Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun
http://blog.ruediger-braun.net

Heinrich-Heine-Universität Düsseldorf

19. Dezember 2014

Erwartungswert

Der Erwartungswert ist derjenige Wert, den man im Mittel beobachten würde, wenn man das Experiment sehr oft wiederholt.

Bei einer Lotterie ist der Erwartungswert der Betrag, bei dem die Lotterie fair wäre, bei dem also weder der Spieler noch der Betreiber langfristig Geld verdienen würde.

Spiel 77

Klasse	Ziffern	Gewinn k	$P(X = k)$	$k \cdot P(X = k)$
I	7	177 777.00€	0.000 000 1	0.018€
II	6	77 777.00€	0.000 001 0	0.078€
III	5	7 777.00€	0.000 010 0	0.078€
IV	4	777.00€	0.000 100 0	0.078€
V	3	77.00€	0.001 000 0	0.077€
VI	2	17.00€	0.010 000 0	0.170€
VII	1	5.00€	0.100 000 0	0.500€

$$E(X) = 0.998€$$

Erwartungswert

- X eine diskrete Zufallsvariable
- Der *Erwartungswert* von X ist

$$E(X) = \sum_k P(X = k) \cdot k$$

- Die Summe läuft über alle Werte k von X

Beispiel: Würfel

Sei X die Augenzahl eines fairen Würfels.

$$\begin{aligned} E(X) &= \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 \\ &= \frac{21}{6} = \frac{7}{2} = 3.5 \end{aligned}$$

Im Mittel zeigt ein fairer Würfel 3.5 Augen

Variante des Spiels 77

- Der Einsatz beträgt 2.50€. Bei welchem Hauptgewinn wäre das Spiel fair?
- Hauptgewinn sei J. Dann

$$\begin{aligned} E(X) &= 0.000\,000\,1 \cdot J + 0.000\,001 \cdot 77\,7777 \\ &+ 0.000\,01 \cdot 7\,777 + 0.000\,1 \cdot 777 \\ &+ 0.001 \cdot 77 + 0.01 \cdot 17 + 0.1 \cdot 5 \\ &= 0.000\,000\,1 \cdot J + 0.980 \end{aligned}$$

- Das soll gleich 2.50 sein. Also $0.000\,000\,1 \cdot J = 1.52$

$$J = \frac{1.52}{0.000\,000\,1} = 15\,200\,000$$

- Spiel 77 ist fair bei einem Hauptgewinn von 15.2 Millionen €

Varianz und Streuung

Die *Varianz* einer Zufallsvariablen X ist definiert als

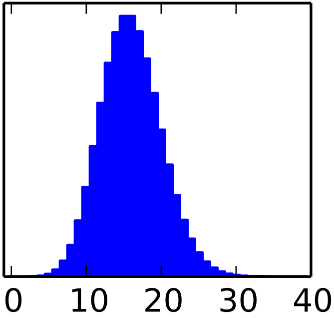
$$Var(X) = \sum_{k=0}^{\infty} (k - \mu)^2 \cdot P(X = k)$$

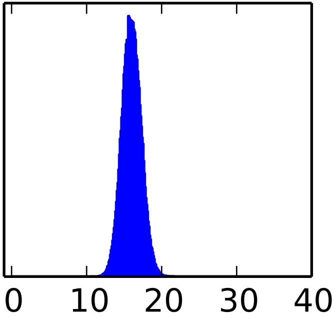
wobei $\mu = E(X)$.

Die *Standardabweichung* oder *Streuung* von X ist definiert als die Wurzel aus der Varianz

$$\sigma = \sqrt{Var(X)}$$

Gleicher Erwartungswert, unterschiedliche Streuung





Links: Verteilung X_1 mit $E(X_1) = 16$ und $\sigma = 4$

Rechts: Verteilung X_2 mit $E(X_2) = 16$ und $\sigma = 1.26$

Modell vs. Datensatz

Datensatz	Modell
arithmetisches Mittel	Erwartungswert
empirische Varianz	Varianz
Stichprobenstreuung	Streuung

Rechenregeln

Rechenregeln für den Erwartungswert

- Für jede Zahl c und jede Zufallsvariable X ist $E(c \cdot X) = c \cdot E(X)$
- Für Zufallsvariablen $X_1, \ldots, X_n$ ist $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

Rechenregeln für die Varianz

- Für jede Zahl a und jede Zufallsvariable X gilt $Var(a + X) = Var(X)$
- Für Zahl c und jede Zufallsvariable X gilt $Var(c \cdot X) = c^2 \cdot Var(X)$
- Für jede Zufallsvariable X gilt $Var(X) = E(X^2) - E(X)^2$

