

Statistik für Naturwissenschaftler 431.010

Woche 04-06: Wahrscheinlichkeitsrechnung (mit R Unterstützung)

Ass.-Prof. Dr. Wolfgang Trutschnig






Fachbereich Mathematik
Universität Salzburg
www.trutschnig.net

Salzburg, April 2016



Beispiel (Münzwurf & Würfeln)

- ▶ Eine 1-Euro Münze wird geworfen.
- ▶ Wie groß ist die Wahrscheinlichkeit, dass 'Zahl' kommt ?
- ▶ Wie groß ist die Wahrscheinlichkeit, dass 'Kopf' kommt ?

- ▶ Ein (fairer) Würfel wird geworfen
- ▶ Wie groß ist die Wahrscheinlichkeit, dass  gewürfelt wird ?
- ▶ Wie groß ist die Wahrscheinlichkeit, dass  gewürfelt wird ?
- ▶ Wie groß ist die Wahrscheinlichkeit, dass  oder  gewürfelt wird ?
- ▶ Wie groß ist die Wahrscheinlichkeit, dass nicht  gewürfelt wird ?

- ▶ Woher kennen wir alle diese Wahrscheinlichkeiten ?



Münzwurf und Würfeln

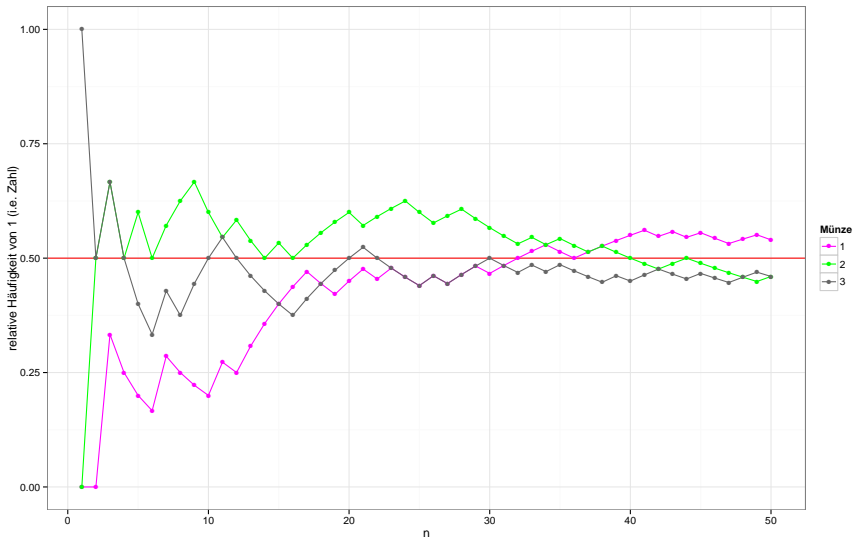
- ▶ Wenn wir 100.000 mal eine 1-Euro Münze werfen, wird in etwa in 50% der Fälle Kopf kommen.
- ▶ Überprüfung 'händisch' ?
- ▶ Lassen R für uns Münze werfen:


```
x<-sample(c("Kopf","Zahl"),size=100000,replace=TRUE)
table(x)
summary(x)
```
- ▶ Ergebnis beim ersten Durchlauf: 50.020 mal Kopf, 49.980 mal Zahl
- ▶ Ergebnis beim zweiten Durchlauf: 49.948 mal Kopf, 50.052 mal Zahl
- ▶ Handlicher: Ergebnis des Münzwurfs entweder 0 (Kopf) oder 1 (Zahl)
- ▶ In R:

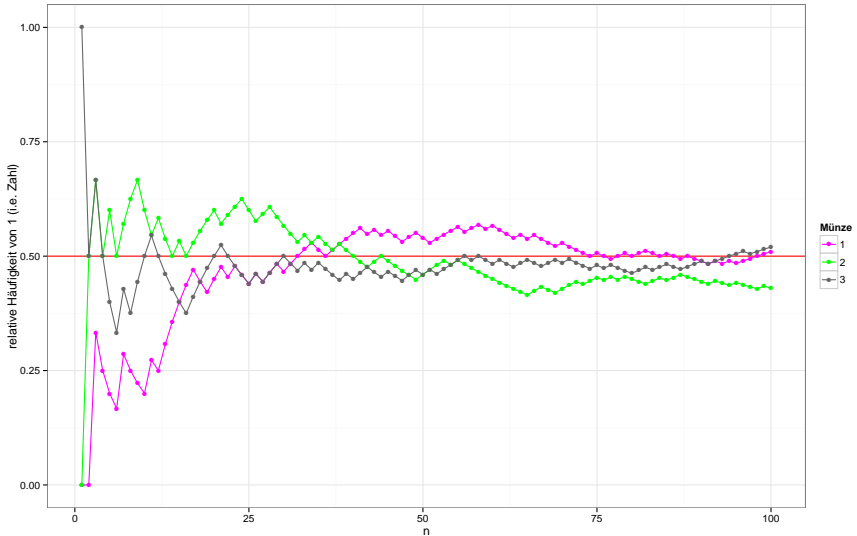
```
x<-sample(c(0,1),size=100000,replace=TRUE)
table(x)
mean(x)
```
- ▶ Bezeichnet x_1, \dots, x_n die so erhaltene Stichproben mit $n = 100.000$ dann beobachten wir also $h_n(\{1\}) = \bar{x}_n \approx \frac{1}{2}$



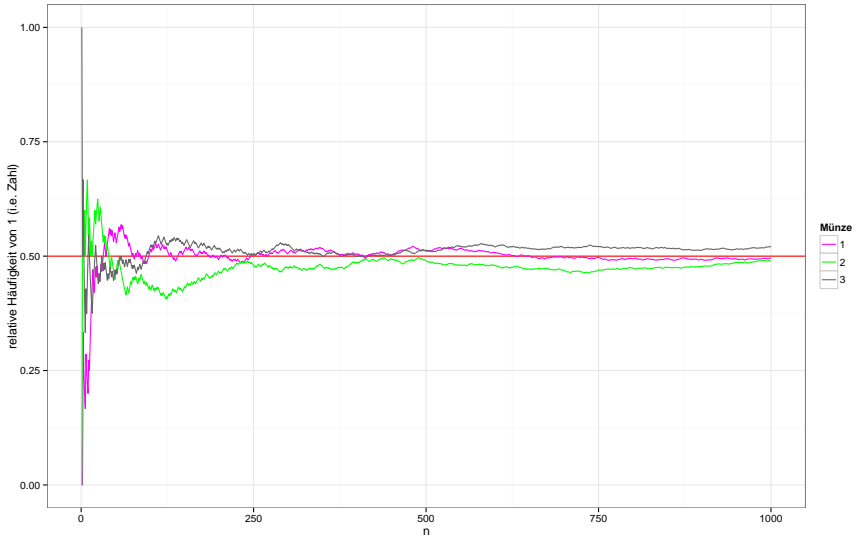
3 Münzen werden jeweils 100.000 mal geworfen – erste 50 Würfe



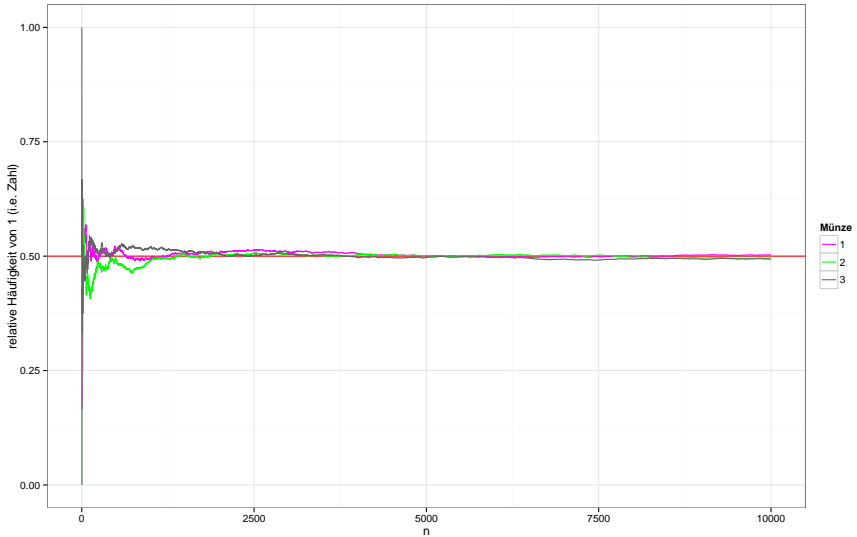
3 Münzen werden jeweils 100.000 mal geworfen – erste 100 Würfe









