Stochastik für die Informatik

Wintersemester 2012/13

Anton Wakolbinger

StofI-Webseite:

http://ismi.math.uni-

frankfurt.de/wakolbinger/teaching/StofI1213/

StofI-Webseite:

http://ismi.math.uni-

frankfurt.de/wakolbinger/teaching/StofI1213/

Oder:

google $\longrightarrow$ “Stochastik für die Informatik 2012/13”

$\longrightarrow$ Hit 1 oder Hit 2

StofI-Webseite:

http://ismi.math.uni-

frankfurt.de/wakolbinger/teaching/StofI1213/

Oder:

google $\longrightarrow$ “Stochastik für die Informatik 2012/13”

$\longrightarrow$ Hit 1 oder Hit 2

Oder:

google $\longrightarrow$ “Wakolbinger” $\longrightarrow$ Hit 1

Tutorien

Übungsgruppen:

Mo.	10-12 Uhr	Raum 711 klein	Benjamin Straub
Di.	08-10 Uhr	Raum 711 klein	Christopher Paulus Imanto
Mi.	10-12 Uhr	Raum 310	Philipp Klein
Do.	10-12 Uhr	Raum 310	Anna Meiser
Fr.	10-12 Uhr	Raum 902	Xiao Yin.

Offenes Tutorium:

Di. 13-14 Uhr, Mi. 9-10 Uhr, Lernzentrum Informatik, Ute Lenz.

Anmeldung zu einer Übungsgruppe:

elektronisch über die StofI - Webseite

bis Mittwoch 17. Oktober 2012, 24 Uhr

Freitags: Ausgabe des Übungsblatts.

Tipps zu den Übungsaufgaben:

in den Tutorien (ab nächster Woche)

Termin für die Abgabe der schriftlichen Lösungen der “S-Aufgaben”:

zwei Wochen nach Ausgabe

In der Woche drauf

werden die Lösungen in den Tutorien besprochen.

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.

Abschlussklausur:

Freitag, 15. Februar 2013, 12:15-13:45 Uhr.

Dabei können 100 Klausurpunkte erreicht werden.

Die Note errechnet sich aus

der Summe der Anzahl der erreichten Klausurpunkte

plus der Anzahl der erreichten Bonuspunkte.

Beträgt diese Summe mindestens 50,

gilt die Abschlussprüfung über die Veranstaltung

als bestanden.

Lehrbuch:

Götz Kersting, Anton Wakolbinger

Elementare Stochastik, Birkhäuser, 2. Aufl. 2010

Semesterausleihe möglich aus der

Bibliothek des Mathematischen Seminars,

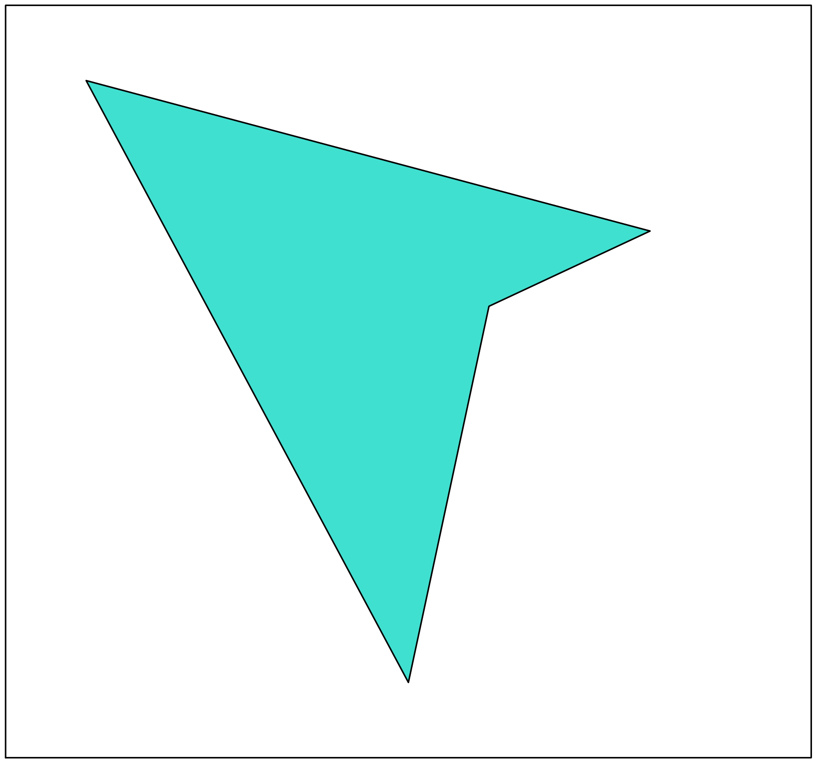
Robert-Mayer-Str. 8, 4. Stock

Vorlesung 1a

Ereignisse, Wahrscheinlichkeiten,

Zufallsvariable, Verteilungen:

Ein Beispiel



Wie groß ist der Anteil der blauen Fläche

an der Fläche des Quadrats?

Angenommen wir haben ein Werkzeug, mit dem man

einen **rein zufälligen Punkt** aus dem Quadrat wählen kann

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

X ist *rein zufälliger Punkt aus* S

bedeutet:

Für jede “messbare” Teilmenge A von S ist

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

lies:

die Wahrscheinlichkeit, dass X in A fällt

ist der Anteil der Fläche von A an der Fläche von S.

$\mathbf{P}$ steht für **probabilitas** (= Wahrscheinlichkeit)

Wir bekommen damit nicht nur *einen* rein zufälligen Punkt, sondern sogar

eine *rein zufällige Folge* $(X_1, X_2, ...)$

von Punkten in S.

Sei $A \subset S$.

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

zählt, ob X_i in A fällt.

Dabei ist

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

Z_i ist der *Indikator des Ereignisses “X_i fällt in A”*.

Alternative Schreibweise: $I_{\{X_i \in A\}}$

Die zufällige Zahl

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

ist ein *Schätzer* für die Wahrscheinlichkeit

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

(und damit für den gefragten Flächenanteil).

Es sei verraten:

Der wahre Anteil der blauen Fläche am Quadrat ist

$$p = 0.195$$

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

$$p = 0.195$$

Damit hat M gar keine Chance, exakt auf p zu fallen, denn

der *Wertebereich* von M

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

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

Eingangs hatten wir uns erst einmal

eine Realisierung von M verschafft.

Wie “zuverlässig” ist dieser Schätzwert?

Die Verteilung von M ist bestimmt durch Gewichte

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

Davon hatten wir uns ein Bild gemacht,

indem wir 1000 “unabhängige Kopien” von M erzeugten

und in einem *Histogramm* darstellen,

wie oft welche Ausgänge realisert wurden.

Die 101 möglichen Ausgänge von M

sind (bei weitem) nicht gleich wahrscheinlich:

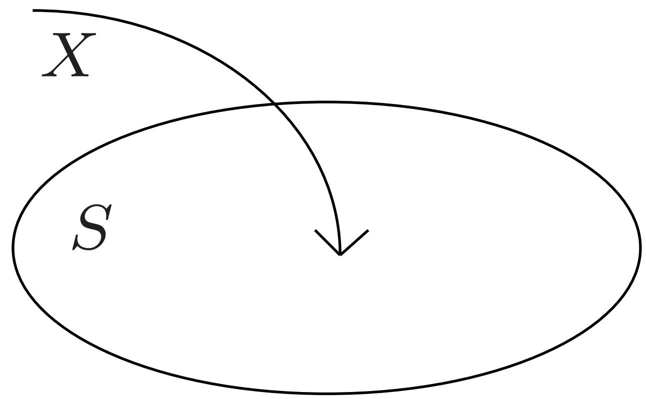
die Verteilung von M “ist um p konzentriert”.