Unabhängigkeit von Zufallsvariablen

Zwei diskrete Zufallsvariable X und Y sind *stochastisch unabhängig*, wenn für alle möglichen Werte k und m

$$P(X = k, Y = m) = P(X = k) \cdot P(Y = m)$$

Die Unabhängigkeit muss durch die Versuchsplanung gesichert werden

Zusätzliche Rechenregeln für unabhängige Zufallsvariable

Produktformel für den Erwartungswert: X und Y seien **unabhängige** Zufallsvariable. Dann

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Summenformel für die Varianz: X und Y seien **unabhängige** Zufallsvariable. Dann

$$Var(X + Y) = Var(X) + Var(Y)$$

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:

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

- 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 Elemente aus 4 auszuwählen

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

- Möglichkeiten, 3 Elemente aus 10 auszuwählen

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

- Möglichkeiten, 7 Elemente 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

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

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: fairer Würfel

- Erfolg: Wurf einer 6
- Erfolgswahrscheinlichkeit im Einzelfall: $p = \frac{1}{6}$
- Misserfolg: Wurf 1,2,3,4,5
- Misserfolgswahrscheinlichkeit im Einzelfall: $q = 1 - p = \frac{5}{6}$
- Gesucht: Wahrscheinlichkeit von

$$A = \text{“genau 2 Erfolge bei 5 Würfen”}$$

- Antwort $B_{5,\,1/6}(2) = \binom{5}{2} p^2 (1-p)^3 = 10 \cdot \frac{1}{36} \cdot \frac{125}{216} = 0.1608$
- Wie kommt das zustande?

Binomialverteilung: Beispiel

e: Erfolg, m: Misserfolg, $q = 1 - p$

$$
\begin{aligned}
P(eemmm) &= p \cdot p \cdot q \cdot q \cdot q = p^2 \cdot q^3 \\
P(ememm) &= p \cdot q \cdot p \cdot q \cdot q = p^2 \cdot q^3 \\
P(emmem) &= p \cdot q \cdot q \cdot p \cdot q = p^2 \cdot q^3 \\
P(emmme) &= p \cdot q \cdot q \cdot q \cdot p = p^2 \cdot q^3 \\
P(meemm) &= q \cdot p \cdot p \cdot q \cdot q = p^2 \cdot q^3 \\
P(memem) &= q \cdot p \cdot q \cdot p \cdot q = p^2 \cdot q^3 \\
P(memme) &= q \cdot p \cdot q \cdot q \cdot p = p^2 \cdot q^3 \\
P(mmeem) &= q \cdot q \cdot p \cdot p \cdot q = p^2 \cdot q^3 \\
P(mmeme) &= q \cdot q \cdot p \cdot q \cdot p = p^2 \cdot q^3 \\
P(mmmee) &= q \cdot q \cdot q \cdot p \cdot p = p^2 \cdot q^3
\end{aligned}
$$

$P(A)$ ist dann die Summe, $P(A) = 10 \cdot p^2 \cdot q^3$.

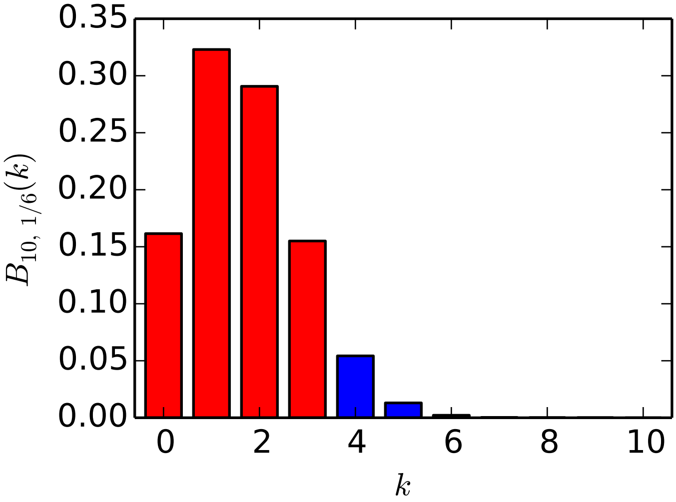
Antwort: Mit Wahrscheinlichkeit

$$B_{5,\,1/6}(0) + B_{5,\,1/6}(1) + B_{5,\,1/6}(2) = 0.9645$$

werden nicht mehr als 3 Sechsen beobachtet.
Jetzt dasselbe für $n = 10$

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

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:

$$\begin{aligned} P &= \sum_{k=0}^{3} B_{10,\,1/6}(k) \\ &= B_{10,\,1/6}(0) + B_{10,\,1/6}(1) + B_{10,\,1/6}(2) + B_{10,\,1/6}(3) \\ &= 0.16151 + 0.32301 + 0.29071 + 0.15504 \\ &= 0.93027 \end{aligned}$$