3 Münzen werden jeweils 100.000 mal geworfen – erste 1.000 Würfe



3 Münzen werden jeweils 100.000 mal geworfen – erste 10.000 Würfe

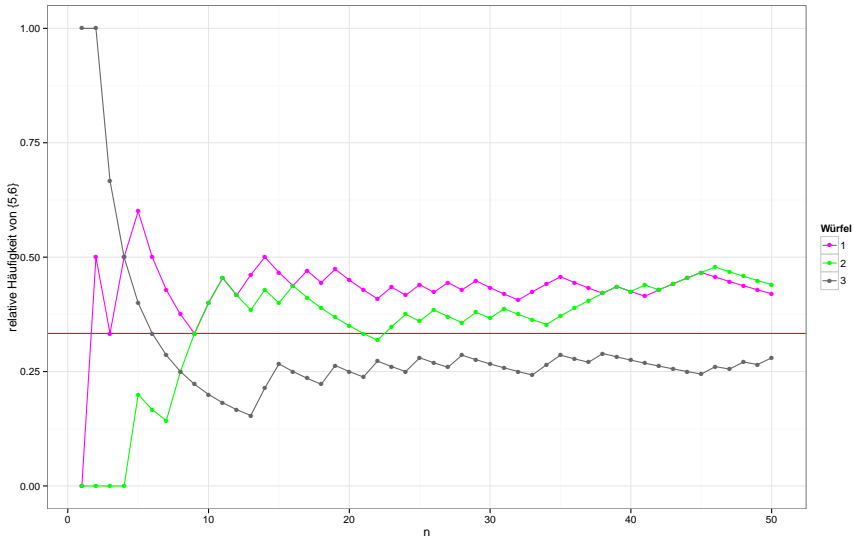


- ▶ Wir würfeln 100.000 mal - wie oft wird in etwa  oder  das Ergebnis sein ?
- ▶ Identifizieren Augensummen mit Zahlen und lassen R für uns würfeln: `x<-sample(c(1,2,3,4,5,6), size=100000,replace=TRUE)`
`table(x)`
- ▶ Ergebnis beim ersten Durchlauf: 16.603 mal ; 16.568 mal ;
- ▶ Insgesamt tritt also 33.171 mal das Ereignis $\{\text{die } 5, 6\}$ auf
- ▶ Ergebnis beim zweiten Durchlauf: 16.523 mal ; 16.662 mal ;
- ▶ Insgesamt tritt beim zweiten Durchlauf also 33.185 mal das Ereignis $\{\text{die } 5, 6\}$ auf
- ▶ Bezeichnet x_1, \dots, x_n die so erhaltene Stichproben mit $n = 100.000$, dann beobachten wir also $h_n(\{\text{die } 5, 6\}) \approx \frac{2}{6} = \frac{1}{3}$.
- ▶ Wie kann $h_n(\{\text{die } 5, 6\})$ einfachstmöglich in R berechnet werden ?
- ▶ `z<-ifelse(x==5 | x==6,1,0)`
`mean(z)`

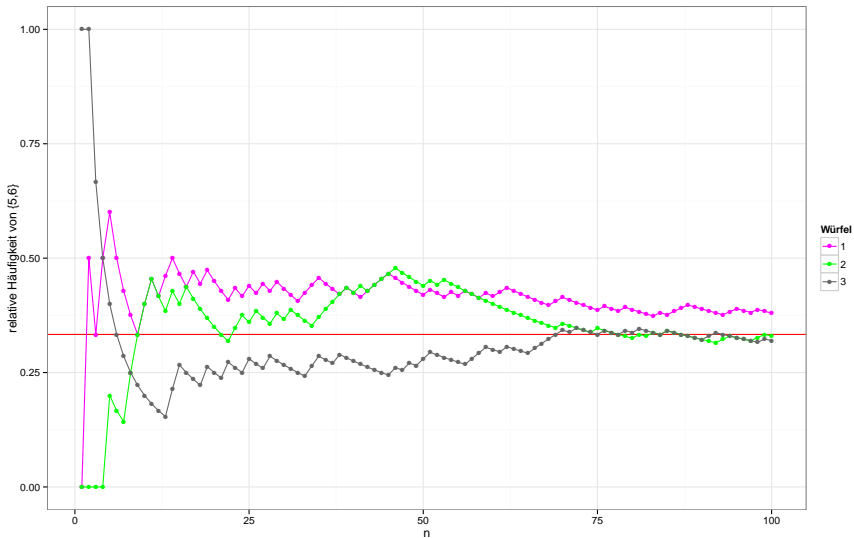


Münzwurf und Würfeln

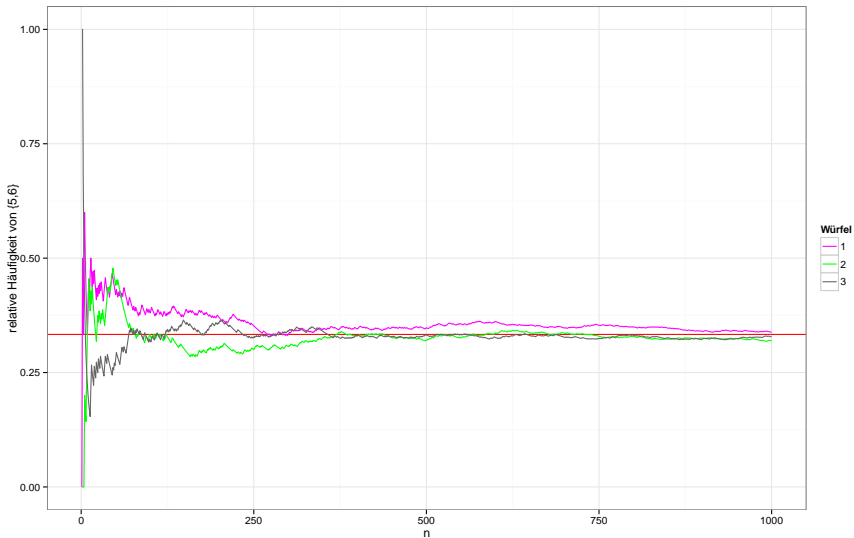
3 Würfel werden jeweils 100.000 mal geworfen; relative Häufigkeit von {5,6} – erste 50 Würfe



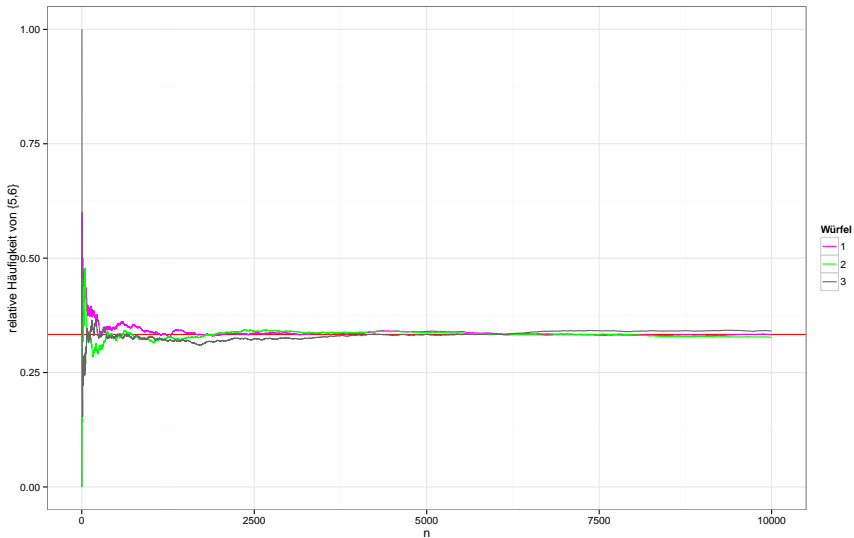
3 Würfel werden jeweils 100.000 mal geworfen; relative Häufigkeit von {5,6} – erste 100 Würfe



