Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

09. Dezember 2011

Laplace-Verteilung

Die Laplace-Verteilung ist diejenige Verteilung, bei der alle Elementarereignisse dieselbe Wahrscheinlichkeit aufweisen.
Wir bezeichnen mit $|A|$ die Anzahl der Elemente von A. Dann

$$P(A) = \frac{|A|}{|\Omega|}$$

Beispiel: Der Würf einer fairen Münze realisiert die Laplace-Verteilung auf dem zweielementigen Ereignisraum $\Omega = \{A, Z\}$, wobei A=Adler und Z=Zahl

Wurf zweier fairer Würfel

Der Wurf zweier fairer Würfel realisiert die Laplace-Verteilung auf dem Ereignisraum $\Omega = \{1, \ldots, 6\}^2$.

$$P(\text{“Sechserpasch”}) = \frac{1}{36}$$

$$P(\text{“eine 3 und eine 4”}) = \frac{2}{36} = \frac{1}{18}$$

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\end{aligned}$$

Einfaches Beispiel

Ein fairer Würfel wird dreimal geworfen. Mit welcher Wahrscheinlichkeit fällt dreimal die 6?

$$|\Omega| = 6^3 = 216$$
$$A = \text{“drei Sechsen”} = \{(6,6,6)\}$$

Also

$$|A| = 1$$

und daher

$$P(A) = \frac{1}{216} = 0.004630$$

Trick: Übergang zum Komplementärereignis

Ein fairer Würfel wird dreimal geworfen. Mit welcher Wahrscheinlichkeit fällt mindestens eine 6?

$$A = \text{“mindestens eine Sechs”}$$
$$A^c = \text{“keine Sechs”} = \{1,2,3,4,5\}^3$$

Also

$$|A^c| = 5^3 = 125$$
$$P(A^c) = \frac{125}{216} = 0.5787$$

Schließlich

$$P(A) = 1 - P(A^c) = 1 - 0.5787 = 0.4213$$

Diversitätsindex nach Simpson

Der *Diversitätsindex* nach Simpson ist die Wahrscheinlichkeit, dass zwei aus einer Artengemeinschaft zufällig ausgewählte Individuen derselben Art angehören. Je kleiner er ist, umso größer ist die Diversität.

Wir berechnen ihn für den Fall zweier Arten S_1 und S_2 mit n_1 bzw. n_2 Individuen.

Der Ereignisraum Ω besteht aus allen Auswahlen von zwei verschiedenen Individuen aus insgesamt $n_1 + n_2$ Individuen. Da dasselbe Individuum nicht zweimal gewählt werden kann, gilt

$$|\Omega| = (n_1 + n_2) \cdot (n_1 + n_2 - 1)$$

Diversitätsindex, Fortsetzung

Das Ereignis, dessen Wahrscheinlichkeit gesucht ist, ist

$$E = A \cup B$$

wobei

$$A = \text{“beide gehören zu } S_1\text{”}$$
$$B = \text{“beide gehören zu } S_2\text{”}$$

Wie oben

$$|A| = n_1 \cdot (n_1 - 1) \qquad \text{und} \qquad |B| = n_2 \cdot (n_2 - 1)$$

Daher

$$P(A) = \frac{n_1 \cdot (n_1 - 1)}{(n_1 + n_2) \cdot (n_1 + n_2 - 1)}$$
$$P(B) = \frac{n_2 \cdot (n_2 - 1)}{(n_1 + n_2) \cdot (n_1 + n_2 - 1)}$$

Diversitätsindex für zwei Arten

A und B sind disjunkt. Daher $P(E) = P(A) + P(B)$. Der Diversitätsindex für zwei Arten beträgt

$$P(E) = \frac{n_1 \cdot (n_1 - 1) + n_2 \cdot (n_2 - 1)}{(n_1 + n_2) \cdot (n_1 + n_2 - 1)}$$