Für solche Fragen gibt es Tabellen der kumulierten Binomialverteilung

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

in Abhängigkeit von r

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

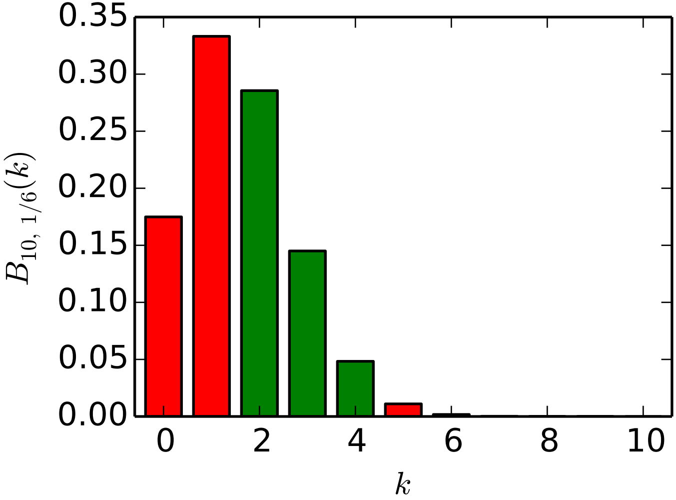
Tabelle der Werte $\sum_{k=0}^{r} B_{n,\,p}$ für $n = 10$

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

- $\sum_{k=0}^{3} B_{10,\,0.18}(k) = 0.91166$
- $\sum_{k=4}^{10} B_{10,\,0.17}(k) = 1 - \sum_{k=0}^{3} B_{10,\,0.17}(k) = 1 - 0.92585 = 0.07415$
- $\sum_{k=2}^{4} B_{10,\,0.16}(k) = \sum_{k=0}^{4} B_{10,\,0.16}(k) - \sum_{k=0}^{1} B_{10,\,0.16}(k) = 0.98699 - 0.50805 = 0.47894$
- 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

Skizze zum dritten Beispiel



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 $B_{47,\,0.85}$

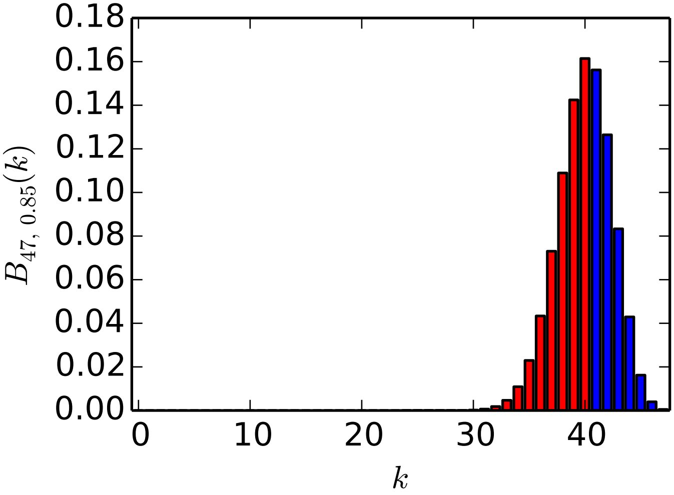


Tabelle der Werte $\sum_{k=0}^{r} B_{n,p}(k)$ für $n = 47$

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 Parasiten, Fortsetzung

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

Beispiel Pharmapräparat

- Beispiel: 47 Mäuse sind erkrankt
- Ein Präparat mit Heilungswahrscheinlichkeit 88% wird eingesetzt
- Mit welcher Wahrscheinlichkeit werden mindestens 40 Mäuse geheilt?
- Die Wahrscheinlichkeit, dass genau 40 Mäuse geheilt werden, wird gegeben durch die Binomialverteilung $B_{47,\,0.88}(40)$
- Die Wahrscheinlichkeit, dass mindestens 40 Mäuse geheilt werden, beträgt