3 Würfel werden jeweils 100.000 mal geworfen; relative Häufigkeit von {5,6} – erste 1.000 Würfe



3 Würfel werden jeweils 100.000 mal geworfen; relative Häufigkeit von {5,6} – erste 10.000 Würfe



- ▶ **Wahrscheinlichkeiten sind Grenzwerte von relativen Häufigkeiten !!**
- ▶ → Frequentistischer Wahrscheinlichkeitsbegriff

Formalisierung:

- ▶ Mögliche Ausgänge beim Münzwurf: $\Omega = \{\text{Kopf}, \text{Zahl}\}$
- ▶ Wahrscheinlichkeiten: $\mathbb{P}(\{\text{Kopf}\}) = \frac{1}{2} = \mathbb{P}(\{\text{Zahl}\})$
- ▶ Mögliche Ausgänge beim Würfeln:

$$\Omega = \left\{ \begin{array}{|c|} \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \end{array}, \begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline \end{array} \right\}$$

- ▶ Wahrscheinlichkeiten: Alle sechs Ausgänge haben die Wahrscheinlichkeit $\frac{1}{6}$, i.e. alle Ausgänge sind gleich wahrscheinlich
- ▶ In beiden Fällen haben wir die Wahrscheinlichkeiten von $E \subseteq \Omega$ berechnet als 'Anzahl gute' durch 'Anzahl mögliche' Ausgänge, i.e.

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega}$$



Definition

Sei die Menge Ω aller Ausgänge eines Zufallsexperiments endlich und jeder Ausgang in Ω gleich wahrscheinlich, i.e. $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ und $\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n}$. Dann heißt (Ω, \mathbb{P}) Laplace Raum und für jedes $E \subseteq \Omega$ wird $\mathbb{P}(E)$, die Wahrscheinlichkeit von E , berechnet durch

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega}$$

Beispiel

- ▶ Eine Münze wird zwei mal geworfen
- ▶ Die Menge Ω aller möglichen Versuchsausgänge ist gegeben durch:

$$\Omega = \{(\text{Kopf, Kopf}), (\text{Kopf, Zahl}), (\text{Zahl, Kopf}), (\text{Zahl, Zahl})\}$$

- ▶ $\#\Omega = 4$ und alle Versuchsausgänge sind gleich wahrscheinlich



Beispiel (cont.)

- ▶ Gesucht: Wahrscheinlichkeit dafür, dass der erste und der zweite Wurf unterschiedlich ausgehen
- ▶ In diesem Fall ist $E = \{(Kopf, Zahl), (Zahl, Kopf)\}$
- ▶ Wir erhalten daher $\mathbb{P}(E) = \frac{2}{4} = \frac{1}{2}$
- ▶ Siehe R-Code 'R-codes_GEO_04.R'- Simulation des 'Experiments'
- ▶ Interpretation: Wenn wir das Experiment 'zwei mal Münze werfen' 100.000 mal wiederholen beobachten wir in ca. 50.000 Fällen, dass der erste und der zweite Wurf unterschiedlich ausgehen.

- ▶ Gesucht: Wahrscheinlichkeit dafür, dass nie Kopf kommt
- ▶ In diesem Fall ist $E = \{(Zahl, Zahl)\}$
- ▶ Wir erhalten daher $\mathbb{P}(E) = \frac{1}{4}$
- ▶ Siehe R-Code 'R-codes_GEO_04.R' - Simulation des 'Experiments'



Beispiel

- ▶ Gesucht: Wahrscheinlichkeit dafür, dass eine zufällig aus dem RTR-Datensatz gewählte ('gezogene') Messung mit Android Betriebssystem durchgeführt wurde
- ▶ Einfachstes Modell $\Omega = \{1, 2, 3, \dots, 110.337\}$ (warum ?)
- ▶ Alle Versuchsausgänge sind gleich wahrscheinlich
- ▶ $E = \{i \in \Omega : \text{Messung } i \text{ wurde mit Android device durchgeführt}\}$
- ▶ Wir erhalten mit Hilfe von R un schwer $\#E = 75.093$
- ▶ Also insgesamt $\mathbb{P}(E) = \frac{\#E}{\#\Omega} = \frac{75.093}{110.337} \approx 0,6806$
- ▶ Wenn wir also 100.000 mal zufällig eine Messung aus dem RTR-Datensatz 'ziehen', erhalten wir in ca. 68% der Fälle eine Messung, die mit einem Android device durchgeführt wurde.
- ▶ Wie könnten diese 100.000 Ziehungen mit R umgesetzt werden ?
- ▶ Siehe R-Code 'R-codes_GEO_04.R' - Simulation des 'Experiments'



Beispiel

- ▶ Eine Münze wird drei mal geworfen
- ▶ Gesucht: Wahrscheinlichkeit dafür, dass genau ein mal 'Kopf' kommt
- ▶ Wir bestimmen als erstes wieder die Menge Ω aller möglichen Versuchsausgänge
- ▶ Zur Vereinfachung der Notation schreiben wir 0 falls 'Kopf' kommt und 1 falls Zahl kommt und erhalten

$$\Omega = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

- ▶ Alle Versuchsausgänge sind gleich wahrscheinlich
- ▶ Das gesuchte Ereignis E lässt sich damit schreiben als
 $E = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$



- ▶ Wir erhalten also

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega} = \frac{3}{8}$$

- ▶ Korrekte Interpretation: Wenn wir das Experiment 'drei mal Münze werfen' 100.000 mal wiederholen beobachten wir in ca. 37.500 Fällen, dass genau ein mal 'Kopf' kommt.
- ▶ Korrekte Interpretation: Wenn wir das Experiment 'drei mal Münze werfen' 300.000 mal wiederholen beobachten wir in ca. $\frac{3}{8} \cdot 300.000 = 112.500$ Fällen, dass genau ein mal 'Kopf' kommt.
- ▶ Simulation des Experiments mittels R - Übungsaufgabe



Beispiel (Wiederholung)

- ▶ Eine Münze wird $n = 3$ mal geworfen
- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau ein mal 'Zahl' kommt**
- ▶ Zur Vereinfachung der Notation schreiben wir 1 falls 'Zahl' kommt und 0 falls Kopf kommt und erhalten

$$\Omega = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

- ▶ Das gesuchte Ereignis E_1 lässt sich damit schreiben als $E_1 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$
- ▶ Wir erhalten also

$$\mathbb{P}(E_1) = \frac{\#E_1}{\#\Omega} = \frac{3}{8} = \frac{3}{2^3}$$



Beispiel (Fortsetzung)

- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau zwei mal 'Zahl' kommt**
- ▶ Das gesuchte Ereignis E_2 lässt sich damit schreiben als $E_2 = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$
- ▶ Wir erhalten also

$$\mathbb{P}(E_2) = \frac{\#E_2}{\#\Omega} = \frac{3}{8} = \frac{3}{2^3}$$

- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau drei mal 'Zahl' kommt**
- ▶ Das gesuchte Ereignis E_3 lässt sich damit schreiben als $E_3 = \{(1, 1, 1)\}$
- ▶ Wir erhalten also

$$\mathbb{P}(E_3) = \frac{\#E_3}{\#\Omega} = \frac{1}{8} = \frac{1}{2^3}$$



Beispiel (Fortsetzung)

- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass kein mal 'Zahl' kommt**
- ▶ Das gesuchte Ereignis E_0 lässt damit schreiben als $E_0 = \{(0, 0, 0)\}$
- ▶ Wir erhalten also

$$\mathbb{P}(E_0) = \frac{\#E_0}{\#\Omega} = \frac{1}{8} = \frac{1}{2^3}$$

- ▶ Haben damit alle möglichen Ausgänge betrachtet.
- ▶ Offensichtlich

$$\mathbb{P}(E_0) + \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) = 1$$



Beispiel (Fortsetzung)

- ▶ Interpretation der Wahrscheinlichkeiten:
- ▶ X ...Anzahl von Einsern
- ▶ Wenn wir das Experiment 'drei mal Münze werfen' 100.000 mal wiederholen erhalten wir ca.