Für große Werte von n_1 und n_2

$$P(E) \cong \frac{n_1^2 + n_2^2}{(n_1 + n_2)^2}$$

Bei großen Populationen mehrerer, gleich häufiger Arten ist der Kehrwert des Diversitätsindex gleich der Anzahl der Arten

Beispiel zum Diversitätsindex

Für ein Waldgebiet wird die Mäusepopulation wie folgt geschätzt

- 500 Rötelmäuse
- 150 Feldmäuse

Der Diversitätsindex ist 0.6444

Stochastische Unabhängigkeit

- Zwei Ereignisse A und B heißen *(stochastisch) unabhängig*, wenn
$$P(A \cap B) = P(A) \cdot P(B)$$
- In konkreten Experimenten wird die Unabhängigkeit durch den Versuchsaufbau gewährleistet

Beispiel für abhängige Ereignisse

Zweifacher Wurf eines fairen Würfels

$$A = \text{“Erster Wurf eine 3”}$$
$$B = \text{“Augensumme 8”} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$
$$A \cap B = \{(3,5)\}$$

Folgende Wahrscheinlichkeiten

$$P(A) = \frac{1}{6} \qquad P(B) = \frac{5}{36} \qquad P(A \cap B) = \frac{1}{36}$$

Aber

$$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{5}{36} = 0.02315 \neq 0.02778 = \frac{1}{36}$$

Also sind die beiden Ereignisse stochastisch abhängig.

Erfolglose unabhängige Versuchswiederholungen

- Derselbe Versuch wird unabhängig n-mal wiederholt
- Jeder einzelne Versuch gelingt mit Wahrscheinlichkeit p
- Mit welcher Wahrscheinlichkeit misslingen alle n Versuche?
- Ein einzelner Versuch misslingt mit Wahrscheinlichkeit $1 - p$
- Zwei unabhängige Versuche misslingen mit Wahrscheinlichkeit $(1 - p)^2$ gemeinsam
- n unabhängige Versuche misslingen mit Wahrscheinlichkeit $(1 - p)^n$ gemeinsam

Beispiel: Mindestens ein Erfolg

- 1 000 Versuche mit Erfolgswahrscheinlichkeit $p = 0.003$
- Mit welcher Wahrscheinlichkeit gibt es mindestens einen Erfolg?
- Übergang zum Komplement: Mit welcher Wahrscheinlichkeit gibt es keinen Erfolg?
- Misserfolgswahrscheinlichkeit im Einzelfall $1 - p = 0.997$
- Wahrscheinlichkeit von 1 000 unabhängigen Misserfolgen:

$$(1 - p)^{1\,000} = 0.997^{1\,000} = 0.04956$$

- Wahrscheinlichkeit mindestens eines Erfolges

$$1 - 0.04956 = 0.95044$$

Zufallsworte

- Gegeben ist ein Alphabet
- Jeder Buchstabe das Alphabets besitzt eine Wahrscheinlichkeit
- Zufallsworte entstehen dadurch, dass Buchstaben unabhängig gezogen und aneinander gereiht werden
- Die Wahrscheinlichkeit eines Worts ist gleich dem Produkt der Wahrscheinlichkeiten seiner Buchstaben

Das genetische Alphabet

- Vier verschiedene Buchstaben mit unterschiedlicher Häufigkeit

Base	Häufigkeit
a	40%
c	22%
g	17%
t	21%

- Beispiele:
  - $P(acac) = 0.4 \cdot 0.22 \cdot 0.4 \cdot 0.22 = 0.007744$
  - $P(gtgt) = 0.17 \cdot 0.21 \cdot 0.17 \cdot 0.21 = 0.001275$

cg

- Problem: In echter DNA sind aufeinander folgende Buchstaben nicht unabhängig
- Untersuche Kombinationen zweier aufeinander folgender Basen
- In Organismen tritt die Kombinationen *cg* nur in 1% aller Fälle auf
- In einem Zufallswort wäre die Häufigkeit

$$0.22 \cdot 0.17 = 0.0374$$

- Bei den anderen Dinukleotiden ist der Unterschied weniger ausgeprägt.

