

Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

10. November 2010

- 1 Bedingte Wahrscheinlichkeit
 - Satz von der totalen Wahrscheinlichkeit
 - Bayessche Formel

- 2 Diskrete Zufallsvariable
 - Grundprinzipien
 - Binomialverteilung
 - Dreiecksverteilung
 - Geometrische Verteilung
 - Poissonverteilung

Satz von der totalen Wahrscheinlichkeit

- Bekannt:
 - totale Wahrscheinlichkeit $P(B)$ und damit auch $P(B^c)$
 - bedingte Wahrscheinlichkeiten $P(A|B)$ und $P(A|B^c)$
- Gesucht: totale Wahrscheinlichkeit $P(A)$

$$\begin{aligned}P(A) &= P(A \cap B) + P(A \cap B^c) \\&= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \\&= P(A|B) \cdot P(B) + P(A|B^c) \cdot (1 - P(B))\end{aligned}$$

Satz von der totalen Wahrscheinlichkeit, allgemeine Version

- $B_1, \dots, B_n$ bilden eine Zerlegung von Ω , d. h.

$$\Omega = B_1 \cup B_2 \cup \dots \cup B_n$$

wobei $B_j \cap B_k = \emptyset$ für alle Wahlen von verschiedenen j und k

- Bekannt:
 - totale Wahrscheinlichkeiten $P(B_1), P(B_2), \dots, P(B_n)$
 - bedingte Wahrscheinlichkeiten $P(A|B_1), P(A|B_2), \dots, P(A|B_n)$
- Gesucht: totale Wahrscheinlichkeit $P(A)$

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \dots + P(A|B_n) \cdot P(B_n)$$

Röntgenreihenuntersuchung auf TB

In den 1960-er Jahren wurden Röntgenreihenuntersuchungen durchgeführt. Beispielhafte Daten:

- bei 94% aller Erkrankten schlägt der Test an
- bei 1% der Gesunden schlägt der Test an
- 99.8% aller Probanden sind gesund
- Welches Ereignis nennen wir A und welches B ?
- B ist das Ereignis, dessen totale Wahrscheinlichkeit bekannt ist

Zufällig herausgegriffener Proband

A : "Verdacht auf TB"

B : "an TB erkrankt"

Röntgenreihenuntersuchungen, Fortsetzung

- $P(B) = 0.002$ (totale Wahrscheinlichkeit)
- $P(A|B) = 0.94$ (bedingte Wahrscheinlichkeit)
- $P(A|B^c) = 0.01$ (bedingte Wahrscheinlichkeit)
- Satz von der totalen Wahrscheinlichkeit

$$\begin{aligned}P(A) &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \\&= 0.94 \cdot 0.002 + 0.01 \cdot 0.998 \\&= 0.00188 + 0.00998 \\&= 0.01186\end{aligned}$$

- 1.186% aller Probanden verlassen die Untersuchung mit einem Verdacht

Bayessche Formel

- Bekannt:
 - totale Wahrscheinlichkeit $P(B)$
 - bedingte Wahrscheinlichkeiten $P(A|B)$ und $P(A|B^c)$
 - totale Wahrscheinlichkeit $P(A)$ aus dem Satz von der totalen Wahrscheinlichkeit
- Gesucht: bedingte Wahrscheinlichkeit $P(B|A)$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Bayessche Formel, allgemeine Version

- $B_1, B_2, \dots, B_n$ bilden eine Zerlegung von Ω
- Bekannt:
 - totale Wahrscheinlichkeiten $P(B_1), P(B_2), \dots, P(B_n)$
 - bedingte Wahrscheinlichkeiten $P(A|B_1), P(A|B_2), \dots, P(A|B_n)$
 - totale Wahrscheinlichkeit $P(A)$ aus dem Satz von der totalen Wahrscheinlichkeit
- Gesucht: bedingte Wahrscheinlichkeiten $P(B_1|A), P(B_2|A), \dots, P(B_n|A)$

$$P(B_1|A) = \frac{P(A|B_1) \cdot P(B_1)}{P(A)}$$

- $P(A)$ bestimmt man dabei über den Satz von der totalen Wahrscheinlichkeit
- und genauso für $B_2, \dots, B_n$

Röntgenreihenuntersuchung, Fortsetzung

- Mit welcher Wahrscheinlichkeit wird ein Kranker nicht entdeckt?
- Mit welcher Wahrscheinlichkeit ist eine Verdachtsdiagnose falsch?

erste Frage

Mit welcher Wahrscheinlichkeit wird ein Kranker nicht entdeckt?

