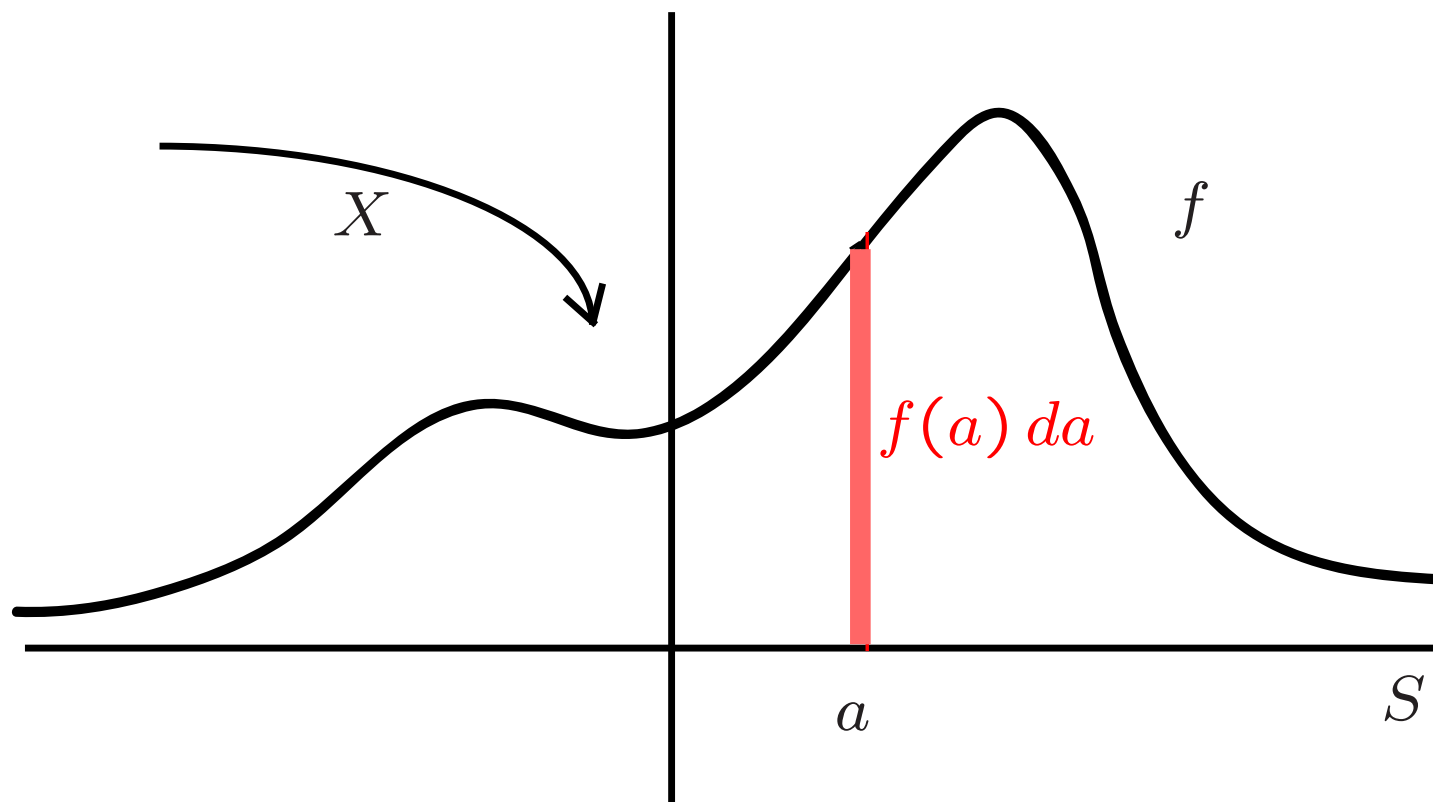
Sei S ein (offenes, abgeschlossenes oder halboffenes)
Intervall mit linker Grenze l und rechter Grenze r ,
und X eine $\overline{\mathbb{R}}$ -wertige Zufallsvariable,
für die $\{X \in S\}$ das sichere Ereignis ist.
(Dabei ist $l = -\infty$ und/oder $r = +\infty$ erlaubt.)

Das Analogon zu den Verteilungsgewichten $\rho(a)$ ist jetzt
gegeben durch “infinitesimale Gewichte” $f(a) da$, wobei

$f : S \rightarrow \mathbb{R}_+$ eine “integrierbare”
(z.B. stückweise stetige) Funktion ist mit

$$\int_l^r f(a) da = 1.$$





Sei X eine Zufallsvariable mit Wertebereich S .

Gilt für alle Intervalle $[b, c] \subset S$ die Gleichung

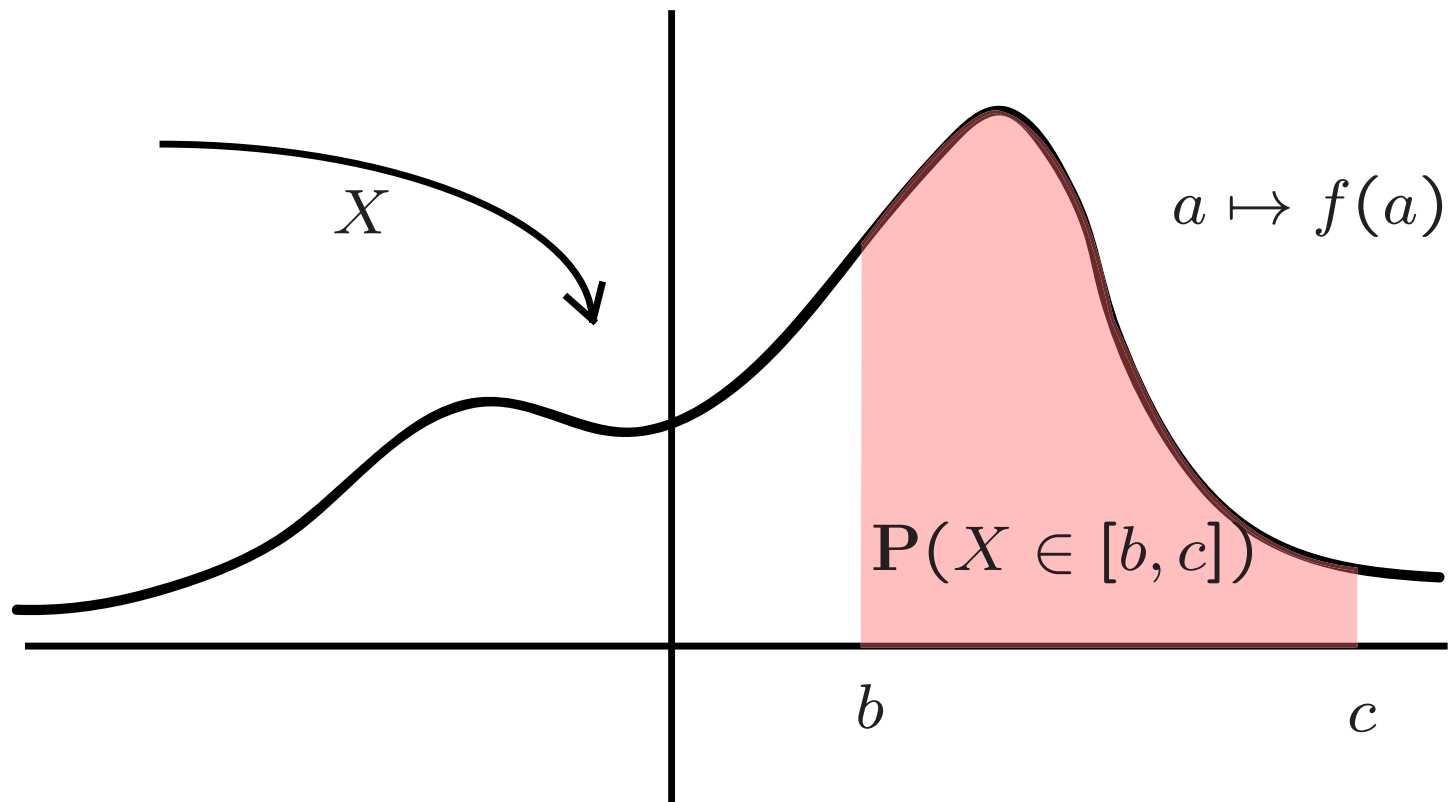
$$\mathbf{P}(X \in [b, c]) = \int_b^c f(a) da ,$$

so sagten wir, dass

X die *Dichte* $f(a) da$ besitzt*,

und nennen f *Dichtefunktion* (der Verteilung) von X .

*Genauer spricht man von der Dichte bzgl des natürlichen Längenmaßes (des Lebesguemaßes) auf $\mathbb{R}$



Wir schreiben dann kurz

$$\mathbf{P}(X \in da) = f(a) da, \quad a \in S,$$

$$\mathbf{P}(X \in da) = f(a) da.$$

Der Ausdruck da taucht hier in zwei Bedeutungen auf:
links als infinitesimales Raumstück da (um den Punkt a)
und rechts als dessen (infinitesimale) Länge da .

$$\mathbf{P}(X \in da) = f(a) da.$$

Der Ausdruck da taucht hier in zwei Bedeutungen auf:
 links als **infinitesimales Raumstück** da (um den Punkt a)
 und rechts als **dessen (infinitesimale) Länge** da .

Diese Gleichung bekommt ihre exakte Bedeutung
 “unter dem Integral”:

$$\mathbf{P}(X \in [c, d]) = \int_c^d f(a) da$$

Beispiele:

Eine auf dem Intervall $[0, 2]$
uniform verteilte Zufallsvariable

hat die Dichte $\frac{1}{2} da, \quad 0 \leq a \leq 2.$

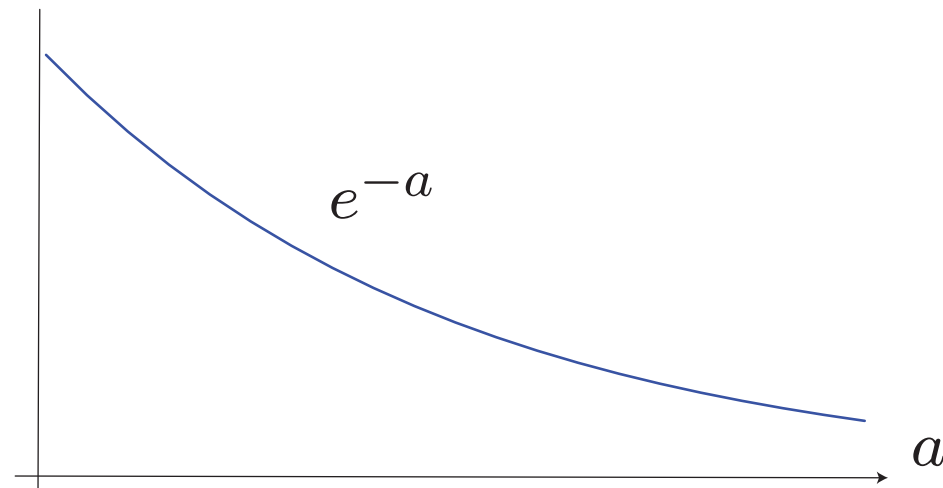
Eine auf einem endlichen Intervall $S = [l, r]$
uniform verteilte Zufallsvariable

hat die Dichte $\frac{1}{r - l} da, \quad a \in S.$

Die Bedingung $\int_S f(a) da = 1$ kann auch erfüllt sein,
wenn S unendlichen Inhalt hat.

Man denke an das Eingangsbeispiel

$$S = [0, \infty); \quad f(a) = e^{-a}, \quad a \geq 0.$$



Merke:

Für eine Zufallsvariable X mit Dichte $f(a) da$

ist für jedes $b \in \mathbb{R}$

$$\mathbf{P}(X = b) = \int_b^b f(a) da = 0.$$

Also gilt (mit naheliegender Schreibweise)

für $b \leq c \in \mathbb{R}$:

$$\int_{(b,c]} f(a) da = \int_{[b,c]} f(a) da = \int_b^c f(a) da.$$

Hat die Zufallsvariable X eine Dichte, so gilt
für jede endliche oder abzählbar unendliche Menge A :

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a) = 0.$$

Insbesondere ist X dann nicht diskret.

Umgekehrt gilt also:

Eine diskrete reellwertige Zufallsvariable besitzt keine Dichte
(bzg. des natürlichen Längenmaßes auf $\mathbb{R}$).

Vorlesung 4b

Zufallsvariable mit Dichten

Teil 3

Verteilungsfunktionen

Wieder sei S ein Intervall in $\mathbb{R}$, und X eine (diskrete oder kontinuierliche) Zufallsvariable mit Wertebereich S .

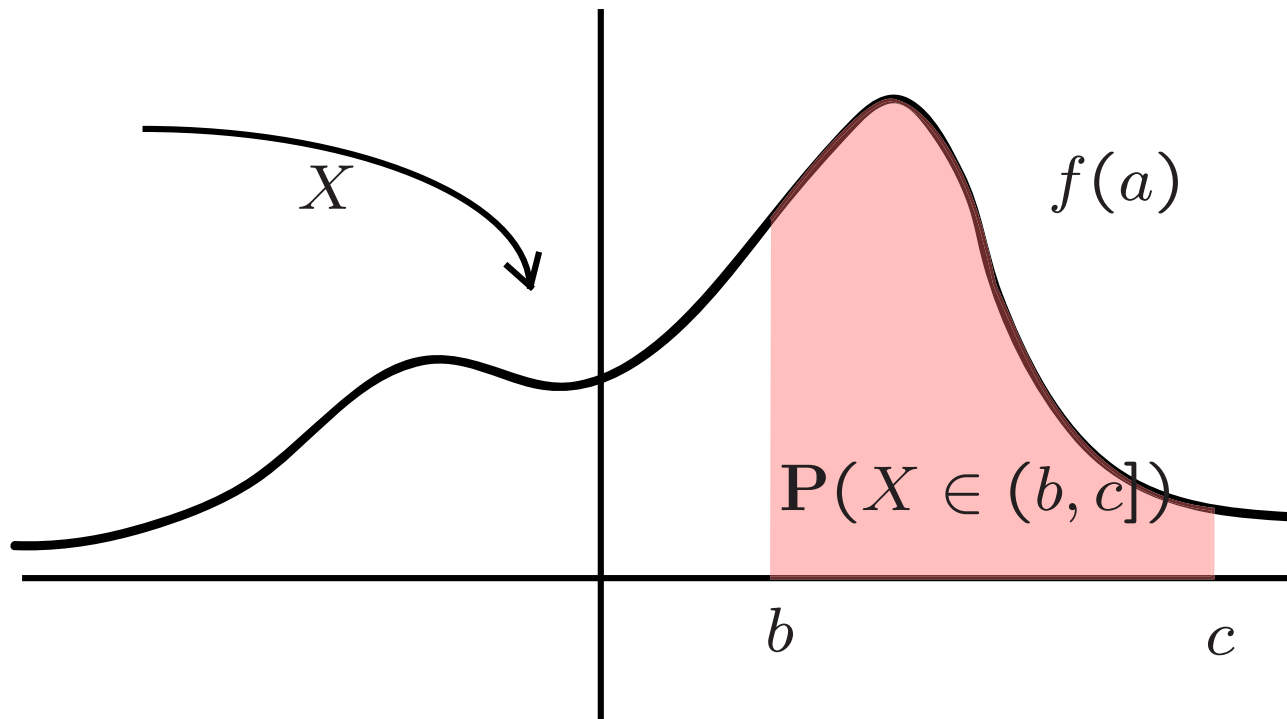
Die Funktion $F(b) := F_X(b) := \mathbf{P}(X \leq b)$, $b \in \mathbb{R}$,
heißt *Verteilungsfunktion* von X .

Hat X die Dichte $f(a) da$, so gilt

$$F(b) = \int_{-\infty}^b f(a) da, \quad b \in \mathbb{R}$$

(mit $f(a) := 0$ für $a \notin S$)

Ist f stetig in a , dann ist $f(a) = F'(a)$.



$$\mathbf{P}(X \leq c) - \mathbf{P}(X \leq b) = \mathbf{P}(b < X \leq c)$$

$$F(c) - F(b) = \int_b^c f(a) da$$

Man findet den Hauptsatz der Differential-und Integralrechnung wieder!

Ein hinreichendes Kriterium für die Existenz einer Dichte.

Sei F die Verteilungsfunktion einer reellwertigen Zufallsvariablen X . Hat F keine Sprünge
*und ist F stückweise stetig differenzierbar**,
dann besitzt X eine Dichte.

Denn für jeden Randpunkt a eines der Intervalle gilt: $P(X = a) = 0$
(ansonsten hätte F in a einen Sprung).

Und innerhalb eines jeden Intervalls gilt nach Voraussetzung der Hauptsatz der Differential-und Integralrechnung, siehe die vorige Folie.

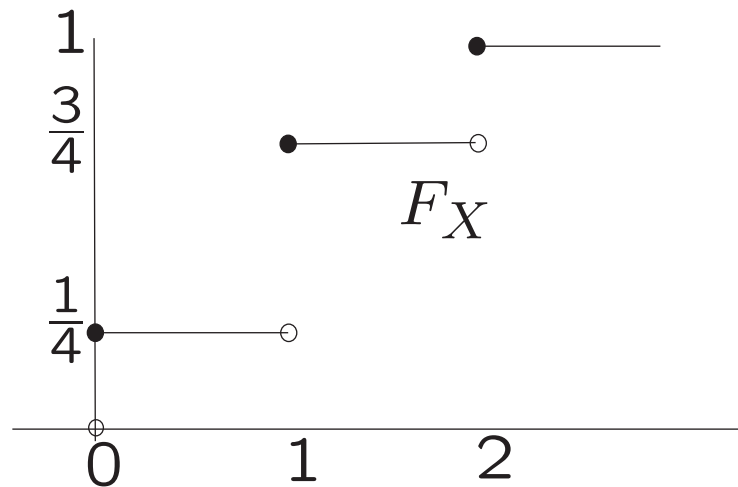
*d.h. es gibt endlich viele disjunkte Intervalle, deren Vereinigung $\mathbb{R}$ ist, so dass F eingeschränkt auf jedes dieser Intervalle eine stetige Ableitung hat.

Für *diskrete* reellwertige Zufallsvariable X
ist F_X stückweise konstant, mit Sprüngen der Höhe

$$\mathbf{P}(X = a), \quad a \in S.$$

Beispiel:

X Binomial(2, 1/2)-verteilt



Vorlesung 5a

Zufallsvariable mit Dichten:

Transformationen, Exponentialverteilung,
Normalverteilung

Teil 1

Transformationen

Drei Beispiele:

A. Sei U uniform verteilt auf $[0, 1]$. Gefragt ist nach Verteilungsfunktion und Dichte von $X := U^2$.

$$F_X(b) = \mathbf{P}(X \leq b) = \mathbf{P}(U \leq \sqrt{b}) = \sqrt{b}, \quad 0 \leq b \leq 1.$$

$$\sqrt{b} \stackrel{!}{=} \int_0^b f(a) da, \quad 0 \leq b \leq 1.$$

$$f(b) = \begin{cases} \frac{1}{2\sqrt{b}}, & 0 < b \leq 1 \\ 0 & \text{sonst} \end{cases} \quad \text{ist Dichtefunktion von } X.$$

B. Sei U uniform verteilt auf $[0, 2]$. Gefragt ist nach Verteilungsfunktion und Dichte von $X := U^2$.

X hat Wertebereich $[0, 4]$.

$$F_X(b) = \mathbf{P}(X \leq b) = \mathbf{P}(U \leq \sqrt{b}) = \frac{1}{2}\sqrt{b}, \quad 0 < b \leq 4.$$

$$\frac{1}{2}\sqrt{b} \stackrel{!}{=} \int_0^b f(a) da, \quad 0 < b \leq 4.$$

$$f(b) = \begin{cases} \frac{1}{4} \frac{1}{\sqrt{b}}, & 0 \leq b \leq 4 \\ 0 & \text{sonst} \end{cases} \quad \text{ist Dichtefunktion von } X.$$

C. Sei U uniform verteilt auf $[0, 1]$.

Gefragt ist nach der Dichte von $X := -\ln U$.

$$\begin{aligned}\mathbf{P}(X \leq b) &= \mathbf{P}(-\ln U \leq b) = \mathbf{P}(\ln U \geq -b) \\ &= \mathbf{P}(U \geq e^{-b}) = \mathbf{P}(U \in [e^{-b}, 1]) \\ &= 1 - e^{-b} \stackrel{!}{=} \int_0^b f(a) da, \quad b \geq 0.\end{aligned}$$

$$f(b) = \begin{cases} e^{-b}, & b \geq 0 \\ 0 & \text{sonst} \end{cases} \quad \text{ist Dichtefunktion von } X.$$

Zufallsvariable X mit der Eigenschaft

$$\mathbf{P}(X > b) = e^{-b}, \quad b \geq 0,$$

sind uns schon (implizit) begegnet

bei der Approximation der Verteilung von pT ;

dabei war T Geom(p)-verteilt mit kleinem p .

Wir sprachen damals von der

Exponentialapproximation der geometrischen Verteilung,

siehe V4a.

Affin lineare Transformation:

X habe Verteilungsfunktion F_X .

Was ist dann die **Verteilungsfunktion von $Y := \beta X + \gamma$** ?

Dabei sei $\beta > 0$.

$$\begin{aligned} & F_Y(b) \\ &= \mathbf{P}(Y \leq b) = \mathbf{P}(\beta X + \gamma \leq b) \\ &= \mathbf{P}(\beta X \leq b - \gamma) = \mathbf{P}\left(X \leq \frac{b - \gamma}{\beta}\right) \\ &= F_X\left(\frac{b - \gamma}{\beta}\right) \end{aligned}$$

X habe Dichte $f_X(a)da$.

Was ist dann die **Dichte von $Y := \beta X + \gamma$** ?

Der Einfachheit halber nehmen wir an: Die Dichtefunktion f_X ist stückweise stetig. Dann ist in allen Stetigkeitspunkten

$$f_Y(b) = F'_Y(b) = F'_X\left(\frac{b - \gamma}{\beta}\right) \frac{1}{\beta}$$

Die Dichte von Y ist somit **$f_Y(b) db = f_X\left(\frac{b - \gamma}{\beta}\right) \frac{db}{\beta}$**

Zum Merken: Schlag nach beim Urbild,
und vergiss den Streckungsfaktor nicht!

Vorlesung 5a

Zufallsvariable mit Dichten:

Transformationen, Exponentialverteilung,
Normalverteilung

Teil 2

Exponentialverteilung

Definition: Die reellwertige Zufallsvariable X heißt
standard-exponentialverteilt, falls

$$\mathbf{P}(X > b) = e^{-b}, \quad b \geq 0.$$

Äquivalent dazu ist:

X ist Zufallsvariable mit Verteilungsfunktion

$$b \mapsto \begin{cases} 0 & \text{für } b < 0 \\ 1 - e^{-b} & \text{für } b \geq 0. \end{cases}$$

Äquivalent dazu ist:

X ist Zufallsvariable mit Dichte $e^{-a} da$, $a \geq 0$.

Ist X standard-exponentialverteilt,
dann hat $\frac{X}{\alpha}$ die Dichte

$$\alpha e^{-\alpha b} db, \quad b \geq 0.$$

Eine Zufallsvariable Y mit dieser Dichte heit
exponentialverteilt mit Parameter α , kurz $\text{Exp}(\alpha)$ -verteilt.

Merke: Ist Y $\text{Exp}(\alpha)$ -verteilt, dann ist αY $\text{Exp}(1)$ -verteilt.

Von der geometrischen zur Exponentialverteilung

Betrachten wir eine $\text{Geom}(p)$ -verteilte Zufallsvariable Y mit

$$p = \frac{1}{100}.$$

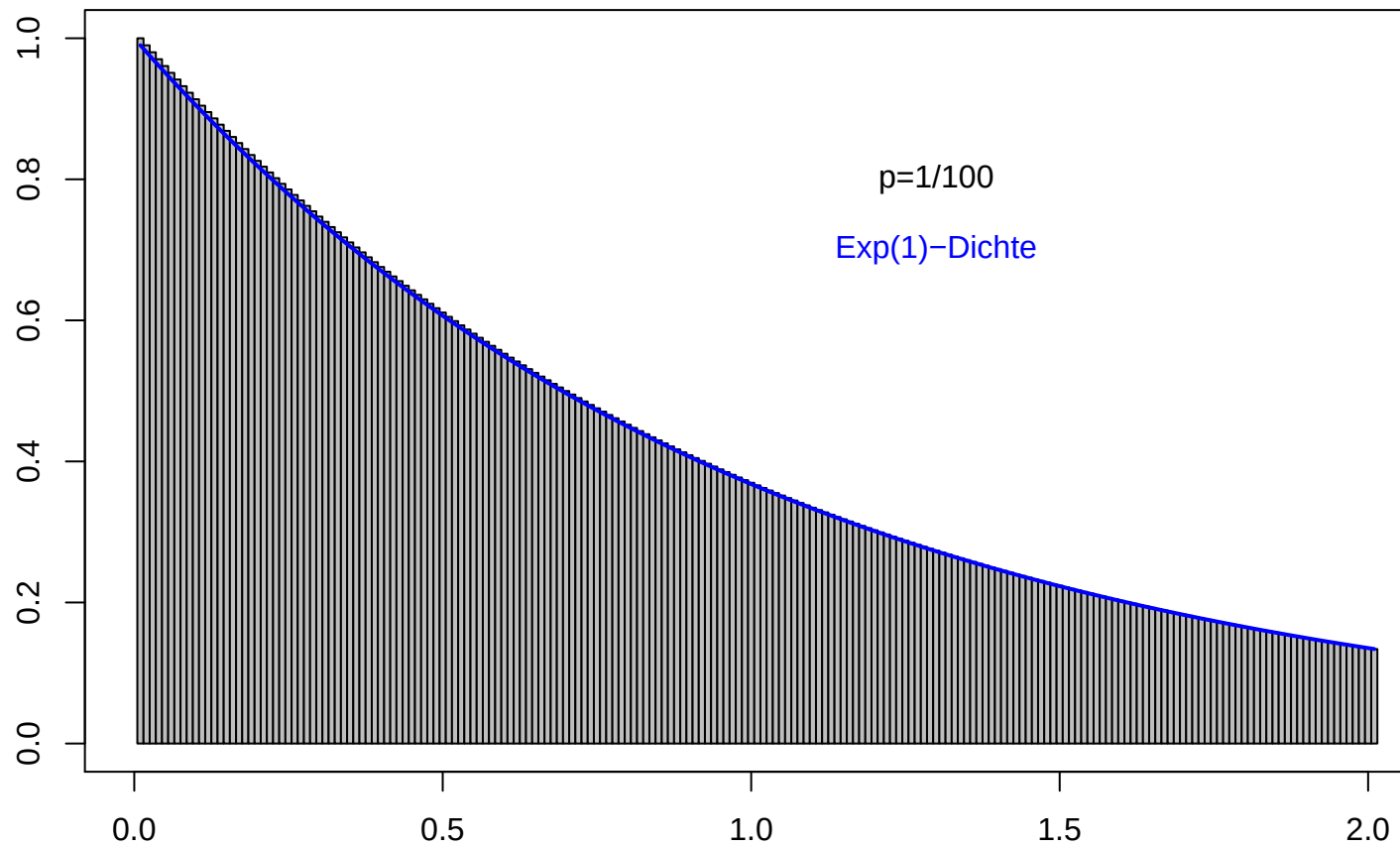
Es gilt: $\mathbf{E}[Y] = 100$, also hat $\frac{Y}{100}$ Erwartungswert 1.

$$\begin{aligned} \mathbf{P}\left(\frac{Y}{100} = \frac{k}{100}\right) &= \mathbf{P}(Y = k) \\ &= \frac{1}{100} \left(1 - \frac{1}{100}\right)^{k-1} = \frac{1}{100} \left(\left(1 - \frac{1}{100}\right)^{100}\right)^{(k-1)/100} \\ &\approx \frac{1}{100} e^{-(k-1)/100} \approx \mathbf{P}\left(X \in \left[\frac{k-1}{100}, \frac{k}{100}\right]\right) \end{aligned}$$

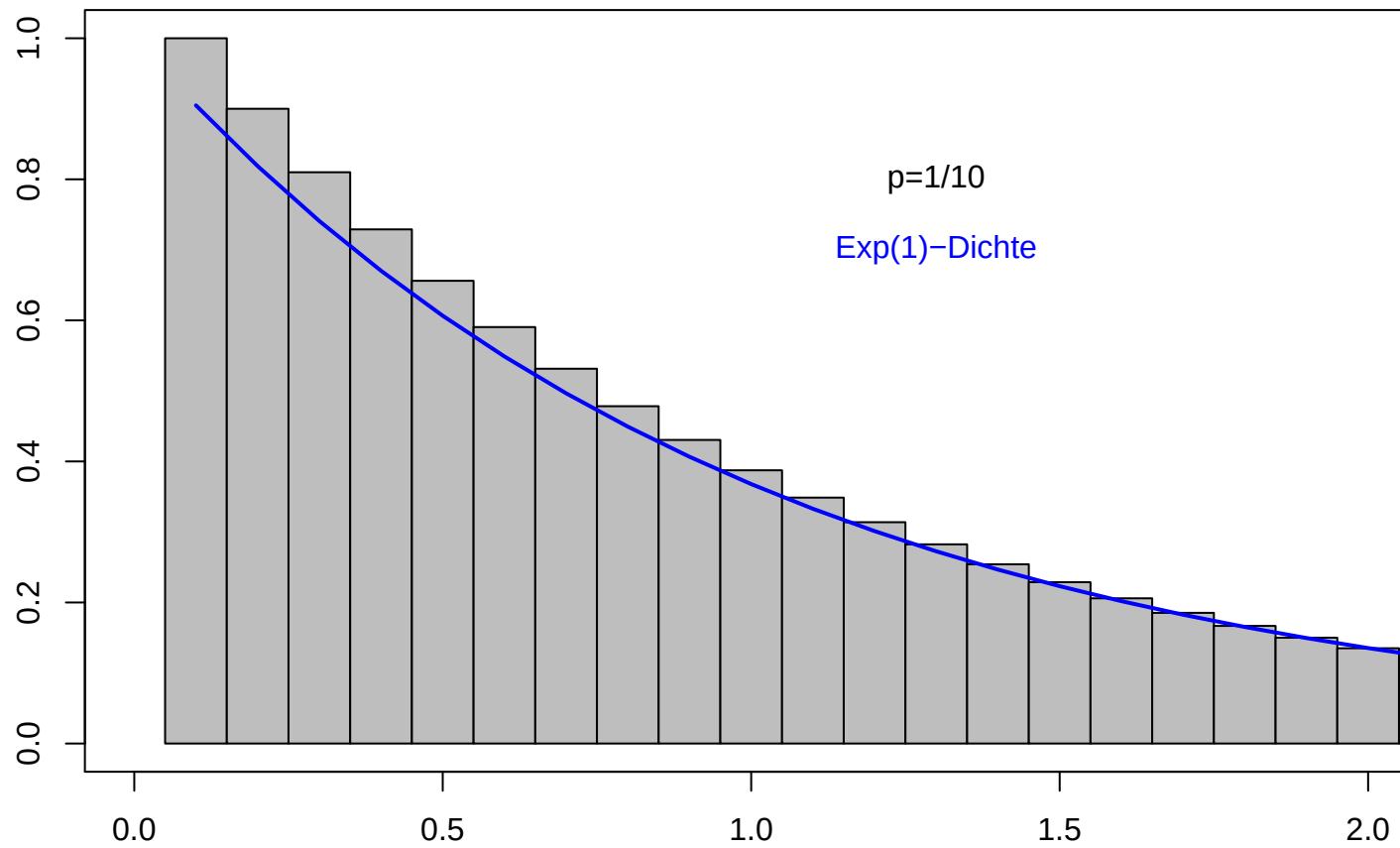
für ein standard-exponentialverteiltes X .

Das wird durch das folgende Bild veranschaulicht:

Gewichte des p -fachen einer Geom(p)-verteilten ZV



Gewichte des p -fachen einer Geom(p)-verteilten ZV



Salopp gesprochen:

Für kleines p hat eine $\text{Geom}(p)$ -verteilte Zufallsvariable Y
einen großen Erwartungswert, nämlich $\frac{1}{p}$.

Man holt die Verteilung von Y zurück ins Schaubild,
indem man pY betrachtet.

Wir erinnern an die “Exponentialapproximation”:
(Vorlesung 4a und Buch Seite 42):

Ist für jedes $n \in \mathbb{N}$ die Zufallsvariable X_n geometrisch verteilt
mit $\mathbf{E}[X_n] \rightarrow \infty$ für $n \rightarrow \infty$, dann gilt:

$$\mathbf{P}\left(\frac{X_n}{\mathbf{E}[X_n]} > t\right) \rightarrow e^{-t} \quad \text{für alle } t \in \mathbb{R}_+.$$

$$\mathbf{P}\left(\frac{X_n}{\mathbf{E}[X_n]} > t\right) \rightarrow e^{-t} \quad \text{für alle } t \in \mathbb{R}_+.$$

Also:

$$\mathbf{P}\left(\frac{X_n}{\mathbf{E}[X_n]} > t\right) \rightarrow \mathbf{P}(X > t), \quad t \in \mathbb{R}_+,$$

wobei X eine standard-exponentialverteilte Zufallsvariable ist.

Man sagt dafür auch:

Die Folge der Zufallsvariablen $X_n/\mathbf{E}[X_n]$

konvergiert in Verteilung

gegen die Zufallsvariable X .

Vorlesung 5a

Zufallsvariable mit Dichten:

Transformationen, Exponentialverteilung,
Normalverteilung

Teil 3

Erwartungswert und Transformationsformel

Für diskrete reellwertige Zufallsvariable hatten wir

$$\mu = \mathbf{E}[X] = \sum_{a \in S} a \rho(a)$$

Das hat sein Analogon im Fall reellwertiger ZV'er mit Dichten:

Den Verteilungsgewichten $\rho(a)$ entspricht die Dichte $f(a) da$.

Und aus der Summe wird ein Integral:

$$\mu = \mathbf{E}[X] = \int_l^r a f(a) da$$

Im Diskreten hatten wir für $h : S \rightarrow \mathbb{R}$
die Transformationsformel

$$\mathbf{E}[h(X)] = \sum_{a \in S} h(a) \rho(a).$$

Analog gilt im Fall mit Dichten:

$$\mathbf{E}[h(X)] = \int_l^r h(a) f(a) da$$

**Der Erwartungswert einer
standard-exponentialverteilten Zufallsvariablen X :**

$$\mathbf{E}[X] = \int_0^{\infty} x e^{-x} dx$$

Mit partieller Integration

$$\int u v' = uv - \int u'v$$

ergibt sich

$$\int_0^{\infty} x e^{-x} dx = -xe^{-x} \Big|_0^{\infty} + \int_0^{\infty} 1 \cdot e^{-x} dx = 1$$

Also:

$$\boxed{\mathbf{E}[X] = 1.}$$

**Der Erwartungswert des Quadrates einer
standard-exponentialverteilten Zufallsvariablen X :**

$$\mathbf{E} [X^2] = \int_0^{\infty} x^2 e^{-x} dx$$

Wieder mit partieller Integration:

$$\int_0^{\infty} x^2 e^{-x} dx = -x^2 e^{-x} \Big|_0^{\infty} + \int_0^{\infty} 2x \cdot e^{-x} dx = 2$$

Also:

$$\boxed{\mathbf{E} [X^2] = 2.}$$

Der Erwartungswert einer $\text{Exp}(\alpha)$ -verteilten Zufallsvariablen Y :

Wir wissen schon: Ist Y $\text{Exp}(\alpha)$ -verteilt,
dann ist αY $\text{Exp}(1)$ -verteilt.

Also ist

$$\mathbf{E}[Y] = \frac{1}{\alpha} \mathbf{E}[\alpha Y] = \frac{1}{\alpha} \cdot 1.$$

$$\boxed{\mathbf{E}[Y] = \frac{1}{\alpha}.}$$

Vorlesung 5a

Zufallsvariable mit Dichten:

Transformationen, Exponentialverteilung,
Normalverteilung

Teil 4

Die Standardnormalverteilung auf $\mathbb{R}$

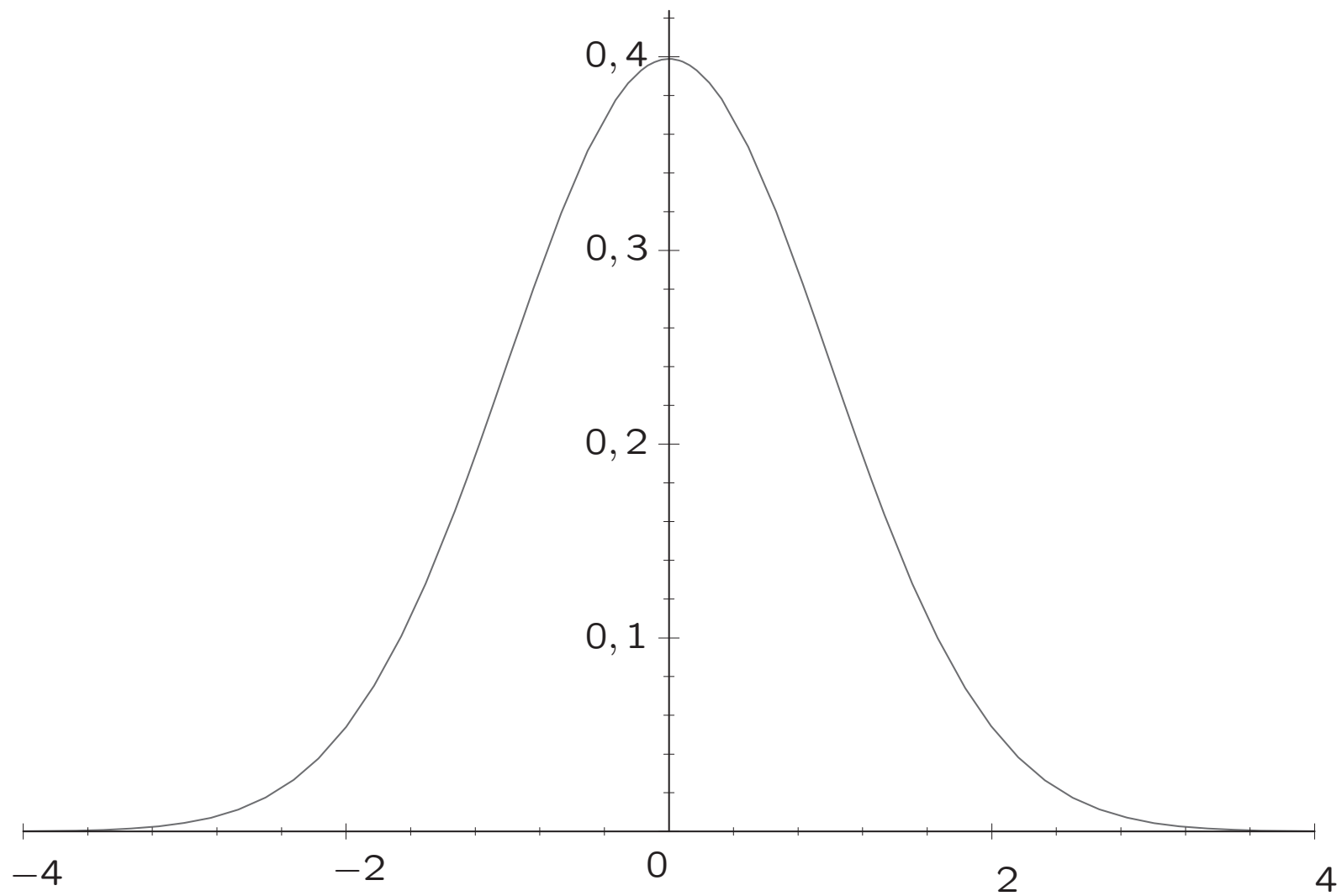
Die folgende Definition beinhaltet
die wichtigste Verteilung der Stochastik:
Eine $\mathbb{R}$ -wertige Zufallsvariable Z mit Dichte

$$\varphi(a) da := \frac{1}{\sqrt{2\pi}} e^{-a^2/2} da$$

heißt **standard-normalverteilt**.

Wir werden bald auf geometrischem Wege sehen, dass gilt
(vgl. die erste Aufgabe auf S. 72 im Buch):

$$\int_{-\infty}^{+\infty} e^{-a^2/2} da = \sqrt{2\pi}, \quad \text{also} \quad \int_{-\infty}^{+\infty} \varphi(a) da = 1.$$



Für ein standard-normalverteiltes Z ist

$$\mathbf{E}[Z] = 0, \quad \mathbf{E}[Z^2] = 1.$$

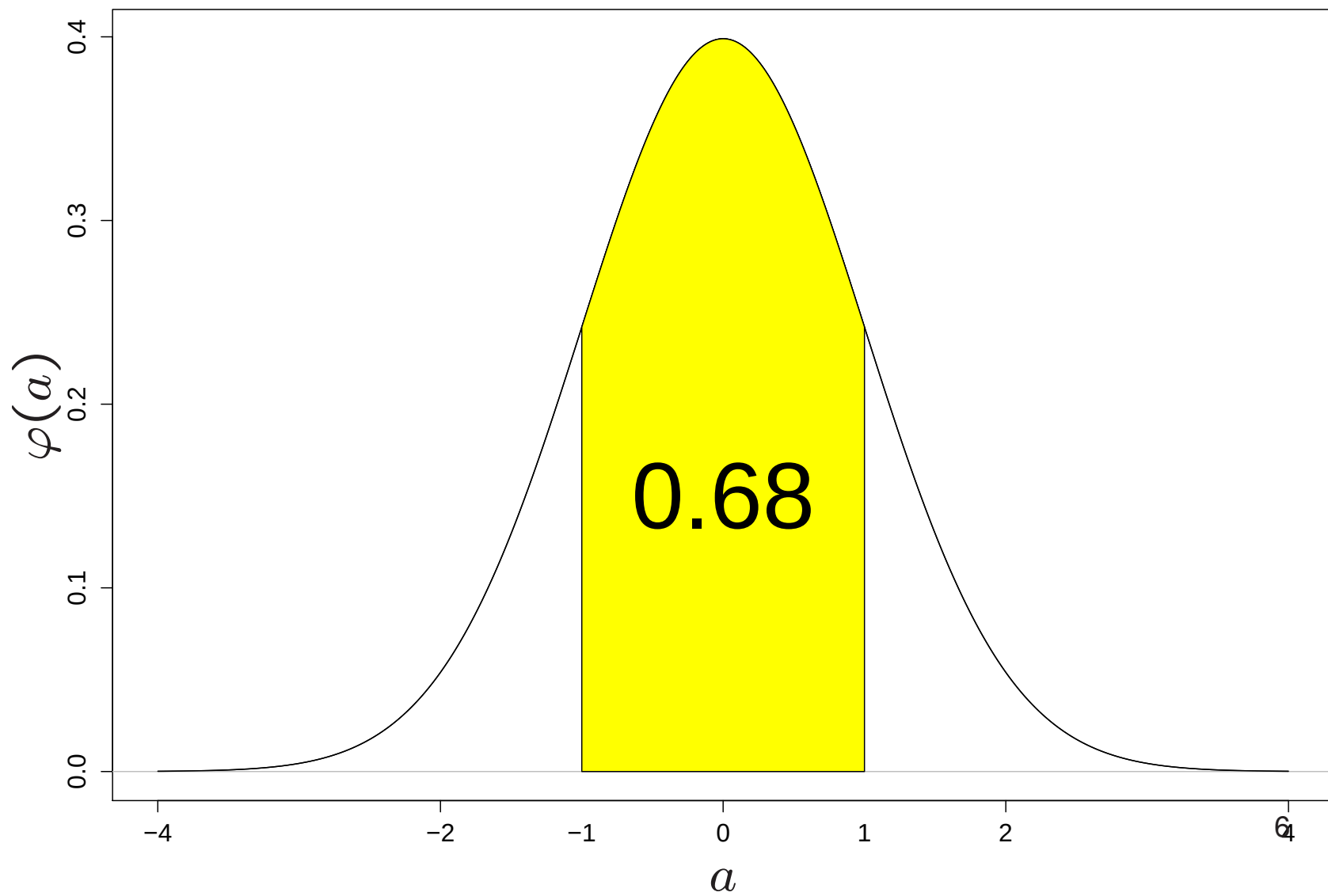
Denn aus Symmetriegründen ist $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a e^{-a^2/2} da = 0$,

und mit partieller Integration bekommt man

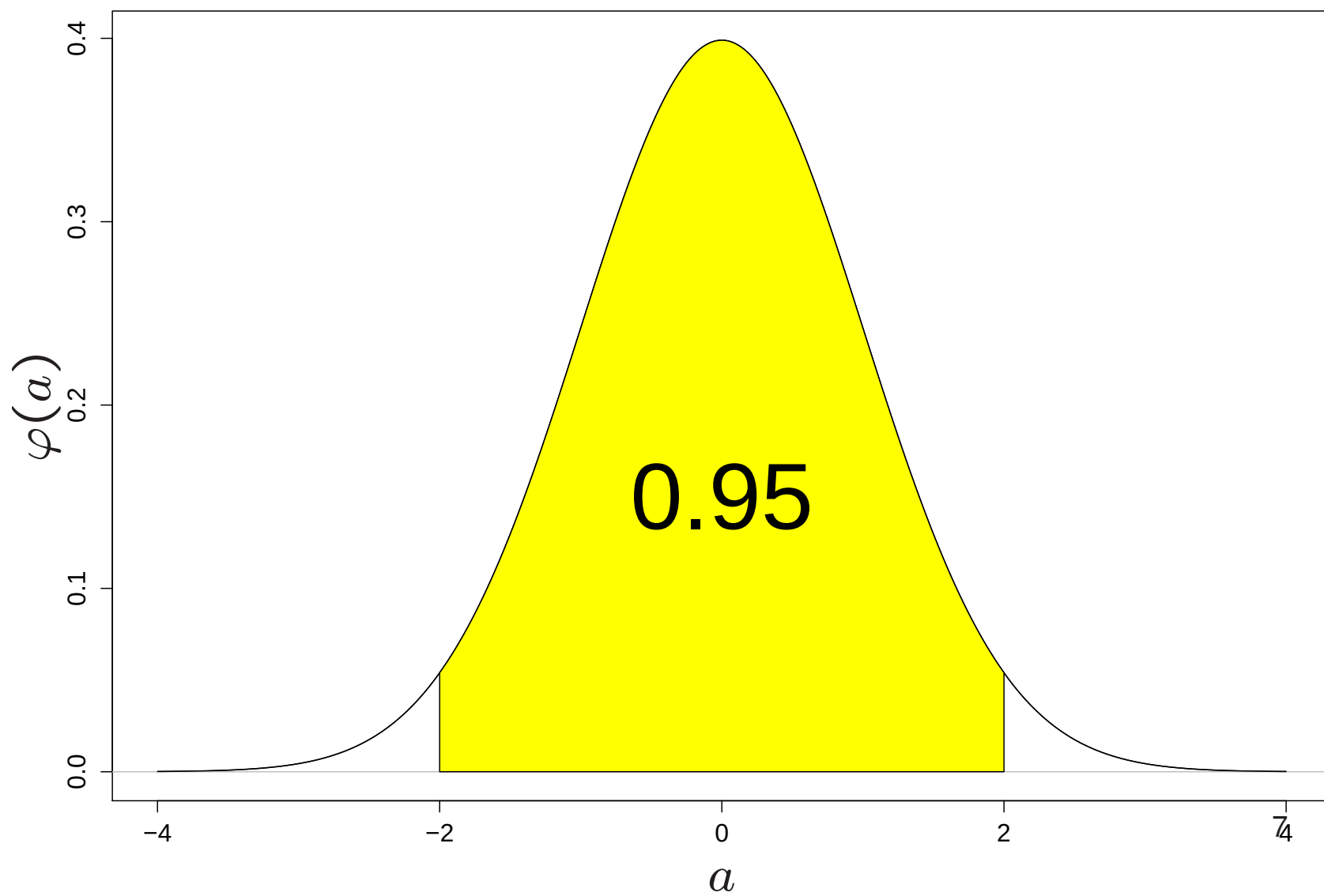
$$\begin{aligned} \int_{-\infty}^{\infty} a^2 e^{-a^2/2} da &= \int_{-\infty}^{\infty} a a e^{-a^2/2} da \\ &= -a e^{-a^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-a^2/2} da = 0 + \sqrt{2\pi}. \end{aligned}$$

Zwei für die Praxis wichtige Zahlen:

$$\mathbf{P}(|Z| < 1) \approx 0.68$$



$$P(|Z| < 2) \approx 0.95$$



$$\Phi(b) := \int_{-\infty}^b \varphi(a) da$$

ist die Standard-Normalverteilungsfunktion.

Es gibt für sie keinen expliziten analytischen Ausdruck
(der ohne die Formulierung als Integral bzw. Stammfunktion auskommt).

Der R-Befehl dafür ist `pnorm(b)` .

Affin lineare Transformation einer standard-normalverteilten ZV'e

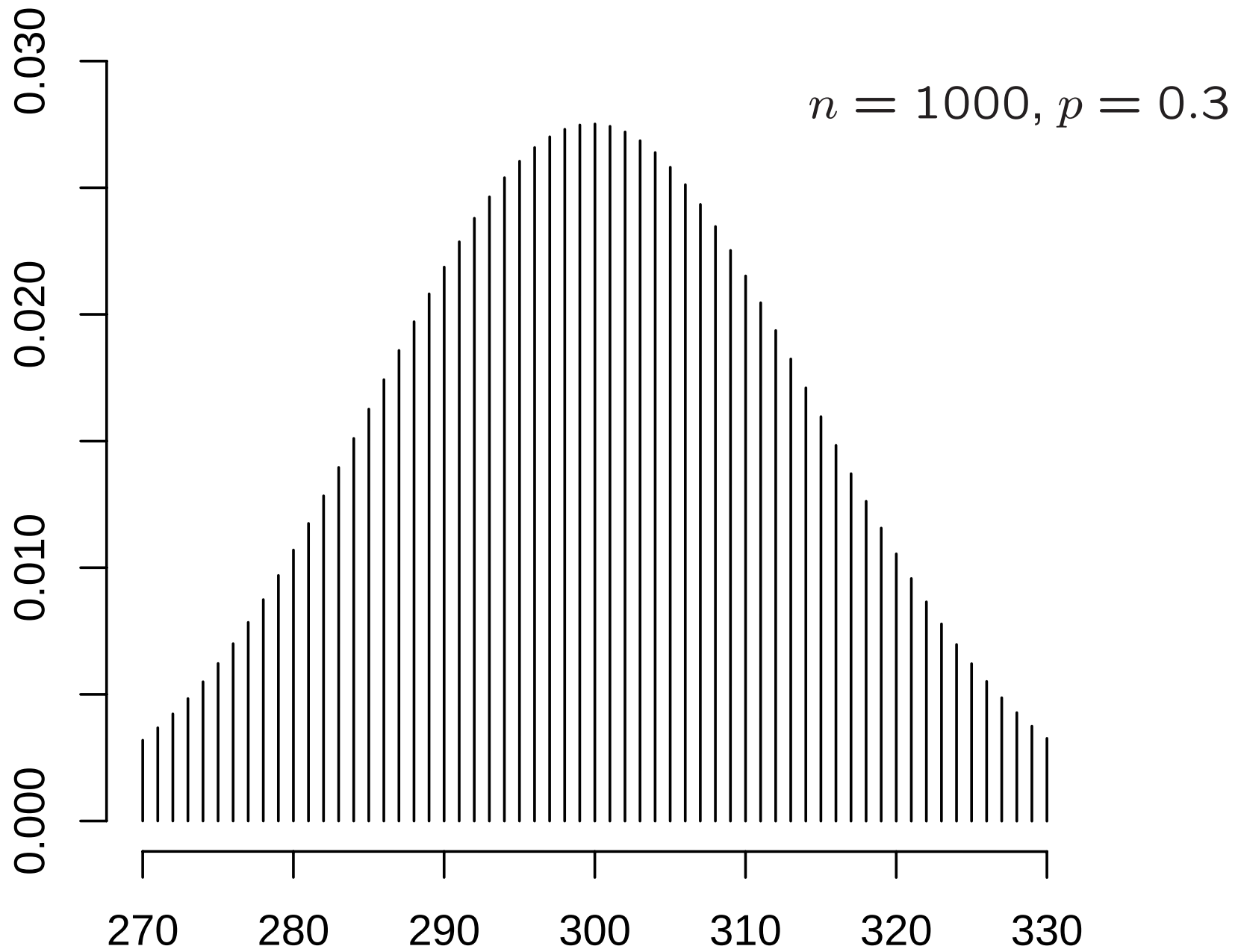
Sei Z standard-normalverteilt, $\mu \in \mathbb{R}$, $\sigma > 0$. Dann hat

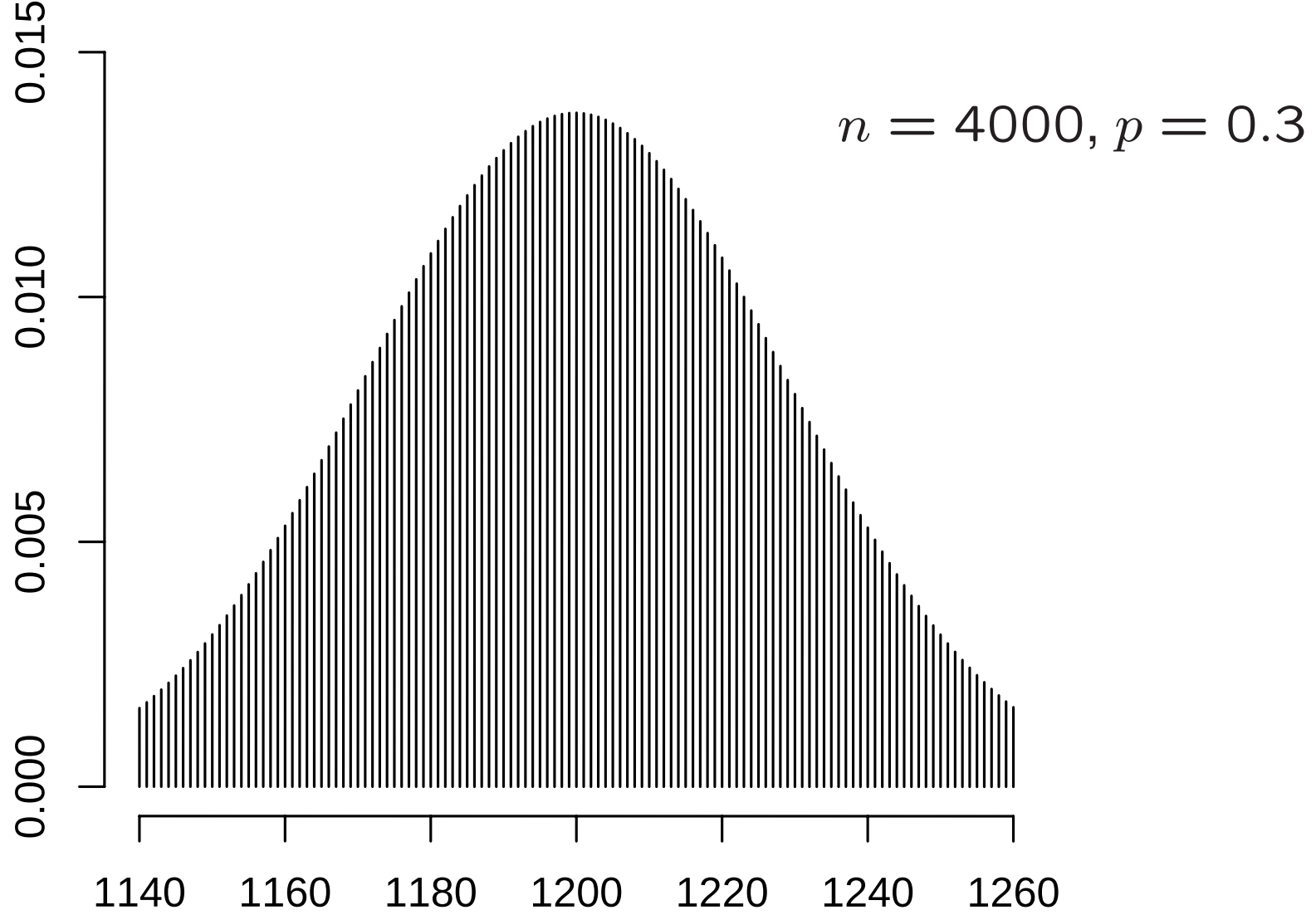
$$X := \sigma Z + \mu :$$

die Dichte

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

Die Gaußsche Glockenkurve sieht man auch entstehen aus den Binomialgewichten mit großem n und großem npq , wenn man sie “geeignet ins Bild bringt”:





Mehr dazu bald in der Vorlesung!

Vorlesung 5b

Unabhängigkeit

Teil 1

Zwei (diskrete) Zufallsvariable

(Buch S. 61)

Zwei binäre Zufallsvariable: Ausgeglichene Verhältnisse?

Der Wertebereich von $X = (X_1, X_2)$ sei $\{o, u\} \times \{\ell, r\}$

	ℓ	r
o		
u		

Wie sollten die Verteilungsgewichte der 4 Ausgänge
 $\{(o, \ell), (o, r), (u, \ell), (u, r)\}$ aussehen,
damit es gerechtfertigt ist zu sagen:
 X_1 und X_2 sind unabhängig?

Beispiel:

	ℓ	r
o	$2/6$	$1/6$
u	$1/6$	$2/6$

Das Chancenverhältnis von ℓ zu r ,
d.h. das Verhältnis $\mathbf{P}(X_2 = \ell) : \mathbf{P}(X_2 = r)$,
ist $(\frac{2}{6} + \frac{1}{6}) : (\frac{1}{6} + \frac{2}{6}) = 1 : 1$.

So wird man bezüglich X_2 wetten,
wenn man nichts über den Ausgang von X_1 weiß.

Beispiel:

	ℓ	r
o	$2/6$	$1/6$
u	$1/6$	$2/6$

Wenn man aber weiß, dass das Ereignis $\{X_1 = o\}$ eintritt,
wird man bezüglich des Ausgangs von X_2
anders wetten als 1 : 1.

Das Chancenverhältnis von ℓ zu r ist dann 2 : 1.

Die Verhältnisse in den Zeilen sind nicht ausgeglichen!

Beispiel: Ausgeglichene Verhältnisse

Der Wertebereich von $X = (X_1, X_2)$ sei $\{o, u\} \times \{\ell, r\}$

	ℓ	r
o	$2/9$	$4/9$
u	$1/9$	$2/9$

Hier sind die Chancenverhältnisse von ℓ zu r
in allen Zeilen gleich -
unabhängig vom Ausgang von X_1 .

Definition

Zufallsvariable X_1, X_2 heißen (stochastisch) *unabhängig*,
wenn für alle Ereignisse $\{X_1 \in A_1\}, \{X_2 \in A_2\}$ gilt:

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2)$$

(“Produktformel für Wahrscheinlichkeiten”)

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2)$$

Zum Merken :

Für jedes Paar von Teilmengen $A_2, A'_2 \subset S_2$ gilt:

Die Verhältnisse

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2) : \mathbf{P}(X_1 \in A_1, X_2 \in A'_2)$$

hängen nicht ab von $A_1 (\subset S_1)$

(und sind in diesem Sinn “ausgeglichen”).

Beispiel: Zweimaliges (gewöhnliches) Würfeln (X_1, X_2) :

X_1 und X_2 sind unabhängig. In der Tat:

Seien $A_1, A_2 \subset \{1, \dots, 6\}$ mit $\#A_1 =: m_1$, $\#A_2 =: m_2$.

Dann ist $\#A_1 \times A_2 = m_1 \cdot m_2$,

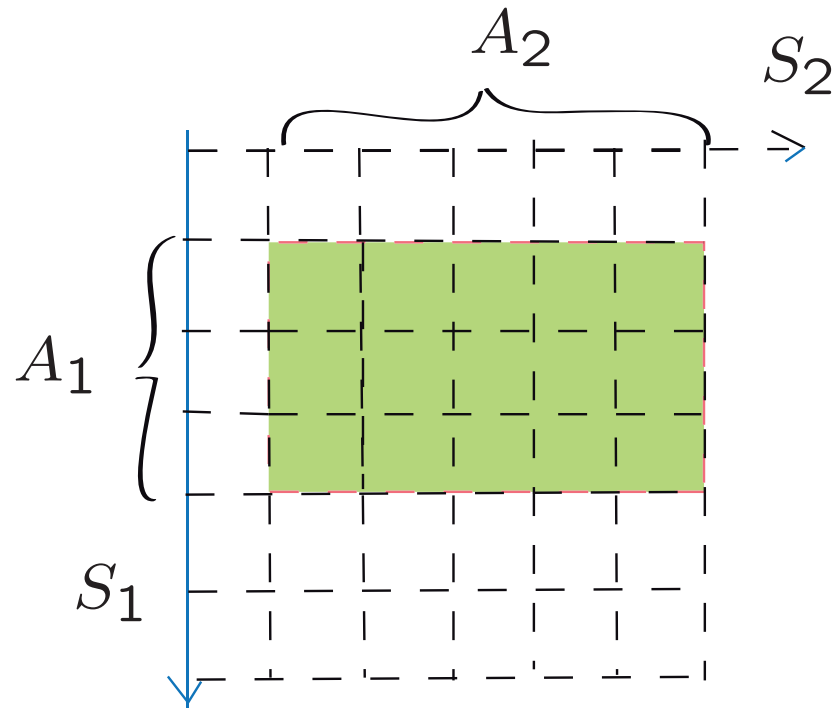
also

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2)$$

$$= \mathbf{P}((X_1, X_2) \in A_1 \times A_2)$$

$$= \frac{m_1 m_2}{36} = \frac{m_1}{6} \cdot \frac{m_2}{6}$$

$$= \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2).$$



$$\begin{aligned}
 & \mathbf{P}(X_1 \in A_1, X_2 \in A_2) \\
 &= \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2).
 \end{aligned}$$

Für die Prüfung der Unabhängigkeit *diskreter* Zufallsvariabler
reicht es fürs Nachprüfen der Produktformel,
statt der *Teilmengen* $A_1 \subset S_1$, $A_2 \subset S_2$
die einzelnen *Ausgänge* $a_1 \in S_1$, $a_2 \in S_2$ zu betrachten:

Für diskrete Zufallsvariable gilt: X_1, X_2 sind unabhängig
genau dann,

wenn für alle $a_1 \in S_1$ und $a_2 \in S_2$ gilt:

$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1) \mathbf{P}(X_2 = a_2)$$

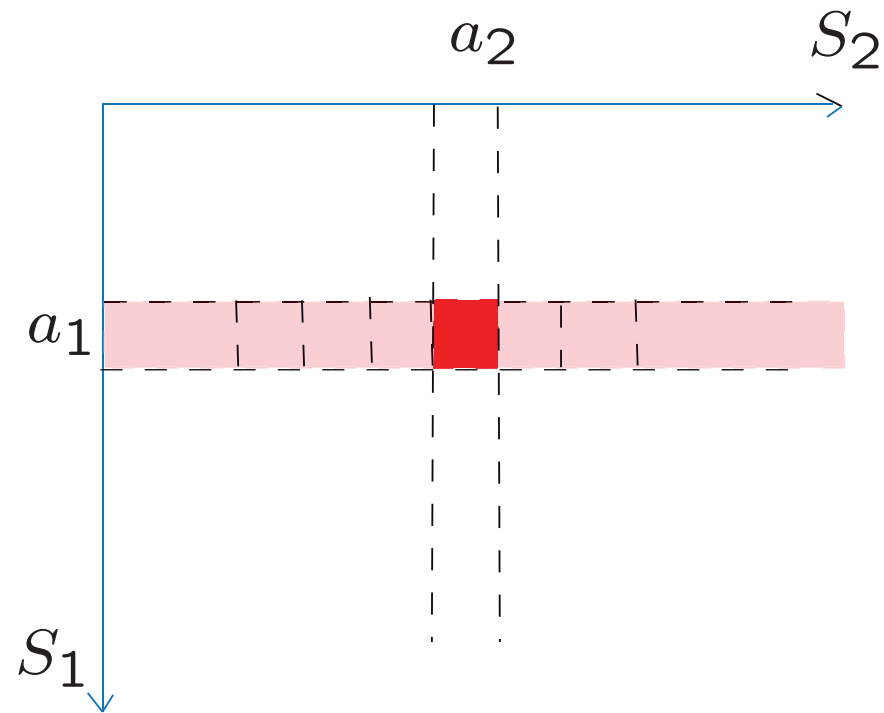
Denn:

“ $\implies$ ” ist klar (wähle $A_1 := \{a_1\}$, $A_2 := \{a_2\}$).

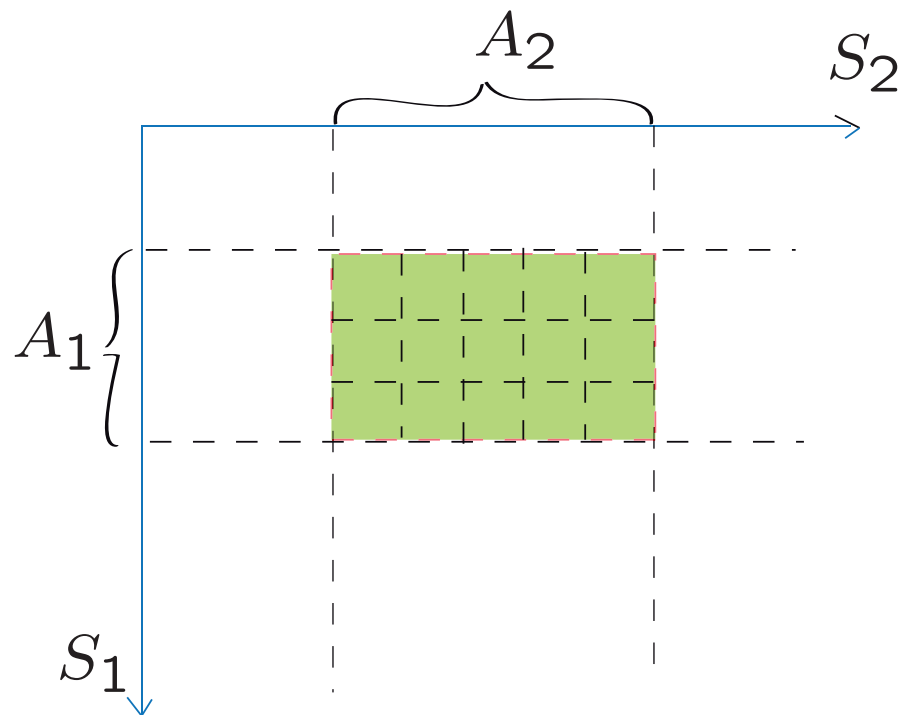
$$\text{“}\Leftarrow\text{”}: \mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \sum_{a_1 \in A_1, a_2 \in A_2} \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1 \in A_1, a_2 \in A_2} \mathbf{P}(X_1 = a_1) \mathbf{P}(X_2 = a_2)$$

$$= \sum_{a_1 \in A_1} \mathbf{P}(X_1 = a_1) \sum_{a_2 \in A_2} \mathbf{P}(X_2 = a_2) = \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2).$$



$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1)\mathbf{P}(X_2 = a_2)$$



$$\begin{aligned} \mathbf{P}(X_1 \in A_1, X_2 \in A_2) \\ = \mathbf{P}(X_1 \in A_1)\mathbf{P}(X_2 \in A_2). \end{aligned}$$

Ein handliches Kriterium für die Unabhängigkeit zweier diskreter Zufallsvariabler ist die Proportionalität der Zeilen* der Matrix ihrer gemeinsamen Verteilungsgewichte $\rho(a_1, a_2)$.

In der Tat ist das hinreichend: dann ist nämlich

$\rho(a_1, \cdot)$ (das ist die Zeile namens a_1) ja auch proportional zu $\rho_2(\cdot) := \sum_{a'_1 \in S_1} \rho(a'_1, \cdot)$ (das ist die Summe der Zeilen), also gilt

$$(*) \quad \rho(a_1, a_2) = k(a_1)\rho_2(a_2)$$

für irgendwelche nichtnegativen Zahlen $k(a_1)$.

Summiert man (*) über $a_2 \in S_2$, so erhält man $\rho_1(a_1) = k(a_1)$.

Damit wird (*) zur Produktformel von Folie 11.

Umgekehrt impliziert die Produktformel von Folie 11 die Proportionalität der Zeilen* der Matrix $(\rho(a_1, a_2))_{a_1 \in S_1, a_2 \in S_2}$ (Übung).

*oder auch der Spalten

Vorlesung 5b

Unabhängigkeit

Teil 2

Produktformel für Erwartungswerte

(Buch S. 61)

Sind X_1, X_2 unabhängig mit Werten in S_1 bzw. S_2 ,
dann gilt für Mengen $A_1 \subset S_1, A_2 \subset S_2$:

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \mathbf{P}(X_1 \in A_1) \mathbf{P}(X_2 \in A_2)$$

Anders geschrieben:

$$\mathbf{E}[1_{A_1}(X_1) \cdot 1_{A_2}(X_2)] = \mathbf{E}[1_{A_1}(X_1)] \mathbf{E}[1_{A_2}(X_2)]$$

In Worten:

Der Erwartungswert des Produktes von $1_{A_i}(X_i)$, $i = 1, 2$
ist das Produkt der Erwartungswerte.

Wir werden gleich sehen, dass allgemeiner gilt:

Sind X_1, X_2 unabhängig,
und h_1, h_2 reellwertige “Verarbeitungen”, dann gilt:

Der Erwartungswert des Produktes $h_1(X_1) \cdot h_2(X_2)$
ist gleich dem Produkt der Erwartungswerte.

Genauer:

Satz:

X_1, X_2 unabhängige ZV'e mit Zielbereichen S_1, S_2 ,
 h_1, h_2 Abbildungen von S_1 bzw. S_2 in die reellen Zahlen.

Haben $h_1(X_1)$ und $h_2(X_2)$ endlichen Erwartungswert,
so folgt

$$\mathbf{E}[h_1(X_1)h_2(X_2)] = \mathbf{E}[h_1(X_1)] \mathbf{E}[h_2(X_2)] .$$

(“Produktformel für Erwartungswerte”)

Beweis (hier nur für diskrete ZV'e):

$$\mathbf{E}[h_1(X_1)h_2(X_2)]$$

$$= \sum_{a_1, a_2} h_1(a_1)h_2(a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1, a_2} h_1(a_1) \mathbf{P}(X_1 = a_1) h_2(a_2) \mathbf{P}(X_2 = a_2)$$

$$= \sum_{a_1} h_1(a_1) \mathbf{P}(X_1 = a_1) \sum_{a_2} h_2(a_2) \mathbf{P}(X_2 = a_2)$$

$$= \mathbf{E}[h_1(X_1)] \mathbf{E}[h_2(X_2)] \quad \square$$

Vorlesung 5b

Unabhängigkeit

Teil 3

Mehrere Zufallsvariable

(Buch S. 64-66)

Zufallsvariable $X_1, \dots, X_n$ mit Zielbereichen $S_1, \dots, S_n$
heißen

(stochastisch) *unabhängig*, falls für alle Ereignisse $\{X_i \in A_i\}$
folgende Produktformel gilt:

$$\mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbf{P}(X_1 \in A_1) \cdots \mathbf{P}(X_n \in A_n) .$$

Unabhängigkeit von abzählbar unendlich vielen Zufallsvariablen:

Sei $X_1, X_2, \dots$ eine Folge von Zufallsvariablen.

Definition:

Die Zufallsvariablen $X_1, X_2, \dots$ sind unabhängig
: $\iff$ für jedes n sind $X_1, \dots, X_n$ unabhängig.

Beispiele:

Fortgesetzter Münzwurf, fortgesetztes Würfeln

Für diskrete Zufallsvariable $X_1, \dots, X_n$
ist die **Unabhängigkeit** gleichbedeutend mit der
Produktform der Verteilungsgewichte:

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = \rho_1(a_1) \cdots \rho_n(a_n)$$

Die $\rho_i(a_i)$ sind dann die Verteilungsgewichte von X_i .

Vorlesung 5b

Unabhängigkeit

Teil 4

Unabhängigkeit von Ereignissen

(Buch S. 67-68)

Unabhängigkeit von Ereignissen

(Buch S. 67)

Ereignisse $E_1, \dots, E_n$ heißen **unabhängig**
: $\Longleftrightarrow I_{E_1}, \dots, I_{E_n}$ sind **unabhängig**.

Satz:

Für die Unabhängigkeit von $E_1, \dots, E_n$ reicht aus, dass

$$\mathbf{P}(E_{i_1} \cap \dots \cap E_{i_k}) = \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k})$$

für alle $1 \leq i_1 < \dots < i_k \leq n$.

Einen eleganten Beweis führt man über eine Rechnung
mit Indikatorvariablen (ähnlich wie bei der
Einschluss-Ausschlussformel), siehe Buch Seite 67.

Der Beweis des Satzes wird in der Vorlesung nicht geführt.

Korollar zum vorigen Satz:

Die Unabhängigkeit zweier Ereignisse E_1, E_2
ist äquivalent zur Produktformel

$$\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1) \mathbf{P}(E_2)$$

Und die Unabhängigkeit dreier Ereignisse E_1, E_2, E_3 ist äquivalent dazu, dass *beide* der folgenden Bedingungen a) und b) erfüllt sind:

$$\begin{aligned} \text{a)} \quad & \mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1)\mathbf{P}(E_2), \\ & \mathbf{P}(E_1 \cap E_3) = \mathbf{P}(E_1)\mathbf{P}(E_3), \\ & \mathbf{P}(E_2 \cap E_3) = \mathbf{P}(E_2)\mathbf{P}(E_3). \end{aligned}$$

$$\text{b)} \quad \mathbf{P}(E_1 \cap E_2 \cap E_3) = \mathbf{P}(E_1)\mathbf{P}(E_2)\mathbf{P}(E_3)$$

a) oder b) allein reichen i.a. nicht für die Unabhängigkeit, wie das folgende Beispiel zeigt:

Beispiel:

(Z_1, Z_2) sei ein zweifacher $\frac{1}{2}$ -Münzwurf,

$$E_1 := \{Z_1 = 1\}, \quad E_2 := \{Z_2 = 1\}, \quad E_3 := \{Z_1 = Z_2\}.$$

E_1, E_2, E_3 sind paarweise unabhängig (warum?),
aber nicht unabhängig:

das Ereignis $E_1 \cap E_2$ zieht das Ereignis E_3 nach sich! Also:

$$\mathbf{P}(E_1 \cap E_2 \cap E_3) = \mathbf{P}(E_1 \cap E_2) > \mathbf{P}(E_1 \cap E_2)\mathbf{P}(E_3);$$

die Ereignisse E_1, E_2, E_3 sind somit nicht unabhängig.

Also reicht Bedingung a) allein nicht
für die Unabhängigkeit der drei Ereignisse!

Auch Bedingung b) allein reicht nicht:

In den Übungen werden wir ein Beispiel von drei Ereignissen E_1, E_2, E_3 sehen, die nicht unabhängig sind, aber für die

$$\mathbf{P}(E_1 \cap E_2 \cap E_3) = \mathbf{P}(E_1)\mathbf{P}(E_2)\mathbf{P}(E_3)$$

gilt.

Vorlesung 5b

Unabhängigkeit

Teil 5

Unabhängige Teilbeobachtungen abhängiger Inputs

(vgl. Buch S. 68, Bsp. 1)

Sind X_1 und X_2 unabhängig,

dann auch $h_1(X_1)$ und $h_2(X_2)$.

Denn:

$$\begin{aligned} & \mathbf{P}(h_1(X_1) \in B_1, h_2(X_2) \in B_2) \\ &= \mathbf{P}(X_1 \in h_1^{-1}(B_1), X_2 \in h_2^{-1}(B_2)) \\ &= \mathbf{P}(X_1 \in h_1^{-1}(B_1)) \mathbf{P}(X_2 \in h_2^{-1}(B_2)) \\ &= \mathbf{P}(h_1(X_1) \in B_1) \mathbf{P}(h_2(X_2) \in B_2). \quad \square \end{aligned}$$

Durch den Übergang zu
“Teilbeobachtungen” $h_1(X_1)$ und $h_2(X_2)$
können aber auch aus abhängigen Zufallsvariablen X_1, X_2
voneinander unabhängige Zufallsvariable
 $h_1(X_1), h_2(X_2)$ entstehen:

Gewisse **Teilaspekte** von **abhängigen Zufallsvariablen**
können **unabhängig** sein:

Beispiel:

(X_1, X_2) seien rein zufällige “Zwei aus $\{1, 2, \dots, 32\}$ ”.

Offenbar sind X_1 und X_2 nicht unabhängig.

Aber: die Ereignisse

$$E_1 := \{X_1 \in \{1, 9, 17, 25\}\}, \quad E_2 := \{1 \leq X_2 \leq 8\}.$$

sind **unabhängig**.

Denn

$$\mathbf{P}(E_1) = \frac{1}{8}, \mathbf{P}(E_2) = \frac{1}{4}, \mathbf{P}(E_1 \cap E_2) = \frac{1 \cdot 7 + 3 \cdot 8}{32 \cdot 31} = \frac{1}{32}.$$

Als Anknüpfung an Teil 1 der heutigen Vorlesung geben wir hier auch die Tafel der 4 Wahrscheinlichkeiten

$P(E_1 \cup E_2)$, $P(E_1 \cup E_2^c)$, $P(E_1^c \cup E_2)$, $P(E_1^c \cup E_2^c)$:

	E_2	E_2^c
E_1	$\frac{1 \cdot 7 + 3 \cdot 8}{32 \cdot 31}$	$\frac{1 \cdot 24 + 3 \cdot 23}{32 \cdot 31}$
E_1^c	$\frac{7 \cdot 7 + 21 \cdot 8}{32 \cdot 31}$	$\frac{7 \cdot 24 + 21 \cdot 23}{32 \cdot 31}$

Man sieht: Die Zeilen stehen im Verhältnis 1 : 7,
die Spalten stehen im Verhältnis 1 : 3.

Das kommt daher, dass der relative Anteil von $A_1 := \{1, 9, 17, 25\}$ in $A_2 := \{1, 2, \dots, 8\}$ ebenso groß ist wie der von A_1 in A_2^c .

Vorlesung 5b

Unabhängigkeit

Teil 6

Abhängige Verquickungen unabhängiger Bausteine

(vgl. Buch S. 68, Bsp. 2)

Ein Beispiel für “indirekte Abhängigkeiten”:

Wir haben zwei gezinkte Münzen, mit $P(\text{Kopf} = 1) = 0.9$,
und eine faire Münze (mit $P(\text{Kopf} = 1) = 0.5$).

Jede der drei Münzen wird einmal geworfen;
die zufälligen Ergebnisse sind G, H für die beiden gezinkten
und F für die faire Münze.

Lernt man aus der Information, ob F so wie G ausfällt, etwas
für die Prognose, ob F so wie H ausfällt?

In der Tat lernt man etwas, denn:

Wenn F wie G ausfällt, fällt F eher als Kopf aus, und dann
fällt H wohl auch eher so aus wie F

Hier ist eine mathematische Analyse:

G und H seien einfache p -Münzwurfe,

F sei ein einfacher $1/2$ -Münzwurf;

G, H, F seien unabhängig. Dann gilt:

$$\mathbf{P}(G = F) = p\frac{1}{2} + q\frac{1}{2} = \frac{1}{2}, \quad \text{analog: } \mathbf{P}(H = F) = \frac{1}{2}$$

$$\mathbf{P}(G = F, H = F) = (p^2 + q^2)\frac{1}{2}$$

$$p^2 + q^2 = \frac{1}{2}((p + q)^2 + (p - q)^2) = \frac{1}{2}(1 + (p - q)^2) \geq \frac{1}{2},$$

mit “=” genau dann wenn $p = \frac{1}{2}$.

Für $p \neq \frac{1}{2}$ sind $\{G = F\}$ und $\{Y = F\}$ “positiv korreliert”!

Positiv korrelierte Ereignisse

sind Ereignisse E_1, E_2 mit der Eigenschaft

$$(*) \quad \mathbf{P}(E_1 \cap E_2) > \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$$

Offenbar sind dann E_1 und E_2 nicht unabhängig.

Was bedeutet die Beziehung $(*)$ anschaulich?

Schreiben wir $\mathbf{P}(E_1) =: p_1$, $\mathbf{P}(E_2) =: p_2$.

Dann ist

$$\begin{aligned} (*) &\iff \mathbf{E}[I_{E_1}I_{E_2} - p_1p_2] > 0 \\ &\iff \mathbf{E}[I_{E_1}I_{E_2} - p_1I_{E_2} - p_2I_{E_1} + p_1p_2] > 0 \\ &\iff \mathbf{E}[(I_{E_1} - p_1)(I_{E_2} - p_2)] > 0. \end{aligned}$$

E_1 und E_2 haben dann also die Tendenz,
gemeinsam einzutreten oder gemeinsam nicht einzutreten.
Solche Ereignisse E_1 und E_2 nennt man **positiv korreliert**.

Vorlesung 6a

Varianz und Kovarianz

Teil 1

Varianz und Standardabweichung: Elementare Eigenschaften

(Buch S. 24)

X sei reellwertige Zufallsvariable
mit endlichem Erwartungswert μ .

Die **Varianz** von X ist definiert als

$$\text{Var}[X] := \mathbf{E}[(X - \mu)^2],$$

die erwartete quadratische Abweichung
der Zufallsvariablen X von ihrem Erwartungswert μ .

Statt $\text{Var}[X]$ schreiben wir auch

$$\text{Var}X$$

oder

$$\sigma_X^2$$

oder (wenn klar ist, welche Zufallsvariable gemeint ist)

$$\sigma^2.$$

Wie ändert sich die Varianz,
wenn man X um eine Konstante verschiebt?

$$\text{Var}[X + d] = \mathbb{E}[(X + d) - (\mu + d)]^2 = \text{Var}X$$

Dann bleibt die Varianz gleich!

Und wenn man X mit einer Konstanten multipliziert
("skaliert")?

$$\text{Var}[cX] = \mathbb{E}[(cX - c\mu)^2] = c^2 \text{Var}X$$

Der Faktor tritt **quadratisch** heraus!

Definition.

Die **Standardabweichung (Streuung)** von X
ist die **Wurzel aus der Varianz**:

$$\sigma := \sigma_X := \sqrt{\text{Var } X}$$

$$= \sqrt{\mathbf{E}[(X - \mu)^2]}.$$

Sie gibt an, mit welcher „typischen Abweichung“
der Zufallsvariablen X von ihrem Erwartungswert
man “rechnen” sollte.

Es gilt:

$$\sigma_{X+d} = \sigma_X,$$

$$\sigma_{cX} = c \sigma_X.$$

Man sagt: σ ist ein *Skalenparameter*.

Für Zufallsvariable mit endlichem Erwartungswert gilt:

$$\text{Var}[X] = 0 \iff \mathbf{P}(X = \mathbf{E}[X]) = 1 .$$

Die Äquivalenz sieht man aus der Gleichheit
 $\text{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$
zusammen mit dem
Satz über die Positivität des Erwartungswertes.

Wenn es einen Ausgang (in diesem Fall eine Zahl) $\bar{a}$ gibt mit

$$\mathbf{P}(X = \bar{a}) = 1$$

sagt man auch:

X ist fast sicher konstant.

Wie der Erwartungswert ist auch die Varianz von X
durch die Verteilung von X bestimmt:

Hat X Erwartungswert μ , so gilt

im diskreten Fall

(mit den Verteilungsgewichten $\rho(a)$, $a \in S \subset \mathbb{R}$)

$$\text{Var } X = \sum_{a \in S} (a - \mu)^2 \rho(a)$$

und falls X Dichte $f(a)da$, $a \in \mathbb{R}$, besitzt:

$$\text{Var } X = \int_{\mathbb{R}} (a - \mu)^2 f(a) da .$$

Einfache Beispiele

Beispiel 1:

Eine faire Münze wird dreimal geworfen.

$X = Z_1 + Z_2 + Z_3$ ist die “Anzahl Köpfe”, $E[X] = \frac{3}{2}$.

Var [X]

$$\begin{aligned} &= \frac{1}{8}(0 - \frac{3}{2})^2 + \frac{3}{8}(1 - \frac{3}{2})^2 + \frac{3}{8}(2 - \frac{3}{2})^2 + \frac{1}{8}(3 - \frac{3}{2})^2 \\ &= \frac{1}{8} \cdot \frac{9 + 3 + 3 + 9}{4} = 3 \cdot \frac{1}{4} \end{aligned}$$

Beispiel 2:

Eine p -Münze wird einmal geworfen.

$$\mathbf{P}(Z = 1) = p, \quad \mathbf{P}(Z = 0) = q$$

$$\mathbf{Var}[Z]$$

$$= q(0 - p)^2 + p(1 - p)^2 = qp^2 + p^2q$$

$$= pq(p + q) = pq.$$

Beispiel 3:

Anzahl der Erfolge beim zweimaligen p -Münzwurf.

$$\text{Var}[Z_1 + Z_2] = ?$$

Wir rechnen mit Zufallsvariablen:

$$\begin{aligned} \mathbf{E}[(Z_1 + Z_2 - 2p)^2] &= \mathbf{E}[(Z_1 - p + Z_2 - p)^2] \\ &= \mathbf{E}[(Z_1 - p)^2 + (Z_2 - p)^2 + 2(Z_1 - p)(Z_2 - p)] \\ &= \mathbf{E}[(Z_1 - p)^2] + \mathbf{E}[(Z_2 - p)^2] + 2\mathbf{E}[(Z_1 - p)(Z_2 - p)] \end{aligned}$$

Der **letzte Term** verschwindet wegen der Produktformel

(V5b2), denn Z_1 und Z_2 sind unabhängig. Also:

$$\text{Var}[Z_1 + Z_2] = \text{Var}[Z_1] + \text{Var}[Z_2] = 2pq.$$

Vorlesung 6a

Varianz und Kovarianz

Teil 2

Summen unabhängiger Zufallsvariabler:

Varianz der Binomialverteilung,
 $\sqrt{n}$ -Gesetz, Chebyshev-Ungleichung und
Schwaches Gesetz der großen Zahlen

(Buch S. 50, S. 26, S74)

Die Rechnung auf der letzten Folie von Teil 1 weist direkt den Weg zum

Satz: $X_1, \dots, X_n$ seien paarweise unabhängige reellwertige Zufallsvariable mit endlichen Erwartungswerten. Dann gilt:

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n].$$

Beweis: Mit $\mu_i := \mathbf{E}[X_i]$ ist nach Definition der Varianz

$$\text{Var}[X_1 + \dots + X_n] = \mathbf{E} \left[\left(\sum_{i=1}^n (X_i - \mu_i) \right)^2 \right]$$

Wegen der Linearität des EW ist dies

$$\sum_{i,j=1}^n \mathbf{E} [(X_i - \mu_i)(X_j - \mu_j)].$$

Die Summanden mit $i \neq j$ sind $= 0$ wegen der Produktformel (V5b2), die “Diagonalterme” ($i = j$) summieren zu $\text{Var}[X_1] + \dots + \text{Var}[X_n]$. $\square$

Als unmittelbare Folgerung aus diesem Satz
zusammen mit Bsp. 2 in Teil 1 ergibt sich

Die Varianz der $\text{Bin}(n, p)$ -Verteilung ist npq .

Das $\sqrt{n}$ -Gesetz für die Standardabweichung

folgt aus der Additivität der Varianz unabhängiger ZV'er:

Seien $X_1, \dots, X_n$ unabhängig
und identisch verteilt mit Varianz σ^2 .

Dann gilt für die Varianz
des Mittelwerts $M_n := \frac{1}{n}(X_1 + \dots + X_n)$:

$$\text{Var}[M_n] = \frac{1}{n^2} \cdot n\sigma^2 = \frac{1}{n}\sigma^2.$$

Man hat somit das berühmte $\sqrt{n}$ -Gesetz:

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}\sigma.$$

Die Ungleichung von Chebyshev

liefert eine Quantifizierung der anschaulichen Botschaft

“Je weniger eine reellwertige Zufallsvariable streut,
mit um so größerer Wahrscheinlichkeit
fällt sie nahe zu ihrem Erwartungswert aus.”

