Vorlesung 1b

Wahrscheinlichkeit von Kollisionen

bei der wiederholten rein zufälligen Wahl

aus endlich vielen möglichen Ausgängen

(Buch S. 1-5)

Teil Eins:

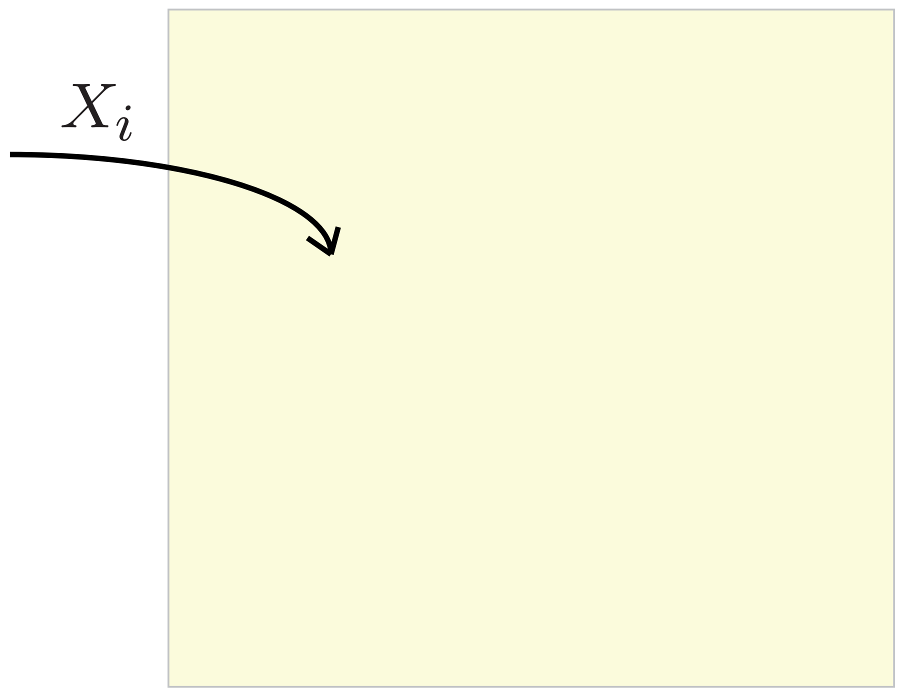
Vom Modell zur Formel

1. Die Fragestellung

– in verschiedenen Verpackungen

n-mal wiederholt wird rein zufällig

ein Pixel aus $g = 10^6$ Pixeln gewählt.



$$i = 1, \ldots, n$$

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

$n = 100$?

$n = 1000$?

$n = 10000$?

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 $1, 2, \ldots, g$.

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.