DinoDNA aus Crichton's Jurassic Park

```
gcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgc
ggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcg
tgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggc
tgctcacgctgtacctatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtg
ccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaa
agtaggacaggtgccggcagcgctctgggtcattttcggcgaggaccgctttcgctggag
atcggcctgtcgcttgcggtattcggaatcttgcacgccctcgctcaagccttcgtcact
ccaaacgtttcggcgagaagcaggccattatcgccggcatggcggccgacgcgctgggct
ggcgttcgcgacgcgaggctggatggccttccccattatgattcttctcgcttccggcgg
cccgcgttgcaggccatgctgtccaggcaggtagatgacgaccatcagggacagcttcaa
cggctcttaccagcctaacttcgatcactggaccgctgatcgtcacggcgatttatgccg
cacatggacgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaa
caagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaa
gcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcggg
ctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctg
acgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtcca
acacgacttaacgggttggcatggattgtaggcgccgccctataccttgtctgcctcccc
gcggtgcatggagccgggccacctcgacctgaatggaagccggcggcacctcgctaacgg
ccaagaattggagccaatcaattcttgcggagaactgtgaatgcgcaaaccaacccttgg
ccatcgcgtccgccatctccagcagccgcacgcggcgcatctcgggcagcgttgggtcct
```

DinoDNA aus Crichton's Jurassic Park

cg cg cg cg cg cgc
g cg cg cg cg
cg cg cg cg cg cg
cg cg cg cg
cg cg cg cg cg cg cg
cg cg cg cg cg cg
cg cg cg cg cg cg cg
cg cg cg cg cg cg cg cgcg
cg cgcg cgcg cg cg cg
cgcg cg
cg cg cg cg cg cg cg
cgcg cg cg cg
cg cg cg
cg cg cg cg cg cg
cg cg cg cg
cg cg cg cg cg cg cg
cg cg cg cg
cg cg cg cg cg cg cg
cg cg
cgcg cg cg cgcg cg cg cg

Auswertung der DinoDNA

- Anteil c: 33%
- Anteil g: 27%
- Anteil cg: 9.6%
- $0.33 \cdot 0.27 = 0.089 = 8.9\%$
- Anteil cg in Organismen: 1%

Die Sequenz aus dem Buch von Crichton ist vermutlich ein Zufallswort.

Fakultät, Beispiel

- In der Lottotrommel sind 49 Kugeln. **Alle** Kugeln werden gezogen, die Reihenfolge wird notiert. Wie viele Möglichkeiten gibt es?
- Es gibt 49 Möglichkeiten für die erste Kugel, 48 für die zweite, 47 für die dritte, ..., 2 für die 48-te (also die vorletzte), und 1 für die letzte
- Anzahl der Möglichkeiten:

$$\begin{aligned} & 49 \cdot 48 \cdot 47 \cdots 3 \cdot 2 \cdot 1 \\ &= 608281864034267560872252163321295376887552831379210240000000000 \\ &= 6.083 \cdot 10^{62} \end{aligned}$$

- Diese Zahl schreibt man 49!

Fakultät

- Die Zahl
$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$
bezeichnet man als *Fakultät* von n
- Sie gibt die Anzahl der Möglichkeiten an, n verschiedene Objekte anzuordnen
- Jede solche Anordnung bezeichnet man als *Permutation*
- Beispiele
$$1! = 1 \qquad 2! = 2 \qquad 3! = 6 \qquad 4! = 24 \qquad 5! = 120$$
- Außerdem definiert man $0! = 1$

Zahlenbeispiele

- $6! = 720$
- $12! = 479\,001\,600$
- $22! = 1.124 \cdot 10^{21}$
- $69! = 1.711 \cdot 10^{98}$
- $70! = 1.198 \cdot 10^{100}$

Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge

- Aus der Lottotrommel werden 6 Kugeln gezogen
- Anzahl der Möglichkeiten **unter** Beachtung der Reihenfolge

$$49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = 10\,068\,347\,520$$

- Taschenrechner Taste $\boxed{\text{nPr}}$

Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge

- Aus der Lottotrommel werden 6 Kugeln gezogen
- Anzahl der Möglichkeiten **ohne** Beachtung der Reihenfolge

$$\frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13\,983\,816$$

- Diese Zahl ist gleich

$$\frac{49!}{6! \cdot 43!}$$

- Sie heißt "49 über 6" und man schreibt sie $\binom{49}{6}$
- Taschenrechner: Taste $\boxed{\text{nCr}}$

Binomialkoeffizienten

n bezeichne die Gesamtzahl der Objekte, und k bezeichne die Anzahl der Züge.

- $$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+2) \cdot (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1}$$

  ist die Anzahl der möglichen Auswahlen von k Objekten aus n-Objekten.
- Die Zahl $\binom{n}{k}$ heißt *Binomialkoeffizient.* Man sagt "n über k".

weitere Beispiele für Binomialkoeffizienten

- Möglichkeiten, 2 Elementen aus 4 auszuwählen

$$\binom{4}{2} = \frac{4 \cdot 3}{2 \cdot 1} = 6$$

- Möglichkeiten, 3 Elementen aus 10 auszuwählen

$$\binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$$

- Möglichkeiten, 7 Elementen aus 10 auszuwählen

$$\binom{10}{7} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 120$$

Rechenregeln

Für jedes n und jedes $k \leq n$ gelten

- $$\binom{n}{0} = 1$$
- $$\binom{n}{1} = n$$
- $$\binom{n}{k} = \binom{n}{n-k}$$
- $$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

Pascalsches Dreieck

Das Pascalsche Dreieck zeigt die Binomialkoeffizienten, beginnend mit $\binom{0}{0}$

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

Urnenmodelle

Jeweils n Kugeln in der Urne, davon werden k gezogen

Urnenmodell I: Ziehen mit Zurücklegen unter Beachtung der Reihenfolge

$$\text{Anzahl} = n^k$$

Urnenmodell II: Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge

$$\text{Anzahl} = n \cdot (n-1) \cdots (n-k+1)$$

Urnenmodell III: Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge

$$\text{Anzahl} = \binom{n}{k}$$

Beispiel für Urnenmodell I

Frage: Wieviele Kombinationen hat ein dreistelliges Zahlenschloss?

Antwort:

- 3 Züge aus Urne mit 10 Elementen, nämlich den 10 Ziffern
- Ziffern können mehrfach auftreten: Ziehen mit Zurücklegen
- Reihenfolge spielt eine Rolle
- also $10^3 = 1000$ Möglichkeiten

Beispiel für Urnenmodell II

Frage: Bei einem Rennen starten 36 Läufer. Wie viele Kombinationen, das Siegertreppchen zu besetzen, gibt es?

Antwort:

- 3 Züge aus Urne mit 36 Elementen
- jeder Läufer gewinnt maximal eine Medaille: Ziehen ohne Zurücklegen
- Reihenfolge spielt eine Rolle
- also
$$36 \cdot 35 \cdot 34 = 42\,840$$
Möglichkeiten

Beispiel für Urnenmodell III

Frage: Wie viele Möglichkeiten gibt es, drei Mutationen auf 150 Basenpaare zu verteilen?

Antwort:

- 3 Züge aus Urne mit 150 Elementen
- Stelle ist entweder mutiert oder nicht: Ziehen ohne Zurücklegen
- Reihenfolge spielt keine Rolle
- also
$$\binom{150}{3} = 551\,300$$
Möglichkeiten

Beispiel: Lotterie

Eine Lotterie wird nach den folgenden Regeln abgehalten:

In einem Mischbehälter befinden sich 70 Kugeln, nämlich je 7 mit der Ziffern 0 bis 9. Nacheinander werden 7 Kugeln entnommen und in einer Reihe so ausgelegt, dass sie eine 7-stellige Zahl bilden. Wer diese Zahl auf seinem Los hat, erhält den Hauptgewinn.