Die Ungleichung von Chebyshev:

Y sei eine reellwertige Zufallsvariable
mit endlichem Erwartungswert μ .

Dann gilt für alle $\varepsilon > 0$:

$$\mathbf{P}(|Y - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{Var}[Y]$$

Beweis:

Mit $X := (Y - \mu)^2$ ist die **Behauptung** äquivalent zu

$$\mathbf{P}(X \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} \mathbf{E}[X].$$

Das aber folgt aus der Ungleichung von Markov. $\square$

Das Schwache Gesetz der Großen Zahlen

ist eine unmittelbare Folgerung aus dem $\sqrt{n}$ -Gesetz zusammen mit der Ungleichung von Chebyshev:

Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit Erwartungswert μ und endlicher Varianz. Dann gilt für die Mittelwerte

$$M_n := \frac{1}{n}(X_1 + \dots + X_n):$$
$$\mathbf{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{Var}[M_n] \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Vorlesung 6a

Varianz und Kovarianz

Teil 4

Umrechnung von $\text{Var}[X]$,
Varianz der Poissonverteilung

(Buch S. 24, 29)

Manchmal verwenden wir die “hilfreiche Formel”:

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

Beweis:

$$\begin{aligned}\text{E}[(X - \mu)^2] &= \text{E}[X^2 - 2\mu X + \mu^2] \\ &= \text{E}[X^2] - 2\mu\text{E}[X] + \mu^2 = \text{E}[X^2] - \mu^2\end{aligned}$$

(wegen Linearität des Erwartungswertes)

Zur Erinnerung:

Die Poissonverteilung mit Parameter λ
entsteht als Grenzwert von Binomialverteilungen mit

$$n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda.$$

Weil dann npq gegen λ konvergiert,
steht zu vermuten:

Die Varianz einer $\text{Pois}(\lambda)$ -verteilten Zufallsvariablen X ist λ .

Beweis durch Rechnung:

$$\begin{aligned}\mathbf{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2.\end{aligned}$$

Also:

$$\mathbf{E}[X^2] = \mathbf{E}[X(X-1)] + \mathbf{E}[X] = \lambda^2 + \lambda,$$

$$\mathbf{Var}[\mathbf{X}] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \lambda.$$

Vorlesung 6a

Varianz und Kovarianz

Teil 4

Die Varianz einer Summe von ZV'en und
die Kovarianz von zwei ZV'en

(Buch S. 60)

Beim zweifachen p -Münzwurf Z_1, Z_2
ergab sich aus der Unabhängigkeit der Z_i :
$$\text{Var}[Z_1 + Z_2] = \text{Var}[Z_1] + \text{Var}[Z_2].$$

Wie “streuen” Summen von
nicht unabhängigen Zufallsgrößen?

Wie steht’s mit der
Varianz einer Summe von Zufallsvariablen?

$$\begin{aligned}\text{Var}[X + Y] &= \mathbf{E}[(X - \mu_X) + (Y - \mu_Y)]^2 \\ &= \mathbf{E}[(X - \mu_X)^2] + \mathbf{E}[(Y - \mu_Y)^2] + 2\mathbf{E}[(X - \mu_X)(Y - \mu_Y)]\end{aligned}$$

Mit der Definition der **Kovarianz**

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$$

bekommen wir

$$\boxed{\text{Var}[X + Y] = \text{Var } X + \text{Var } Y + 2\text{Cov}[X, Y].}$$

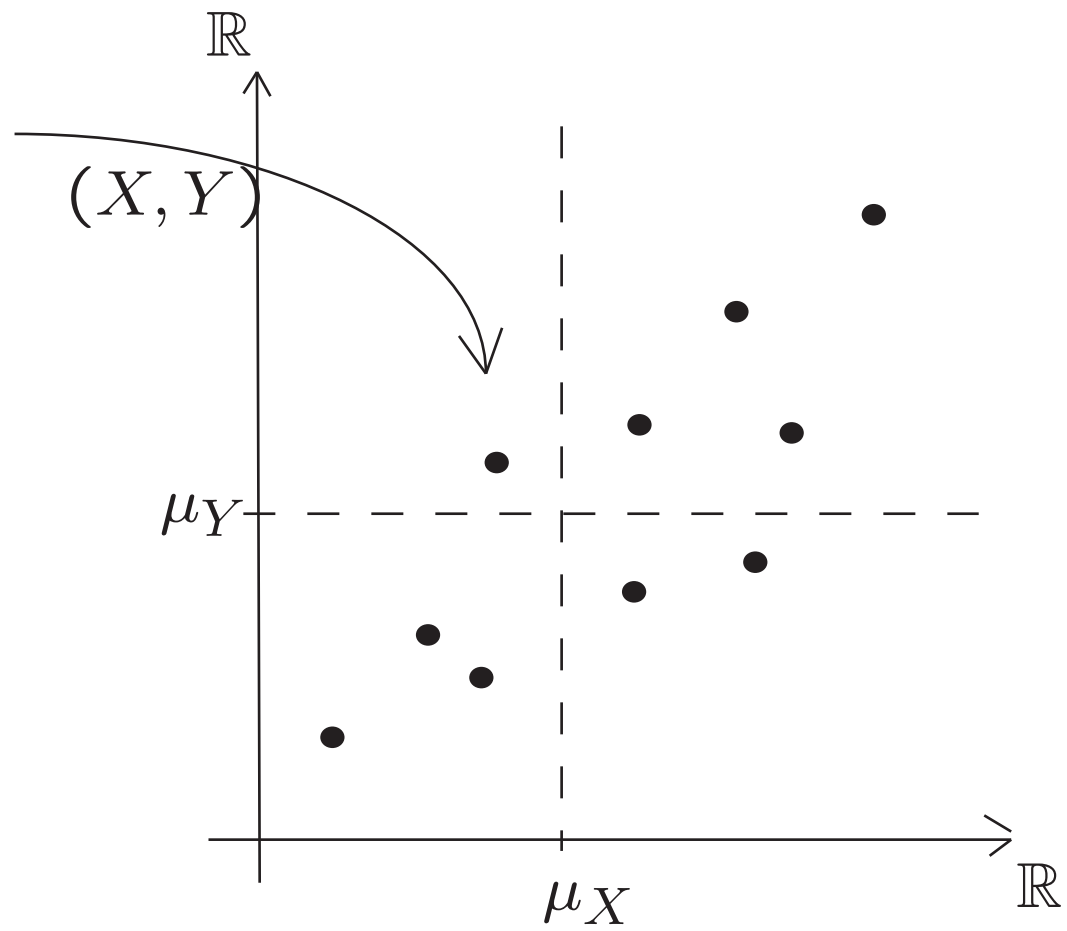
Die Kovarianz

$$\text{Cov}[X, Y] = \text{E}[(X - \mu_X)(Y - \mu_Y)]$$

ist positiv,

wenn X und Y die Tendenz haben,
gemeinsam über bzw. gemeinsam unter
ihrem Erwartungswert auszufallen.

(Größere Abweichungen fallen dabei mehr ins Gewicht.)



Ist $\text{Cov}[X, Y]$

$= 0$,	dann nennt man X, Y	unkorreliert
> 0 ,	...	positiv korreliert
< 0 ,	...	negativ korreliert.

Zwei Spezialfälle:

$$Y = X :$$

$$\text{Cov}[X, Y] = \mathbf{E}[(X - \mu_X)(X - \mu_X)] = \text{Var}[X]$$

$$Y = -X :$$

$$\text{Cov}[X, Y] = \mathbf{E}[(X - \mu_X)(-X + \mu_X)] = -\text{Var}[X]$$

$$\text{Var}[X + Y] = \text{Var}X + \text{Var}Y + 2\text{Cov}[X, Y].$$

Ganz analog ergibt sich:

$$\begin{aligned} & \text{Var}[Z_1 + \cdots + Z_n] \\ &= \text{Var} Z_1 + \cdots + \text{Var} Z_n + 2 \sum_{i < j} \text{Cov}[Z_i, Z_j] \end{aligned}$$

Wir halten fest:

Sind $X_1, \dots, X_n$ reellwertige Zufallsvariable
mit endlicher Varianz und

$$\text{Cov}[X_i, X_j] = 0 \quad \text{für } i \neq j$$

(man sagt dafür auch: die X_i sind *paarweise unkorreliert*)

dann gilt:

$$\text{Var}[X_1 + \dots + X_n] = \text{Var } X_1 + \dots + \text{Var } X_n$$

Eine Umformung von $\text{Cov}[X, Y]$:

$$\begin{aligned}\mathbf{E}[(X - \mu_X)(Y - \mu_Y)] &= \mathbf{E}[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y] \\ &= \mathbf{E}[XY] - \mu_X\mu_Y\end{aligned}$$

wegen der Linearität des Erwartungswertes.

So bekommen wir die (manchmal) hilfreiche Formel

$$\boxed{\text{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]}$$

$$\text{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

Speziell für zwei Ereignisse E_1, E_2 :

$$\text{Cov}[I_{E_1}, I_{E_2}] = \mathbf{P}(E_1 \cap E_2) - \mathbf{P}(E_1)\mathbf{P}(E_2).$$

Vgl. die Definition
der positiven/negativen Korreliertheit von Ereignissen
in V5b6.

$$\text{Cov}[X, Y] = \text{E}[XY] - \text{E}[X]\text{E}[Y]$$

Aus der Multiplikationsformel für den Erwartungswert
sehen wir

Unabhängige Zufallsvariable X und Y sind unkorreliert.

Speziell ist für unabhängige Zufallsvariable
mit endlichen Varianzen

die Varianz der Summe gleich der Summe der Varianzen.

Vorlesung 6a

Varianz und Kovarianz

Teil 5

Die Varianz der hypergeometrischen Verteilung

(Buch S. 32 und S. 61)

Ein Beispiel für die Anwendung der Formel

$$\text{Var}[Z_1 + \cdots + Z_n] = \\ \text{Var } Z_1 + \cdots + \text{Var } Z_n + 2 \sum_{i < j} \text{Cov}[Z_i, Z_j]$$

liefert das Zählen der “Erfolge”
beim Ziehen ohne Zurücklegen.

In einer Urne sind r rote und b blaue Kugeln.

Es wird n -mal ohne Zurücklegen gezogen.

$X :=$ Anzahl der gezogenen roten Kugeln.

$$\text{Var}[X] = ?$$

Zur Erinnerung:

Mit $g := r + b$ ist

$$\mathbf{P}(X = k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{g}{n}}, \quad k = 0, \dots, r.$$

X heißt **hypergeometrisch verteilt** mit Parametern n , g und r .

Erwartungswert und Varianz kann man direkt
über die Verteilungsgewichte ausrechnen (siehe Buch S. 32).

Es geht auch eleganter (vgl Buch S. 50/51):

Wir betrachten dazu die Zufallsvariable Z_i , die
... den Wert 1 annimmt, falls die i -te gezogene Kugel rot ist,
... und sonst den Wert 0.

Man sagt dafür auch:

Z_i ist die *Indikatorvariable*
(kurz: der *Indikator*)
des Ereignisses $\{i\text{-te gezogene Kugel rot}\}$.

$$X := Z_1 + \cdots + Z_n$$

$$\mathbf{E}[Z_i] = \mathbf{P}(Z_i = 1) = p, \quad \text{mit}$$

$$p := \frac{r}{g} \quad \text{der Anteil der roten Kugeln in der Urne.}$$

$$\text{Also: } \mathbf{E}[X] = np$$

(vgl. V3a4)

Und wie stehts mit der Varianz von X ?

$$X := Z_1 + \cdots + Z_n$$

$$\text{Var } X = \text{Var } Z_1 + \cdots + \text{Var } Z_n + 2 \sum_{1 \leq i < j \leq n} \text{Cov}[Z_i, Z_j]$$

Sei $g = r + b$ die Gesamtanzahl der Kugeln,
 $p := \frac{r}{g}$ der Anteil der roten Kugeln in der Urne,

$$q := 1 - p.$$

$$\text{Var } Z_i = pq.$$

$$\text{Cov}[Z_i, Z_j] = ?$$

Ein eleganter Weg zur Berechnung von $\text{Cov}[Z_i, Z_j]$:

Wir ziehen in Gedanken, bis die Urne leer ist

(d.h. wir setzen $n = g$.)

Wir ziehen in Gedanken, bis die Urne leer ist.

Dann ist

$$Z_1 + \cdots + Z_g = r,$$

also

$$\text{Var}[Z_1 + \cdots + Z_g] = 0.$$

$$0 = \text{Var } Z_1 + \cdots + \text{Var } Z_g + 2 \sum_{1 \leq i < j \leq g} \text{Cov}[Z_i, Z_j], \quad \text{d.h.}$$

$$0 = gpq + g(g-1)\text{Cov}[Z_1, Z_2], \quad \text{d.h.}$$

$$\text{Cov}[Z_1, Z_2] = -\frac{1}{g-1}pq$$

$$X = Z_1 + \cdots + Z_n$$

$$\text{Var } X = \text{Var } Z_1 + \cdots + \text{Var } Z_n + 2 \sum_{1 \leq i < j \leq n} \text{Cov}[Z_i, Z_j]$$

$$= n \text{Var } Z_1 + n(n-1) \text{Cov}[Z_1, Z_2]$$

$$= npq - n(n-1) \frac{1}{g-1} pq$$

$$= npq \left(1 - \frac{n-1}{g-1} \right) = npq \frac{g-n}{g-1}. \quad \square$$

Fazit:

Die Varianz von $\text{Hyp}(n, g, pg)$ ist

$$npq \frac{g - n}{g - 1}.$$

Zusammenfassung

des Wichtigsten aus V6a

$$\text{Var}[X] := \text{E}[(X - \mu)^2]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

Die Varianz einer Summe von unkorrelierten ZV'en
ist gleich der Summe der Varianzen,
die Varianz einer Summe von negativ korrelierten ZV'en
ist kleiner als die Summe der Varianzen.

Die Varianz von $\text{Bin}(n, p)$ ist npq .

Die Varianz von $\text{Hyp}(n, g, pg)$ ist $npq \frac{g-n}{g-1}$. Die Varianz einer $\text{Poisson}(\lambda)$ -verteilten Zufallsvariablen ist so groß wie ihr Erwartungswert, nämlich λ .

Ungleichung von Chebyshev:

$$\mathbf{P}(|Y - \mu| \geq \varepsilon \sigma_Y)) \leq \frac{1}{\varepsilon^2}$$

$$\begin{aligned}\text{Cov}[X, Y] &:= \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]\end{aligned}$$

Speziell für Indikatorvariable:

$$\begin{aligned}\text{Cov}[I_{E_1}, I_{E_2}] \\ = \mathbf{P}(E_1 \cap E_2) - \mathbf{P}(E_1)\mathbf{P}(E_2).\end{aligned}$$

Vorlesung 6b

Die Normalverteilung

Teil 1:

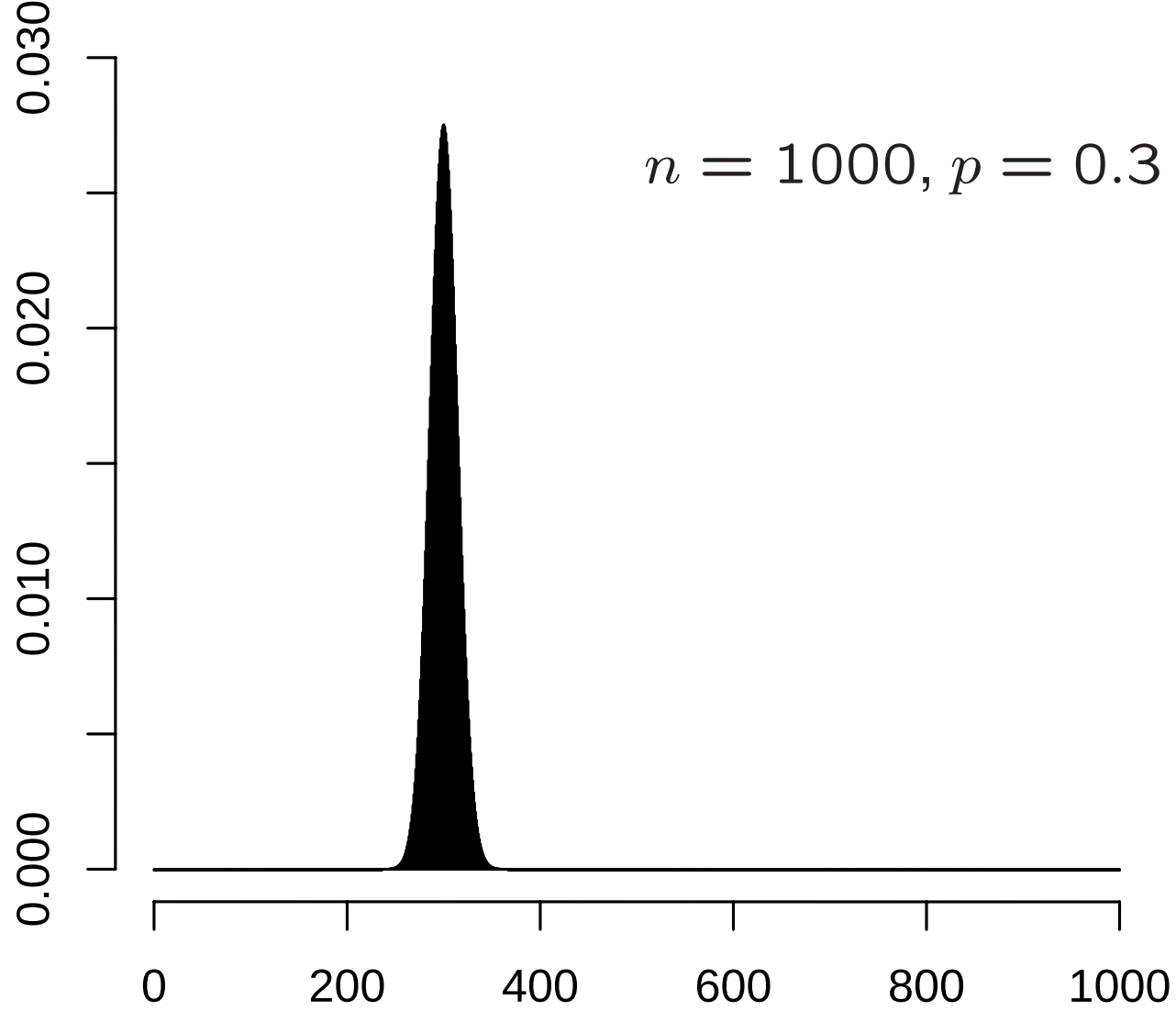
Von den Binomialgewichten
zur Gauß'schen Glockenkurve

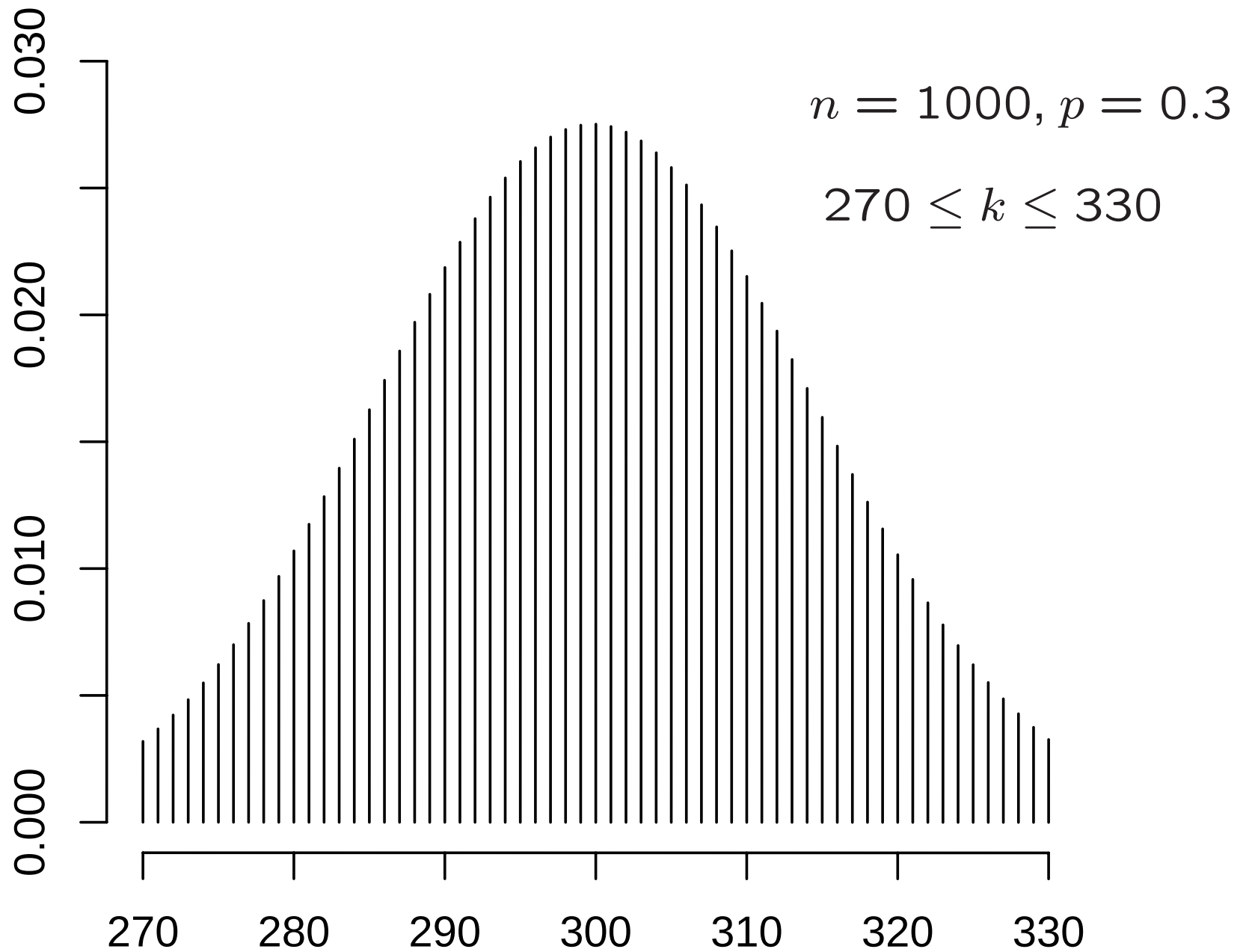
Binomialverteilungen

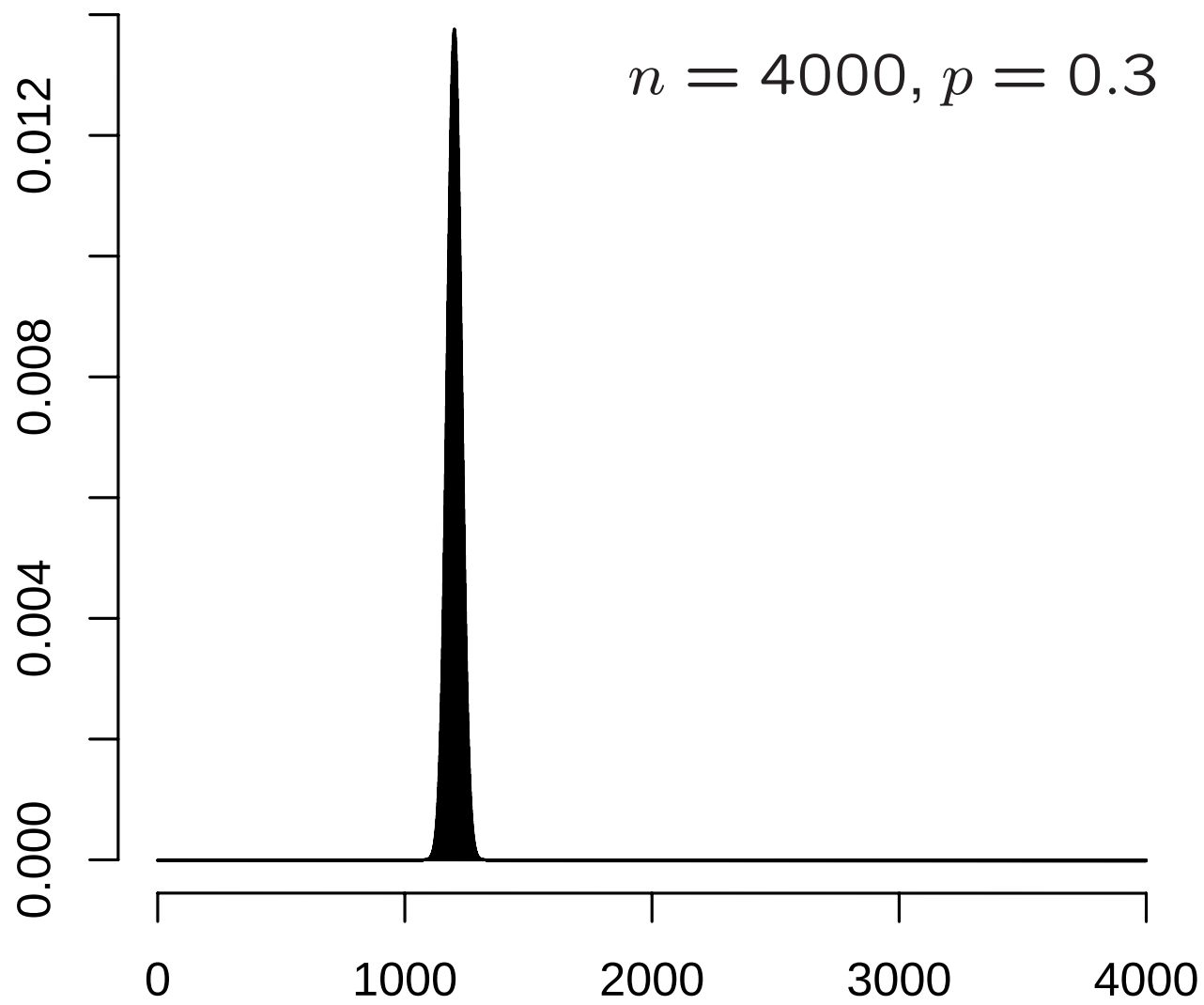
mit großem Erwartungswert und großer Varianz

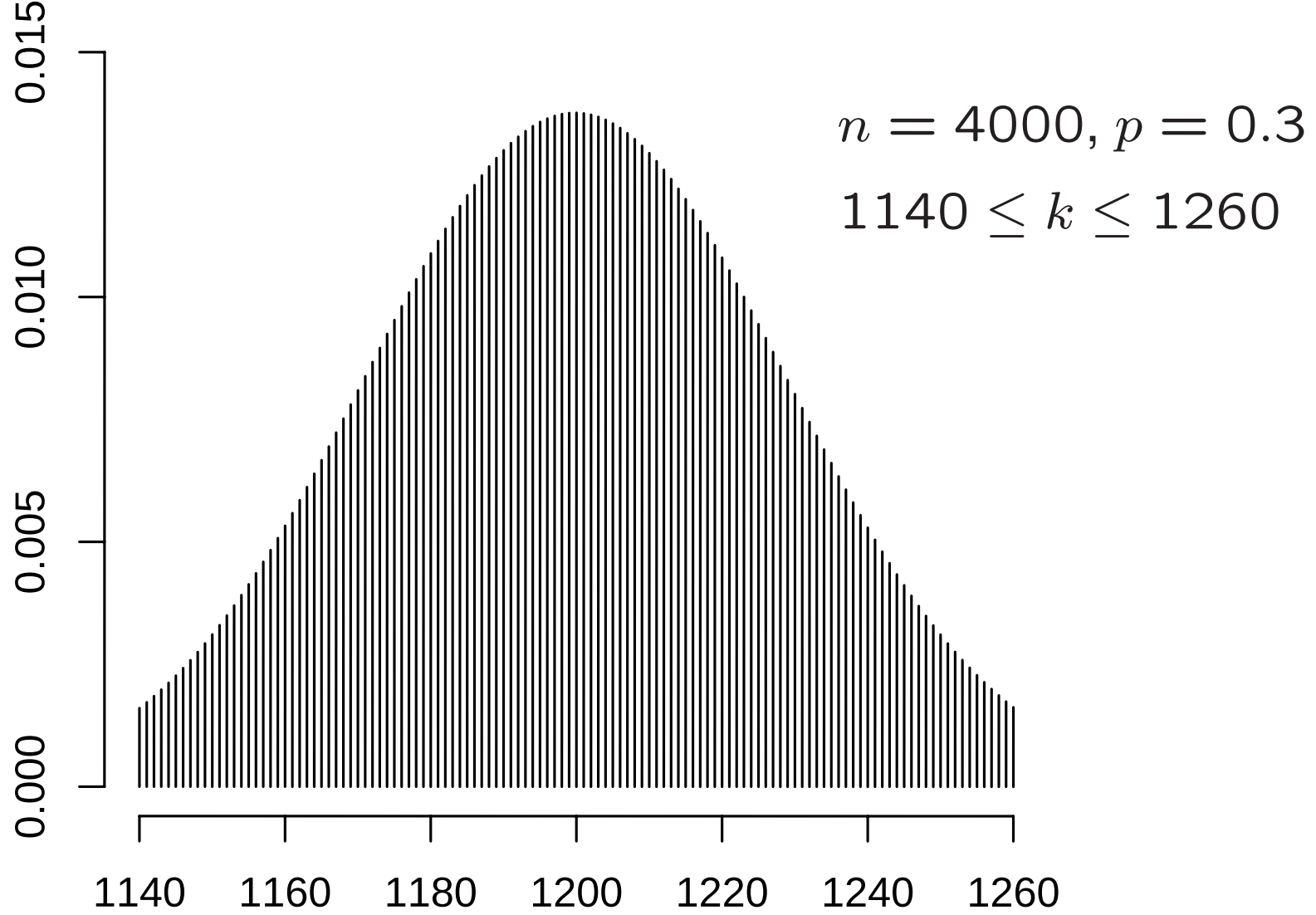
Wir wissen schon aus V4b4:
Für großes n und kleines p ,
so dass $np \approx npq \approx \lambda$,
ist die Binomialverteilung $\text{Bin}(n, p)$
approximativ gleich $\text{Pois}(\lambda)$.

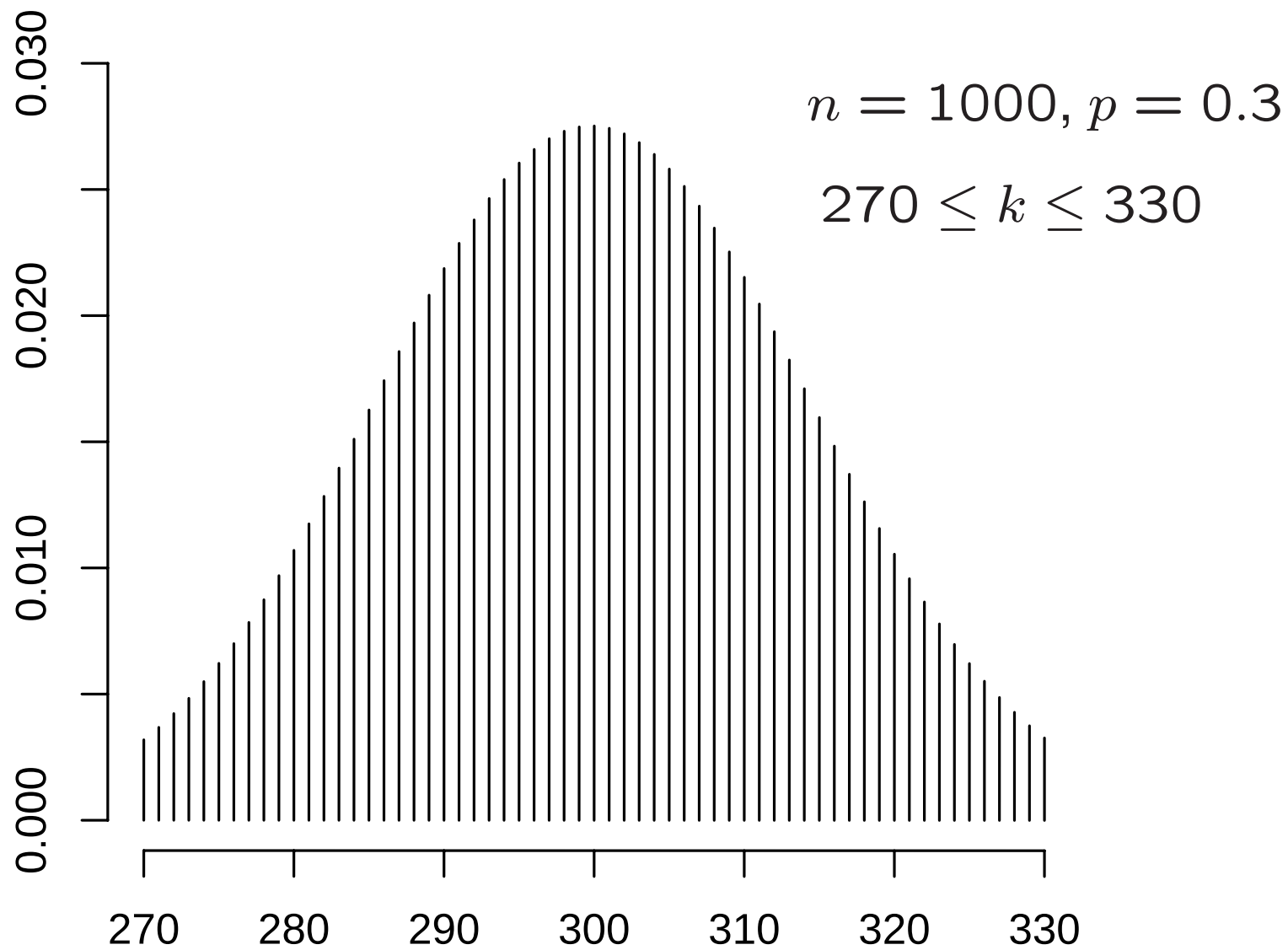
Wie aber sieht die $\text{Bin}(n, p)$ -Verteilung
mit großem n und großem npq aus? ?











Binomialverteilungen mit großem n und großer Varianz npq
sehen “glockenförmig” aus,
wenn man sie geeignet “ins Bild holt”.

Approximation der Binomialgewichte
mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{k - np}{\sqrt{npq}} \right)^2 \right) .$$

Approximation der Binomialgewichte
mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{k - np}{\sqrt{npq}} \right)^2 \right)$$

$$\mu := np, \quad \sigma := \sqrt{npq}$$

Approximation der Binomialgewichte
mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{k - \mu}{\sigma} \right)^2 \right)$$

$$\mu := np, \quad \sigma := \sqrt{npq}$$

Approximation der Binomialgewichte
mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{k - \mu}{\sigma} \right)^2 \right)$$
$$= \frac{1}{\sigma} \varphi \left(\frac{k - \mu}{\sigma} \right)$$

mit $\mu := np$, $\sigma := \sqrt{npq}$ und

$$\varphi(a) := \frac{1}{\sqrt{2\pi}} e^{-a^2/2}, \quad a \in \mathbb{R}.$$

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) \quad (*)$$

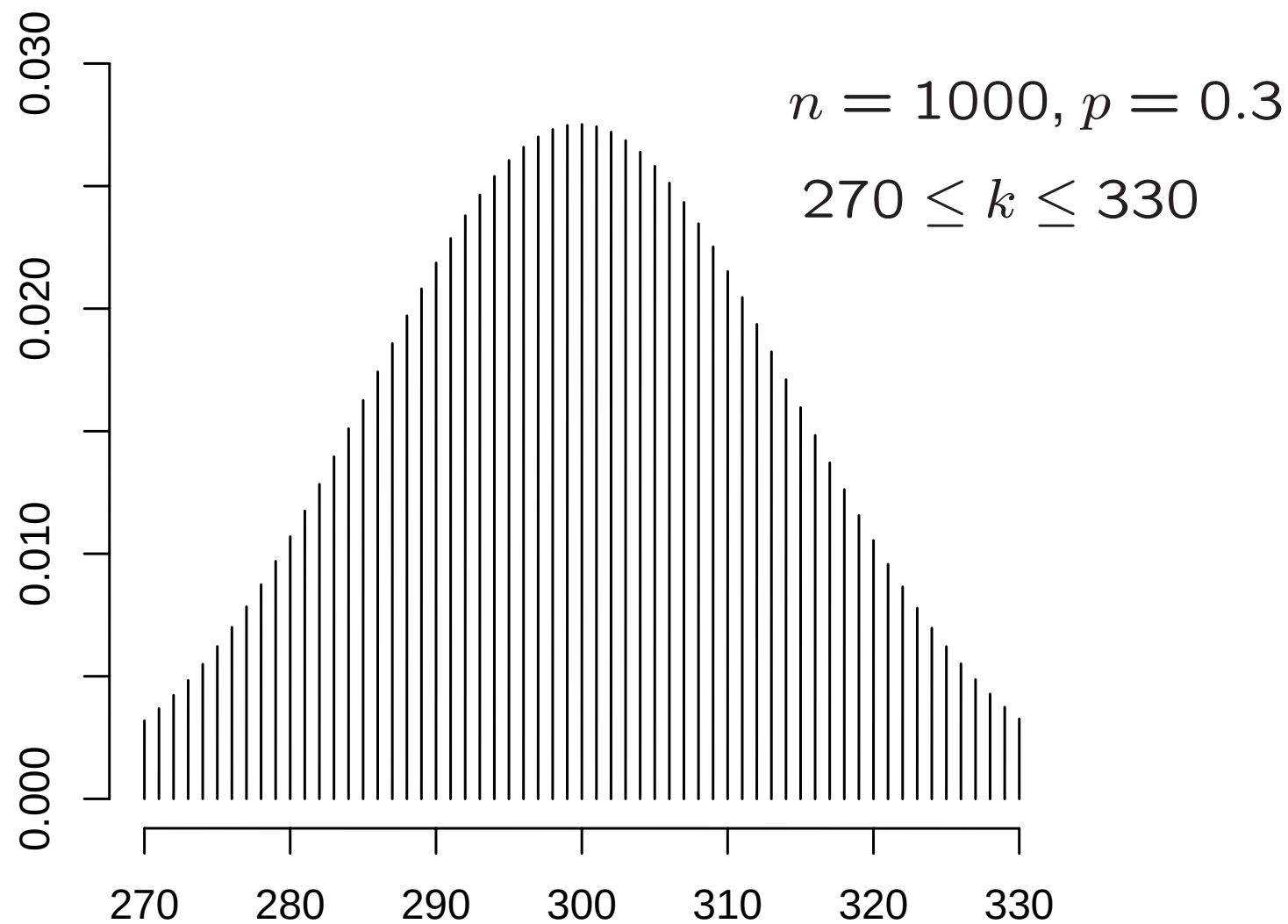
$$\text{mit } \mu := np, \quad \sigma := \sqrt{npq} \quad \text{und} \\ \varphi(a) := \frac{1}{\sqrt{2\pi}} e^{-a^2/2}, \quad a \in \mathbb{R}.$$

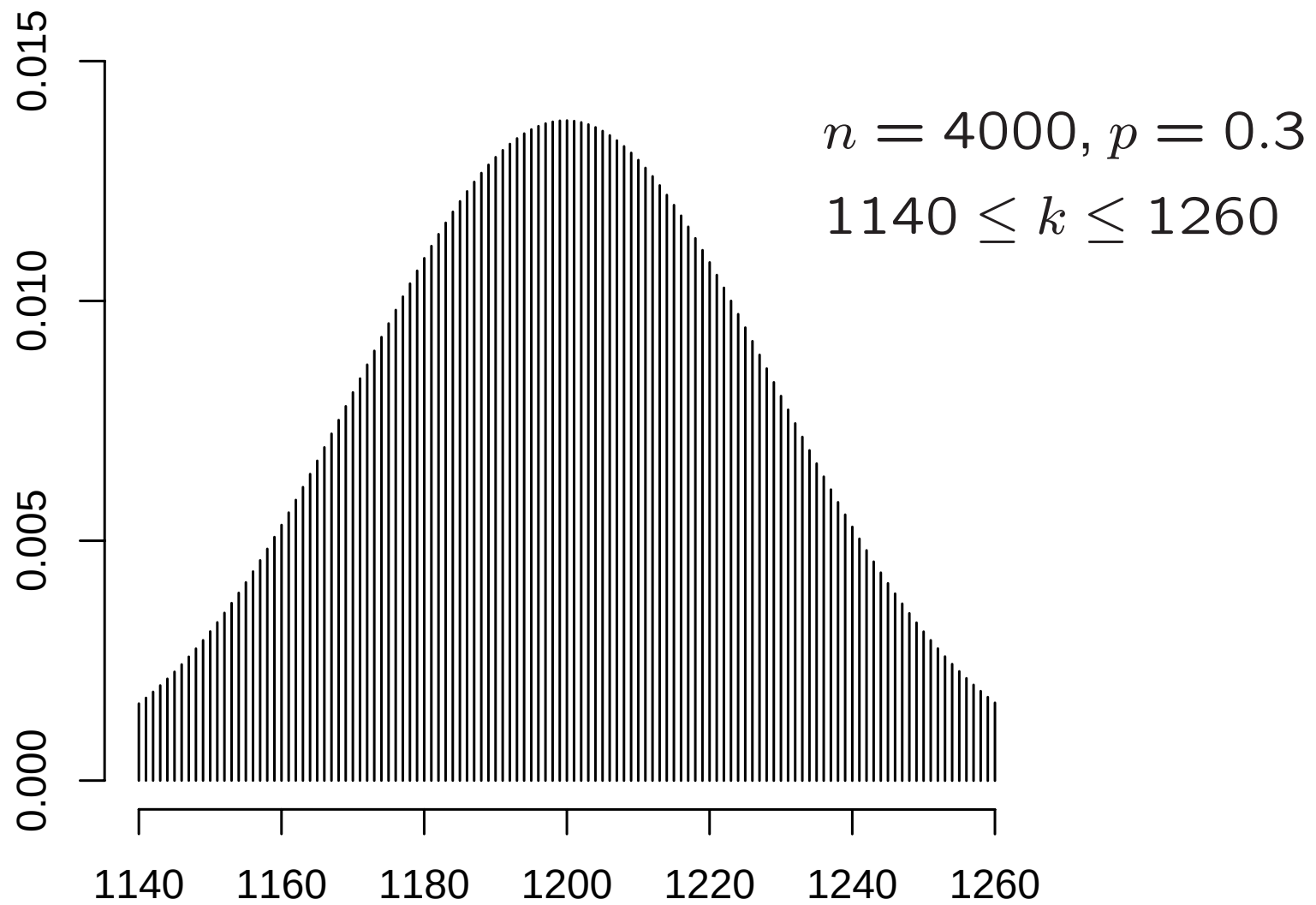
Wenn man n vervierfacht, verdoppelt sich σ .

Approximativ gilt dann in der Darstellung (*):

Im Bereich von einer Standardabweichung um das Zentrum μ
bringt man doppelt so viele k unter.

Das Gewicht jedes einzelnen k halbiert sich.





Die Ähnlichkeit ist unverkennbar:



Carl Friedrich Gauss (1777-1855)

Vorlesung 6b

Die Normalverteilung

Teil 2:

Die Standardnormalverteilung auf $\mathbb{R}$
und ihre Schwestern $N(\mu, \sigma^2)$

Wir hatten in V5a4 definiert:

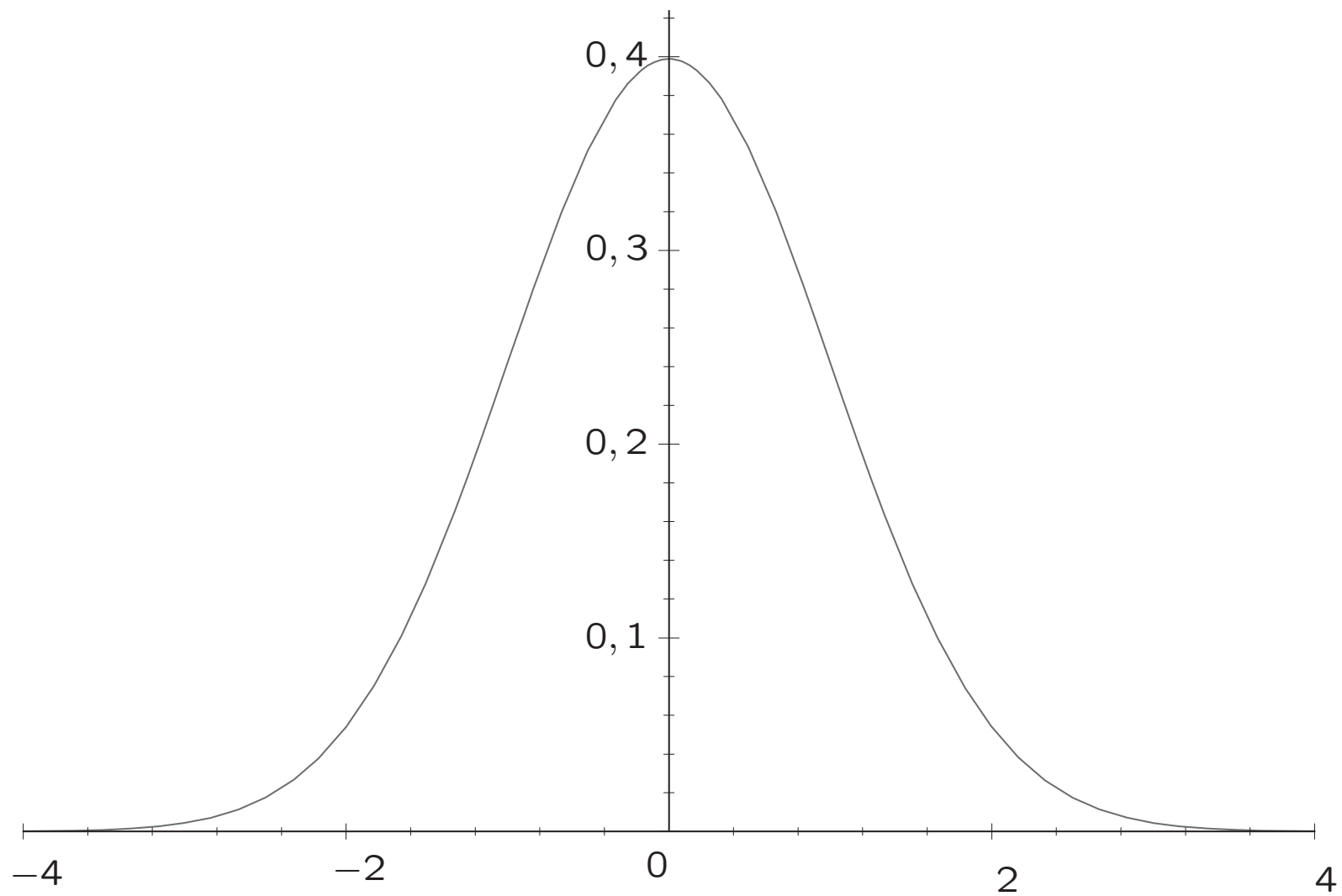
Eine $\mathbb{R}$ -wertige Zufallsvariable Z mit Dichte

$$\varphi(a) da := \frac{1}{\sqrt{2\pi}} e^{-a^2/2} da$$

heißt **standard-normalverteilt**.

Es gilt (vgl. Teil 4 der heutigen VL):

$$\int_{-\infty}^{+\infty} \varphi(a) da = 1.$$



Für ein standard-normalverteiltes Z
hatten wir in V5a4 festgestellt:

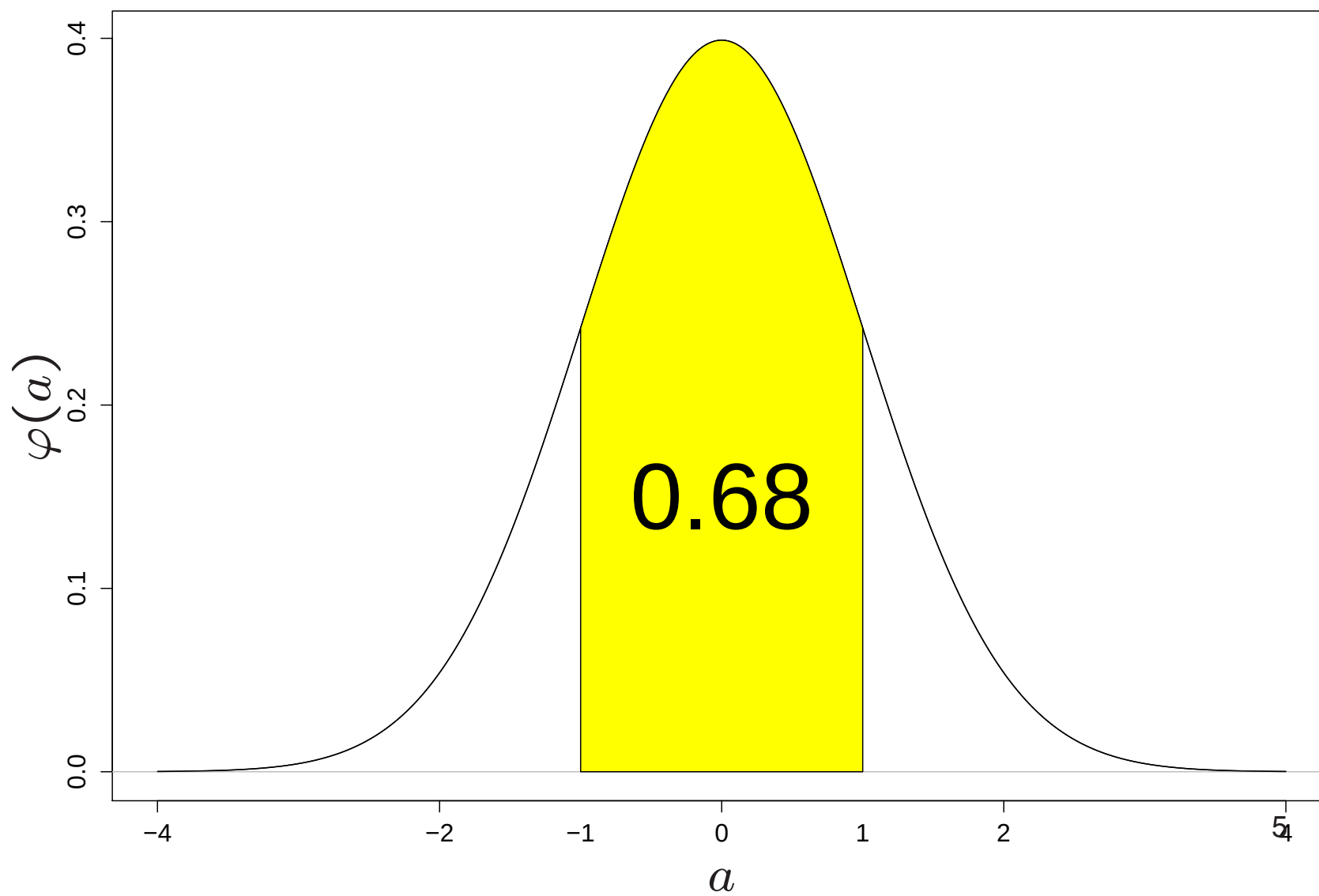
$$\mathbf{E}[Z] = 0, \quad \mathbf{E}[Z^2] = 1.$$

Also:

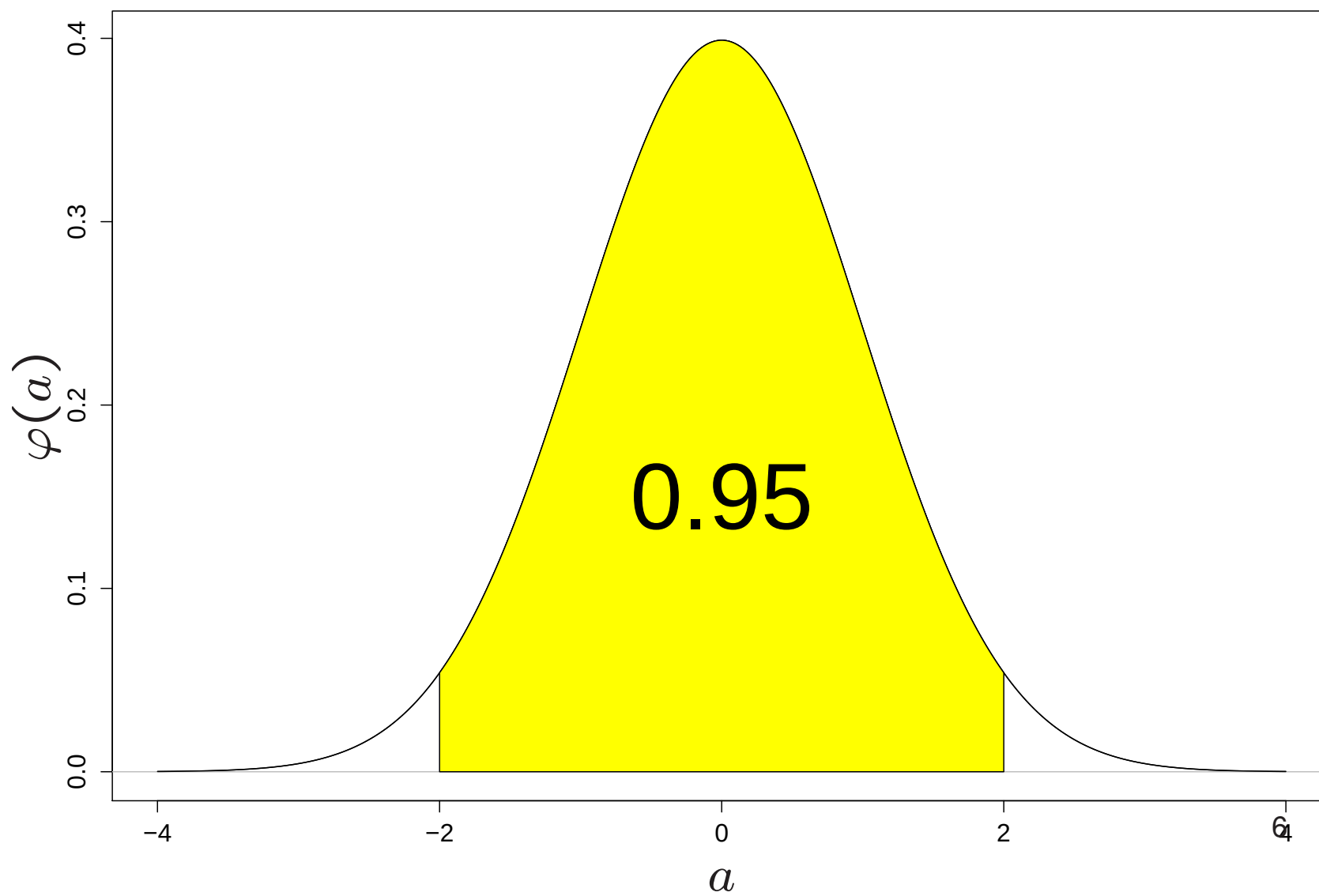
$$\mathbf{Var} [Z] = 1.$$

Zwei für die Praxis wichtige Zahlen:

$$\mathbf{P}(|Z| < 1) \approx 0.68$$



$$P(|Z| < 2) \approx 0.95$$



Sei Z standard-normalverteilt, $\mu \in \mathbb{R}$, $\sigma > 0$. Dann gilt für

$$X := \sigma Z + \mu :$$

$$\mathbf{E}[X] = \mu, \quad \mathbf{Var}[X] = \sigma^2,$$

und die Dichte von X ist $\varphi_{\mu, \sigma^2}(a) da$ mit

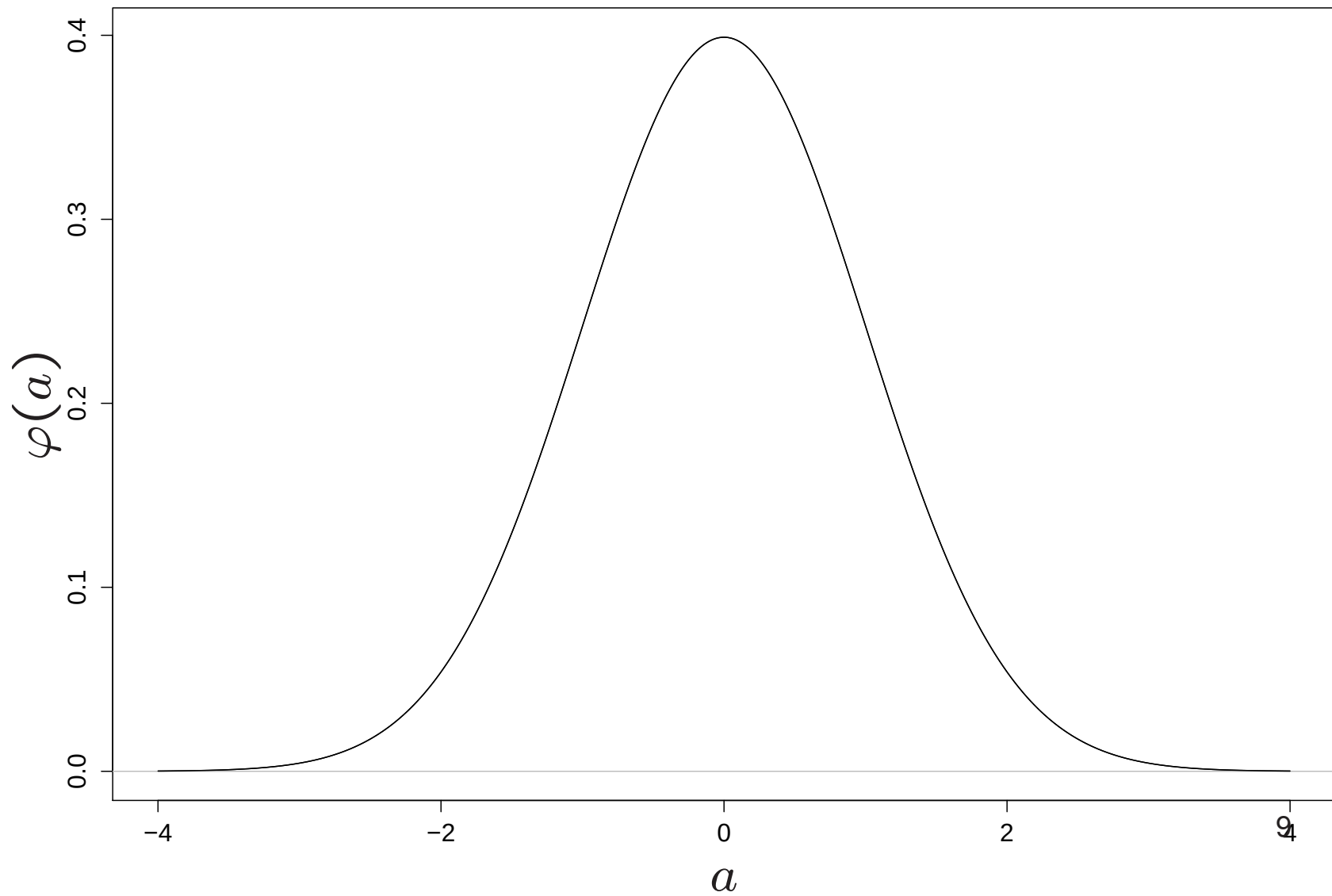
$$\varphi_{\mu, \sigma^2}(a) := \frac{1}{\sigma} \varphi\left(\frac{a - \mu}{\sigma}\right) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(a - \mu)^2}{2\sigma^2}}.$$

Eine Zufallsvariable mit Dichte $\varphi_{\mu, \sigma^2}(a) da$
heißt **normalverteilt mit Erwartungswert μ und Varianz σ^2** ,
kurz
 $N(\mu, \sigma^2)$ -verteilt.

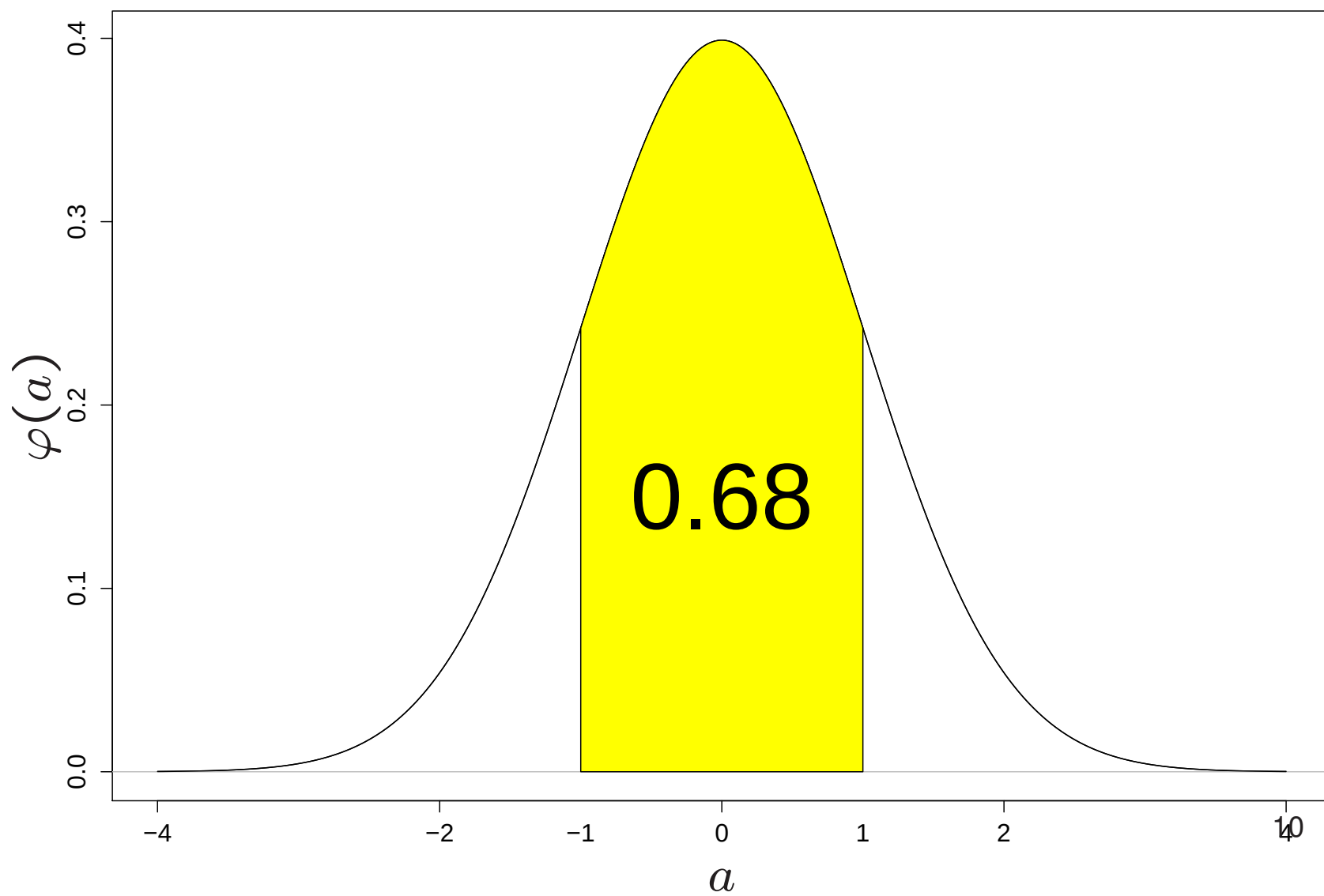
Ist Y $N(\mu, \sigma^2)$ -verteilt,

dann ist $\frac{Y - \mu}{\sigma}$ standard-normalverteilt.

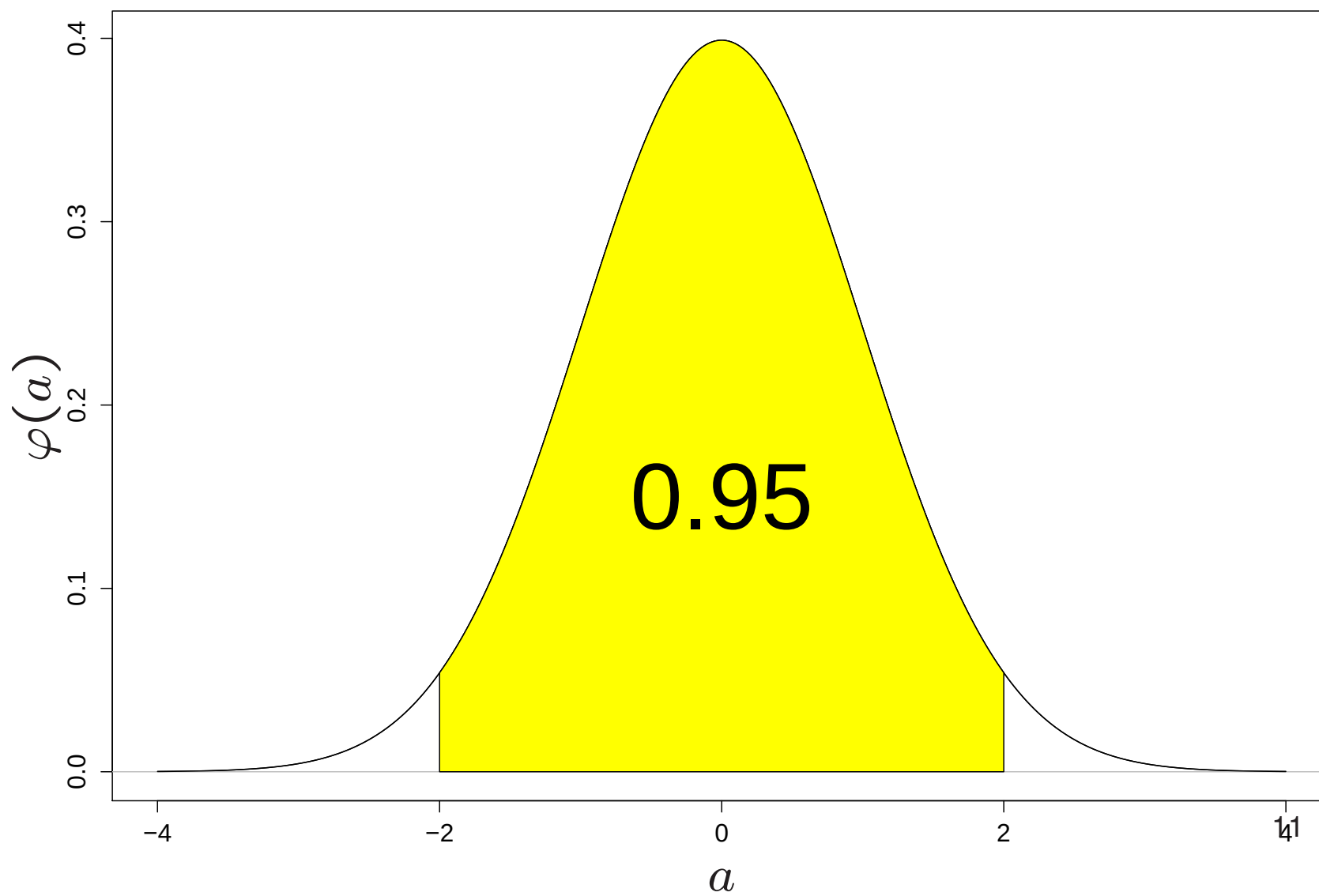
Dichtefunktion φ der Standard-Normalverteilung



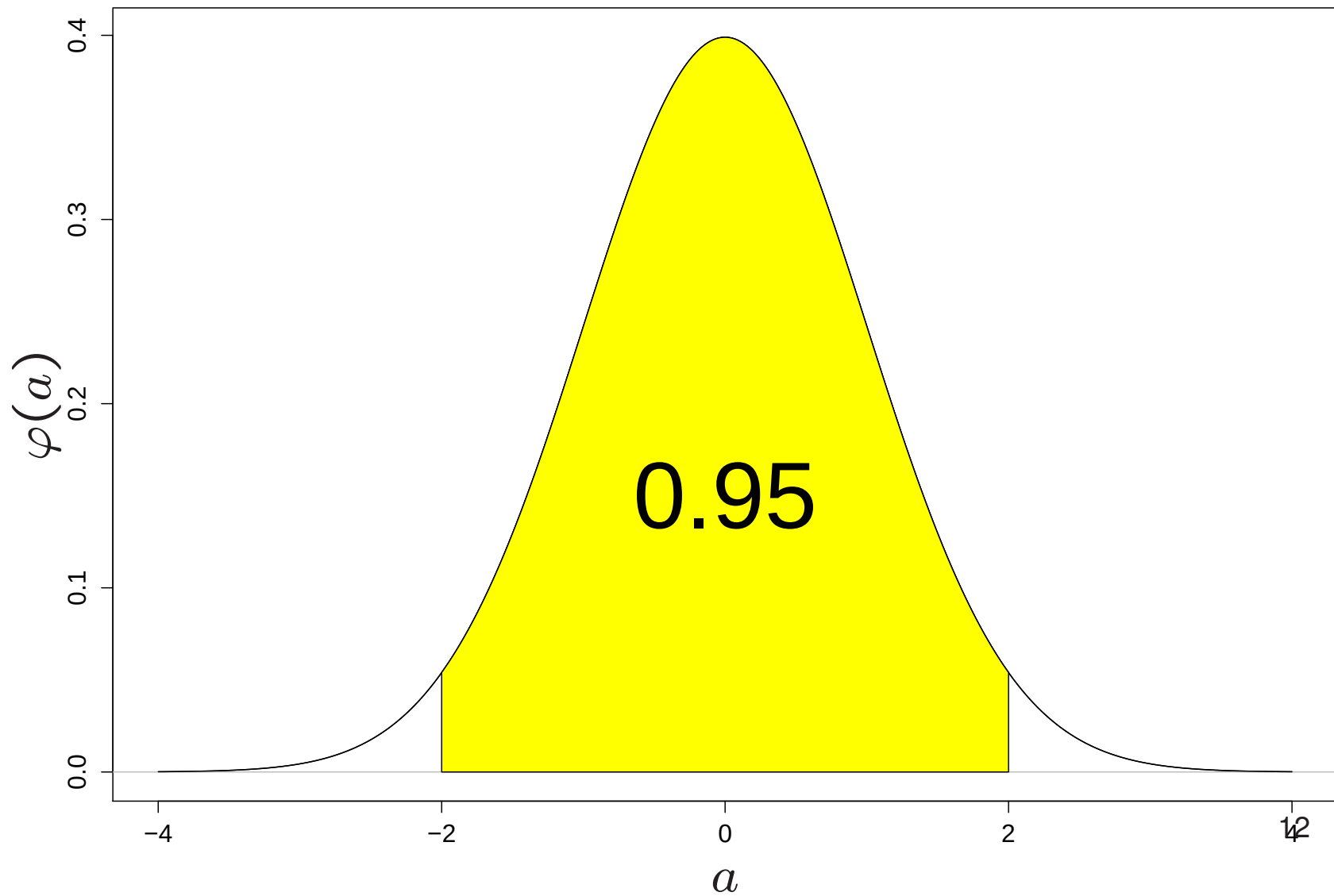
$$\mathbf{P}(|Z| < 1) \approx 0.68$$



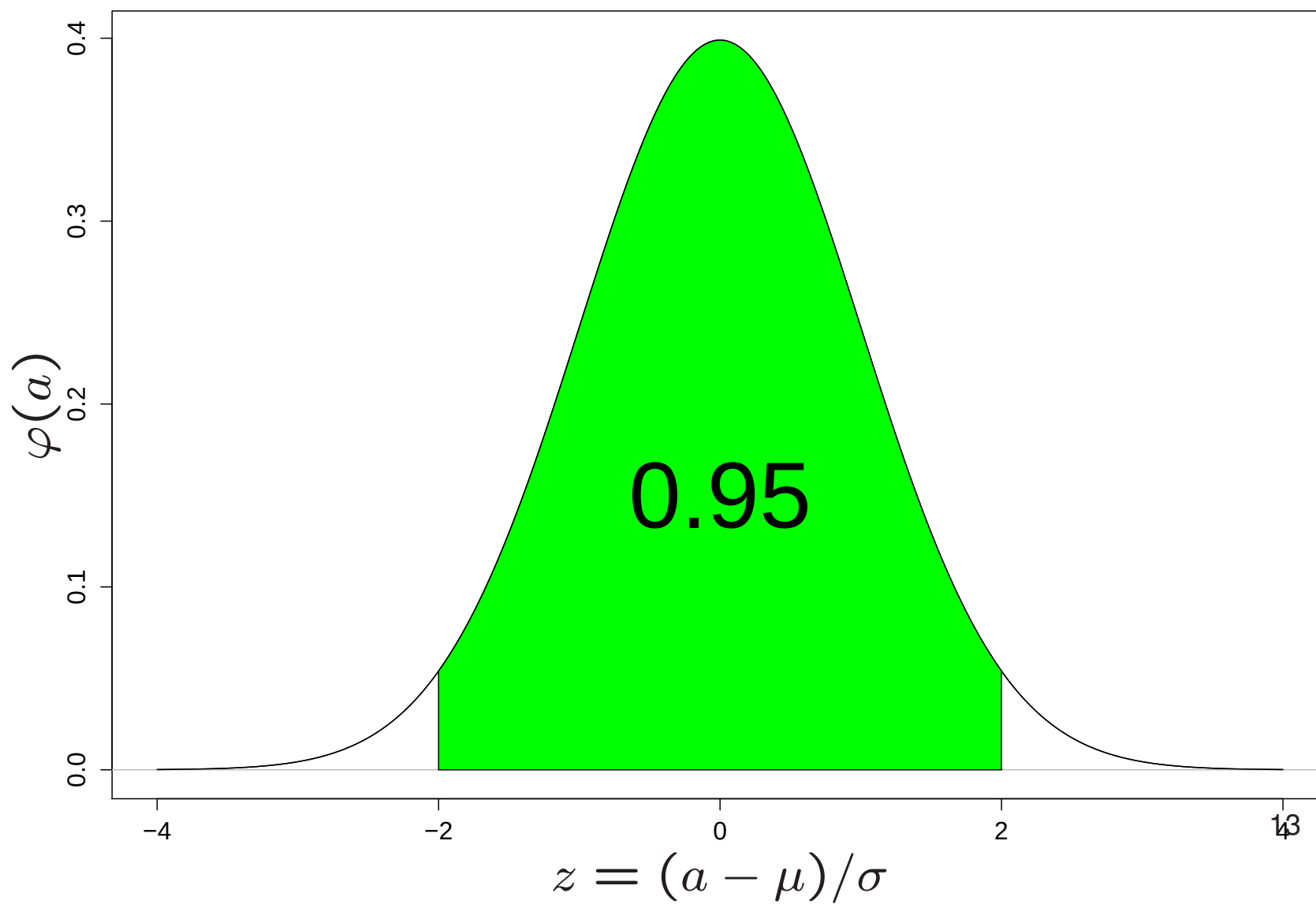
$$P(|Z| < 2) \approx 0.95$$



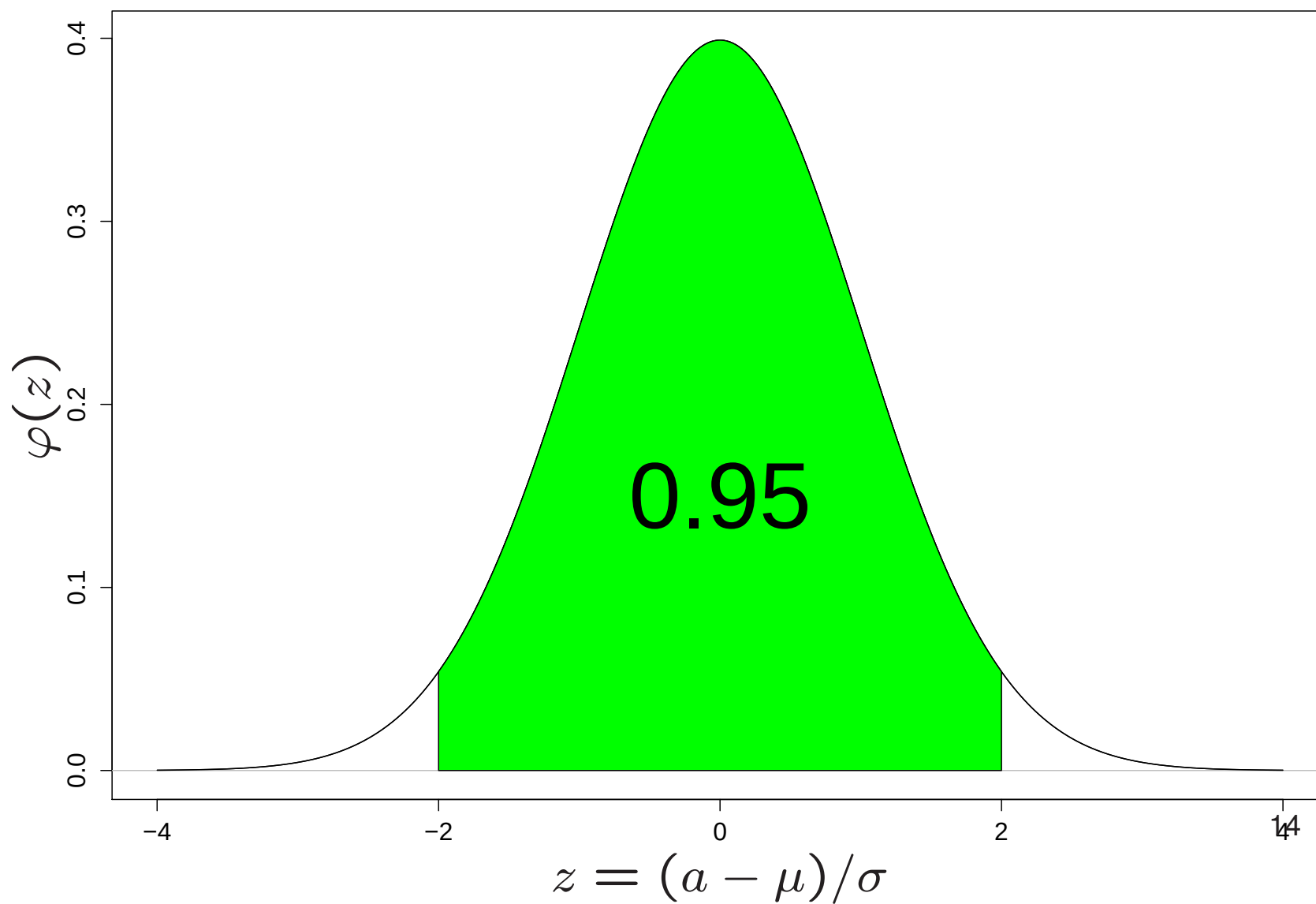
Und für $N(\mu, \sigma)$ -verteilte Zufallsgrößen X ?

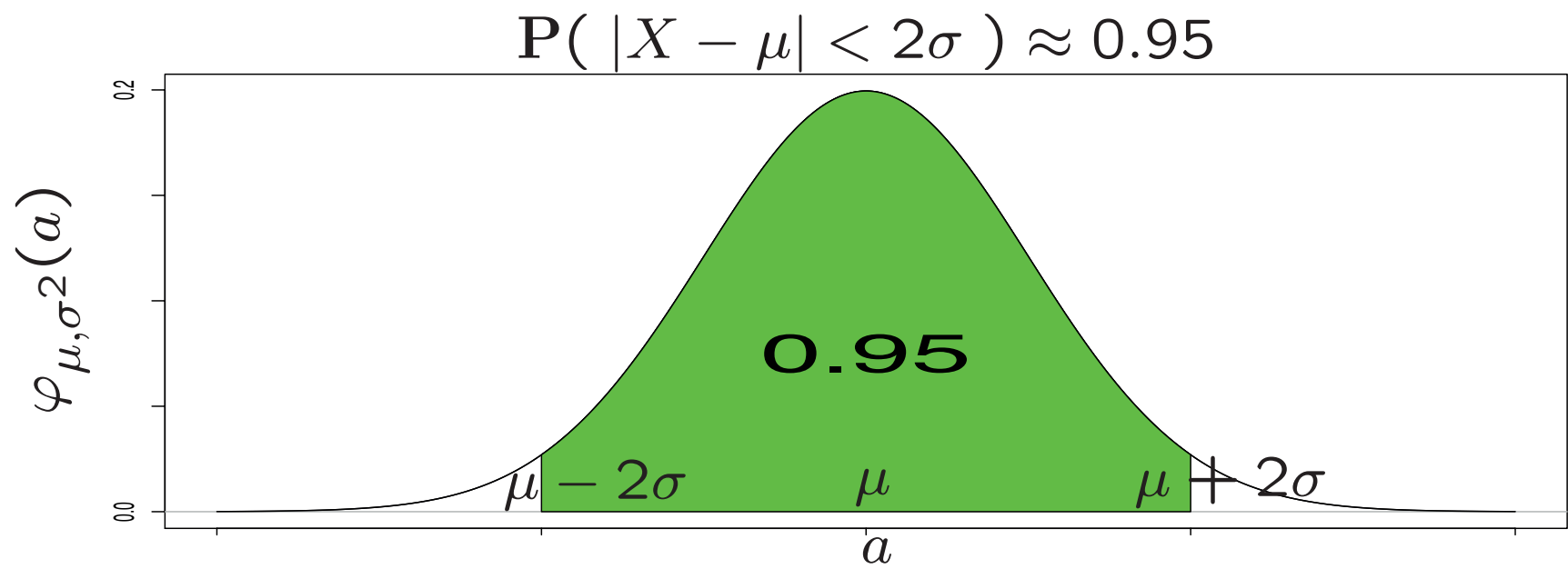


Dasselbe in grün.

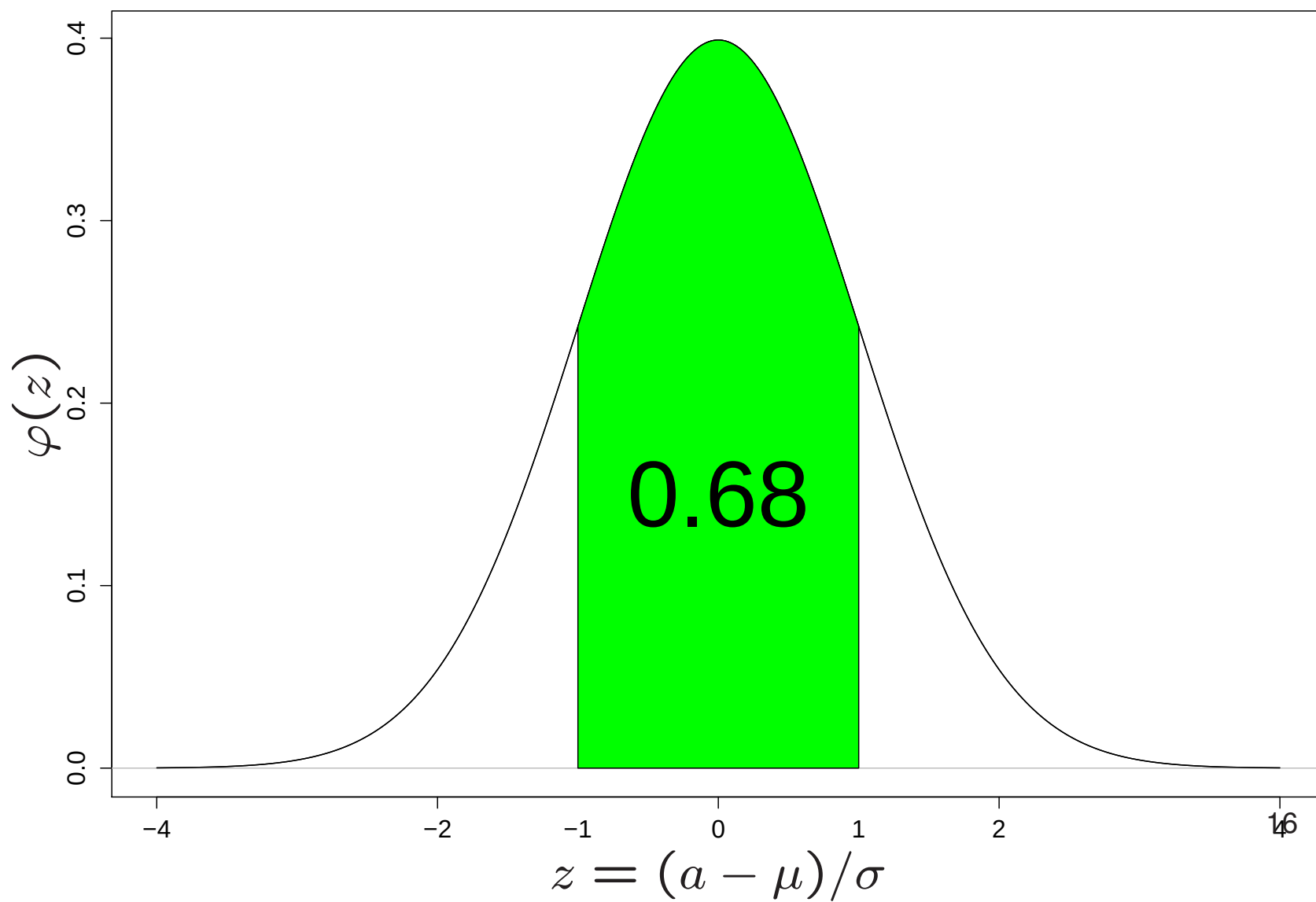


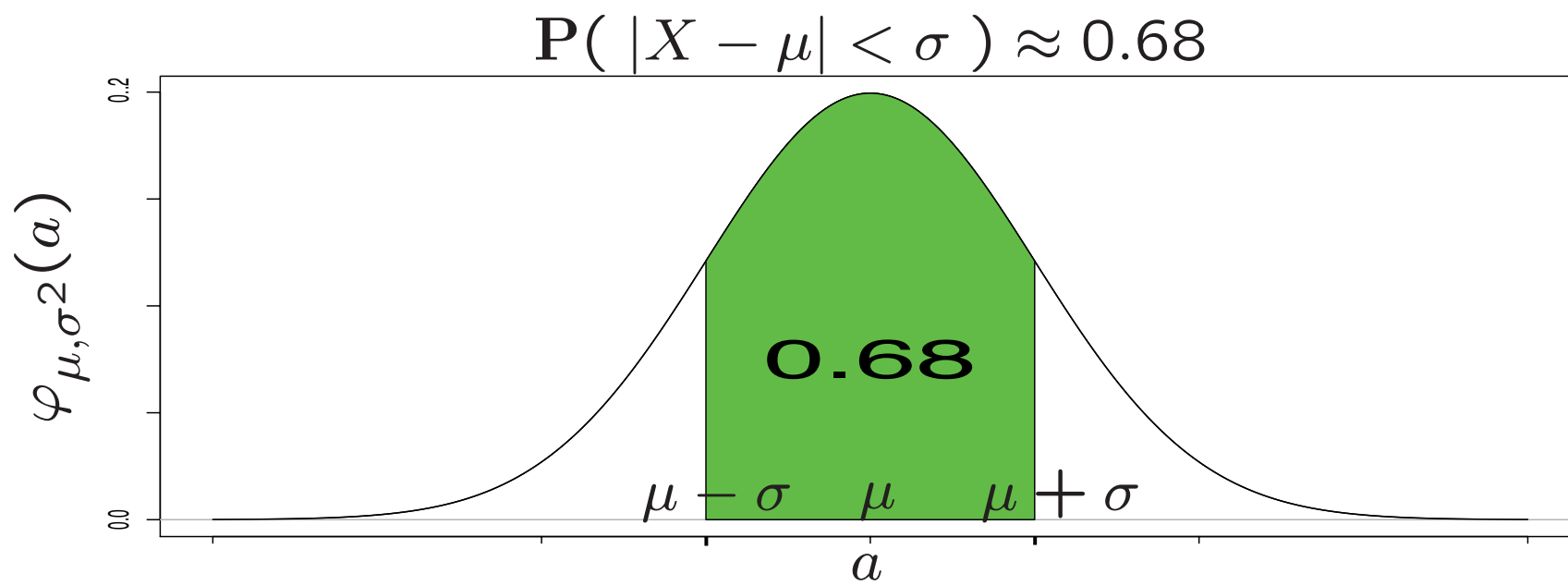
$$\mathbf{P}(|X - \mu|/\sigma < 2) \approx 0.95$$





$$\mathbf{P}(|X - \mu|/\sigma < 1) \approx 0.68$$





Approximation der Gewichte der Binomialverteilung
durch die Dichtefunktion der Normalverteilung

für große $\mu := np$ und $\sigma := \sqrt{npq}$:

Sei X_n Binomial(n, p)-verteilt. Dann gilt (siehe Teil 1):

$$\mathbf{P}(X_n = k) \approx \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) = \varphi_{\mu, \sigma^2}(k) \approx \int_{k - \frac{1}{2}}^{k + \frac{1}{2}} \varphi_{\mu, \sigma^2}(a) da$$

$$\approx \mathbf{P}(k - \tfrac{1}{2} \leq X \leq k + \tfrac{1}{2})$$

für eine $N(\mu, \sigma^2)$ -verteilte ZV'e X .

