Mathematik für Biologen

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

Heinrich-Heine-Universität Düsseldorf

13. Dezember 2013

Teil II

Wahrscheinlichkeitstheorie

Zufallsereignisse

Wahrscheinlichkeiten I

Frage: Wie groß ist die Wahrscheinlichkeit, dass hier im Hörsaal gleich viele männliche und weibliche Studierende sitzen?

Antwort: Das ist keine gute Frage.

Wahrscheinlichkeiten II

- Aus allen Studierenden der Mathematik werden zufällig vier ausgewählt. Mit welcher Wahrscheinlichkeit sind zwei weiblich und zwei männlich?
- Die Mathematik hat ein ausgeglichenes Geschlechterverhältnis.
- Es gibt die folgenden 16 Möglichkeiten:

(m,m,m,m)	(m,m,m,w)	(m,m,w,m)	**(m,m,w,w)**
(m,w,m,m)	**(m,w,m,w)**	**(m,w,w,m)**	(m,w,w,w)
(w,m,m,m)	**(w,m,m,w)**	**(w,m,w,m)**	(w,m,w,w)
(w,w,m,m)	(w,w,m,w)	(w,w,w,m)	(w,w,w,w)

- 6 der 16 gleichwahrscheinlichen Versuchsausgängen gehören zur gesuchten Gruppe.
- Also ist die Wahrscheinlichkeit gleich $6/16 = 0.375$

Wahrscheinlichkeit

- Was ist eine Wahrscheinlichkeit?
- Eine Wahrscheinlichkeit ist eine Modellannahme.
- Modellannahmen kommen her von
  - beobachteten relativen Häufigkeiten
  - abstrakten Überlegungen (z.B. faire Münze)
  - Konstruktion aus Teilsystemen (z.B. mehrfacher Münzwurf)
- Überprüfung des Modells am Experiment

Modelle

- Die Wissenschaft arbeitet mit Modellen
- Das Forschungsobjekt der Naturwissenschaften kann meist nicht vollständig verstanden werden, weil es
  - zu komplex ist (Organismus eines Säugetiers)
  - zu schlecht zu beobachten ist (Atomkern, Galaxie)
  - beides (Zellstoffwechsel)
- Also macht man sich ein Modell; zu den Modellannahmen gehören oft auch Wahrscheinlichkeiten
- Die wissenschaftliche Methode besteht darin, dass man Vorhersagen aus dem Modell ableitet und diese im Experiment überprüft (Falsifizierbarkeit)

Elementarereignisse

- Welchen Objekten können Wahrscheinlichkeiten zugeordnet werden?
- Diese Objekte sind die Elementarereignisse
- Ein Elementarereignis ist ein Versuchsergebnis, das sich nicht wiederum aus Versuchsergebnissen zusammensetzt
- Beispiel: Bei dem Experiment "Zufällige Auswahl von vier Studierenden der Mathematik" gibt es 16 Elementarereignisse

(m,m,m,m)	(m,m,m,w)	(m,m,w,m)	(m,m,w,w)
(m,w,m,m)	(m,w,m,w)	(m,w,w,m)	(m,w,w,w)
(w,m,m,m)	(w,m,m,w)	(w,m,w,m)	(w,m,w,w)
(w,w,m,m)	(w,w,m,w)	(w,w,w,m)	(w,w,w,w)

- Die Aussage "Die Mathematik hat ein ausgeglichenes Geschlechterverhältnis" bedeutet, dass jedes dieser Elementarereignisse die Wahrscheinlichkeit $1/16$ besitzt

Ereignisse

- Elementarereignisse werden zu Ereignissen zusammengefasst
- Im Beispiel interessiert

$$A = \{(m, m, w, w), (m, w, m, w), (m, w, w, m), (w, m, m, w), (w, m, w, m), (w, w, m, m)\}$$

- Elementarereignisse werden häufig durch den griechischen Buchstaben ω bezeichnet
- Ereignisse durch große lateinische Buchstaben
- Spezielle Ereignisse:
  - Das *sichere Ereignis* besteht aus allen möglichen Elementarereignissen für das gegebene Experiment. Es wird mit Ω bezeichnet. Ω heißt auch *Ereignisraum.*
  - Das *unmögliche Ereignis* ist leer. Es wird mit $\emptyset$ bezeichnet.