$$P(A^c|B) = 1 - P(A|B) = 1 - 0.94 = 0.06$$

Die Wahrscheinlichkeit, dass ein Kranker für gesund gehalten wird, beträgt 6%

zweite Frage

Mit welcher Wahrscheinlichkeit ist eine Verdachtsdiagnose falsch?

$$\begin{aligned}P(B^c|A) &= 1 - P(B|A) \\&= 1 - \frac{P(A|B) \cdot P(B)}{P(A)} \\&= 1 - \frac{0.94 \cdot 0.002}{0.01186} \\&= 1 - 0.1585 \\&= 0.8415\end{aligned}$$

Wer mit Verdachtsdiagnose aus der Röntgenreihenuntersuchung kam, war mit nahezu 85% Wahrscheinlichkeit **gesund**.

Fehldiagnose

- Mit welcher Wahrscheinlichkeit kommt es zu einer Fehldiagnose?
- Das ist eine totale Wahrscheinlichkeit, nämlich

$$\begin{aligned}P(A \cap B^c) + P(A^c \cap B) &= P(A|B^c) \cdot P(B^c) + P(A^c|B) \cdot P(B) \\&= 0.06 \cdot 0.002 + 0.01 \cdot 0.998 \\&= 0.0101\end{aligned}$$

- Die Wahrscheinlichkeit einer Fehldiagnose beträgt 1.01%

Beispiel: betrügerisches Casino

► Überspringen

- Ein Casino benutzt in 0.1% aller Fälle gezinkte Würfel, bei denen die Wahrscheinlichkeit einer 6 gleich 50% ist. Es wird auf den Wurf dreier Würfel gewettet.
- Der Wurf (6, 6, 6) wird beobachtet
- Wie wahrscheinlich ist es, dass die gezinkten Würfel benutzt wurden?

A : “(6, 6, 6) beobachtet”

B : “gezinkte Würfel”

- $P(B) = 0.001$
- $P(A|B) = 0.5^3 = 0.125$
- $P(A|B^c) = \left(\frac{1}{6}\right)^3 = 0.004630$

Casino, Fortsetzung

$$\begin{aligned}P(A) &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \\&= 0.125 \cdot 0.001 + 0.004630 \cdot 0.999 \\&= 0.000125 + 0.004625 \\&= 0.004750\end{aligned}$$

Bei 0.475% aller Würfe wird (6,6,6) beobachtet.

Casino, Fortsetzung

$$\begin{aligned}P(B|A) &= \frac{P(A|B) \cdot P(B)}{P(A)} \\&= \frac{0.125 \cdot 0.001}{0.004750} \\&= 0.02632\end{aligned}$$

(6, 6, 6) ist nur in 2.6% der Fälle das Resultat der gezinkten Würfel

Realistische Version

Casino	Gensequenzierer
gezinkter Würfel	verunreinigte Probe

Beispiel zur allgemeinen Formel von Bayes

- Äpfel auf Plantage

Art	Häufigkeit
Klarapfel	40%
Delicious	55%
Elstar	5%

- Arten sind unterschiedlich anfällig für Schädling

Art	W'keit, dass befallen
Klarapfel	3%
Delicious	5%
Elstar	1%

- Frage: Ich habe einen zufällig ausgewählten, gesunden Apfel vor mir. Mit welcher Wahrscheinlichkeit ist es ein Klarapfel?

Beispiel Apfelplantage: Definition der Ereignisse

Für einen zufällig ausgewählten Apfel definiere Ereignisse

B_1		Apfel ist ein Klarapfel
B_2		Apfel ist ein Delicious
B_3		Apfel ist ein Elstar
A		Apfel ist befallen

Bekannte Wahrscheinlichkeiten

$$P(B_1) = 0.40$$

$$P(A|B_1) = 0.03$$

$$P(B_2) = 0.55$$

$$P(A|B_2) = 0.05$$

$$P(B_3) = 0.05$$

$$P(A|B_3) = 0.01$$

Satz von der totalen Wahrscheinlichkeit

$$\begin{aligned}
 P(A) &= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3) \\
 &= 0.03 \cdot 0.40 + 0.05 \cdot 0.55 + 0.01 \cdot 0.05 = 0.04
 \end{aligned}$$

4% aller Äpfel sind befallen.

