Binomialverteilung und Anwendungen

Dr. Elke Warmuth

Sommersemester 2018

Bernoulli-Experiment

Vorgänge mit zufälligem Ergebnis, bei denen nur zwischen Erfolg (1) und Misserfolg (0) unterschieden wird, heißen Bernoulli-Experimente oder Bernoulli-Versuche.

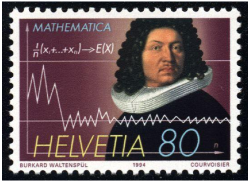
Bernoulli-Kette

Wird ein Bernoulli-Experiment mit derselben Erfolgswahrscheinlichkeit n mal **unabhängig** voneinander ausgeführt, so entsteht eine **Bernoulli-Kette** der **Länge** n mit der **Erfolgswahrscheinlichkeit** p.

Beispiele und Gegenbeispiele

- 10-mal Würfeln, Erfolg – 6
- 5-mal Ziehen mit Zurücklegen, Erfolg – rote Kugel
- Tagesmitteltemperatur an aufeinanderfolgenden Tagen im Juli in Berlin, Erfolg – Tagesmitteltemperatur über 18,5° Celsius
- Multiple-Choice-Test, Erfolg – richtige Antwort
- Spielschein 13er-Wette, Erfolg – richtiger Tipp
- kleine Stichproben ohne Zurücklegen aus großen Grundgesamtheiten, Erfolg – Merkmal liegt vor
- 1000 aufeinanderfolgende Buchstaben eines deutschen Textes auswerten, Erfolg – Vokal
- Elfmeterschüsse beim Elfmeterschießen am Ende eines Fußballspiels, Erfolg – Tor

Jakob Bernoulli (1654-1705): Ars conjectandi (1713)



Schweizer Briefmarke 1994
Quelle: www.fh-friedberg.de/.../marke04_09_bild01.jpg"

Bernoulli-Ketten u.a. als Rahmen für den Beweis des Gesetzes der großen Zahlen

`Diese Entdeckung gilt mir mehr, als wenn ich gar die Quadratur des Kreises geliefert hätte; denn wenn diese auch gänzlich gefunden würde, so wäre sie doch sehr wenig nütz.`

Quelle: F. Barth, R. Haller: Stochastik Leistungskurs. München: Ehrenwirth, 1985.

- BERNOULLI-Experiment: $\Omega = \{0, 1\}, P(1) = p,\ 1 \doteq$ Erfolg
- BERNOULLI-Kette für $n = 4$ und p beliebig aus $[0; 1]$
  - Ergebnismenge: $\Omega = \{w = (w_1, w_2, w_3, w_4) : w_i \in \{0, 1\}\}$
  - $P(\text{nur Erfolge}) = P((1, 1, 1, 1)) = p \cdot p \cdot \ldots \cdot p = p^4$

    (Unabhängigkeit)
  - $P(\text{nur Misserfolge}) = P((0, 0, 0, 0))$

    $= (1 - p) \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) = (1 - p)^4$
  - Einzelwahrscheinlichkeiten allgemein:

    $P((w_1, w_2, w_3, w_4)) = p^{\text{Anzahl der Einsen}} \cdot (1-p)^{\text{Anzahl der Nullen}}$
  - $P(\text{genau ein Erfolg}) = 4 \cdot p \cdot (1 - p)^3$
  - $P(\text{genau } k \text{ Erfolge}) = \binom{4}{k} p^k (1 - p)^{4-k}, k = 0, 1, 2, 3, 4$

- BERNOULLI-Kette für beliebige n und p:
  - Ergebnismenge: $\Omega = \{w = (w_1, \dots, w_n) : w_i \in \{0, 1\}\}$
  - Einzelwahrscheinlichkeiten:

    $P((w_1, \dots, w_n)) = p^{\text{Anzahl der Einsen}} \cdot (1-p)^{\text{Anzahl der Nullen}}$

    (Unabhängigkeit)
  - $P(\text{nur Erfolge}) = P((1, 1, \dots, 1)) = p \cdot p \cdot \dots \cdot p = p^n$

    $P(\text{nur Misserfolge}) = P((0, 0, \dots, 0)) =$
    $(1-p) \cdot (1-p) \cdot \dots \cdot (1-p) = (1-p)^n$