Vorlesung 6b

Die Normalverteilung

Teil 3

Intermezzo: Dichten von Produktform

(Buch S. 69/70)

Für diskrete Zufallsvariable
war die Unabhängigkeit von X_1 und X_2
äquivalent zur Produktform der gemeinsamen Gewichte:

$$\rho(a_1, a_2) = \rho_1(a_1) \rho_2(a_2).$$

Für Zufallsvariable mit Dichten
ist die Unabhängigkeit von X_1 und X_2
äquivalent zur Produktform der gemeinsamen Dichte:

$$f(a_1, a_2) da_1 da_2 = f_1(a_1) da_1 f_2(a_2) da_2$$

Allgemeiner gilt der
Satz über die Unabhängigkeit von Zv'en mit Dichten

$X_1, \dots, X_n$ seien reellwertige Zufallsvariable,
 $f_1, \dots, f_n$ seien Dichtefunktionen.

Dann sind äquivalent:

(i) $X_1, \dots, X_n$ sind unabhängig,
und X_i hat die Dichte $f_i(a_i) da_i$, $i = 1, \dots, n$.

(ii) $(X_1, \dots, X_n)$ hat die Dichte

$$f_1(a_1) \cdots f_n(a_n) da_1 \dots da_n$$

Beispiel:

Die uniforme Verteilung
auf dem Einheitsquadrat

X_1, X_2 seien unabhängig und uniform verteilt auf $[0, 1]$.

Dann hat (X_1, X_2) die Dichte

$$1_{[0,1]}(a_1) da_1 \cdot 1_{[0,1]}(a_2) da_2$$

$$= 1_{[0,1] \times [0,1]}(a_1, a_2) da_1 da_2,$$

und ist somit uniform verteilt auf $[0, 1] \times [0, 1]$.

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 1:

Die Standardnormalverteilung auf $\mathbb{R}^2$

(vgl. Buch S. 71)

Zur Erinnerung:

Eine $\mathbb{R}$ -wertige Zufallsvariable Z mit Dichte

$$\varphi(a) da := \frac{1}{\sqrt{2\pi}} e^{-a^2/2} da$$

heißt **standard-normalverteilt** (auf $\mathbb{R}^1$).

Wichtige Beobachtung:

Z_1, Z_2 seien standard-normalverteilt und unabhängig.

(Z_1, Z_2) hat dann die Dichte

$$\varphi(a_1) da_1 \varphi(a_2) da_2$$

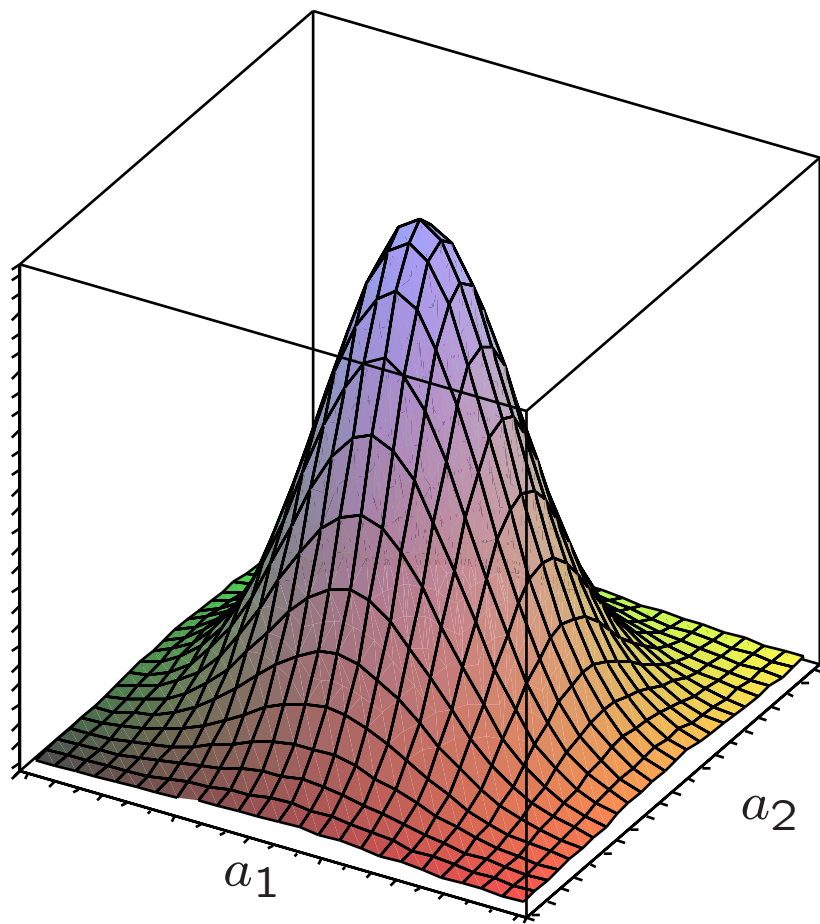
$$= \frac{1}{\sqrt{2\pi}} e^{-a_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-a_2^2/2} da_1 da_2$$

$$= \frac{1}{2\pi} e^{-|a|^2/2} da, \quad a = (a_1, a_2) \in \mathbb{R}^2, \quad |a|^2 := a_1^2 + a_2^2.$$

Die Dichte ist rotationssymmetrisch!

$$f(a_1, a_2) =$$

$$\frac{1}{2\pi} e^{-(a_1^2 + a_2^2)/2}$$



Definition:

Eine $\mathbb{R}^2$ -wertige Zufallsvariable Z mit Dichte

$$f(a) da = \frac{1}{2\pi} e^{-|a|^2/2} da, \quad a \in \mathbb{R}^2,$$

heißt **standard-normalverteilt auf $\mathbb{R}^2$** .

Aufgabe:

Berechne $\mathbf{P}(|Z|^2 > b)$

für ein auf $\mathbb{R}^2$ standard-normalverteiltes Z
und $b > 0$.

Anders formuliert:

Berechne $\mathbf{P}(Z \in A)$

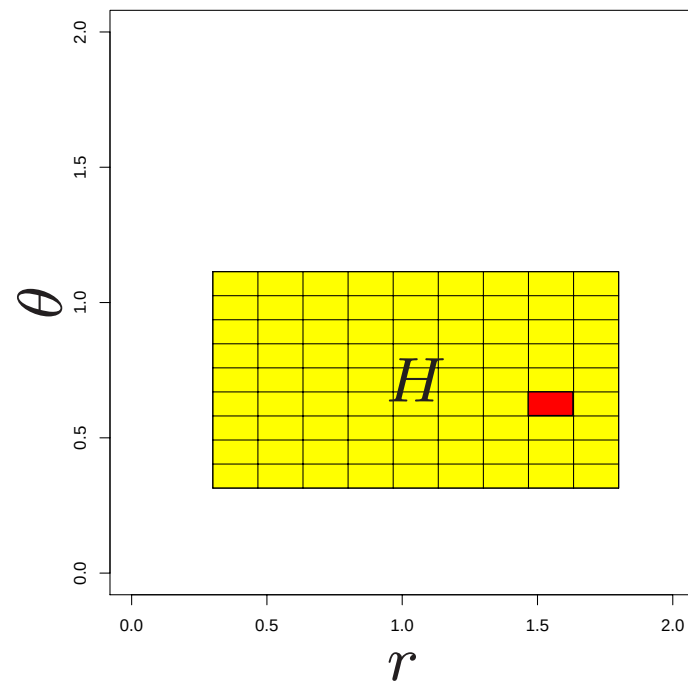
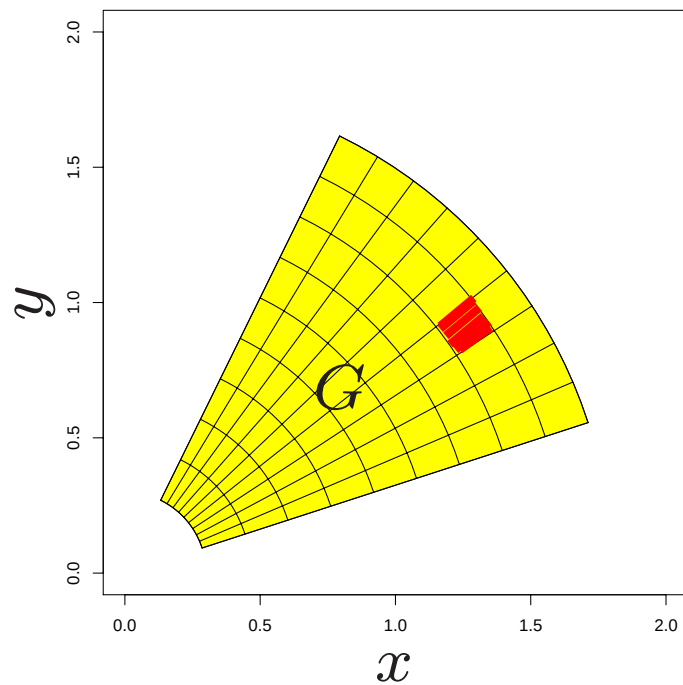
mit $A := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > b\}$

Zu berechnen ist also

$$\frac{1}{2\pi} \iint_A e^{-(x^2+y^2)/2} dx dy$$

Zur Illustration der Polarkoordinatentransformation:

$$\iint_G e^{-(x^2+y^2)/2} dx dy = \iint_H e^{-r^2/2} r dr d\theta$$



Speziell mit $G := A$ und $H := (\sqrt{b}, \infty) \times [0, 2\pi)$:

$$\begin{aligned} \iint_{\{(x,y): x^2+y^2 > b\}} e^{-(x^2+y^2)/2} dx dy &= \int_0^{2\pi} \int_{\sqrt{b}}^{\infty} e^{-r^2/2} r dr d\theta \\ &= 2\pi(-e^{-r^2/2}) \Big|_{\sqrt{b}}^{\infty} = 2\pi e^{-b/2}. \end{aligned}$$

$$\begin{aligned} \mathbf{P}(|Z|^2 > b) &= \mathbf{P}(Z \in A) \\ &= \frac{1}{2\pi} \iint_A e^{-(x^2+y^2)/2} dx dy = e^{-b/2}. \end{aligned}$$

$|Z|^2$ ist $\text{Exp}(\frac{1}{2})$ -verteilt!

Eine Folgerung aus der Rotationssymmetrie

(vgl Buch S. 71)

Z_1 und Z_2 seien unabhängig und standard-normalverteilt.

Wir wollen einsehen, dass dann auch

$$\frac{Z_1 + Z_2}{\sqrt{2}}$$

standard-normalverteilt ist.

Dafür gibt es ein schönes geometrisches Argument.

Sei $\vec{e}_1, \vec{e}_2$ die Standardbasis in $\mathbb{R}^2$.

Wir fassen das zufällige Zahlenpaar $Z = (Z_1, Z_2)$ auf als die Koordinaten des zufälligen Vektors

$$\vec{Z} = Z_1 \vec{e}_1 + Z_2 \vec{e}_2$$

und stellen fest:

- Z_1 ist die Koordinate von $\vec{Z}$ bgl. des Einheitsvektors $\vec{e}_1$.
 - $\frac{Z_1 + Z_2}{\sqrt{2}}$ ist die Koordinate von $\vec{Z}$ bgl. des Einheitsvektors
- $$\vec{u}_{\text{diag}} := \frac{1}{\sqrt{2}}\vec{e}_1 + \frac{1}{\sqrt{2}}\vec{e}_2.$$
- Die Verteilung von $\vec{Z}$ ist rotationssymmetrisch

Also ist Z_1 so verteilt wie $\frac{Z_1 + Z_2}{\sqrt{2}}$.

Und was dem Einheitsvektor $\vec{u}_{\text{diag}}$ recht ist,
ist jedem Einheitsvektor $\vec{u} = \tau_1\vec{e}_1 + \tau_2\vec{e}_2$ billig !

Anders gesagt:

Sind Z_1, Z_2 unabhängig und $N(0, 1)$ -verteilt,
dann gilt für jedes Zahlenpaar (τ_1, τ_2) mit $\tau_1^2 + \tau_2^2 = 1$:

$Y := \tau_1 Z_1 + \tau_2 Z_2$ ist $N(0, 1)$ -verteilt.

(Denn Y ist die Koordinate von $\vec{Z} = Z_1 \vec{e}_1 + Z_2 \vec{e}_2$
zum Einheitsvektor $\vec{u} := \tau_1 \vec{e}_1 + \tau_2 \vec{e}_2$.)

Eine wichtige Folgerung hieraus:

Die Summe von
unabhängigen, normalverteilten Zufallsvariablen
ist wieder normalverteilt.

Denn für Z_1, Z_2 unabhängig und $N(0, 1)$ -verteilt gilt:

$$\sigma_1 Z_1 + \sigma_2 Z_2 = \sqrt{\sigma_1^2 + \sigma_2^2} \left(\frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_2 \right)$$

ist $N(0, \sigma_1^2 + \sigma_2^2)$ -verteilt.

und

$$(\sigma_1 Z_1 + \mu_1) + (\sigma_2 Z_2 + \mu_2) = (\sigma_1 Z_1 + \sigma_2 Z_2) + (\mu_1 + \mu_2).$$

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 2:

Die Standardnormalverteilung auf $\mathbb{R}^n$

(vgl. Buch S. 71)

In Teil 3 der Vorlesung 6b formulierten wir den
Satz über die Unabhängigkeit von ZV'en mit Dichten:

$X_1, \dots, X_n$ seien reellwertige Zufallsvariable,
 $f_1, \dots, f_n$ seien Dichtefunktionen.

Dann sind äquivalent:

(i) $X_1, \dots, X_n$ sind unabhängig,
und X_i hat die Dichte $f_i(a_i) da_i$, $i = 1, \dots, n$.

(ii) $(X_1, \dots, X_n)$ hat die Dichte

$$f_1(a_1) \cdots f_n(a_n) da_1 \dots da_n$$

Dazu passt die folgende Situation:

Sei $Z := (Z_1, \dots, Z_n)$. Dann gilt:

$Z_1, \dots, Z_n$ sind unabhängig und $N(0, 1)$ -verteilt

$\Longleftrightarrow$

$$\mathbf{P}(Z \in da) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{|a|^2}{2}\right) da, \quad a \in \mathbb{R}^n,$$

$$\text{mit } |a|^2 := a_1^2 + \dots + a_n^2.$$

Z heißt dann *standard-normalverteilt auf* $\mathbb{R}^n$.

Die Dichte der Standard-Normalverteilung auf $\mathbb{R}^n$
ist rotationssymmetrisch.

Analog zum Fall $n = 2$ gilt deshalb:

Ist $Z = (Z_1, \dots, Z_n)$ standard-normalverteilt auf $\mathbb{R}^n$

dann ist $Y := \frac{Z_1 + \dots + Z_n}{\sqrt{n}}$ $N(0, 1)$ -verteilt.

Denn Y ist die Koordinate von $\vec{Z} = Z_1\vec{e}_1 + \dots + Z_n\vec{e}_n$
zum Einheitsvektor $\frac{1}{\sqrt{n}}(1, \dots, 1)$.

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 3

Die Botschaft des Zentralen Grenzwertsatzes

(Buch S. 77)

Zur Erinnerung:

$(Z_1, \dots, Z_n)$ heißt standard-normalverteilt im $\mathbb{R}^n$

$:\Leftrightarrow$

$Z_1, \dots, Z_n$ sind unabhängig und $N(0,1)$ -verteilt

Auch $\frac{Z_1 + \dots + Z_n}{\sqrt{n}}$ ist dann $N(0,1)$ -verteilt!

Der Zentrale Grenzwertsatz liefert
eine gewaltige Weiterung der vorigen Aussage
(asymptotisch für große n):

Zentraler Grenzwertsatz (Version (0,1))

salopp formuliert:

“Die durch $\sqrt{n}$ geteilte Summe von
 n unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 $\mathbb{R}$ -wertigen Zufallsvariablen
mit Erwartungswert 0 und Varianz 1
ist für große n annähernd standard-normalverteilt”

Satz (ZGWS, “Version (0,1)”)

Seien $X_1, X_2, \dots$
unabhängige und identisch verteilte Zufallsvariable
mit Erwartungswert 0 und Varianz 1.

Dann gilt für alle $c < d \in \mathbb{R}$

$$\mathbf{P}\left(\frac{X_1 + \dots + X_n}{\sqrt{n}} \in [c, d]\right) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z \in [c, d]).$$

Dabei ist Z standard-normalverteilt.

Jetzt: (μ, σ^2) statt $(0, 1)$:

Hat X den Erwartungswert μ und die Varianz σ^2 ,
dann hat $\frac{X - \mu}{\sigma}$ Erwartungswert 0 und Varianz 1.

Man nennt $\frac{X - \mu}{\sigma}$ auch die *Standardisierung* von X .

Sind $X_1, X_2, \dots$ unabhängige Kopien von X ,
dann ergibt der ZGWS (Version $(0,1)$)
angewandt auf die Standardisierungen der X_i :

Zentraler Grenzwertsatz (Klassische Version)

Seien $X_1, X_2, \dots$

unabhängige und identisch verteilte Zufallsvariable mit endlichem Erwartungswert μ und endlicher Varianz $\sigma^2 > 0$.

Dann gilt für alle $c < d \in \mathbb{R}$

$$\mathbf{P}\left(\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} \in [c, d]\right) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z \in [c, d]).$$

Dabei ist Z standard-normalverteilt.

Der Münzwurf passt in den Rahmen des klassischen ZGWS:

(X_i) sei eine Bernoulli-Folge zum Parameter p ,

also insbesondere:

$X_1, X_2, \dots$ unabhängig und identisch verteilt mit

$$\mathbf{E}[X_i] = \mathbf{P}(X_i = 1) = p, \quad \mathbf{Var}[X_i] = pq$$

Dann gilt für alle $c < d \in \mathbb{R}$:

$$\mathbf{P} \left(\frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \in [c, d] \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_c^d e^{-a^2/2} da.$$

Das ist der Satz von de Moivre (1733, für $p = 1/2$)

und Laplace (1812, für allgemeines p).

Der klassische Zentrale Grenzwertsatz in anschaulichen Worten:

“Die standardisierte Summe von VIELEN
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 $\mathbb{R}$ -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Manchmal noch praktischer so:

“Die standardisierte Summe von VIELEN
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 $\mathbb{R}$ -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Beispiel:

$X_1, X_2, \dots$ seien unabhängig und identisch verteilt
mit Erwartungswert μ und Varianz σ^2

Dann ist für großes n die Zufallsvariable

$$M_n := \frac{X_1 + \dots + X_n}{n}$$

approximativ $N(\mu, \frac{\sigma^2}{n})$ -verteilt.

Spezialfall des vorigen Beispiels:

Ziehen mit Zurücklegen

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

$$\mu := \frac{1}{g} \sum_{a \in S} h(a) \quad \dots \text{Populationsmittelwert}$$

$$\sigma^2 := \frac{1}{g} \sum_{a \in S} (h(a) - \mu)^2 \quad \dots \text{Populationsvarianz}$$

$J_1, J_2, \dots$ seien unabhängig und uniform auf S verteilt

Die Zufallsvariablen $X_i := h(J_i)$, $i = 1, 2, \dots$

passen dann in den Rahmen des vorigen Beispiels.

“Die standardisierte Summe von VIELEN
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 $\mathbb{R}$ -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Diese Aussage bleibt auch
unter schwächeren Bedingungen bestehen,
sowohl was die Unabhängigkeit,
als auch was die identische Verteiltheit betrifft.

Eine Botschaft zum Mitnehmen ins Leben
(salopp formuliert):

“ Die Summe von vielen
annähernd unabhängigen Zufallsvariablen,
die nicht notwendig identisch verteilt, aber
ungefähr von derselben Größenordnung sind,
ist annähernd normalverteilt.”

Beispiel: **Ziehen ohne Zurücklegen.**

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

$J_1, J_2, \dots$ rein zufälliges Ziehen **ohne** Zurücklegen aus S .

Dann ist für großes n und großes $g - n$

$$M_n := \frac{1}{n} \sum_{i=1}^n h(J_i)$$

annähernd normalverteilt.

Beispiel: **Ziehen ohne Zurücklegen.**

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

$J_1, J_2, \dots$ rein zufälliges Ziehen **ohne** Zurücklegen aus S .

Dann ist für großes n und großes $g - n$

$$M_n := \frac{1}{n} \sum_{i=1}^n h(J_i)$$

annähernd normalverteilt.

(Zur Berechnung von $\mathbf{Var}[M_n]$ aus der Populationsvarianz
siehe ÜA 24 + ein Übungsbeispiel auf dem Weihnachtsblatt.)

Zentraler Grenzwertsatz: Meilensteine in seiner Geschichte

Abraham de Moivre:

Der faire Münzwurf (1733)

Pierre-Simon Laplace:

Allgemeine binomiale Zufallsgrößen (1812)

Pafnuty Lvovich **Chebyshev**:

Skizze eines Beweises für den allgemeinen Fall (1887)

Aleksandr Mikhailovich **Lyapunov**:

“Klassischer” zentraler Grenzwertsatz (1901)

Noch allgemeiner (1906)

Andrei Andreyevich **Markov**:

weitere Verallgemeinerungen (~ 1910)

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 4

Erlebnis:

Via Monte Carlo zur Gaußschen Glockenkurve

Nehmen wir an,
die am Ende von Teil 3 genannten Herren hätten sich
auf ihre vielen anderen Interessen
beschränkt.

ZENTRALER GRENZWERTSATZ

Unbekannt.

Könnten wir ihn entdecken?

Wie kämen wir auf die “Glockenkurve”?

Warum gerade $e^{-x^2/2}$?



e

Ein Ausflug mit Brooks Ferebee

Ein Beispiel: Summen von
unabhängigen **uniform** verteilten Zufallsvariablen

Wir denken an

Rundungsfehler bei Addition

In Wirklichkeit

$$\pi =$$

3.141592653589793238462643383279502884197169399375105...

Im Rechner

$$\pi \leftarrow 3.14159265358979$$

MODELL

Zahl = Rechnerdarstellung + Rundungsfehler.

$$A = a^{[R]} + \varepsilon X \quad \varepsilon = 10^{-15}$$

Annahme: X uniform verteilt auf $[-0.5, 0.5]$.

$$\sum_{i=1}^n A_i = ?$$

$$\sum_{i=1}^n A_i = \sum_{i=1}^n a_i^{[R]} + \varepsilon \sum_{i=1}^n X_i$$

Wie groß ist der Fehler?

$$\sum_{i=1}^n X_i \approx ?$$

Der Zentrale Grenzwertsatz gibt die Auskunft:

$$\sum_{i=1}^n X_i \text{ ist}$$

für große n

approximativ $N(0, n \sigma_{X_1}^2)$ -verteilt.

Ein Beispiel:

$X_1, X_2, \dots$ unabhängig
und uniform auf $[-0.5, 0.5]$ verteilt

Empirische Verteilung von

$$S_n := X_1 + \dots + X_n$$

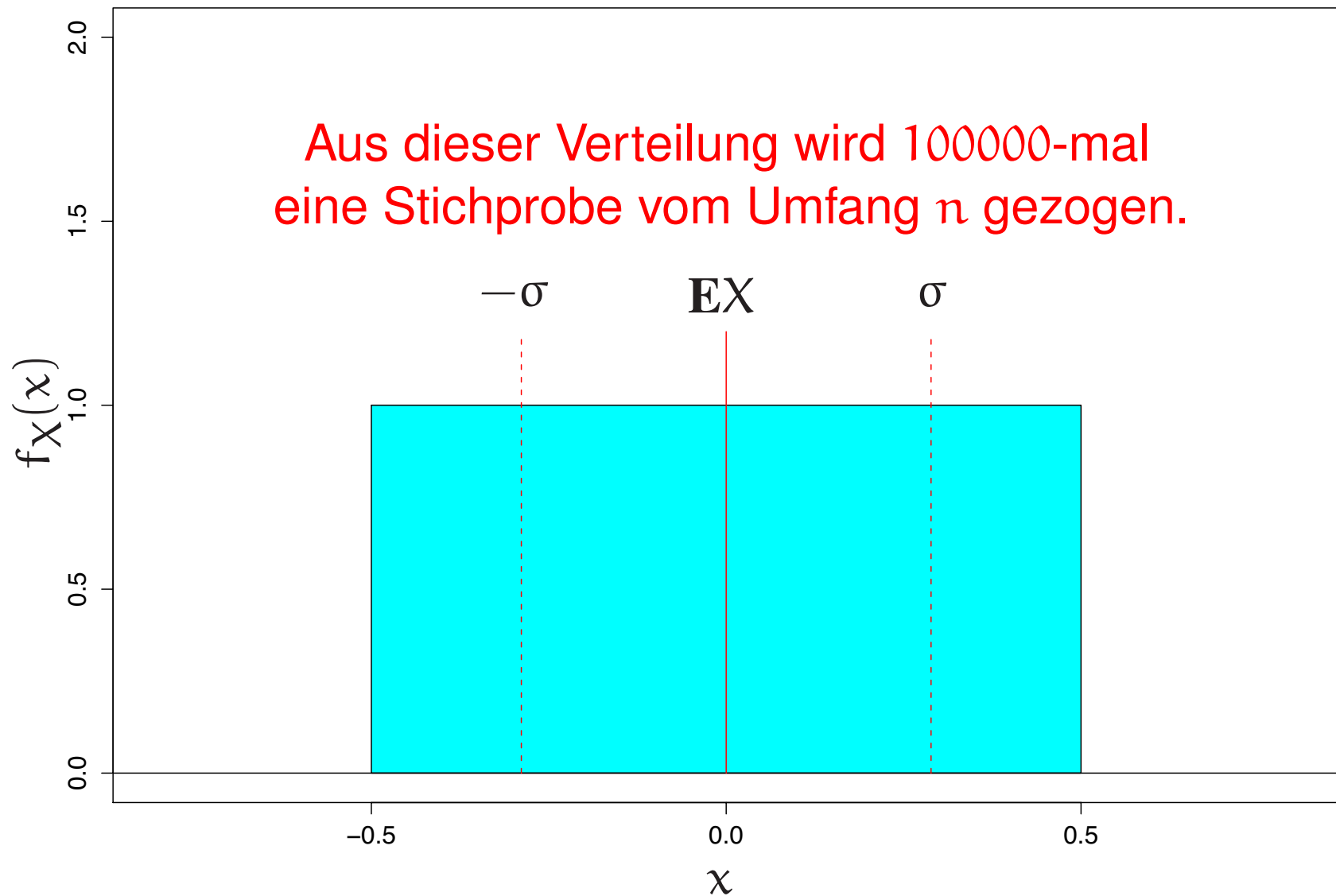
100000 Simulationen

jeweils für

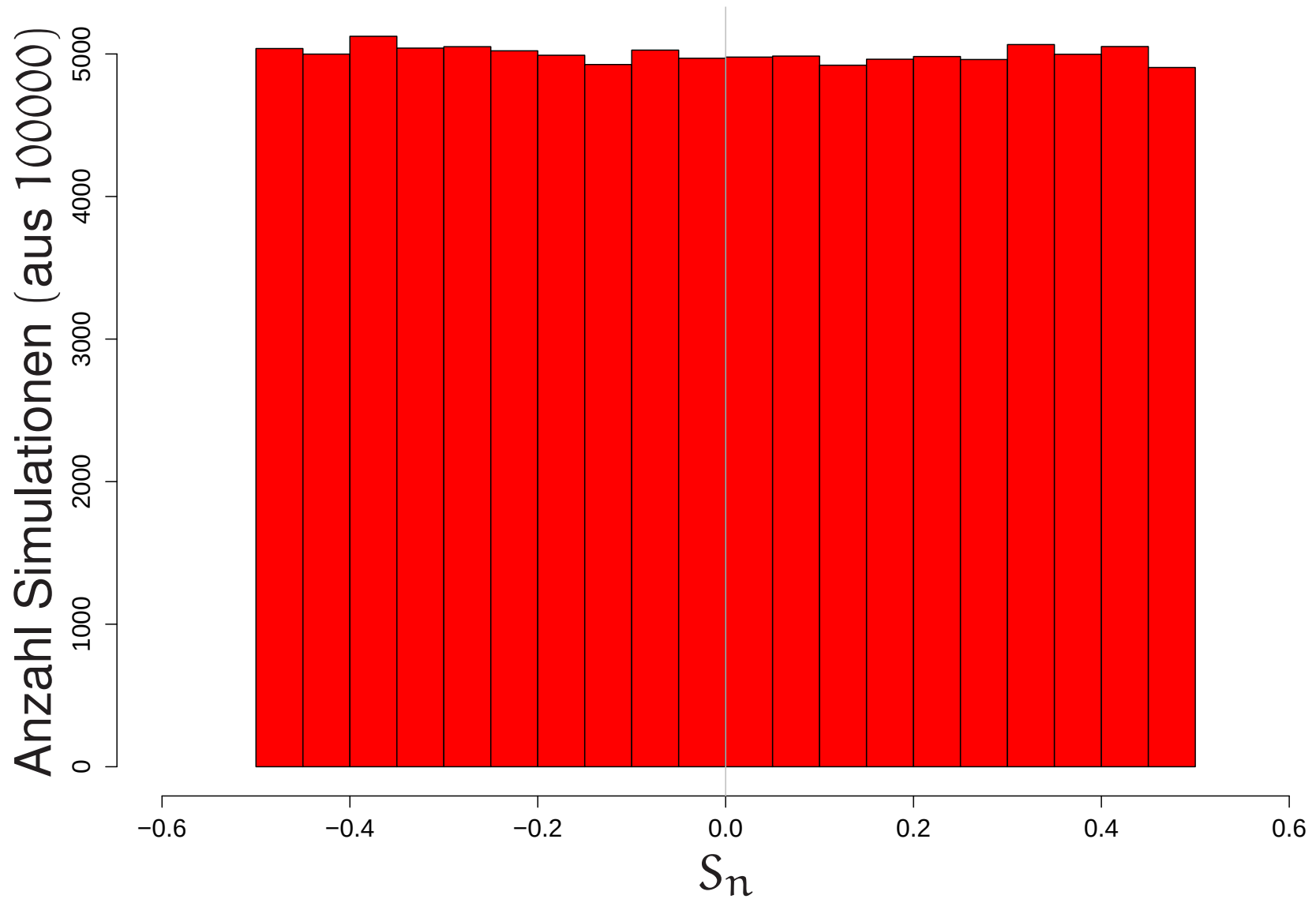
$$n = 1, 2, \dots, 10$$

$$n = 15, 20, 30, \dots, 100$$

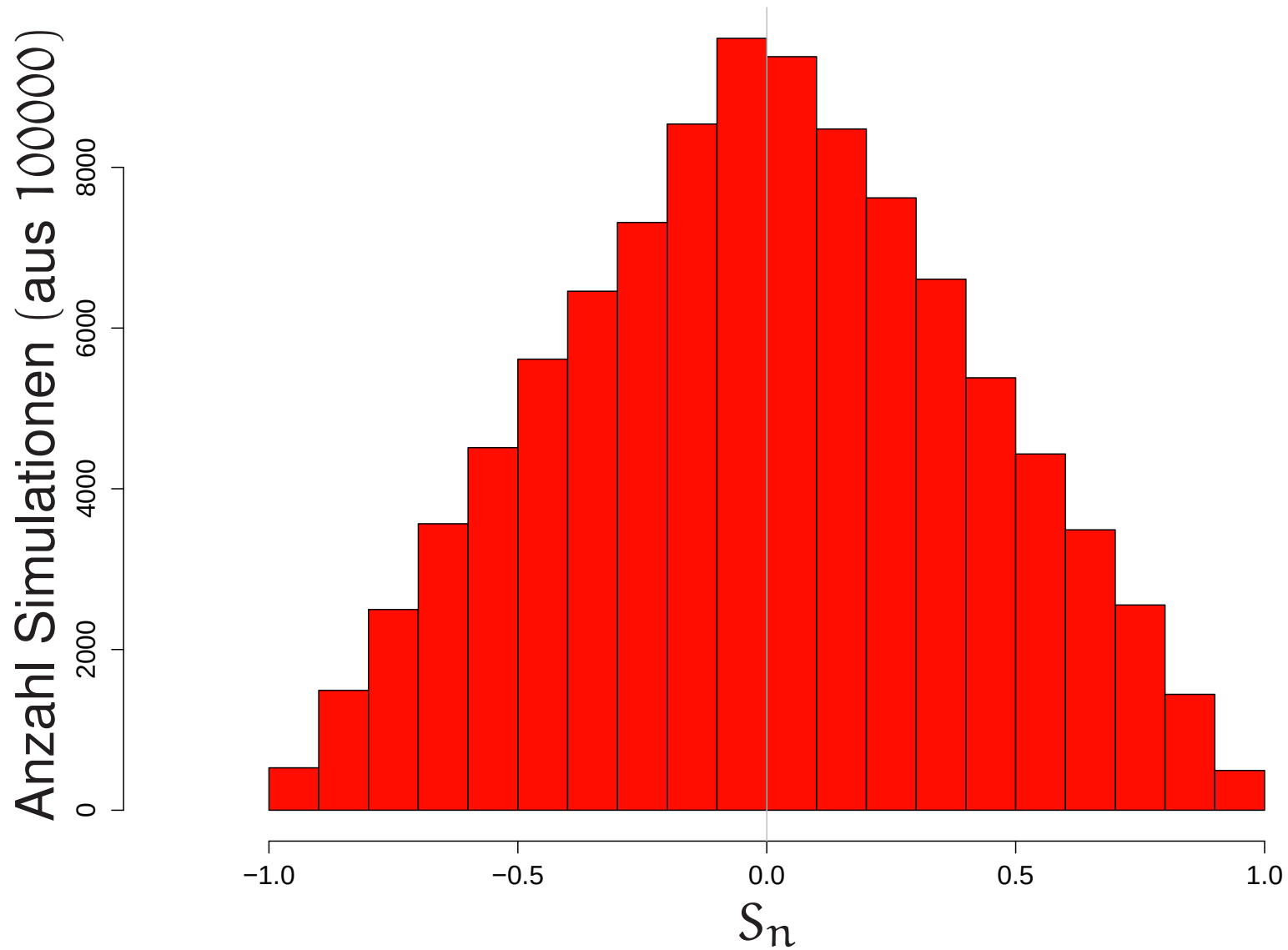
Dichtefunktion f_X der Verteilung von X



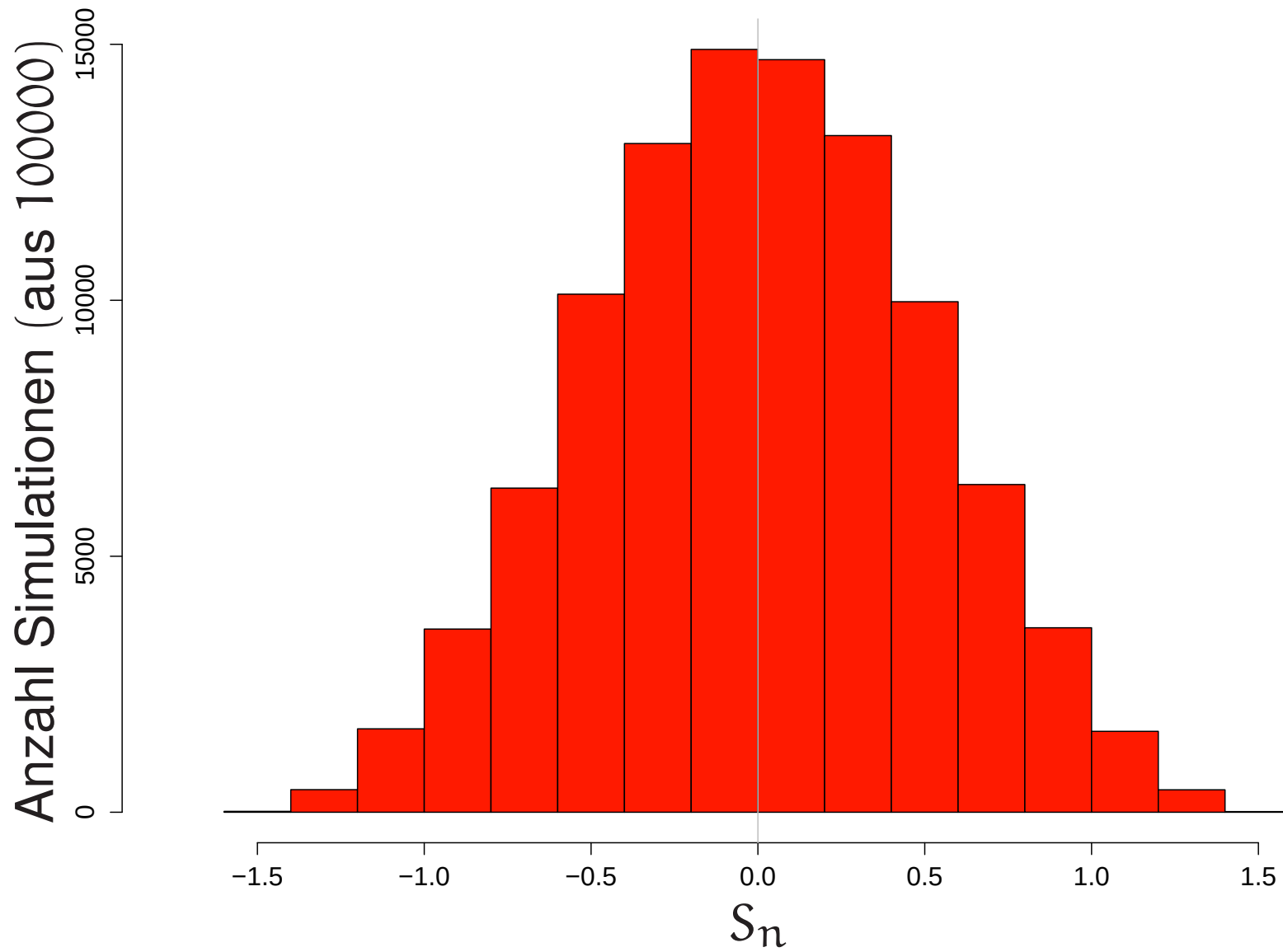
Verteilung von $S_1 = X_1$ ($n = 1$)



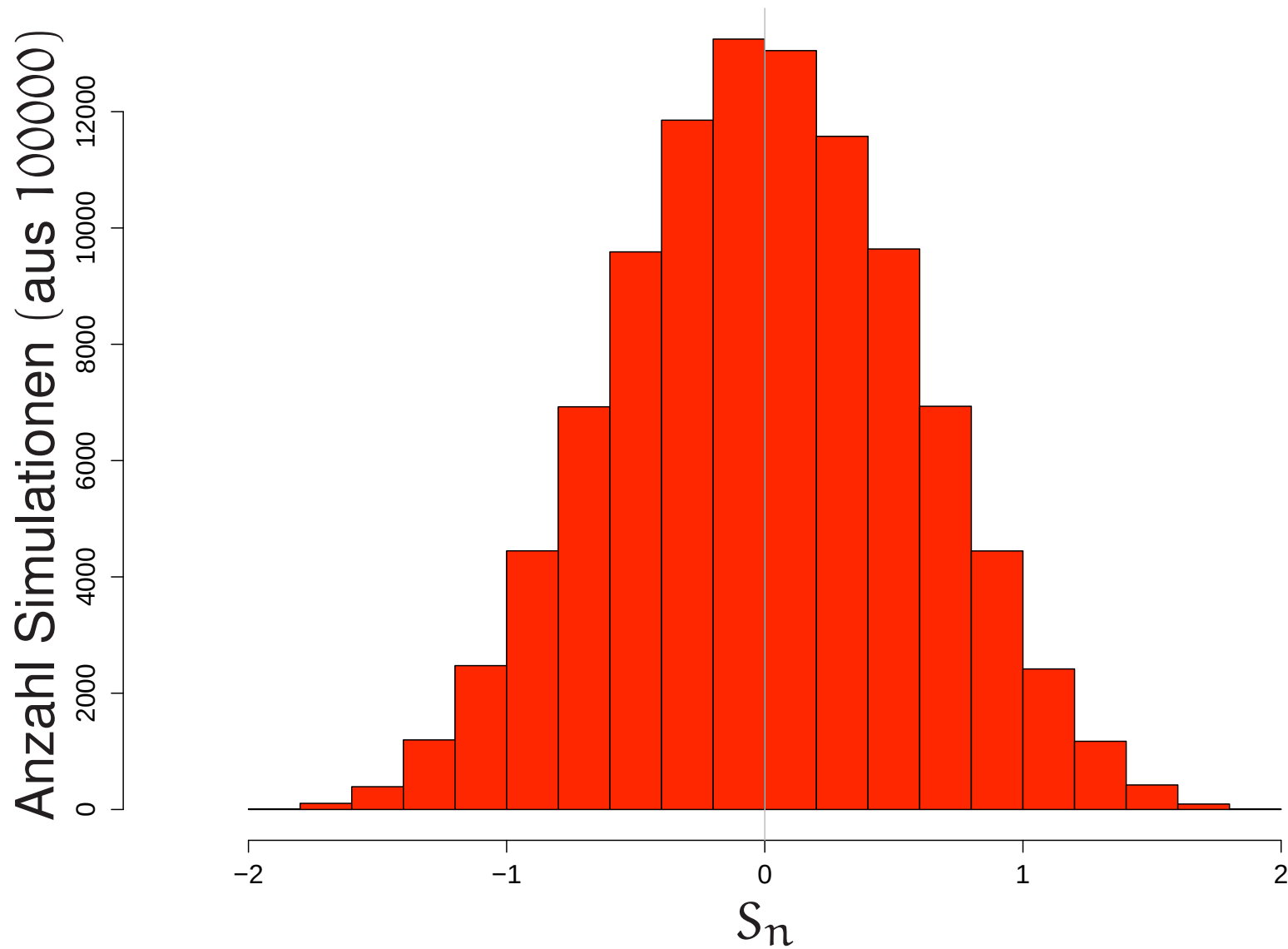
Verteilung von $S_2 = X_1 + X_2$ ($n = 2$)



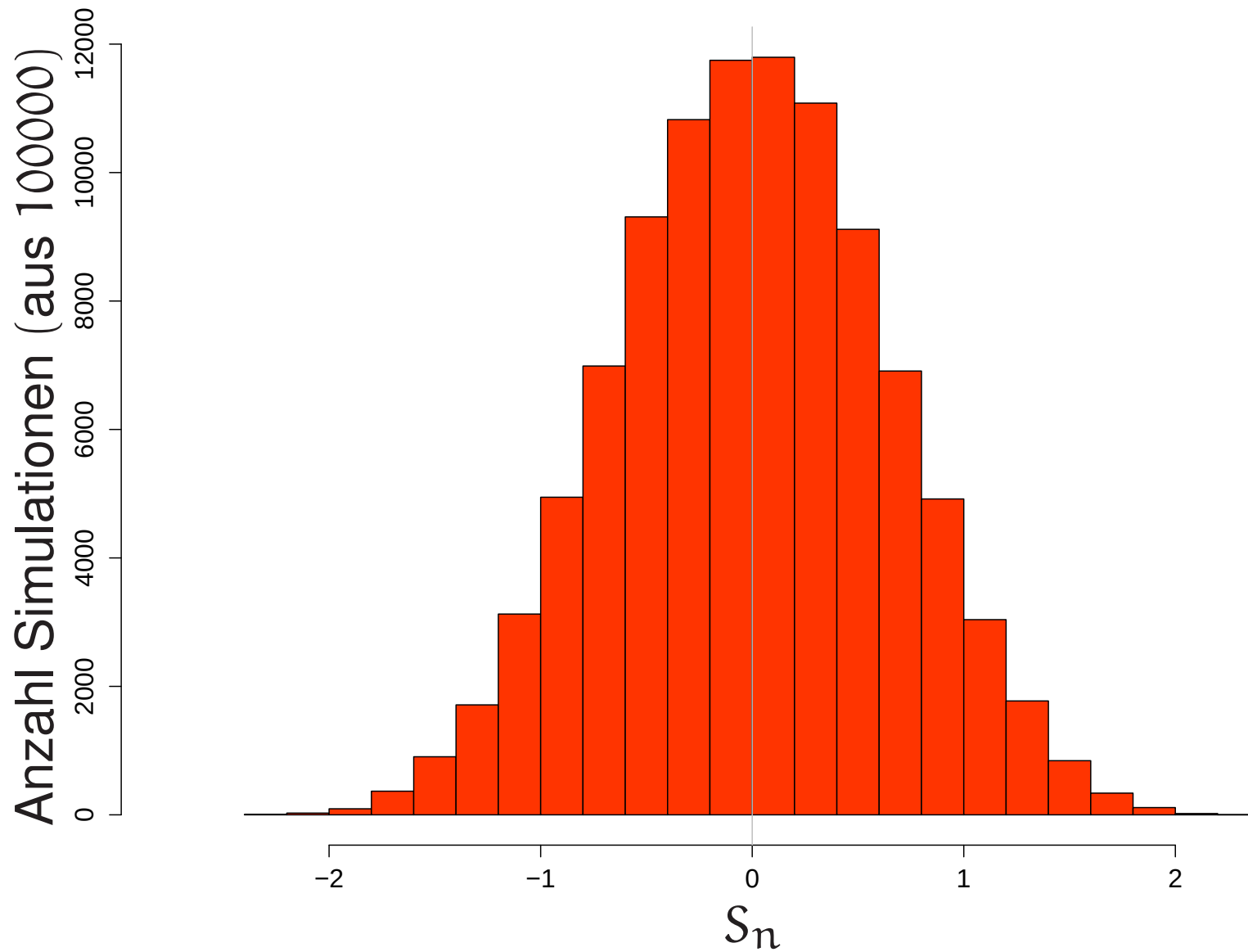
Verteilung von S_n ($n = 3$)



Verteilung von S_n ($n = 4$)



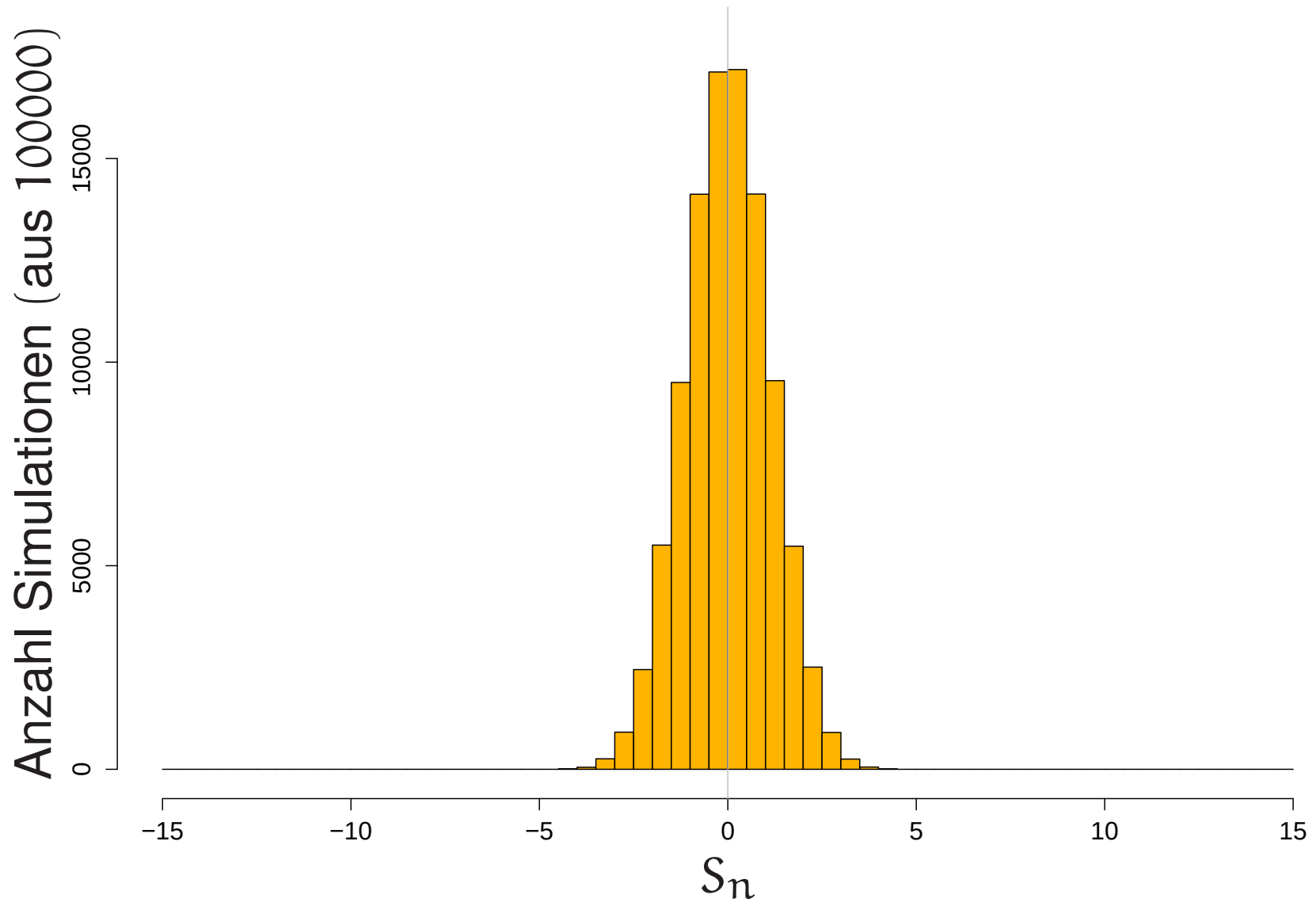
Verteilung von S_n ($n = 5$)



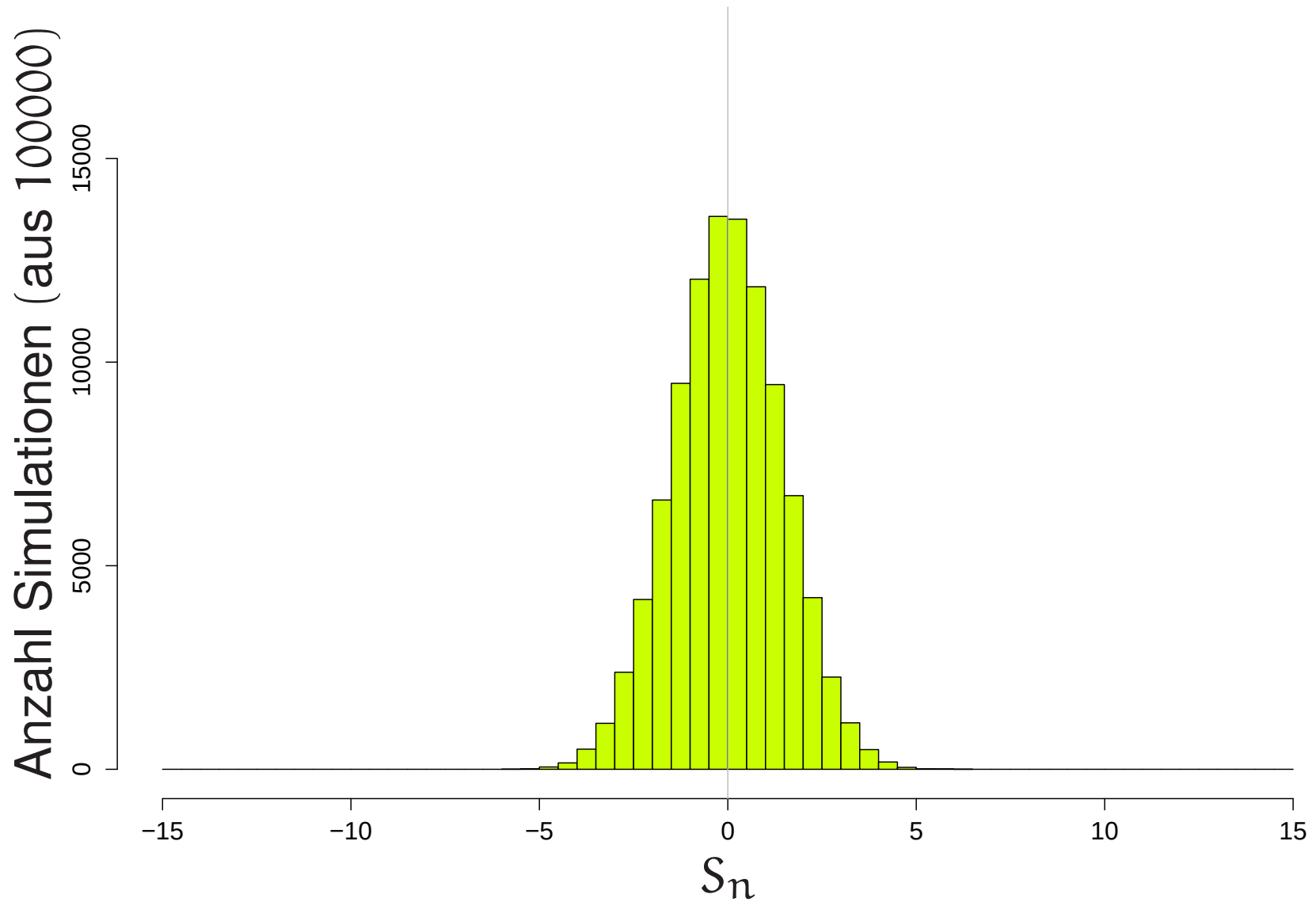
Bisher: dynamische Skalierung der Breite.

Ab jetzt: feste Skalierung der Breite,
mit dem Intervall $[-15, + 15]$

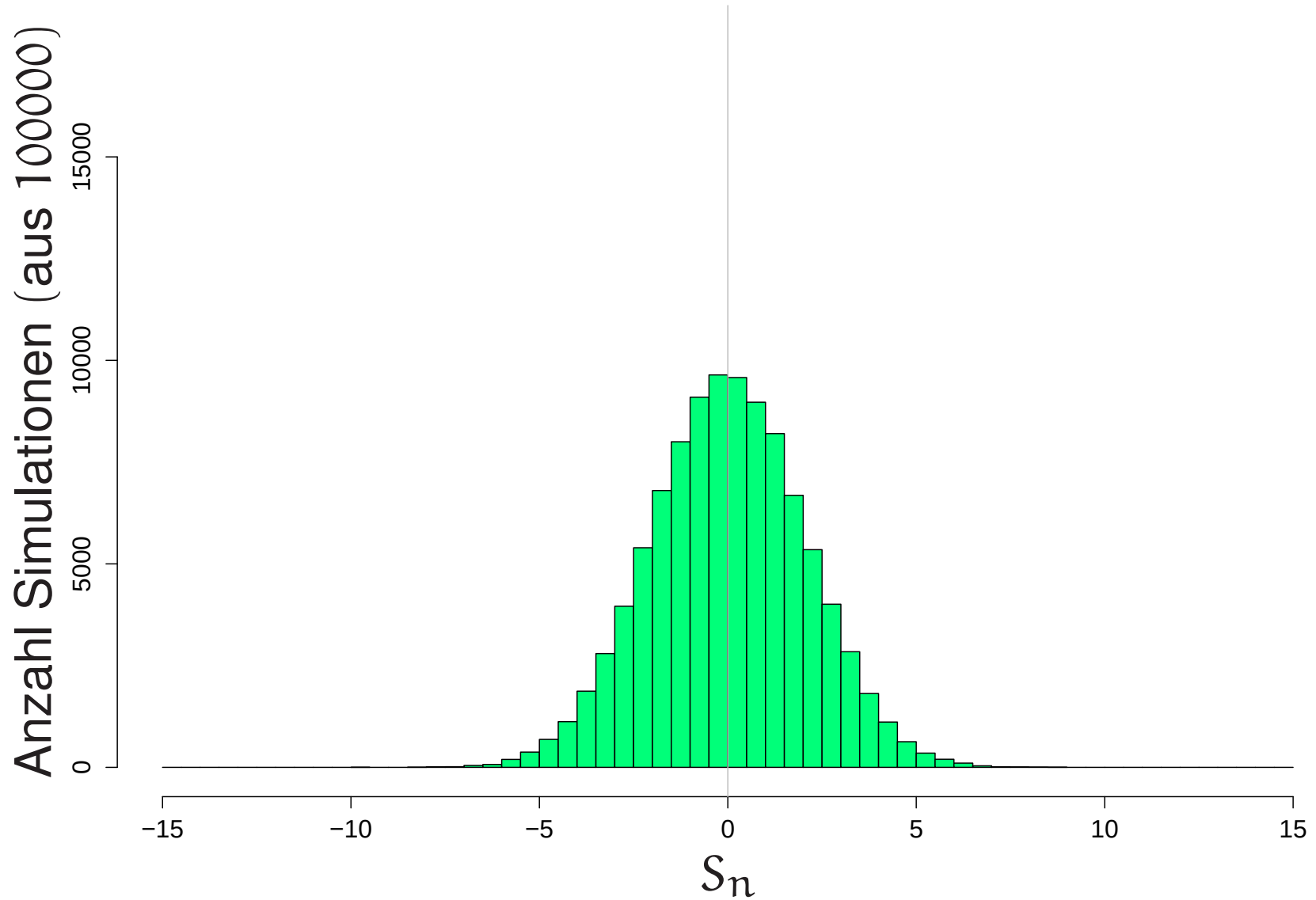
Verteilung von S_n ($n = 15$)



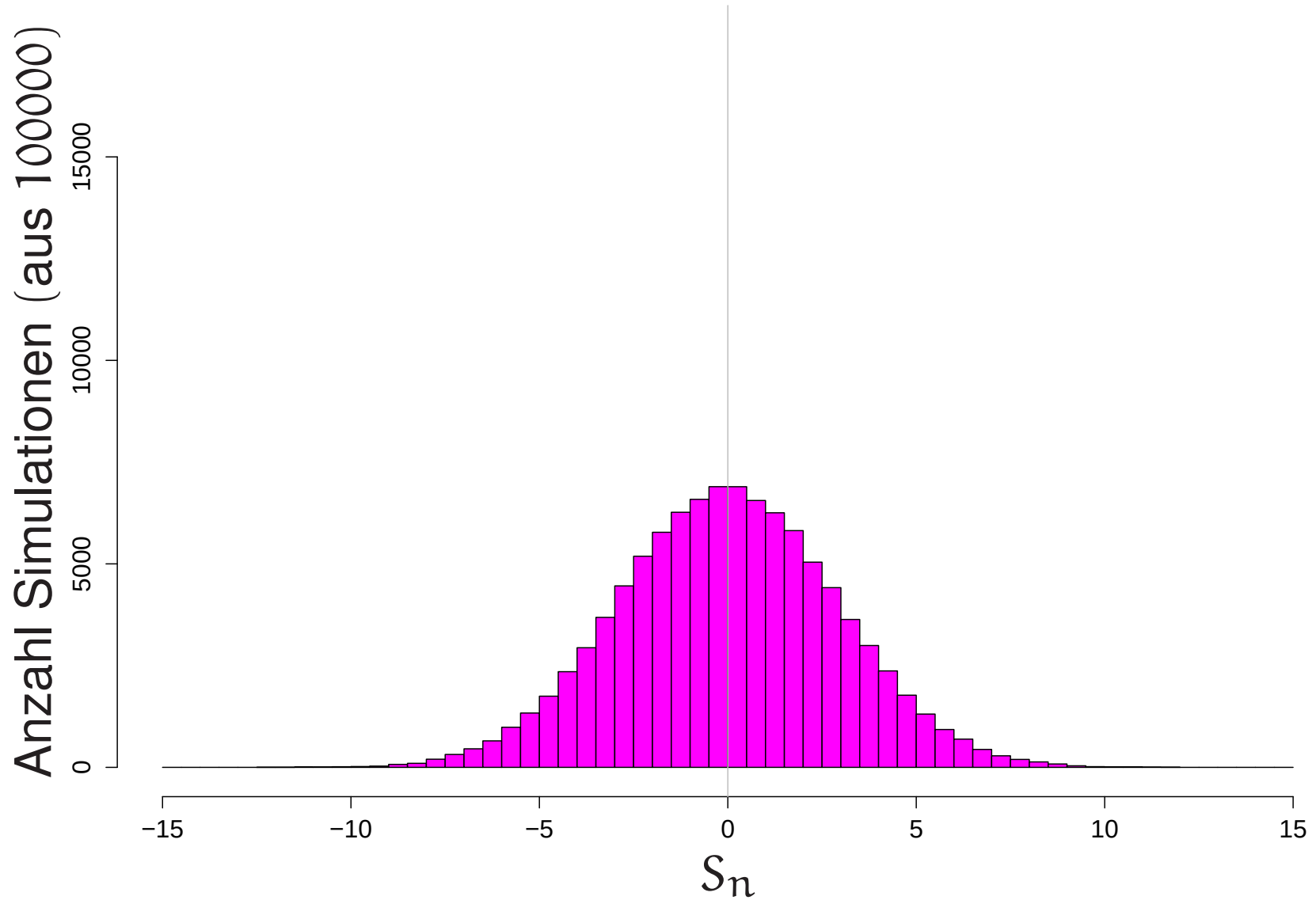
Verteilung von S_n ($n = 25$)



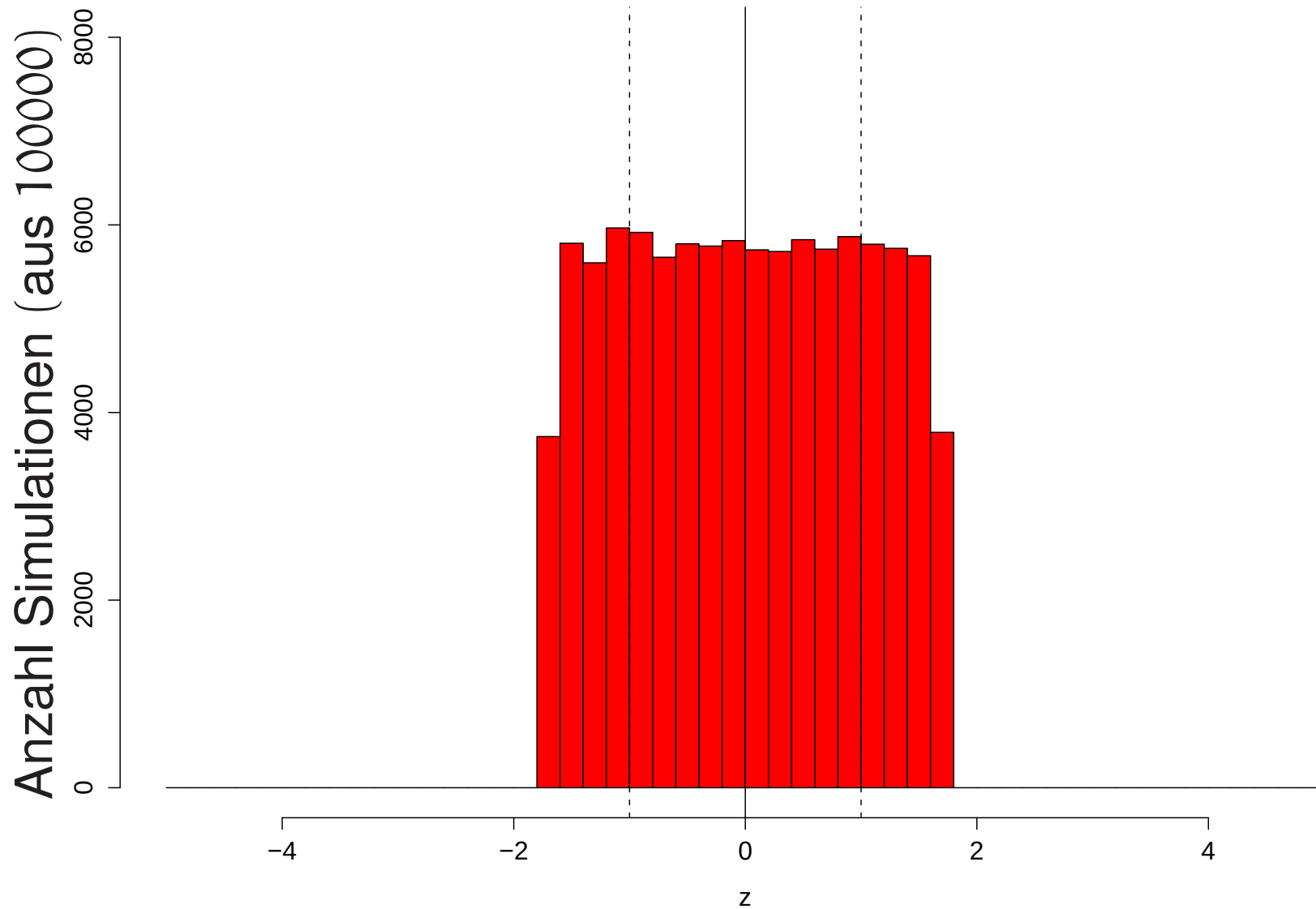
Verteilung von S_n ($n = 50$)



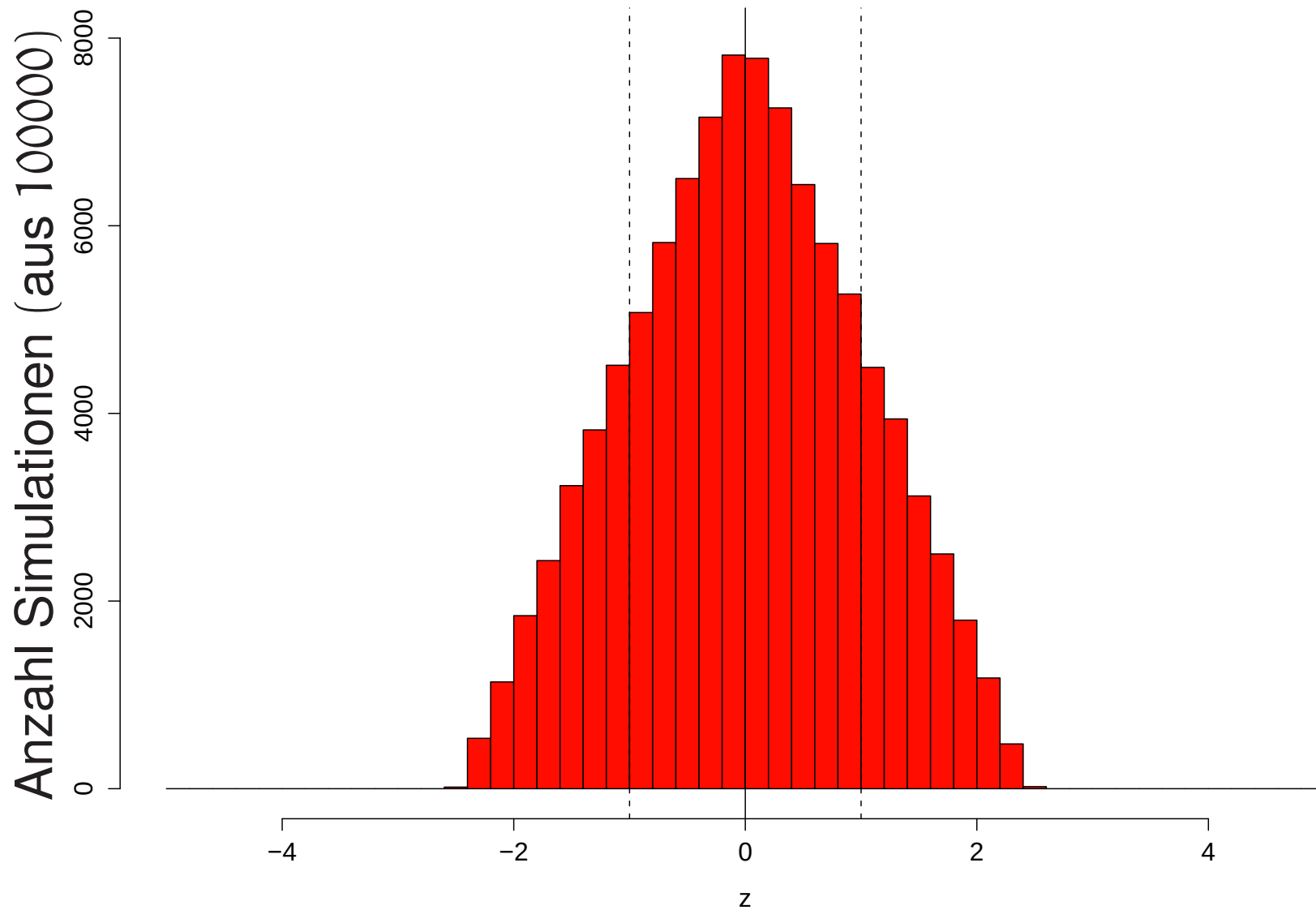
Verteilung von S_n ($n = 100$)



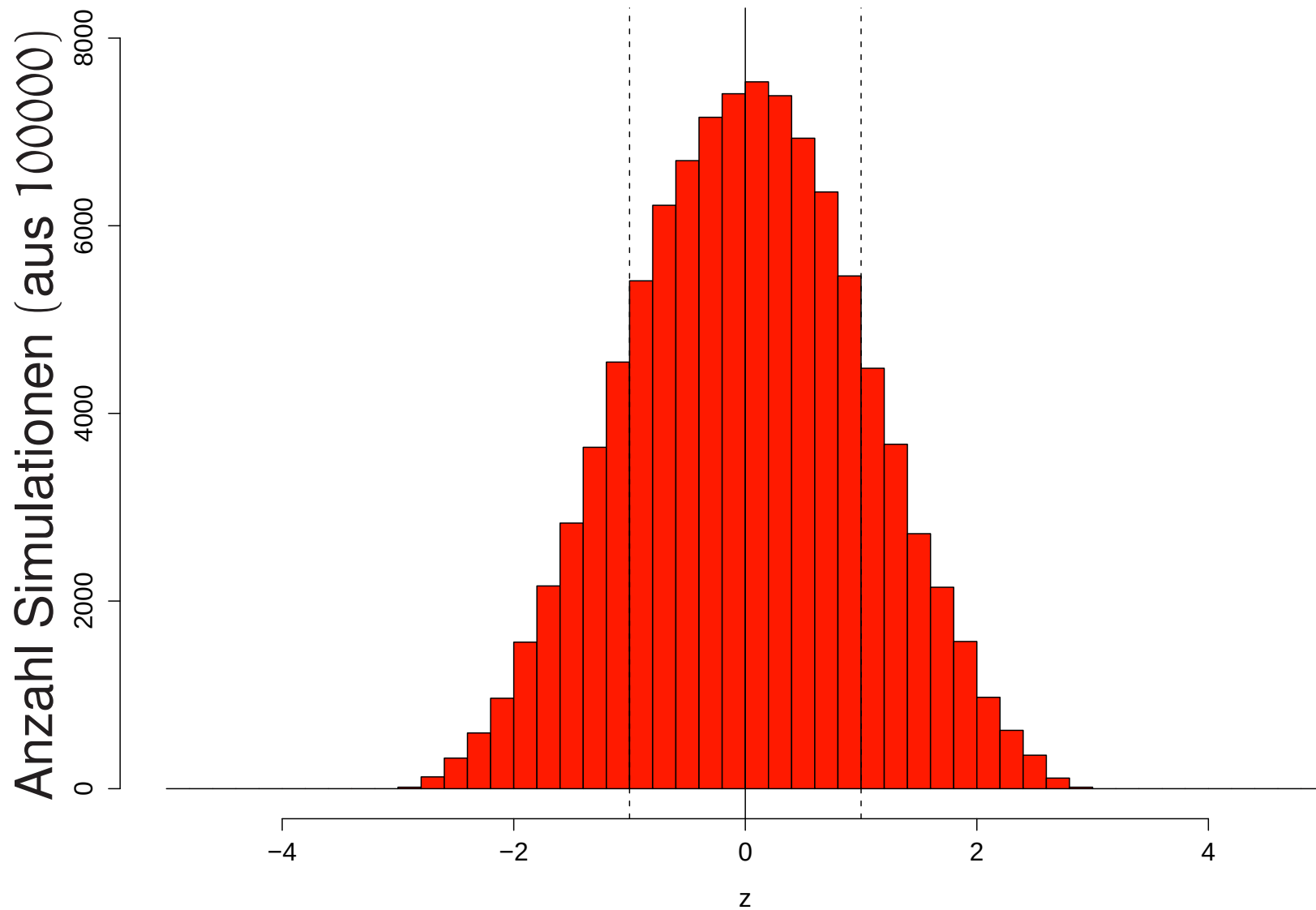
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 1)$



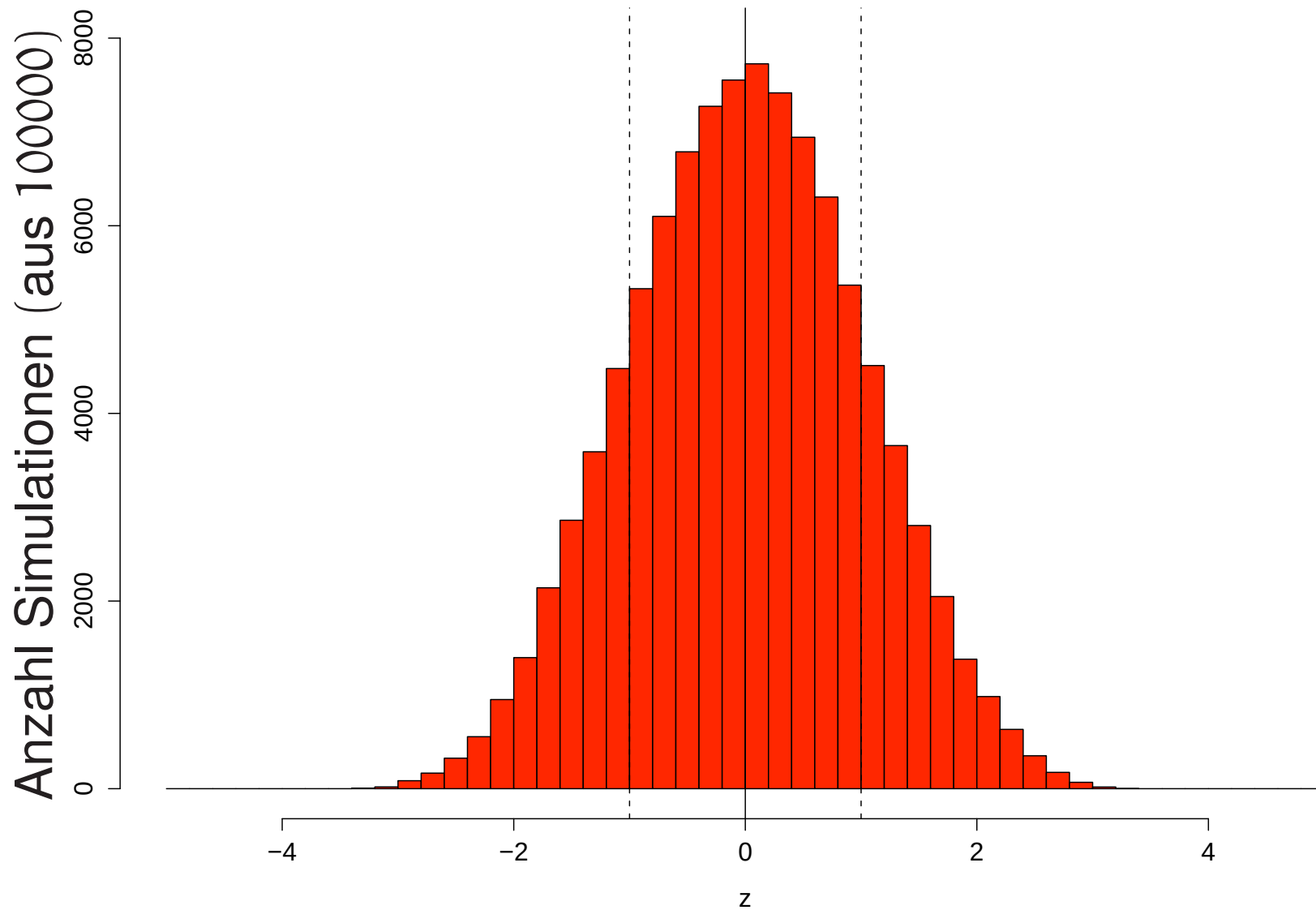
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 2)$



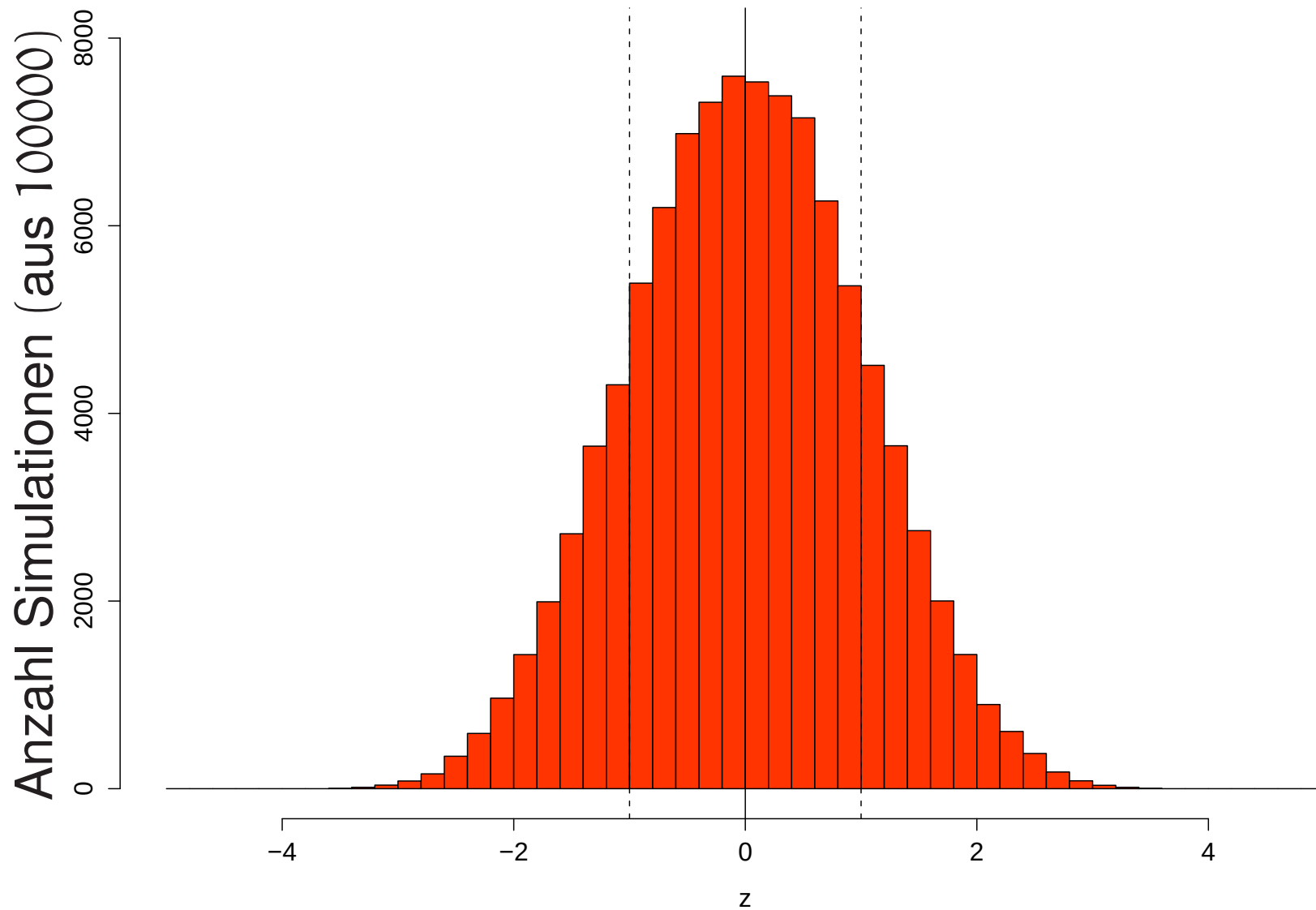
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 3)$



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 4)$



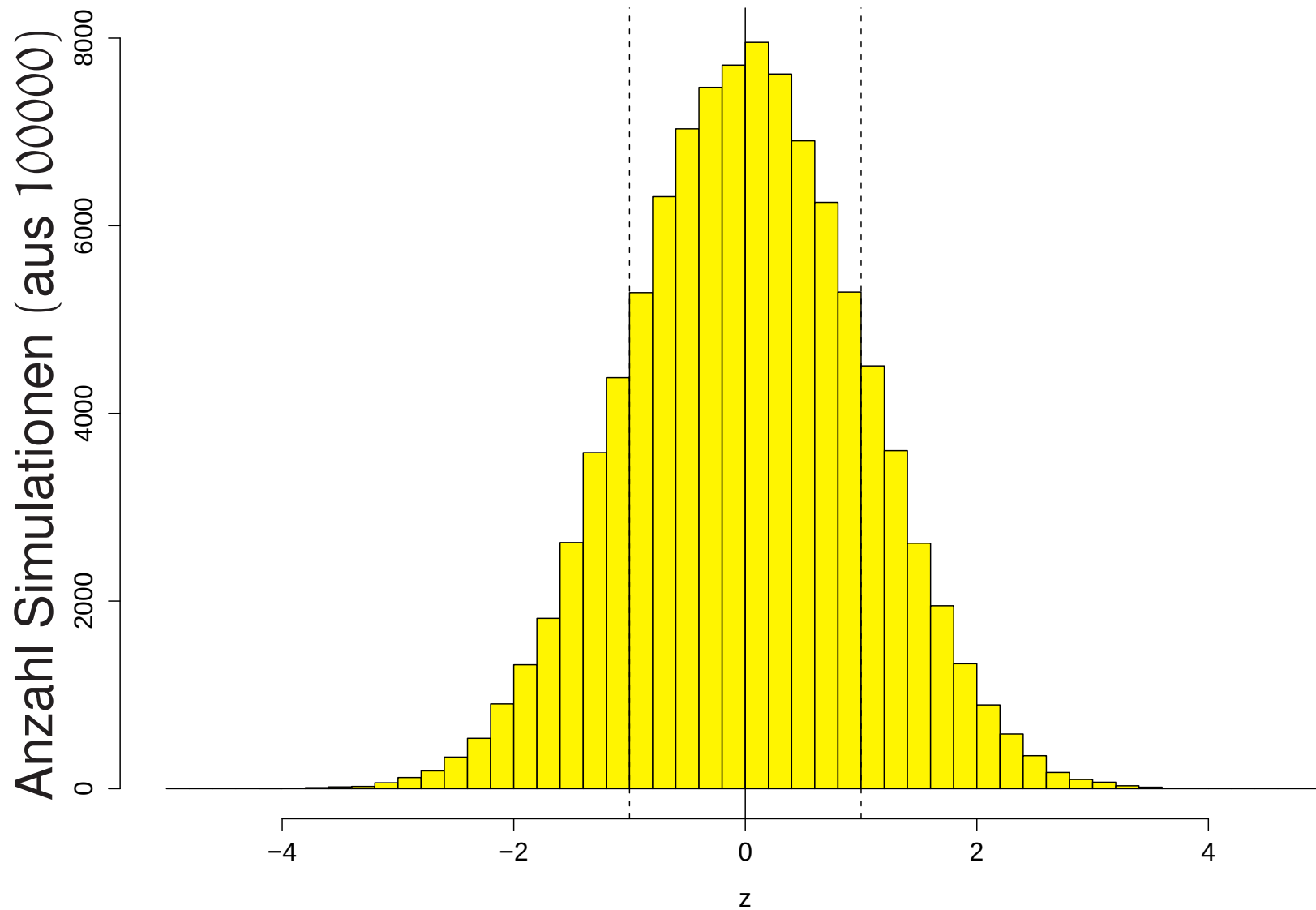
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 5)$



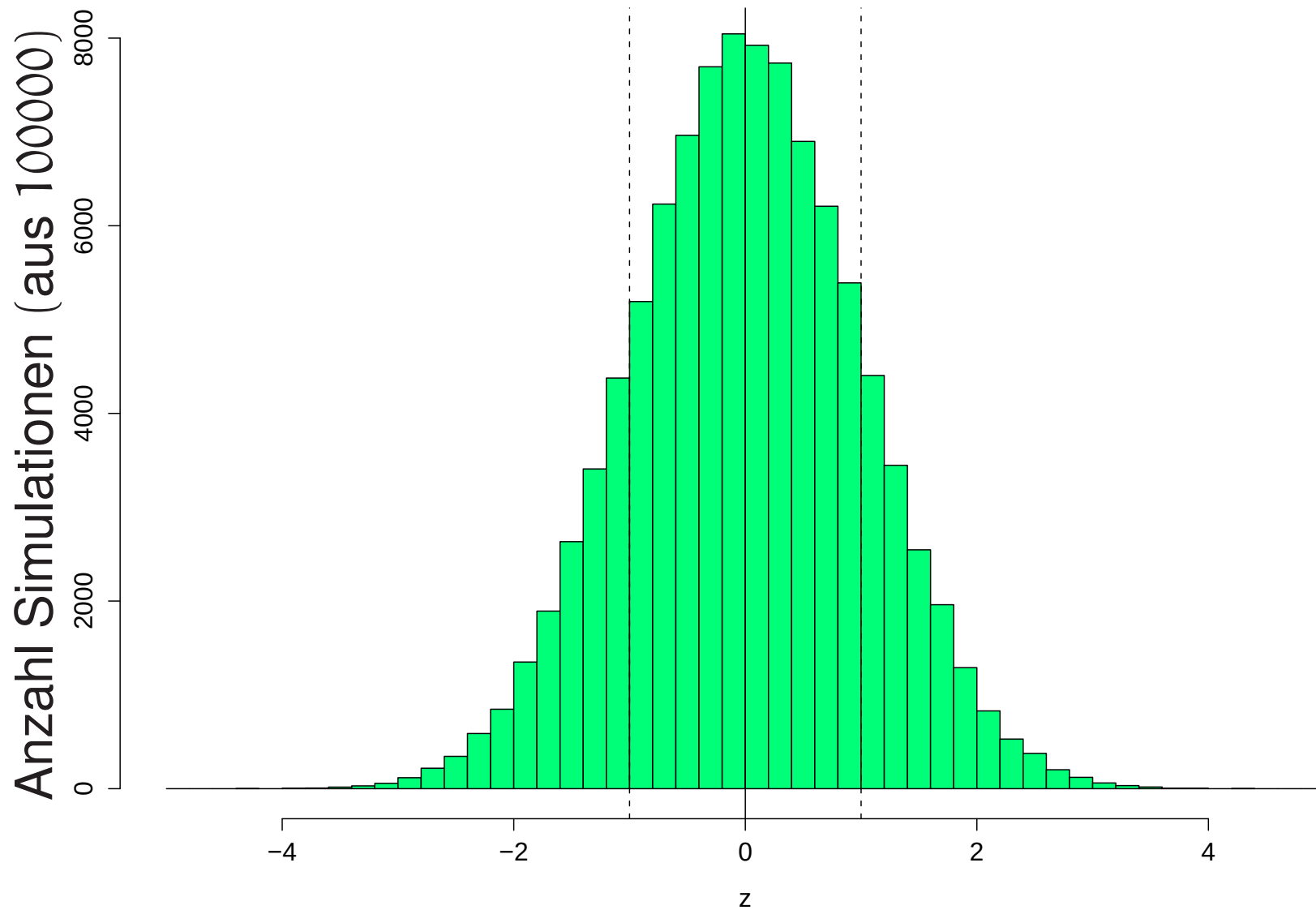
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 10)$



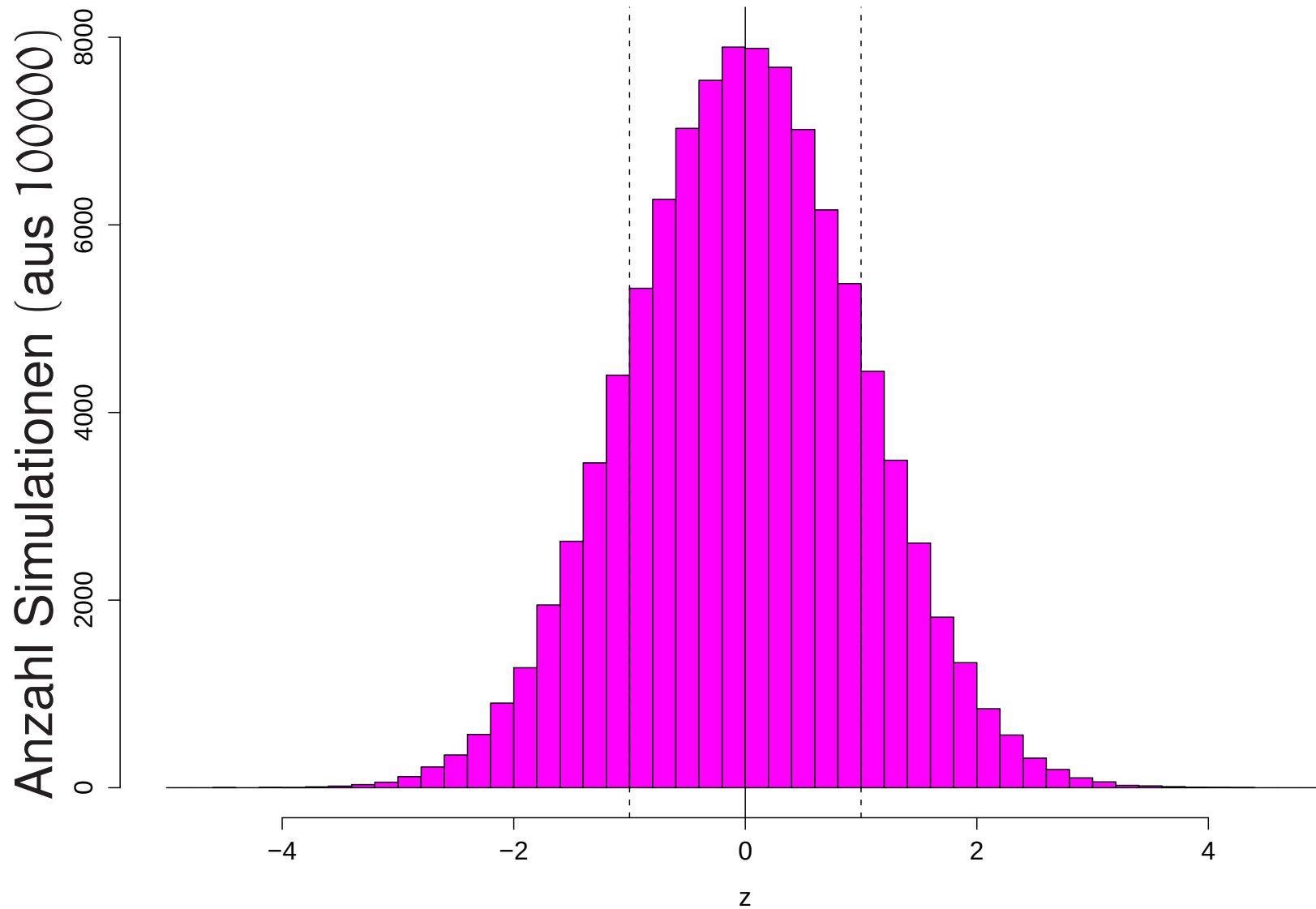
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 20)$



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 50)$



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 100)$



Die Verteilung von
 Z_{100}
ist glockenförmig.