Beispiel Apfelplantage: Fortsetzung

- Ein zufällig herausgegriffener Apfel ist befallen. Mit welcher Wahrscheinlichkeit ist es ein Klarapfel?
- Gesucht ist $P(B_1|A)$

$$\begin{aligned}P(B_1|A) &= \frac{P(A|B_1) \cdot P(B_1)}{P(A)} \\&= \frac{0.03 \cdot 0.40}{0.04} \\&= 0.30\end{aligned}$$

- Ein zufällig herausgegriffener befallener Apfel ist mit einer Wahrscheinlichkeit von 30% ein Klarapfel

Beispiel Apfelplantage: Fortsetzung

- Ein zufällig herausgegriffener Apfel ist gesund. Mit welcher Wahrscheinlichkeit ist es ein Klarapfel?
- Gesucht ist $P(B_1|A^c)$

$$\begin{aligned}P(B_1|A^c) &= \frac{P(A^c|B_1) \cdot P(B_1)}{P(A^c)} \\&= \frac{0.97 \cdot 0.40}{0.96} \\&= 0.4042\end{aligned}$$

- Ein zufällig herausgegriffener gesunder Apfel ist mit einer Wahrscheinlichkeit von 40.42% ein Klarapfel

Zufallsvariable

- Zufallsexperiment wird durchgeführt, dessen Ergebnis ein Wert ist
- Das ist der Wert der Zufallsvariablen
- Zufallsvariablen heißen meist X , Y
- Mathematisch ausgedrückt: Eine Zufallsvariable ordnet jedem Elementarereignis ω eine Zahl $X(\omega)$ zu
- Beispiel 10-facher Wurf eines fairen Würfels: Die Anzahl der Sechsen definiert eine Zufallsvariable X
- Zufallsvariable lenken den Blick auf die interessanten Daten, indem sie die Elementarereignisse ausblenden

Diskrete Zufallsvariablen

- Eine Zufallsvariable heißt *diskret*, wenn alle ihre Werte ganze Zahlen sind
- Eine Zufallsvariable heißt *stetig*, wenn sie auch gebrochene Werte annehmen kann

Interpretation

Wahrscheinlichkeitstheorie	Experiment
Zufallsvariable X	Messvorrichtung
Ereignisraum Ω	Menge aller möglichen Versuchsabläufe
Elementarereignis ω	beobachteter Versuchsablauf
Wert $X(\omega)$	beobachteter Messwert

Schreibweisen

X eine Zufallsvariable auf Ω . Wir schreiben zur Abkürzung (hierbei sind a und b irgendwelche Zahlen):

$$\{X = a\} = \{\text{alle Elementarereignisse } \omega, \text{ für die } X(\omega) = a\}$$

$$\{X \leq a\} = \{\text{alle Elementarereignisse } \omega, \text{ für die } X(\omega) \leq a\}$$

$$\{a < X \leq b\} = \{\text{alle Elementarereignisse } \omega, \text{ für die } a < X(\omega) \leq b\}$$

usw.

Beispiel zur Schreibweise

- Dreifacher Wurf einer fairen Münze, also $\Omega = \{A, Z\}^3$
- X bezeichne die Anzahl der Würfe mit "Adler". Dann kann X die Zahlen 0,1,2 und 3 annehmen
- $\{X = 2\} = \{(A, A, Z), (A, Z, A), (Z, A, A)\}$
- $P(X = 2) = \frac{3}{8} = 0.375$
- Statt $P(X = 2)$ schreibt man auch $P_X(2)$
- Dann ist P_X eine Verteilung auf den ganzen Zahlen

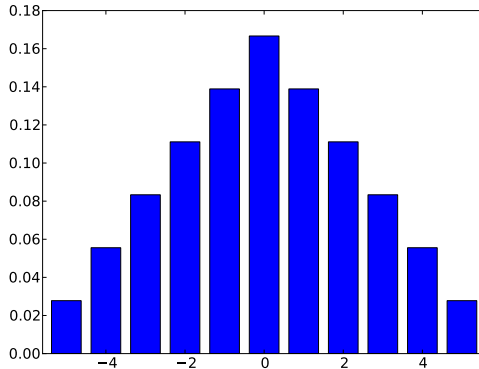
Binomialverteilung als P_X

- Beispiel $\Omega =$ "Serien von n ja/nein-Experimenten"
- Erfolgswahrscheinlichkeit im Einzelfall sei p
- X zählt die Zahl der Erfolge
- $P(X = k) = P_X(k) = B_{n,p}(k)$