Wie groß sind die Gewinnchancen der Lose 1234567 und 9999999?

Beispiel: Lotterie

Ziehen ohne Zurücklegen: Ereignisraum

$$\Omega = \{(a_1, \ldots, a_7) \mid 1 \leq a_i \leq 70,\ a_i \neq a_j \text{ für } i \neq j\}$$

mit der Interpretation, dass beispielsweise die Kugeln $a_{05}, a_{15}, \ldots, a_{65}$ jeweils eine 5 zeigen.

$$|\Omega| = \frac{70!}{63!} = 70 \cdot 69 \cdot 68 \cdots 64 = 6\,041\,824\,588\,800$$

Es sei A das Ereignis, dass 1234567 gewinnt. $|A| = 7^7$.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{7^7}{70 \cdot 69 \cdot 68 \cdots 64} \cong 0.000\,000\,136307$$

Beispiel: Lotterie

Es sei B das Ereignis, dass 9999999 gewinnt. $|B| = 7!$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{7!}{70 \cdot 69 \cdot 68 \cdots 64} \cong 0.000\,000\,000\,834$$

Vergleich:

$$P(A) = 0.000\,000\,136\,307 \cong 1.36 \cdot 10^{-7}$$
$$P(B) \cong 0.000\,000\,000\,834 \cong 8.34 \cdot 10^{-10}$$

Binomialverteilung

- n unabhängige Wiederholungen eines ja/nein-Experiments
- das Ergebnis "ja" bezeichnet man als "Erfolg"
- die Wahrscheinlichkeit von "ja" sei p, genannt "Erfolgswahrscheinlichkeit im Einzelfall"
- $B_{n,p}(k)$ ist die Wahrscheinlichkeit von genau k Erfolgen

$$B_{n,p}(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

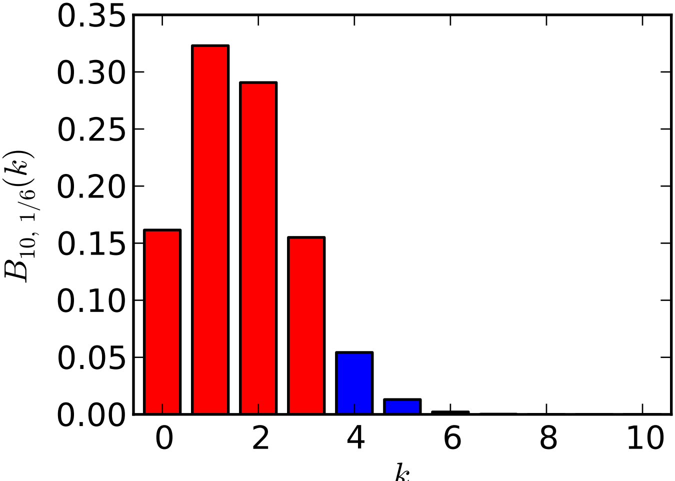
- Beispiel:

$$B_{10,\,1/6}(2) = \binom{10}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^8 = 45 \cdot 0.02778 \cdot 0.2326 = 0.2907$$

- $B_{n,p}(n) = p^n$ und $B_{n,p}(0) = (1-p)^n$

Stabdiagramm von $B_{10,\,1/6}$, Fortsetzung

Rote Fläche ist die Antwort auf die Frage:
Mit welcher Wahrscheinlichkeit fallen beim 10-fachen Wurf eines fairen Würfels nicht mehr als 3 Sechsen?