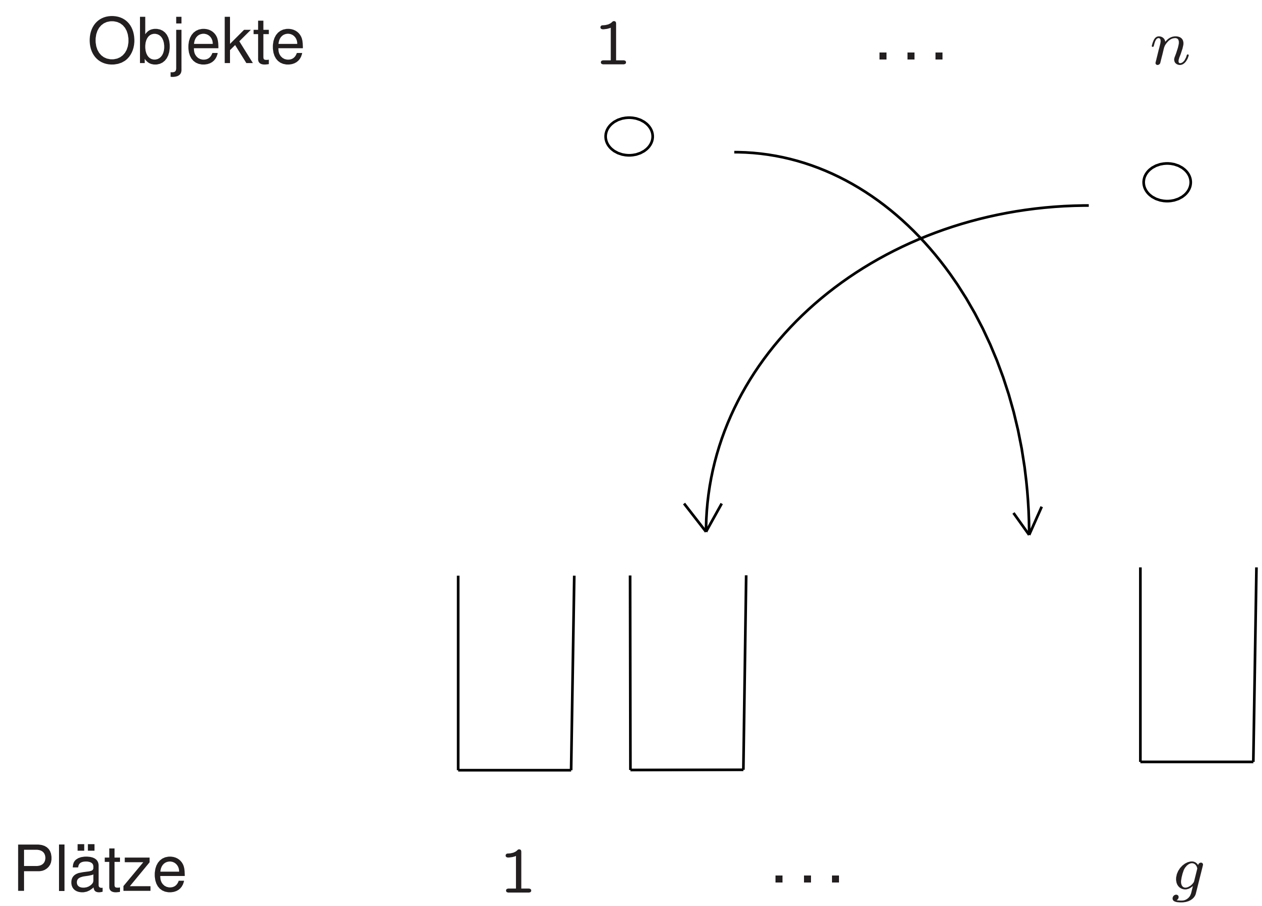
Jedes Objekt wird auf einen

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 $\quad (\in \{1, 2, \ldots, 365\})$

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

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

wobei a_i den für das i-te Objekt gewählte Platz bezeichnet

$(1 \leq a_i \leq g)$.

Die Menge der möglichen Wahlprotokolle ist

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

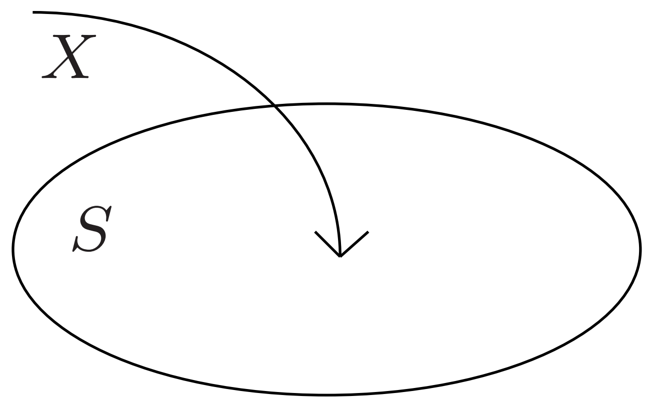
die Menge aller n-tupel $(a_1, \ldots, a_n)$

mit Einträgen (Komponenten)

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

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 $(a_1, \ldots, a_n)$ unserer Menge S

besteht auch die Zufallsvariable X aus n Komponenten:

$$X = (X_1, \ldots, X_n) \; .$$

Wir interessieren uns für das *Ereignis*,

dass keine zwei Komponenten von X gleich sind.

Dieses Ereignis schreiben wir als

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

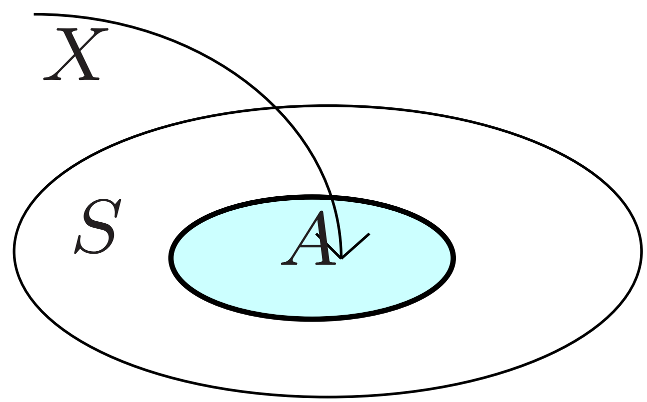
oder auch als

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

mit der Teilmenge

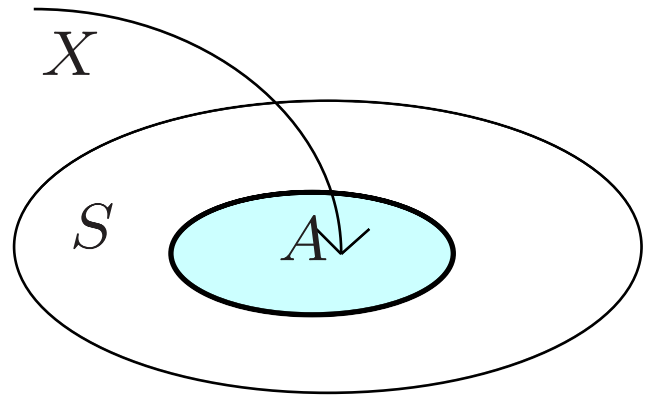
$$A := \{a \in S \,:\, a_i \neq a_j \text{ für alle } i \neq j\}\ .$$