Beispiel Dreiecksverteilung

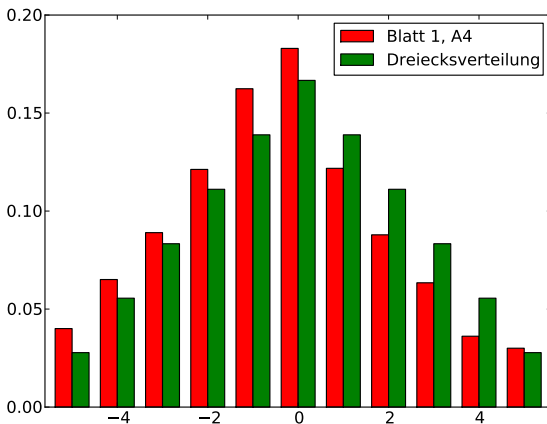
- zweifacher Wurf eines fairen Würfels,
 $X =$ Differenz zwischen erstem und zweitem Wurf
- $\{X = 5\} = \{(6, 1)\}$
- $\{X = 4\} = \{(6, 2), (5, 1)\}$
- $\{X = 3\} = \{(6, 3), (5, 2), (4, 1)\}$
- $\{X = 0\} = \{(6, 6), (5, 5), (4, 4), (3, 3), (2, 2), (1, 1)\}$
- $\{X = -5\} = \{(1, 6)\}$
- Diese Verteilung heißt *Dreiecksverteilung*

Dreiecksverteilung



Blatt 1, Aufgabe 4

Relative Häufigkeitsverteilung der Differenz aus erstem und zweitem Wurf



Geometrische Verteilung

Die geometrische Verteilung modelliert Wartezeiten in getakteten Modellen

Geometrische Verteilung, Beispiel

Wurf eines fairen Würfels:

- Erfolgswahrscheinlichkeit im Einzelfall $p = \frac{1}{6}$
- Die Wahrscheinlichkeit, dass im ersten Wurf eine Sechs fällt, ist

$$\frac{1}{6} = 0.1667$$

- Die Wahrscheinlichkeit, dass die erste Sechs im zweiten Wurf fällt, ist

$$\frac{5}{6} \cdot \frac{1}{6} = 0.1389$$

Beispiel, Fortsetzung

Wurf eines fairen Würfels:

- Die Wahrscheinlichkeit, dass die erste Sechs im dritten Wurf fällt, ist

$$\left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} = 0.1157$$

- Die Wahrscheinlichkeit, dass die erste Sechs im k -ten Wurf fällt, ist

$$\left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$$

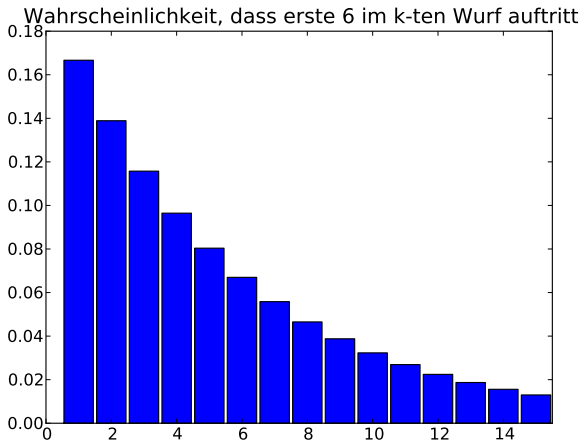
Geometrische Verteilung: Definition

Die *geometrische Verteilung* zum Parameter p wird gegeben durch

$$G_p(k) = (1 - p)^{k-1} \cdot p$$

In einem ja/nein-Experiments, bei dem die Erfolgswahrscheinlichkeit im Einzelfall gleich p ist, gebe die Zufallsvariable X die Nummer des Zugs mit dem ersten Erfolg an. Dann ist X verteilt gemäß G_p .

Stabdiagramm von $G_{1/6}$



Poissonverteilung

Die Poissonverteilung modelliert seltene Ereignisse

Selten bedeutet, dass sich die Einzelereignisse nicht beeinflussen

Poissonverteilung: Definition

Es sei $\lambda > 0$. Die Poissonverteilung zum Parameter λ ist definiert durch

$$P_{\lambda}(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

Poissonverteilung, Bedeutung

Unter den folgenden Voraussetzungen ist eine Zufallsvariable X poissonverteilt zum Parameter λ :

- X zählt das Auftreten eines Ereignisses pro Zähleinheit
- Im Mittel treten λ Ereignisse pro Zähleinheit auf
- Die Ereignisse beeinflussen sich nicht gegenseitig

Beispiel: α -Strahlung

Versuchsaufbau:

- Wir beobachten die Einschläge von α -Teilchen in eine Zelle
- Im Mittel gebe es einen Einschlag alle 10 Minuten, also 6 Einschläge pro Stunde.
- Welcher Prozentsatz der Zellen übersteht eine Stunde ohne Einschlag?

Es ist $\lambda = 6$, und wir suchen den Wert von P_λ für $k = 0$

$$P_\lambda(0) = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-6} = 0.002479$$

Nur ungefähr jede 400-te Zelle bleibt unbeschädigt

Stabdiagramm von P_6 