Um welche Glockenkurve handelt es sich genau?

Das nehmen wir im nächsten Teil unter die Lupe.

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 5

Warum gerade e^{-cx^2} ?

Am Ende von Teil 4 hatten wir (für $n = 100$) betrachtet:

$$Z_n := (S_n - \mathbf{E}S_n) / \sigma_{S_n} ,$$

$$\text{mit } S_n := X_1 + \cdots + X_n,$$

$X_1, X_2, \dots$ unabhängig und uniform auf $[-0.5, 0.5]$.

Die Verteilung von Z_n war glockenförmig.

Was ist deren genaue Form für $n \rightarrow \infty$?

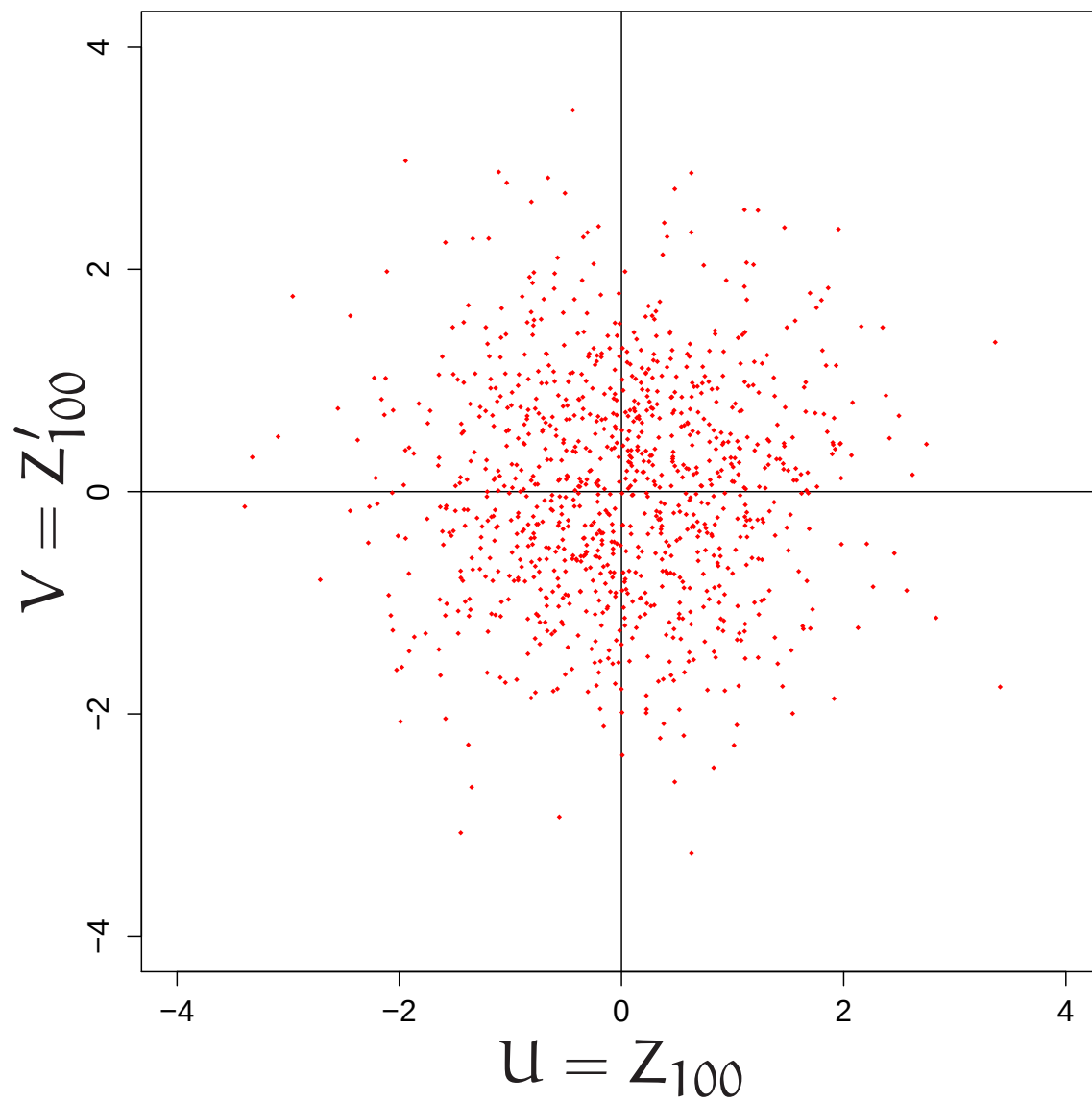
Glücklicher Einfall:

Nimm zwei unabhängige Kopien

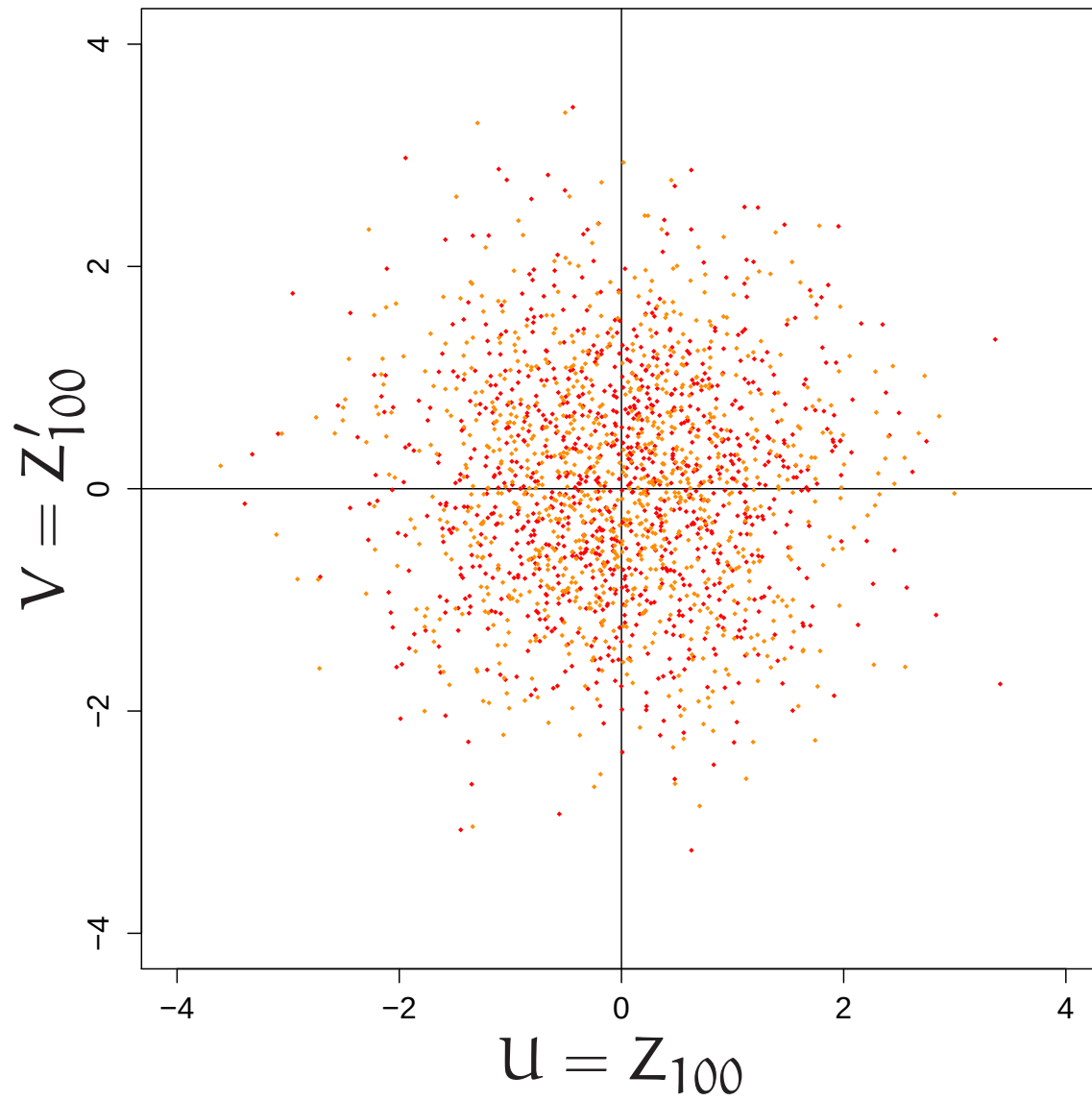
$$(U, V) := (Z_{100}, Z'_{100})$$

Wie sieht die gemeinsame Verteilung
von U und V aus?

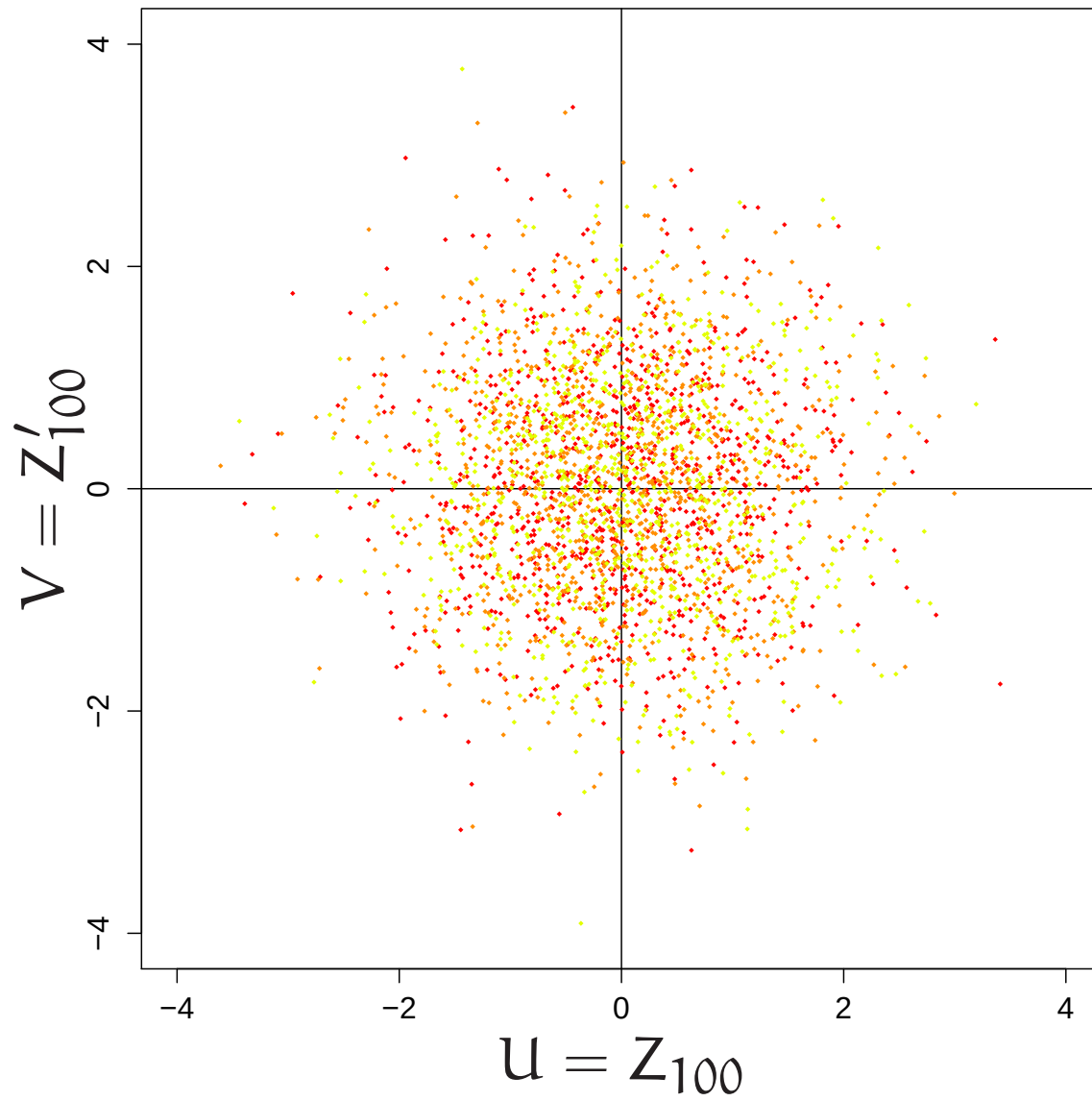
1000 Simulationen



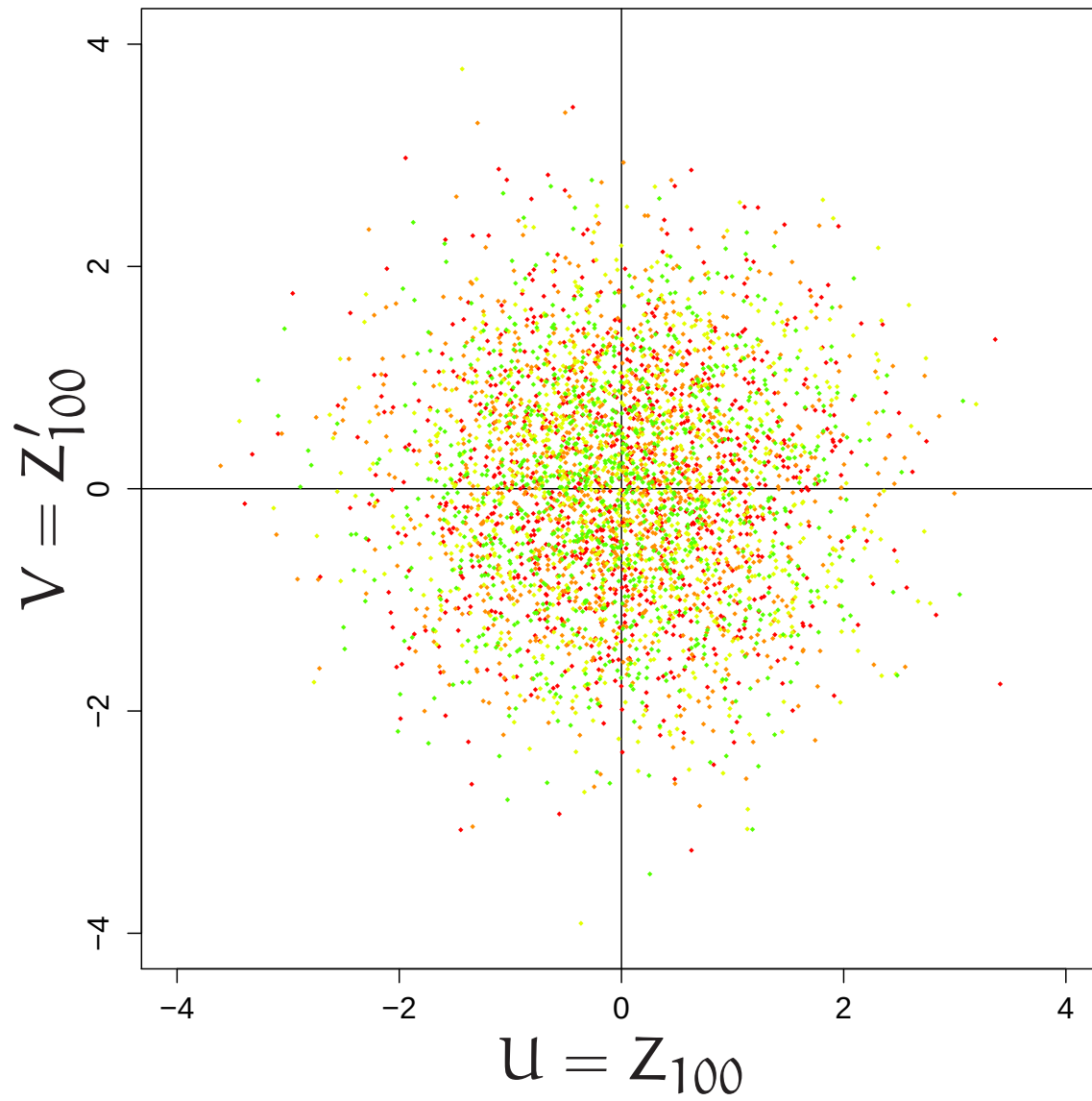
2000 Simulationen



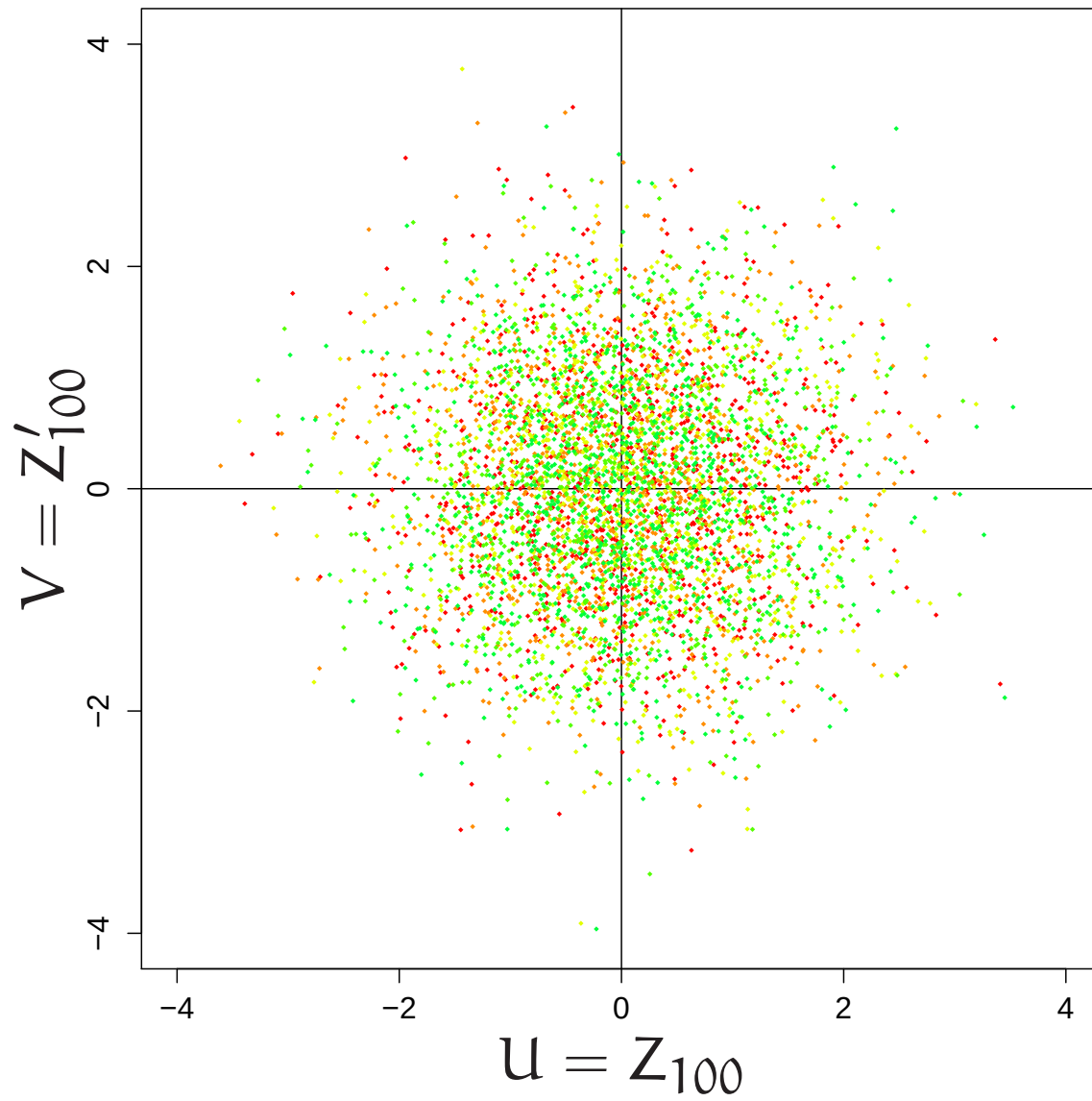
3000 Simulationen



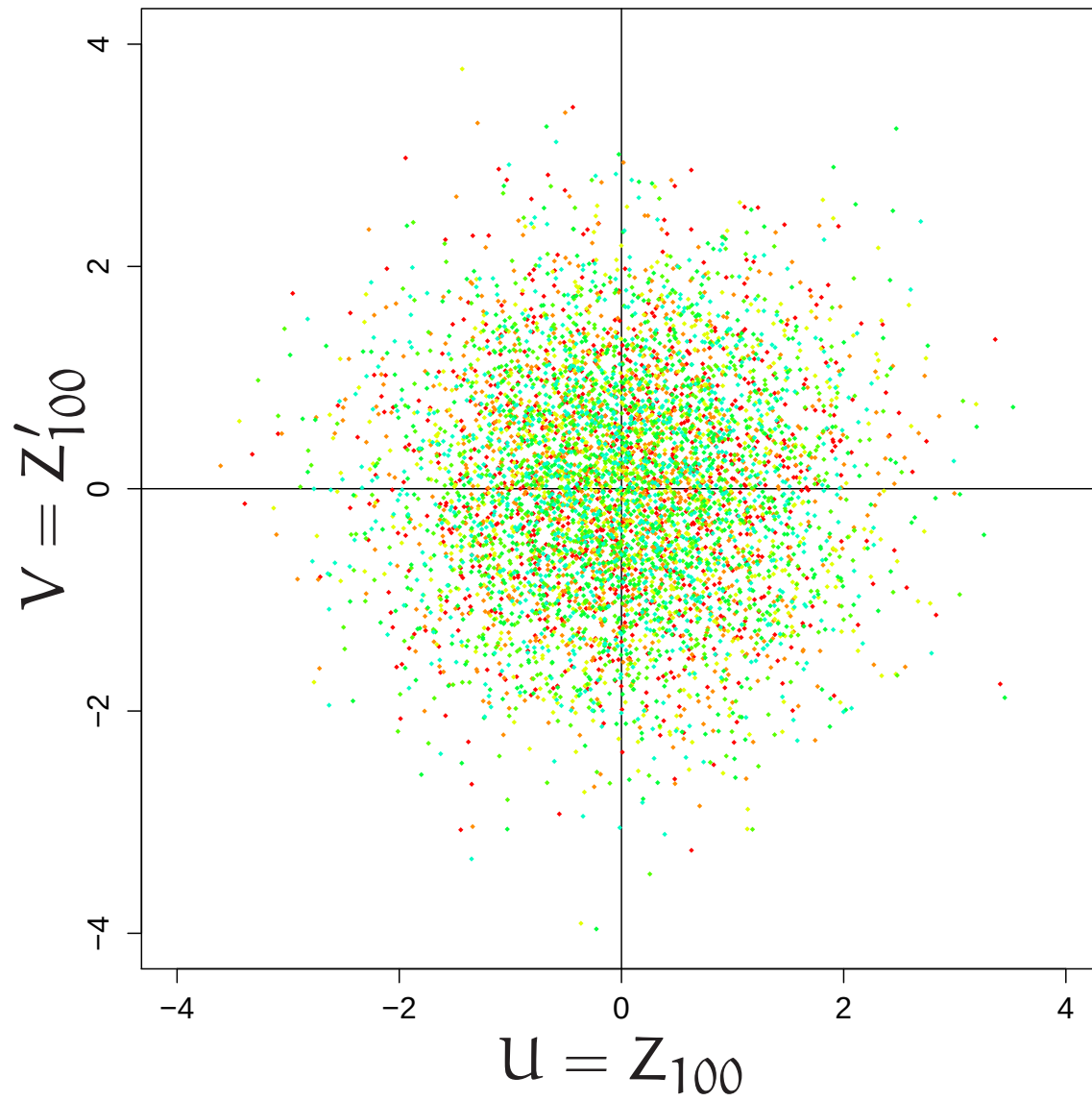
4000 Simulationen



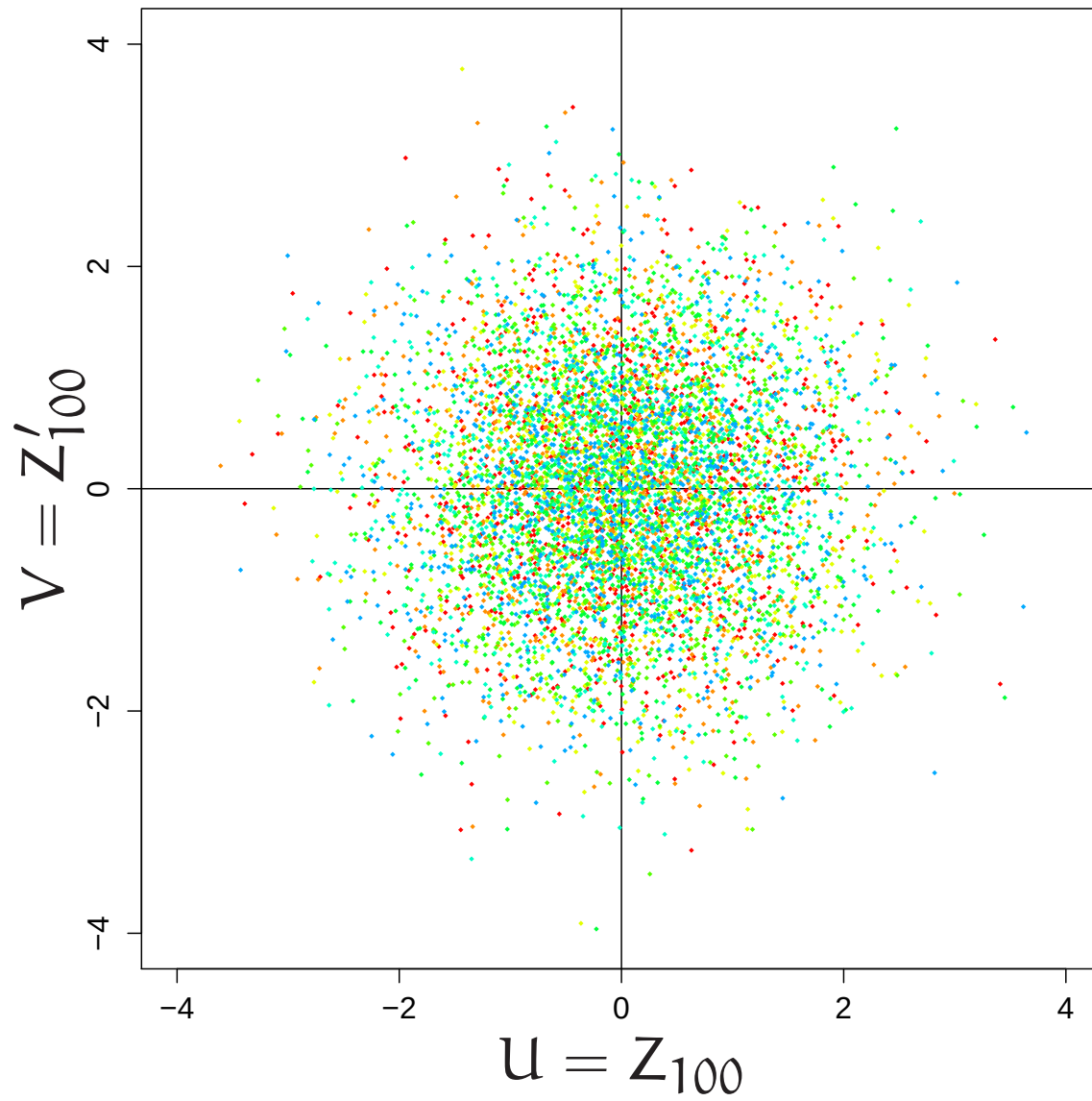
5000 Simulationen



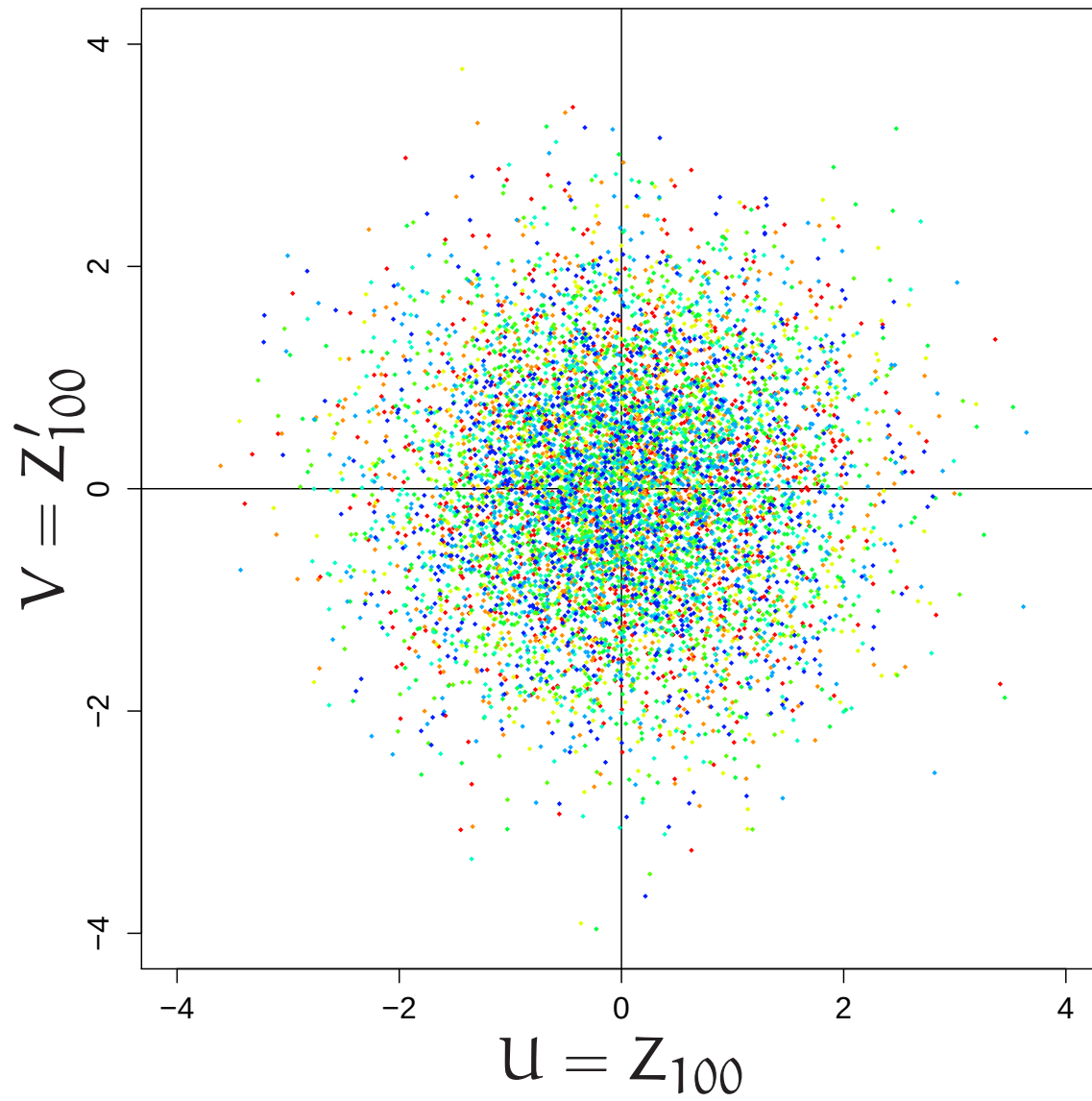
6000 Simulationen



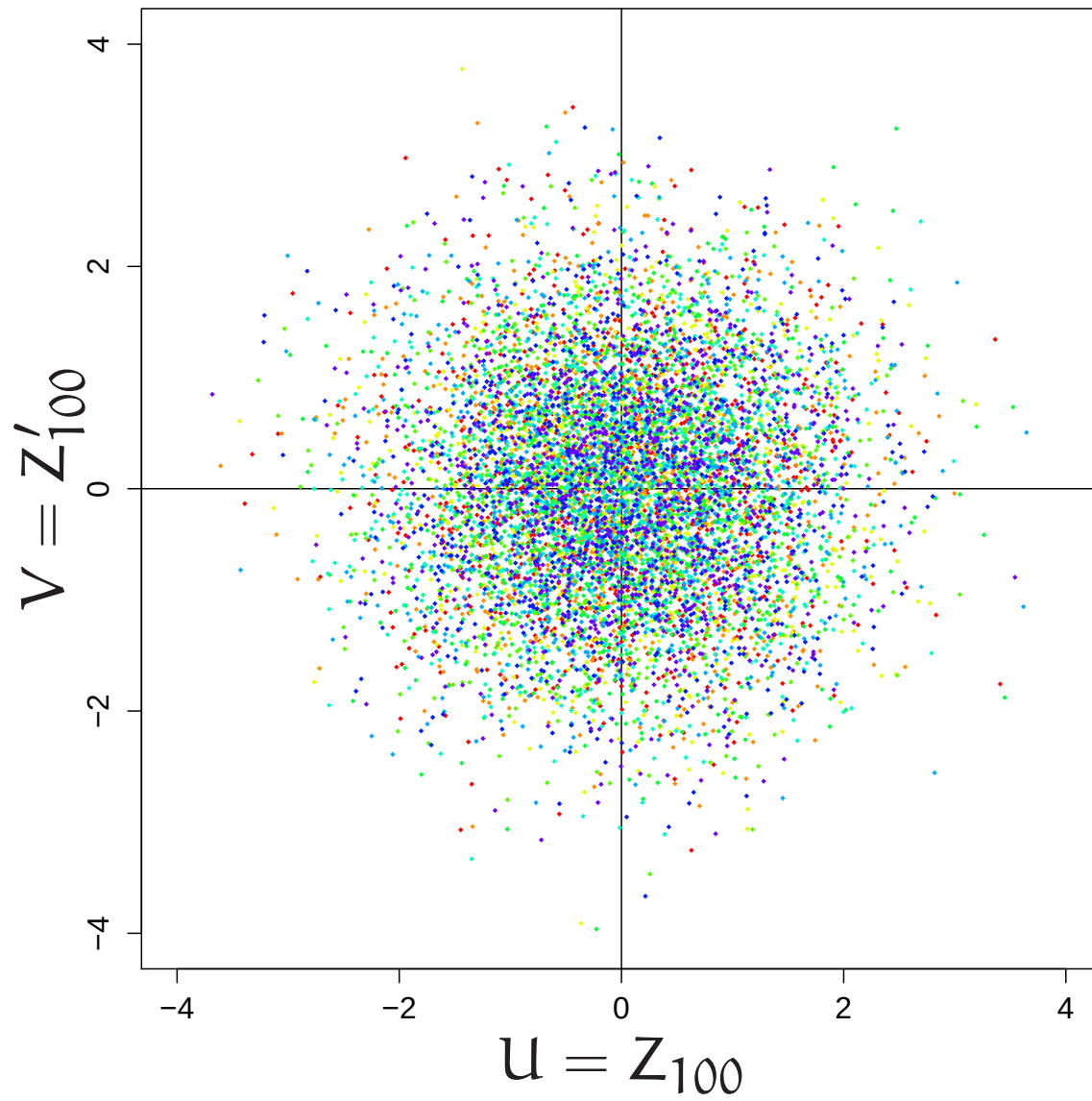
7000 Simulationen



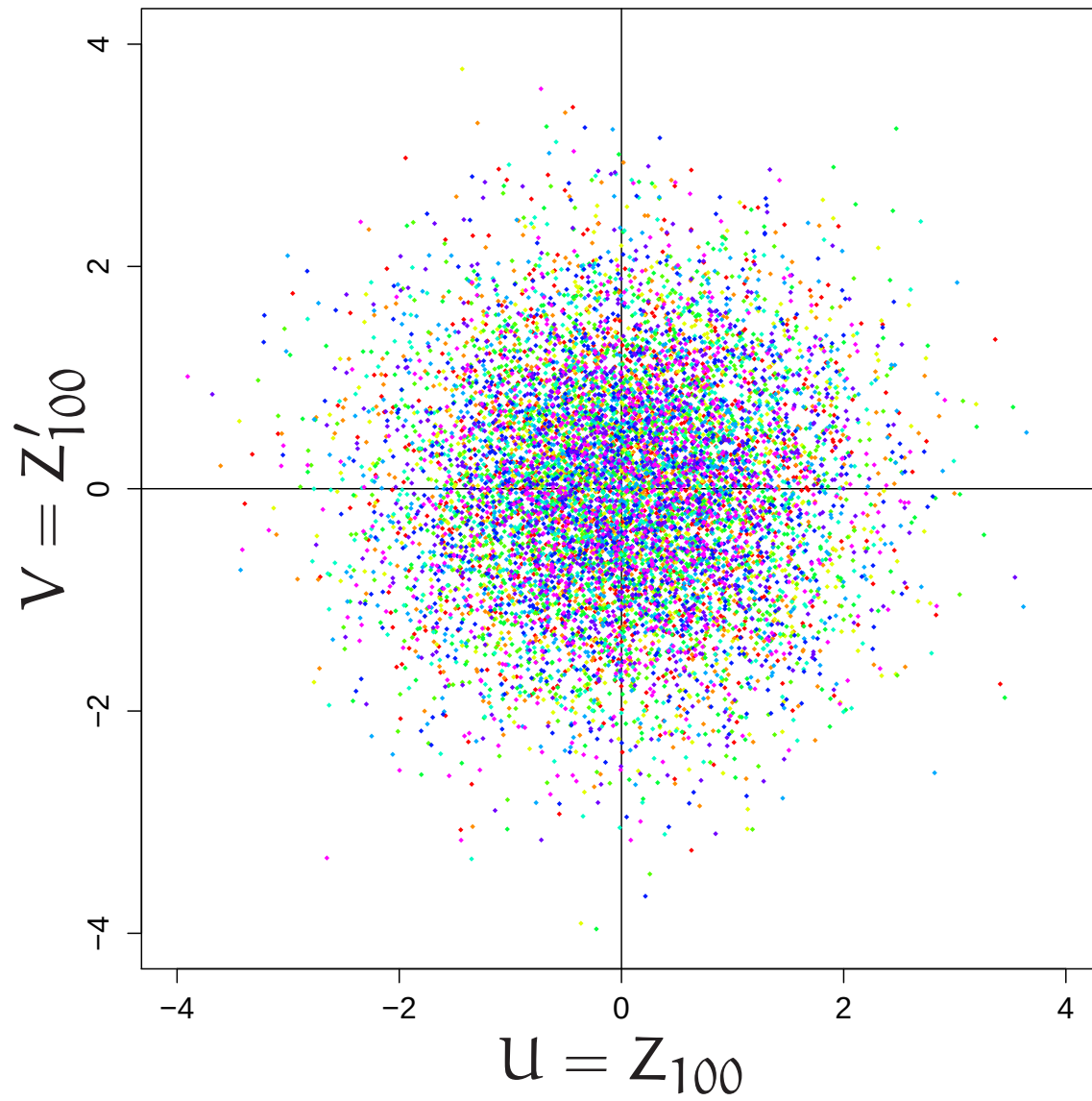
8000 Simulationen



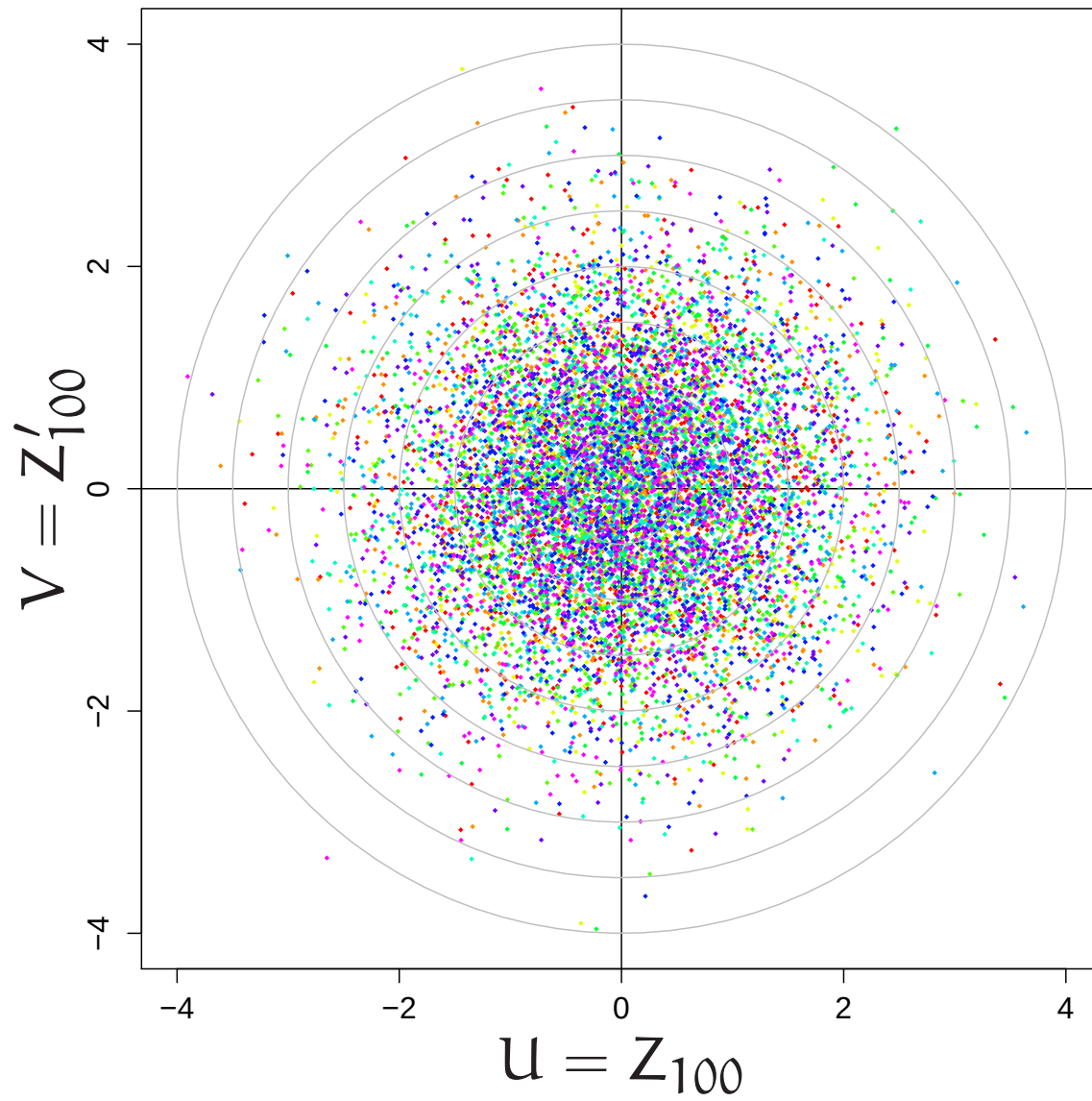
9000 Simulationen



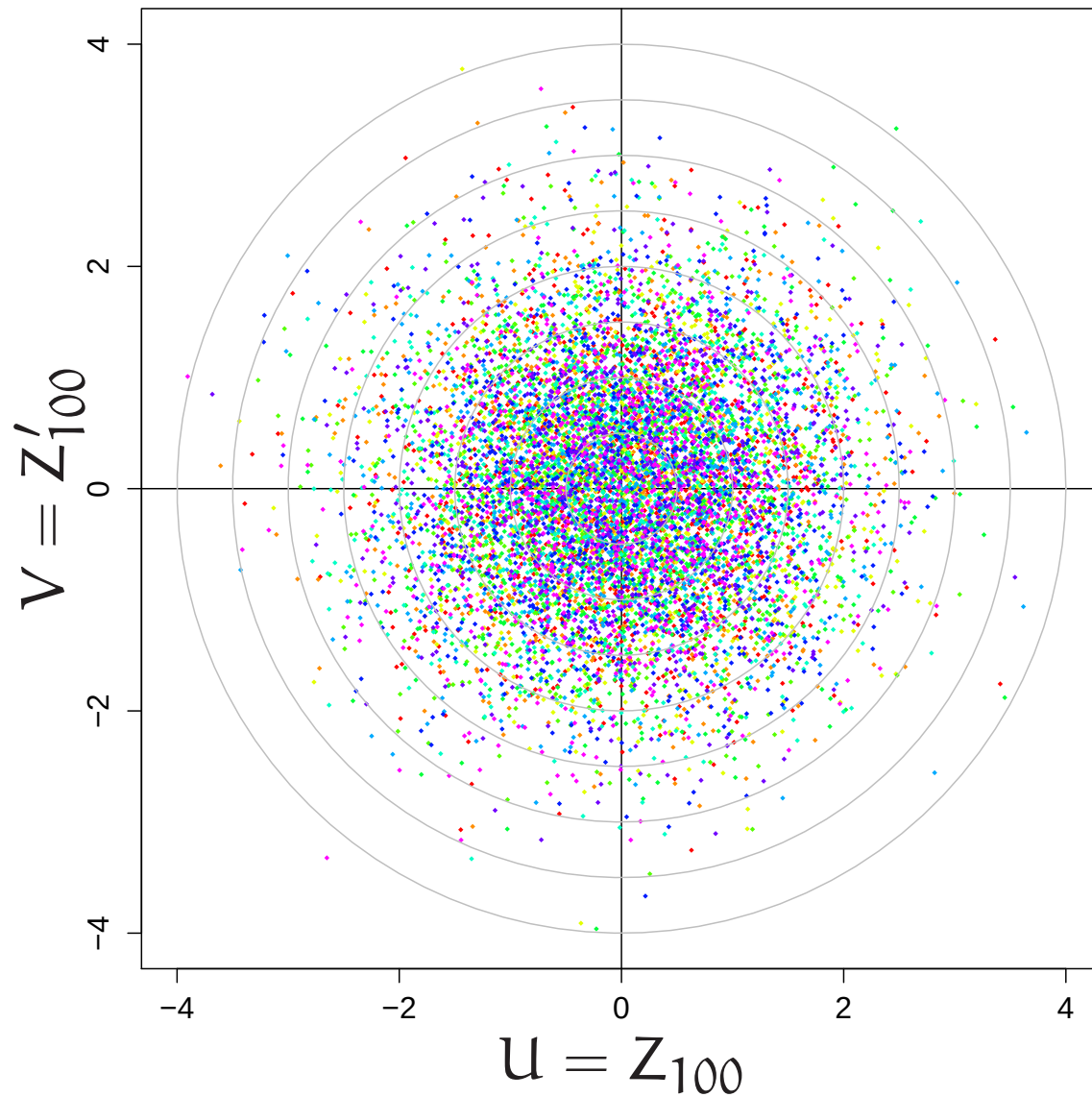
10000 Simulationen



10000 Simulationen



Die Verteilung von (U, V) ist annähernd rotationssymmetrisch!



Eine Charakterisierung der zweidimensionalen Standardnormalverteilung

Behauptung:

Aus “ U und V unabhängig und identisch verteilt’

und

“Verteilung von (U, V) rotationssymmetrisch”

folgt,

dass U und V normalverteilt sind:

$$f_U(x) = f_V(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}$$

Denn:

U, V *unabhängig* bedeutet:

$$(*) \quad f_{(U,V)}(a, b) = f_U(a) f_V(b)$$

$f_{(U,V)}$ *rotationssymmetrisch* heißt: es existiert ein g mit

$$(**) \quad f_{(U,V)}(a, b) = g(r), \quad r := \sqrt{a^2 + b^2}.$$

Mit $f_U = f_V =: h$ folgt aus $(*)$ und $(**)$:

$$h(a)h(b) = g(r), \quad r = \sqrt{a^2 + b^2}$$

$$h(a) h(b) = g(r), \quad r = \sqrt{a^2 + b^2}$$

Die zwei Paare (a, b) und $(0, \sqrt{a^2 + b^2})$ haben dasselbe r .

Also:

$$h(a) h(b) = h(0) h(\sqrt{a^2 + b^2})$$

Das ist eine Gleichung für h . Eine Lösung hiervon ist:

$$h(x) = e^{-x^2}$$

Denn

$$e^{-a^2} e^{-b^2} = 1 \cdot e^{-(a^2+b^2)}$$

$$h(a) h(b) = h(0) h(\sqrt{a^2 + b^2})$$

Wie sehen *alle* die h aus, die diese Beziehung erfüllen?

$$w(u) := h(\sqrt{u}), \quad u \geq 0, \text{ erfüllt}$$

$$w(a^2)w(b^2) = w(0)w(a^2 + b^2), \quad a, b \in \mathbb{R}.$$

$$w(u) w(v) = k_0 w(u + v), \quad u, v \geq 0$$

hat als allgemeine Lösung

$$w(u) = k_0 e^{-k_1 u}, \quad k_1 \in \mathbb{R}.$$

Daraus folgt:

$$h(a) = w(a^2) = k_0 e^{-k_1 a^2}.$$

FAZIT

Der Zentrale Grenzwertsatz
lässt sich erraten
(in konkreten Fällen,
mit etwas Glück).

Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 1

Die Varianz-Kovarianz-Ungleichung

(Buch S. 56)

Wir erinnern an die
Definition der Kovarianz:

Für reellwertige Zufallsvariable X, Y
mit $\mathbf{E}[X^2] < \infty$ und $\mathbf{E}[Y^2] < \infty$ ist

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

Insbesondere ist also

$$\mathbf{Cov}[X, X] = \mathbf{Var}[X].$$

Wichtige Eigenschaften der Kovarianz:

Die Kovarianz ist

- im Fall von zwei gleichen Einträgen *nichtnegativ*:

$$\text{Cov}[X, X] \geq 0$$

- in den beiden Einträgen *symmetrisch*:

$$\text{Cov}[X, Y] = \text{Cov}[Y, X]$$

- *bilinear*, d.h. in jedem einzelnen Eintrag linear:

$$\text{Cov}[c_1 X_1 + c_2 X_2, Y] = c_1 \text{Cov}[X_1, Y] + c_2 \text{Cov}[X_2, Y]$$

Wir beweisen jetzt die

Kovarianz-Varianz-Ungleichung:

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var} X} \sqrt{\text{Var} Y}$$

Mit $G := X - \mu_X$, $H := Y - \mu_Y$

folgt die Behauptung aus der

Cauchy-Schwarz Ungleichung:

Für reellwertige Zufallsvariable G, H mit $\mathbf{E}[G^2], \mathbf{E}[H^2] < \infty$
ist

$$|\mathbf{E}[GH]| \leq \sqrt{\mathbf{E}[G^2]} \sqrt{\mathbf{E}[H^2]} .$$

Beweis der Cauchy-Schwarz Ungleichung:

Fall 1: $E[G^2] > 0$ und $E[H^2] > 0$.

Zu zeigen ist $\left| E \left[\frac{G}{\sqrt{E[G^2]}} \frac{H}{\sqrt{E[H^2]}} \right] \right| \leq 1$.

Für $U := G/\sqrt{E[G^2]}$, $V := H/\sqrt{E[H^2]}$

ist also zu zeigen: $|E[UV]| \leq 1$.

Aus $\pm 2UV \leq U^2 + V^2$ folgt (mit Monotonie des EW)
 $\pm E[UV] \leq \frac{1}{2}(E[U^2] + E[V^2]) = 1$.

Fall 2: $\mathbf{E}[G^2] = 0$.

Dann folgt

mit dem Satz von der Positivität des Erwartungswertes:

$$\mathbf{P}(G^2 = 0) = 1,$$

also

$$\mathbf{P}(GH = 0) = 1$$

und

$$\mathbf{E}[GH] = 0,$$

und damit

$$\pm \mathbf{E}[GH] = 0 \leq 0 = \sqrt{\mathbf{E}[G^2]} \sqrt{\mathbf{E}[H^2]}. \quad \square$$

Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 2

Der Korrelationskoeffizient

(Buch S. 62)

Definition.

Für zwei Zufallsvariable X, Y
mit positiven, endlichen Varianzen ist

$$\kappa_{XY} := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var} X} \sqrt{\text{Var} Y}}$$

der **Korrelationskoeffizient** von X und Y
(kurz auch: *die Korrelation* von X und Y).

Aus der Kovarianz-Varianz-Ungleichung folgt sofort:

$$-1 \leq \kappa_{XY} \leq 1.$$

Fünf prominente Zahlen
zur (teilweisen) Beschreibung der Verteilung
eines zufälligen Paares (X, Y) in $\mathbb{R} \times \mathbb{R}$:

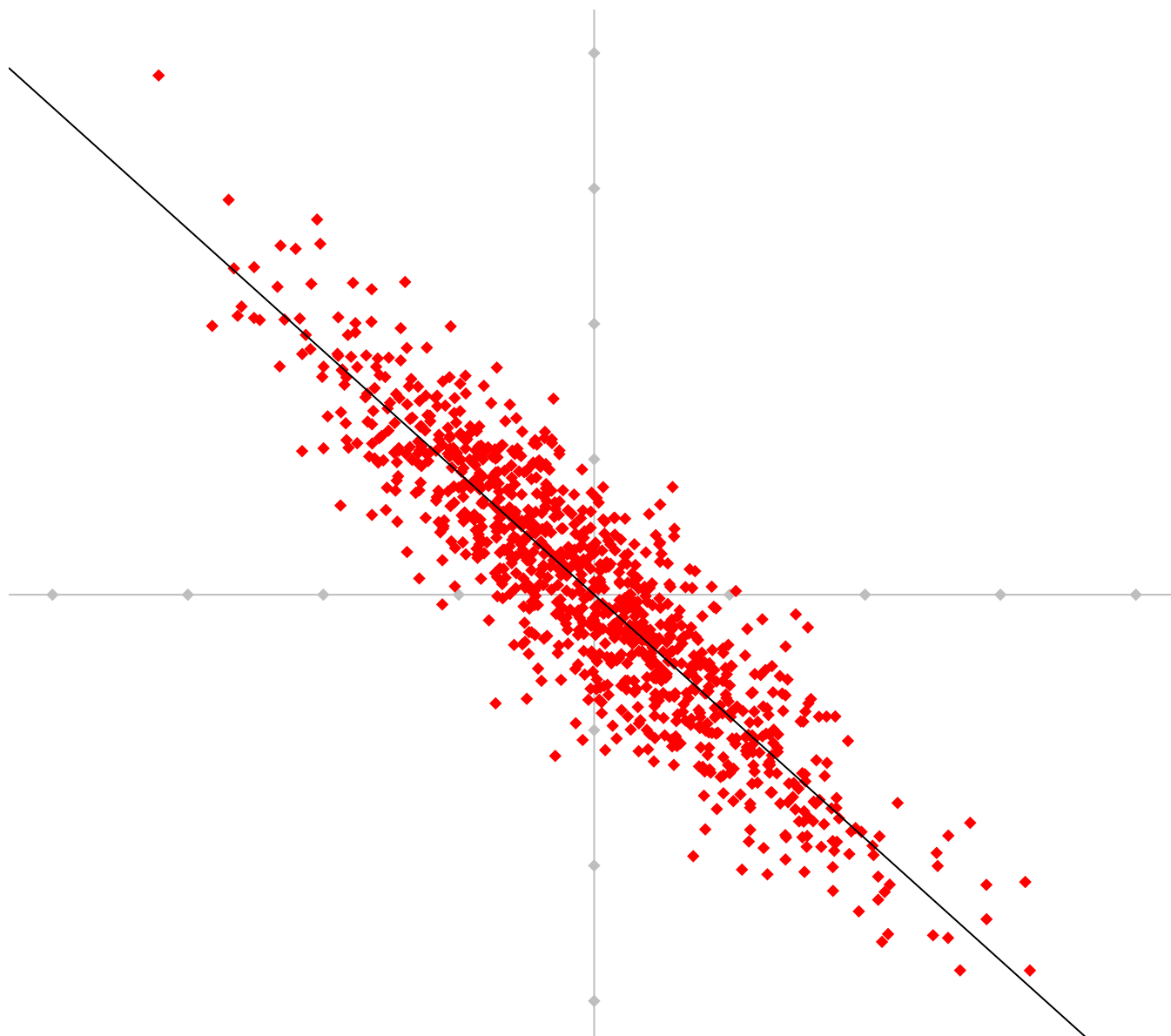
μ_X und μ_Y : die Erwartungswerte von X und Y

σ_X und σ_Y : die Standardabweichungen von X und Y

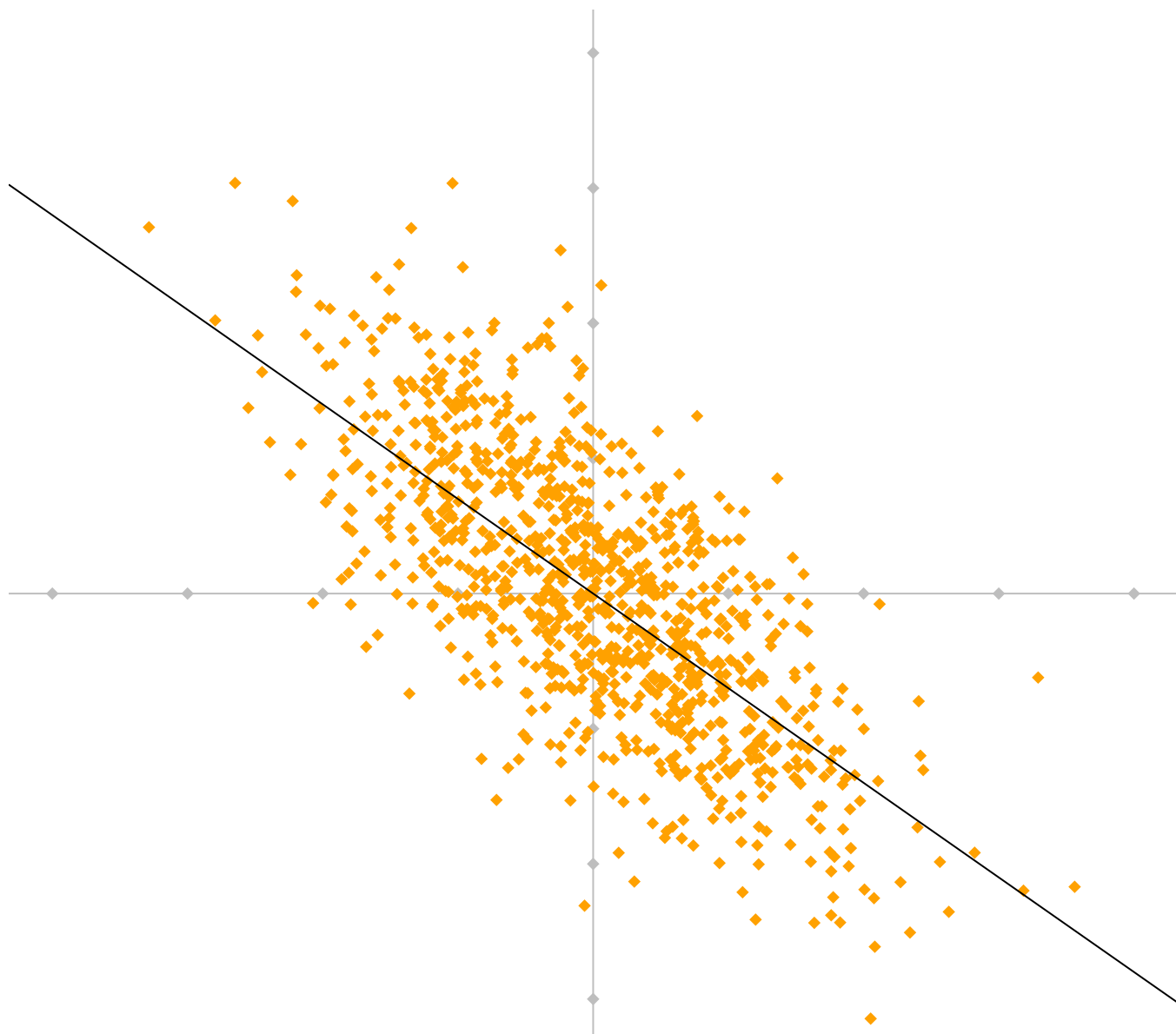
κ_{XY} : der Korrelationskoeffizient von X und Y

Die folgenden 11 Bilder zeigen jeweils die
Realisierungen von 1000 unabhängige Kopien (X_i, Y_i)
eines zufälligen Paares (X, Y) , mit
 $X \sim N(0, 1)$ -verteilt, $Y \sim N(0, 1)$ -verteilt, und
 $\kappa_{XY} = -0.9, -0.7, \dots, -0.1, 0, 0.1, \dots, 0.7, 0.9$,
zusammen mit der
Geraden durch den Ursprung mit Anstieg κ_{XY} .

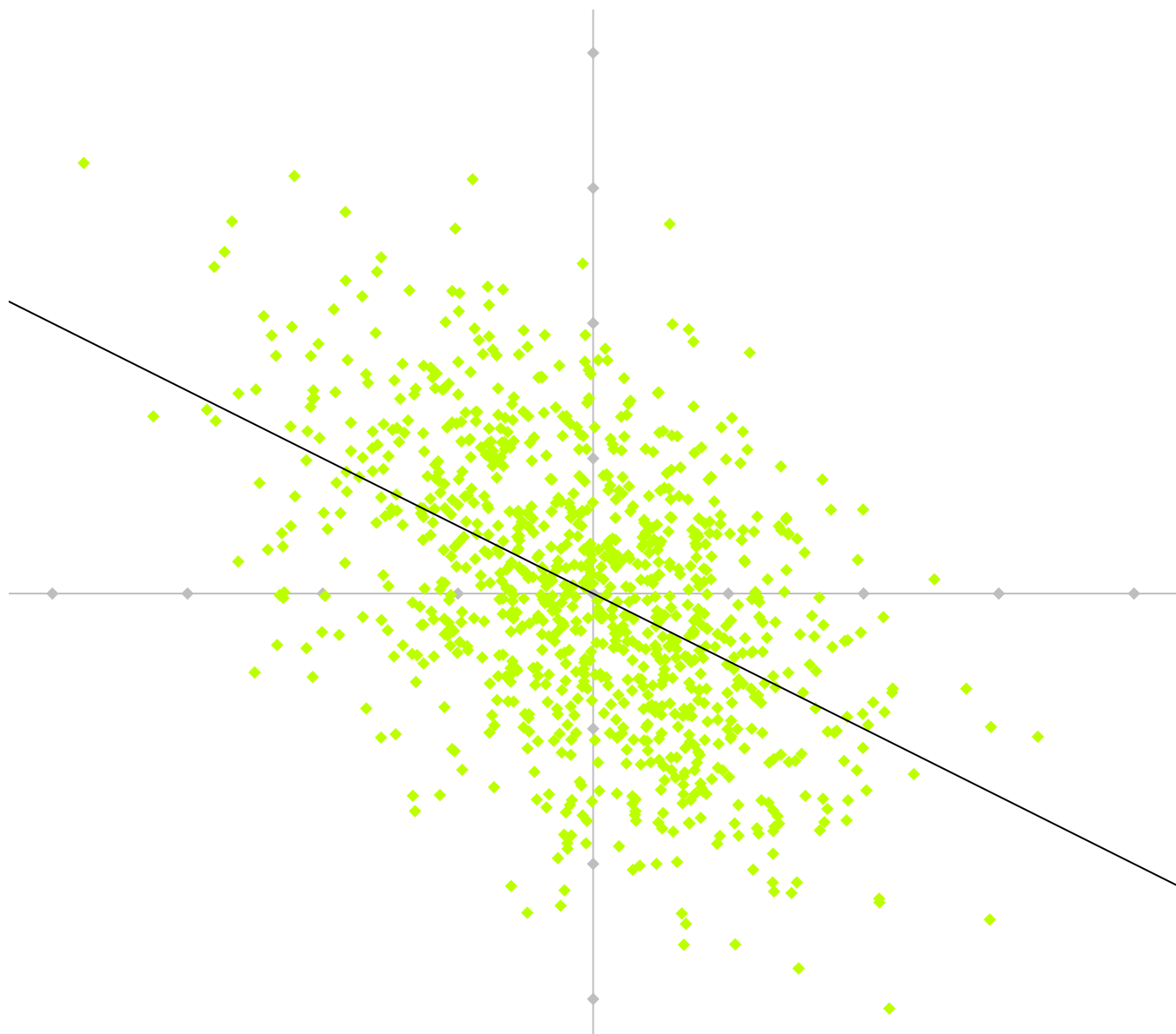
Korrelation = - 0.9



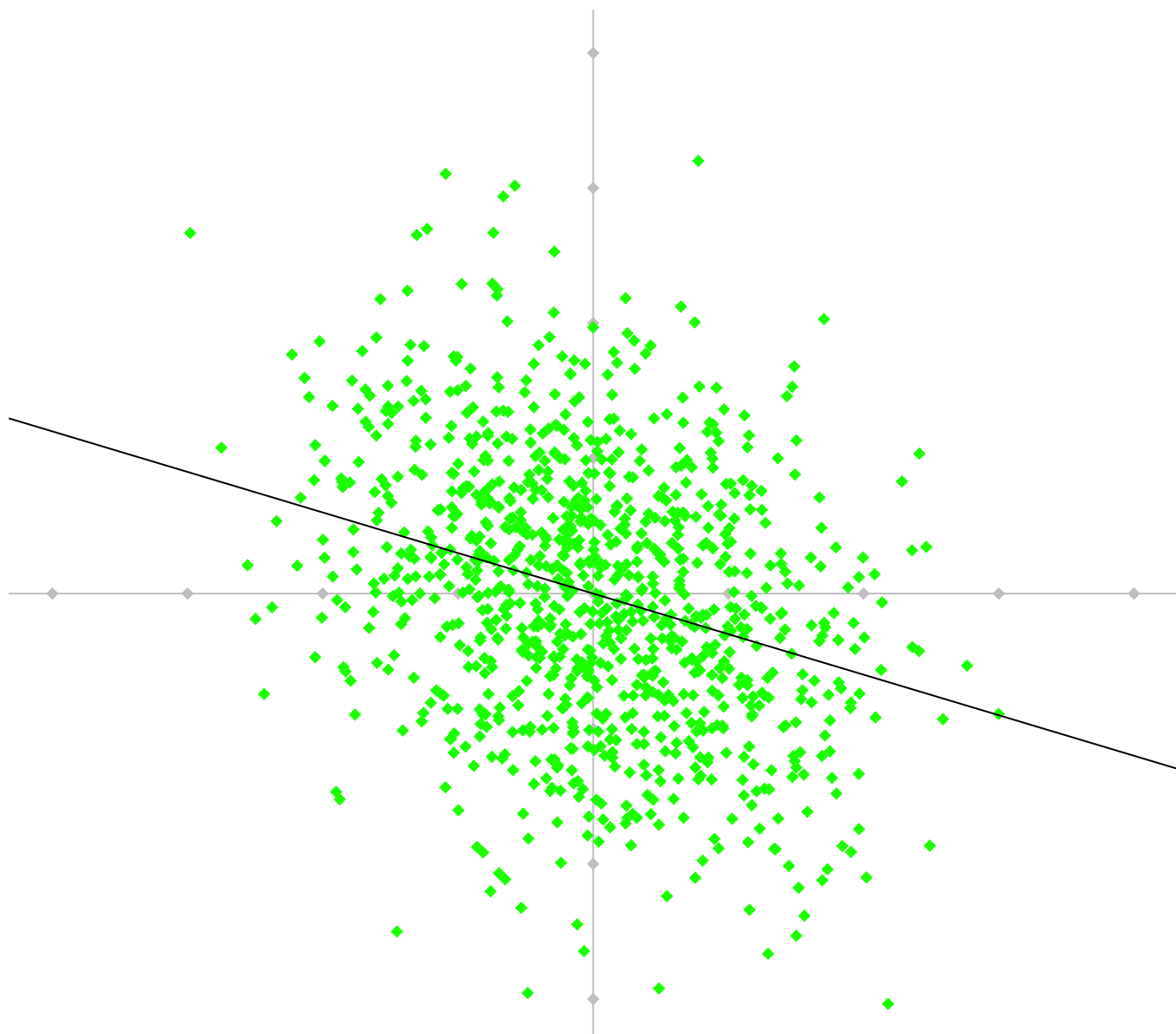
Korrelation = - 0.7



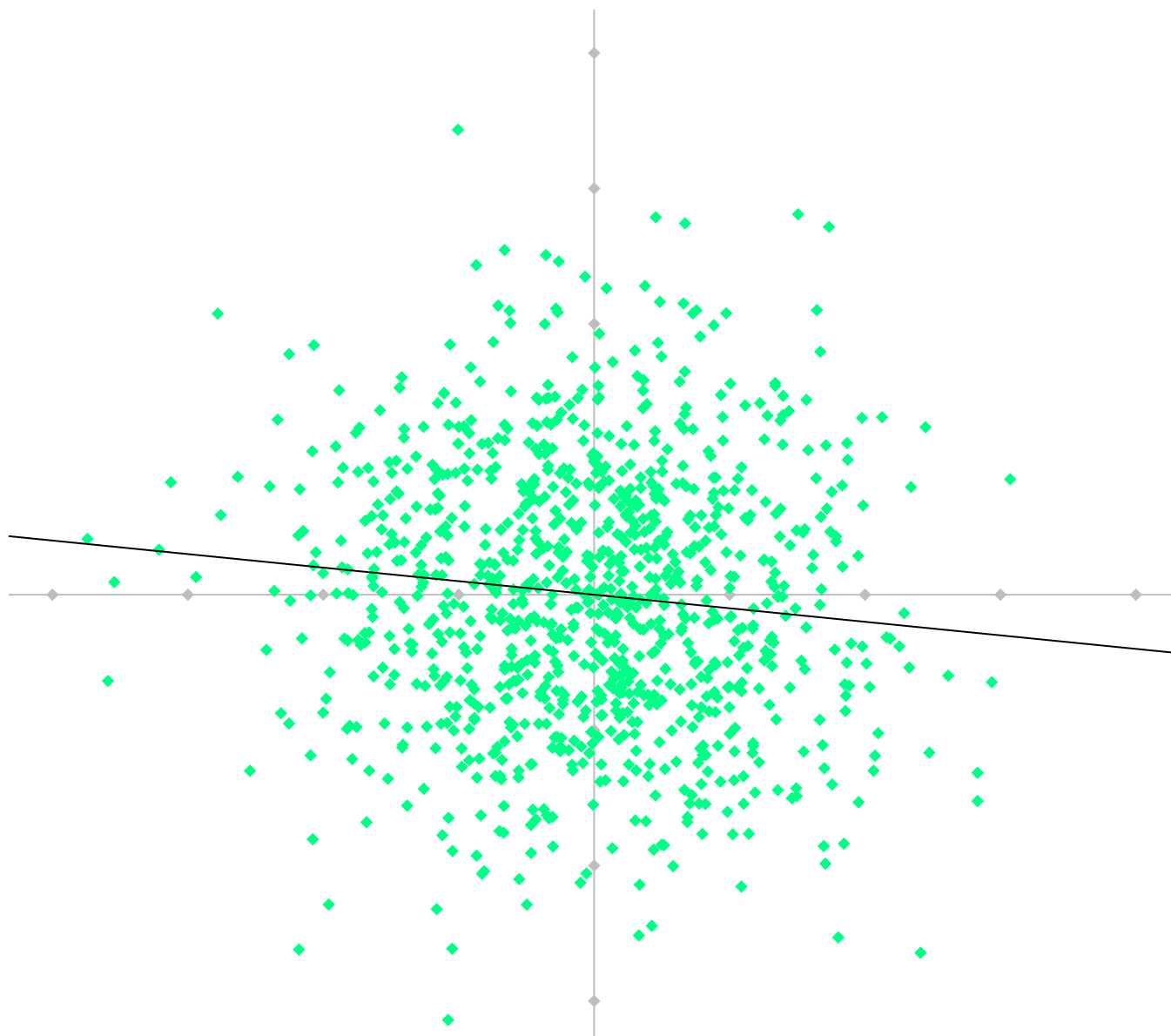
Korrelation = - 0.5



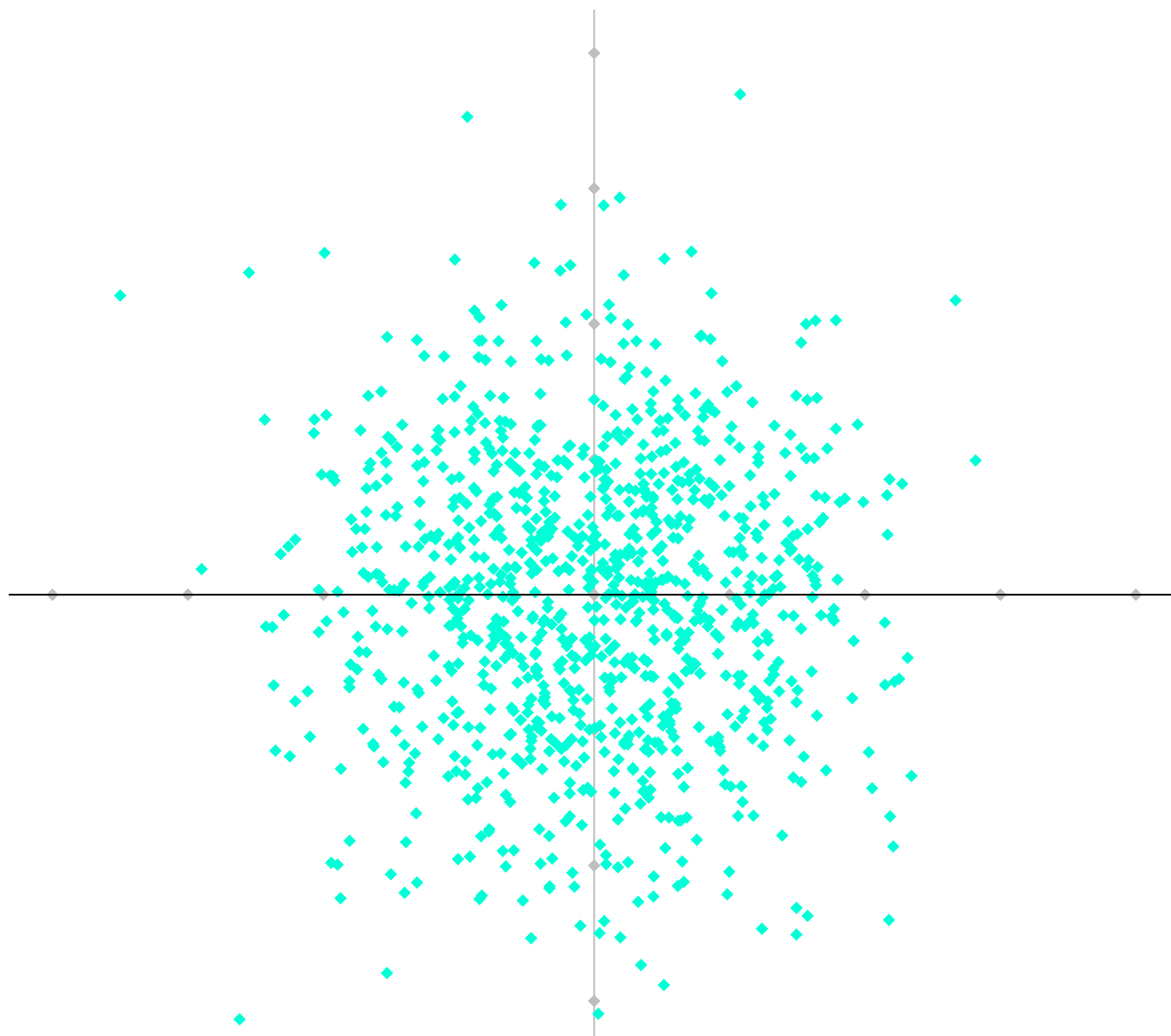
Korrelation = - 0.3



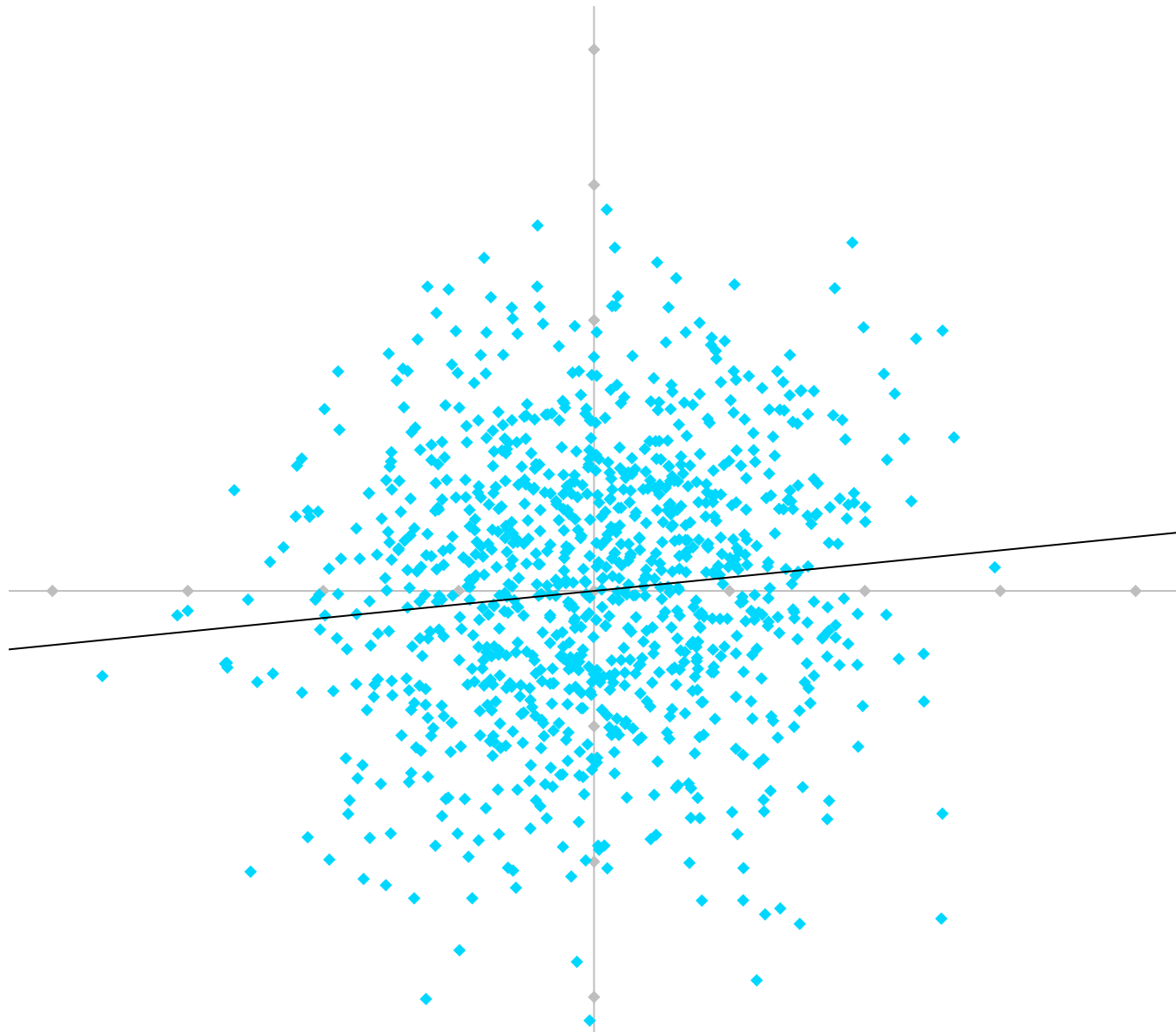
Korrelation = - 0.1



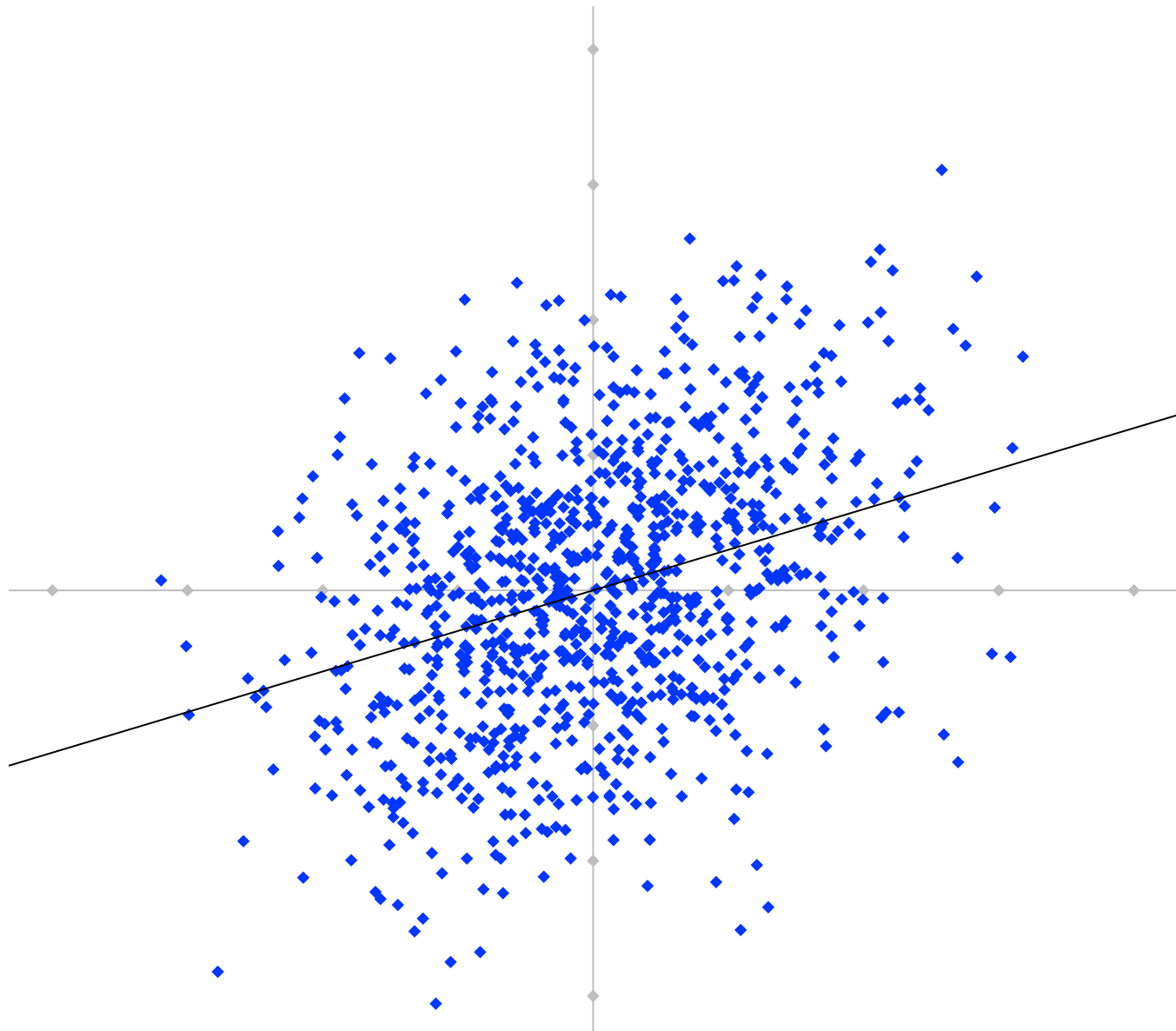
Korrelation = 0



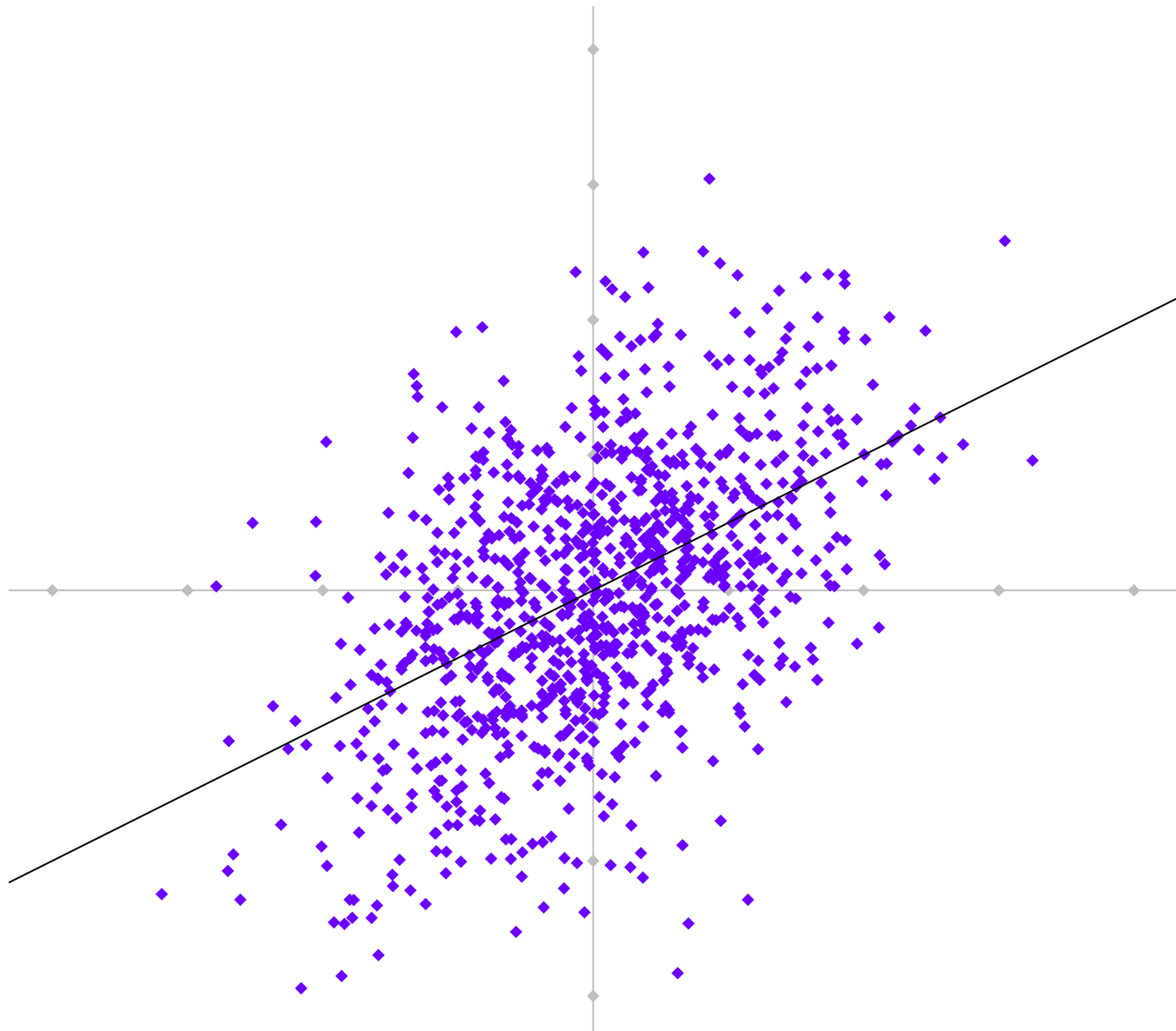
Korrelation = 0.1



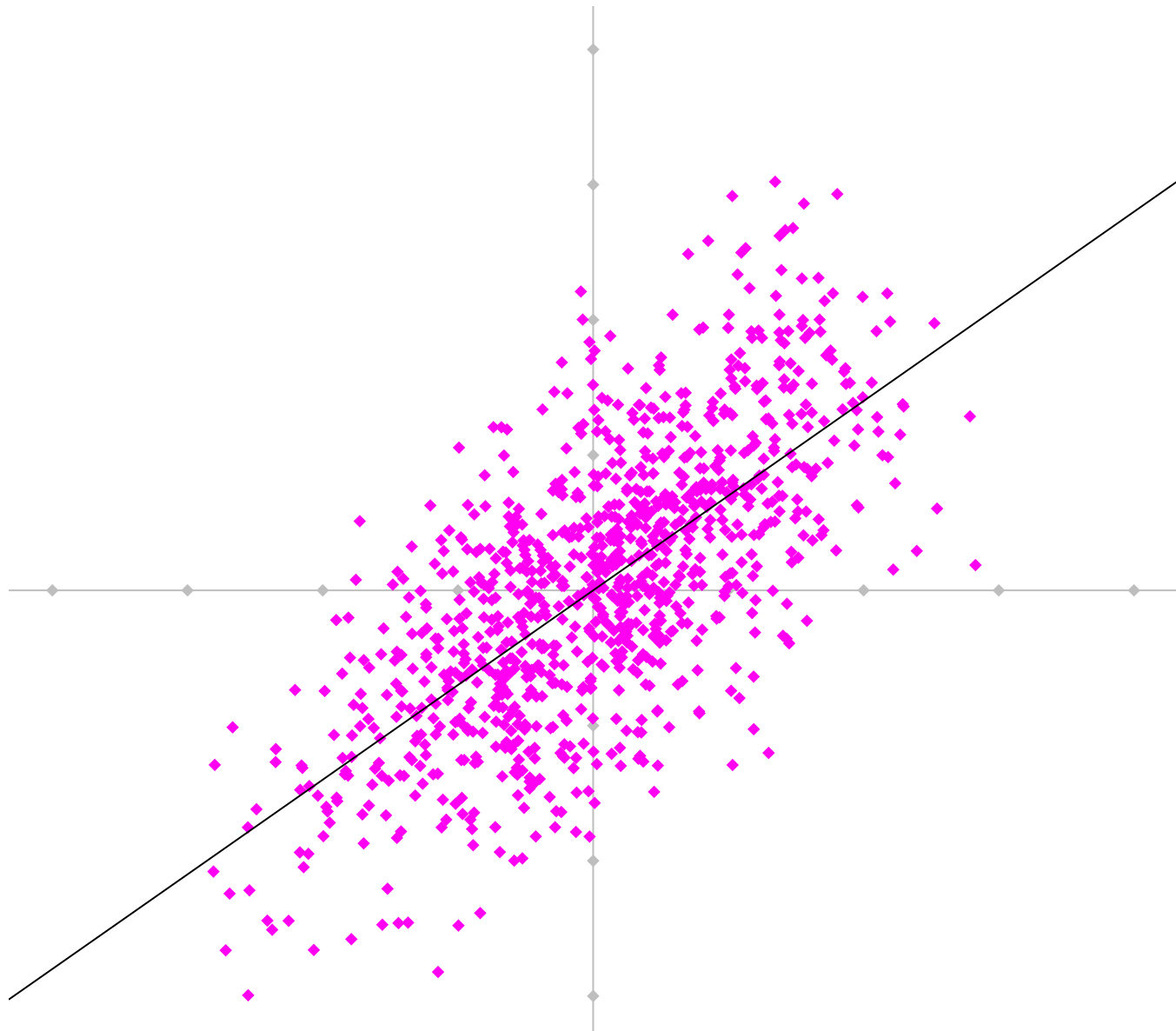
Korrelation = 0.3



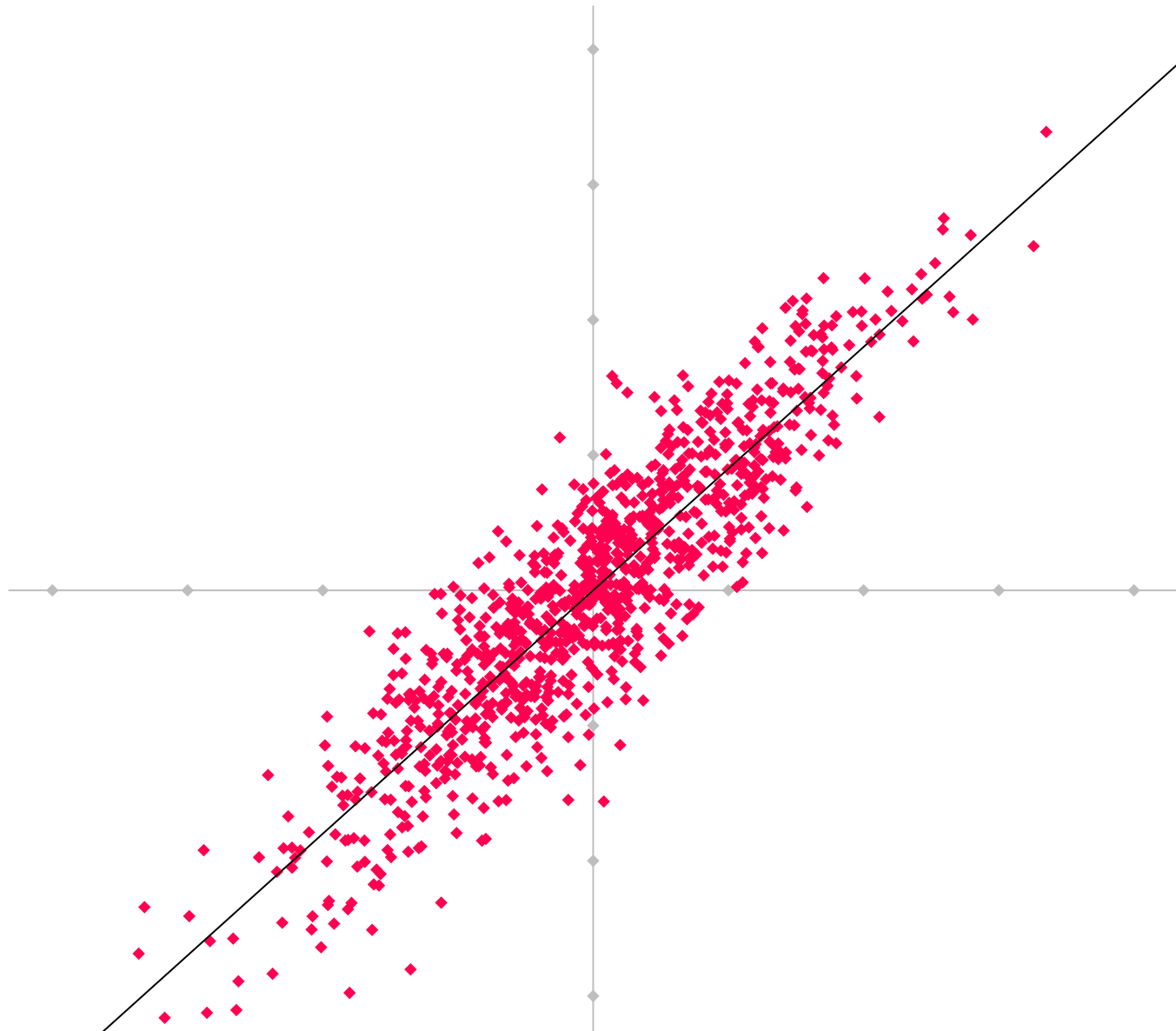
Korrelation = 0.5



Korrelation = 0.7



Korrelation = 0.9



Hier kommt das Rezept,
wie in den 9 vorangegangenen Bildern
jeweils die 1000 Punkte (X_i, Y_i) erzeugt wurden:

Beispiel:

Gemeinsam normalverteilte Zufallsvariable X und Y

mit $\sigma_X^2 = \sigma_Y^2 = 1$ und Korrelationskoeffizient κ :

Z_1, Z_2 seien unabhängig und standard-normalverteilt.

$$X := Z_1, \quad Y := \kappa Z_1 + \sqrt{1 - \kappa^2} Z_2.$$

Aus der Bilinearität der Kovarianz folgt:

$$\text{Cov}[X, Y] = \text{Cov}[Z_1, \kappa Z_1] + 0 = \kappa.$$

$$\sigma_X^2 = 1, \quad \sigma_Y^2 = \kappa^2 + (1 - \kappa^2) = 1.$$

$$\text{Also: } \kappa_{XY} = \kappa.$$

Auch Y ist (standard-) normalverteilt
(siehe V7a1 Folie 12).