X	Häufigkeit
0	12.500
1	37.500
2	37.500
3	12.500







- ▶ Simulation in R liefert die folgenden Resultate (siehe R-codes_GEO_05.R)

X	Häufigkeit
0	12.609
1	37.382
2	37.565
3	12.444

- ▶ → Run simulations in R



Beispiel (3 mal Würfeln)

- ▶ Würfel wird $n = 3$ Mal geworfen
- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau ein mal  oder  kommt**
- ▶ Zur Vereinfachung der Notation schreiben wir 1 falls  oder  kommt und 0 falls nicht  oder  kommt und erhalten

$$\Omega = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\} \quad (\text{Baum aufzeichnen!})$$

- ▶ Das gesuchte Ereignis E_1 ('gute Äste') lässt sich damit schreiben als $E_1 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$
- ▶ Gilt wiederum

$$\mathbb{P}(E_1) = \frac{\#E_1}{\#\Omega} = \frac{3}{8} = \frac{3}{2^3} ?$$







- ▶ Wo liegt der Denkfehler ?



Beispiel (3 mal Würfeln, cont.)

- ▶ Wie groß ist die Wahrscheinlichkeit, dass $(0, 0, 1)$ auftritt (Baum!) ?
- ▶ $\mathbb{P}(\{(0, 0, 1)\}) = \frac{2}{3} \frac{2}{3} \frac{1}{3} = \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^1$
- ▶ Analog erhalten wir $\mathbb{P}(\{(0, 1, 0)\}) = \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^1 = \mathbb{P}(\{(1, 0, 0)\})$, also



$$\mathbb{P}(E_1) = 3 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2$$

- ▶ 3 Äste; **1** mal tritt  oder  auf; **2** Mal tritt nicht  oder  auf
- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau zwei mal  oder  kommt**
- ▶ Das gesuchte Ereignis E_2 ('gute Äste') lässt sich damit schreiben als $E_2 = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$; wir erhalten (analog zu vorher)



$$\mathbb{P}(E_2) = 3 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1$$



Beispiel (3 mal Würfeln, cont.)

- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau drei mal  oder  kommt**
- ▶ Das gesuchte Ereignis E_3 ('gute Äste') lässt sich damit schreiben als $E_3 = \{(1, 1, 1)\}$
- ▶ Wir erhalten (analog zu vorher)

$$\mathbb{P}(E_3) = 1 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^0$$

- ▶ **Gesucht: Wahrscheinlichkeit dafür, dass genau null mal  oder  kommt**
- ▶ Das gesuchte Ereignis E_0 ('gute Äste') lässt sich damit schreiben als $E_0 = \{(1, 1, 1)\}$ und wir erhalten

$$\mathbb{P}(E_0) = 1 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3$$





Beispiel (3 mal Würfeln, cont.)

- ▶ Struktur klar bis auf die Faktoren **3,3,1,1**
- ▶ $@E_1$: Eine Eins auf drei mögliche Stellen (Würfe) aufteilen $\rightarrow 3$ Möglichkeiten
- ▶ $@E_2$: Zwei Einser auf drei mögliche Stellen (Würfe) aufteilen $\rightarrow 3$ Möglichkeiten
- ▶ $@E_3$: Drei Einser auf drei mögliche Stellen (Würfe) aufteilen $\rightarrow 1$ Möglichkeit
- ▶ $@E_0$: Null Einser auf drei mögliche Stellen (Würfe) aufteilen $\rightarrow 1$ Möglichkeit
- ▶ **Es gibt $\binom{n}{k}$ Möglichkeiten, k Objekte aus einer Menge von n verschiedenen Objekten auszuwählen (ohne Zurücklegen)**
- ▶ $\binom{n}{k}$ heißt *Binomialkoeffizient* und ist wie folgt definiert

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



Beispiel (3 mal Würfeln, cont.)

- ▶ $i! = i \cdot (i - 1) \cdot (i - 2) \cdot \dots \cdot 2 \cdot 1$ heißt Fakultät von i ; $0! = 1$
- ▶ Bsp: $\binom{3}{1} = \frac{3!}{1!2!} = 3$, $\binom{3}{2} = \frac{3!}{2!1!} = 3$
- ▶ Bezeichne X , wie oft bei 3-maligem Würfeln  oder  kommt, dann gilt insgesamt

$$\mathbb{P}(X = k) = \binom{3}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{3-k} \quad (1)$$

Beispiel (10 mal Würfeln, cont.)

- ▶ $n = 10$ mal Würfeln, X ...Anzahl, wie oft  oder  kommt

$$\mathbb{P}(X = 4) = \binom{10}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^6 \approx 0.22$$

- ▶ In R `dbinom(x=4,size=10,prob=1/3)`



Satz

Ein 'Experiment', das mit Wahrscheinlichkeit $p \in [0, 1]$ erfolgreich ist, wird n mal wiederholt. Die (Zufalls-) Variable X zählt die Anzahl der erfolgreichen Ausgänge. Dann ist die Wahrscheinlichkeit dafür, dass genau k Erfolge eintreten gegeben durch

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in \{0, 1, \dots, n\}.$$

Die (Zufalls-) Variable X nennen wir binomialverteilt mit Parameter n (Anzahl Versuche) und Wahrscheinlichkeit p , und schreiben $X \sim \text{Binom}(n, p)$.



Beispiel (Ölbohrungen, 'wildcat' exploration)

- ▶ Man weiß aus historischen Daten, dass unter gewissen geologischen Rahmenbedingungen, ca. 12% aller Bohrungen erfolgreich sind.
- ▶ **Gesucht:** Wahrscheinlichkeit dafür, bei 10 Bohrungen genau ein mal erfolgreich zu sein.
- ▶ Wir erhalten analog zu vorhin

$$\mathbb{P}(X = 1) = \binom{10}{1} (0.12)^1 (0.88)^9 \approx 0.38$$

- ▶ In R `dbinom(x=1,size=10,prob=0.12)`
- ▶ Für zwei Erfolge erhalten wir

$$\mathbb{P}(X = 2) = \binom{10}{2} (0.12)^2 (0.88)^8 \approx 0.23$$



Beispiel (Ölbohrungen, 'wildcat' exploration, cont.)

- ▶ **Gesucht:** Wahrscheinlichkeit dafür, bei 10 Bohrungen höchstens zwei mal erfolgreich zu sein.
- ▶ Wir erhalten analog zu vorhin

$$\begin{aligned} \mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0} (0.12)^0 (0.88)^{10} + \binom{10}{1} (0.12)^1 (0.88)^9 \\ &\quad + \binom{10}{2} (0.12)^2 (0.88)^8 \approx 0.8913 \end{aligned}$$

- ▶ In R `p<-dbinom(x=0:2,size=10,prob=0.12); sum(p)`



- ▶ In den vorigen Beispielen war $p = \frac{1}{2}$, $p = \frac{1}{3}$ und $p = 0.12$, daraus wurden dann Wahrscheinlichkeiten berechnet
- ▶ In der Praxis ist die Erfolgswahrscheinlichkeit p unbekannt und man möchte p basierend auf Daten *schätzen*
- ▶ Bsp: Gegeben ist die Stichprobe 1, 0, 1, 1, 0, 1, 0, 0, 1, 1*
- ▶ 6 Erfolge, 4 Misserfolge
- ▶ Wie würden Sie p schätzen?
- ▶ Naheliegend: $\hat{p} = \frac{6}{10} = 0.6$
- ▶ Muss der echte Wert genau 0.6 sein? Antwort: Nein
- ▶ Oft gibt man daher ein sogenanntes *Konfidenzintervall* an, i.e. ein Intervall, das den Parameter in 95% der Fälle enthält.
- ▶ In R `binom.test(6,10)`
- ▶ Liefert: 95 percent confidence interval : 0.2623781 0.8784477

*Denken Sie an eine Meinungsumfrage



- ▶ Die sogenannte *Poisson-Verteilung* ist in vielen Fällen ein akzeptables Modell für die Anzahl von Ereignissen, die innerhalb eines gewissen Zeitintervalls auftreten.
- ▶ Beispiele:
 - ▶ Anzahl von Erdbeben (in einer bestimmten Region) pro Jahr/Dekade/Jahrhundert
 - ▶ Anzahl von Hochwasserereignissen pro Jahr
 - ▶ Anzahl von in einer Versicherung eintreffenden Schadensmeldungen verursacht durch Wintersturm/Sommersturm/Hagel pro Jahr
 - ▶ Anzahl emittierte Partikel einer radioaktiven Substanz pro Zeiteinheit
 - ▶ Anzahl Anrufe in Telefonhotline pro Tag



Definition

Eine (Zufalls-) Variable X heißt *Poisson-verteilt* mit Parameter (Intensität) $\lambda > 0$ falls sie Werte in $\mathbb{N}_0 = \{0, 1, \dots\}$ annimmt und falls gilt

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \{0, 1, 2, \dots\}.$$

Wir schreiben im Folgenden kurz $X \sim \text{Pois}(\lambda)$.