Kumulierte Binomialverteilung

Mit welcher Wahrscheinlichkeit fallen beim 10-fachen Wurf eines fairen Würfels nicht mehr als 3 Sechsen?
Antwort:

$$P = \sum_{k=0}^{3} B_{10,\,1/6}(k) = 0.93027$$

Für solche Fragen gibt es Tabellen der kumulierten Binomialverteilung

$$\sum_{k=0}^{r} B_{n,\,p}(k)$$

in Abhängigkeit von r

Tabelle der kumulierten $B_{10,\,p}$

$$\text{Tabelle der Werte } \sum_{k=0}^{r} B_{n,\,p} \text{ für } n = 10$$

r	p	0.15	0.16	$\frac{1}{6}$	0.17	0.18	0.19
0	0.	19687	17490	16151	15516	13745	12158
1		54430	50805	48452	47296	43916	40676
2		82020	79360	77523	76587	73720	70778
3		95003	93864	93027	92585	91166	89607
4		99013	98699	98454	98320	97868	97337
5		99862	99804	99756	99729	99633	99512
6		99987	99979	99973	99970	99956	99938
7		99999	99999	99998	99998	99996	99995

Lesehinweise für kumulierte Tabellen

- $\sum_{k=0}^{3} B_{10,\,0.18}(k) = 0.91166$
- freie Felder oberhalb der Tabelle sind 0 im Rahmen der Tabellengenauigkeit
- freie Felder unterhalb der Tabelle sind 1 im Rahmen der Tabellengenauigkeit
- Tabellen erhalten Sie von mir

Beispiel Parasiten

- Bestimmte Fische erkranken mit 85% Wahrscheinlichkeit an einem Parasiten
- 47 Fische werden untersucht. Mit welcher Wahrscheinlichkeit sind höchstens 40 davon erkrankt?
- Gesucht

$$\sum_{k=0}^{40} B_{47,\,0.85}(k)$$

Graph von $B_{47,\,0.85}$

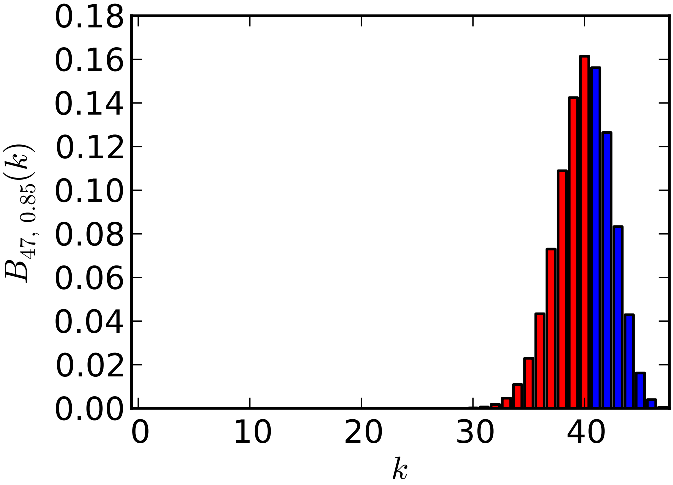


Tabelle der Werte $\sum_{k=0}^{r} B_{n,p}(k)$ für $n = 47$

r	p	0.85	0.86	0.87	0.88	0.89
27	0.	00001				
28		00002	00001			
29		00008	00003	00001		
30		00029	00012	00005	00002	00001
31		00093	00043	00018	00007	00002
32		00274	00137	00063	00026	00010
33		00742	00398	00199	00091	00038
34		01832	01060	00573	00286	00130
35		04128	02571	01503	00817	00408
36		08463	05663	03578	02115	01156
37		15768	11311	07707	04946	02957
38		26660	20441	14978	10408	06792
39		40904	33384	26208	19651	13952
40		57047	49285	41238	33208	25538
41		72665	65962	58411	50182	41543
42		85309	80597	74830	67964	60042
43		93639	91050	87606	83128	77447
44		97931	96887	95379	93236	90248
45		99552	99278	98847	98178	97153
46		99952	99917	99856	99754	99582

Beispiel Pharmapräparat, Fortsetzung

Die Wahrscheinlichkeit, dass höchstens 40 Fische erkrankt sind, ist gleich 0.57047