Eine Interpretation von κ^2 :

Wir werden sehen:

Für jedes zufällige Paar (X, Y) reellwertiger Zufallsvariabler mit endlichen Varianzen gilt:

κ^2 ist ein Maß dafür, um wieviel besser man Y durch eine affin lineare Funktion von X vorhersagen kann:

$$Y = \beta_1 X + \beta_0 + \text{“Fehler”},$$

als durch eine Konstante:

$$Y = c + \text{“Fehler”}.$$

(Die “Güte der Vorhersage” bezieht sich auf die Kleinheit des **erwarteten quadrierten “Fehlers” (mean square error).**)

Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 3

Beste affin lineare Vorhersage:
die Regressionsgerade

(Buch S. 62-64)

Wie am Ende von Teil 2 angekündigt, wollen wir einsehen:

Das Quadrat des Korrelationskoeffizienten ist ein Maß dafür,
um wieviel besser man Y
durch eine affin lineare Funktion von X vorhersagen kann:

$$Y = \beta_1 X + \beta_0 + \text{“Fehler”},$$

als durch eine Konstante:

$$Y = c + \text{“Fehler”}.$$

Dazu fragen wir erst einmal:

Durch welche **Konstante** wird die Zufallsvariable Y
(im Sinn des erwarteten quadratischen Fehlers)
am besten vorhergesagt?

Die Antwort ist:

Durch ihren **Erwartungswert $E[Y]$** !

Denn:

$$\mathbf{E}[(Y - c)^2] =$$

$$\mathbf{E}[(Y - c)^2] = \mathbf{E}[(Y - \mu_Y + \mu_Y - c)^2]$$

$$= \mathbf{E}[(Y - \mu_Y)^2] + 2\mathbf{E}[(Y - \mu_Y)(\mu_Y - c)] + (\mu_Y - c)^2$$

$$= \sigma_Y^2 + 0 + (\mu_Y - c)^2.$$

Das wird minimiert von

$$c = \mu_Y$$

und hat den Minimalwert

$$\sigma_Y^2.$$

Jetzt fragen wir:

Durch welche **affin lineare Funktion** von X ,

$$\beta_1 X + \beta_0,$$

wird die Zufallsvariable Y

(wieder im Sinn des erwarteten quadratischen Fehlers)

am besten vorhergesagt?

Genauer:

Für welche Zahlen β_1, β_0 wird

$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2]$ minimal?

Wie wir gleich sehen werden, ist die Lösung:

$$\beta_1 := \frac{\sigma_Y}{\sigma_X} \kappa_{XY}$$

und β_0 so, dass $\mu_Y = \beta_1 \mu_X + \beta_0$.

M. a. W.: β_0 so, dass der Punkt (μ_X, μ_Y)
auf der Geraden $y = \beta_1 x + \beta_0$ liegt.

Wir nennen diese Gerade
die **Regressionsgerade** für Y auf der Basis von X .

Wir nennen diese Gerade
die **Regressionsgerade** für Y auf der Basis von X .

Wir begründen jetzt die Behauptung über β_0 und β_1 :

Wir begründen jetzt die Behauptung über β_0 und β_1 :

$$\begin{aligned} & \mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \\ &= \mathbf{Var}[Y - \beta_1 X - \beta_0] + (\mathbf{E}[Y - \beta_1 X - \beta_0])^2 \\ &= \mathbf{Var}[Y - \beta_1 X] + (\mu_Y - \beta_1 \mu_X - \beta_0)^2 \end{aligned}$$

Wir begründen jetzt die Behauptung über β_0 und β_1 :

$$\begin{aligned} & \mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \\ &= \mathbf{Var}[Y - \beta_1 X - \beta_0] + (\mathbf{E}[Y - \beta_1 X - \beta_0])^2 \\ &= \mathbf{Var}[Y - \beta_1 X] + (\mu_Y - \beta_1 \mu_X - \beta_0)^2 \end{aligned}$$

Der **zweite Summand** ist Null für $\beta_0 = \mu_Y - \beta_1 \mu_X$.

Damit haben wir schon mal **die eine Bedingung** gefunden.

Für welches β_1 wird **der erste Summand** minimal?

Im Rest des Abschnittes schreiben wir um der besseren Lesbarkeit willen κ statt κ_{XY} .

$$\text{Var}[Y - \beta_1 X] = \text{Var}[Y] - 2\beta_1 \text{Cov}[X, Y] + \beta_1^2 \text{Var}[X]$$

$$= \sigma_Y^2 - 2\beta_1 \kappa \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2$$

$$= \sigma_Y^2 - \sigma_Y^2 \kappa^2 + (\sigma_Y^2 \kappa^2 - 2\beta_1 \kappa \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2)$$

$$\mathbf{Var}[Y - \beta_1 X] = \mathbf{Var}[Y] - 2\beta_1 \mathbf{Cov}[X, Y] + \beta_1^2 \mathbf{Var}[X]$$

$$= \sigma_Y^2 - 2\beta_1 \kappa \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2$$

$$= \sigma_Y^2 - \sigma_Y^2 \kappa^2 + (\sigma_Y \kappa - \beta_1 \sigma_X)^2$$

Der rechte Summand wird Null für

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \kappa.$$

Und der Minimalwert von $\mathbf{Var}[Y - \beta_1 X]$ ist $\sigma_Y^2(1 - \kappa^2)$.

Damit ist auch der Minimalwert von $\text{Var}[Y - \beta_1 X - \beta_0 \cdot 1]$ gleich $\sigma_Y^2(1 - \kappa^2)$.

Der Minimalwert von $\text{Var}[Y - c \cdot 1]$ war σ_Y^2 .

Die **Verbesserung der Approximation** (“Vorhersage”) von Y im quadratischen Mittel, wenn man zu den Vielfachen von 1 die Vielfachen von X dazunimmt, **beträgt**

$$\sigma_Y^2 - \sigma_Y^2(1 - \kappa^2) = \kappa^2 \sigma_Y^2.$$

Also ist der Anteil von σ_Y^2 ,

der von den Vielfachen von X zusätzlich zu den Vielfachen von 1 “erklärt” wird, gleich $\kappa^2 \sigma_Y^2$.

Wir halten fest: Die Minimierungsaufgabe

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \stackrel{!}{=} \min$$

für die *beste affin lineare Vorhersage von Y*
auf der Basis von X

(im Sinn des quadratischen Mittels)

hat die Lösung

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \kappa, \quad \mu_Y = \beta_1 \mu_X + \beta_0$$

und den Minimalwert $(1 - \kappa^2) \sigma_Y^2$.

Beispiel: “Welche Gerade passt am besten?”

Die Methode der kleinsten Quadrate.

$(x_1, y_1), \dots, (x_n, y_n)$ seien n verschiedene Punkte im $\mathbb{R}^2$.

Für welche β_0, β_1 wird $\sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2$ minimal?

Beispiel: “Welche Gerade passt am besten?”

Die Methode der kleinsten Quadrate.

$(x_1, y_1), \dots, (x_n, y_n)$ seien n verschiedene Punkte im $\mathbb{R}^2$.

Für welche β_0, β_1 wird $\sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2$ minimal?

Diese Aufgabe interpretieren wir stochastisch. Sei J uniform

verteilt auf $\{1, \dots, n\}$ und $(X, Y) := (x_J, y_J)$, also

$$\mathbf{P}((X, Y) = (x_i, y_i)) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Diese Aufgabe interpretieren wir stochastisch. Sei J uniform verteilt auf $\{1, \dots, n\}$ und $(X, Y) := (x_J, y_J)$, also

$$\mathbf{P}((X, Y) = (x_i, y_i)) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Dann ist

$$\mathbf{E}[X] = \frac{1}{n} \sum x_i =: \bar{x}, \quad \mathbf{E}[Y] = \frac{1}{n} \sum y_i =: \bar{y},$$

$$\sigma_X^2 = \frac{1}{n} \sum (x_i - \bar{x})^2, \quad \sigma_Y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2,$$

$$\mathbf{Cov}[X, Y] = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}),$$

$$\kappa := \kappa_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}.$$

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2$$

wird, wie wir gezeigt haben, minimiert durch

$$\beta_1 := \frac{\sigma_Y}{\sigma_X} \kappa = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

und β_0 so, dass $\bar{y} = \beta_1 \bar{x} + \beta_0$.

Diese Gerade $y = \beta_1 x + \beta_0$ heißt die
Regressionsgerade zu den Punkten (x_i, y_i) , $i = 1, \dots, n$,
 oder auch *die mit der Methode der kleinsten Quadrate
 gefundene Ausgleichsgerade*.

Statistik für Biologen

Korrelation

Wie messen wir
die Stärke
der Abhängigkeit
zwischen Variablen?

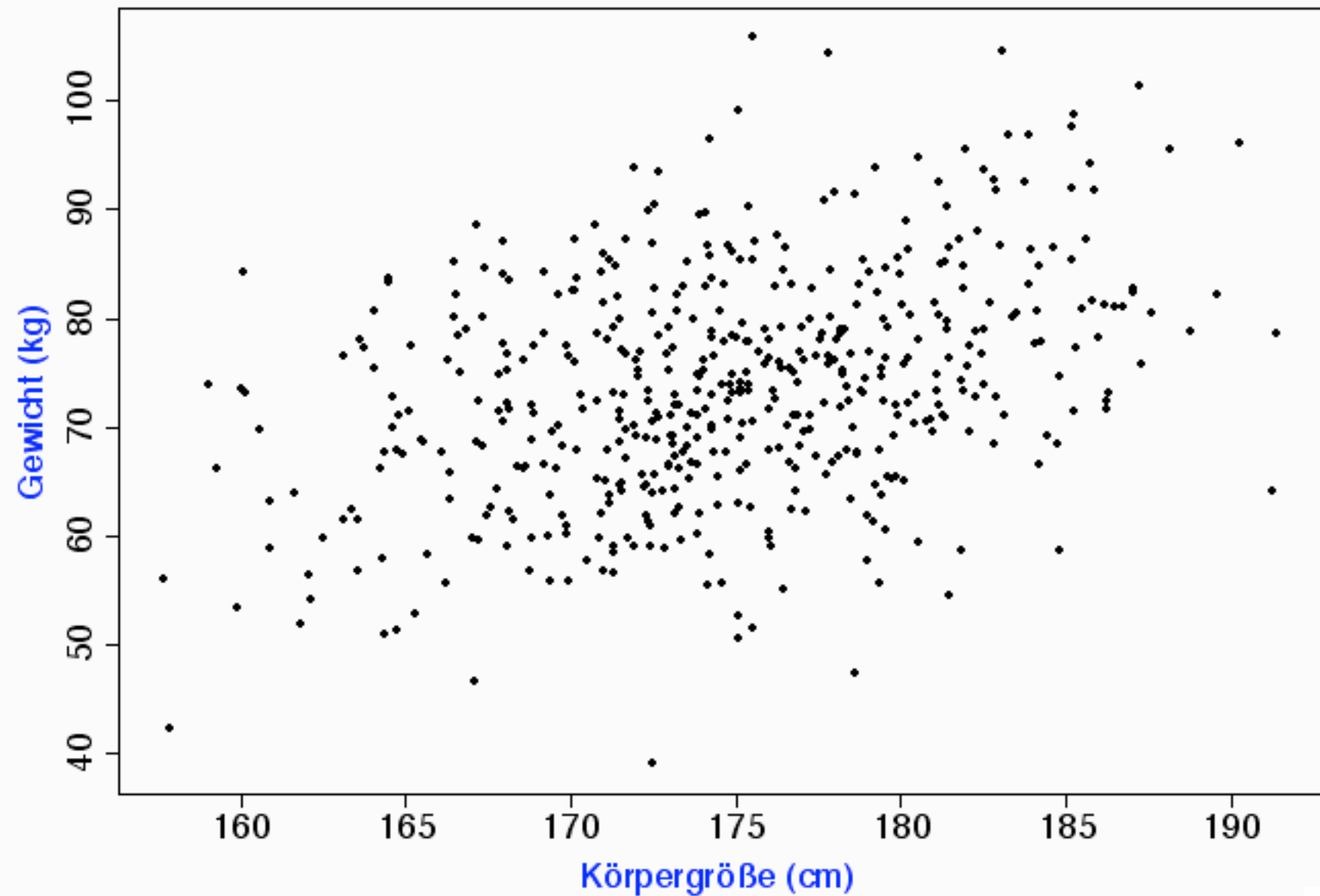
Beispiel:

Körpergröße

und

Gewicht

Männliche US-Amerikaner 18–24 J. (n=500)



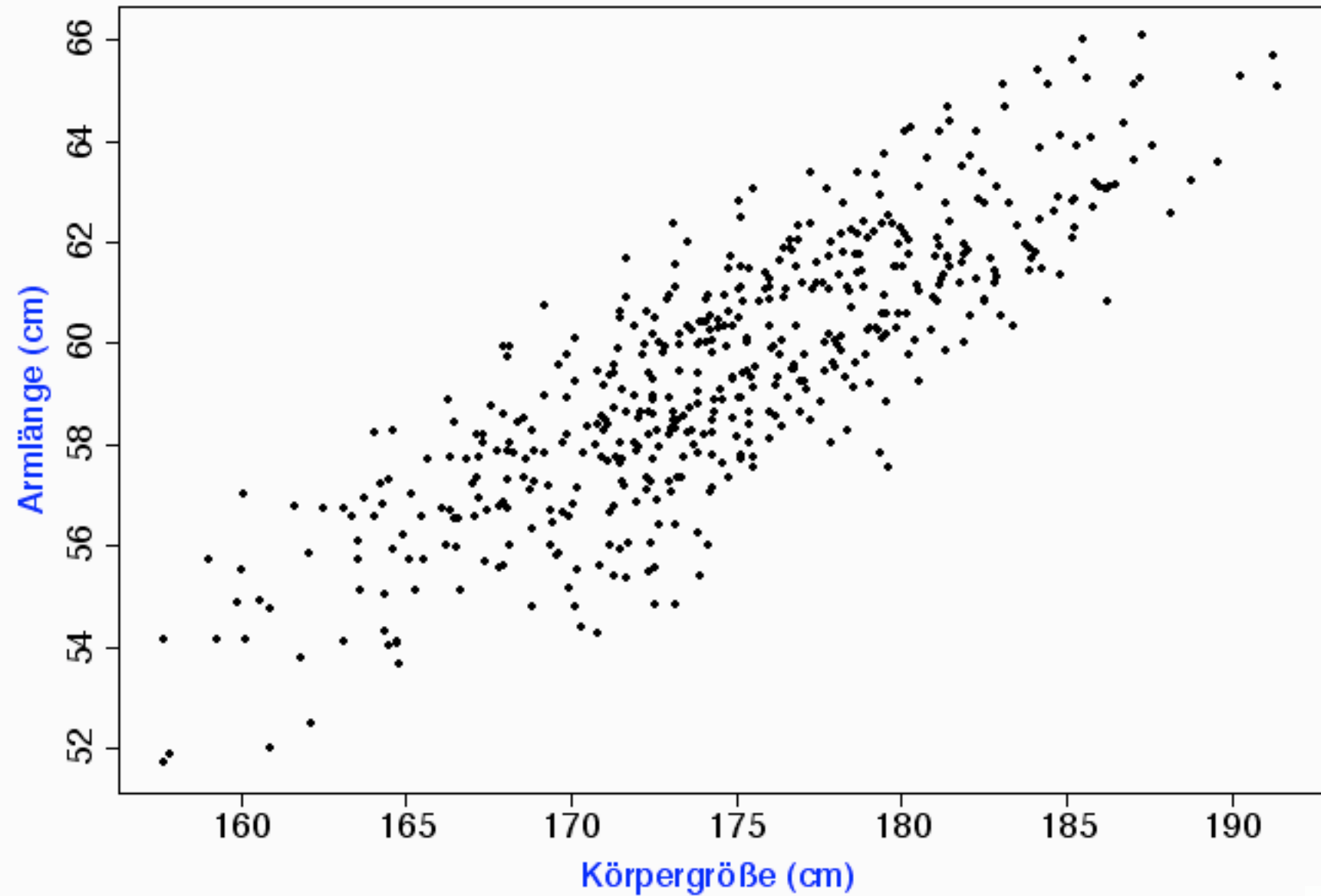
Ein klarer Zusammenhang
aber nicht sehr stark.

Körpergröße

und

Armlänge

Männliche US-Amerikaner 18–24 J. (n=500)

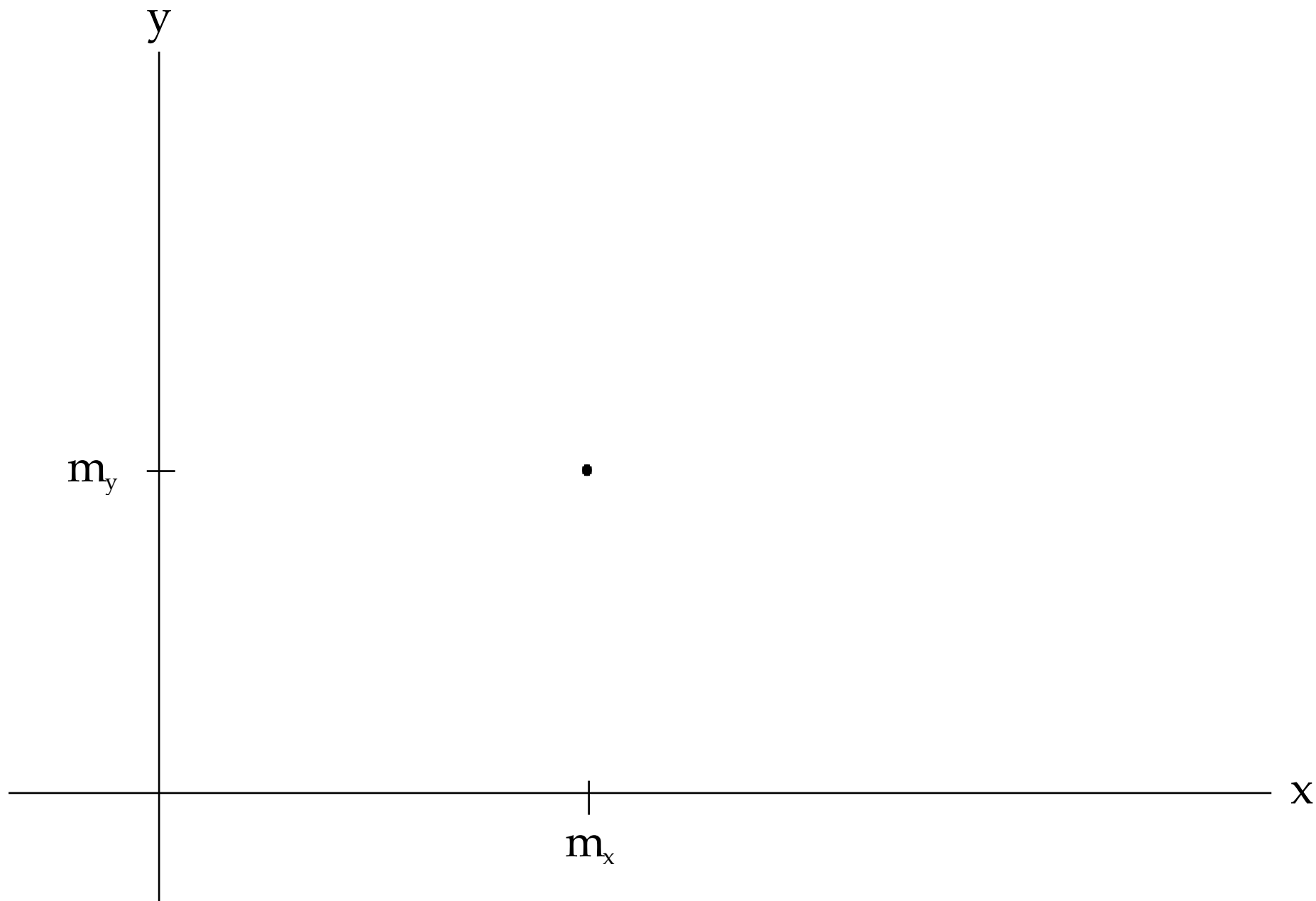


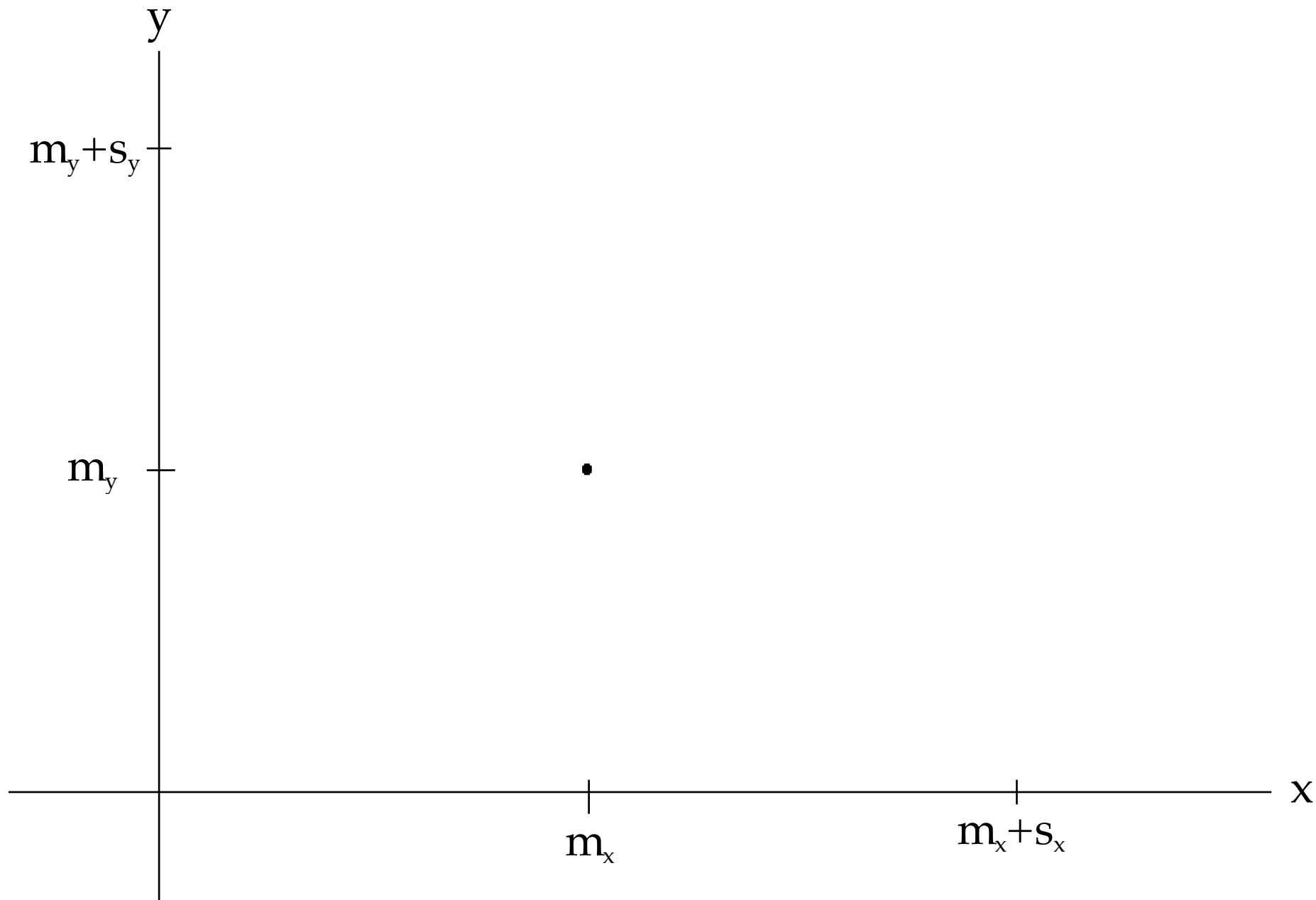
Hier ist der Zusammenhang
viel stärker.

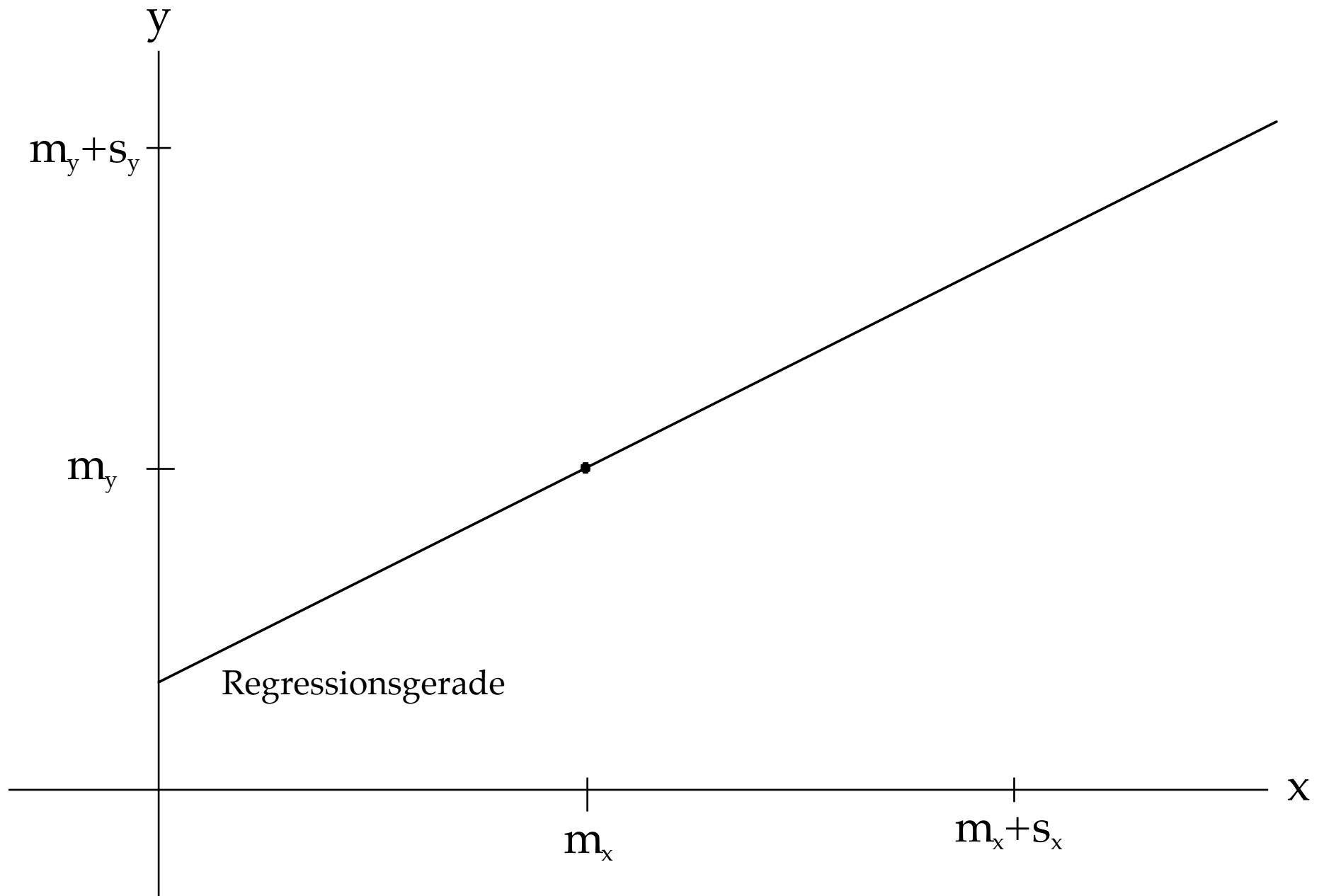
Wie können wir
die Stärke des Zusammenhangs
quantitativ
erfassen?

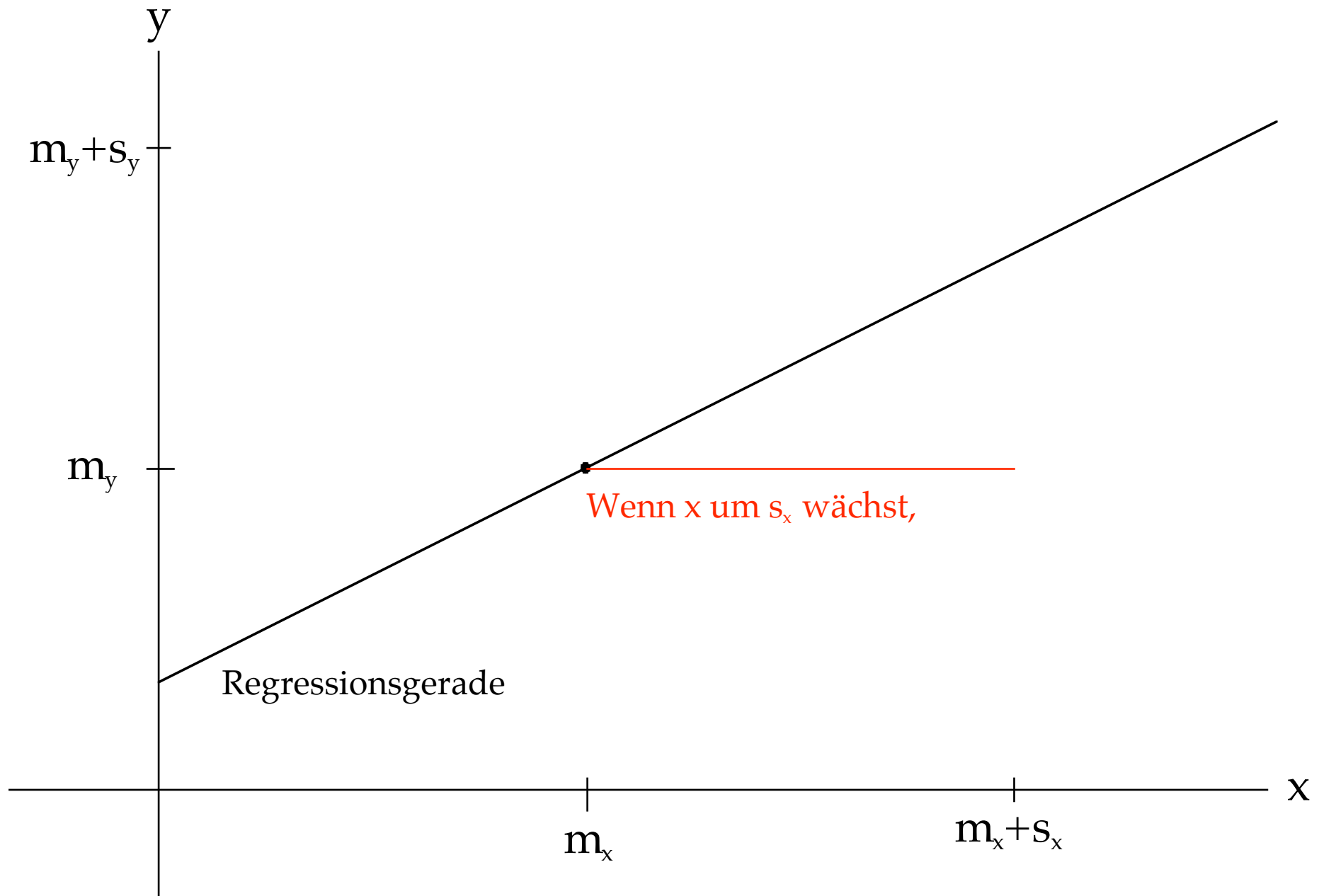
Der Korrelationskoeffizient

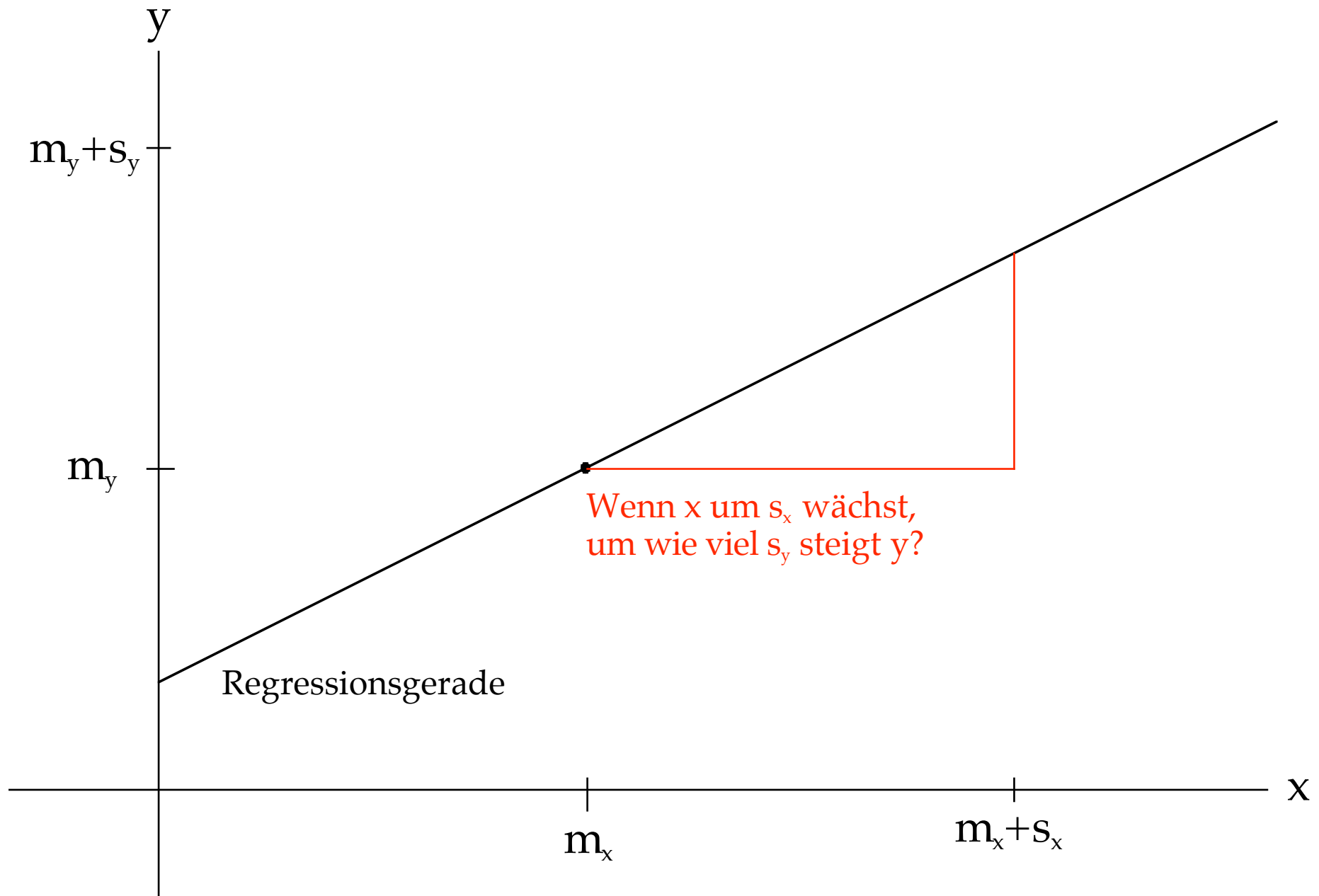
r

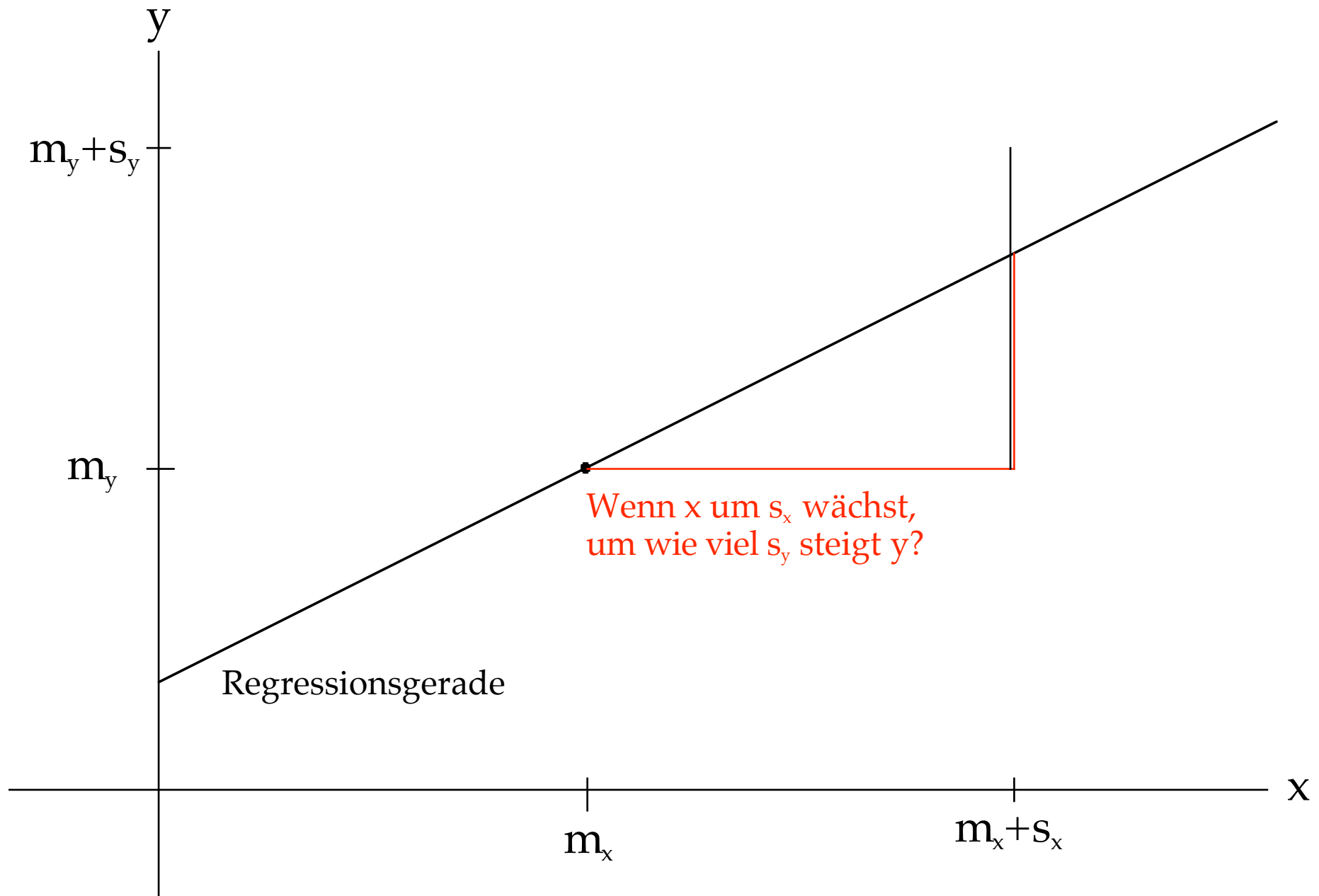


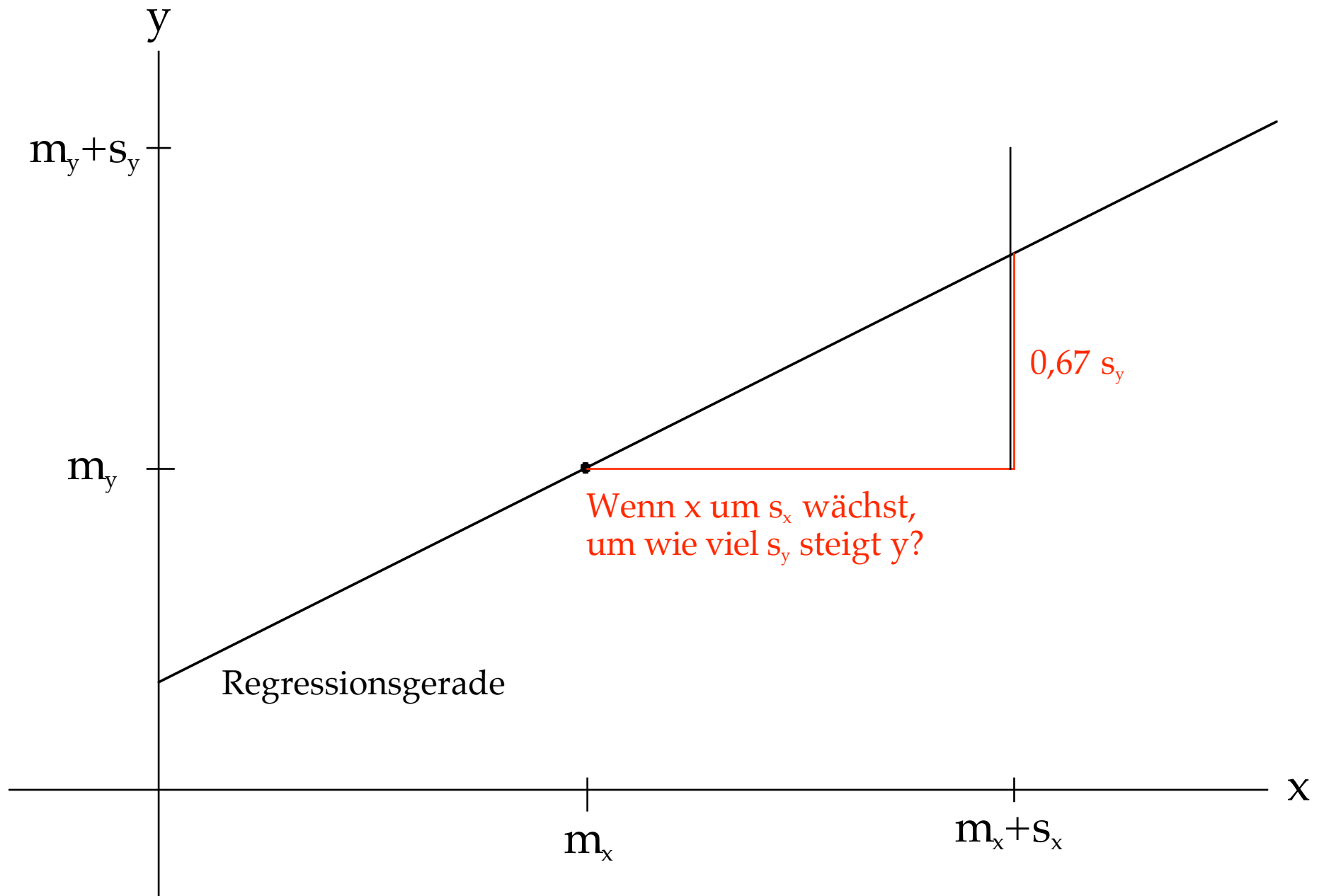


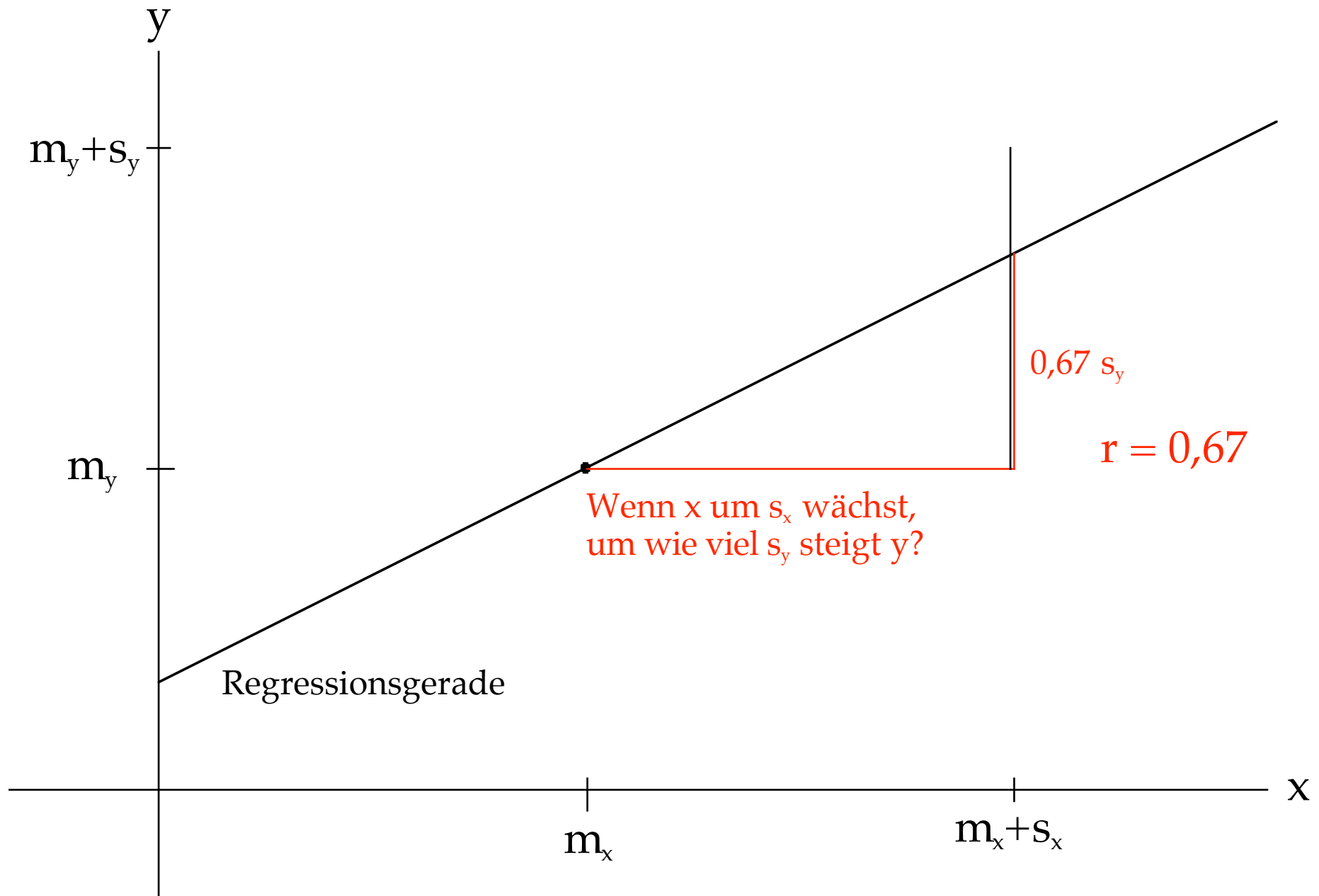












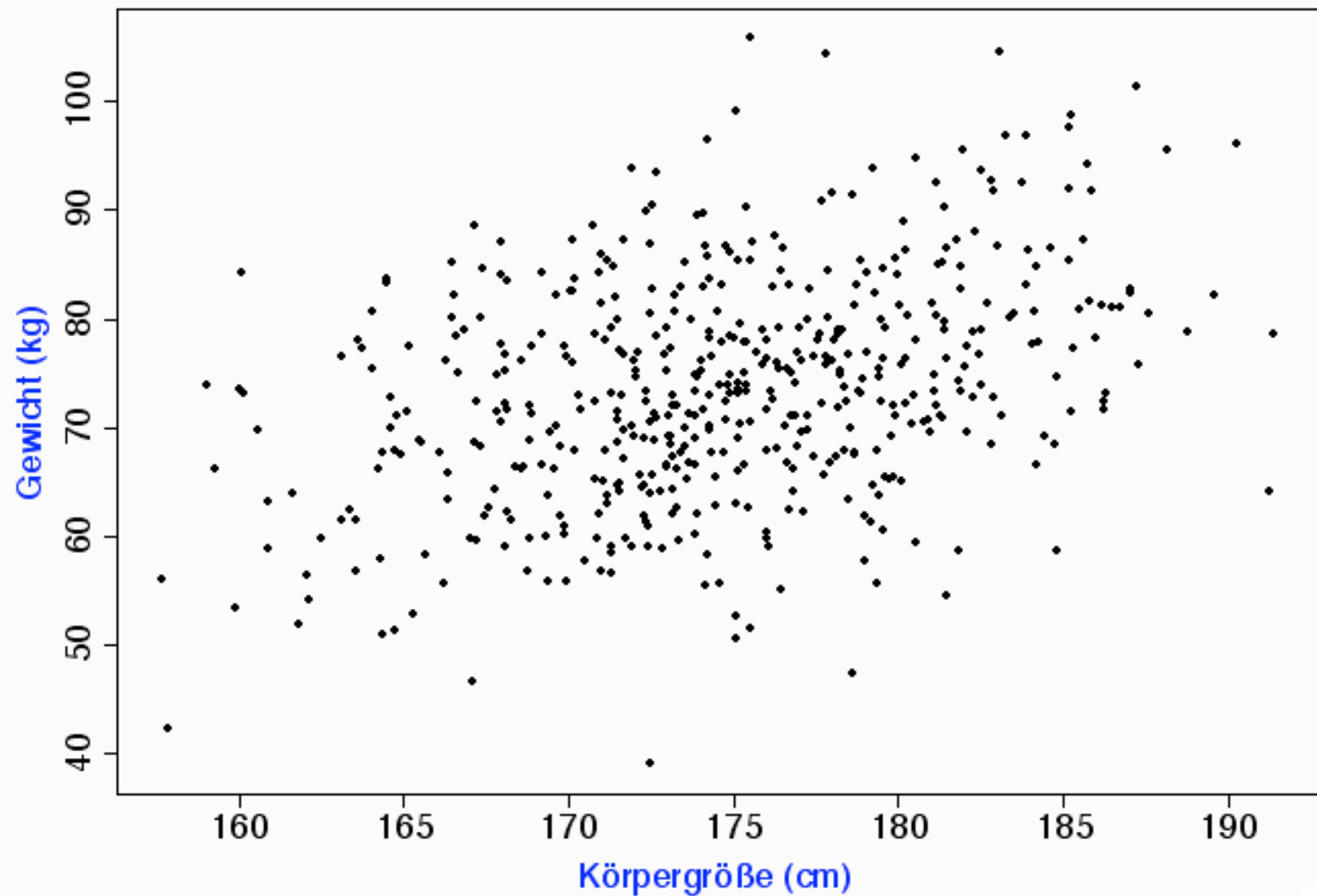
Wenn x
um s_x größer wird,
wird y
im Schnitt um
 $r \cdot s_y$
größer.

Körpergröße

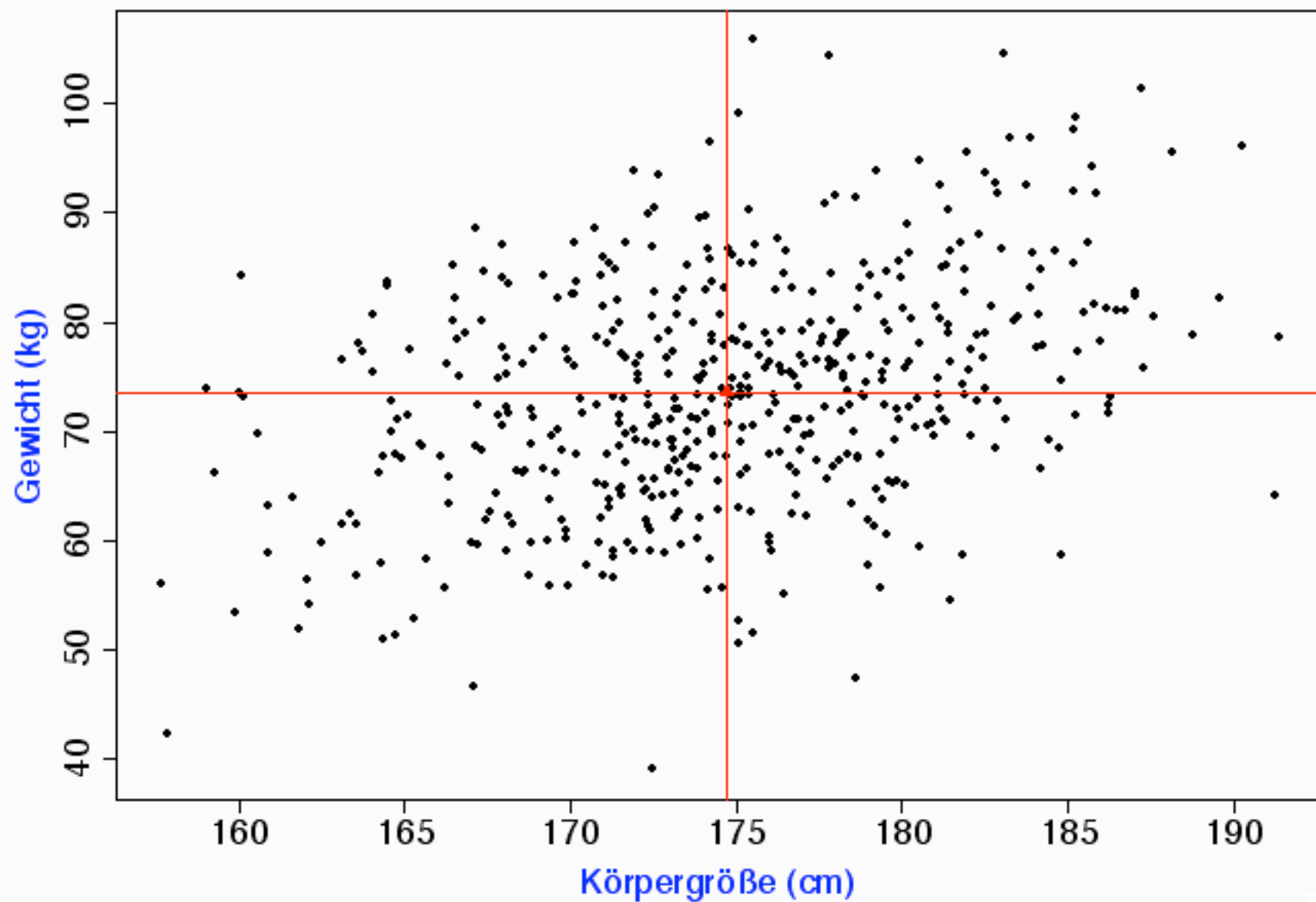
und

Gewicht

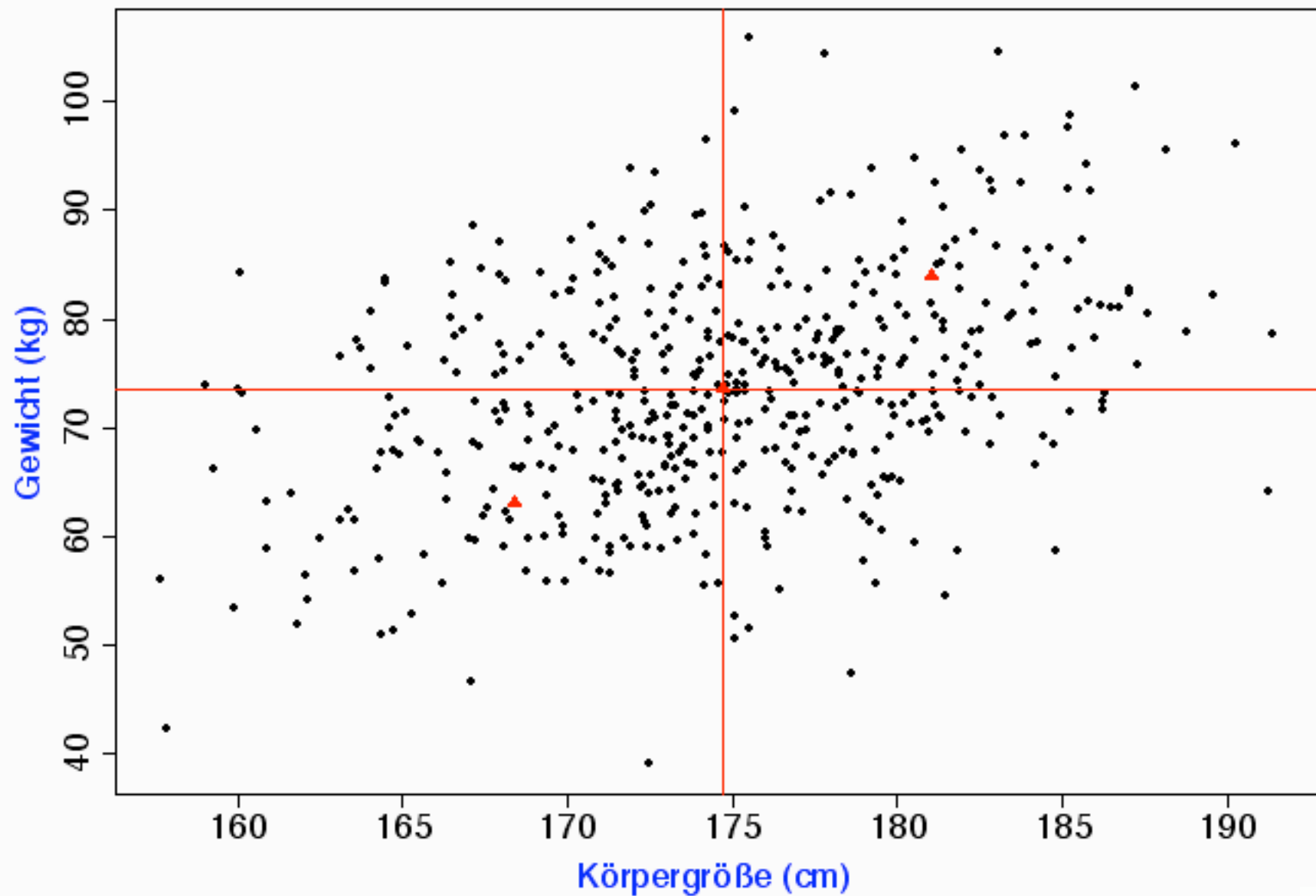
Männliche US-Amerikaner 18–24 J. (n=500)



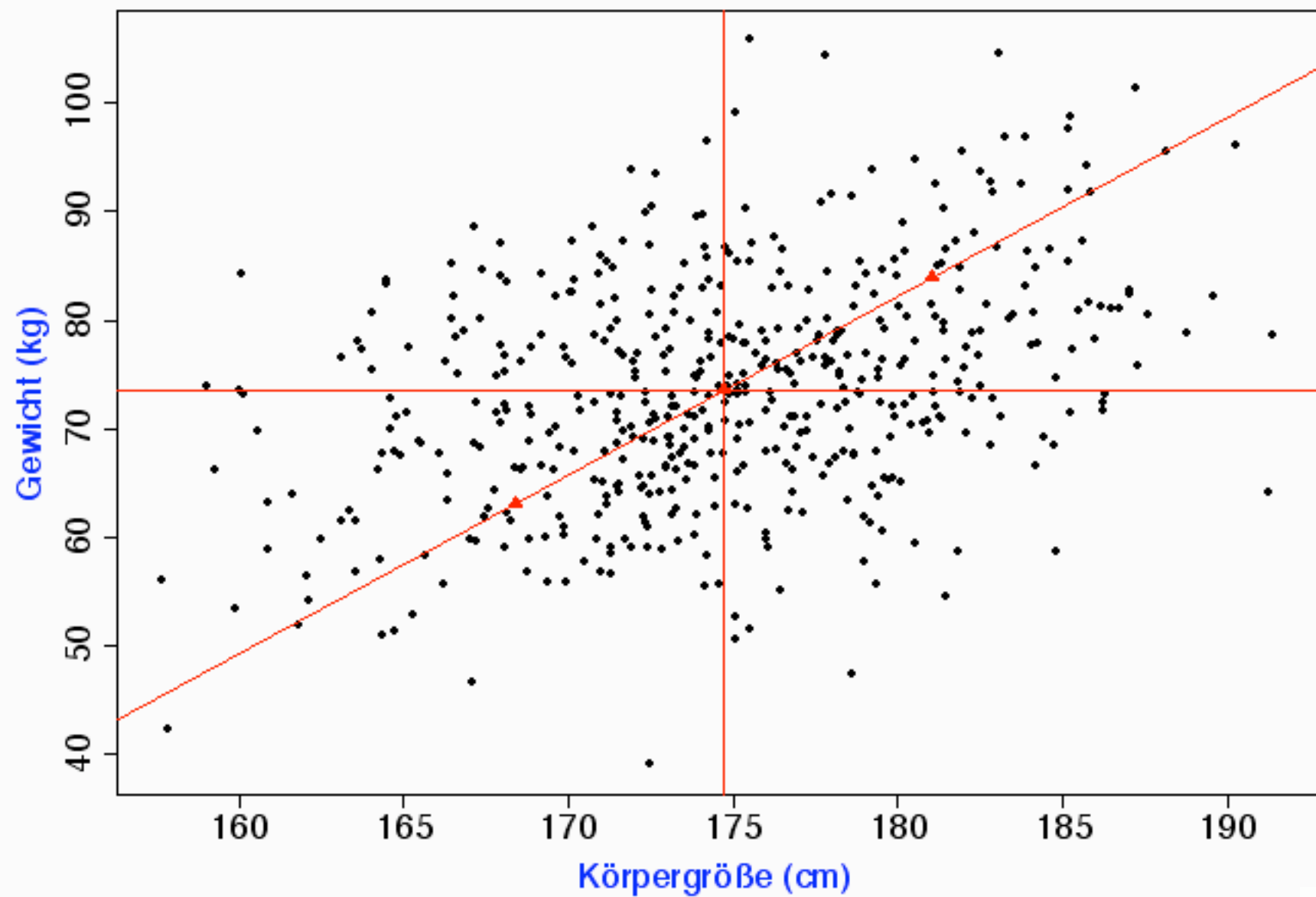
Mittelwerte



Mittelwerte $\pm$ Standardabweichungen



Mittelwerte $\pm$ Standardabweichungen



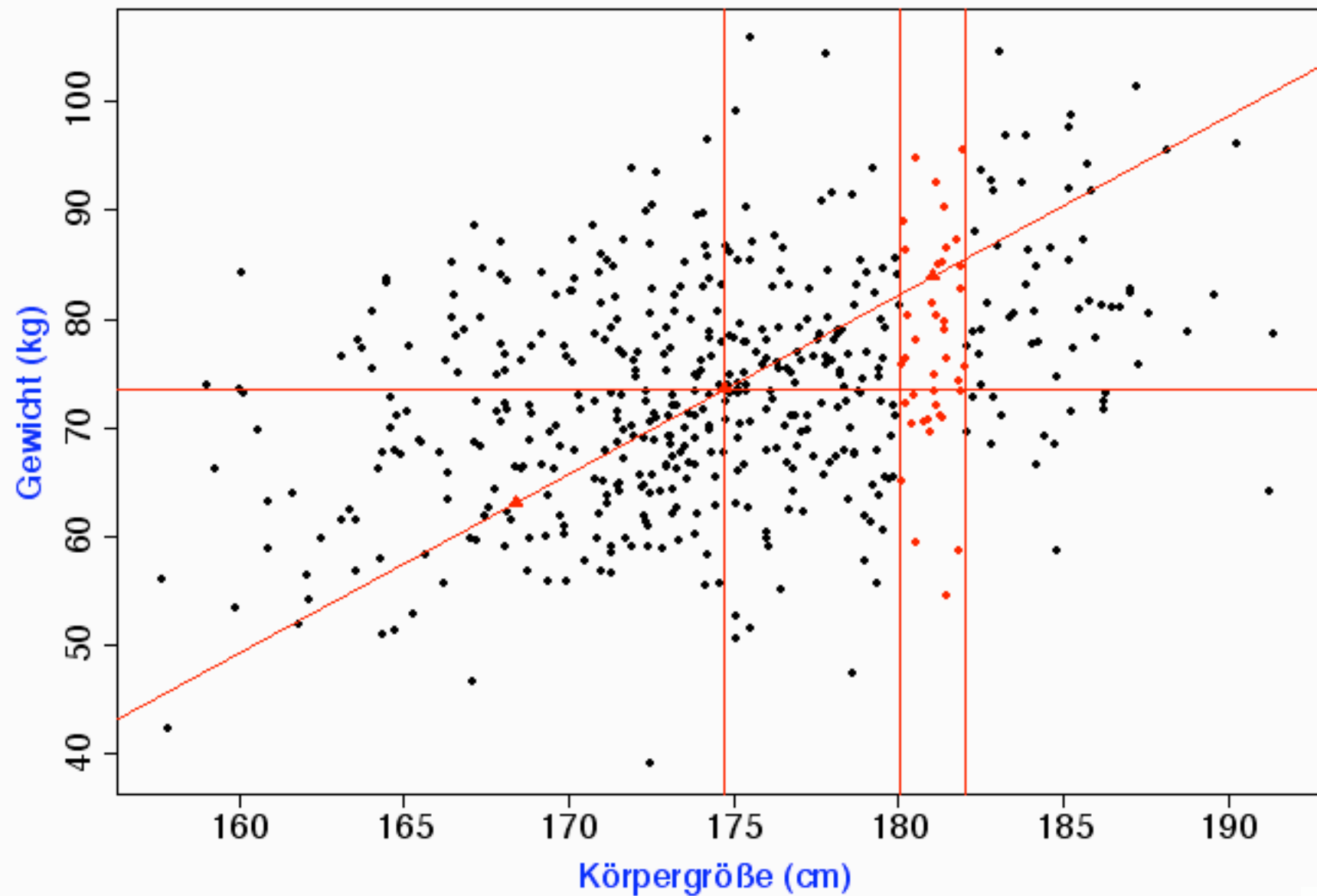
Ein Mann ist
eine Standardabweichung größer
als der Durchschnitt.

Was wiegt er?

(Im Schnitt)

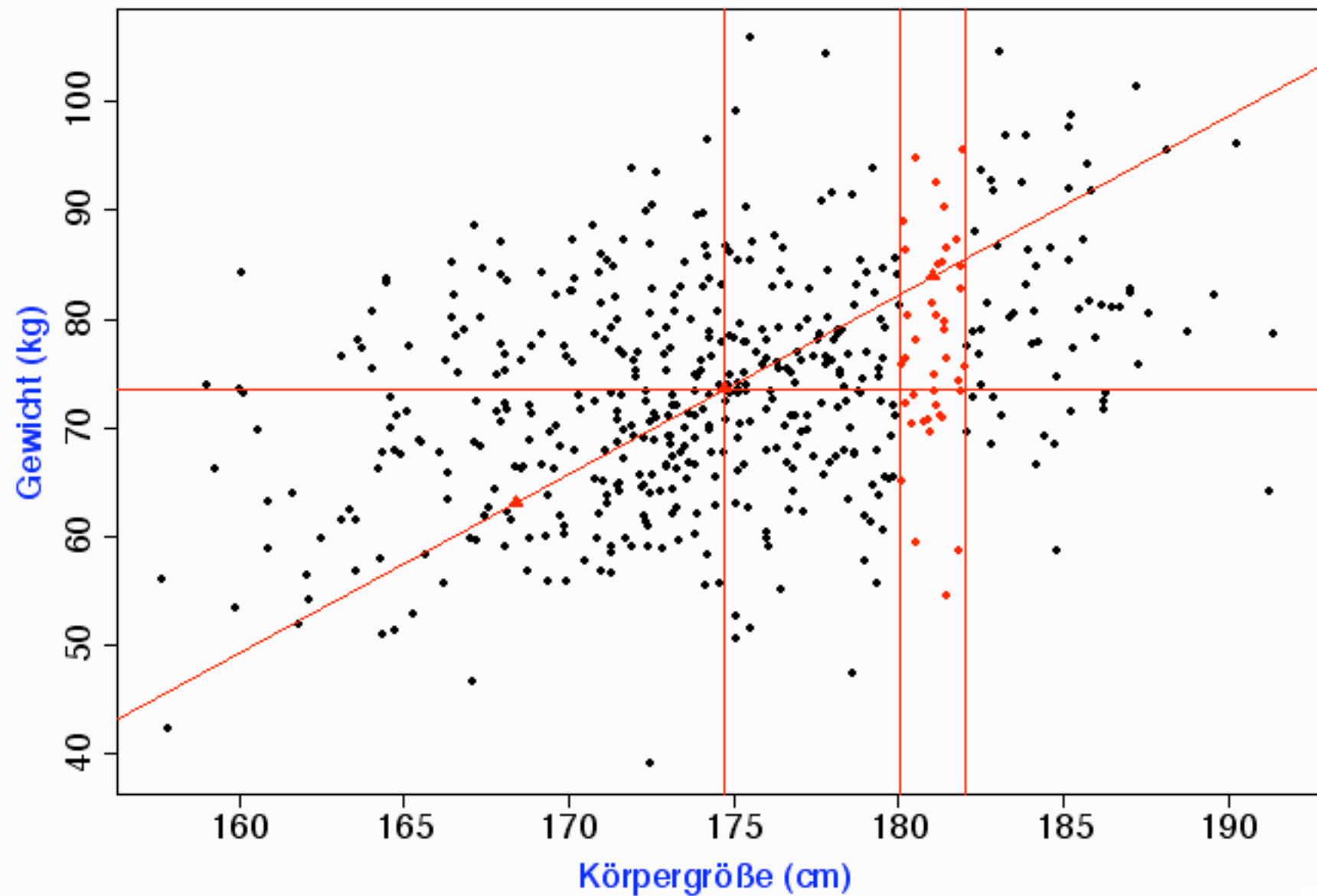
(Wie viele Standardabweichungen
schwerer als der Durchschnitt?)

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



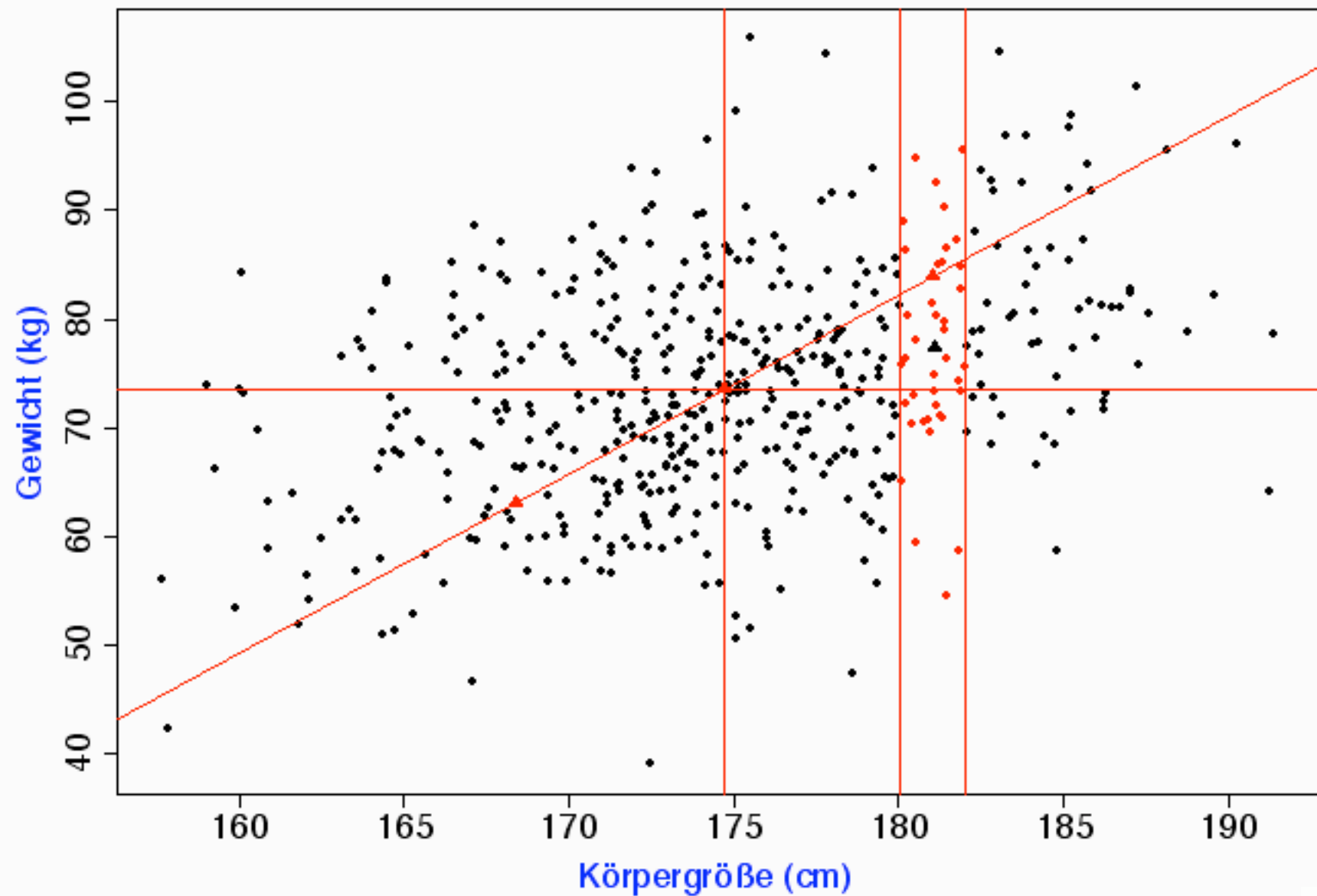
Die meisten Punkte
liegen unterhalb der „Diagonale“.

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



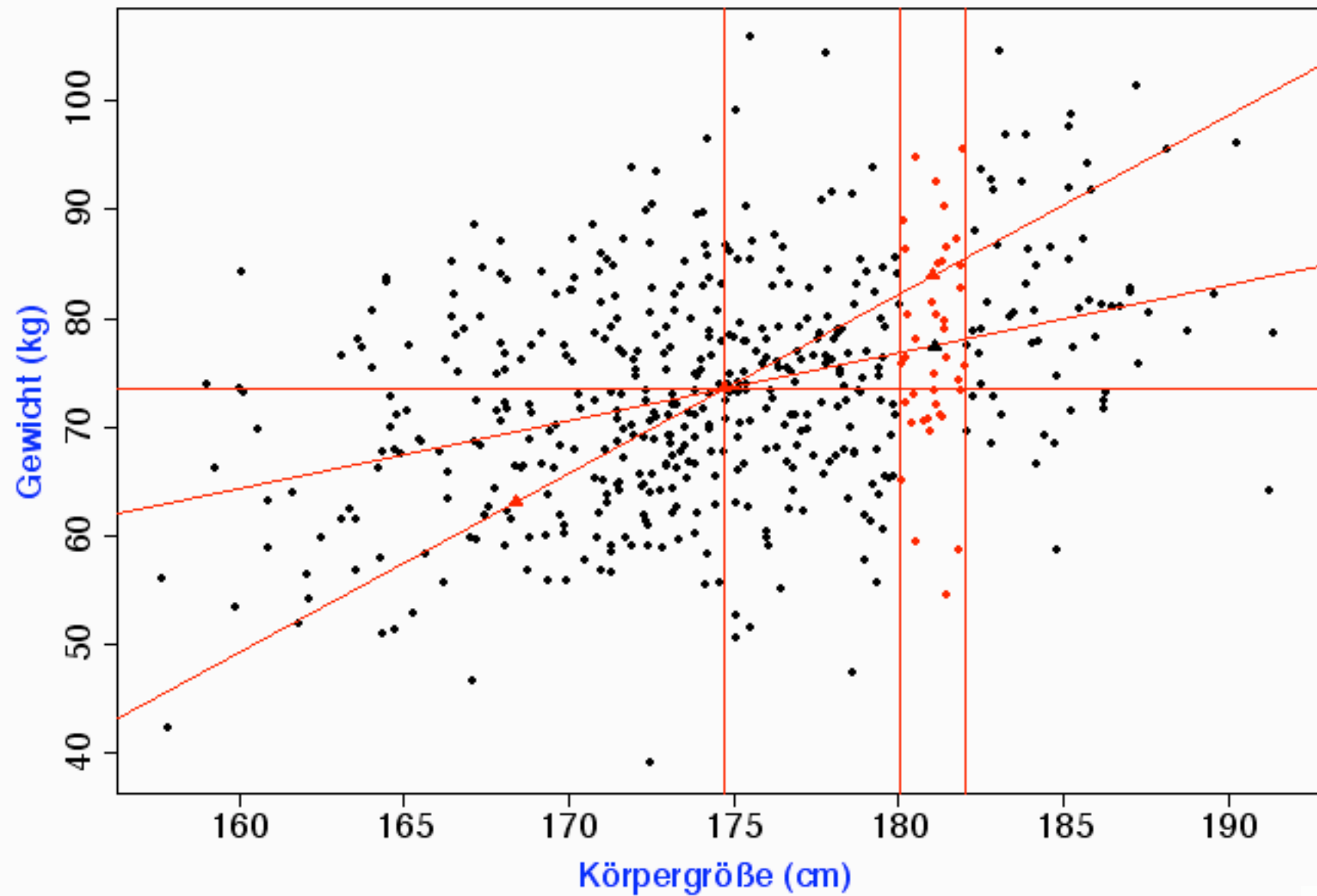
Der Mittelwert auch.

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)

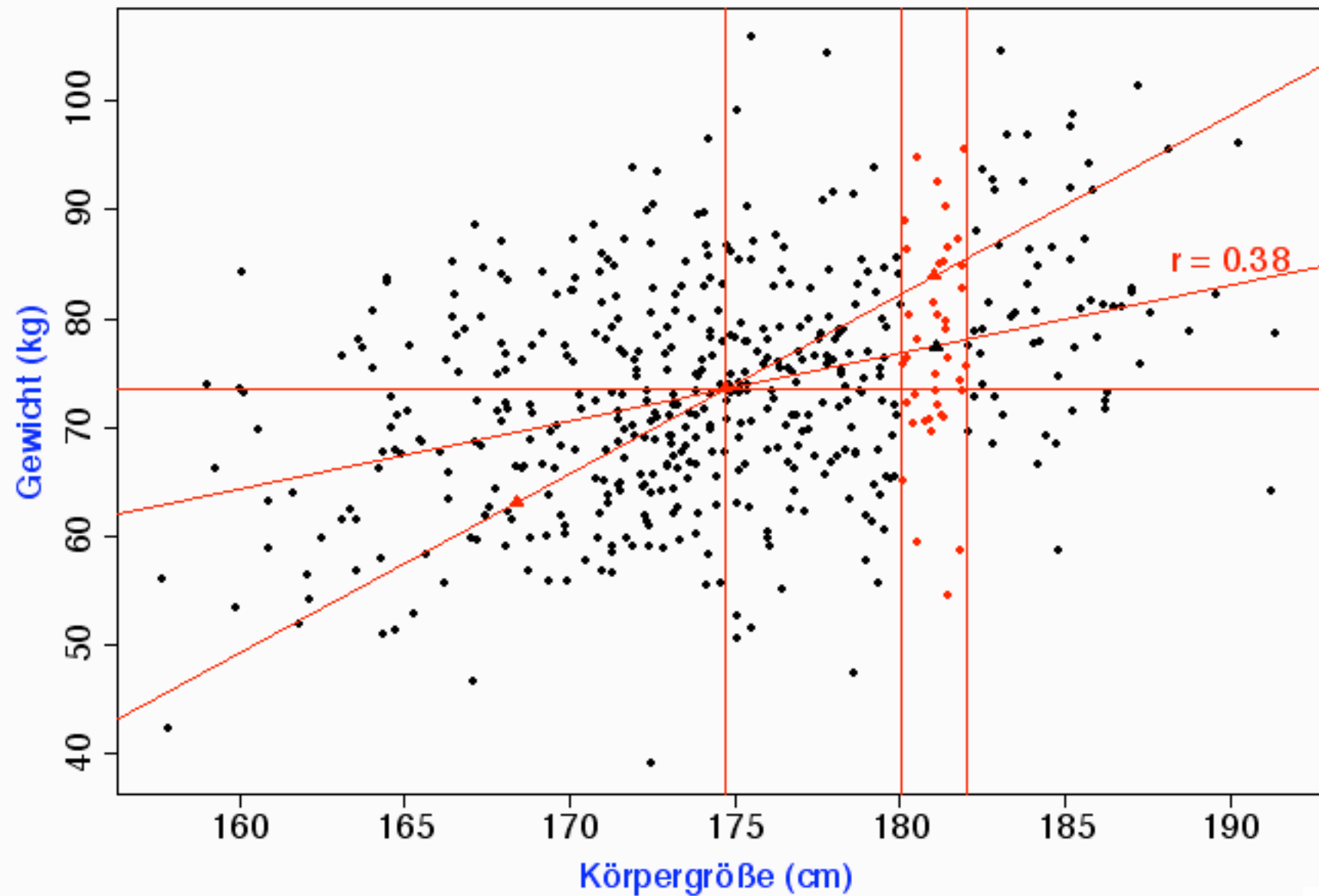


Die Regressionsgerade

Die Regressionsgerade



Die Regressionsgerade



$$r = 0,38$$

INTERPRETATION

Wenn x um s_x größer wird,
wird y im Schnitt
um $0,38 s_y$ größer.