    $P(\text{genau ein Erfolg}) = n \cdot p \cdot (1-p)^{n-1}$

    $P(\text{genau } k \text{ Erfolge}) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \dots, n$

    (BERNOULLI-Formel)

realer Vorgang als BERNOULLI-Kette:

- Unabhängigkeit der Teilvorgänge
- gleichbleibende Erfolgswahrscheinlichkeit
- BERNOULLI-Experimente zeitlich parallel oder nacheinander

in der Regel **idealisierende** Annahmen, deshalb **Modellkritik** wichtig

aber

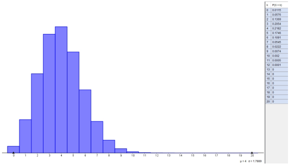
- auch einfache Modelle können helfen Einsichten zu gewinnen
- einfache Modelle als erster Schritt

PS: Wir lassen die Pfadregeln hinter uns.

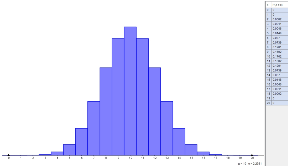
Es sei X die Anzahl der Erfolge in einer BERNOULLI-Kette der Länge n mit der Erfolgswahrscheinlichkeit p. Die Verteilung von X ist gegeben durch

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots, n$$

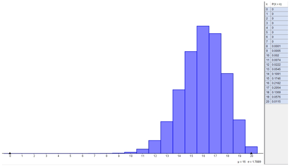
Diese Verteilung heißt Binomialverteilung mit den Parametern n und p, abgekürzt $X \sim B(n; p)$.



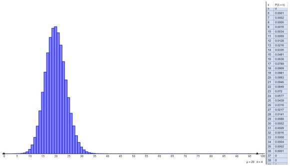
$$n = 20; p = 0,2$$



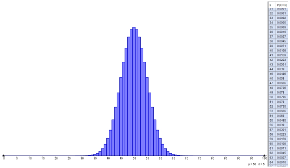
$$n = 20; p = 0,5$$



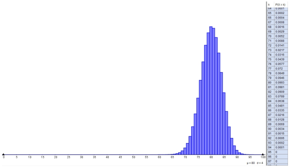
$$n = 20; p = 0,8$$



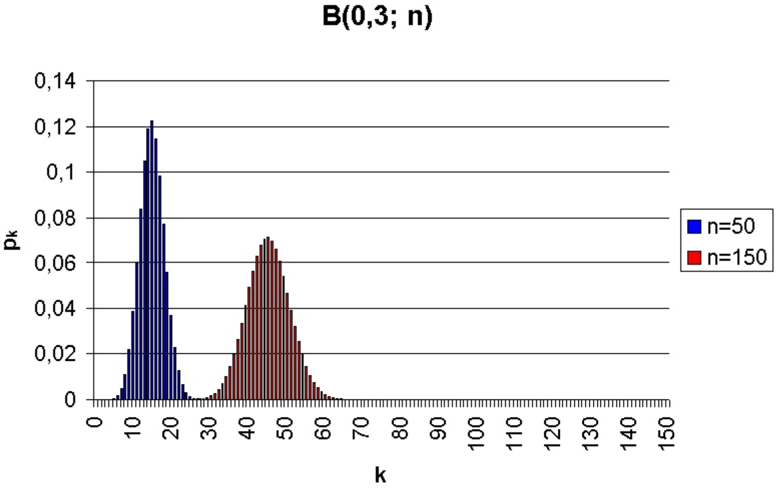
$n = 100; p = 0,2$



$n = 100; p = 0,5$



$n = 100; p = 0,8$


B(0,3; n)
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
pk
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
k
n=50
n=150

Satz

Es sei X die Anzahl der Erfolge in einer BERNOULLI-*Kette der Länge n mit der Erfolgswahrscheinlichkeit p. Dann gilt*

1. *Am wahrscheinlichsten sind etwa $n \cdot p$ Erfolge.*
2. $E(X) = n \cdot p$
3. $E\left((X - n \cdot p)^2\right) = n \cdot p \cdot (1 - p)$

Die Kenngröße $E\left((X - n \cdot p)^2\right)$ heißt **Varianz**, die Wurzel aus der Varianz heißt **Standardabweichung** und wird mit σ bezeichnet.