$$\sum_{k=40}^{47} B_{47,\,0.88}(k) = 1 - \sum_{k=0}^{39} B_{47,\,0.88}(k)$$

Graph von $B_{47,\,0.88}$

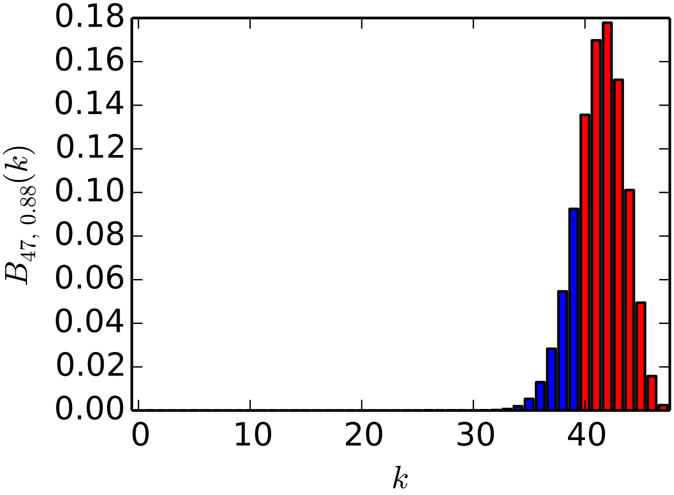


Tabelle der Werte $\sum_{k=0}^{r} B_{n,p}(k)$ für $n = 47$

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

Pharmapräparat, Fortsetzung

Die Wahrscheinlichkeit, dass mindestens 40 Mäuse geheilt werden, ist

$$1 - \sum_{k=0}^{39} B_{47,\,0.88}(k) = 1 - 0.19651 = 0.80349$$

Einzelnes ja/nein-Experiment

Zufallsvariable X nimmt mit Wahrscheinlichkeit p den Wert 1 und mit Wahrscheinlichkeit $1-p$ den Wert 0 an

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{1} k \cdot P(X=k) \\
&= 0 \cdot (1-p) + 1 \cdot p \\
&= p \\
Var(X) &= \sum_{k=0}^{1} (k - E(X))^2 \cdot P(X=k) \\
&= p^2 \cdot (1-p) + (1-p)^2 \cdot p \\
&= p \cdot (1-p) \cdot (p + (1-p)) \\
&= p \cdot (1-p)
\end{aligned}
$$

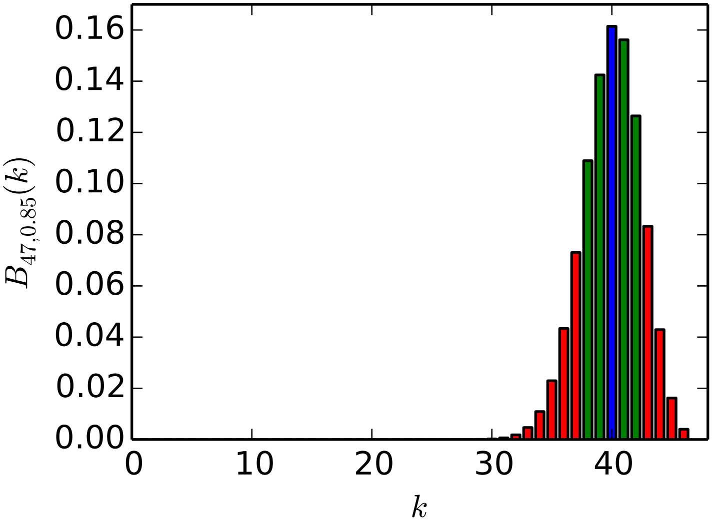
Erwartungswert der Binomialverteilung

- $X_1, X_2, \ldots, X_n$ seien Zufallsvariable
- jedes X_j nimmt mit Wahrscheinlichkeit p den Wert 1 und mit Wahrscheinlichkeit $1-p$ den Wert 0 an
- Interpretation $X_j = 1$ ist "Erfolg"
- $Y = X_1 + X_2 + \cdots + X_n$ ist die Anzahl der Erfolge
- Y ist verteilt gemäß $B_{n,p}$
- $E(X_1) = E(X_2) = \cdots = E(X_n) = p$
- also $E(Y) = n \cdot p$
- Der Erwartungswert einer $B_{n,p}$-verteilten Zufallsvariable beträgt $n \cdot p$

Varianz der Binomialverteilung

- $X_1, X_2, \dots, X_n$ seien unabhängige Zufallsvariable
- jedes X_j nimmt mit Wahrscheinlichkeit p den Wert 1 und mit Wahrscheinlichkeit $1-p$ den Wert 0 an
- $Y = X_1 + X_2 + \dots + X_n$ ist verteilt gemäß $B_{n,p}$
- $Var(X_1) = Var(X_2) = \dots = Var(X_n) = p \cdot (1-p)$
- also $Var(Y) = n \cdot p \cdot (1-p)$
- Die Varianz einer $B_{n,p}$-verteilten Zufallsvariable beträgt $n \cdot p \cdot (1-p)$

Stabdiagramm der Binomialverteilung



Der Erwartungswert ist blau, alles innerhalb der Standardabweichung ist grün