Beispiel (Erdbeben, Poisson-Verteilung)

- ▶ Angenommen, die Anzahl X von jährlichen Erdbeben der Stärke ≥ 5 ist $\text{Pois}(6)$ -verteilt.
- ▶ Wie groß ist die Wahrscheinlichkeit dafür, dass in einem Jahr kein Erdbeben der Stärke ≥ 5 auftritt ?
- ▶ Wir erhalten

$$\mathbb{P}(X = 0) = e^{-6} \frac{6^0}{0!} \approx 0.00248$$



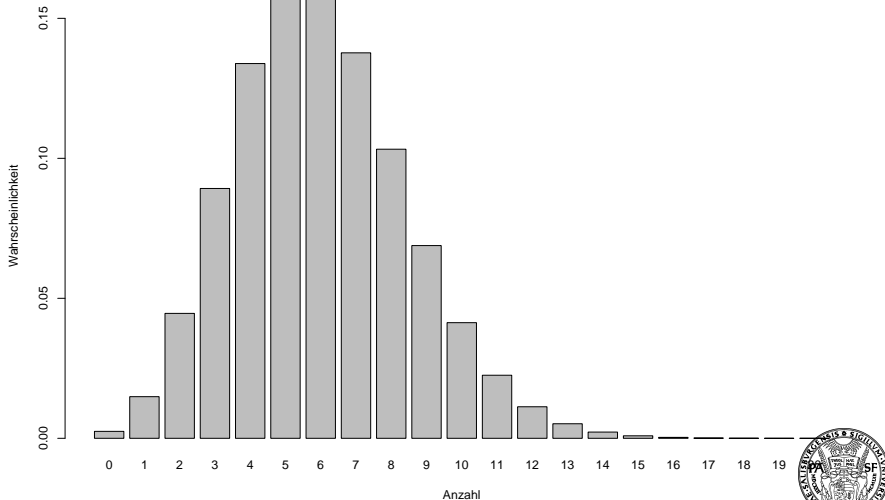
Beispiel (Erdbeben, Poisson-Verteilung, cont.)

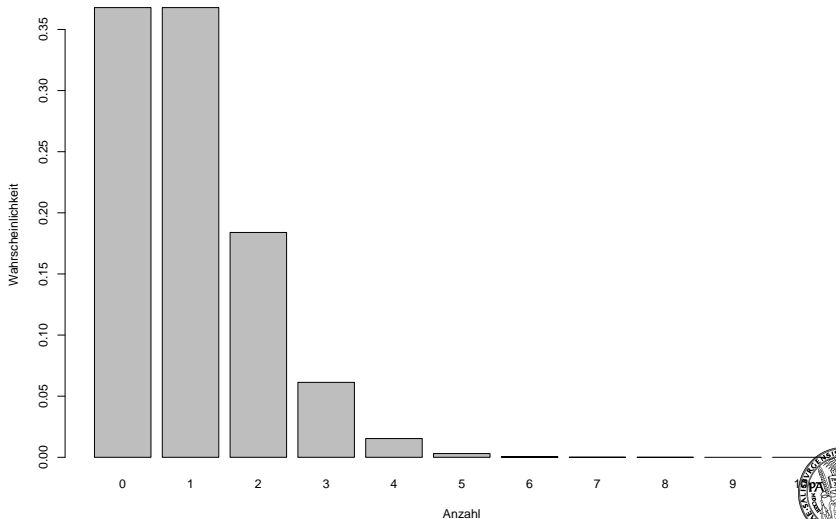
- ▶ Wie groß ist die Wahrscheinlichkeit dafür, dass in einem Jahr höchstens drei Erdbeben der Stärke ≥ 5 auftreten ?
- ▶ Wir erhalten

$$\begin{aligned} \mathbb{P}(X \leq 3) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) \\ &= e^{-6} \cdot \left(\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!} + \frac{6^3}{3!} \right) \approx 0.1512 \end{aligned}$$

- ▶ Interpretation ? In ca. 15% der Jahre treten höchstens drei Erdbeben der Stärke ≥ 5 auf.
- ▶ Berechnung in R: $\mathbb{P}(X = 3)$ `dpois(3, lambda=6)`
- ▶ Berechnung von $\mathbb{P}(X \leq 3)$: `sum(dpois(0:3, lambda=6))` oder (noch einfacher) `ppois(3, lambda=6)` (siehe Übungen)

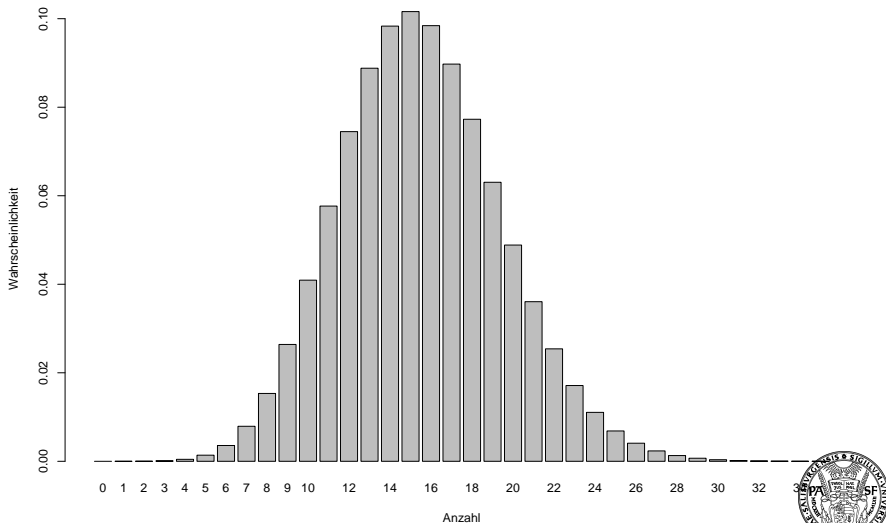


Poisson Verteilung mit Parameter $\lambda=6$ 

Poisson Verteilung mit Parameter $\lambda=1$ 



Poisson Verteilung

Poisson Verteilung mit Parameter $\lambda=15.5$ 

- ▶ Angenommen, wir wissen, dass die jährliche Anzahl X von Erdbeben Poisson-verteilt ist, wir kennen aber λ nicht.
- ▶ Wie können wir λ basierend auf historischen jährlichen Anzahlen x_1, \dots, x_n 'schätzen' ?
- ▶ Nähern uns der Antwort mit Hilfe von Simulationen
- ▶ Simulieren in R Stichproben von $X \sim \text{Pois}(6)$ und versuchen, aus den Daten den echten Parameter $\lambda = 6$ 'herauszufiltern'.
- ▶ `x <- rpois(n=1000, lambda=6); barplot(table(x))`
- ▶ Berechnen zusätzlich das Stichprobenmittel \bar{x}_n
- ▶ Siehe 'R-codes_GEO_06.R'
- ▶ Wichtige Beobachtung: \bar{x}_n liegt sehr nahe beim echten Parameter $\lambda = 6$ - Glück gehabt ?
- ▶ Was passiert wenn wir die sample size n erhöhen oder verringern ?
- ▶ Was passiert für andere Werte von λ ?