Interpretationen von $E(X) = np$

- Vorhersage für durchschnittliche Anzahl von Erfolgen
- wahrscheinlichste Anzahl von Erfolgen – Besonderheit der Binomialverteilung
- Lage der Verteilung

Interpretationen von $Var(X) = np(1-p)$

- Vorhersage für empirische Streuung s^2
- Breite der Verteilung
- maximal für $p = 0,5$
- minimal für $p = 0$ oder $p = 1$

$k\sigma$-Intervalle für die Anzahl der Erfolge S_n

Mit $\sigma_n = \sqrt{np(1-p)}$ gilt für hinreichend große n

$$P\left(np - k\sigma_n \leq S_n \leq np + k\sigma_n\right) \approx \begin{cases} 0,683 & k = 1 \\ 0,954 & k = 2 \\ 0,997 & k = 3 \end{cases}$$

- Wahrscheinlichkeiten fest, Lage und Länge der Intervalle hängen von n und p ab.
  - kleine Standardabweichung $\Rightarrow$ kurze Intervalle
  - große Standardabweichung $\Rightarrow$ lange Intervalle
  - sehr schöne Interpretation von Erwartungswert und Standardabweichung
- **Faustregel** für Anwendung der Näherung $np(1-p) > 9$

Fixieren $k = 2$. Für hinreichend große n gilt

$\sqrt{n}$-Gesetz für die Anzahl der Erfolge S_n

$$P\left(np - 2\sqrt{p(1-p)}\,\sqrt{n} \leq S_n \leq np + 2\sqrt{p(1-p)}\,\sqrt{n}\right) \approx 0,95.$$

- Schwankungen der **absoluten** Häufigkeiten um ihren Erwartungswert np sind von der **Größenordnung** $\sqrt{n}$ und wachsen demzufolge, wenn n wächst.
- Liegt die beobachtete Anzahl von Erfolgen außerhalb des 2σ-Intervalls, dann ist dies ein **signifikante** Abweichung vom Erwartungswert np auf dem Signifikanzniveau 0,05 eingetreten.

$$P\left(np - 2\sqrt{p(1-p)}\,\sqrt{n} \leq S_n \leq np + 2\sqrt{p(1-p)}\,\sqrt{n}\right) \approx 0,95$$

Division durch n liefert:

$\frac{1}{\sqrt{n}}$-**Gesetz** für die relative Häufigkeit der Erfolge $\frac{S_n}{n}$

Für hinreichend große gilt

$$P\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{S_n}{n} \leq p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 0,95.$$

- Schwankungen der **relativen** Häufigkeiten um ihren Erwartungswert p sind von der **Größenordnung** $\frac{1}{\sqrt{n}}$ und fallen demzufolge, wenn n wächst.
- Gesetz der großen Zahlen in der Modellebene

Merkregel für hinreichend große n

Bei n unabhängigen Versuchen unterscheidet sich die relative Häufigkeit $h_n(A)$ eines Ereignisses A von der Wahrscheinlichkeit $P(A)$ mit einer Sicherheit von mindestens 95% höchstens um $\frac{1}{\sqrt{n}}$.

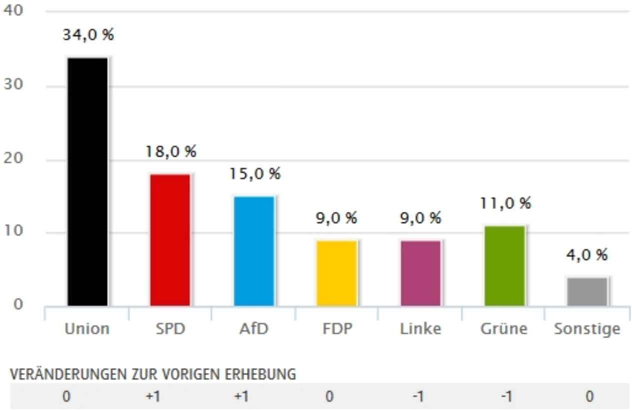
Beispiele:

n	20	100	200	400	800	2000
$\frac{1}{\sqrt{n}}$	0,22	0,10	0,07	0,05	0,04	0,02

Anwendungen

Sonntagsfrage

Überbuchung



Quelle: `https://www.infratest-dimap.de/umfragen-analysen/bundesweit/sonntagsfrage/`, 06.03.2018