Zusammenfassung

der wichtigsten Begriffe

der ersten Stunde:

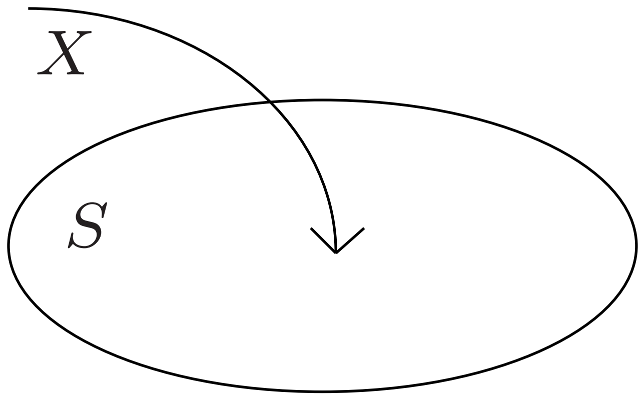
Ein Logo der Elementaren Stochastik:



X ...zufällige Wahl eines Elements aus S

S ...Menge von möglichen Ausgängen

Ein Logo der Elementaren Stochastik:

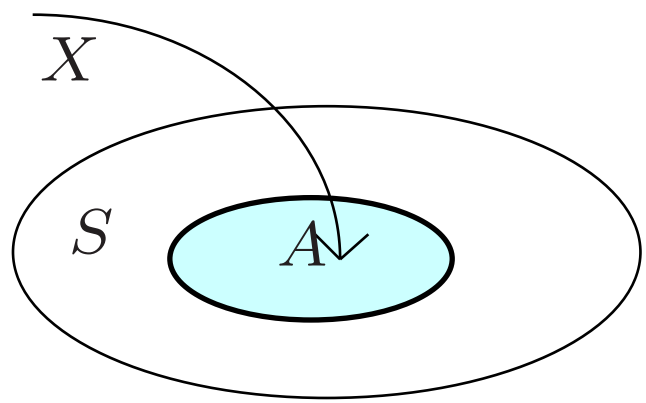


X . . . **Zufallsvariable**

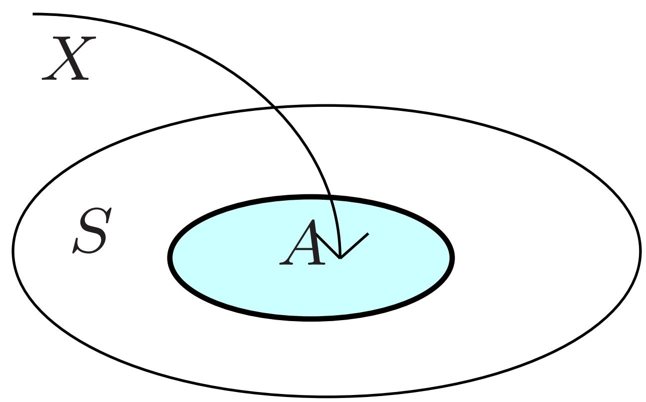
mit *Zielbereich (Wertebereich)* S

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 *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 $\mathbf{P}(\{X \in A\})$ scheiben wir kurz:

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

Zur Abrundung und

zur Einstimmung auf die zweite Vorlesung:

Ein Beispiel zum Bekanntschaftmachen

mit den grundlegenden Begriffen

Zufallsvariable, Zielbereich,

Ereignis,

Wahrscheinlichkeit

$n = 25$ Individuen,

$r = 365$ Plätze.