Beobachtungen:

- ▶ Für jeden Parameterwert λ liegt das Stichprobenmittel \bar{x}_n nahe bei λ falls n nicht zu klein ist.
- ▶ **Je größer der Stichprobenumfang desto näher liegt \bar{x}_n bei λ .**
- ▶ **Mehr Information** (größere Stichprobe/mehr Beobachtungen) **heißt bessere Schätzung von λ .**
- ▶ Der echte Parameter kann (nicht perfekt aber für großes n gut) aus den Daten 'herausgelesen' (geschätzt) werden.
- ▶ → Grundprinzip der (parametrischen) Statistik !
- ▶ 'Herauslesen' relevanter (zu Grunde liegender) Kenngrößen bzw. Parameter aus Beobachtungen
- ▶ Können die Parameter der Binomialverteilung aus Stichproben geschätzt werden ?
- ▶ → Meinungsbildung wiederum mit Hilfe von Simulationen (siehe Übungsaufgaben)



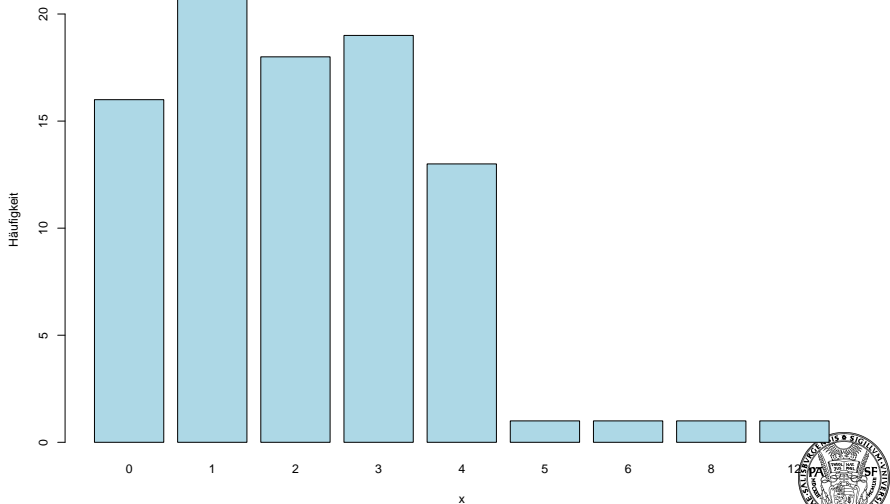
Beispiel (Erdbeben, Poisson-Verteilung, realistischer Datensatz)

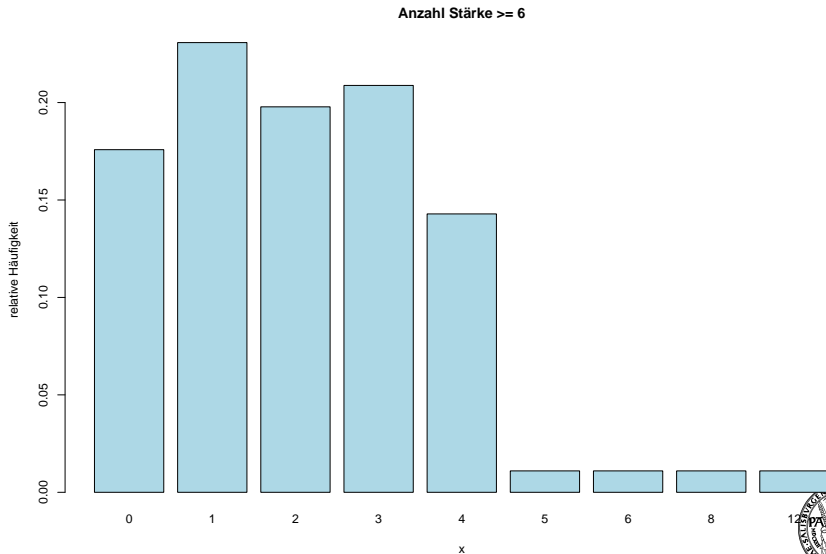
- ▶ Der Datensatz 'Erdbeben.RData' enthält die auf <http://www.usgs.gov/> aufgelisteten Erbeben der Stärke ≥ 6 in den Jahren 1900-1990 inkl. Stärke und Region:

ymd	year	month	day	region	magnitude
1900-10-09	1900	10	09	Kodiak Island, Alaska	7.70
1901-03-03	1901	03	03	Parkfield, California	6.40
1901-12-31	1901	12	31	Cook Inlet, Alaska	7.10
1902-04-19	1902	04	19	Guatemala	7.50
1902-12-16	1902	12	16	eastern Uzbekistan (Turkestan)	6.40
1903-04-28	1903	04	28	Turkey	7.00

- ▶ Der Datensatz wird in den Übungen analysiert
- ▶ Als ersten Schritt nehmen wir an, dass die Anzahl X von jährlichen Erbeben der Stärke ≥ 6 Poisson-verteilt ist und schätzen den Parameter λ .
- ▶ Siehe 'R-codes_GEO_06.R'