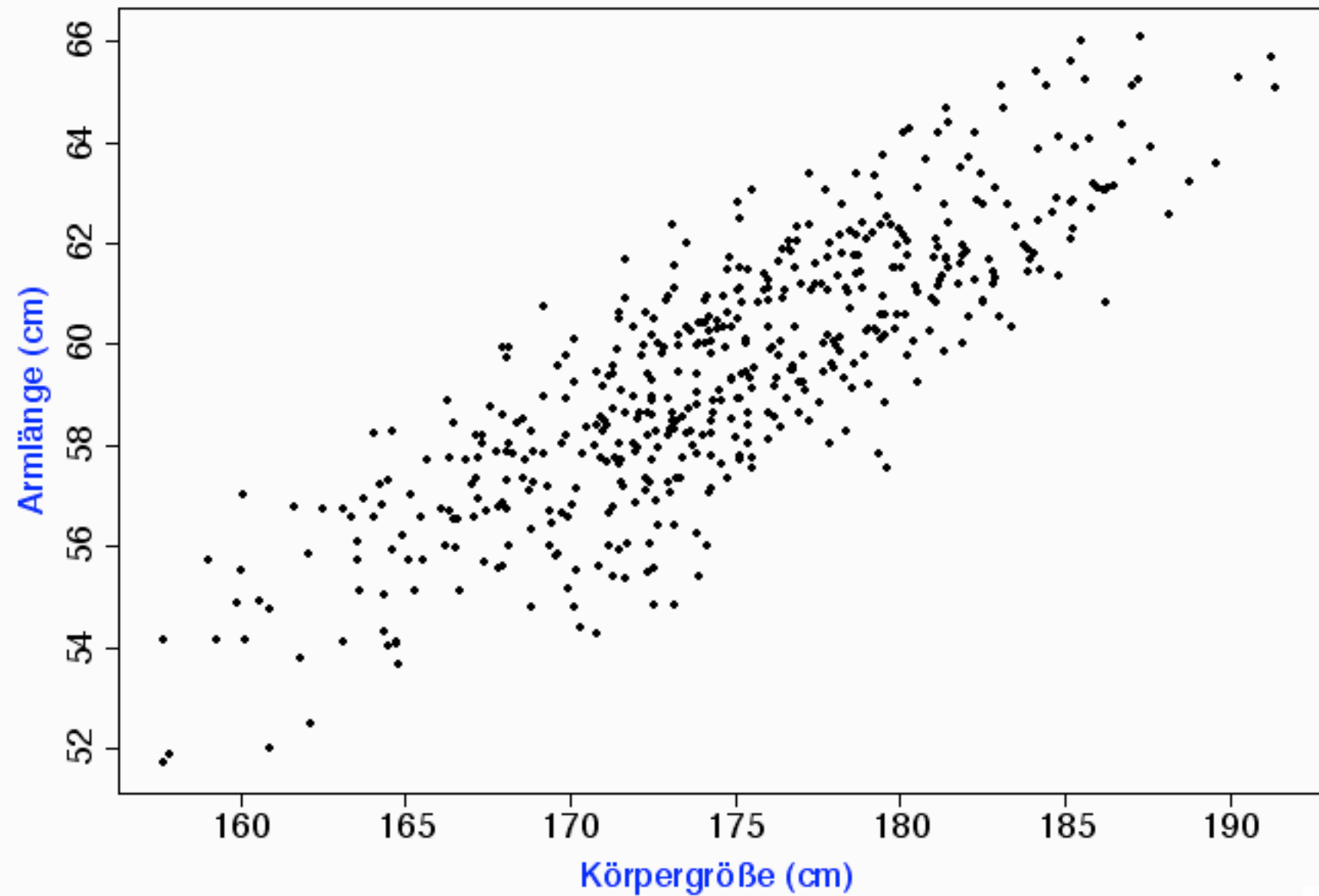
Zum Vergleich:

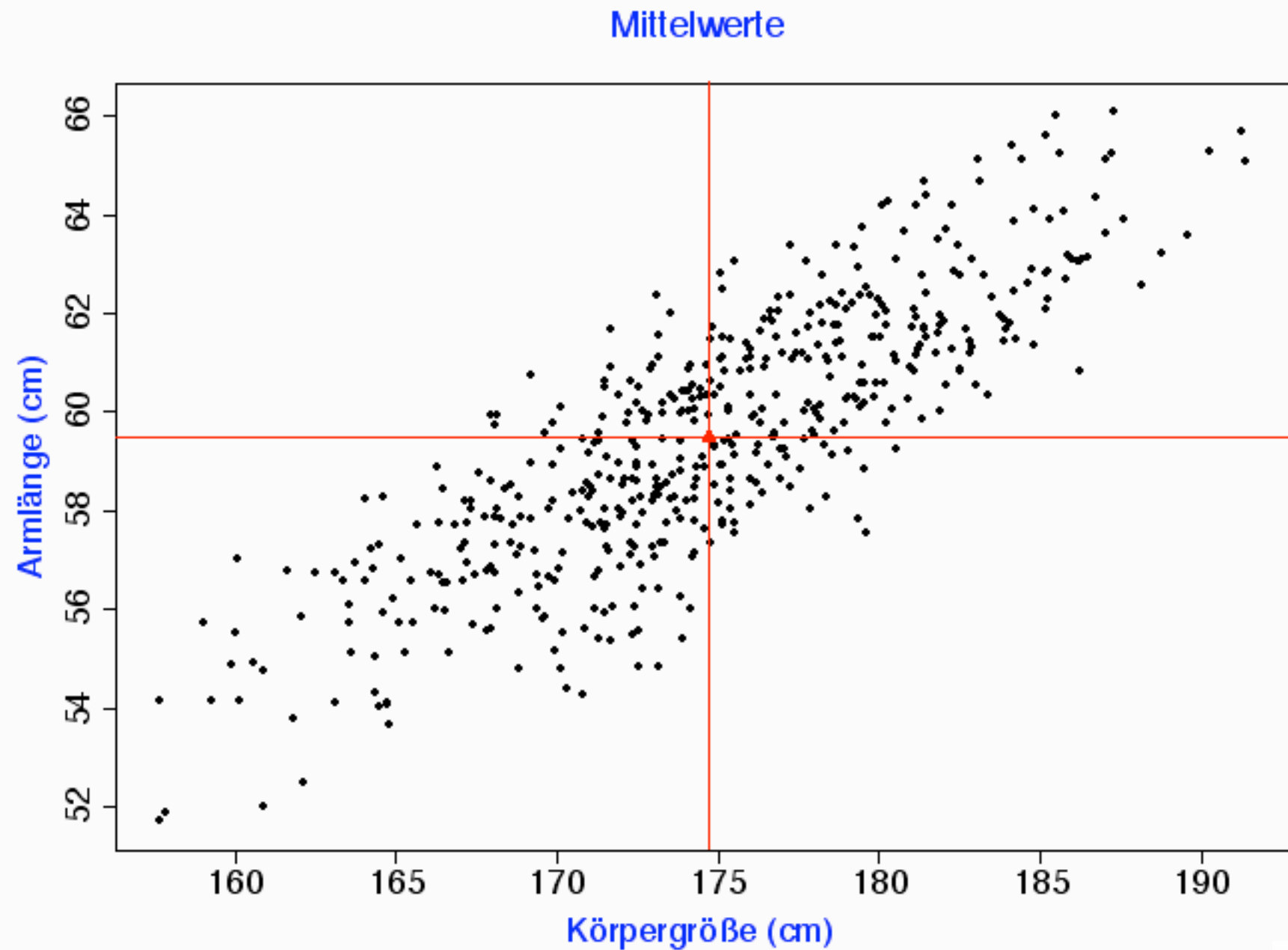
Körpergröße

und

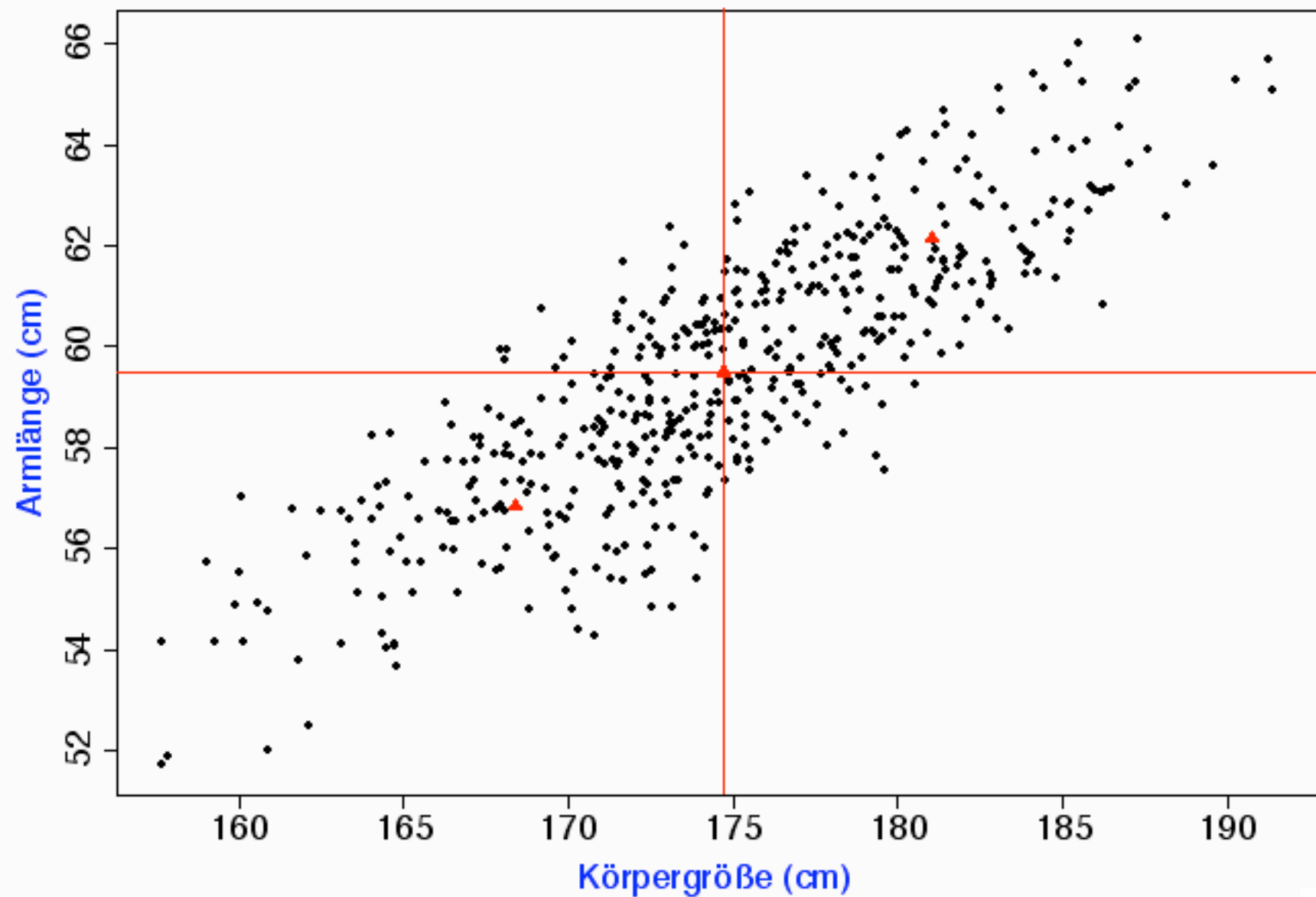
Armlänge

Männliche US-Amerikaner 18–24 J. (n=500)

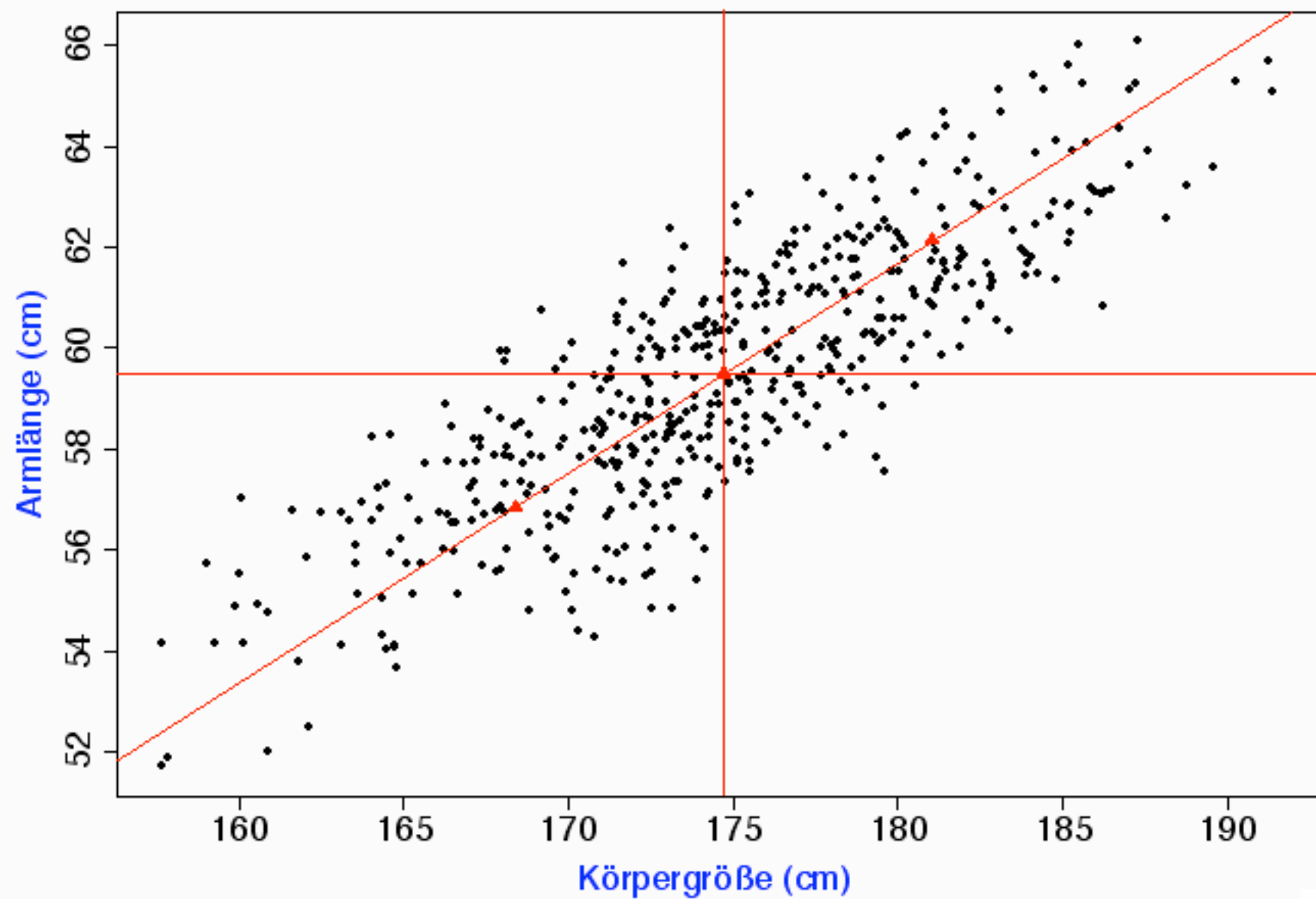




Mittelwerte $\pm$ Standardabweichungen

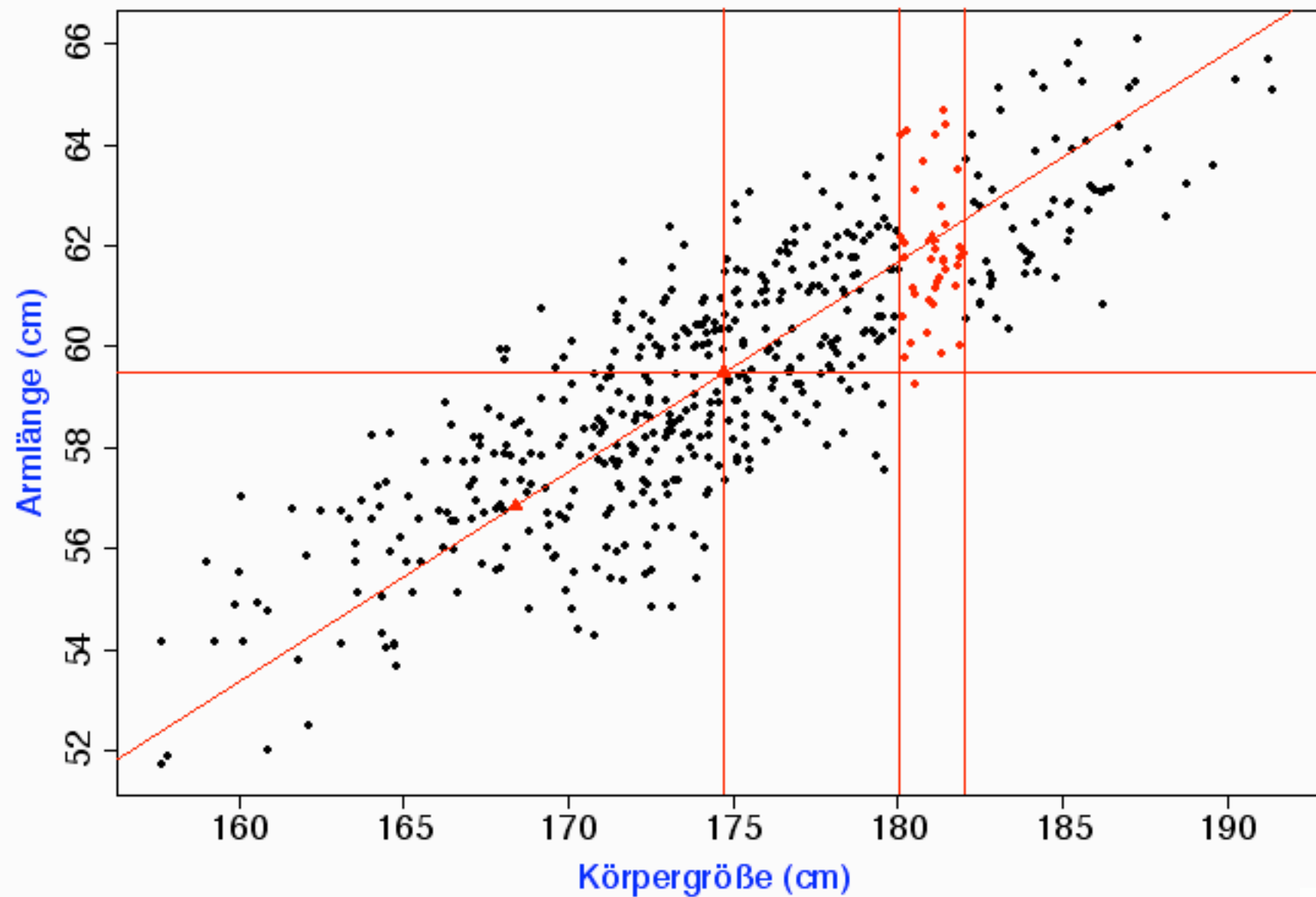


Mittelwerte $\pm$ Standardabweichungen

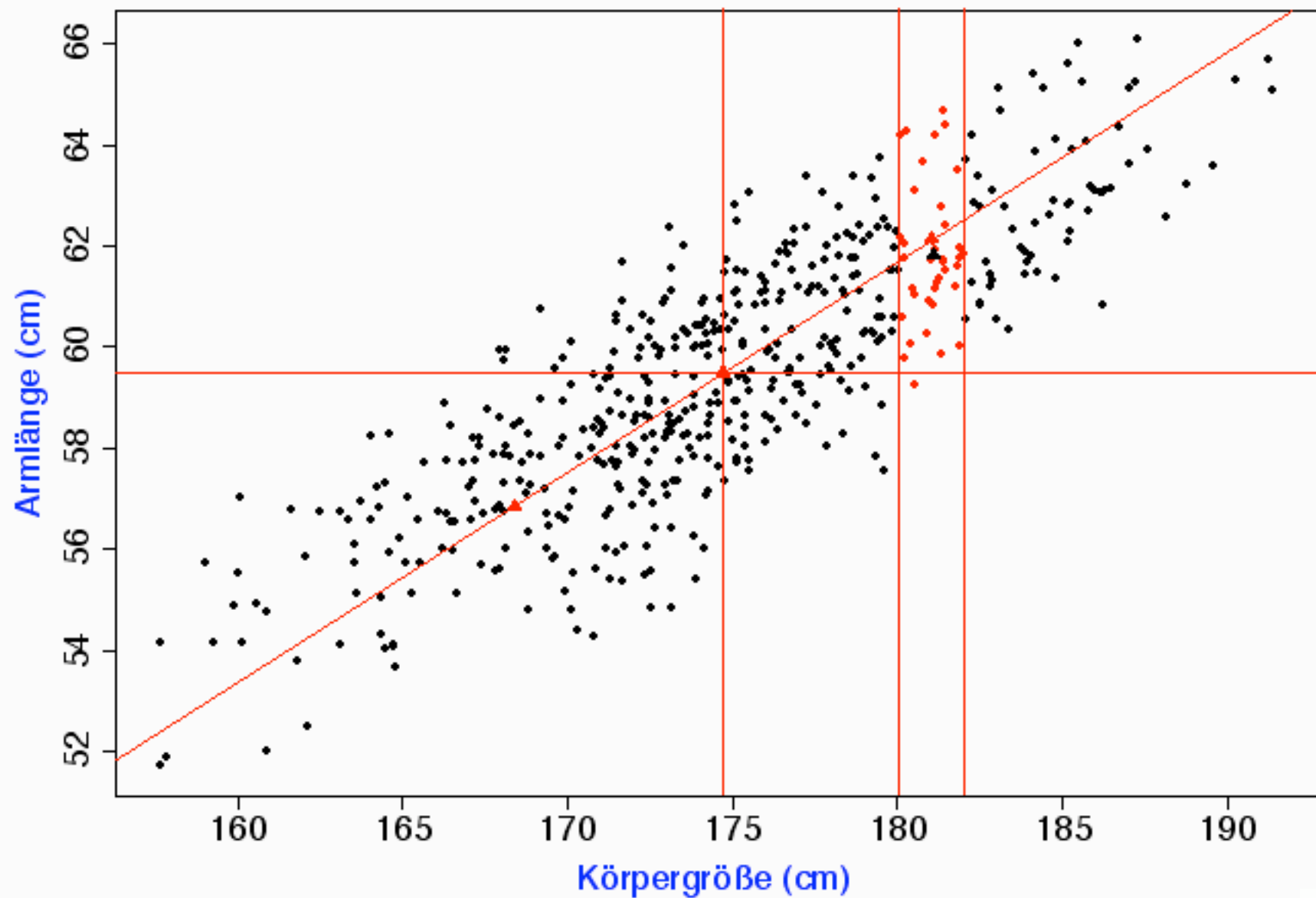


Wie groß im Schnitt
ist die Armlänge
eines Mannes,
dessen Körpergröße
um eine Standardabweichung
größer als der Durchschnitt ist?

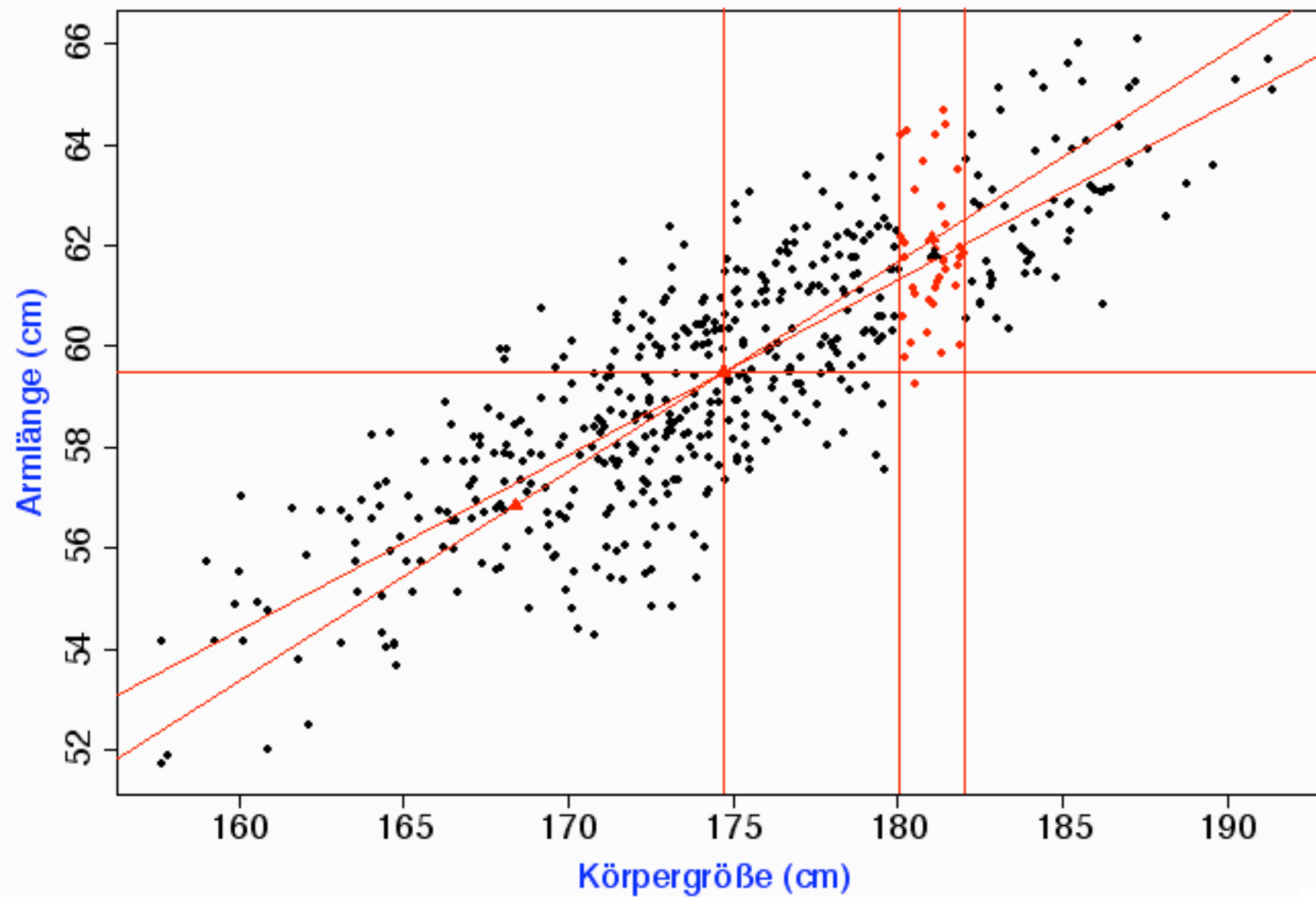
Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



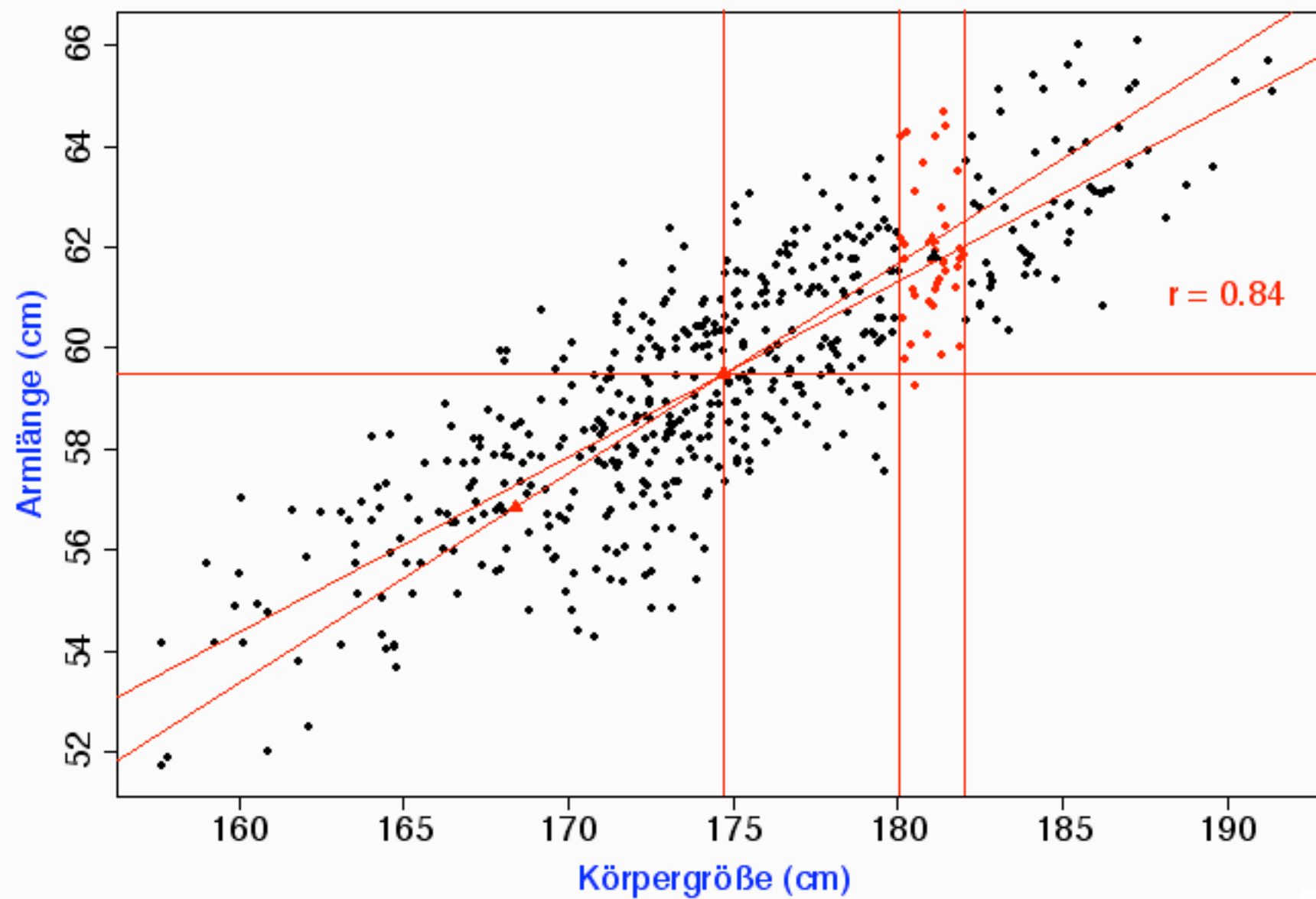
Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



Die Regressionsgerade



Die Regressionsgerade



$$r = 0,84$$

Wenn x
um s_x größer wird,
wird y
im Schnitt
um $0,84 s_y$ größer.

Eigenschaften

von

r

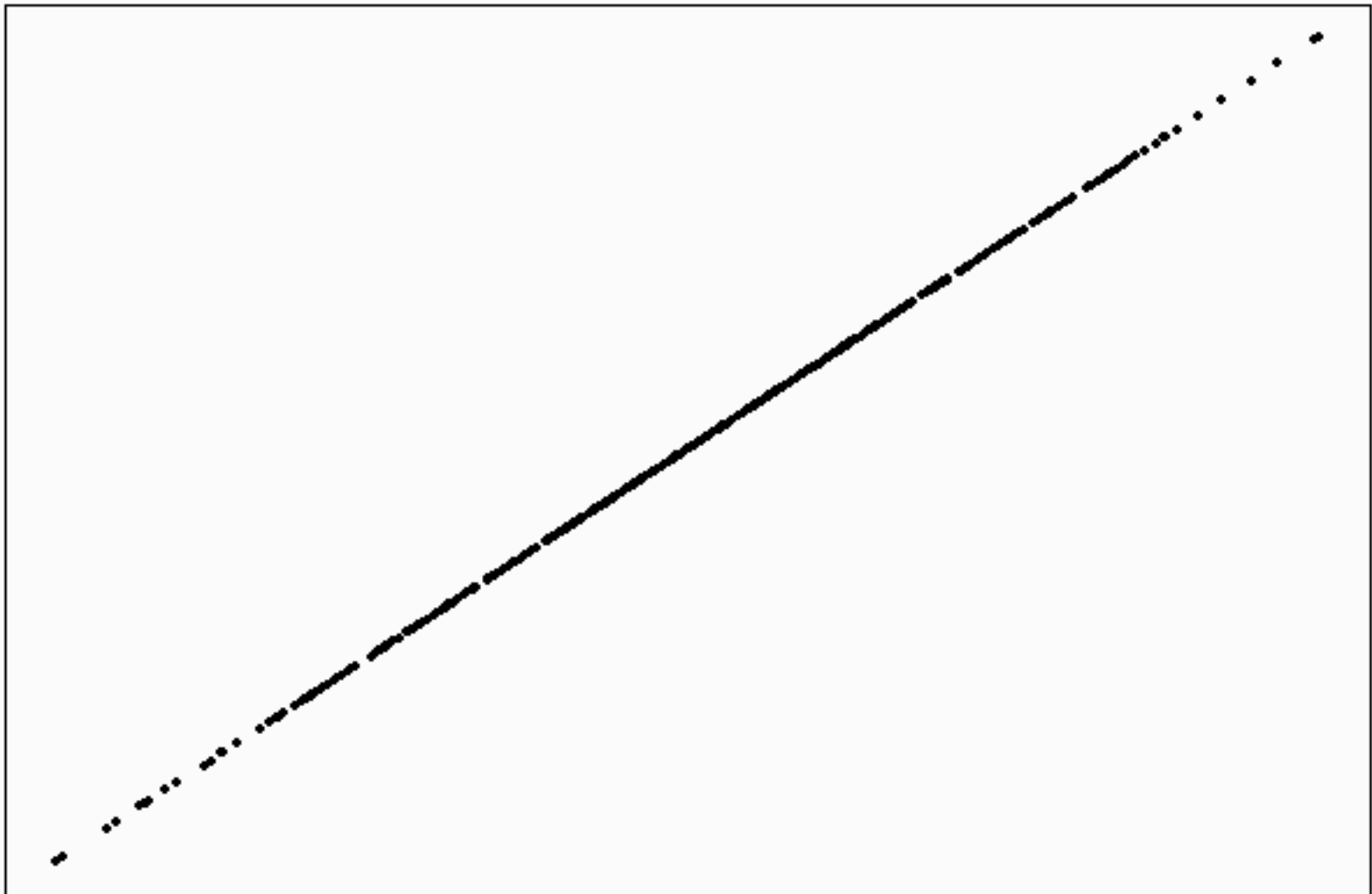
$$-1 \leq r \leq 1$$

$$r = 1:$$

Alle Punkte

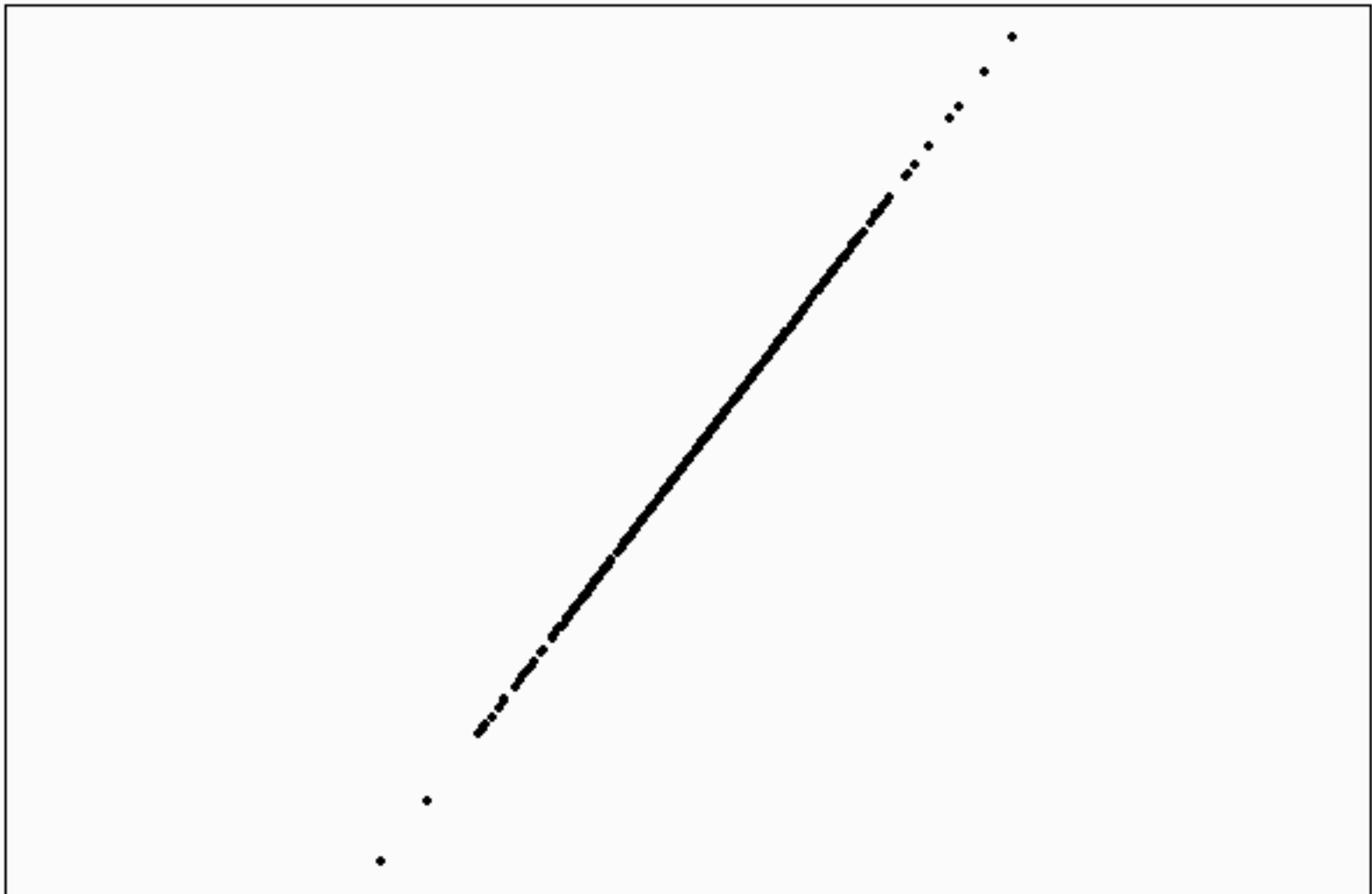
liegen auf einer steigenden Gerade.

$$r = 1$$

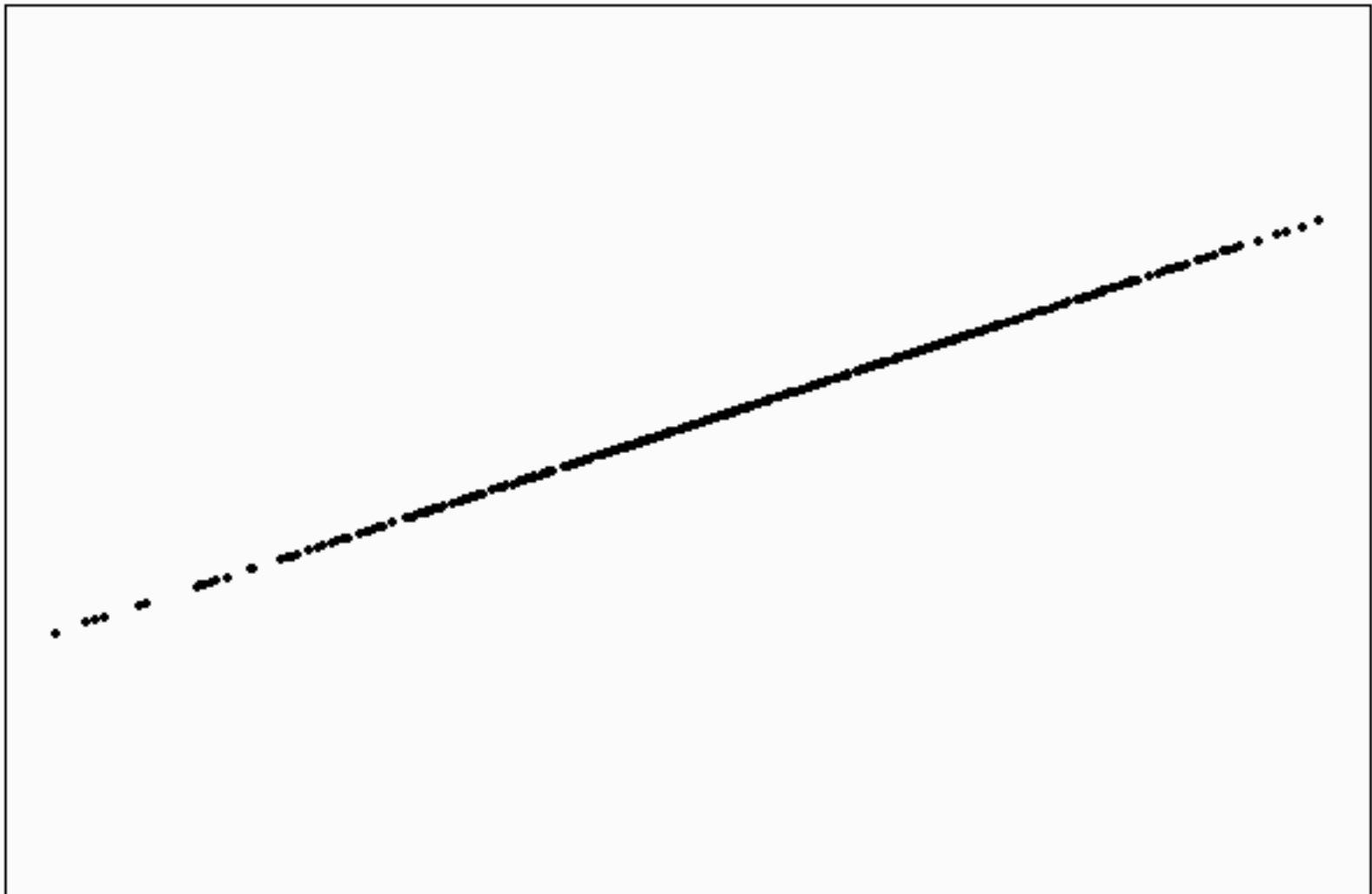


Die Steigung der Gerade,
solange sie positiv ist,
ist egal.

$$r = 1$$



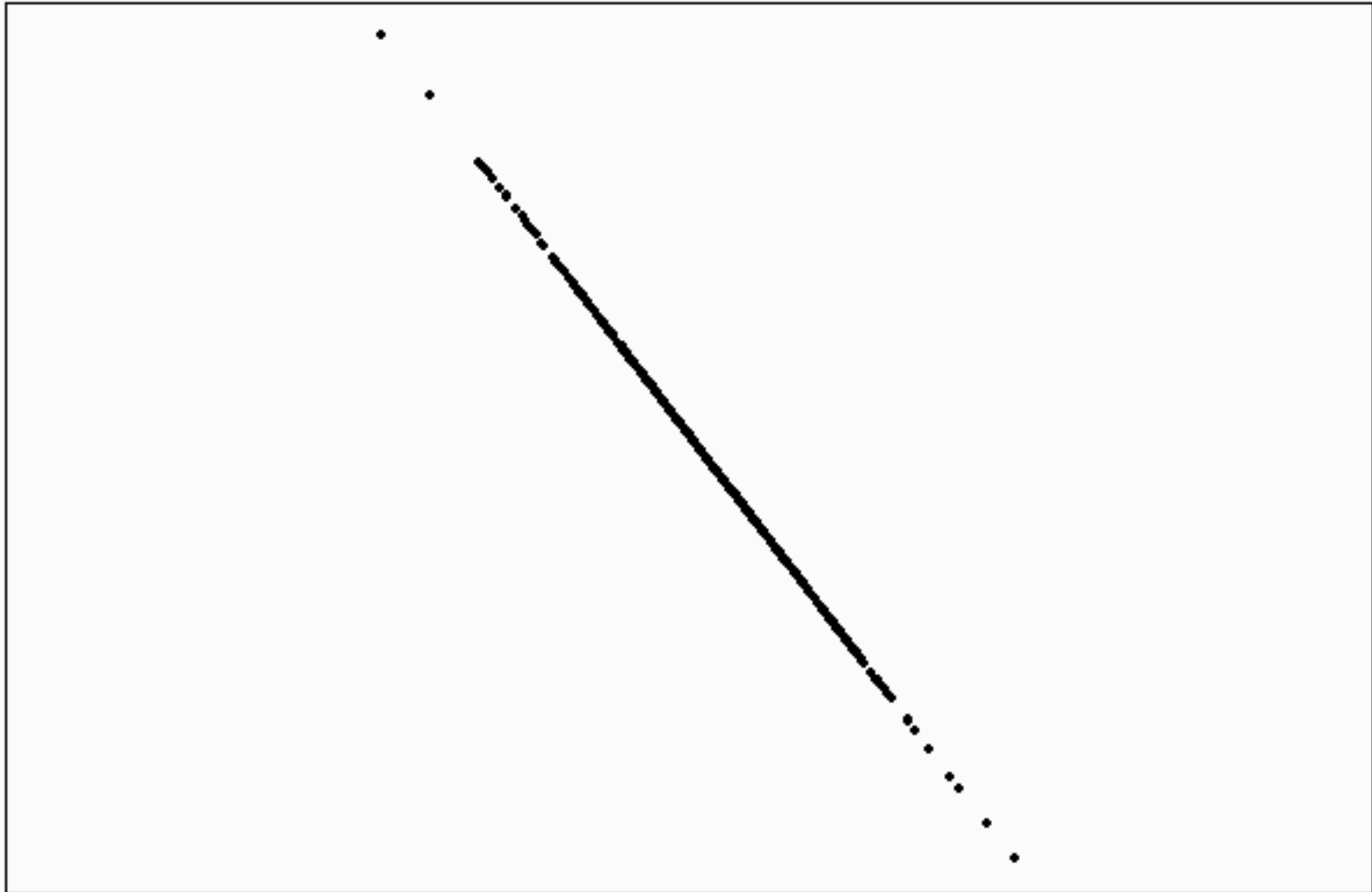
$$r = 1$$



$$r = -1:$$

Alle Punkte
liegen auf einer
fallenden
Gerade.

$$r = -1$$



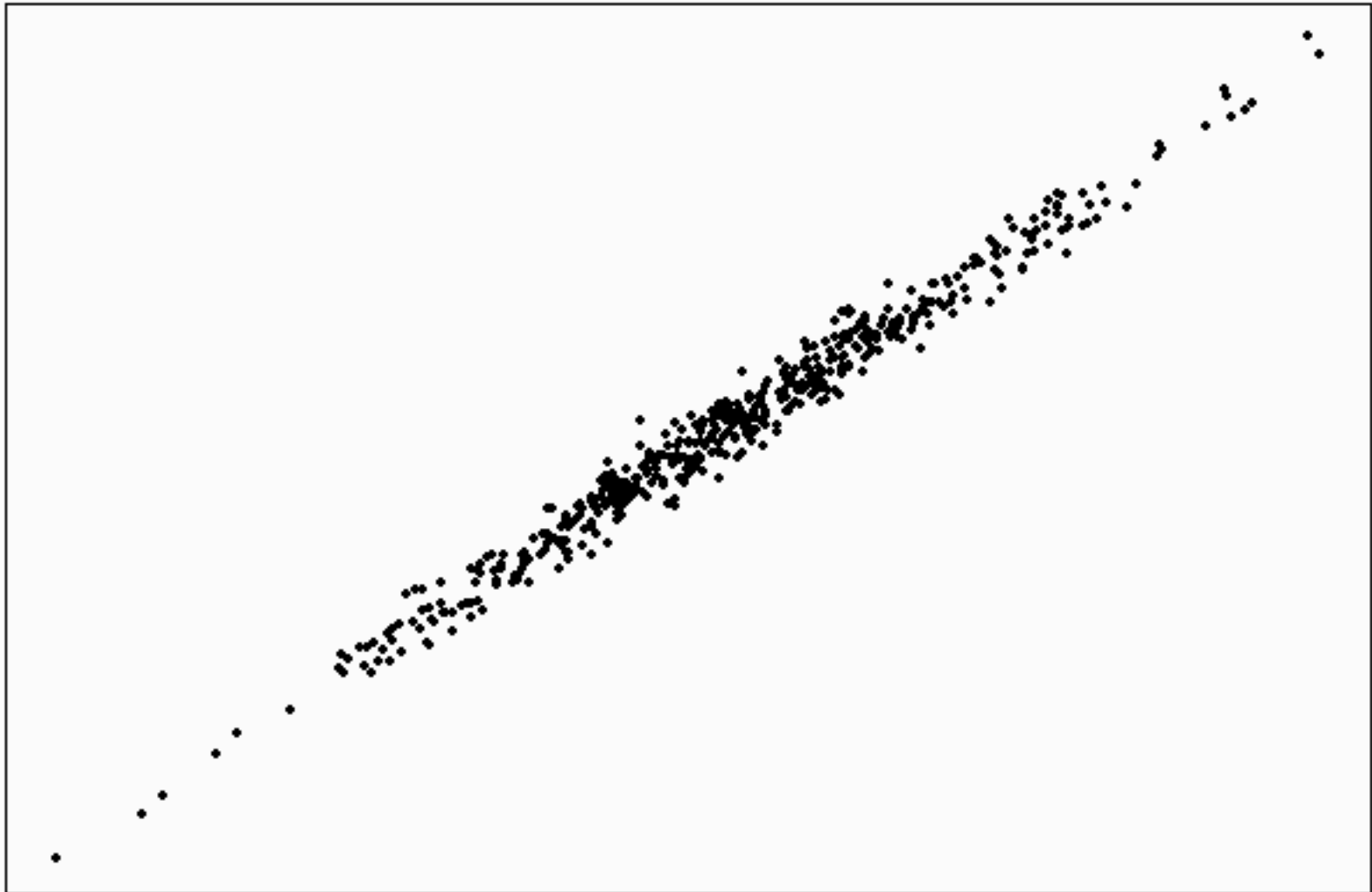
$$r \approx 1$$

Die Punktwolke

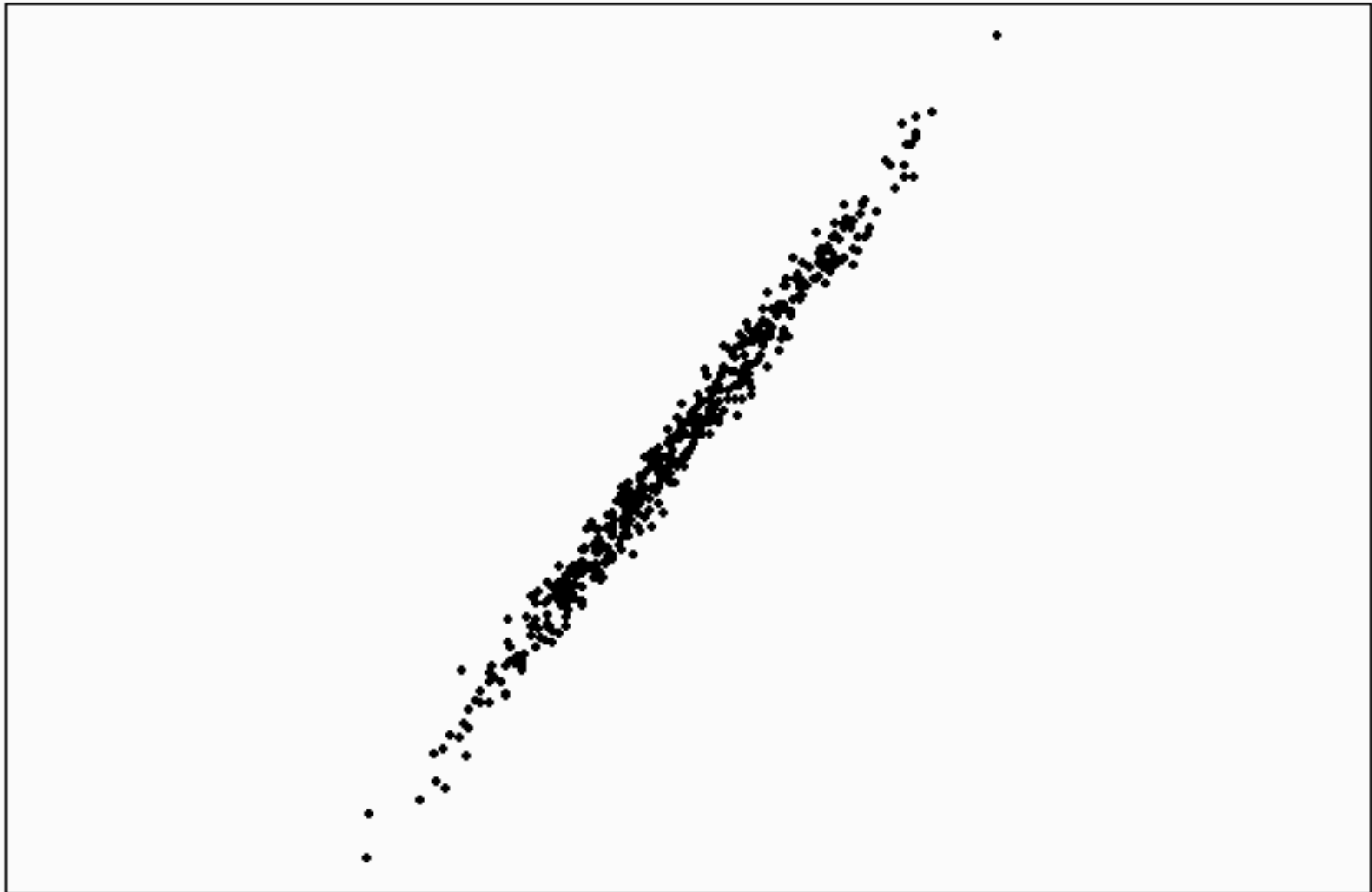
liegt eng

um eine steigende Gerade.

$r = 0.99$

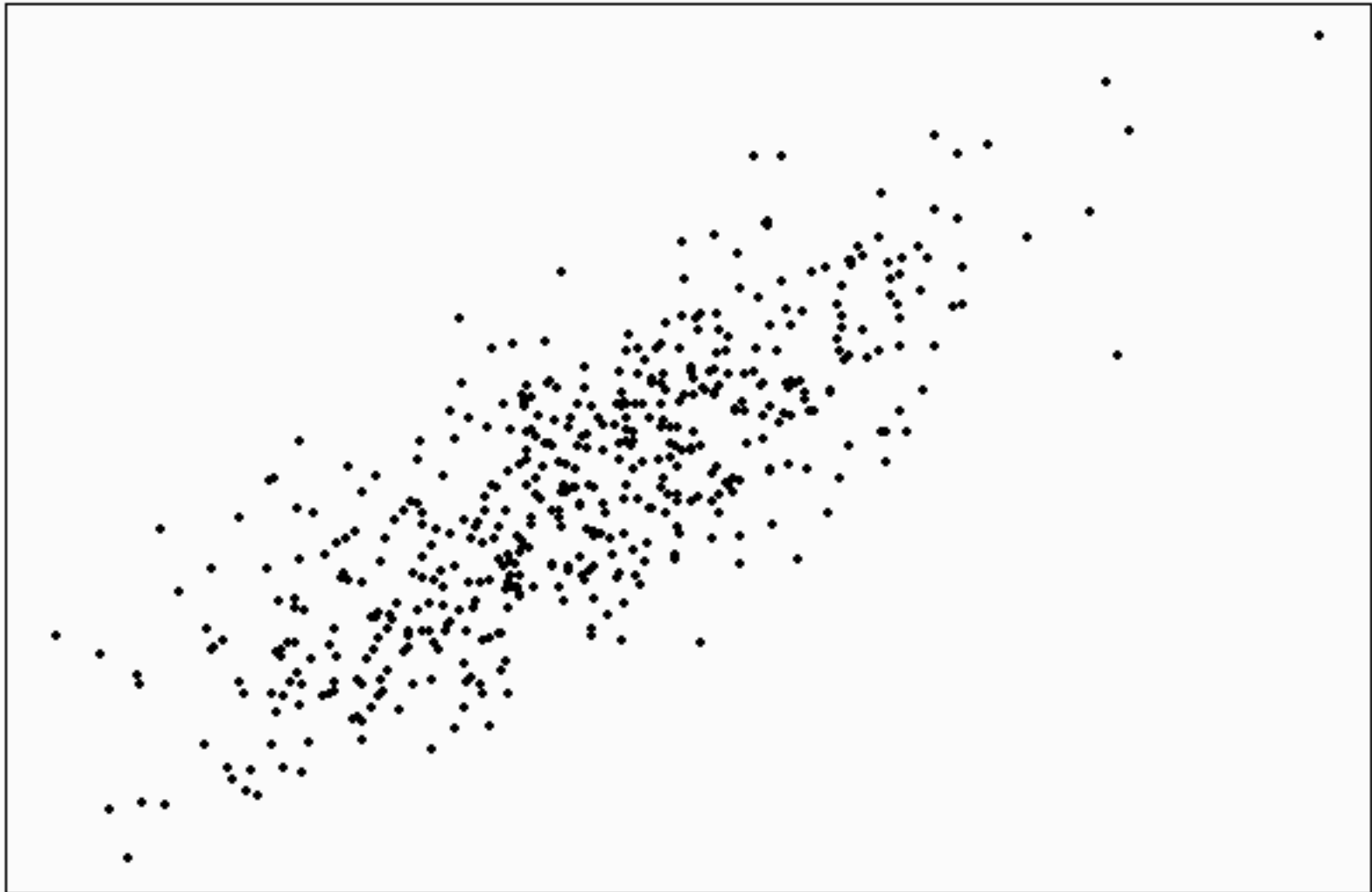


$r = 0.99$

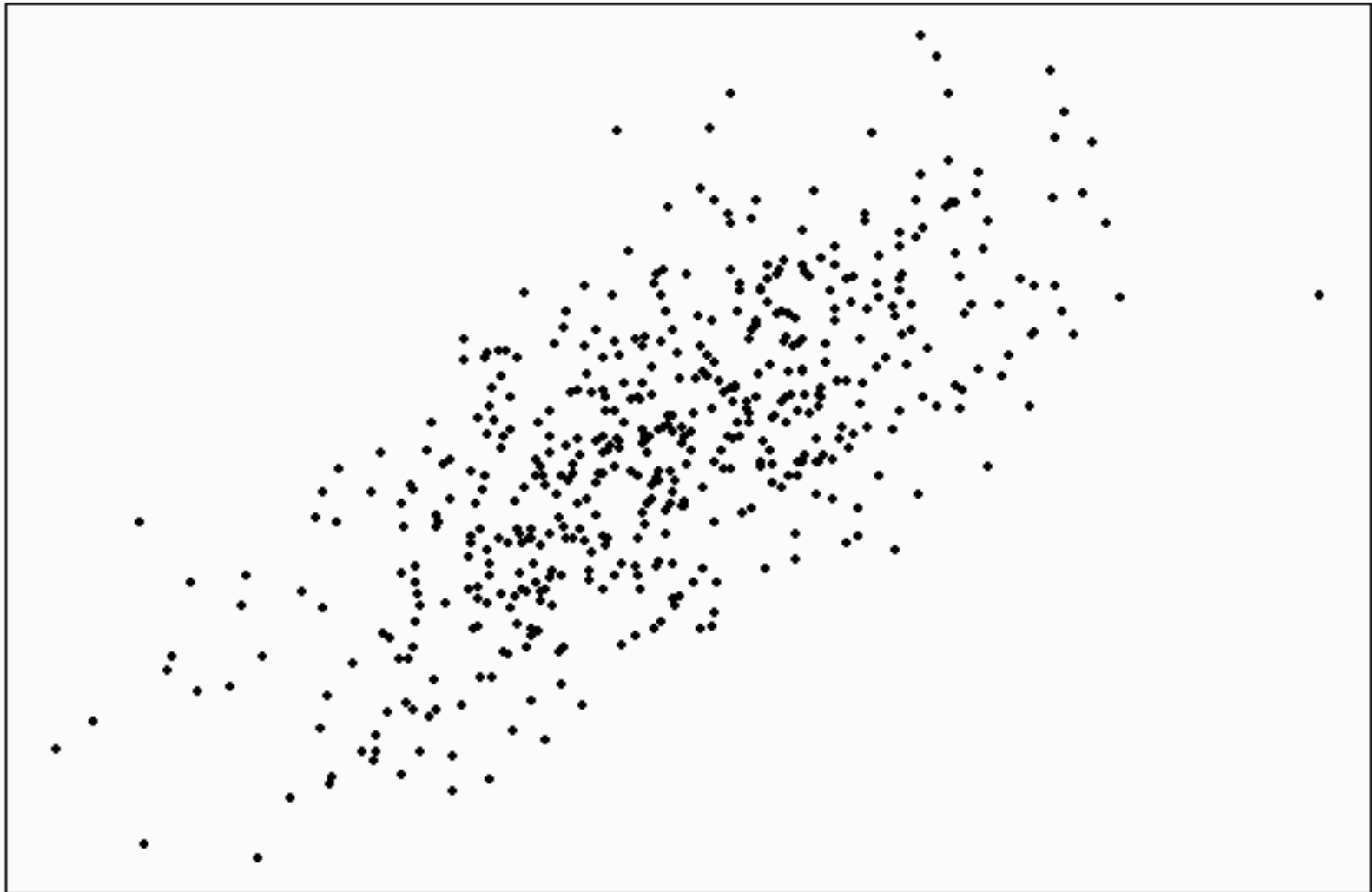


Wenn r kleiner wird,
streut die Punktwolke
immer mehr
um die Regressionsgerade.

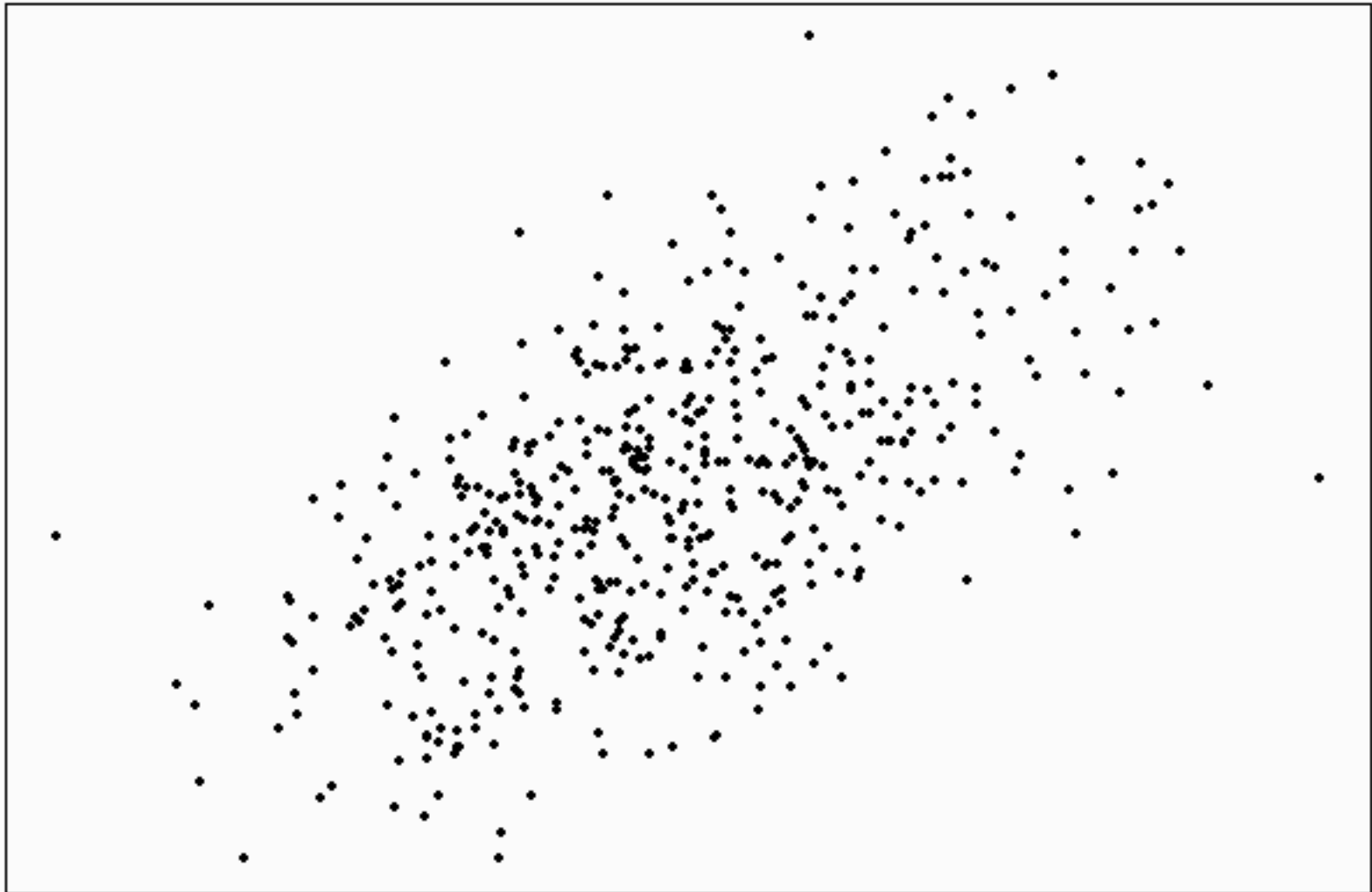
$r = 0.8$



$r = 0.7$



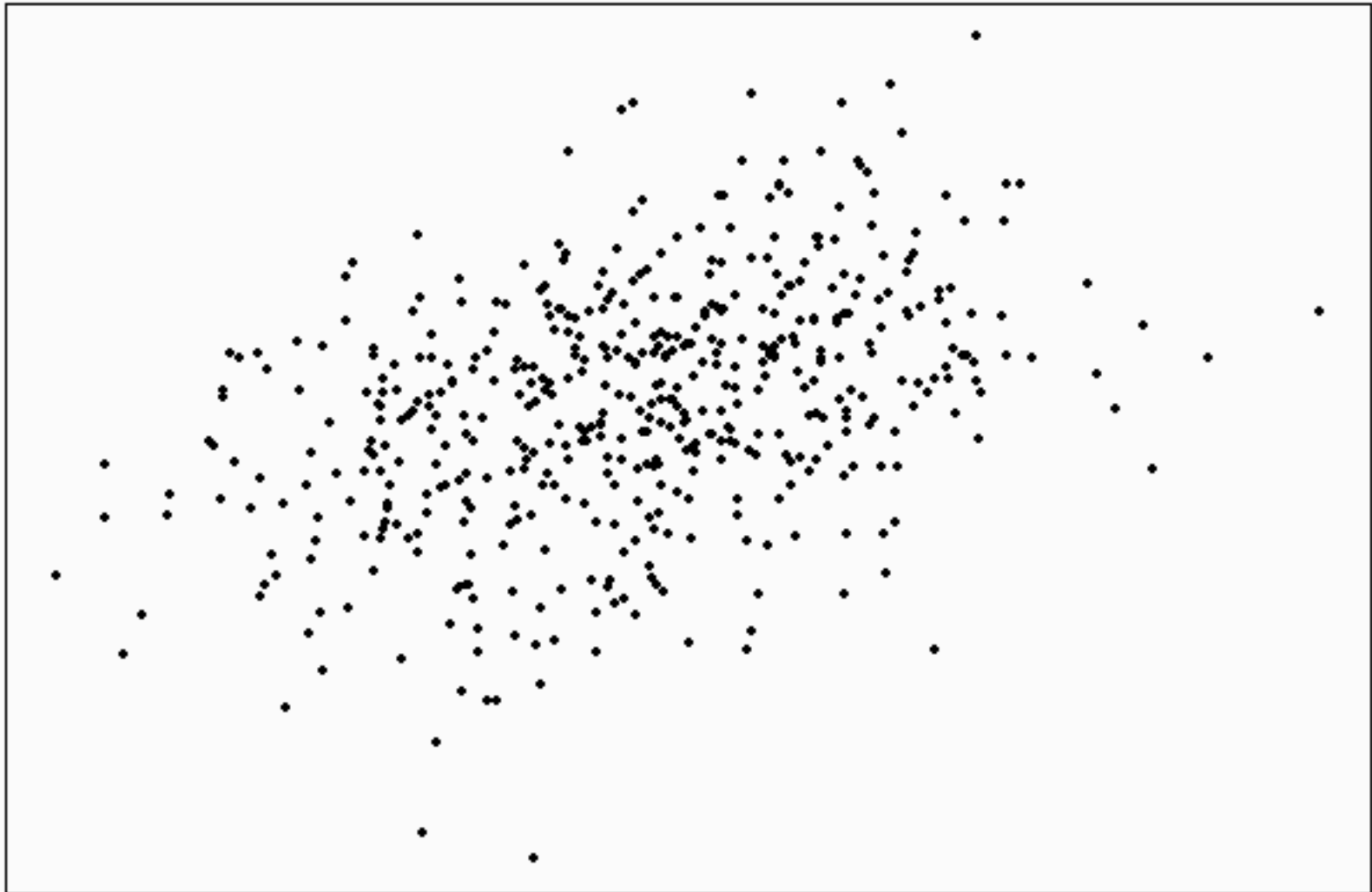
$r = 0.6$



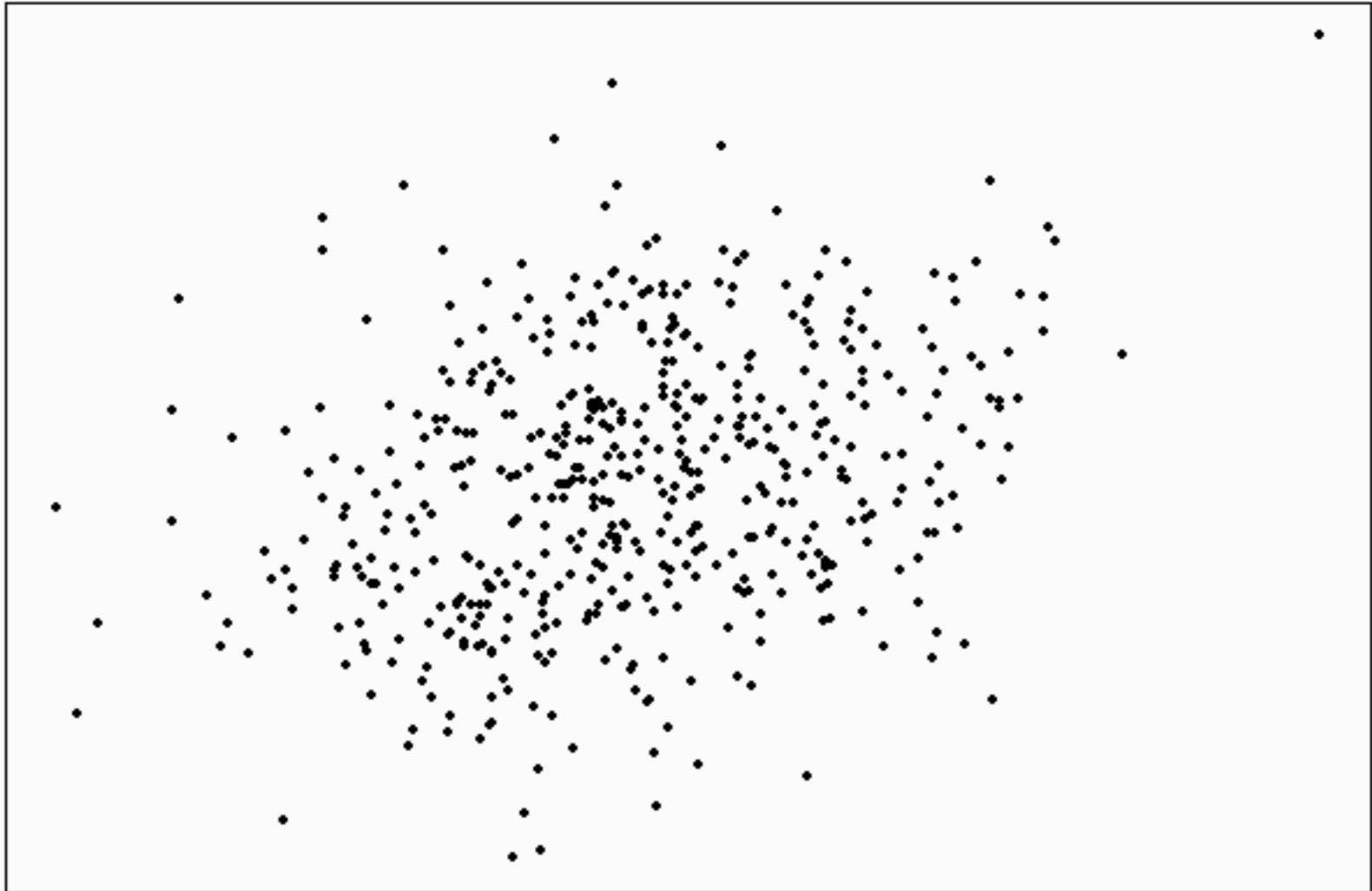
$r = 0.5$



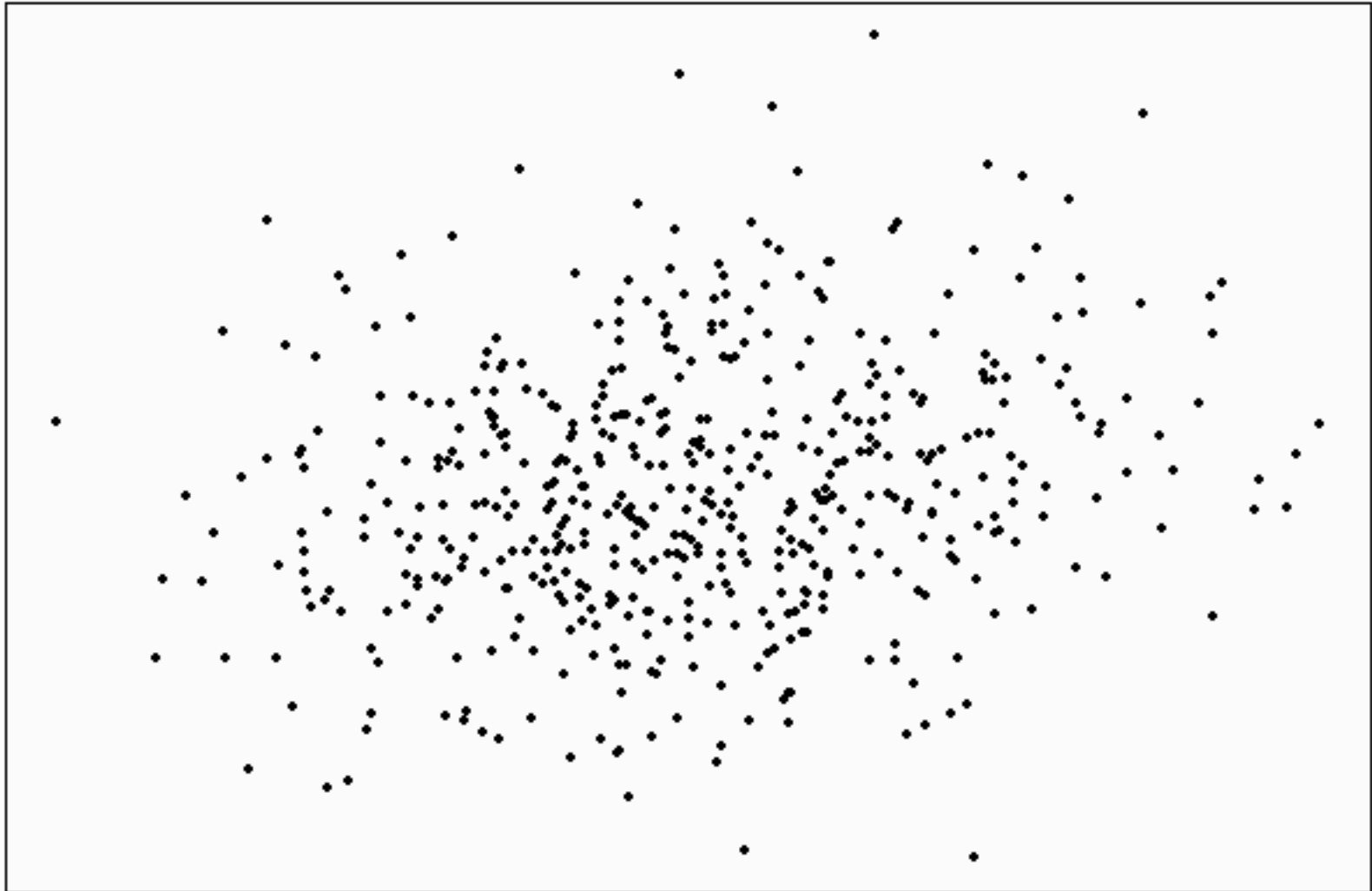
$r = 0.4$



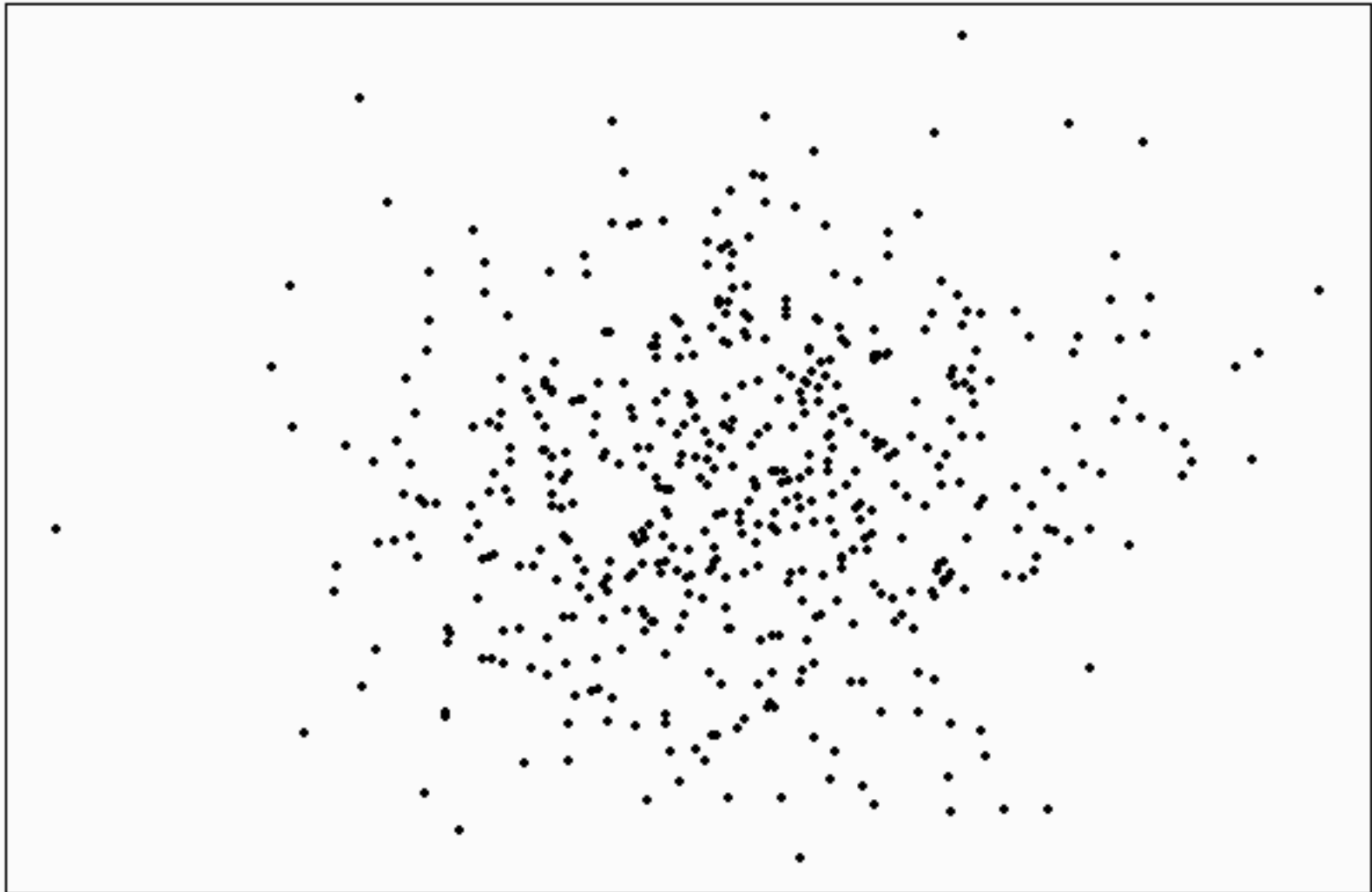
$r = 0.3$



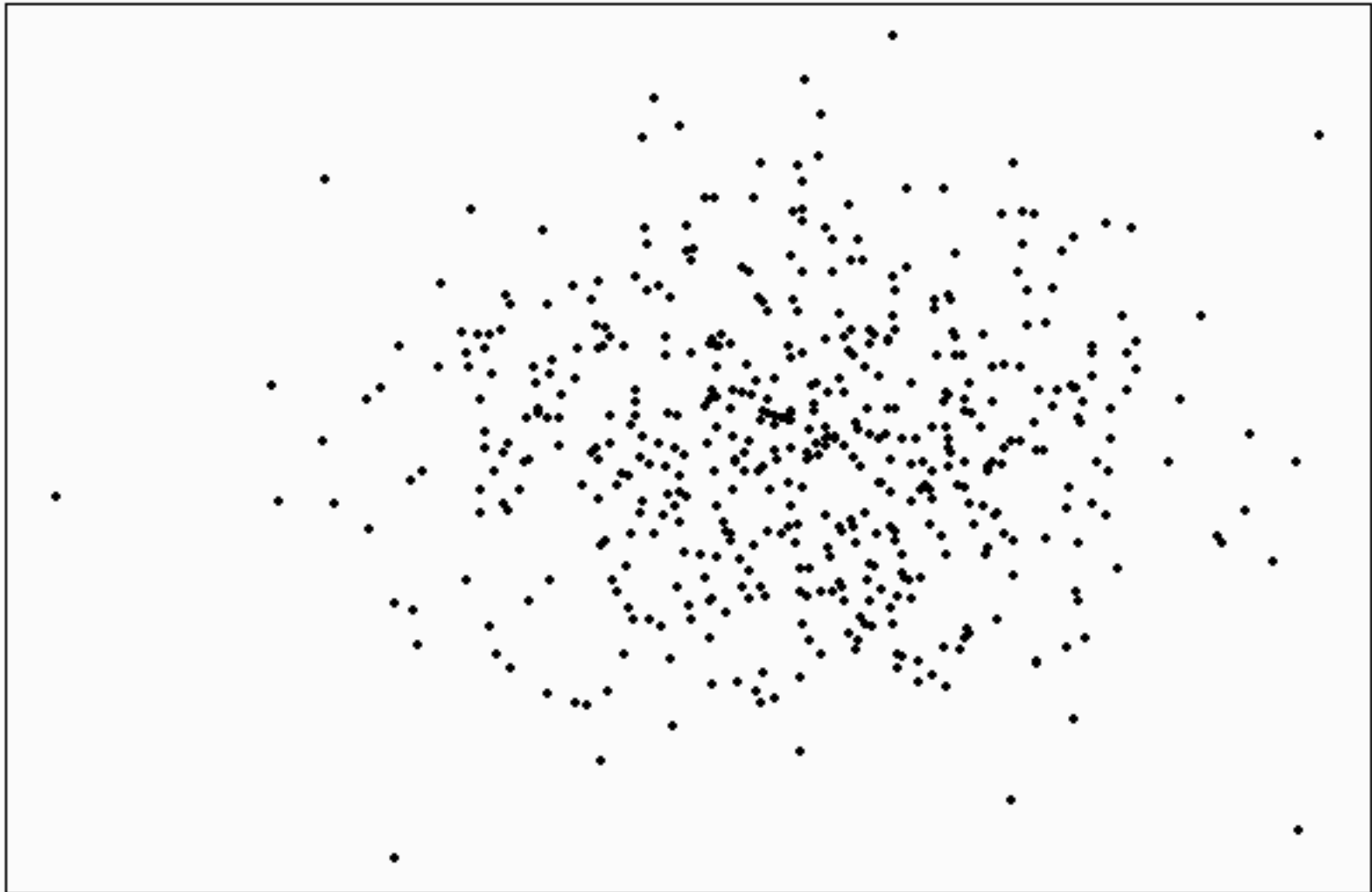
$r = 0.2$



$r = 0.1$



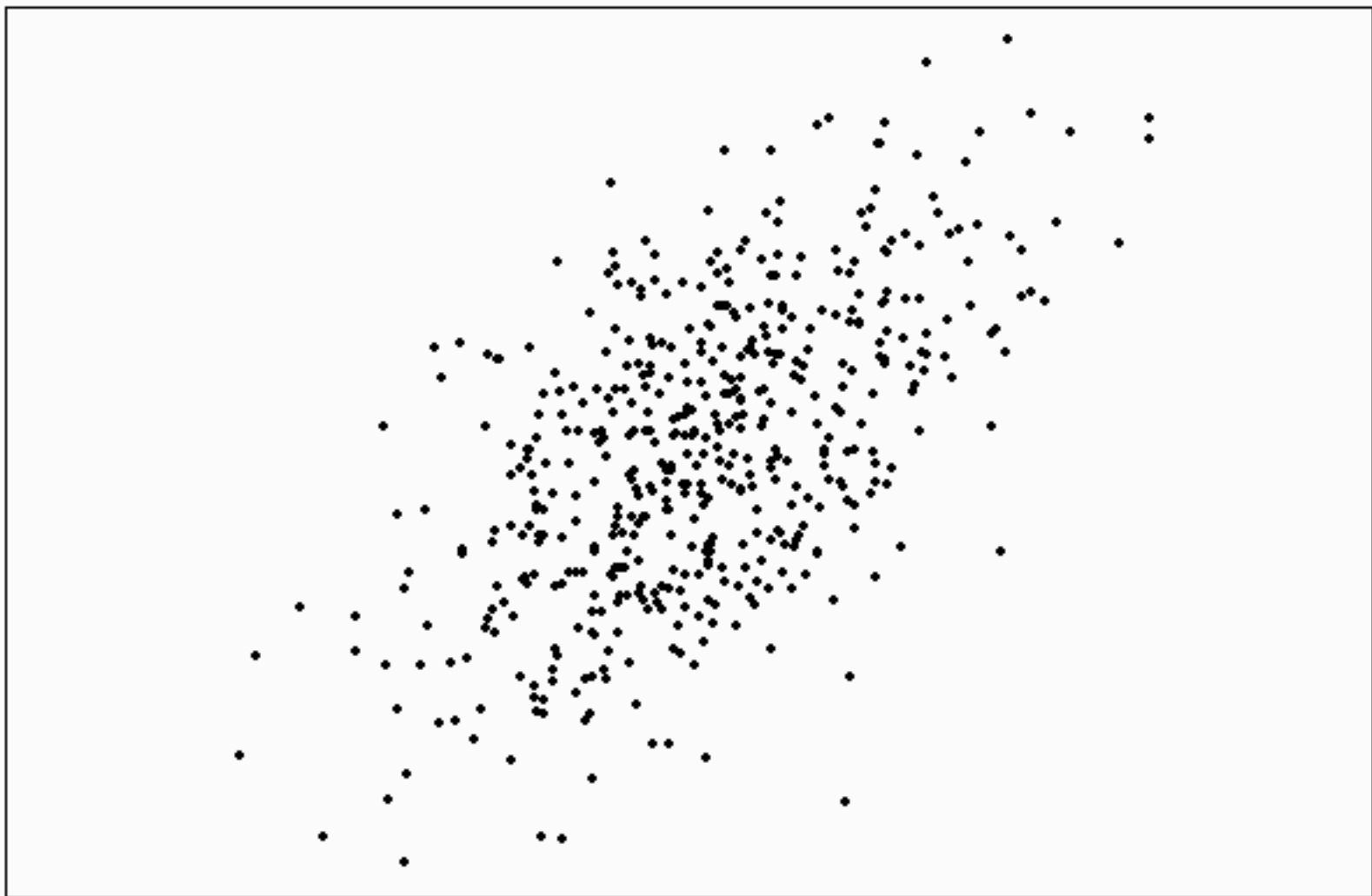
$$r = 0$$



$$r > 0:$$

Wenn x größer wird,
wird y größer.