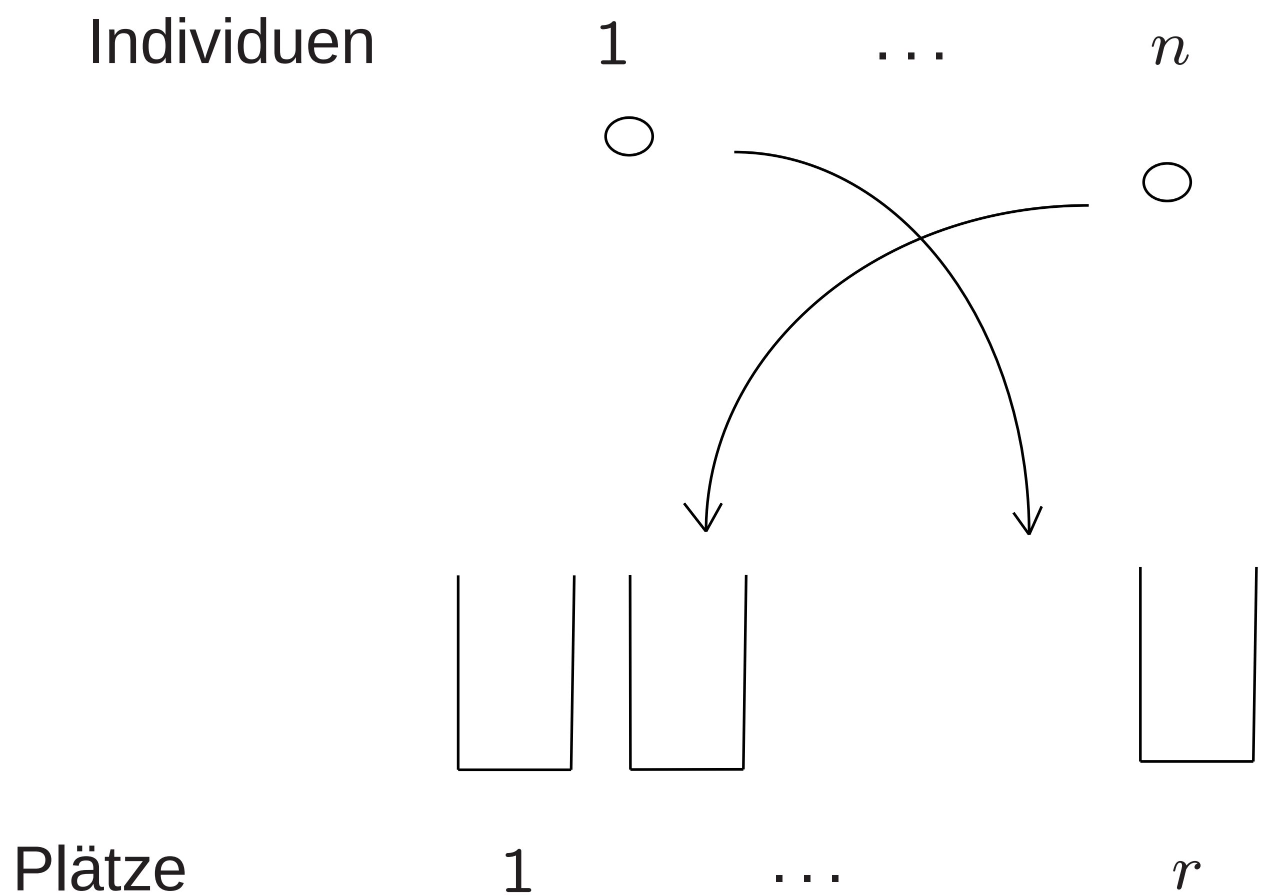
Jedes Individuum wird auf einen

rein zufällig ausgewählten Platz gesetzt.

(Mehrfachbelegungen sind erlaubt!)

Wie wahrscheinlich ist es, dass dabei

keine Mehrfachbelegung auftritt?


Individuen
1
...
n
Plätze
1
...
r

Ein anderer Blick:

Jedes von n Individuen

ist mit je einem von r möglichen Kennzeichen versehen,

das vom Zufall bestimmt ist.

Wie stehen die Chancen,

dass keine der zwei Individuen gleich gekennzeichnet sind?

(Zitat aus “dem Buch”, Seite 1)

In der Informatik denkt man bei den Individuen an Daten

und spricht bei den Kennzeichen

von Hash-Werten oder Fingerabdrücken.

Populäre Version:

$n = 25$ Leute auf einer Party

Kennzeichen . . . Geburtstag ($\in \{1, 2, \ldots, 365\}$)

Wie wahrscheinlich ist es,

dass keine zwei Leute am selben Tag Geburtstag haben?

Die Individuen denken wir uns mit 1 bis n

und die Kennzeichen mit 1 bis r nummeriert.