Wahrscheinlichkeitsverteilung

- Eine Wahrscheinlichkeitsverteilung ordnet jedem Elementarereignis ω eine Zahl $P(\omega) \geq 0$ zu
- Für ein Ereignis $A = \{\omega_1, \dots, \omega_n\}$ ist die Wahrscheinlichkeit dann $P(A) = P(\omega_1) + \cdots + P(\omega_n)$
- Damit P eine Wahrscheinlichkeitsverteilung ist, muss $P(\Omega) = 1$ gelten
- $P(\emptyset) = 0$ gilt dann automatisch

Beispiele für Ereignisse

- Zweifacher Wurf eines Würfels: Der Ereignisraum ist

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

- Mit *Pasch* bezeichnet man das Ereignis

$$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

- Das Ereignis "eine 3 und eine 4" ist

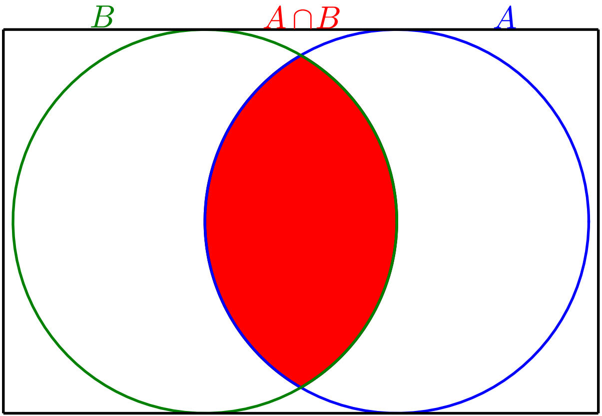
$$B = \{(3,4),(4,3)\}$$

Konstruktionen

Aus einfachen Ereignissen werden komplexere aufgebaut.

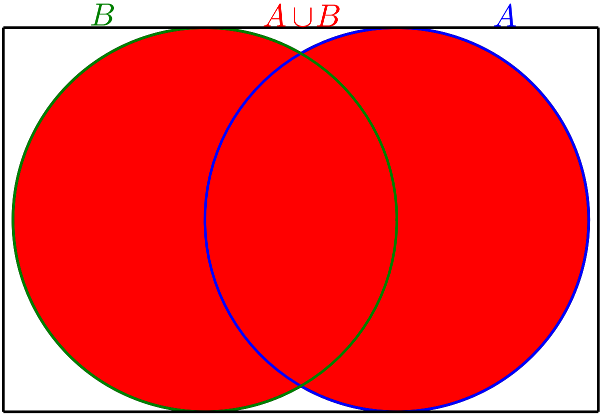
- Durchschnitt
- Vereinigung
- Differenz
- Komplement
- Produkt

Durchschnitt zweier Ereignisse



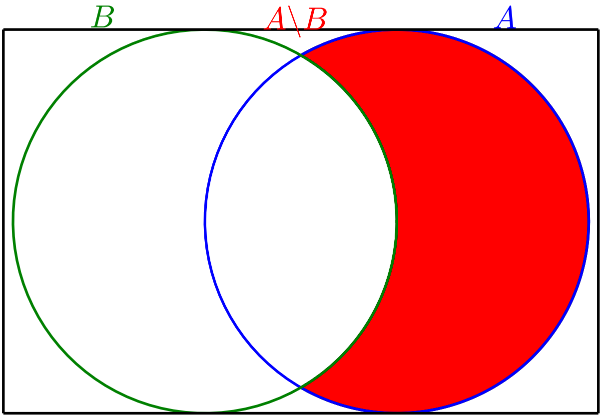
Der Durchschnitt $A \cap B$ zweier Ereignisse A und B besteht aus allen Elementarereignissen, die sowohl zu A als auch zu B gehören.