$$P\left(|h_n - p| \leq 2\sqrt{p(1-p)}\frac{1}{\sqrt{n}}\right) \approx 0,95$$

$n = 1500$ und $p = 0,30$ liefern $2\sqrt{p(1-p)}\frac{1}{\sqrt{n}} = 0,024$

$n = 1500$ und $p = 0,15$ liefern $2\sqrt{p(1-p)}\frac{1}{\sqrt{n}} = 0,018$

Probleme mit Wahlumfragen und Wahlprognosen

- zufällige Stichprobe aus rund 62 Millionen Wahlberechtigten
  historisches Beispiel: amerikanische Präsidentschaftswahl 1936: Demokrat Roosevelt siegt völlig überraschend mit großem Vorsprung.
  Prognosen stützen sich auf Stichprobe aus Telefonanschlüssen und KfZ-Anmeldungen.
- Erreichbarkeit der Zielpersonen
- Wähler sagen u.U. nicht die Wahrheit
  Beispiel: 2010: 74% der Deutschen wollen ihre Organe spenden, 25 % besitzen Organsspendeausweis
- Problem Erst- und Zweitstimme, Zweitstimmen unterschätzt
- Wahlprognosen als Meinungsbildungsinstrumente

„Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife“ (Beschluss der KMK vom 18.10.2012):

Auf einer bestimmten Strecke verwendet eine Fluggesellschaft Flugzeuge mit 100 Plätzen, die vor Flugantritt gebucht und bezahlt werden. Die Flüge auf dieser Strecke sind im Voraus stets ausgebucht. Allerdings werden anschließend im Mittel 10% der gebuchten Plätze kurzfristig storniert (d.h. von den Leuten, die gebucht haben, wieder abgesagt). Von einer Person, die tatsächlich fliegt, nimmt die Fluggesellschaft 200 Euro ein, bei einer Stornierung wegen teilweiser Erstattung nur 100 Euro.

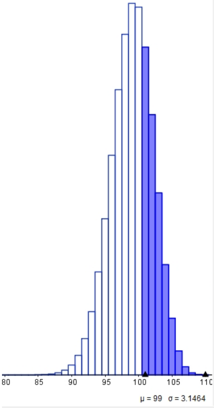
a) Nennen Sie eine Annahme, so dass die mögliche Anzahl der Stornierungen bei einem Flug als binomialverteilt modelliert werden kann. Beschreiben Sie eine reale Situation, in der diese Annahme nicht zutrifft.

Weiter: Im Folgenden wird nun angenommen, dass die mögliche Anzahl der Stornierungen bei einem Flug tatsächlich binomialverteilt ist.
Vorbildlich!

Flugzeug hat 100 Plätze, $n = 110$ werden verkauft,
10% werden im Mittel storniert.
Wie groß ist die Wahrscheinlichkeit, dass es zu Überbuchungen kommt (Ereignis $\ddot{U}$)?

- Was ist eine geeignete Zufallsgröße?
  X – Anzahl der Fluggäste, die erscheinen
- **Modellannahme:** $X \sim B(100; 0,9)$
- Überbuchungswahrscheinlichkeit: $P(\ddot{U}) = P(X > 100)$
- Orientierung mit 2σ-Intervall:$E(X) = 110 \cdot 0,9 = 99, \sigma \approx 3,15$

  $\Rightarrow$ 2σ-Intervall $[93, 105]$ mit rund 95% Wahrscheinlichkeit
- Schätzen Sie die Überbuchungswahrscheinlichkeit!



k	P(X = k)
76	0
77	0
78	0
79	0
80	0
81	0
82	0
83	0
84	0
85	0
86	0.0001
87	0.0003
88	0.0007
89	0.0016
90	0.0033
91	0.0066
92	0.0123
93	0.0214
94	0.0349
95	0.0529
96	0.0744
97	0.0966
98	0.1153
99	0.1258
100	0.1246
101	0.111
102	0.0881
103	0.0616
104	0.0373
105	0.0192
106	0.0081
107	0.0027
108	0.0007
109	0.0001
110	0

$$P(\ddot{U}) = P(X > 100) \approx 0,33$$