Ein Ausgang der Kennzeichnung lässt sich beschreiben

durch das n-tupel

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

wobei a_i das Kennzeichen des i-ten Individuums bezeichnet

$(1 \le a_i \le r)$.

Die Menge der möglichen Ausgänge der Kennzeichnung ist

$$S := \{1, \ldots, r\}^n ,$$

die Menge aller n-tupel $(a_1, \ldots, a_n)$ mit $a_i \in \{1, \ldots, r\}$.

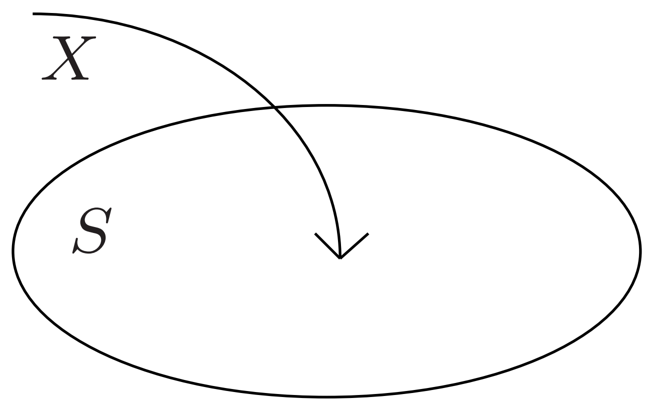
Manchmal schreiben wir dafür auch

$$S := \{1, \ldots, r\}^{\{1,\ldots,n\}},$$

die Menge aller Abbildungen von $\{1, \ldots, n\}$ nach $\{1, \ldots, r\}$.

Den zufälligen Ausgang der Kennzeichnung

beschreiben wir durch eine *Zufallsvariable* X.



X kommt durch zufällige Wahl

eines Elementes aus S zustande.

Die Menge S heißt *Zielbereich* 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

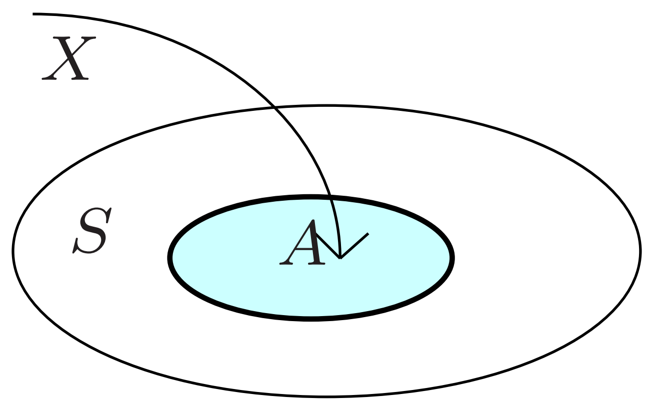
$$\{X_i \neq X_j \text{ für alle } i \neq j\}$$

oder auch als

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

mit der Teilmenge

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



Wie kommt man zur **Wahrscheinlichkeit**

des Ereignisses $\{X \in A\}$?

Zur Erinnerung:

Wahrscheinlichkeiten gehören zu Ereignissen

und messen deren Chance einzutreten

mit einer Zahl zwischen 0 und 1:

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

Zwei einleuchtende Regeln

für das Rechnen mit Wahrscheinlichkeiten:

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

d.h. das *sichere Ereignis* $\{X \in S\}$ hat Wahrscheinlichkeit 1

und (bei endlich vielen möglichen Ausgängen, d.h. $\#S < \infty$)

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

(Additivität.)

Um die Wahrscheinlichkeit $\mathbf{P}(X \in A)$ berechnen zu können,

muss man eine Modellannahme treffen.

Unsere Modellannahme:

rein zufällige 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.

Also:

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

Und

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

Wir haben nun die Aufgabe des Abzählens der Mengen

$$S := \{1, \ldots, r\}^n = \{(a_1, \ldots, a_n) : 1 \le a_i \le r\}$$

und

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

$$\#S = r^n$$

$$S = \{(a_1, \ldots, a_n) : 1 \leq a_i \leq r\}$$

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

$$\#A = ?$$

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

Also:

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

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