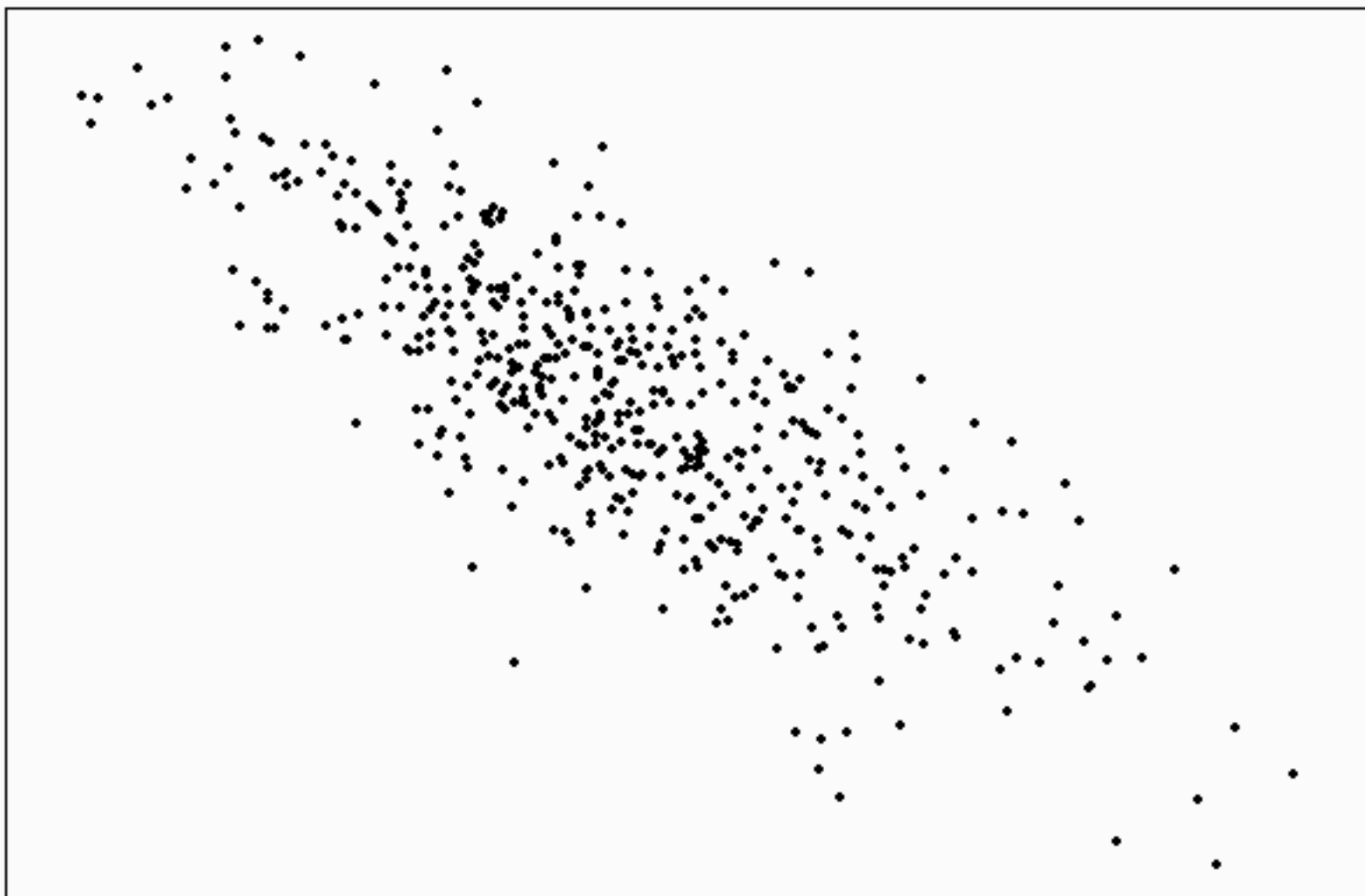
$r = 0.58$



$$r < 0$$

Wenn x größer wird,
wird y kleiner.

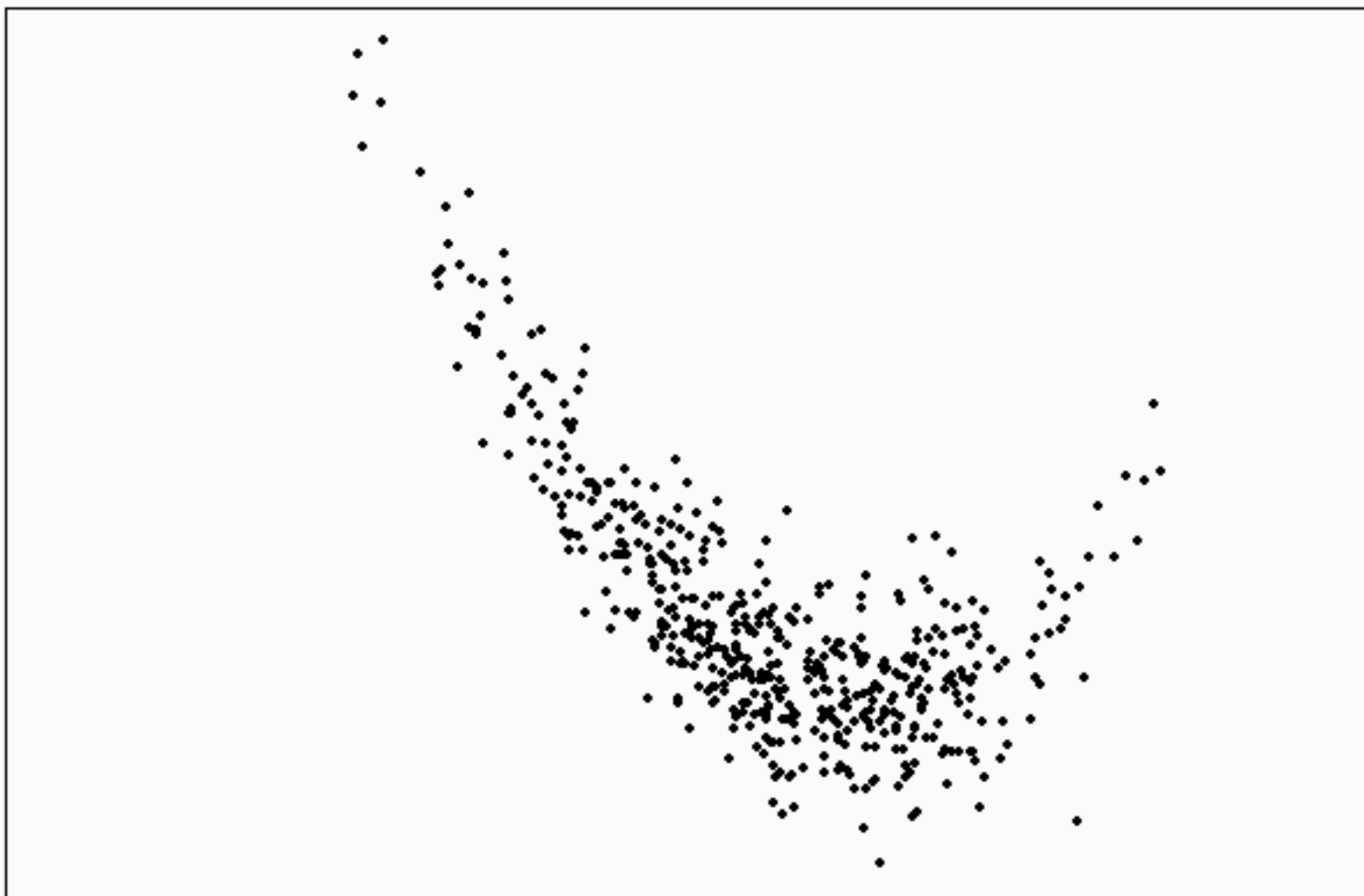
$r = -0.78$



r misst die Stärke
des **linearen** Zusammenhangs
zwischen X und Y .

Wenn der Zusammenhang
nicht ungefähr linear ist,
ist r irreführend.

$$r = -0.67$$



$$r < 0$$

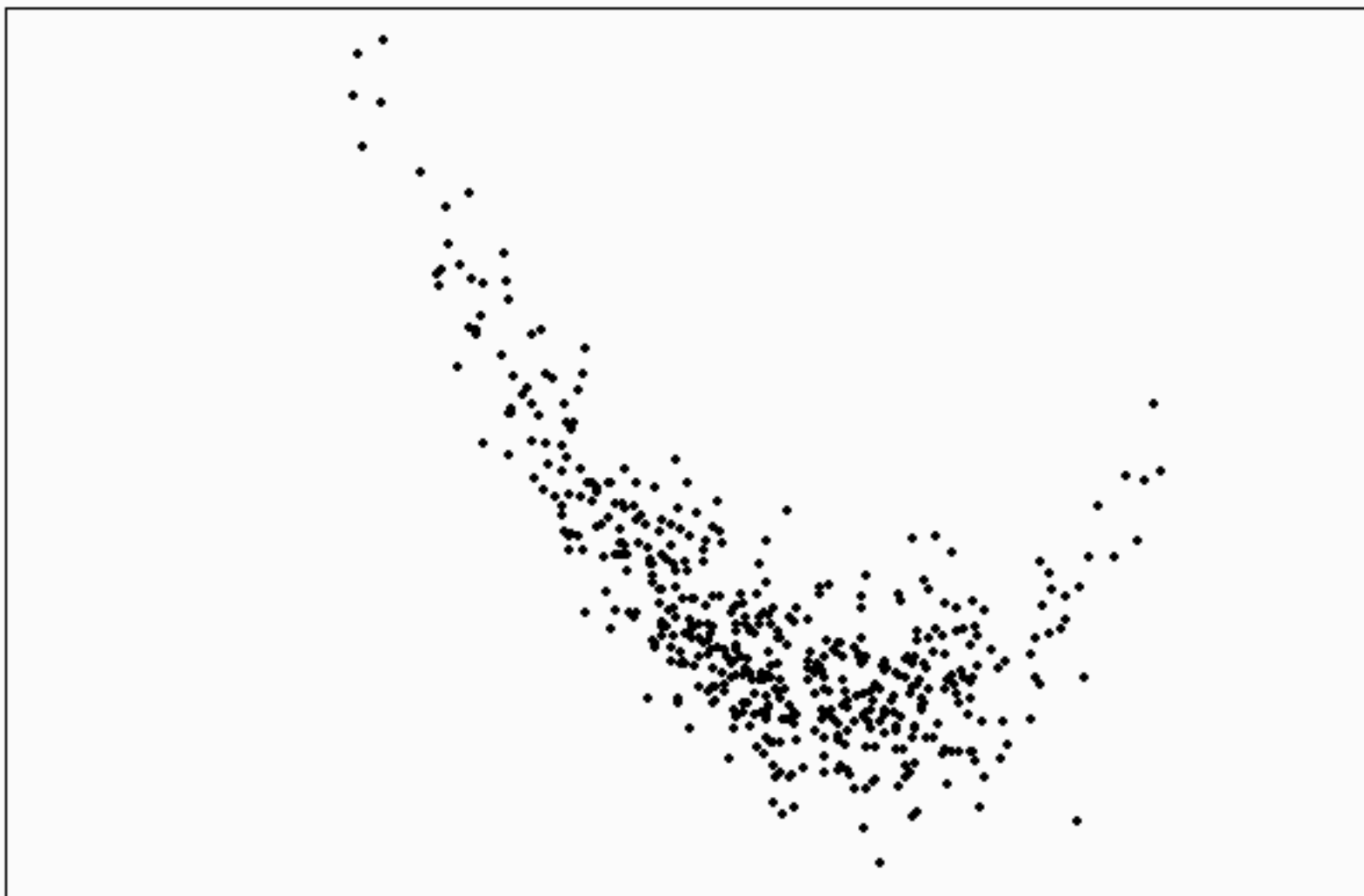
$$r < 0$$

aber

für große X

steigt Y .

$$r = -0.67$$



In solchen Fällen

sollte man

r nicht benutzen.

Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 1:

Übergangswahrscheinlichkeiten

Stellen wir uns ein zufälliges Paar $X = (X_1, X_2)$ vor,
das auf zweistufige Weise zustande kommt:

es gibt eine **Regel**, die besagt, wie X_2 verteilt ist,
gegeben dass X_1 den Ausgang a_1 hat.

Beispiel A:

In Stufe 1 entscheiden wir uns
entweder für einen fairen Würfel

oder für einen gezinkten:

drei von dessen 6 Seiten sind mit der Augenzahl 5,
die anderen drei sind mit der Augenzahl 6 beschriftet.

X_1 := der Typ des gewählten Würfels (fair oder gezinkt)

X_2 := die dann (in Stufe 2) geworfene Augenzahl.

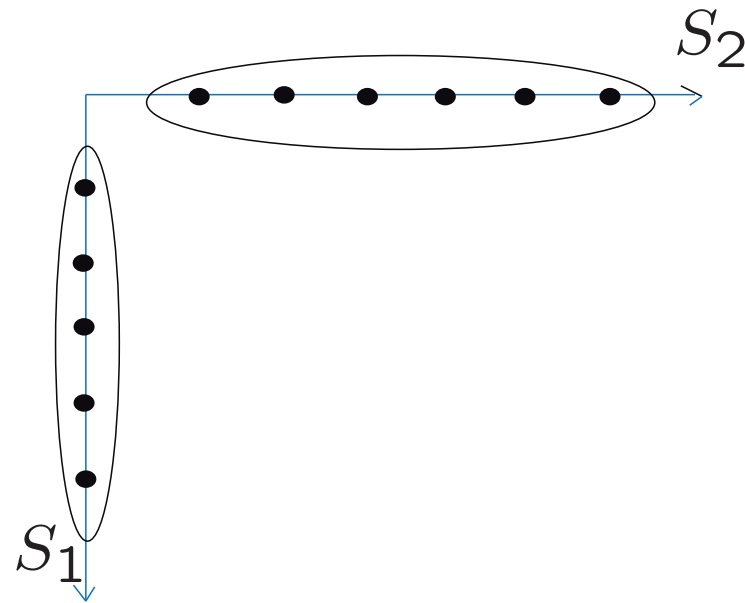
Wenn in Stufe 1 der **faire** Würfel gewählt wird,
dann sind die **Verteilungsgewichte** von X_2

$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

Wenn in Stufe 1 der **gezinkte** Würfel gewählt wird,
dann sind die **Verteilungsgewichte** von X_2

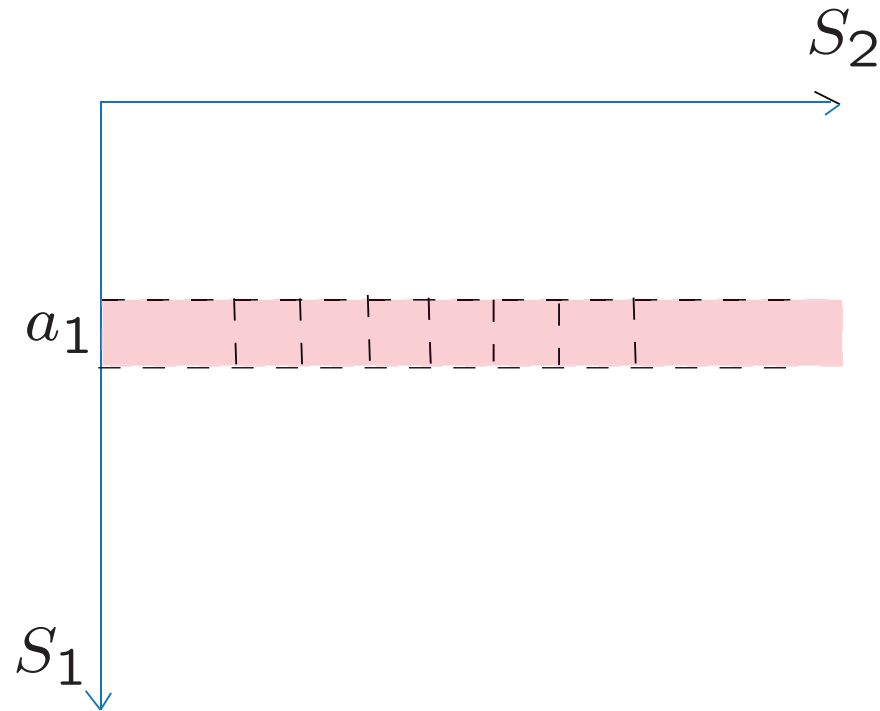
$$0 \quad 0 \quad 0 \quad 0 \quad \frac{1}{2} \quad \frac{1}{2}$$

Ein allgemeiner Rahmen.



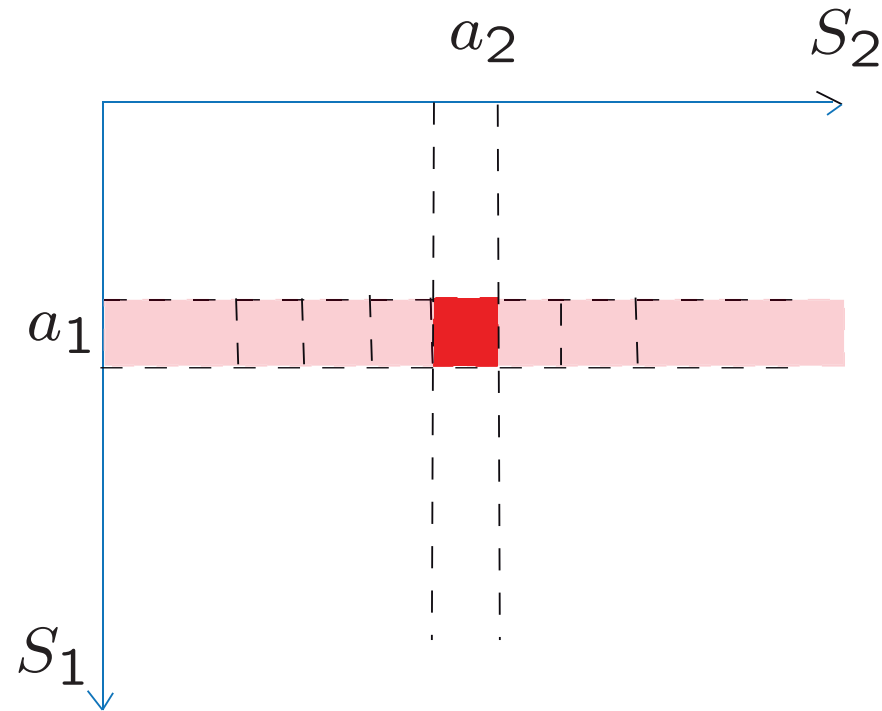
S_1 und S_2 seien zwei Wertebereiche.

Stellen wir uns vor: **Wenn** in Stufe 1
die Wahl auf das Element $a_1 \in S_1$ fällt,
dann landet man in der mit a_1 bezeichneten Zeile.



S_1 und S_2 seien zwei Wertebereiche.

Stellen wir uns vor: **Wenn** in Stufe 1
die Wahl auf das Element $a_1 \in S_1$ fällt,
dann landet man in der mit a_1 bezeichneten **Zeile**.



Wenn **dann** in Stufe 2
die Wahl auf das Element $a_2 \in S_2$ fällt,
landet man in dem mit (a_1, a_2) bezeichneten **Feld**
(Zeile a_1 , Spalte a_2)

Beispiel B:

$$S_1 = \{1, 2, 3\}, S_2 = \{0, 1, 2, 3, 4\}.$$

In Stufe 1

wählen wir eine Zahl X_1 aus $\{1, 2, 3\}$.

In Stufe 2 verschieben wir das in Stufe 1 erzielte Ergebnis
mit W'keit $1/2$ um eins nach rechts
und mit W'kt $1/2$ um eins nach links.

Mit anderen Worten:

Gegeben $\{X_1 = a_1\}$

ist X_2 uniform verteilt auf $\{a_1 - 1, a_1 + 1\}$.

Beispiel C:

$$S_1 = S_2 = \mathbb{R}.$$

In Stufe 1

stellt sich eine reelle Zahl X_1 ein.

In Stufe 2 wird dazu

eine unabhängige standard-normalverteilte ZV'e addiert:

Mit anderen Worten:

Gegeben $\{X_1 = a_1\}$

hat X_2 die Verteilung $N(a_1, 1)$.

(Dieses Beispiel weist über den diskreten Fall hinaus.)

Für die
W'keit des Ereignisses $\{X_2 \in A_2\}$ gegeben $\{X_1 = a_1\}$
schreiben wir
 $P_{a_1}(X_2 \in A_2)$
und sprechen von den
Übergangswahrscheinlichkeiten.

Für diskretes S_2

reicht es, die einzelnen Ausgänge a_2 zu betrachten:

$$P(a_1, a_2) = \mathbf{P}_{a_1}(X_2 = a_2), \quad a_1 \in S_1, a_2 \in S_2$$

ist die sogenannte **Übergangsmatrix**.

Man hat dann:

$$P(a_1, A_2) := \sum_{a_2 \in A_2} P(a_1, a_2) = \mathbf{P}_{a_1}(X_2 \in A_2)$$

und kurz:

$$P(a_1, \cdot) = \mathbf{P}_{a_1}(X_2 \in \cdot).$$

Für kontinuierliches $S_2 \subset \mathbb{R}$ hat man
üblicherweise wieder Dichten (statt der Gewichte):
 $g_{a_1}(a_2) da_2 = \mathbf{P}_{a_1}(X_2 \in da_2)$, $a_1 \in S_1$, $a_2 \in S_2$,
heißen die **Übergangsdichten**.

Man hat dann z.B. für Intervalle $A_2 = [c, d] \subset S_2$:

$$\mathbf{P}_{a_1}(X_2 \in A_2) = \int_c^d g_{a_1}(a_2) da_2.$$

Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 2:

Von der Startverteilung
mit den Übergangswahrscheinlichkeiten
zur gemeinsame Verteilung

(Buch S. 87-89, 91-92)

Zweistufige Zufallsexperimente - der diskrete Fall:

Sei $X = (X_1, X_2)$ ein zufälliges Paar
mit diskretem Zielbereich $S = S_1 \times S_2$.

Für jedes $a_1 \in S_1$ sei $P(a_1, \cdot)$ eine Verteilung auf S_2 .

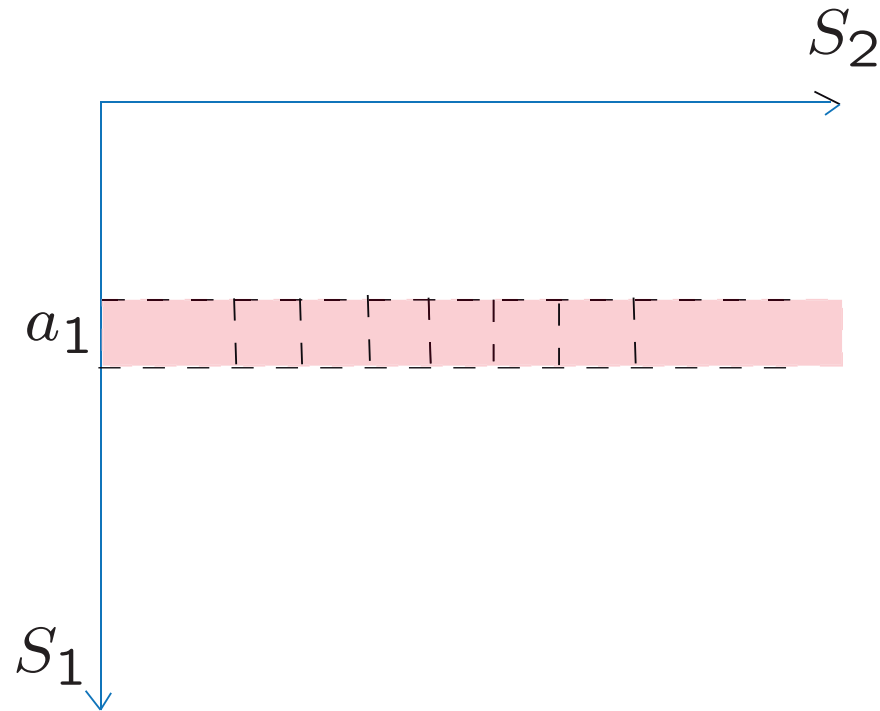
Damit ist hier gemeint, dass $P(a_1, a_2)$, $a_2 \in S_2$, Verteilungsgewichte auf S_2 sind, also nichtnegative Zahlen, die zu 1 summieren.

Vorstellung: gegeben $\{X_1 = a_1\}$

hat die Zufallsvariable X_2 die Verteilung $P(a_1, \cdot)$.

Schreibweise:

$$P(a_1, a_2) =: \mathbf{P}_{a_1}(X_2 = a_2) .$$

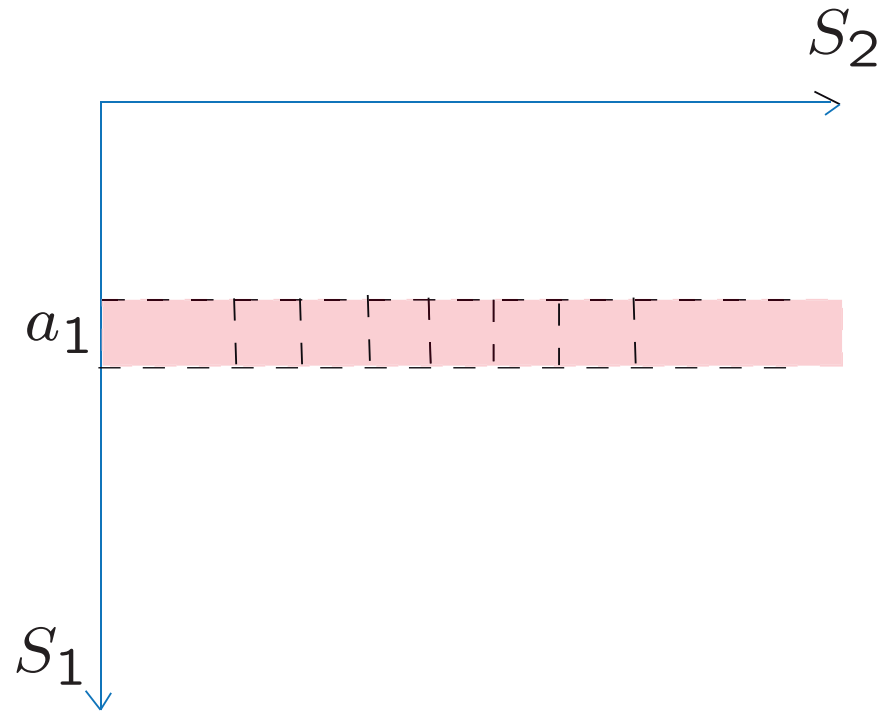


Jetzt bringen wir den Zufall auch in der erste Stufe ins Spiel.

Sei X_1 eine S_1 -wertige Zufallsvariable mit **Verteilung** ρ_1 .

Mit Wahrscheinlichkeit $\rho_1(a_1)$ fällt X_1 auf a_1

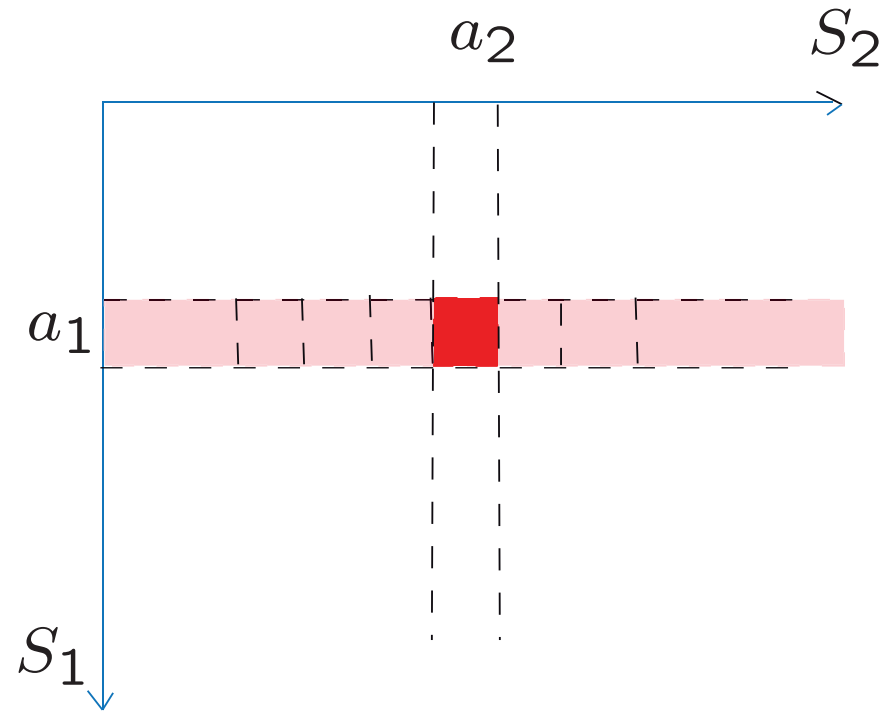
.... damit landet man in der Zeile namens a_1 .



Für jedes $a_1 \in S_1$ sei $P(a_1, \cdot)$ eine Verteilung auf S_2

Das heißt im hier betrachteten diskreten Fall:

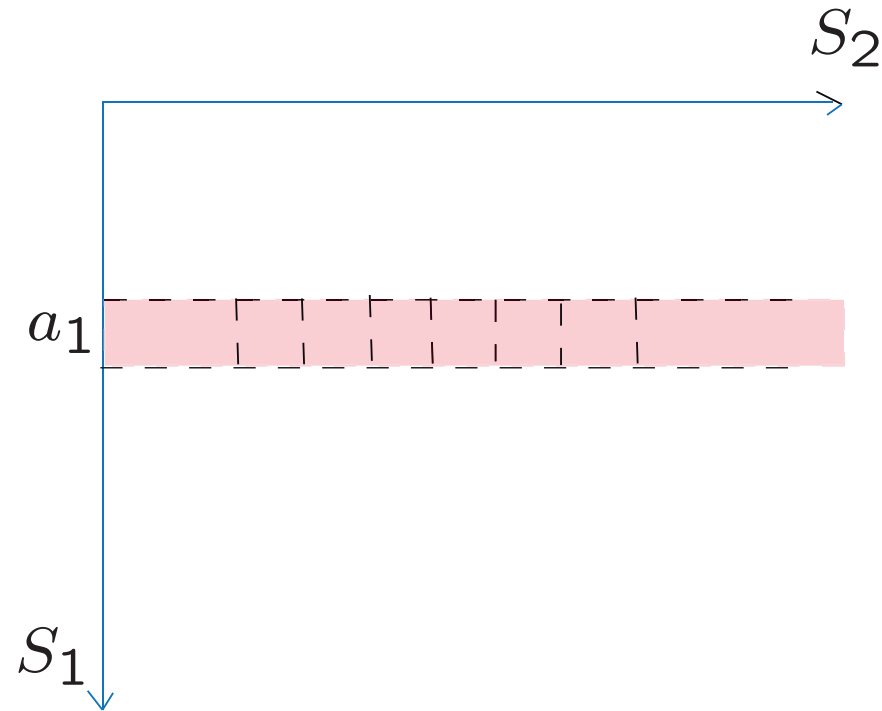
$P(a_1, a_2)$, $a_2 \in S_2$, sind Verteilungsgewichte auf S_2 ,
also nichtnegative Zahlen, die zu 1 summieren.



Vorstellung:

Gegeben $\{X_1 = a_1\}$

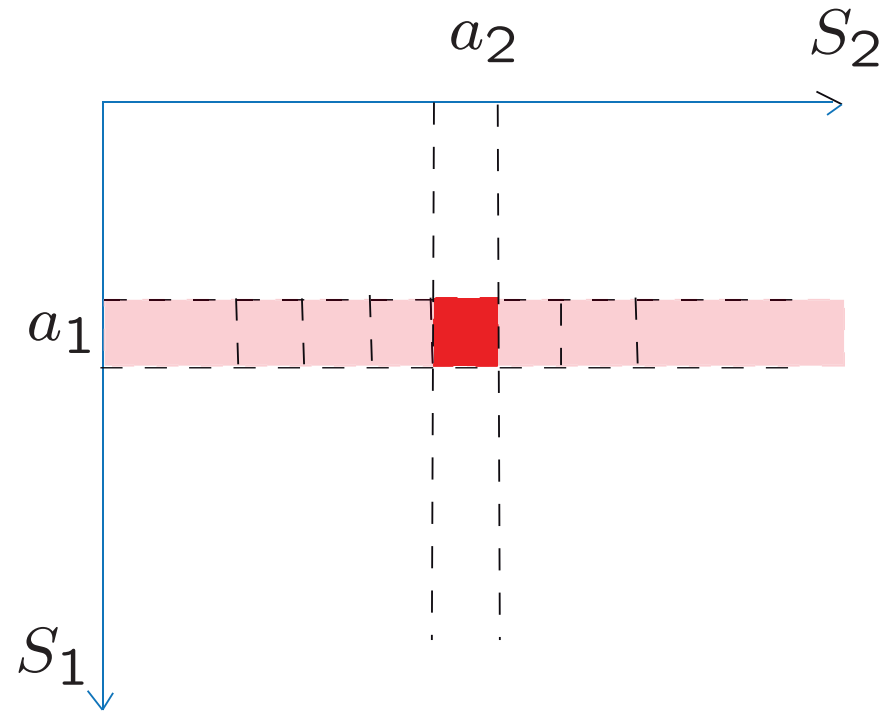
hat das Ereignis $\{X_2 = a_2\}$ die W'keit $P(a_1, a_2)$.



Anders gesagt:

Gegeben $\{X_1 = a_1\}$

hat die Zufallsvariable X_2 die Verteilung $P(a_1, \cdot)$.



Das Gewicht $\rho_1(a_1)$ aus Stufe 1
 wird in Stufe 2 gemäß $P(a_1, \cdot)$
 auf die Zeile a_1 aufgeteilt:

$$P(X_1 = a_1, X_2 = a_2) = \rho_1(a_1) P(a_1, a_2).$$

Aus der Verteilung von X_1
(der sogenannten *Startverteilung*, hier bezeichnet mit ρ_1)

und den Verteilungen $P(a_1, \cdot)$
(den sogenannten *Übergangsverteilungen*)

gewinnt man die *gemeinsame Verteilung* von X_1 und X_2
mit den Gewichten

$$\nu(a_1, a_2) = \rho_1(a_1) P(a_1, a_2)$$

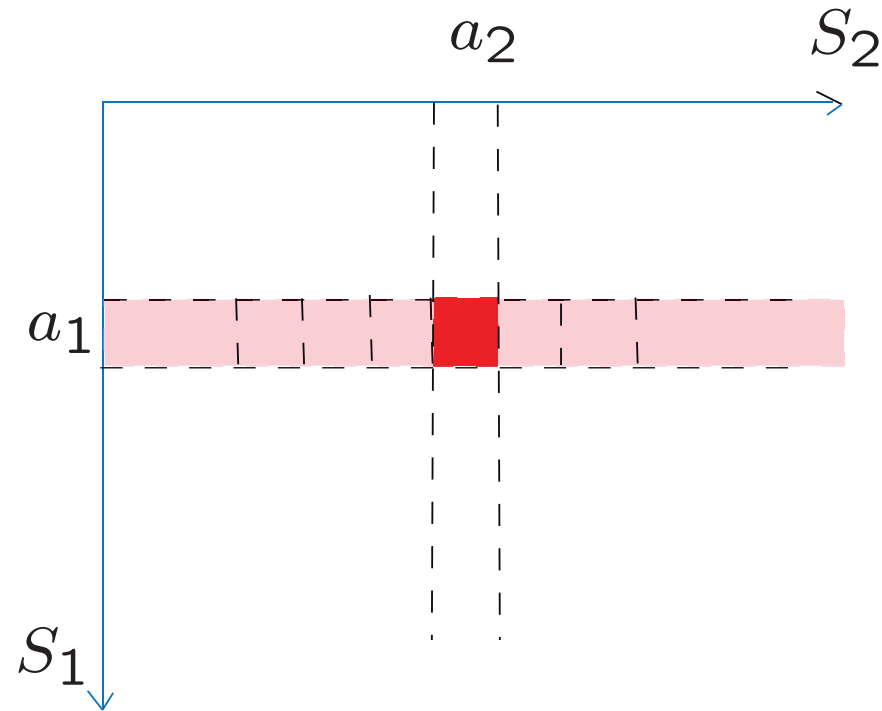
$$\nu(a_1, a_2) = \rho_1(a_1) P(a_1, a_2)$$

kann man auch schreiben als

$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) .$$

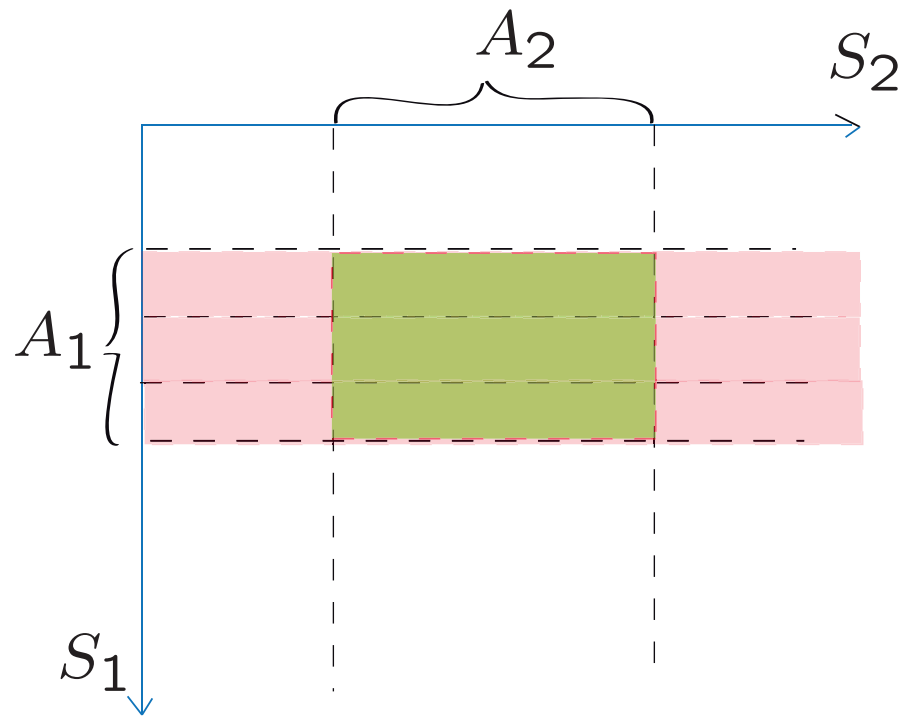
Das ist die einfachste Variante der sogenannten
Multiplikationsregel.

Später werden wir das Analogon für n
statt 2 Zufallsvariable kennenlernen.



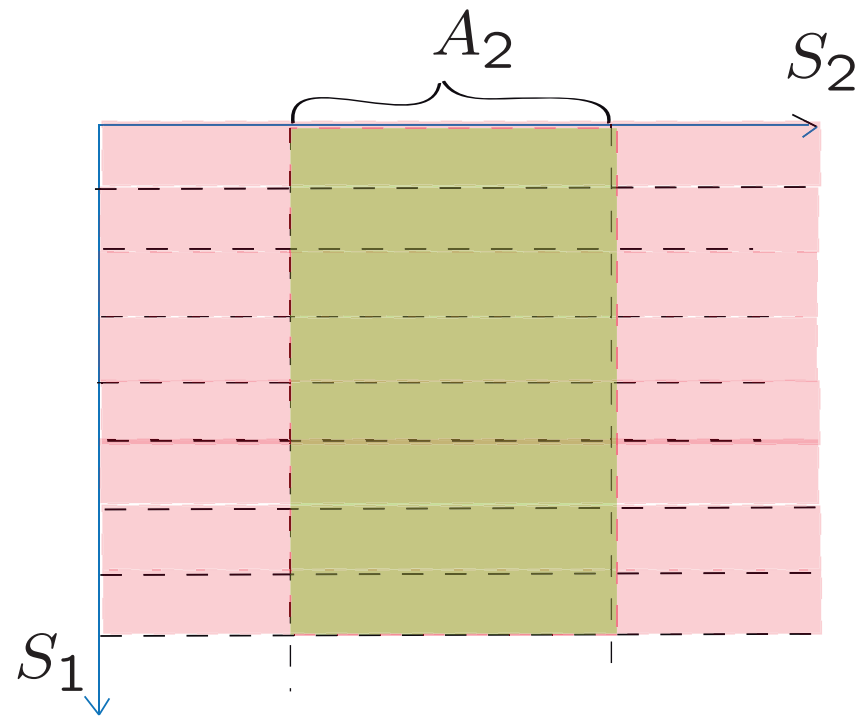
$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) .$$

Summiert über $(a_1, a_2) \in A_1 \times A_2$ erhält man daraus:



$$\begin{aligned}
 & \mathbf{P}(X_1 \in A_1, X_2 \in A_2) \\
 &= \sum_{a_1 \in A_1} \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 \in A_2).
 \end{aligned}$$

Speziell mit $A_1 := S_1$ ergibt sich:



$$\mathbf{P}(X_2 \in A_2) = \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 \in A_2).$$

“Formel von der totalen Wahrscheinlichkeit”

“Zerlegung nach dem ersten Schritt”

$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) .$$

Wir bemerken: Genau dann
hängen die Verteilungen $\mathbf{P}_{a_1}(X_2 \in \cdot)$ nicht von a_1 ab,
wenn X_1 und X_2 unabhängig sind.

$$\nu(a_1, a_2) = \rho_1(a_1) P(a_1, a_2)$$

$P(a_1, a_2)$, $a_1 \in S_1$, $a_2 \in S_2$
sind die Einträge der sogenannten
Übergangsmatrix P .

Jede einzelne Zeilensumme von P ist 1.

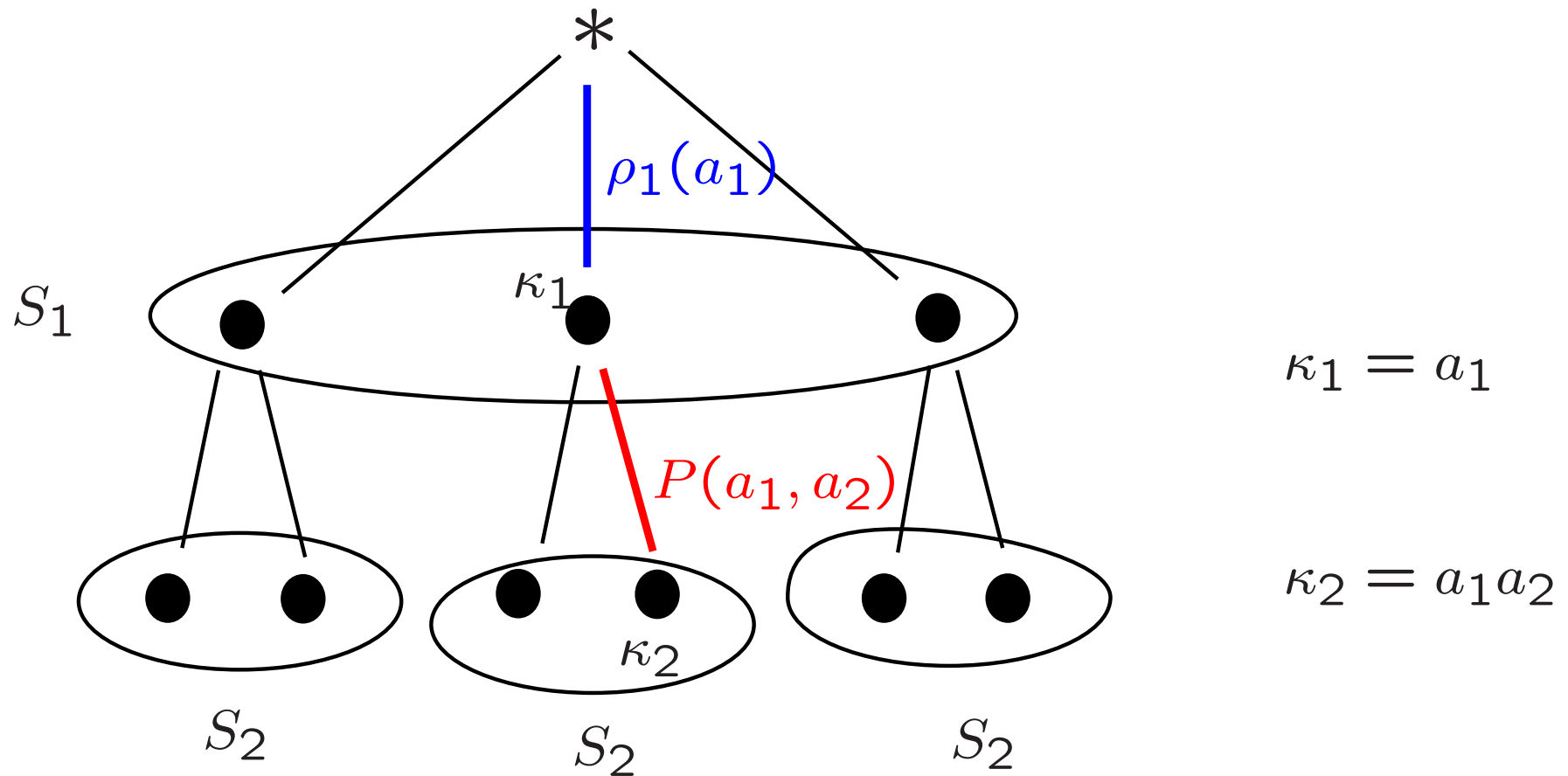
Die Zeilensummen der Matrix $\nu(.,.)$
ergeben die Einträge $\rho_1(.)$.

Die Gesamtsumme aller $\nu(a_1, a_2)$ ist 1.

Veranschaulichung durch einen Baum

Ein zweistufiges Zufallsexperiment
kann in seiner Abfolge
durch einen *Baum der Tiefe 2* veranschaulicht werden.

Beispiel mit $\#S_1 = 3$, $\#S_2 = 2$:



$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \rho_1(a_1)P(a_1, a_2).$$

(Produkt der Kantengewichte entlang des Weges von $*$ zum Knoten κ_2)

Zweistufige Zufallsexperimente - kontinuierlicher Fall:

Sei $X = (X_1, X_2)$ ein zufälliges Paar
mit Zielbereich $S = S_1 \times S_2$.

S_1 und S_2 seien Intervalle in $\mathbb{R}$.

Für jedes $a_1 \in S_1$ sei

$g_{a_1}(a_2) da_2$, $a_2 \in S_2$, eine Dichte auf S_2 .

Vorstellung: gegeben $\{X_1 = a_1\}$

hat die Zufallsvariable X_2 die Dichte $g_{a_1}(a_2) da_2$.

Schreibweise:

$$\mathbf{P}_{a_1}(X_2 \in da_2) = g_{a_1}(a_2) da_2$$

Die S_1 -wertige Zufallsvariable X_1 habe die Dichte

$$f_1(a_1) da_1, a_1 \in S_1.$$

Dann hat (X_1, X_2) die gemeinsame Dichte

$$\mathbf{P}(X_1 \in da_1, X_2 \in da_2) = f_1(a_1) g_{a_1}(a_2) da_1 da_2$$

Integriert über $A_1 \times A_2$ ergibt das:

$$\mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \int_{A_1} f_1(a_1) \int_{A_2} g_{a_1}(a_2) da_2 da_1$$

Speziell mit $A_1 = S_1$ folgt als Dichte von X_2 :

$$\mathbf{P}(X_2 \in da_2) = \left(\int_{S_1} f_1(a_1) g_{a_1}(a_2) da_1 \right) da_2$$

Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 3:

Addieren von unabhängigen Zufallsvariablen
– zweistufig aufgefasst.

Y und Z seien unabhängige $\mathbb{R}$ -wertige
Zufallsvariable.

Wie ist $Y + Z$ verteilt?

Wir können $Y + Z$ auffassen als
zweite Stufe eines Zufallsexperiments.

Die erste Stufe ist $X_1 := Y$.

Gegeben $\{Y = a\}$ ist

$$X_2 := Y + Z$$

so verteilt wie $a + Z$.

Der diskrete Fall:

Y und Z seien unabhängige $\mathbb{Z}$ -wertige Zufallsvariable.

$$\begin{aligned}\mathbf{P}(Y = a, Y + Z = b) &= \mathbf{P}(Y = a, Z = b - a) \\ &= \mathbf{P}(Y = a) \mathbf{P}(Z = b - a)\end{aligned}$$

Summation über a ergibt die “totale Wahrscheinlichkeit”:

$$\mathbf{P}(Y + Z = b) = \sum_a \mathbf{P}(Y = a) \mathbf{P}(Z = b - a)$$

(Zerlegung von $\mathbf{P}(Y + Z = b)$ nach den Ausgängen von Y ,
“Zerlegung nach dem ersten Schritt”)

$$\mathbf{P}(Y + Z = b) = \sum_a \mathbf{P}(Y = a) \mathbf{P}(Z = b - a)$$

Beispiel:

Y, Z unabhängig und $\text{Geom}(p)$ -verteilt

$$\begin{aligned} \mathbf{P}(Y + Z = b) &= \sum_{a=1}^{b-1} pq^{a-1} pq^{b-a-1} \\ &= (b-1)p^2 q^{b-2}, \quad b = 2, 3, \dots \end{aligned}$$

Dies sind die Gewichte der sogenannten

negativen Binomialverteilung mit Parametern 2, p

Diese ist die Verteilung der Anzahl der Versuche in einem p -Münzwurf bis einschließlich zum zweiten Erfolg.

Der Fall mit Dichten (Buch S. 92)

Haben Y und Z die Dichten $f(y) dy$ und $g(z) dz$,
so bekommt man analog
die gemeinsame Dichte von $(Y, Y + Z)$:

$$\begin{aligned}\mathbf{P}(Y \in da, Y + Z \in db) &= \mathbf{P}(Y \in da, a + Z \in db) \\ &= \mathbf{P}(Y \in da) \mathbf{P}(a + Z \in db) \\ &= f(a)da \, g(b - a) db\end{aligned}$$

$$\mathbf{P}(Y \in da, Y + Z \in db) = f(a) g(b - a) da db$$

Integration über a gibt die Dichte von $Y + Z$:

$$\mathbf{P}(Y + Z \in db) = \left(\int f(a) g(b - a) da \right) db .$$

Beispiel: Für Y und Z unabhängig und Exp(1)-verteilt ist

$$\begin{aligned} \mathbf{P}(Y + Z \in db) &= \left(\int_0^b e^{-a} e^{-(b-a)} da \right) db \\ &= b e^{-b} db, \quad b \geq 0. \end{aligned}$$

(Dichte der Gamma(2)-Verteilung)

Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 4:

Bedingte Verteilung

(Buch S. 111)

Bisher legten wir das Hauptaugenmerk auf den

Aufbau der gemeinsamen Verteilung von X_1 und X_2
aus der Verteilung von X_1
und Übergangswahrscheinlichkeiten

über die Multiplikationsregel:

$$P(X_1 = a_1, X_2 = a_2) := \rho_1(a_1) P(a_1, a_2)$$

im diskreten Fall,

$$P(X_1 \in da_1, X_2 \in da_2) := f_1(a_1) da_1 g_{a_1}(a_2) da_2$$

im Fall mit Dichten.

Jetzt gehen wir den **Weg zurück**:

Wir gehen aus von der **gemeinsamen Verteilung**
und berechnen aus ihr die **bedingte Verteilung**

Es gelte

$$P(X_1 = a_1, X_2 = a_2) = \nu(a_1, a_2)$$

im diskreten Fall,

$$P(X_1 \in da_1, X_2 \in da_2) = f(a_1, a_2) da_1 da_2$$

im Fall mit Dichten.

Wie man die **Verteilung von X_1** aus der
(gemeinsamen) Verteilung von (X_1, X_2) bekommt,
wissen wir schon:

Im diskreten Fall:

$$\rho_1(a_1) = \mathbf{P}(X_1 = a_1) = \sum_{a_2 \in S_2} \nu(a_1, a_2).$$

Im Fall mit Dichten:

$$f_1(a_1) da_1 = \mathbf{P}(X_1 \in da_1) = \left(\int_{S_2} f(a_1, a_2) da_2 \right) da_1.$$

Die *bedingte Verteilung* von X_2 gegeben $\{X_1 = a_1\}$ bekommt man durch Umstellen der Multiplikationsregel:

$$\mathbf{P}(X_2 \in A_2 \mid X_1 = a_1) := \frac{\mathbf{P}(X_1 = a_1, X_2 \in A_2)}{\mathbf{P}(X_1 = a_1)}$$

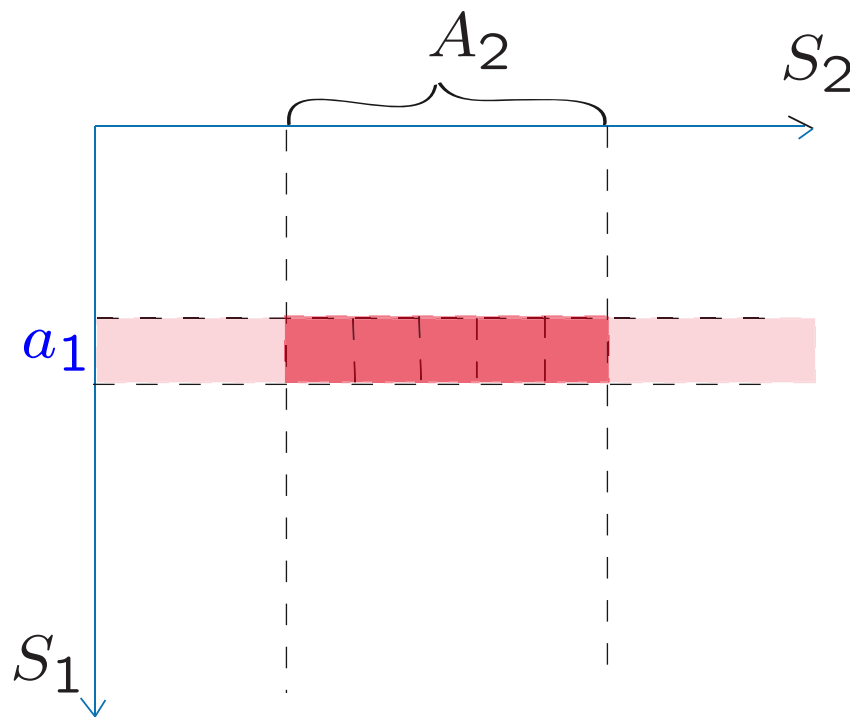
im Fall von diskretem X_1 ,

und im Fall mit Dichten

als *bedingte Dichte* von X_2 gegeben $\{X_1 = a_1\}$:

$$\mathbf{P}(X_2 \in da_2 \mid X_1 = a_1) := \frac{f(a_1, a_2)}{f_1(a_1)} da_2.$$

In der Matrix der gemeinsamen Verteilungsgewichte
 $\nu(a_1, a_2) = \mathbf{P}(X_1 = a_1, X_2 = a_2)$
ist $\mathbf{P}(X_2 \in A_2 \mid X_1 = a_1)$ das relative Gewicht von A_2
bezogen auf das Gesamtgewicht der Zeile a_1



Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 5:

Bedingte Verteilung mit X_2 als erster Stufe:

Die Formel von Bayes

(Buch S. 111-112, 116)

Bei der Untersuchung von zwei Zufallsvariablen X_1, X_2
kann man immer
zu einer **2-stufigen Betrachtungsweise** übergehen.

Man kann dabei wählen,
ob man X_1 als die erste Stufe auffasst oder als die zweite.
Wählt man X_1 als die zweite und X_2 als die erste Stufe,
dann schreibt sich die Multiplikationsformel als

$$\begin{aligned} & \mathbf{P}(X_1 = a_1, X_2 = a_2) \\ &= \mathbf{P}(X_2 = a_2) \mathbf{P}(X_1 = a_1 | X_2 = a_2). \end{aligned}$$

Mit der *Formel für die totale Wahrscheinlichkeit*

$$\mathbf{P}(X_2 = a_2) = \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{P}(X_2 = a_2 | X_1 = a_1)$$

ergibt sich daraus eine Version der *Formel von Bayes*:

$$\mathbf{P}(X_1 = a_1 | X_2 = a_2) = \frac{\mathbf{P}(X_1 = a_1) \mathbf{P}(X_2 = a_2 | X_1 = a_1)}{\sum_{a' \in S_1} \mathbf{P}(X_1 = a') \mathbf{P}(X_2 = a_2 | X_1 = a')}$$

Beispiel: Erfolgreiche Würfe beim Münzwurf:

Wann kamen die erfolgreichen unter den ersten zehn Würfeln,
gegeben man kennt ihre Anzahl?

Bei einem 10-maligen p -Münzwurf sei

K die Anzahl der Erfolge,

und $G \subset \{1, \dots, 10\}$ die zufällige Menge der Zeiten,
zu denen die Erfolge eintreten.

Die bedingte Verteilung von K gegeben $G = a$ ist klar:

$$P(K = \#a \mid G = a) = 1.$$

Was ist die bedingte Verteilung von G , gegeben $\{K = 4\}$?

Für jede 4-elementige Teilmenge a von $\{1, \dots, 10\}$ ist
$$\mathbf{P}(\{G = a\} \cap \{K = 4\}) = \mathbf{P}(G = a) = p^4(1 - p)^6.$$

Das hängt nicht von a ab.

Damit hängt auch $\mathbf{P}(G = a | K = 4)$ nicht von a ab,

also ist $\mathbf{P}(G \in \cdot | K = 4)$

die uniforme Verteilung

auf den 4-elementigen Teilmengen von $\{1, \dots, 10\}$.

Beispiel: Wie war der erste Schritt?

Es seien Y und Z unabhängige $\mathbb{Z}$ -wertige Zufallsvariable und

$$X_1 := Y, X_2 := Y + Z.$$

Wir haben gesehen:

Die bedingte Verteilung von $Y + Z$, gegeben $\{Y = a\}$,
ist die Verteilung von $a + Z$.

Was ergibt sich für die bedingte Verteilung von Y ,
gegeben $\{Y + Z = b\}$?

“Wie war der erste Schritt?”

Die Gewichte der **gemeinsame Verteilung** von Y und $Y + Z$ kennen wir schon aus Teil 3.

Die bedingte Verteilung von Y , gegeben $Y + Z = b$, hat somit die Gewichte

$$\mathbf{P}(Y = a \mid Y + Z = b) = \frac{\mathbf{P}(Y = a) \mathbf{P}(Z = b - a)}{\mathbf{P}(Y + Z = b)}.$$

Ein instruktiver Spezialfall:

Y und Z seien unabhängig und $\text{Geom}(p)$ -verteilt.

Dann ist

$$\mathbf{P}(Y = a) \mathbf{P}(Z = b - a) = q^{a-1} p q^{(b-a)-1} p = p^2 q^{b-2}.$$

Dieses hängt nicht von a ab. Also hängt auch

$$\mathbf{P}(Y = a | Y + Z = b) = \frac{p^2 q^{b-2}}{\mathbf{P}(Y + Z = b)} \text{ nicht von } a \text{ ab.}$$

Deshalb ist die bedingte Verteilung von Y gegeben

$$\{Y + Z = b\}$$

die uniforme Verteilung auf $\{1, \dots, b - 1\}$.

Das ist auch ohne Rechnung plausibel:

Gegeben, dass in einem p -Münzwurf
der zweite Erfolg beim b -ten Versuch kommt,
ist der Zeitpunkt des ersten Erfolges
uniform verteilt auf $\{1, \dots, b - 1\}$.

Beispiel: Exponentialverteilte Summanden

Y und Z seien unabhängig und $\text{Exp}(1)$ -verteilt.

Was ist die bedingte Dichte von Y , gegeben $\{Y + Z = b\}$?

Die gemeinsame Dichte von Y und $Y + Z$ ist (vgl. F8a3.6)

$$e^{-a}e^{-(b-a)} da db = e^{-b} da db, \quad 0 \leq a \leq b$$

Die Dichte von $Y + Z$ ist somit $\left(\int_0^b da\right)e^{-b} db = be^{-b}db$

Also:

$$\mathbf{P}(Y \in da \mid Y + Z = b) = \frac{e^{-b}}{be^{-b}} da = \frac{1}{b} da, \quad 0 \leq a \leq b.$$

Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 6:

Bedingte Wahrscheinlichkeiten

(Buch S. 115-117)

Definition.

Seien E_1, E_2 Ereignisse. Dann ist die
bedingte Wahrscheinlichkeit von E_2 , gegeben E_1 ,
definiert als

$$\mathbf{P}(E_2 | E_1) := \frac{\mathbf{P}(E_2 \cap E_1)}{\mathbf{P}(E_1)}$$

*... die Wahrscheinlichkeit von E_2 , wenn man schon weiß,
dass E_1 eingetreten ist.*

Dies passt in den Rahmen von Teil 4,
als Verteilungsgewicht $\mathbf{P}(I_{E_2} = 1 | I_{E_1} = 1)$
der bedingten Verteilung von I_{E_2} , gegeben $\{I_{E_1} = 1\}$.

Beispiel:

T sei $\text{Geom}(p)$ -verteilt. Für $k, \ell \in \mathbb{N}_0$ berechnen wir die bedingte Wahrscheinlichkeit von $\{T - k > \ell\}$, gegeben $\{T > k\}$.

Wegen $\{T - k > \ell\} \cap \{T > k\} = \{T > k + \ell\}$ ist

$$\mathbf{P}(T - k > \ell \mid T > k) = q^{k+\ell}/q^k = q^\ell .$$

Die bedingte Verteilung von $T - k$, gegeben $\{T > k\}$,
ist somit (wieder) gleich $\text{Geom}(p)$.

Die Kenntnis, dass T größer als k ausfällt,
ändert also die Verteilung der verbleibenden Wartezeit nicht.

Man nennt diese Eigenschaft die
Gedächtnislosigkeit der geometrischen Verteilung.

Bedingte Verteilung der verbleibenden Wartezeit: Gedächtnislosigkeit der Exponentialverteilung

Für exponentialverteiltes T zum Parameter λ gilt für $r, s > 0$

$$\mathbf{P}(T > r + s \mid T > r) = e^{-\lambda s} .$$

Die bedingte Verteilung von $T - r$, gegeben $\{T > r\}$,
ist somit gleich $\text{Exp}(\lambda)$.

Die Kenntnis, dass T einen Wert größer als r annimmt,
ändert also die Verteilung der verbleibenden Wartezeit nicht.

Für zwei Ereignisse E_1, E_2 lautet die
Formel von Bayes:

$$\mathbf{P}(E_1|E_2) = \frac{\mathbf{P}(E_1)P(E_2|E_1)}{\mathbf{P}(E_1)P(E_2|E_1) + \mathbf{P}(E_1^c)P(E_2|E_1^c)}$$

Beispiel:

In einer Population haben 0.1% eine bestimmte Krankheit.

Bei einer bestimmten Reihenuntersuchung wird
eine kranke Person in 100% der Fälle positiv getestet,
eine gesunde Person in 1%.

Wie groß ist die Wahrscheinlichkeit, dass
eine positiv getestete Person wirklich krank ist?

Hier ist erst einmal ein Rezept
für eine intuitive Überschlagsrechnung:

In einer Population von 1000
sind 999 gesund und einer krank. Von den 999 Gesunden
werden ca 10 (falsch) positiv diagnostiziert.

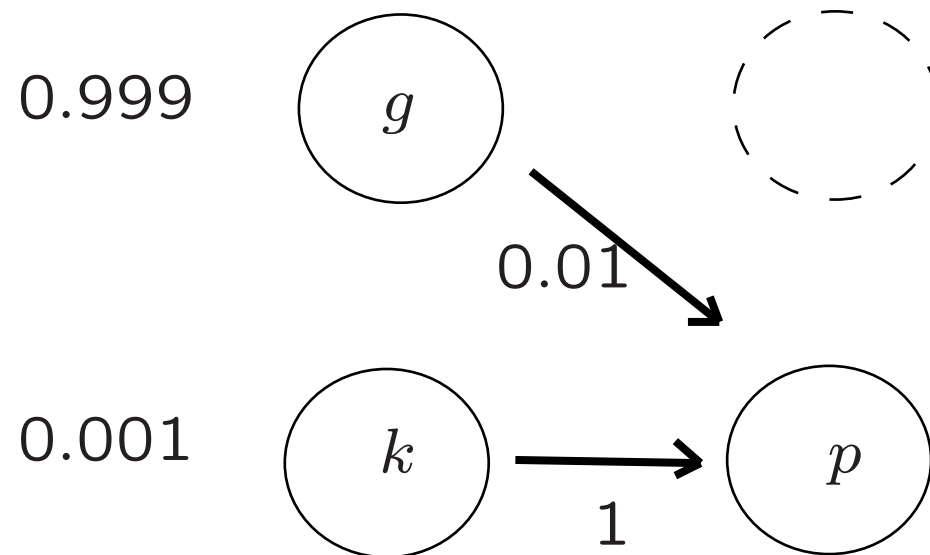
Von 11 positiv Diagnostizierten
ist also im Schnitt nur einer krank.

Hier ist eine **Formalisierung**:

X_1 sei der Gesundheitszustand (mit Werten in $S_1 = \{g, k\}$),

X_2 der Testbefund (mit Werten in $S_2 = \{p, n\}$).

(X_1, X_2) entsteht über ein zweistufiges Experiment:



$$\mathbf{P}(X_1 = k | X_2 = p) = \frac{\mathbf{P}(X_1 = k, X_2 = p)}{\mathbf{P}(X_2 = p)}$$

$$= \frac{0.001 \cdot 1}{0.999 \cdot 0.01 + 0.001 \cdot 1} \approx \frac{1}{11}$$

Vorlesung 8b

Mehrstufige Zufallsexperimente

Teil 1

Mehrstufigkeit und Multiplikationsregel

(Buch S. 94-95)

bisher Zweistufigkeit nur von “heute” zu “morgen”
jetzt zusätzlich von “heute & morgen” zu “übermorgen”, etc.

Für jedes $i = 1, \dots, n - 1$ hat man

Übergangswahrscheinlichkeiten

$$P(a_1 \dots a_i, a_{i+1}) = P_{a_1 \dots a_i}(X_{i+1} = a_{i+1}) ,$$

die angeben,

mit welcher Wahrscheinlichkeit in der $(i + 1)$ -ten Stufe

das Ereignis $\{X_{i+1} = a_{i+1}\}$ eintritt,

gegeben das Eintreten von $\{X_1 = a_1, \dots, X_i = a_i\}$.

Die gemeinsame Verteilung von $X_1, \dots, X_n$
ergibt sich rekursiv als

Multiplikationsregel

$$\begin{aligned} & \mathbf{P}(X_1 = a_1, \dots, X_n = a_n) \\ &= \mathbf{P}(X_1 = a_1, \dots, X_{n-1} = a_{n-1}) P(a_1 \dots a_{n-1}, a_n) \\ & \quad = \dots \\ &= \rho(a_1) P(a_1, a_2) \dots P(a_1 \dots a_{n-1}, a_n) \\ &=: \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) \dots \mathbf{P}_{a_1 \dots a_{n-1}}(X_n = a_n) \end{aligned}$$

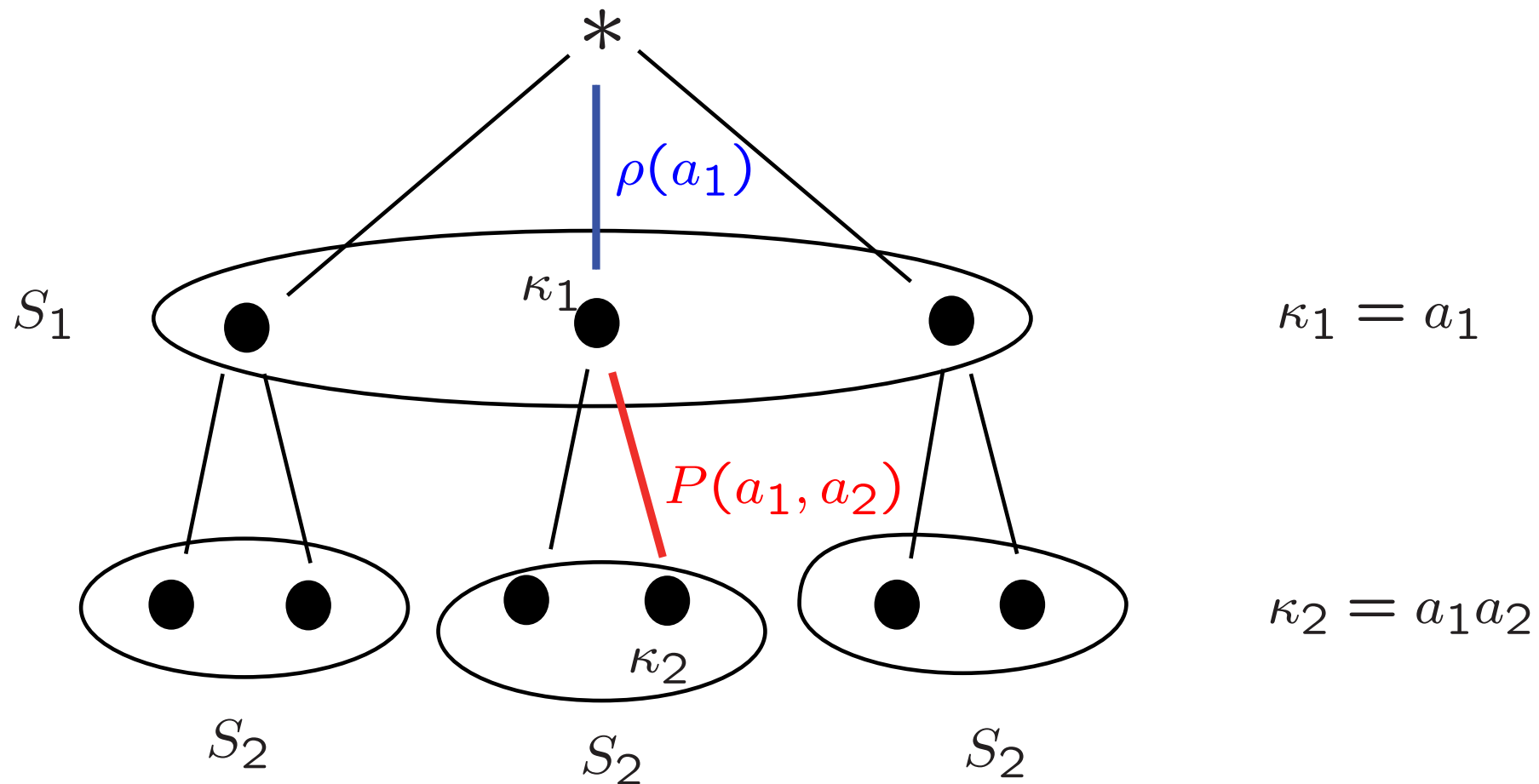
Dabei ist $a_i \in S_i := \text{Wertebereich von } X_i$.

Mehrstufige Zufallsexperimente kann man
durch Bäume veranschaulichen

(Buch S. 95)

Wir erinnern uns an die

**Veranschaulichung von zweistufigen Experimenten
durch *Bäume*:**



$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \rho(a_1)P(a_1, a_2).$$

(Produkt der Kantengewichte von $*$ zum Knoten κ_2)

Was 2 recht ist, ist $i + 1$ billig:

$\kappa_i = a_1 \dots a_i$ ist ein *Knoten der Tiefe i*

Die *Nachfolger* von κ_i sind von der Form $\kappa_i a_{i+1}$

Die *Kanten* des Baums erhalten die Gewichte

$$g(*, \kappa_1) := \rho(a_1), \quad g(\kappa_i, \kappa_{i+1}) := P(a_1 \dots a_i, a_{i+1})$$

mit $\kappa_1 = a_1$, $\kappa_i = a_1 \dots a_i$, $\kappa_{i+1} = a_1 \dots a_{i+1}$.

Die Wahrscheinlichkeit, in einem bestimmten Blatt zu enden,
ergibt sich als Produkt der Kantengewichte
entlang des Weges von der Wurzel zum Blatt.

Vorlesung 8b

Mehrstufige Zufallsexperimente

Teil 2

Die Pólya-Urne

Oder

Wo Tauben sind, fliegen Tauben zu

(Buch S. 94-95)

In einer Urne befinden sich anfangs
eine rote und eine blaue Kugel.

In jedem Schritt wird eine Kugel rein zufällig gezogen und
gemeinsam mit einer zusätzlichen Kugel derselben Farbe
zurückgelegt.

Die Zufallsvariable Z_i mit Werten in $\{0, 1\}$ bezeichne die
im i -ten Zug vorgefundene Farbe (1 für rot, 0 für blau).

Wir nennen dann $(Z_1, \dots, Z_n)$ auch
eine (Standard-) *Pólya-Folge* der Länge n .

$$P(Z_1 = 0) = P(Z_1 = 1) = \frac{1}{2}$$

$$P_0(Z_2 = 0) = \frac{2}{3}$$

(das ist die **W'keit**, beim zweiten Zug **blau** zu ziehen,
wenn beim ersten Zug **blau** gezogen wurde)

$$P_0(Z_2 = 1) = \frac{1}{3}$$

$$P_{00}(Z_3 = 0) = \frac{3}{4}$$

$$P_{01}(Z_3 = 0) = \frac{2}{4}$$

u. s. w.

Die Übergangswahrscheinlichkeiten sind

$$\mathbf{P}_{a_1 \dots a_i}(Z_{i+1} = 0) = \frac{1 + \#\{j : 1 \leq j \leq i, a_j = 0\}}{2 + i}$$

$$\mathbf{P}_{a_1 \dots a_i}(Z_{i+1} = 1) = \frac{1 + \#\{j : 1 \leq j \leq i, a_j = 1\}}{2 + i}$$

Z. B. ist

$$\begin{aligned} \mathbf{P}((Z_1, \dots, Z_8) = (\textcolor{red}{1}, \textcolor{red}{1}, \textcolor{blue}{0}, \textcolor{red}{1}, \textcolor{blue}{0}, \textcolor{blue}{0}, \textcolor{red}{1}, \textcolor{red}{1})) &= \frac{\textcolor{red}{1}}{2} \frac{\textcolor{red}{2}}{3} \frac{\textcolor{blue}{1}}{4} \frac{\textcolor{red}{3}}{5} \frac{\textcolor{blue}{2}}{6} \frac{\textcolor{blue}{3}}{7} \frac{\textcolor{red}{4}}{8} \frac{\textcolor{red}{5}}{9} \\ &= \frac{\textcolor{red}{5!} \textcolor{blue}{3!}}{9!}. \end{aligned}$$

Für $0 \leq k \leq n$ hat jede 01-Zugfolge $(a_1, \dots, a_n)$
mit $a_1 + \dots + a_n = \textcolor{red}{k}$ dieselbe Wahrscheinlichkeit, nämlich

$$(*) \quad \mathbf{P}((Z_1, \dots, Z_n) = (a_1, \dots, a_n)) = \frac{\textcolor{red}{k!} \textcolor{blue}{(n-k)!}}{(n+1)!} = \frac{1}{n+1} \binom{n}{k}^{-1}$$

Für $0 \leq k \leq n$ hat jede 01-Zugfolge $(a_1, \dots, a_n)$ mit $a_1 + \dots + a_n = k$ dieselbe Wahrscheinlichkeit, nämlich

$$(*) \quad \mathbf{P}((Z_1, \dots, Z_n) = (a_1, \dots, a_n)) = \frac{1}{n+1} \binom{n}{k}^{-1}.$$

Es gibt $\binom{n}{k}$ derartige Zugfolgen.

Also ist

$$\mathbf{P}(Z_1 + \dots + Z_n = k) = \frac{1}{n+1}, \quad k = 0, \dots, n.$$

Fazit:

Die Anzahl der
nach n Zügen hinzugekommenen roten Kugeln
ist uniform verteilt in $\{0, 1, \dots, n\}$.

Pólya-Urne mit g Farben:

(Buch S. 95)

Wieder wird in jedem Zug die gezogene Kugel zusammen mit einer gleichfarbigen Kugel zurückgelegt.

Vor dem ersten Zug sind g Kugeln in der Urne,
von jeder Farbe eine.

Wie sind die Zugänge nach n Zügen verteilt?

$$S_{n,g} = \{k = (k_1, \dots, k_g) : k_j \in \mathbb{N}_0, k_1 + \dots + k_g = n\}$$

ist die Menge der Besetzungen von g Plätzen mit n Objekten
(vgl Vorlesung 2a3).

$X_{jn} := \#$ Zugänge der Farbe j nach n Schritten.

$$X_n := (X_{1n}, \dots, X_{gn})$$

ist eine Zufallsvariable mit Wertebereich $S_{n,g}$.

Wie sieht die Verteilung von X_n aus?

Sei $(k_1, \dots, k_g) \in S_{n,g}$,
d.h. $k_j \in \mathbb{N}_0$ mit $k_1 + \dots + k_g = n$.

Man sieht wie im Fall $g = 2$:

Alle möglichen Zugfolgen
von $(1, \dots, 1)$ nach $(1 + k_1, \dots, 1 + k_g)$
haben dieselbe Wahrscheinlichkeit, nämlich

$$\frac{k_1! \cdots k_g!}{g \cdot (g+1) \cdots (n+g-1)} = \frac{k_1! \cdots k_g!}{(n+g-1)!} (g-1)! .$$

Es gibt $\binom{n}{k_1, \dots, k_g} = \frac{n!}{k_1! \dots k_g!}$ solche Zugfolgen. Also ist

$$\begin{aligned} \mathbf{P}(X_{1n} = k_1, \dots, X_{gn} = k_g) &= n! \frac{(g-1)!}{(n+g-1)!} \\ &= \frac{1}{\binom{n+g-1}{n}} \end{aligned}$$

d.h. $(X_{1n}, \dots, X_{gn})$ ist uniform verteilt auf $S_{n,g}$.

Fazit:

Die Pólya-Urne mit Anfangsbesetzung $(1, \dots, 1)$
liefert uniform verteilte Besetzungen!

Vorlesung 9a

Markovketten

Teil 1

Markovketten als
spezielle mehrstufige Zufallsexperimente

(Buch S. 94)

Zur Erinnerung:

Bei mehrstufigen Zufallsexperimenten

hat man für jedes $i = 1, 2, \dots$

Übergangswahrscheinlichkeiten

$$P(a_1 \dots a_i, a_{i+1}) = P_{a_1 \dots a_i}(X_{i+1} = a_{i+1}) ,$$

die angeben,

mit welcher Wahrscheinlichkeit in der $(i + 1)$ -ten Stufe

das Ereignis $\{X_{i+1} = a_{i+1}\}$ eintritt,

gegeben das Eintreten von $\{X_1 = a_1, \dots, X_i = a_i\}$.

Eine wichtige Beispielklasse mehrstufiger Zufallsexperimente:

alle X_i haben ein-und denselben Wertebereich S

und die Übergangswahrscheinlichkeiten der nächsten Stufe

hängen nur von der aktuellen Stufe ab

(und nicht von den vorhergehenden):

$$P(\dots a_{i-2} \textcolor{red}{a_{i-1}}, a_i) = \textcolor{red}{P}(a_{i-1}, a_i)$$

In dem Fall spricht man von einer **Markovkette**
auf dem Zustandsraum S mit Übergangsmatrix $\textcolor{red}{P}$.

Die Stufen sind jetzt mit $i = 0, 1, 2, \dots$ indiziert.

Man denkt sich die **Übergangsmatrix** P als fest
und notiert die **Verteilung** ρ von X_0 (die “Startverteilung”)
als Subskript bei der Wahrscheinlichkeit P .

Also insbesondere:

$$P_{\rho}(X_0 = a_0) = \rho(a_0).$$

Die Multiplikationsregel ergibt:

$$\begin{aligned} & P_{\rho}(X_0 = a_0, \dots, X_n = a_n) \\ &= \rho(a_0) P(a_0, a_1) \cdots P(a_{n-1}, a_n) \end{aligned}$$

Startet die Kette in $a \in S$,
dann ist ρ die auf a konzentrierte Verteilung
(notiert als $\rho = \delta_a$).

Statt P_{δ_a} schreibt man auch P_a
und erhält

$$P_a(X_0 = a) = 1,$$

$$P_a(X_1 = a_1, \dots, X_n = a_n) = P(a, a_1) \cdots P(a_{n-1}, a_n) .$$

Beispiele für Markovketten

(Buch S. 98)

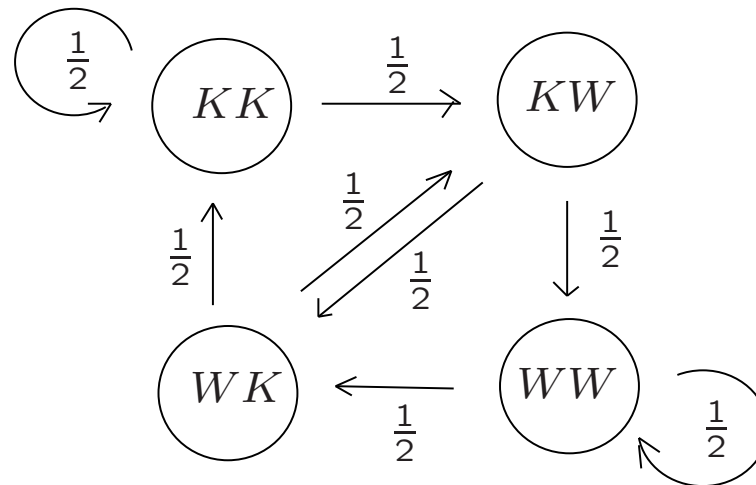
Beispiel 1:

Muster der Länge 2 beim fairen Münzwurf

$Z_0, Z_1, \dots$ unabhängig und uniform verteilt auf $\{K, W\}$,

$$X_n := (Z_n, Z_{n+1}), \quad n = 0, 1, 2, \dots$$

Graph der Übergangswahrscheinlichkeiten:



Beispiel 2:

(p, q) -Irrfahrt auf $\mathbb{Z}$:

Sei $p \in [0, 1]$, $q := 1 - p$, $S := \mathbb{Z}$

$$P(k, k + 1) := p, \quad P(k, k - 1) := q$$

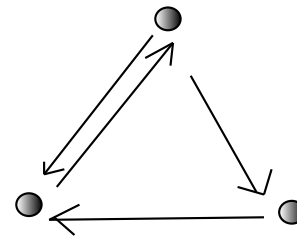
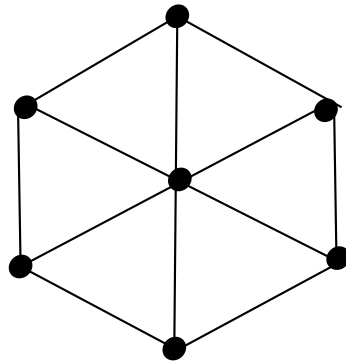
Für die Markovkette X mit Übergangsmatrix P und Start in a ist wegen der Multiplikationsregel $(X_0, X_1, \dots, X_n)$ so verteilt wie $(a, a + Z_1, \dots, a + Z_1 + \dots + Z_n)$, wobei $Z_1, Z_2, \dots$ unabhängig sind mit $\mathbf{P}(Z_i = 1) = p$, $\mathbf{P}(Z_i = -1) = q$.

Daraus sieht man sofort:

$$\mathbf{E}_a[X_n] = \mathbf{E}[a + Z_1 + \dots + Z_n] = a + n(p - q),$$

$$\mathbf{Var}_a[X_n] = \mathbf{Var}[a + Z_1 + \dots + Z_n] = 4npq.$$

Beispiel 3:
Einfache Irrfahrt
auf einem (ungerichteten oder gerichteten) Graphen



$S :=$ die Menge der Knoten.

Der nächste Schritt erfolgt jeweils
zu einem rein zufällig ausgewählten Nachbarn.

Beispiel 4:

Die Pólya-Urne als Markovkette auf $\mathbb{N}^2$

$$S = \{(r, b) : r, b \in \mathbb{N}\} = \mathbb{N}^2$$

$$P((r, b), (r + 1, b)) := \frac{r}{r + b},$$

$$P((r, b), (r, b + 1)) := \frac{b}{r + b}.$$

Das modelliert dieselbe Situation wie in Vorlesung 8b2,
allerdings “sparsamer”:

als aktueller Zustand wird nicht der gesamte bisherige Pfad, sondern nur
die Anzahl der roten und blauen Kugeln in der Urne mitgeführt

Vorlesung 9a

Markovketten

Teil 2

Zerlegung nach dem ersten Schritt

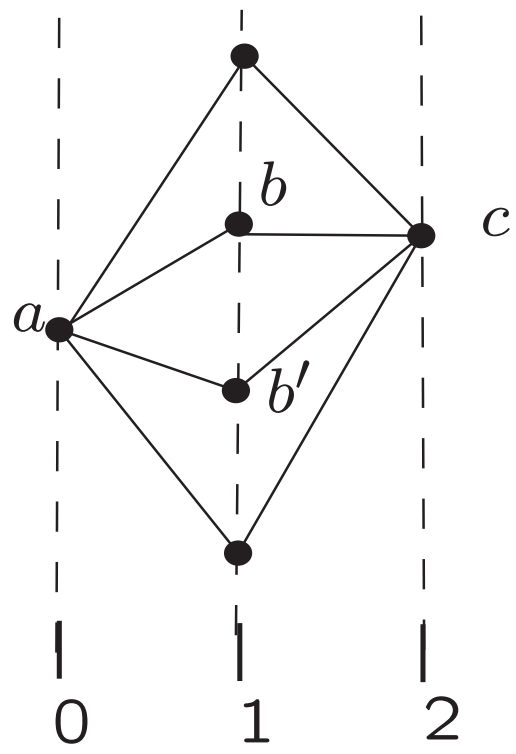
(Buch S. 94)

$$P_a(X_1 = b, X_2 = c) = P(a, b)P(b, c).$$

Summation über $b \in S$:

$$P_a(X_2 = c) = \sum_{b \in S} P(a, b)P(b, c)$$

Zerlegung der Verteilung von X_2 nach X_1 ,
Zerlegung nach dem ersten Schritt



$$\mathbf{P}_a(X_2 = c) = \sum_{b \in S} P(a, b)P(b, c)$$

Die **rechte Seite** lässt sich übrigens
als Matrixprodukt verstehen: Für

$$P^2 := P \cdot P :=$$

das Produkt der Matrix P mit sich selbst

gilt

$$P^2(a, c) = \sum_{b \in S} P(a, b)P(b, c)$$

Zerlegung nach dem ersten Schritt:

jetzt für n statt 2:

$$\begin{aligned} & \mathbf{P}_a(X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) \\ &= P(a, a_1)P(a_1, a_2) \cdots P(a_{n-1}, a_n) \\ &= P(a, a_1) \mathbf{P}_{a_1}(X_1 = a_2, \dots, X_{n-1} = a_n) \end{aligned}$$

$$\begin{aligned}
& \mathbf{P}_a(X_n = c) \\
&= \sum_{b, a_2, \dots, a_{n-1} \in S} P(a, b) P(b, a_2) \cdots P(a_{n-1}, c) \\
&= \sum_{b \in S} P(a, b) \sum_{a_2, \dots, a_{n-1} \in S} P(b, a_2) \cdots P(a_{n-1}, c)
\end{aligned}$$

Damit bekommen wir die

**Zerlegung der Verteilung von X_n
nach dem ersten Schritt:**

$$\mathbf{P}_a(X_n = c) = \sum_{b \in S} P(a, b) \mathbf{P}_b(X_{n-1} = c) .$$

$$\mathbf{P}_a(X_n = c)$$

$$= \sum_{a_1, a_2, \dots, a_{n-1} \in S} P(a, a_1) P(a_1, a_2) \cdots P(a_{n-1}, c)$$

Die **rechte Seite** ist übrigens nichts anderes als der
 Eintrag in Zeile namens a und Spalte namens c
 der Matrix $P^n := P \cdots P$

(also der n -ten Potenz von P bzg. der Matrixmultiplikation)

Vorlesung 9a

Markovketten

Teil 3

Treffwahrscheinlichkeiten

(Buch S. 100-104)

In Teil 2 hatten wir eine Markovkette $(X_i)_{i \in \mathbb{N}_0}$
in Gedanken laufen lassen bis zu einem festen Zeitpunkt n
und nach dem ersten Schritt zerlegt.