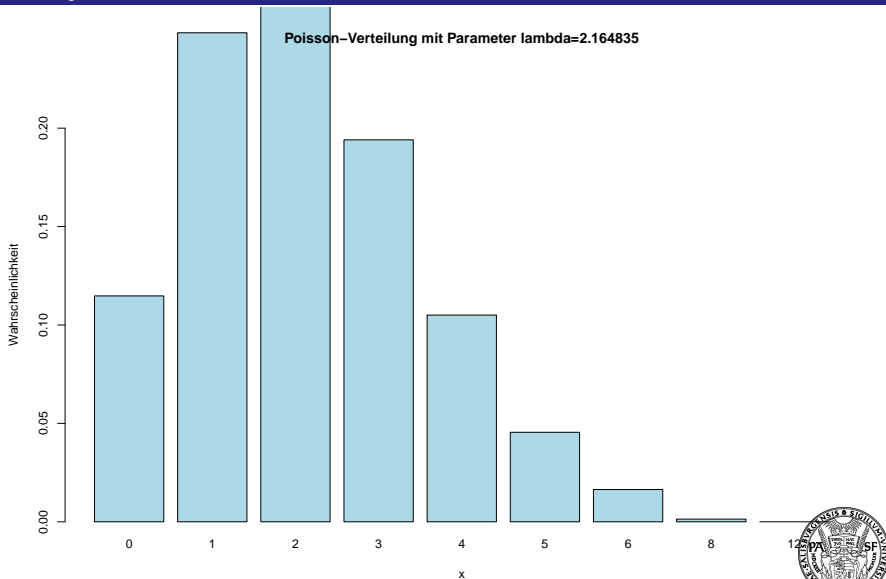


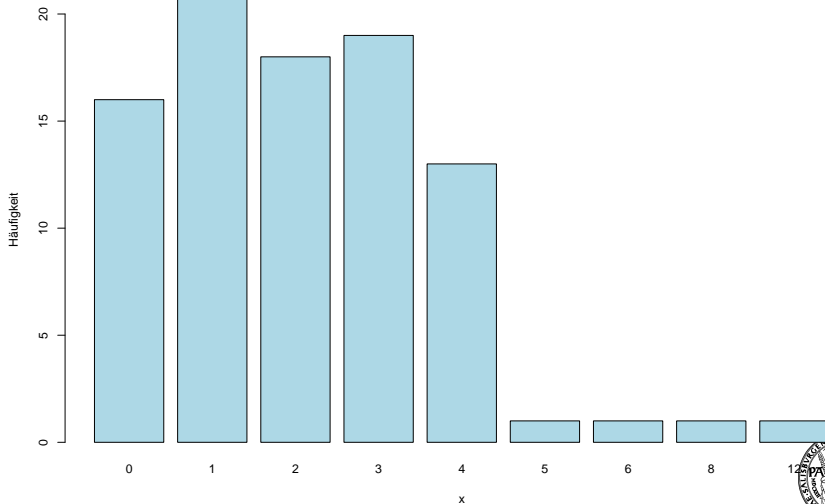
Anzahl jährliche Erdbeben Stärke ≥ 6 

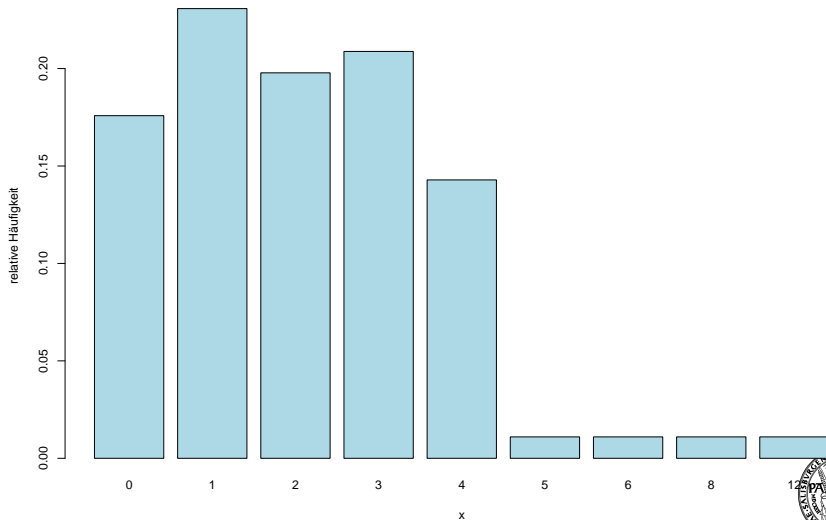




Parameterschätzung

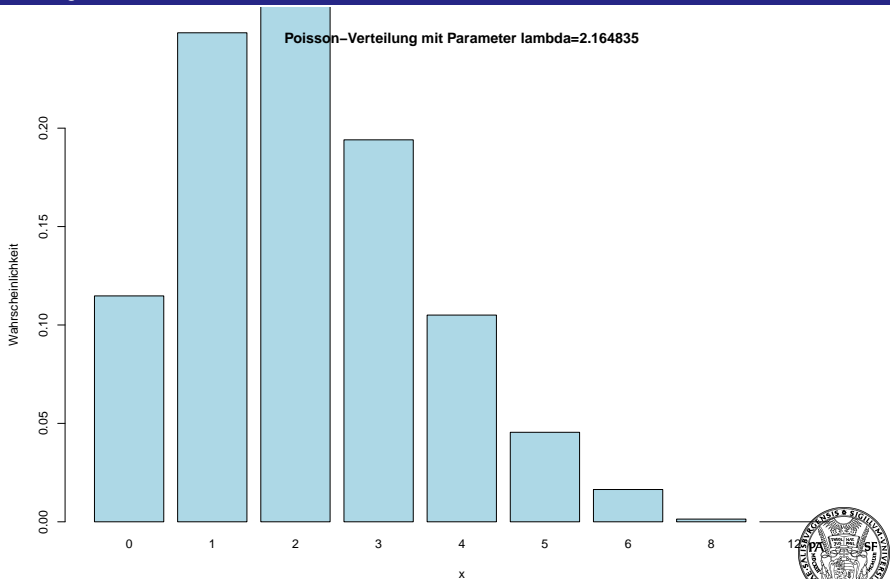


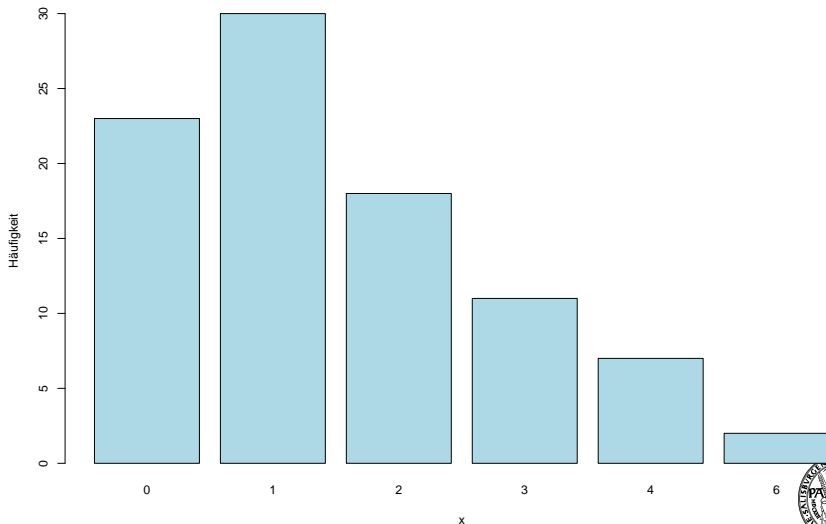
Anzahl jährliche Erdbeben Stärke ≥ 7 

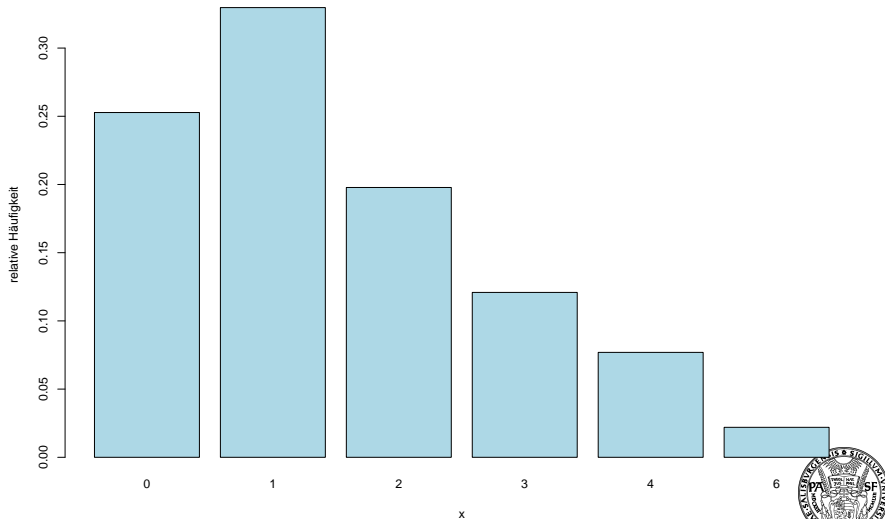
Anzahl jährliche Erdbeben Stärke ≥ 7 

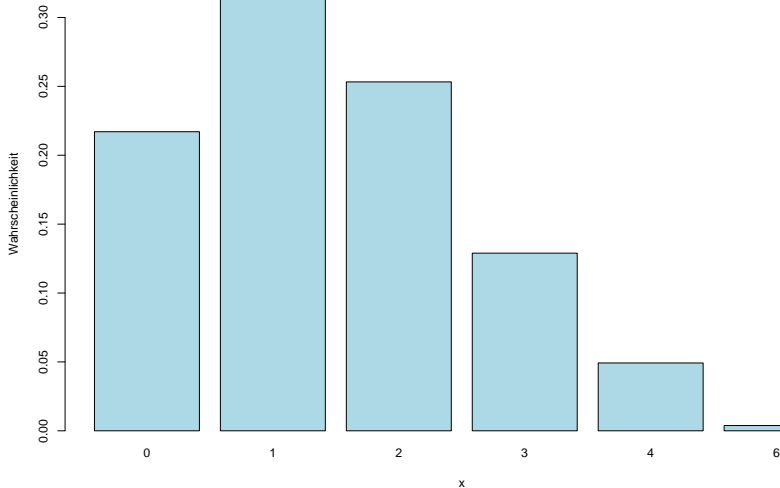


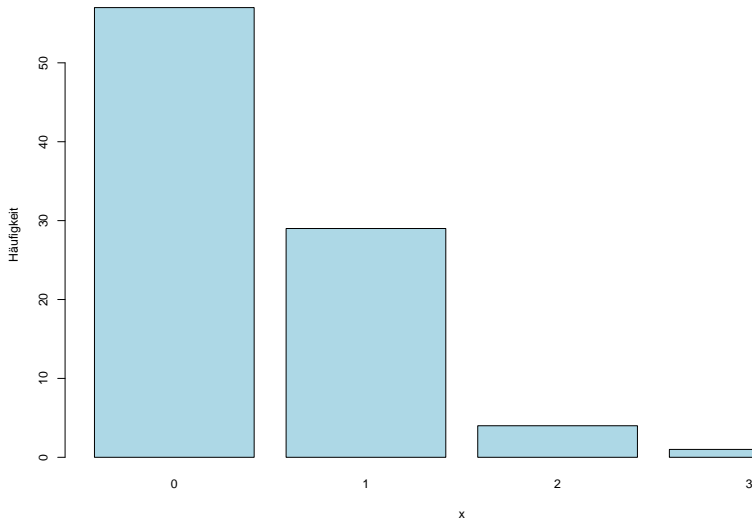
Parameterschätzung

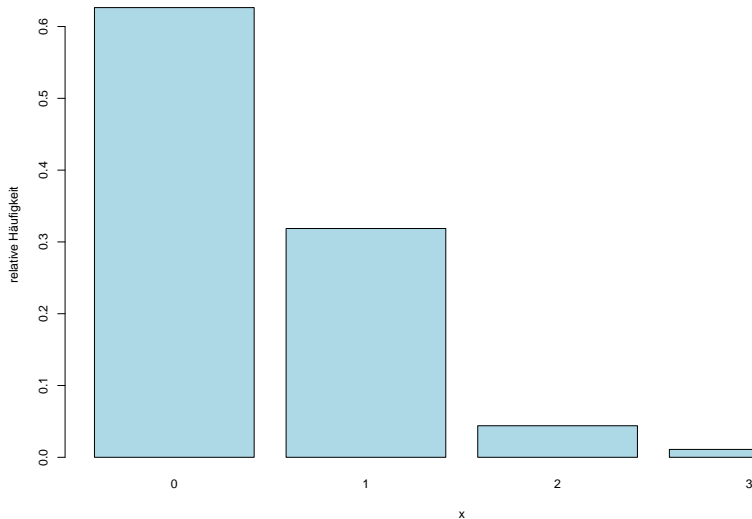


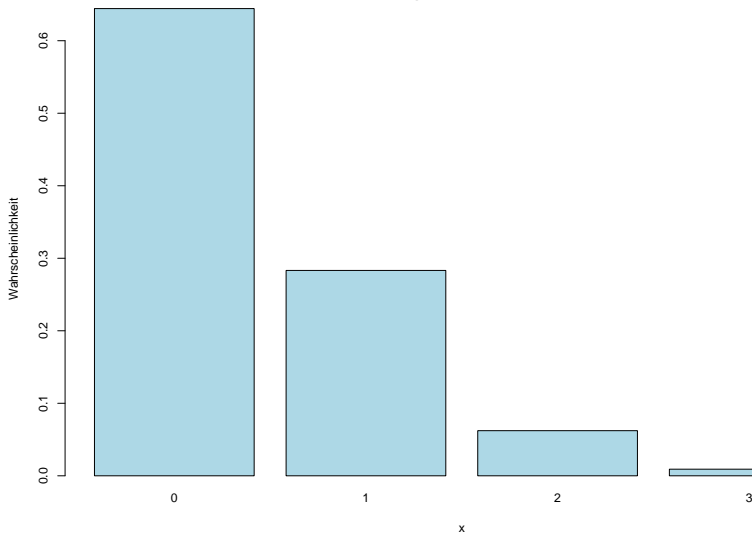
Anzahl jährliche Erdbeben Stärke ≥ 7.3 

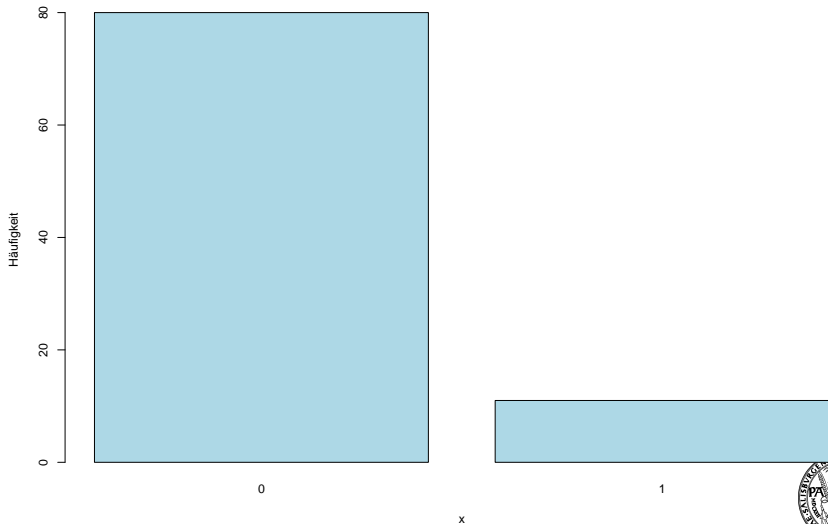
Anzahl jährliche Erdbeben Stärke ≥ 7.3 

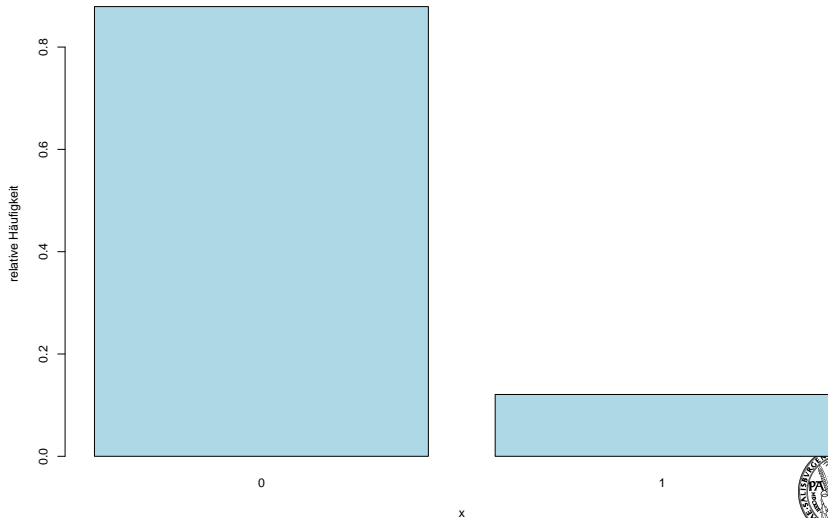
Poisson-Verteilung mit Parameter $\lambda=1.527473$ 

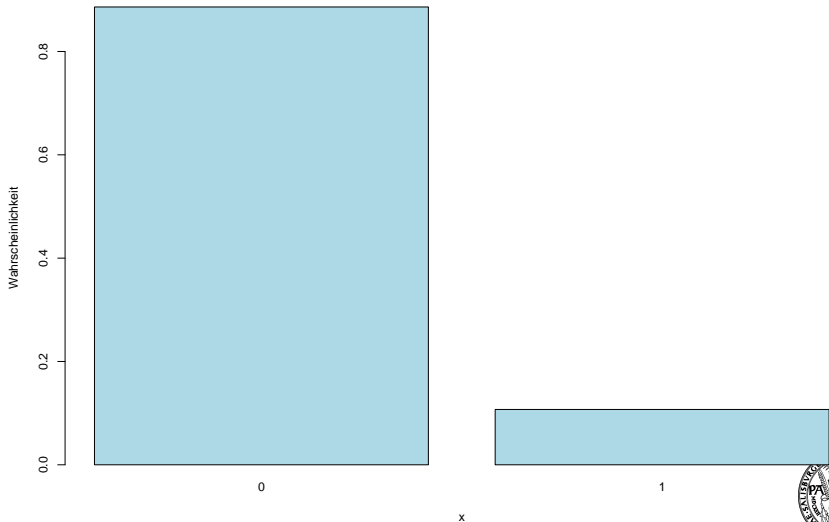
Anzahl jährliche Erdbeben Stärke ≥ 8 

Anzahl jährliche Erdbeben Stärke ≥ 8 

Poisson-Verteilung mit Parameter $\lambda=0.43956$ 

Anzahl jährliche Erdbeben Stärke ≥ 8.5 

Anzahl jährliche Erdbeben Stärke ≥ 8.5 

Poisson-Verteilung mit Parameter $\lambda=0.120879$ 

- ▶ Angenommen, basierend auf der Stichprobe x_1, \dots, x_n wurde der Parameter λ geschätzt und die Poissonverteilung mit Parameter λ passt gut (Grafischer Vergleich, etc.)
- ▶ **Preisfrage: Was haben wir dadurch gewonnen ?**
- ▶ Wir können die Wahrscheinlichkeit auch für bisher nicht beobachtete Ereignisse schätzen
- ▶ NB: Für nicht beobachtete Ereignisse wäre die naheliegende Schätzung sonst naheliegenderweise 0!
- ▶ Bsp: mehr als ein Erdbeben der Stärke ≥ 8.5 :

$$\mathbb{P}(X \geq 2) = 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1)) \approx 0.0067$$