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

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

und

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

$$\#S = g^n$$

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

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

$$\#A = ?$$

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

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

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

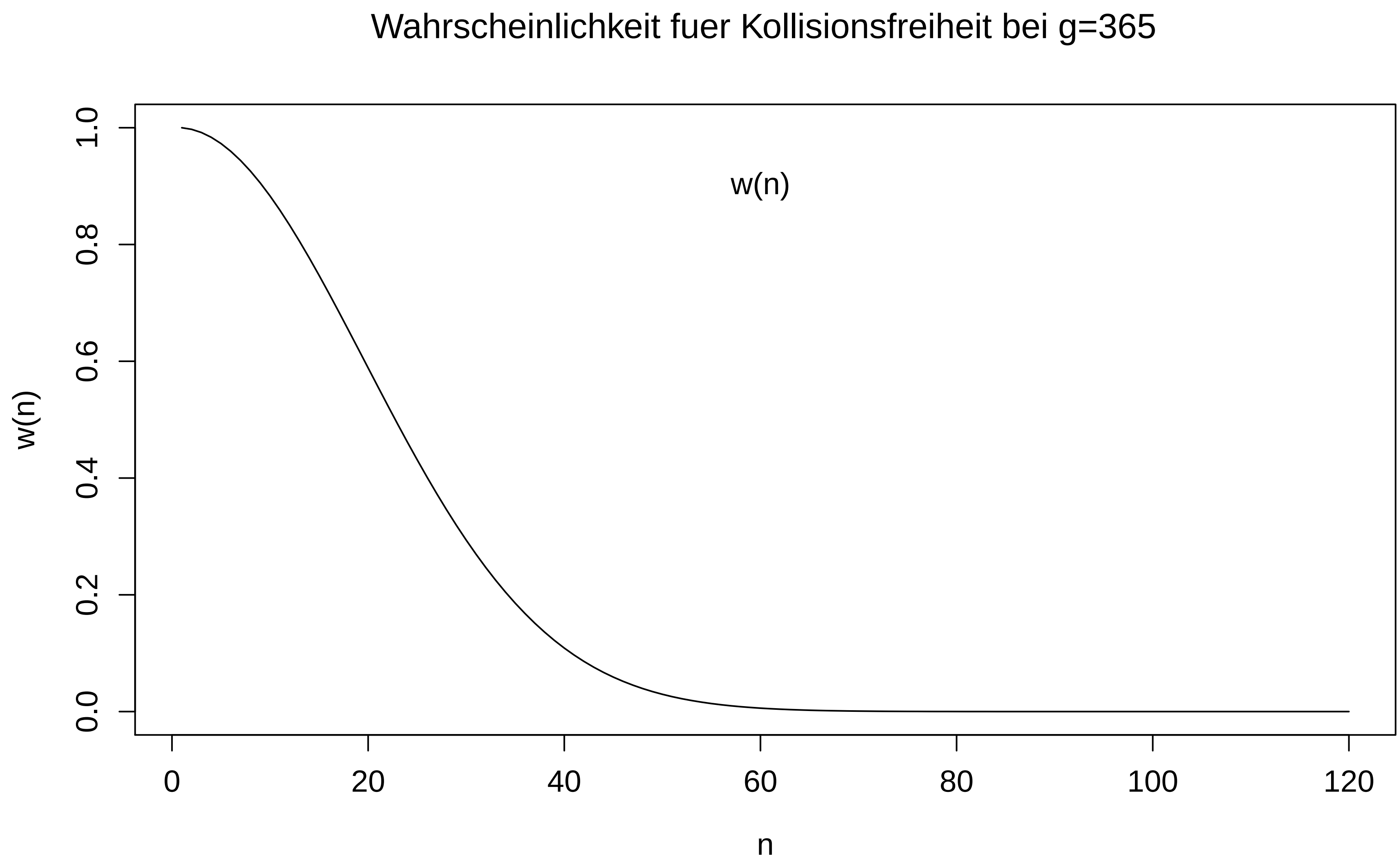
$$= \frac{g-1}{g}\,\frac{g-2}{g}\cdots\frac{g-(n-1)}{g}$$

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

ist die Wahrscheinlichkeit für Kollisionsfreiheit

bei der n-maligen wiederholten rein zufälligen Platzwahl

aus g Plätzen


Wahrscheinlichkeit fuer Kollisionsfreiheit bei g=365
w(n)
1.0
0.8
0.6
0.4
0.2
0.0
w(n)
0
20
40
60
80
100
120
n

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 ($i = 1, 2, \ldots, g+1$)
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”

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

ist somit eine Zufallsvariable mit Wertebereich

$$S :=$$

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

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

ist der *Zeitpunkt der ersten Kollision*.