Jetzt lassen wir sie laufen, bis sie
erstmal eine bestimmte Menge $D \subset S$ trifft,
und zerlegen wieder nach dem ersten Schritt.

Diese Zerlegung eignet sich wunderbar zur Berechnung von
Treffwahrscheinlichkeiten.

Treffwahrscheinlichkeiten

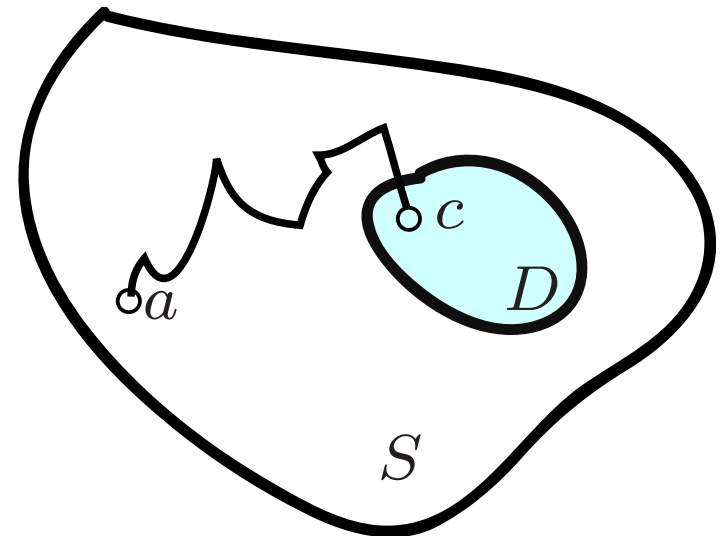
Die Frage:

P sei eine Übergangsmatrix auf der Menge S

X sei Markovkette mit Übergangsmatrix P .

$D \subset S$, $c \in D$ seien fest.

Wie wahrscheinlich ist es,
dass der in $a \in S$ startende Pfad
die Menge D erstmals
im Zustand c trifft?



Treffwahrscheinlichkeiten

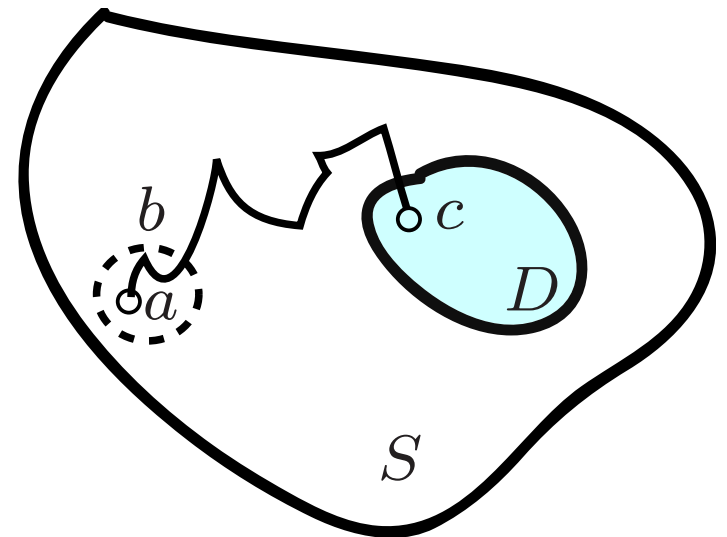
Die Antwort:

Sei $w(a)$ die gesuchte Wahrscheinlichkeit.

Die Zahlen $w(a)$, $a \in S$, erfüllen das Gleichungssystem

$$w(a) = \sum_{b \in S} P(a, b) w(b)$$

für $a \in S \setminus D$



Treffwahrscheinlichkeiten

Die Antwort:

Sei $w(a)$ die gesuchte Wahrscheinlichkeit.

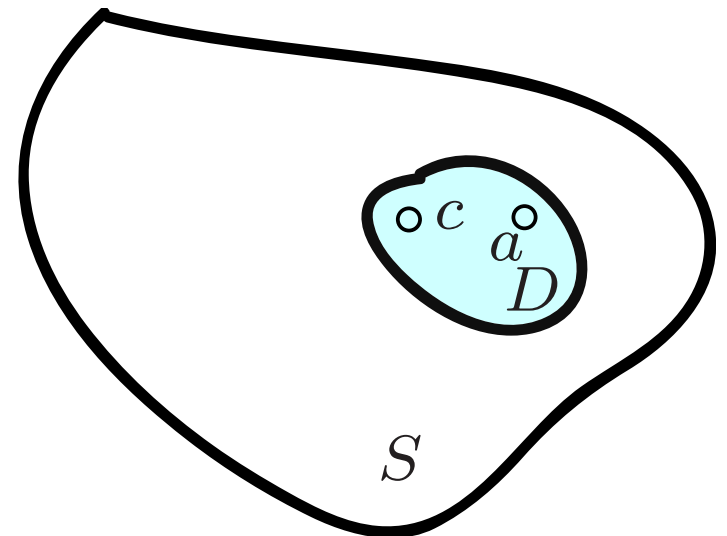
Die Zahlen $w(a)$, $a \in S$, erfüllen das Gleichungssystem

$$w(a) = \sum_{b \in S} P(a, b) w(b)$$

für $a \in S \setminus D$,

$$w(a) = \delta_{ac} \quad \text{für } a \in D.$$

Wir wenden uns jetzt
dem Beweis dieser Aussage zu.



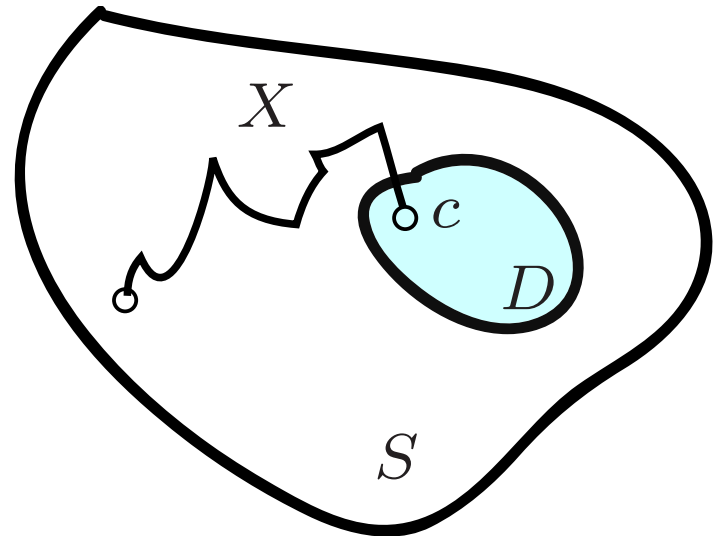
Treffereignisse und erste Treffzeit

$E_n := \{ \text{der zufällige Pfad } X$
trifft die Menge D
erstmalig in n Schritten,
und zwar im Punkt $c \}$

$$T_D := \min\{n : X_n \in D\}$$

$$E_n = \{T_D = n, X_{T_D} = c\}$$

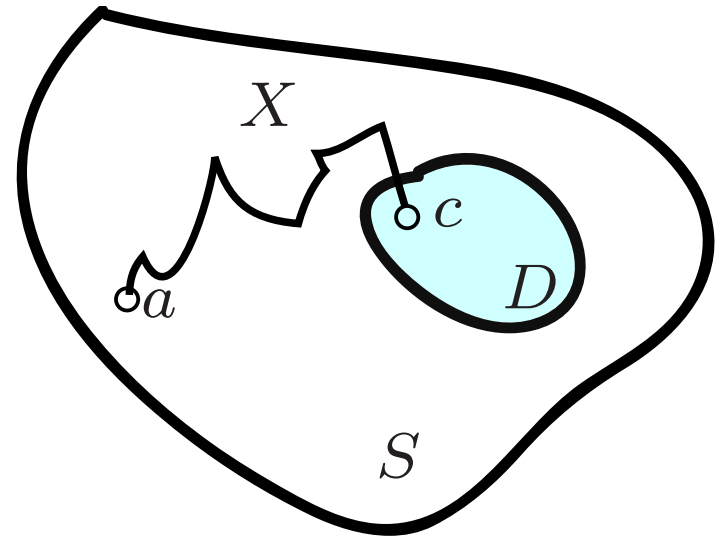
$$E_\infty := \{T_D = \infty\}$$



Erste Treffzeit:

$$T_D := \min\{n : X_n \in D\}$$

$$E_n = \{T_D = n, X_{T_D} = c\}$$



Für $b \in S$ ist

$$\mathbf{P}_b(E_0) = \delta_{bc} := \begin{cases} 1 & \text{für } b = c, \\ 0 & \text{sonst.} \end{cases}$$

Für $a \notin D$ ist

$$\mathbf{P}_a(E_1) = P(a, c) = \sum_{b \in S} P(a, b) \mathbf{P}_b(E_0)$$

$$T_D = \min\{n : X_n \in D\}$$

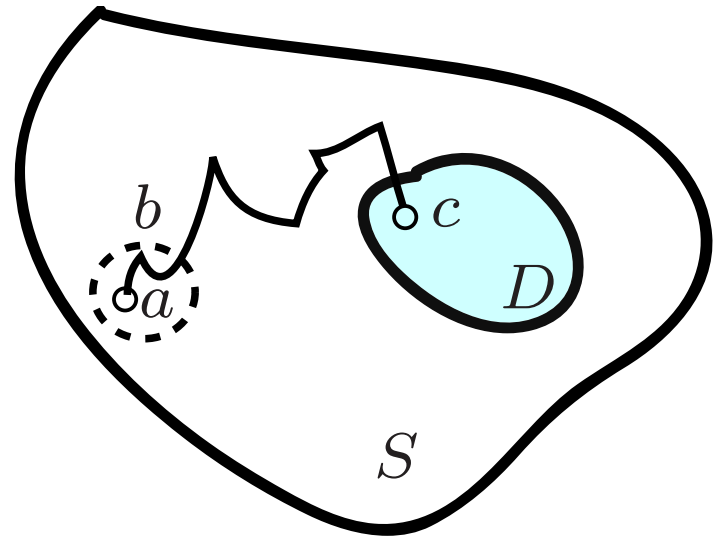
$$E_n = \{T_D = n, X_{T_D} = c\}$$

Für $n \geq 1, a \notin D$:

$$\mathbf{P}_a(E_n) = \sum_{b \in S} P(a, b) \mathbf{P}_b(E_{n-1})$$

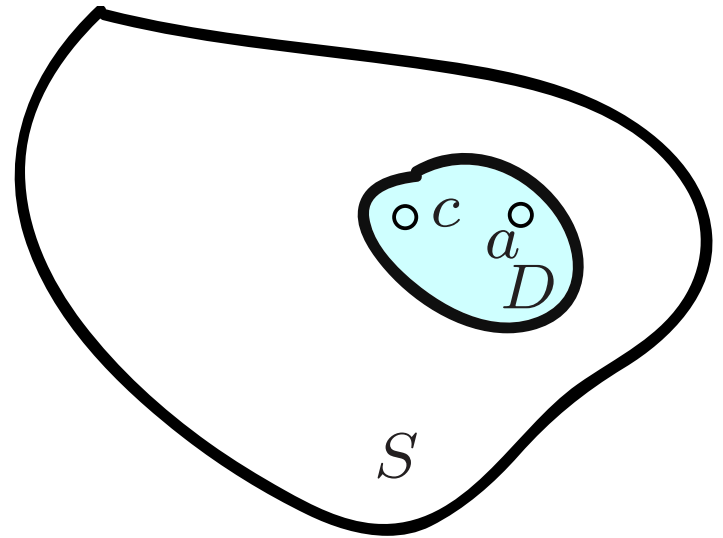
Summation über $n \geq 1$, mit $E := \{T_D < \infty, X_{T_D} = c\}$:

$$\mathbf{P}_a(E) = \sum_{b \in S} P(a, b) \mathbf{P}_b(E), \quad a \notin D$$



$$T_D = \min\{n : X_n \in D\}$$

$$E = \{T_D < \infty, X_{T_D} = c\}$$



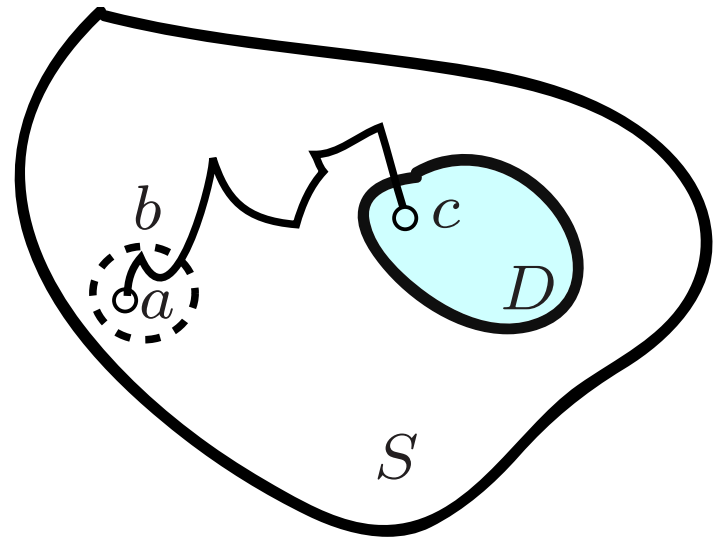
Bei Start in $a \in D$ ist $T_D = 0$ und $X_{T_D} = a$,
also ist für $a \in D$

$$\mathbf{P}_a(E) = \delta_{ac} = \begin{cases} 1 & \text{für } a = c, \\ 0 & \text{sonst.} \end{cases}$$

Treffwahrscheinlichkeiten

$$T_D = \min\{n : X_n \in D\}$$

$$E = \{T_D < \infty, X_{T_D} = c\}$$



Fazit:

Die Abbildung $w : a \mapsto \mathbf{P}_a(E)$ erfüllt das Gleichungssystem

$$\sum_{b \in S} P(a, b) w(b) = w(a) \text{ für } a \in S \setminus D, \quad w(a) = \delta_{ac} \text{ für } a \in D$$

Beispiel A: Gewinn oder Ruin?

Eine einfache Irrfahrt auf $\mathbb{Z}$ starte im Punkt 3.

Mit welcher W'keit erreicht sie den Punkt $c = 10$,
bevor sie zum Nullpunkt kommt?

“Zerlegung nach dem ersten Schritt” und Randbedingungen:

$$w(a) = \frac{1}{2}w(a-1) + \frac{1}{2}w(a+1), \quad a = 1, \dots, c-1,$$
$$w(0) = 0, \quad w(c) = 1.$$

Fazit: Die $w(a)$ liegen auf einer Geraden,

$$w(a) = \beta a + \gamma \quad \text{mit } \gamma = 0, \beta = 1/c.$$

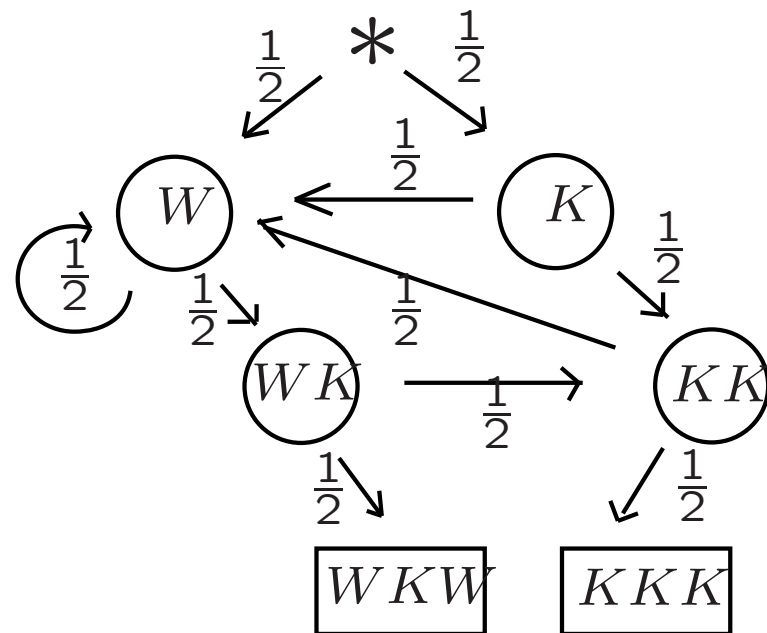
Die gesuchte Wahrscheinlichkeit ist also $3/10$.

Beispiel B

Welches Muster kommt eher?

Mit welcher W'keit kommt beim fairen Münzwurf das Muster KKK früher als das Muster WKW ?

Hier ist ein
“reduzierter Graph”
der relevanten
Zustände
und Übergänge:

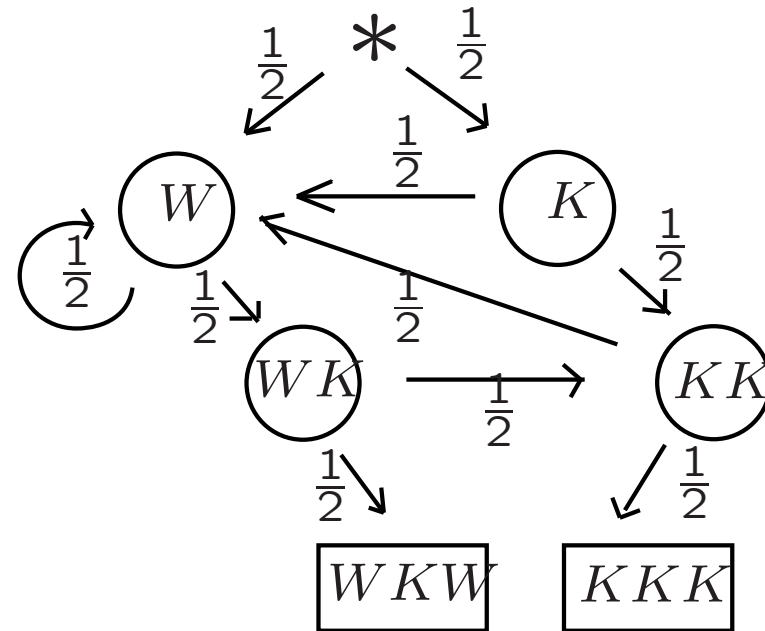


Für $w(a) :=$

$\mathbf{P}_a(\text{Spiel endet in } KKK)$

ergibt sich das

Gleichungssystem



$$w(KK) = \frac{1}{2} + \frac{1}{2}w(W) , \quad w(WK) = \frac{1}{2}w(KK)$$

$$w(K) = \frac{1}{2}w(KK) + \frac{1}{2}w(W) , \quad w(W) = \frac{1}{2}w(WK) + \frac{1}{2}w(W) .$$

und daraus

$$w(W) = \frac{1}{3}, \quad w(K) = \frac{1}{2}, \quad w(*) = \frac{1}{2}w(W) + \frac{1}{2}w(K) = \frac{5}{12} .$$

Beispiel C: **Kommt man je zurück?**

Einfache Irrfahrt auf $\mathbb{Z}$, $D := \{0\}$, $T_0 := T_D$.

$w_a := \mathbf{P}_a(T_0 < \infty)$ erfüllt

$$w(0) = 1 \text{ und } w(a) = \frac{1}{2}w(a-1) + \frac{1}{2}w(a+1), \quad a \neq 0,$$

Also existieren Zahlen $c, d, \tilde{c}, \tilde{d}$ mit

$$w(a) = ca + d \text{ auf } \mathbb{Z}_+ \text{ und } w(a) = \tilde{c}a + \tilde{d} \text{ auf } \mathbb{Z}_-.$$

Wegen $0 \leq w(a) \leq 1$ folgt $w(a) \equiv 1$, also

$$\mathbf{P}_0(\text{Rückkehr in endlicher Zeit}) = \frac{1}{2}w(-1) + \frac{1}{2}w(1) = 1.$$

Vorlesung 9a

Markovketten

Teil 4

Erwartete Treffzeiten

(Buch S. 105-106)

Sei X eine Markovkette mit Zustandsraum S
und Übergangsmatrix P .

Für eine Teilmenge $D \subset S$ ist
$$T_D := \min\{n : n \geq 0, X_n \in D\}$$

die erste Treffzeit von D .

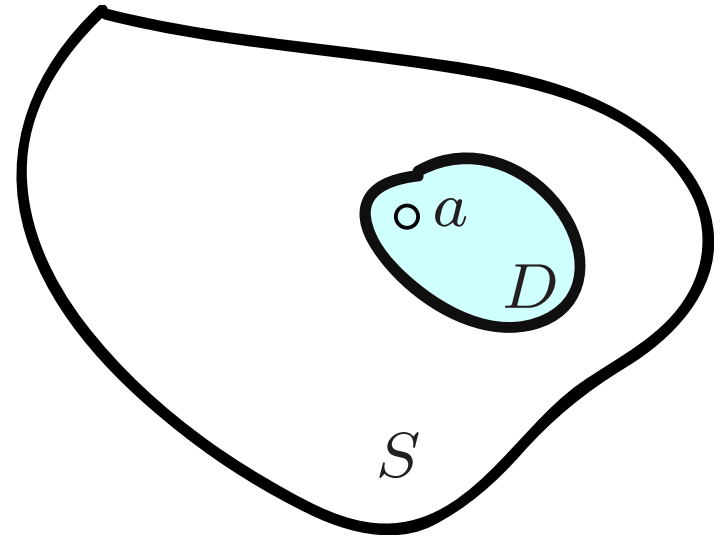
Es geht um die Berechnung von $\mathbf{E}_a[T_D]$:

Für $a \in D$

ist $\mathbf{P}_a(T_D = 0) = 1$,

also

$\mathbf{E}_a[T_D] = 0$.



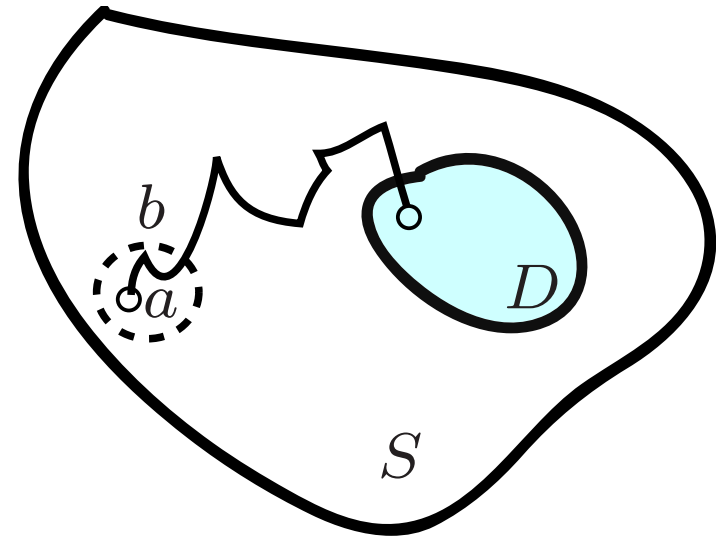
Für $a \notin D$:

Zerlegung von $\mathbf{E}_a[T_D]$
nach dem ersten Schritt:

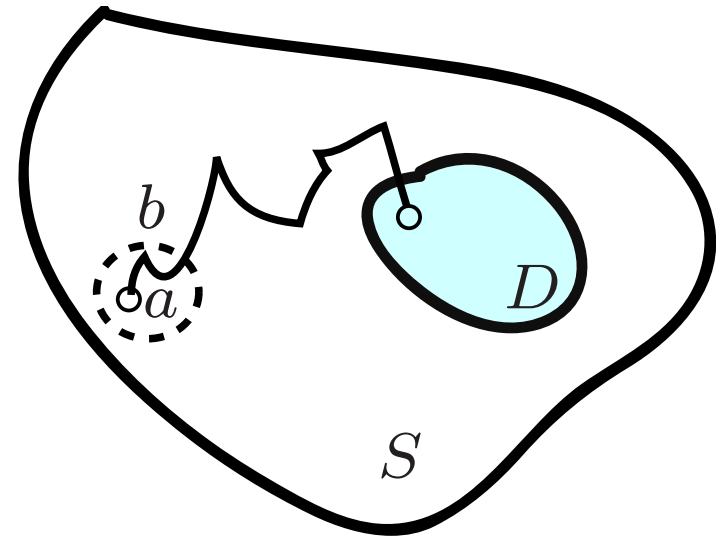
Erst ein Schritt
von a nach b gemäß $P(a, b)$,
dann “Neustart” in b :

$$\mathbf{E}_a[T_D] = 1 + \sum_{b \in S} P(a, b) \mathbf{E}_b[T_D]$$

(Formales Argument hierfür: siehe Buch Seite 106)



Fazit:



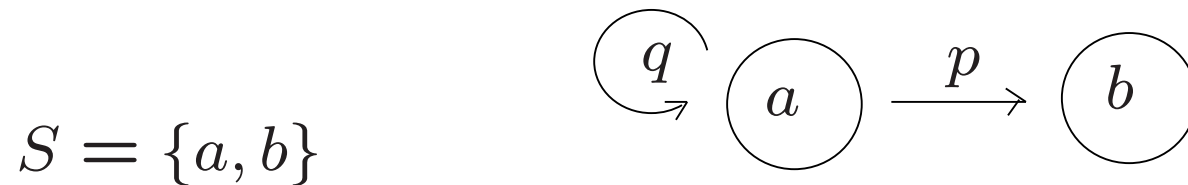
$a \mapsto e(a) := \mathbf{E}_a[T_D]$ erfüllt das Gleichungssystem

$$e(a) = 1 + \sum_{b \in S} P(a, b) e(b) \quad \text{für } a \notin D,$$

$$e(a) = 0 \quad \text{für } a \in D.$$

Beispiel 1:

Erwartete Zeit bis zum ersten Erfolg.



$$\mathbf{E}_a[T_b] = 1 + q\mathbf{E}_a[T_b] + p\mathbf{E}_b[T_b] .$$

Wegen $\mathbf{E}_b[T_b] = 0$ wird dies gelöst durch

$$\mathbf{E}_a[T_b] = 1/p.$$

Beispiel 2:

Einfache Irrfahrt auf den ganzen Zahlen: $\mathbf{E}_0[T_1] = ?$

Zerlegung nach dem ersten Schritt:

$$\mathbf{E}_0[T_1] = 1 + \frac{1}{2}0 + \frac{1}{2}\mathbf{E}_{-1}[T_1]$$

Andererseits gilt:

$$\mathbf{E}_{-1}[T_1] = \mathbf{E}_{-1}[T_0] + \mathbf{E}_0[T_1] = 2\mathbf{E}_0[T_1]$$

Zusammen: $\mathbf{E}_0[T_1] = 1 + \mathbf{E}_0[T_1]$,

mit der Lösung $\mathbf{E}_0[T_1] = \infty$.

Vorlesung 9b

Markovketten II

Teil 1:

Transport von Verteilungen

(Buch S. 107-108)

$(X_n)_{n \in \mathbb{N}_0}$ sei eine Markovkette auf dem Zustandsraum S mit Übergangsw'keit P und Startverteilung ρ_0 .

Dann gilt:

$$\mathbf{P}_{\rho_0}(X_0 = a_0, X_1 = a_1) = \rho_0(a_0)P(a_0, a_1)$$

Summation über $a_0 \in S$ ergibt

$$\mathbf{P}_{\rho_0}(X_1 = a_1) = \sum_{a_0 \in S} \rho_0(a_0)P(a_0, a_1)$$

Die rechte Seite kann man lesen als Produkt

des Zeilenvektors ρ_0 mit der Matrix P :

$$\sum_{a_0 \in S} \rho_0(a_0)P(a_0, a_1) := (\rho_0 P)(a_1)$$

$\mathbf{P}_\rho(X_n = c)$ zerlegt nach X_{n-1} ergibt die **Rekursion**

$$\mathbf{P}_\rho(X_n = c) = \sum_{b \in S} \mathbf{P}_\rho(X_{n-1} = b) P(b, c)$$

Mit

$$\rho_n(\cdot) := \mathbf{P}_\rho(X_n \in \cdot) , \quad n = 0, 1, \dots$$

lautet diese Rekursion

$$\rho_n(c) = \sum_{b \in S} \rho_{n-1}(b) P(b, c) , \quad n \geq 1$$

oder in Vektor-Matrix-Schreibweise, mit ρ_n als Zeilenvektor:

$$\rho_n = \rho_{n-1} P , \quad n \geq 1.$$

Vorlesung 9b

Markovketten II

Teil 2:

Gleichgewichtsverteilungen

(Buch S. 108-109)

Sei P eine Übergangsmatrix auf S
und ρ eine (Start-)Verteilung auf S .

Dann gilt (vgl. Teil 1):

$$\mathbf{P}_\rho(X_0 = a, X_1 = b) = \rho(a)P(a, b), \quad \text{also}$$

$$\mathbf{P}_\rho(X_0 = a) = \rho(a),$$

$$\mathbf{P}_\rho(X_1 = b) = \sum_{a \in S} \rho(a)P(a, b).$$

Für welche Startverteilung ρ ist X_1 so verteilt wie X_0 ?

Eine Verteilung π auf S heißt
Gleichgewichtsverteilung zur Übergangsmatrix P , wenn
eine der folgenden (äquivalenten!) Bedingungen erfüllt ist:

$$(G1) \quad \sum_{a \in S} \pi(a) P(a, b) = \pi(b) , \quad b \in S .$$

$$(G2) \quad \mathbf{P}_{\pi}(X_1 = b) = \mathbf{P}_{\pi}(X_0 = b) , \quad b \in S$$

d.h. unter $\mathbf{P}_{\pi}$ haben X_0 und X_1 dieselbe Verteilung.

Reversible Gleichgewichtsverteilungen

Hinreichend für

(G2)

$$\mathbf{P}_\pi(X_1 = b) = \mathbf{P}_\pi(X_0 = b) , \quad b \in S$$

ist die Bedingung

(R)

$$\mathbf{P}_\pi(X_0 = a, X_1 = b) = \mathbf{P}_\pi(X_0 = b, X_1 = a) , \quad a, b \in S$$

Denn dann ist unter $\mathbf{P}_\pi$

das Paar (X_0, X_1) so verteilt wie (X_1, X_0) ,
also insbesondere X_0 so verteilt wie X_1 .

Gleichbedeutend mit

(R)

$$\mathbf{P}_\pi(X_0 = a, X_1 = b) = \mathbf{P}_\pi(X_0 = b, X_1 = a) , \quad a, b \in S$$

ist

$$\pi(a)P(a, b) = \pi(b)P(b, a) , \quad a, b \in S .$$

π heißt dann *reversible Gleichgewichtsverteilung* zu P .

Ein Beispiel einer nicht reversiblen Gleichgewichtsverteilung:

Zyklische Irrfahrt auf $S = \{a, b, c\}$, mit

$$P(a, b) = P(b, c) = P(c, a) := p,$$

$$P(b, a) = P(c, b) = P(a, c) := 1 - p.$$

Die uniforme Verteilung auf S
ist Gleichgewichtsverteilung zu P .
Nur für $p = 1/2$ ist sie reversibel.

**Die Gleichgewichtsverteilung der
einfachen Irrfahrt auf dem gewöhnlichen Würfel**

$$S = \{0, 1\}^3:$$

Von jedem $a \in S$ geht man in einem Schritt
zu einem rein zufällig ausgewählten Nachbarn.

(Zwei Elemente von S heißen *benachbart*,
wenn sie sich in genau einer Komponente unterscheiden.)

Für benachbarte Knoten a und b ist hier $P(a, b) = 1/3$.

Die uniforme Verteilung auf S
ist reversible Gleichgewichtsverteilung.

Eine wichtige Beispielklasse:

Die einfache Irrfahrt

auf einem ungerichteten, zusammenhängenden Graphen

mit endlicher Knotenmenge S

Von jedem $a \in S$ geht man in einem Schritt
zu einem rein zufällig ausgewählten Nachbarn:

$$P(a, b) = \frac{1}{g(a)},$$

mit b Nachbar von a , $g(a) := \# \text{ Nachbarn von } a$.

Ansatz: $\pi(a) := \frac{1}{c}g(a)$

Diese Verteilung π erfüllt die Reversibilitätsbedingung (R),

denn für benachbarte Knoten a, b gilt:

$$\frac{1}{c}g(a)\frac{1}{g(a)} = \frac{1}{c}g(b)\frac{1}{g(b)}$$

Man kann zeigen (hier ohne Beweis): Es gibt (unter den gegebenen Voraussetzungen) nur **eine** Gleichgewichtsverteilung.

**Fazit: Die Gewichte der Knoten
unter der Gleichgewichtsverteilung
sind proportional zur Anzahl der Nachbarn der Knoten.**

Vorlesung 9b

Markovketten II

Teil 3:

Das Ehrenfest-Modell

(Buch S. 109-110)

veröffentlicht 1909 von Paul und Tatjana Ehrenfest, konzipiert
als Spielzeugmodell für Boltzmanns Statistische Mechanik:

d Teilchen sind verteilt auf eine linke und eine rechte Urne:

ℓ Teilchen links, r Teilchen rechts.

In jedem Schritt wird rein zufällig eines aus den d
ausgewählt und in die andere Urne verfrachtet.

Hat diese Dynamik eine Gleichgewichtsverteilung,
und wenn ja, wie sieht sie aus?

Zur Illustration betrachten wir erst nur den Fall $d = 3$.

Die Übergangsw'keiten für die *Anzahl links* sind dann

(für $\ell = 0, 1, 2, 3$):

$$P(\ell, \ell + 1) = \frac{3 - \ell}{3}, \quad P(\ell, \ell - 1) = \frac{\ell}{3}.$$

Ein eleganter Weg zur Antwort führt über ein *Feinmodell*:

Die Teilchen werden nummeriert mit 1, 2, 3.

$$a_i = \begin{cases} 1 & \text{falls das Teilchen mit Nr. } i \text{ in linker Urne,} \\ 0 & \text{..... in rechter Urne.} \end{cases}$$

$$a := (a_1, a_2, a_3) \in \{0, 1\}^3.$$

Dynamik des Feinmodells:

Eine Nummer $i \in \{1, 2, 3\}$ wird rein zufällig ausgewählt
und das a_i wird “geflippt”
(von 0 nach 1 bzw. von 1 nach 0).

Das ergibt die einfache Irrfahrt auf dem gewöhnlichen Würfel
mit der Knotenmenge $\{0, 1\}^3$ (siehe Teil 2).

Diese hat ein reversibles Gleichgewicht:
die uniforme Verteilung auf $\{0, 1\}^3$
(denn jeder Knoten hat gleich viele Nachbarn).

Vom Feinmodell zum Ehrenfest-Modell kommt man durch
“Zählen der Teilchen links”:

$$h(a) := a_1 + a_2 + a_3.$$

Ist $Z = (Z^{(1)}, Z^{(2)}, Z^{(3)})$ uniform verteilt auf $\{0, 1\}^3$,
dann ist $Z^{(1)} + Z^{(2)} + Z^{(3)}$ Binomial($3, \frac{1}{2}$)-verteilt.

Zur Probe: Die Binomial($3, \frac{1}{2}$)-Verteilung
ist ein reversibles Gleichgewicht für das Ehrenfest-Modell
mit $d = 3$ Kugeln:

$$2^{-3} \binom{3}{\ell} \frac{3 - \ell}{3} = 2^{-3} \binom{3}{\ell + 1} \frac{\ell + 1}{3}. \quad \square$$

Was 3 recht ist, ist einem allgemeinen d billig:

Die Binomial($d, \frac{1}{2}$)-Verteilung
ist ein reversibles Gleichgewicht
für das Ehrenfest-Modell mit d Kugeln.

Dazu werden jetzt die Teilchen durchnummeriert mit $1, \dots, d$.

$$a_i = \begin{cases} 1 & \text{falls das Teilchen mit Nr. } i \text{ in linker Urne,} \\ 0 & \text{..... in rechter Urne.} \end{cases}$$

$$a := (a_1, \dots, a_d) \in \{0, 1\}^d.$$

Dynamik des Feinmodells:

Ein $i \in \{1, \dots, d\}$ wird rein zufällig ausgewählt
und das a_i wird “geflippt”
(von 0 nach 1 bzw. von 1 nach 0).

Das ergibt einfach die Irrfahrt auf dem Würfel
mit der Knotenmenge $\{0, 1\}^d$.

Diese hat ein reversibles Gleichgewicht:
die uniforme Verteilung auf $\{0, 1\}^d$

Vom Feinmodell zum Ehrenfest-Modell kommt man durch

“Zählen der Teilchen links”:

$$h(a) := \sum_{i=1}^d a_i$$

Ist $Z = (Z^{(1)}, \dots, Z^{(d)})$ uniform verteilt auf $\{0, 1\}^d$,

dann ist $\sum_{i=1}^d Z^{(i)}$ Binomial($d, \frac{1}{2}$)-verteilt.

Zur Probe: Die Binomial($d, \frac{1}{2}$)-Verteilung
ist ein reversibles Gleichgewicht für das Ehrenfest-Modell:

$$2^{-d} \binom{d}{\ell} \frac{d - \ell}{d} = 2^{-d} \binom{d}{\ell + 1} \frac{\ell + 1}{d}. \quad \square$$

Vorlesung 10

Schätzen mit Verlass:

Konfidenzintervalle

1. Schätzen von Anteilen

(Buch S. 121-122)

Große Population (♀ und ♂)
mit unbekanntem Weibchenanteil p

In einer Stichprobe vom Umfang $n = 53$
gab es 23 Weibchen.

Wie zuverlässig ist $\frac{23}{53}$ als Schätzung für p ?

Goldene Idee der Statistik:

In einem idealisiert gedachten Szenario
interpretiert man den Schätzwert
als Realisierung einer Zufallsvariablen
und rechnet mit der Variabilität dieser Zufallsvariablen.

In unserem Eingangsbeispiel wird
die Stichprobenziehung (idealisiert!)
als p -Münzwurf gedeutet.

Als *Schätzer* für p betrachten wir die *Zufallsvariable*

$$H := \frac{K}{n},$$

mit $K :=$ Anzahl der “Erfolge”.

H ist die relative Häufigkeit der Erfolge (die “Trefferquote”).

$$\sigma_H = ?$$

K ist $\text{Bin}(n, p)$ -verteilt. Also:

$$\sigma_H = \frac{1}{\sqrt{n}} \sqrt{p(1-p)}$$

Der Zentrale Grenzwertsatz liefert uns:

H ist approximativ normalverteilt

mit Erwartungswert $\mu_H = p$ und Standardabweichung σ_H .

Also insbesondere:

$$\mathbf{P}_p(|p - H| \leq 2\sigma_H) \approx 0.95$$

$$\mathbf{P}_p(p \in [H - 2\sigma_H, H + 2\sigma_H]) \approx 0.95$$

Das zufällige Intervall $[H - 2\sigma_H, H + 2\sigma_H]$ überdeckt den Parameter p mit Wahrscheinlichkeit ≈ 0.95 .

In der Praxis ist auch σ_H
(aus der *einen* vorliegenden Stichprobe) zu schätzen.

Als Schätzer für $\sigma_H = \frac{1}{\sqrt{n}}\sqrt{p(1-p)}$ bietet sich an:

$$\widehat{\sigma}_H := \frac{1}{\sqrt{n}}\sqrt{H(1-H)}$$

$$\mathbf{P}_p(p \in [H - 2\sigma_H, H + 2\sigma_H]) \approx 0.95$$

überträgt sich auf

$$\mathbf{P}_p(p \in [H - 2\widehat{\sigma}_H, H + 2\widehat{\sigma}_H]) \approx 0.95.$$

Das zufällige Intervall

$$I := [H - 2\widehat{\sigma}_H, H + 2\widehat{\sigma}_H]$$

ist ein

Konfidenzintervall für p

mit approximativer Überdeckungswahrscheinlichkeit 0.95

oder kurz ein

approximatives 95%-Konfidenzintervall für p .

In unserem Eingangsbeispiel ($n = 53$, $k = 23$)
hatten die beobachteten Realisierungen von H und $\widehat{\sigma}_H$
die Werte

$$h = 23/53 = 0.43,$$
$$\frac{1}{\sqrt{n}}\sqrt{h(1-h)} = \sqrt{\frac{0.43 \cdot 0.57}{53}} = 0.07.$$

Als Realisierung von $I = [H - 2\widehat{\sigma}_H, H + 2\widehat{\sigma}_H]$ ergab sich
 $[0.43 - 2 \cdot 0.07, 0.43 + 2 \cdot 0.07] = [0.29, 0.57]$.

Man beachte:

Nicht der Parameter p ist zufällig, sondern das Intervall I .

Im Jargon der Statistik wird oft
sowohl der Schätzer (die Zufallsvariable) H
als auch der Schätzwert (die Zahl) h
mit dem Symbol $\hat{p}$ bezeichnet.

Auf der nächsten Folie,
die das Obige zusammenfasst,
steht $\hat{p}$ für die Zufallsvariable H
(vgl. Buch S. 122):

Das zufällige Intervall

$$I := \left[\hat{p} - \frac{2}{\sqrt{n}} \sqrt{\hat{p}(1 - \hat{p})}, \hat{p} + \frac{2}{\sqrt{n}} \sqrt{\hat{p}(1 - \hat{p})} \right]$$

ist ein

Konfidenzintervall für p

mit approximativer Überdeckungswahrscheinlichkeit 0.95

oder kurz ein

approximatives 95%-Konfidenzintervall für p .

Dabei sollte $np(1 - p)$ “nicht zu klein” sein.

Eine Faustregel für die Anwendbarkeit ist: $nh \geq 9$ und $n(1 - h) \geq 9$.

Ein frischer Blick auf Konfidenzintervalle

(samt Wiederholung des Obigen):

n sei fest, und K sei die Anzahl der Treffer bei n Versuchen und Erfolgswahrscheinlichkeit p ,

d.h. es gilt

$$\mathbf{P}_p(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Wie oben setzen wir $H := \frac{K}{n}$, $\hat{p} := \frac{k}{n}$.

Im Geist von Übungsaufgabe 27.S definieren wir folgende
Teilmenge von $\{0, 1, \dots, n\} \times [0, 1]$:

$$M_N := \left\{ (\hat{p}, p) : |\hat{p} - p| \leq 2 \cdot \sqrt{\frac{1}{n} \cdot \hat{p} (1 - \hat{p})} \right\}$$

Für nicht zu kleines npq gilt
wegen der approximativen Normalität von K
(und weil $\sqrt{\frac{1}{n}H(1-H)}$ ein passender Schätzer für σ_H ist):

$$\mathbf{P}_p((H, p) \in M_N) \approx 0.95.$$

Für die Intervalle $I_N(\hat{p}) := \{p : (\hat{p}, p) \in M_N\}$ haben wir für jedes $p \in [0, 1]$ die Ereignisgleichheit

$$\{p \in I_N(H)\} = \{(H, p) \in M_N\}.$$

Also gilt für nicht zu kleines npq :

$$\mathbf{P}_p(p \in I_N(H)) \approx 0.95.$$

So lange die Normalapproximation passt, gilt

$$M_N \approx \left\{ \left(\frac{k}{n}, p \right) : k \text{ liegt zwischen dem 0.025-Quantil und dem 0.975-Quantil der Bin}(n, p)\text{-Verteilung, } 0 \leq p \leq 1. \right\}$$

Für $p = 0$ bzw. $p = 1$ wird die Approximation unbrauchbar:

$\left(\frac{k}{n}, 0 \right)$ gehört zu M_N nur für $k = 0$,
und $\left(\frac{k}{n}, 1 \right)$ gehört zu M_N nur für $k = n$.

Auch für kleine npq wird die Approximation schlecht.

Vorschlag von Clopper und Pearson (1934):

$$M_C := \left\{ \left(\frac{k}{n}, p \right) : k \text{ liegt zwischen dem 0.025-Quantil und dem 0.975-Quantil der Bin}(n, p)\text{-Verteilung, } 0 \leq p \leq 1. \right\}$$

Damit bekommen wir für *alle* $p \in [0, 1]$:

$$\mathbf{P}_p((H, p) \in M_C) \geq 0.95.$$

Nach demselben Muster wie oben setzen wir

$$I_C(\hat{p}) := \{p : (\hat{p}, p) \in M_C\}$$

und bekommen jetzt für *alle* $p \in [0, 1]$:

$$\mathbf{P}_p(p \in I_C(H)) \geq 0.95.$$

Beispiel: Von 1500 Senioren bekam die (per Losentscheid ermittelte) eine Hälfte (die *Behandlungsgruppe*) einen Impfstoff verabreicht und die andere Hälfte (die *Kontrollgruppe*) ein Placebo. Fünf der Senioren haben sich infiziert, alle gehörten zur Kontrollgruppe.

(Quelle:

<https://www.nejm.org/doi/pdf/10.1056/NEJMoa2034577?articleTo>
Table 3, Age group ≥ 75 yr)

Als Realisierung des 95% Clopper-Pearson Konfidenzintervalls für die Erfolgswahrscheinlichkeit des Impfstoffes ergibt sich $[0.478, 1]$ (denn $0.478^5 = 0.025$).

Dieses Intervall schließt den Parameterwert $p = 1/2$ ein: auch wenn alles rein zufällig zugegangen wäre, hätte ein so extremes Ergebnis eine Wahrscheinlichkeit von mehr als 5%.

2. Schätzung des Erwartungswertes einer Verteilung auf $\mathbb{R}$ (Lageschätzung)

$$m := \frac{1}{n}(x_1 + \cdots + x_n)$$

wird gedacht als eine Realisierung der Zufallsvariablen

$$\bar{X} := \frac{1}{n}(X_1 + \cdots + X_n)$$

mit $X_1, \dots, X_n$ unabhängig, identisch verteilt

mit Erwartungswert μ und Standardabweichung σ .

Anders als bei der Anteilschätzung ist hier σ i.a. keine Funktion von μ .

$\bar{X}$ ist ein Schätzer für μ .

$\bar{X}$ hat Erwartungswert μ und Standardabweichung $\sigma/\sqrt{n}$.

Ein Schätzer für σ^2 ist

$$S^2 := \frac{1}{n-1} ((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)$$

Im Jargon der Statistik schreibt man oft s^2 statt S^2 und verwendet s^2 dann auch zur Bezeichnung von Schätzwerten.

Anders als im Buch werden wir hier auf den Folien an der oben definierten Bezeichnung S^2 festhalten und s^2 für die Bezeichnung einer Realisierung von S^2 reservieren.

2a. Großer Stichprobenumfang n

Der Zentrale Grenzwertsatz liefert uns:

$\bar{X} - \mu$ ist approximativ $N(0, \frac{\sigma^2}{n})$ -verteilt.

Bei bekanntem σ ist also für große n

$$\left[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}} \right]$$

ein approximatives 95%-Konfidenzintervall für μ .

In der Praxis hat man auch σ aus den Daten zu schätzen.

Für große n ist auch

$$J := \left[\bar{X} - 2 \frac{\textcolor{red}{S}}{\sqrt{n}}, \bar{X} + 2 \frac{\textcolor{red}{S}}{\sqrt{n}} \right]$$

ein approximatives 95%-Konfidenzintervall für μ

Dieses Intervall enthält keine unbekannten Parameter.

Der beobachtete Wert (die Realisierung) von S ist

$$s := \sqrt{\frac{1}{n-1} \left((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2 \right)}$$

s ist ein Schätzwert für σ .

Die Realisierung von J ist somit $\left[\bar{x} - 2 \frac{s}{\sqrt{n}}, \bar{x} + 2 \frac{s}{\sqrt{n}} \right]$.

2b. Kleiner Stichprobenumfang n

Für kleine n (etwa: $n \leq 10$) und (exakt bzw. annähernd)
normalverteilte X_i

macht man sich zunutze, dass die Verteilung von

$T := \frac{\sqrt{n}(\bar{X} - \mu)}{S}$ (exakt bzw. annähernd) so verteilt ist wie

$$T_{n-1} := \frac{N_0}{\sqrt{\frac{1}{n-1} (N_1^2 + \dots + N_{n-1}^2)}}$$

mit unabhängigen und $N(0, 1)$ verteilten $N_0, \dots, N_{n-1}$.

Satz (W. Gosset (alias “Student”, 1908), R. Fisher (1924))

Sind $X_1, \dots, X_n$ unabhängig und $N(\mu, \sigma^2)$ -verteilt,

dann ist $T := \frac{\sqrt{n}(\bar{X} - \mu)}{S}$ so verteilt wie

$$T_{n-1} := \frac{N_0}{\sqrt{\frac{1}{n-1} (N_1^2 + \dots + N_{n-1}^2)}}$$

mit unabhängigen und $N(0, 1)$ verteilten $N_0, \dots, N_{n-1}$.

Das folgt aus der Rotationssymmetrie
der Standard-Normalverteilung im $\mathbb{R}^n$

mit einem ähnlichen Argument wie dem in F7a1.12, vgl. Buch S. 138.

Satz (W. Gosset (alias “Student”, 1908), R. Fisher (1924))

Sind $X_1, \dots, X_n$ unabhängig und $N(\mu, \sigma^2)$ -verteilt,

dann ist $T := \frac{\sqrt{n}(\bar{X} - \mu)}{S}$ so verteilt wie

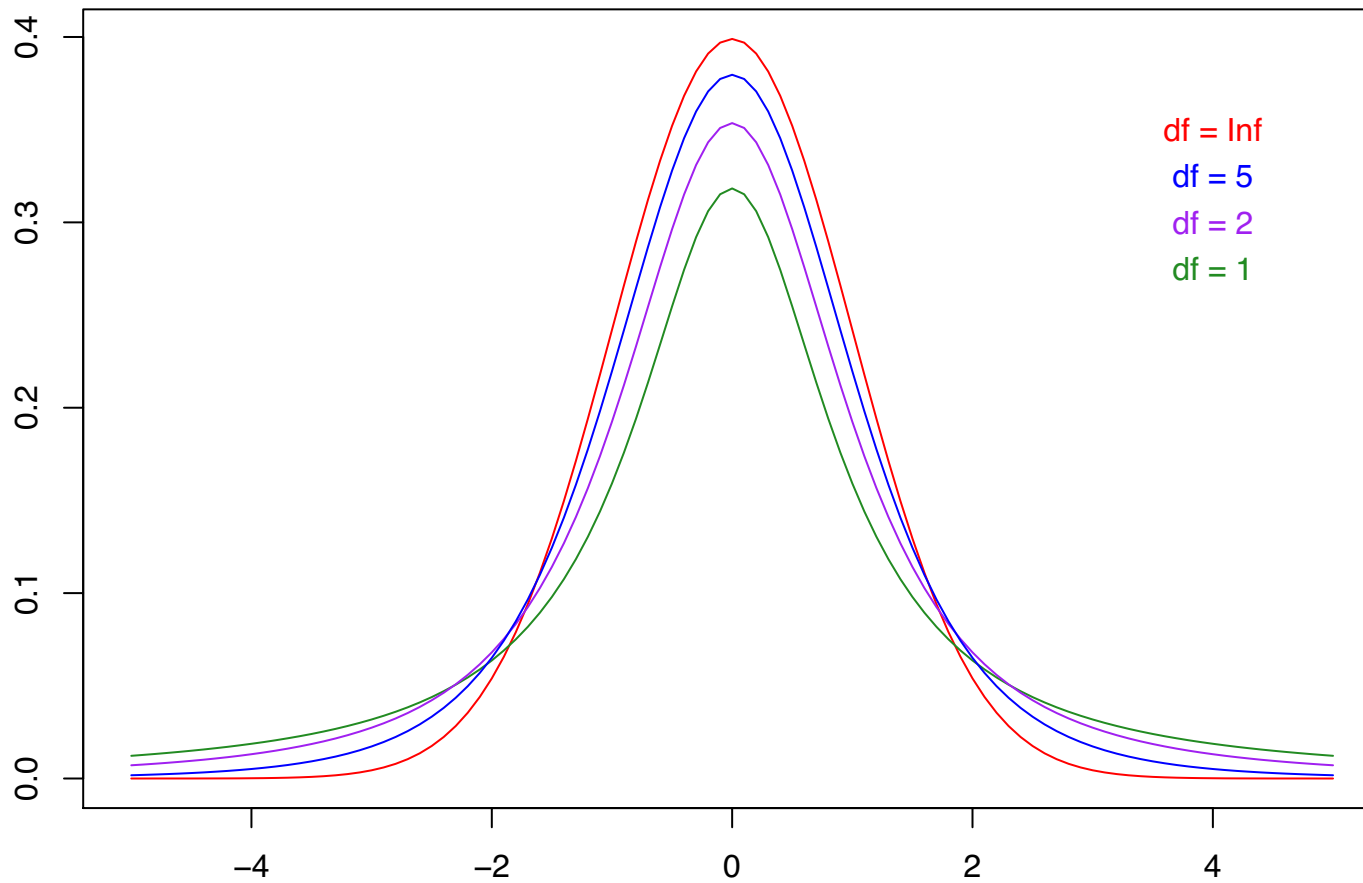
$$T_{n-1} := \frac{N_0}{\sqrt{\frac{1}{n-1} (N_1^2 + \dots + N_{n-1}^2)}}$$

mit unabhängigen und $N(0, 1)$ verteilten $N_0, \dots, N_{n-1}$.

T heißt (die aus $X_1, \dots, X_n$ gewonnene) **t -Statistik**.

Die **Verteilung von T_{n-1}** heißt **t -Verteilung** (oder **Student-Verteilung**) mit $n - 1$ Freiheitsgraden.

Student's t: Dichtefunktionen



“df” steht hier für “degrees of freedom” , d.h. “Freiheitsgrade”

$$T_{n-1} := \frac{N_0}{\sqrt{\frac{1}{n-1} (N_1^2 + \dots + N_{n-1}^2)}}$$

Für $n \rightarrow \infty$ ist T_{n-1} asymptotisch $N(0, 1)$ -verteilt
(Gesetz der großen Zahlen für den Nenner von T_{n-1}).

Je kleiner n , um so mehr schwankt der Nenner, und
um so *breitschultriger* ist die Verteilung von T_{n-1}

Z.B. für $n = 6$: $\mathbf{P}(|T_5| \leq 2.57) = 0.95$.

Satz (W. Gosset (alias “Student”, 1908), R. Fisher (1924))

Sind $X_1, \dots, X_n$ unabhängig und $N(\mu, \sigma^2)$ -verteilt,

dann ist $T := \frac{\sqrt{n}(\bar{X} - \mu)}{S}$ so verteilt wie

$$T_{n-1} := \frac{N_0}{\sqrt{\frac{1}{n-1} (N_1^2 + \dots + N_{n-1}^2)}}$$

mit unabhängigen und $N(0, 1)$ verteilten $N_0, \dots, N_{n-1}$.

Folgerung: Sind $X_1, \dots, X_n$ unabhängig und $N(\mu, \sigma^2)$ -verteilt, dann ist für jedes $c > 0$:

$$\mathbf{P}(|T_{n-1}| \leq c) = \mathbf{P}\left(\left|\frac{\sqrt{n}(\bar{X} - \mu)}{S}\right| \leq c\right) = \mathbf{P}\left(\mu \in \left[\bar{X} - \frac{cS}{\sqrt{n}}, \bar{X} + \frac{cS}{\sqrt{n}}\right]\right)$$

Für ein 95%-Konfidenzintervall bestimme c so, dass sich **für die linke Seite** 0.95 ergibt.

Z.B. für $n = 6$: $\mathbf{P}(|T_5| \leq 2.57) = 0.95$.

Der passende R-Befehl ist `qt(0.975, 5)`, mit der Ausgabe 2.57

Denn: $\mathbf{P}(T_5 \leq 2.57) = 0.975$.

Man sagt: Das 0.975-Quantil der $t(5)$ -Verteilung ist 2.57.

3. Ein Konfidenzintervall für den Median

(Buch S. 128)

Seien $X_1, X_2, \dots, X_n$ unabhängig, mit Verteilung ρ .
Es gibt Situationen, in denen die Schätzung der “Lage” von ρ
über den Stichprobenmittelwert problematisch ist –
etwa wenn ρ so viel Masse “weit draußen” hat,
dass $\sigma^2 = \text{Var}[X_1]$ sehr groß ist.

In diesem Fall ist es günstig,
einen “robusteren” Lageschätzer zu verwenden:

Eine Zahl ν heißt *Median* der Verteilung ρ auf $\mathbb{R}$,
wenn $\rho((-\infty, \nu]) \geq 1/2$ **und** $\rho([\nu, \infty)) \geq 1/2$ gilt.

Wenn es nur eine Zahl ν gibt mit $\rho((-\infty, \nu]) = 1/2$
(also z. B. wenn $\rho([\ell, r]) = 1$ gilt
und ρ eine strikt positive Dichtefunktion besitzt),
dann ist ν **der** Median von ρ .

Eine Zahl ν heißt *Median* der Verteilung ρ auf $\mathbb{R}$,
wenn $\rho((-\infty, \nu]) \geq 1/2$ **und** $\rho([\nu, \infty)) \geq 1/2$ gilt.

Wenn es mehrere Zahlen ν gibt mit $\rho((-\infty, \nu]) = 1/2$,
dann ist jede dieser Zahlen **ein** Median von ρ .

Bsp: Für die uniforme Verteilung auf $[0, 1] \cup [2, 3]$
ist jedes $\nu \in [1, 2]$ ein Median.

Wie schätzt man den Median?

Die *Ordnungsstatistiken* $X_{(1)} \leq \dots \leq X_{(n)}$
sind die aufsteigend geordneten $X_1, \dots, X_n$.

Sei j eine natürliche Zahl mit $0 \leq j < n/2$.

Ein Kandidat für ein **Konfidenzintervall für den Median** ist

$$[X_{(1+j)}, X_{(n-j)}].$$

Z.B. für $j = 0$:

$$\mathbf{P}_\rho(\nu \notin [X_{(1)}, X_{(n)}]) = \mathbf{P}_\rho(X_{(1)} > \nu) + \mathbf{P}_\rho(X_{(n)} < \nu) .$$

$$\mathbf{P}_\rho(X_{(1)} > \nu) \leq 2^{-n}$$

$$\mathbf{P}_\rho(X_{(n)} < \nu) \leq 2^{-n} .$$

Also:

$$\mathbf{P}_\rho(\nu \in [X_{(1)}, X_{(n)}]) \geq 1 - \frac{1}{2^{n-1}} .$$

$$\mathbf{P}_\rho(\nu \in [X_{(1)}, X_{(n)}]) \geq 1 - \frac{1}{2^{n-1}} .$$

Bsp: $n = 6$

$$\mathbf{P}_\rho(\nu \in [X_{(1)}, X_{(6)}]) \geq 1 - \frac{1}{32} = 0.97 :$$

Das Konfidenzintervall $[X_{(1)}, X_{(6)}]$ für den Median hält die Überdeckungswahrscheinlichkeit 0.97 ein.

Vorlesung 11

Kann das Zufall sein?

Beispiele von statistischen Tests

1. Testen von Anteilen

1a. Fishers exakter Test

“Passen die Verhältnisse in den Rahmen?”

(vgl. Buch S. 130/131)



Sir Ronald Fisher

1890-1962

Eine Übungsaufgabe aus Woche 3 des WS 18/19 *

Denken wir uns einen FBR, der aus 9 Vertretern der Informatik und 8 der Mathematik besteht. In einem 5-köpfigen Komitee des FBR findet sich nur ein Vertreter der Informatik.

Wie wahrscheinlich ist eine so extreme Zusammensetzung bei einer rein zufälligen Auswahl?

Genauer: Was ist die Wahrscheinlichkeit, dass *die Anzahl X der Informatiker in einem rein zufällig gebildeten 5-köpfigen Komitee des FBR* mindestens so weit entfernt von $\mu := \mathbb{E}[X]$ ist wie die beobachtete Anzahl 1?

*vgl. dazu die Aufgabe 14 des laufenden Semesters

Die Bilderbuchversion dieser Aufgabe lautet so:

Aus einer Urne mit 9 roten und 8 blauen Kugeln
wurden 5 Kugeln entnommen .

1 davon war rot, und 4 waren blau.

Passt das zur Hypothese
einer rein zufälligen Entnahme der 5 Kugeln?

Oder: *Was ist die Wahrscheinlichkeit eines
mindestens so extremen Ergebnisses wie das Beobachtete?*

$$E[X] = 5 \cdot \frac{9}{9+8} = 2.65$$

Der beobachtete Ausgang **1** hat von 2.65 die Distanz 1.65.

Die möglichen Ausgänge waren 0, 1, 2, 3, 4, 5.

Aus diesen sind die Ausgänge 0, 1 und 5
mindestens so weit von 2.65 entfernt wie **1**.

$$P(X \in \{0, 1, 5\}) = ?$$

$$P(X = k) = \frac{\binom{r}{k} \binom{g-r}{n-k}}{\binom{g}{n}}, \quad k = 0, 1, \dots, n.$$

$$P(X \in \{0, 1, 5\}) = 0.13$$

Unter der Hypothese des rein zufälligen Ziehens kommt ein so extremer Ausgang nur mit W'keit 0.13 vor.

Würde man das Experiment oft wiederholen, so bekäme man in ca 13% der Fälle ein mindestens so extremes Ergebnis.

Einem Vorschlag von R.A. Fisher folgend spricht man hier vom

p-Wert 0.13

(unter der Hypothese des reinen Zufalls).

“Ungefähr jedes zehnte Mal” ist zwar anders als “alltäglich”,
aber auch nicht “richtig selten”.

Salopp ausgedrückt:

Ein p-Wert 0.13 kann uns nicht so richtig stutzig machen.

(und erlaubt uns wohl nicht,
die Hypothese des reinen Zufalls abzulehnen –
eingebürgert hat sich hierfür die Schranke 0.05).

Noch ein Beispiel:

Gleiches Modell, andere Zahlen, frische Anwendung:

Aus einer Urne mit 80 roten und 87 blauen Kugeln
wurden 113 Kugeln entnommen.

40 davon waren rot, und 73 waren blau.

Passt das zur Hypothese, dass die Kugeln
rein zufällig gezogen wurden?

Stimmen die Verhältnisse einigermaßen,
oder fallen sie aus dem Rahmen?

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Unter den 113 gezogenen Kugeln erwartet man ähnliche Verhältnisse wie in der gesamten Urne:

80: 167 für rot, 87 : 167 für blau.

Tatsächlich ergab sich in der Stichprobe für rot ein **sehr** unterdurchschnittliches Ergebnis !

Wie lässt sich das quantifizieren?

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Unter der Hypothese des rein zufälligen Ziehens
ist die Anzahl X der gezogenen roten Kugeln
hypergeometrisch verteilt
mit Parametern $n = 113$, $g = 167$, $r = 80$.

Dafür ergibt sich:

$$E[X] = n \cdot \frac{r}{g} = 54.1 .$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Unter der Hypothese des rein zufälligen Ziehens
ist die Anzahl X der gezogenen roten Kugeln
hypergeometrisch verteilt
mit Parametern $n = 113$, $g = 167$, $r = 80$.

Dafür ergibt sich:

$$E[X] = n \cdot \frac{r}{g} = 54.1 .$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Die Wahrscheinlichkeit, ein Ergebnis zu erhalten,
das mindestens so weit von 54 weg ist
wie der beobachtete Wert 40, ist

$$\begin{aligned}
 & P(|X - 54| \geq |40 - 54|) \\
 &= P(X \leq 40) + P(X \geq 68) \\
 &= 5.57 \cdot 10^{-6} .
 \end{aligned}$$

$$\mathbf{P}(|X - 54| \geq |40 - 54|) = 5.6 \cdot 10^{-6}$$

Was bedeutet das?

Fazit: Angenommen die Hypothese trifft zu.
Dann tritt ein Ergebnis, das so extrem ist wie das beobachtete, gerade 6 mal in einer Million auf.
Damit wird die Hypothese mehr als fragwürdig.

Man nennt die berechnete Wahrscheinlichkeit
den zu den Daten gehörigen p-Wert
oder auch das *beobachtete Signifikanzniveau*,
zu dem die Hypothese abgelehnt wird.

Wie passt unser Urnen-Beispiel in die Welt?

Es geht (wieder) um die Fragestellung

“Passen die Proportionen

– oder sollte man

an der Hypothese der reinen Zufälligkeit zweifeln?”

Wird unsere Einstellung gegenüber Fakten
durch deren Verpackung beeinflusst?

Dazu eine (wahre oder zumindest gut erfundene) Geschichte:

Zu einer bestimmten Krankheit gibt es zwei
Therapiemethoden, eine sanfte (T1) und eine harte (T2).

Eine klinische Studie hatte als Fakten ergeben:

T1 war in 70% der Fälle erfolgreich,
T2 in 80% (allerdings mit mehr Nebenwirkungen).

Am Ende einer Sommerschule wurden diese Fakten
in zwei verschiedene Formen **A** und **B** verpackt,
zusammen mit einer Frage:

A. Die sanfte Therapiemethode T1 brachte
in nicht weniger als 30% der Fälle keinen Heilungserfolg,
wohingegen die harte Therapiemethode T2
in immerhin 80 % der Fälle erfolgreich war.

Welche Therapiemethode würden Sie bevorzugen?

B. Die harte Therapiemethode T2 brachte
in nicht weniger als 20% der Fälle keinen Heilungserfolg,
wohingegen die sanfte Therapiemethode T1
in immerhin 70 % der Fälle erfolgreich war.

Welche Therapiemethode würden Sie bevorzugen?

Von insgesamt 167 Ärzt*inn*en
wurden **rein zufällig 80** ausgewählt,
denen die Botschaft **in der Form A** vermittelt wurde,
die restlichen **87** bekamen die Botschaft **in der Form B**.

Jede(r) hatte sich daraufhin für die Bevorzugung
einer der beiden Therapiemethoden zu entscheiden.

Das Ergebnis war:

	“bin eher für T1”	“bin eher für T2”	Summe
A	40	40	80
B	73	14	87
Summe	113	54	167

Angenommen

die 113 Befürworter der (sanften) Behandlungsmethode T1
und die 54 Befürworter der (harten) Behandlungsmethode T2
sind zu ihrer Einstellung aufgrund der Fakten gekommen
und nicht aufgrund von deren Verpackungen.

Das heißt dann, dass die Zuteilung der

80 Formulare mit der Botschaft in der Form A und der

87 Formulare mit der Botschaft in der Form B

auf die 113 Befürworter von T1

rein zufällig (durch Ziehen ohne Zurücklegen) erfolgt ist.

Für das Testen der Hypothese
“Die Verpackung der Botschaft
hat keinen Einfluss auf die Entscheidung”
eignet sich das vorher besprochene Urnenmodell.

Unter dieser Hypothese
kommt die Aufteilung der 80 + 87 Formulare
auf die 113 Befürworter von T1
und die 54 Befürworter von T2
rein zufällig zustande.

So gesehen kann das Ergebnis “wohl kaum Zufall sein”:

unter unserer Hypothese tritt ein Ausgang,
der so extrem ist wie der beobachtete,
gerade mal 6 mal in einer Million auf.

1b. Testen von Anteilen mit der Normalapproximation

Wenn (wie im vorigen Beispiel)
der Stichprobenumfang n einigermaßen groß ist,
(und beim Ziehen ohne Zurücklegen auch $g - n$ groß ist)

bietet die *Normalapproximation*
eine weitere Möglichkeit des Testens der Hypothese
“Die Verhältnisse passen zueinander”:

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Anteilsschätzung über die Normalapproximation:

X := Anzahl roter Kugeln bei $n = 113$ Zügen
 ohne Zurücklegen aus einer Urne mit $g = 167$ Kugeln,
 von denen 80 rot sind.

$$\mathbb{E}[X] = np$$

mit

$$p = \frac{80}{167}$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

$H = \frac{X}{n}$ ist approximativ normalverteilt

(trotz der schwachen Abhängigkeiten beim Ziehen ohne Zurücklegen), mit

$$\mathbb{E}[H] = \mu_H = p = \frac{80}{167},$$

$$\sigma_H^2 = \frac{p(1-p)}{n} \frac{g-n}{g-1},$$

$$\sigma_H = \sqrt{0.022 \cdot 0.326} = 0.0268$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

$Z := \frac{H - \mu_H}{\sigma_H}$ ist approximativ $N(0, 1)$ -verteilt

Der beobachtete Wert von Z war $z = -4.67$.

$$\mathbf{P}(|Z| > 4.67) = 3 \cdot 10^{-6}$$

ist hier der p-Wert, zu dem die Hypothese abgelehnt wird.

2. Tests der Hypothese $\mu = \mu_0$

“Kann *diese* Verschiebung des Mittelwertes Zufall sein?”

n reelle Messwerte $x_1, \dots, x_n$ haben den Mittelwert m .

Unterscheidet sich der beobachtete Mittelwert m signifikant von einem hypothetischen “Populationsmittelwert” μ_0 ?

Eine Auskunft gibt ein Vergleich

des Unterschiedes $|m - \mu_0|$

mit der geschätzten Standardabweichung

des Stichprobenmittelwertes

$$s/\sqrt{n}.$$

Wir können fragen:

Um weches Vielfache von $s/\sqrt{n}$

unterscheidet sich m von μ_0 ?

Dies erhält seinen theoretischen Unterbau

durch die goldene Idee der Statistik

(man fasse die x_i auf als Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen X_i mit Erwartungswert μ und Varianz σ^2)

und den Zentralen Grenzwertsatz:

Für große n ist der Stichprobenmittelwert $\bar{X}$ approximativ $N\left(\mu, \frac{\sigma^2}{n}\right)$ -verteilt.

2a. Normalapproximation.

Für große n ist

$\bar{X} - \mu$ approximativ $N(0, \frac{\sigma^2}{n})$ -verteilt.

Bei bekanntem σ sei $z := \frac{m - \mu_0}{\sigma/\sqrt{n}}$.

Unter der Hypothese $\mu = \mu_0$ ist

$$\mathbf{P}(|\bar{X} - \mu_0| \geq |m - \mu_0|) \approx \mathbf{P}(|Z| \geq |z|),$$

mit $N(0, 1)$ -verteiletem Z .

In der Praxis ist σ meist unbekannt.

Aber:

Für große n ist S mit großer W'keit nahe bei σ .

Also ist für große n

$T := \frac{\bar{X} - \mu}{S/\sqrt{n}}$ approximativ $N(0, 1)$ -verteilt.

Sei $t := \frac{m - \mu_0}{s/\sqrt{n}}$.

Unter der Hypothese $\mu = \mu_0$ ist für große n

$$\mathbf{P}(|T| \geq |t|) \approx \mathbf{P}(|Z| \geq |t|),$$

mit $N(0, 1)$ -verteiletem Z .

Beispiel:

Ist für großes n der “Wert der t -Statistik”

$$t = \frac{m - \mu_0}{s/\sqrt{n}}$$

gleich 2, dann ergibt sich

$$\mathbf{P}(|Z| \geq 2) \approx 0.05.$$

Man spricht vom **p-Wert** für die Ablehnung der Hypothese $\mu = \mu_0$ zugunsten der Alternative $\mu \neq \mu_0$.

Oft gibt man sich ein *Signifikanzniveau* α vor.

Wenn der p-Wert kleiner als α ist, sagt man: Die Hypothese $\mu = \mu_0$ kann zugunsten der Alternative $\mu \neq \mu_0$ zum Niveau α abgelehnt werden.

Populär ist die Wahl $\alpha = 0.05$.

2b. Der t-Test

Wieder fragen wir:

“Kann *diese* Verschiebung des Mittelwertes Zufall sein?”

Was lässt sich aus der Teststatistik T
bei kleinem Stichprobenumfang n ablesen?

Unter der **zusätzlichen Modellannahme**

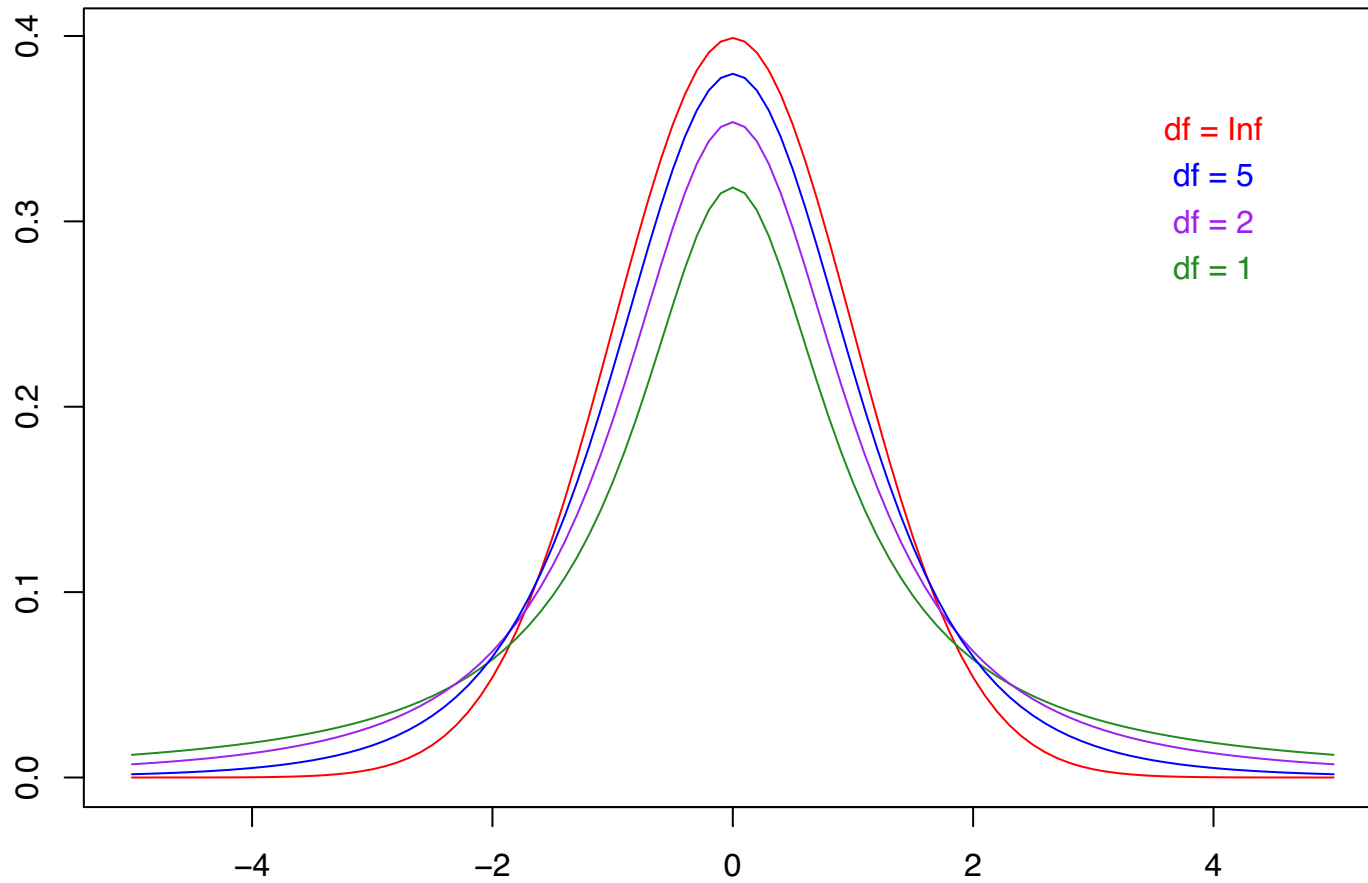
$X_1, \dots, X_n$ sind (unabhängig und) $N(\mu, \sigma^2)$ -verteilt

ist $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ so verteilt wie

$T_{n-1} :=$ eine student-verteilte Zufallsvariable
mit $n - 1$ Freiheitsgraden (degrees of freedom, df),
vgl. F10 Folie 25.

Diese Verteilung kennt R gut (Befehl für Verteilungsfunktion: `pt(q, df)`);
sie lässt sich aus der Rotationssymmetrie der n -dimensionalen
Standard-Normalverteilung auch gut verstehen (Buch Seite 132).

Student's t: Dichtefunktionen



Dichten von T_{df}

Beispiel:

Ist für $n = 16$ der “Wert der t -Statistik”

$$t = \frac{\bar{m} - \mu_0}{s/\sqrt{n}}$$

gleich 2.5, dann ergibt sich

$$\begin{aligned} \mathbf{P}(|T_{15}| \geq 2.5) &= 2\mathbf{P}(T_{15} \geq 2.5) \\ &= 2(1 - \text{pt}(2.5, 15)) = 0.025. \end{aligned}$$

Man spricht vom **p-Wert** für die Ablehnung der Hypothese $\mu = \mu_0$ zugunsten der Alternative $\mu \neq \mu_0$.

Oft gibt man sich ein *Signifikanzniveau* α vor.

Wenn der p-Wert kleiner als α ist, sagt man: Die Hypothese $\mu = \mu_0$ kann zugunsten der Alternative $\mu \neq \mu_0$ zum Niveau α abgelehnt werden.

Populär ist die Wahl $\alpha = 0.05$.

3 Tests der Hypothese $\mu_1 = \mu_2$ bei ungepaarten Strichproben.

“Unterscheiden sich zwei Mittelwerte signifikant?”

Man stelle sich vor: Die Mittelwerte m_x und m_y
von zwei Stichproben des Umfangs n_x und n_y
unterscheiden sich um 0.5 Einheiten: $|m_y - m_x| = 0.5$.

Ist dieser Unterschied signifikant?

Das kommt drauf an ...

Nach bewährtem Rezept vergleichen wir $|m_x - m_y|$
mit seiner geschätzten Standardabweichung f .

Weil wir hier an unabhängige Stichproben denken,
addieren sich die Varianzen:

$$f := \sqrt{f_x^2 + f_y^2}$$

Eine Maßzahl für den “relativen Unterschied” ist also

$$\frac{m_y - m_x}{f}.$$

Anders gefragt: Wie groß ist der beobachtete Wert der
Differenz der Stichprobenmittelwerte,
gemessen in Einheiten der geschätzten Standardabweichung
der Differenz der Stichprobenmittelwerte?

3a. Normalapproximation (bei großen Stichprobenumfängen)

Mit Hilfe der asymptotischen Normalität
wird die Antwort leicht: Unter der Hypothese $\mu_X = \mu_Y$
ist $\frac{m_y - m_x}{f}$ zu lesen als
Realisierung einer annähernd $N(0, 1)$ - verteilten
Zufallsvariablen.

Ist dieser Wert (etwa) 1.96, dann bekommt man 0.05
als p-Wert für die Ablehnung der Hypothese $\mu_X = \mu_Y$.

Dabei interpretiert man wiederum die x_i und die y_j
als Realisierungen von
unabhängigen Zufallsvariablen X_i, Y_j ,
(mit (X_i) identisch verteilt, (Y_j) identisch verteilt)
und fragt nach der Verteilung von

$$T := \frac{\bar{Y} - \bar{X}}{F} = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{S_X^2}{n_x} + \frac{S_Y^2}{n_y}}}$$

$$T := \frac{\bar{Y} - \bar{X}}{F} = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{S_X^2}{n_x} + \frac{S_Y^2}{n_y}}}$$

Für große n_x, n_y ist T annähernd $N(0, 1)$ -verteilt
(wegen des Zentralen Grenzwertsatzes und des Gesetzes
der großen Zahlen).

3b. Der t-Test für zwei ungepaarte Stichproben.

Wir betrachten dieselbe Situation wie in 3a.

Was kann man tun für kleine Stichprobenumfänge n_x, n_y ?
Hier kommt man zumindest unter der zusätzlichen Annahme
weiter, dass die X_i und Y_j normalverteilt sind.

Man kann zeigen, dass T dann annähernd t -verteilt ist mit einer i.a. nicht ganzzahligen Anzahl von Freiheitsgraden.

Die Formel dafür (die man sich nicht merken muss) findet man auf http://en.wikipedia.org/wiki/Student's_t-test im Abschnitt “Equal or unequal sample sizes, unequal variance”

Wichtig ist der praktische Umgang damit in R, zu dem man dort auf die Frage `?t.test` Auskunft bekommt.

4. Der Wilcoxon-Test.

Wie untypisch ist die Lage der Ränge?

Wie eben zuvor geht es um einen Test der Hypothese,
dass zwei Stichproben
aus derselben Verteilung (auf $\mathbb{R}$) kommen,
gegen die Alternative, dass sich die beiden Verteilungen
durch eine Verschiebung unterscheiden.

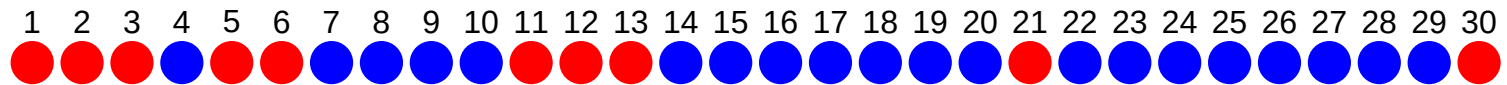
Die folgende Idee kommt ganz
ohne spezielle Verteilungsannahme aus:
Man ordnet die $n_x + n_y$ Werte der Größe nach
und ersetzt sie durch ihre Ränge $R(x_i), R(y_j)$.

(der kleinste Wert bekommt den Rang 1, der zweitkleinste den Rang 2,...).

Dann beobachtet man die *Rangsumme* $w := \sum_{i=1}^{n_x} R(x_i)$

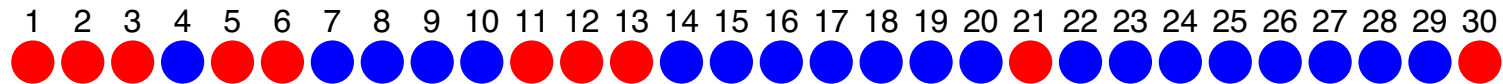
und fragt: Wie wahrscheinlich ist eine
mindestens so “randständige” Rangsumme
bei rein zufälliger Auswahl von n_x Elementen
aus der Menge $\{1, \dots, n_x + n_y\}$?

Die Raenge der x_i und der y_j



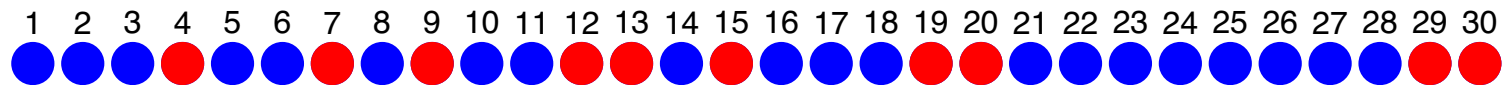
Rangsumme der $x_i = 104$

Die Ränge der x_i und der y_j



Rangsumme der $x_i = 104$

Eine zufaellige Permutation



Rangsumme der x_i in der Permutation = 158

Die “beobachtete” Rangsumme war 104.

Die minimale mögliche Rangsumme
einer “roten Teilstichprobe” ist $1 + \dots + 10 = 55$.

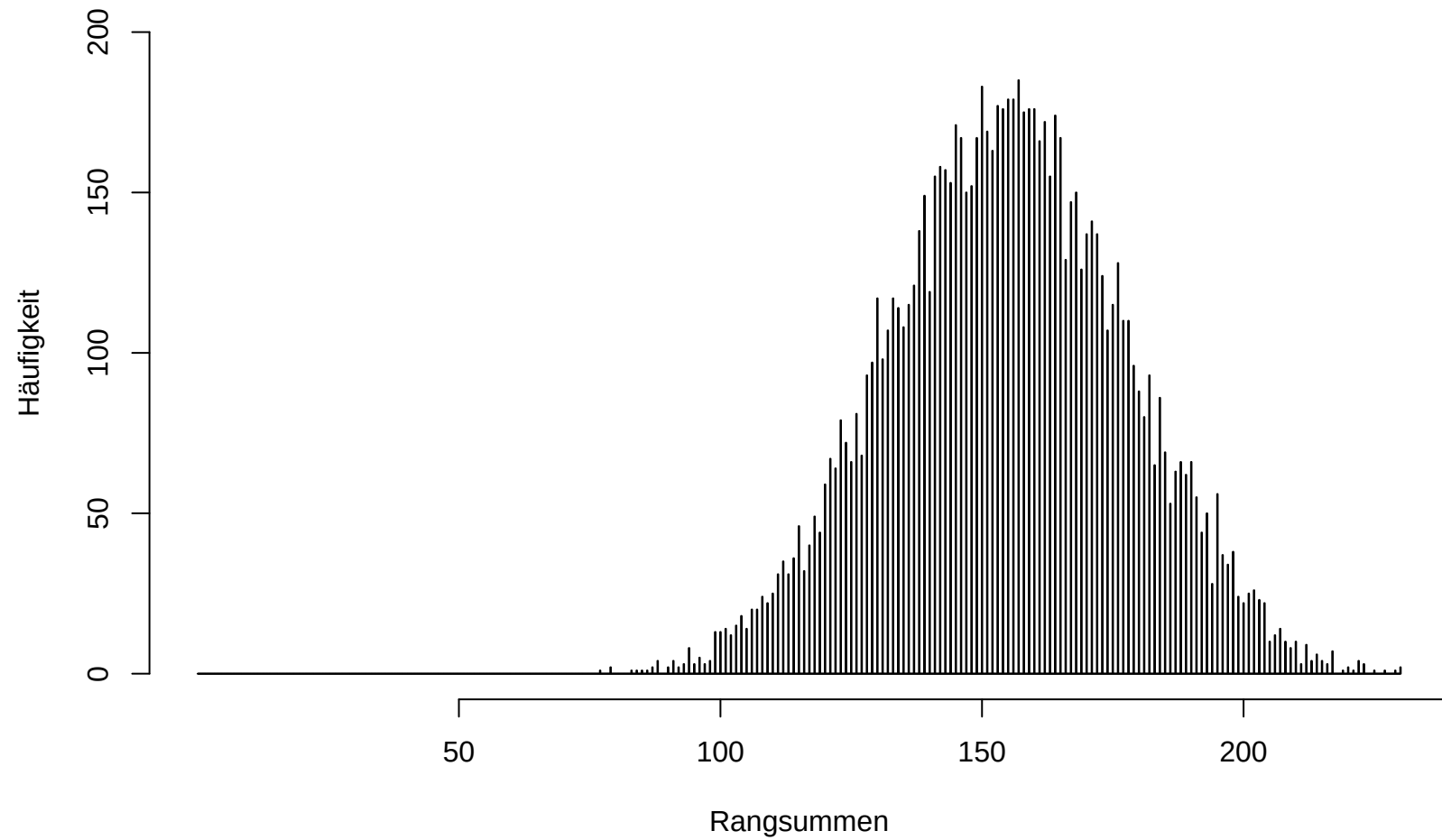
Ihre maximale mögliche Rangsumme ist

$$21 + \dots + 30 = 255.$$

Wir ziehen 10000 mal eine Stichprobe der Größe 10 (aus 30)
und notieren deren Rangsumme.

Der *stochastische p-Wert* ist die relative Häufigkeit
der Ergebnisse, für die sich eine Rangsumme
 ≤ 104 oder $\geq 255 - (104 - 55)$ ergibt.

Rangsummen aus 10000 Permutationen



Rangsummen aus 10000 Permutationen