T ist eine aus $(X_1, \ldots, X_{g+1})$ 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

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

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:

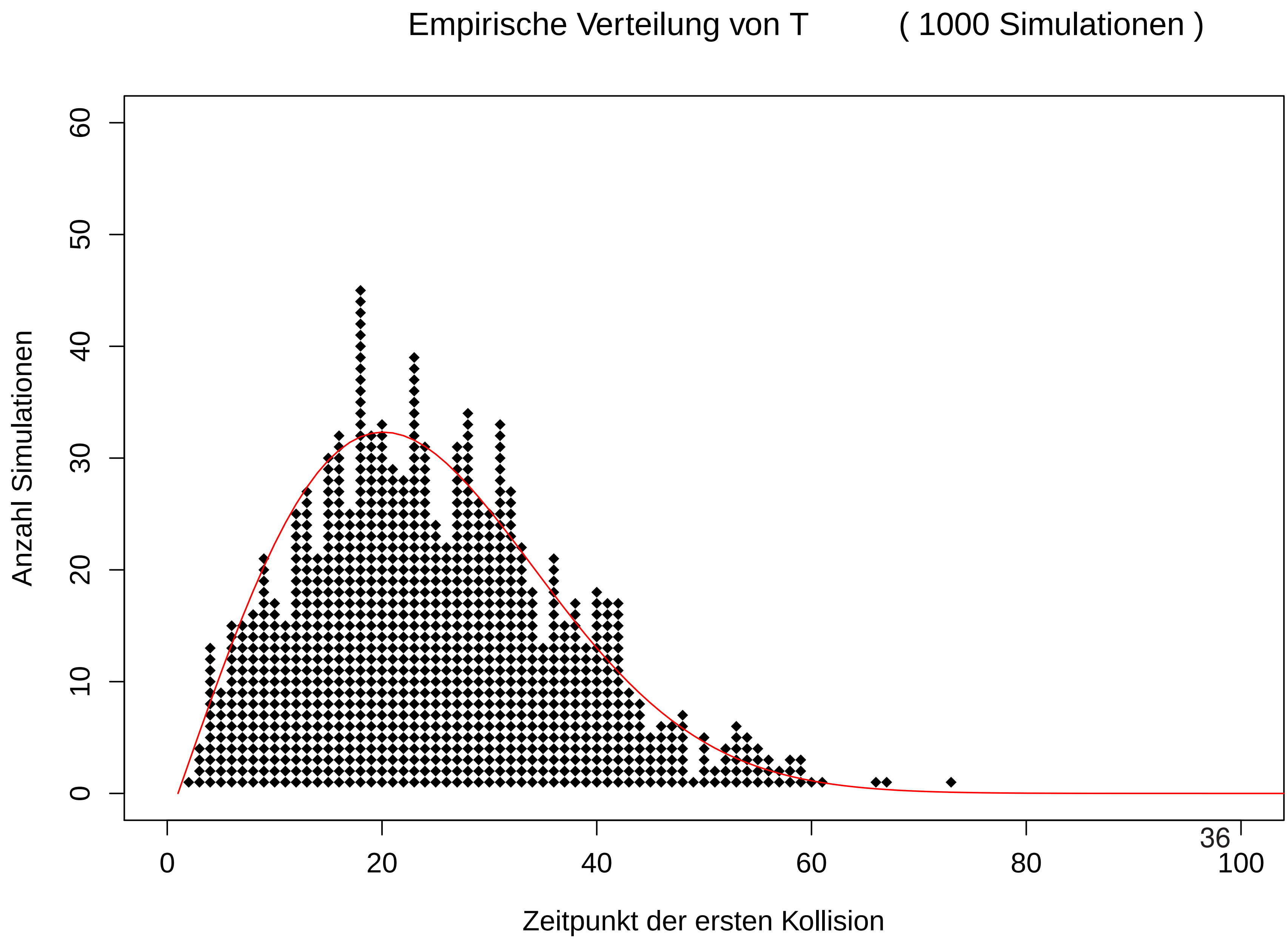
$T = h(X)$ mit passendem $h : S \to \mathbb{N}$.

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)
Anzahl Simulationen
0
10
20
30
40
50
60
0
20
40
60
80
100
Zeitpunkt der ersten Kollision