Vereinigung zweier Ereignisse



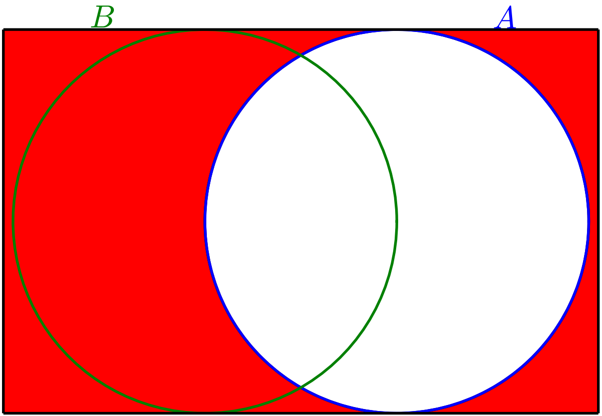
Die Vereinigung $A \cup B$ zweier Ereignisse A und B besteht aus allen Elementarereignissen, die zu mindestens einem der beiden Ereignisse gehören.

Differenz zweier Ereignisse



Die Differenz $A \setminus B$ zweier Ereignisse A und B besteht aus allen Elementarereignissen, die zu A, aber nicht zu B gehören.

Komplement



Das Komplement A^c eines Ereignisses A besteht aus allen Elementarereignissen, die nicht zu A gehören.

Mengensprechweise

Die Menge aller Elementarereignisse ist der Ereignisraum. Seine Teilmengen heißen (Zufalls)-Ereignisse. Die Mengenlehre dient uns als Sprechweise, Ereignisse kurz und zweifelsfrei zu beschreiben.

verbal	mathematisch
Ereignisse A und B treffen ein	$A \cap B$
Ereignis A oder Ereignis B trifft ein	$A \cup B$
Ereignis A trifft nicht ein	A^c
Ereignis A trifft ein, Ereignis B aber nicht	$A \setminus B$
unmögliches Ereignis	$\emptyset$
sicheres Ereignis ($=$ Ereignisraum)	Ω
Elementarereignis ω gehört zu A	$\omega \in A$
Elementarereignis ω gehört nicht zu A	$\omega \notin A$
alle Elementarereignisse von A gehören zu B	$A \subset B$

Beispiele für Mengensprech

$\Omega =$ "Wurf eines Würfels" $= \{1, 2, 3, 4, 5, 6\}$,
$A =$ "ungerade Zahl gewürfelt" $= \{1, 3, 5\}$ und
$B =$ "Zahl kleiner 4 gewürfelt" $= \{1, 2, 3\}$

- $A \cap B = \{1, 3\}$
- $A \cup B = \{1, 2, 3, 5\}$
- $A^c = \{2, 4, 6\}$
- $A \setminus B = \{5\}$
- $A \cup B = \Omega \setminus \{4, 6\}$

Produkt

- Gegeben zwei Ereignisräume Ω_1 und Ω_2 und in jedem ein Ereignis

$$A \subset \Omega_1 \qquad B \subset \Omega_2$$

- Das Produktereignis $A \times B$ besteht aus allen Paaren (a, b) mit $a \in A$ und $b \in B$
- Mathematisch

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}$$

- Es ist ein Ereignis im Ereignisraum $\Omega_1 \times \Omega_2$

Ein Produkt mit 12 Elementen

$$A = \{a_1, a_2, a_3\}$$
$$B = \{b_1, b_2, b_3, b_4\}$$

	b_1	b_2	b_3	b_4
a_1	(a_1, b_1)	(a_1, b_2)	(a_1, b_3)	(a_1, b_4)
a_2	(a_2, b_1)	(a_2, b_2)	(a_2, b_3)	(a_2, b_4)
a_3	(a_3, b_1)	(a_3, b_2)	(a_3, b_3)	(a_3, b_4)

Beispiele für Produktereignisse

- Im Beispiel "Geschlechter von vier Studierenden" ist

$$\begin{aligned}\Omega &= \{m, w\} \times \{m, w\} \times \{m, w\} \times \{m, w\} \\ &= \{m, w\}^4\end{aligned}$$

- Zweifacher Wurf eines Würfels

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

Wahrscheinlichkeiten

Konsistenzregeln

Für jedes Ereignis A gebe es eine Zahl $P(A)$ mit

(P1) $P(A) \geq 0$ für alle A

(P2) $P(\Omega) = 1$

(P3) $P(A \cup B) = P(A) + P(B)$, falls A und B **disjunkte** Ereignisse sind, also keine gemeinsamen Elementarereignisse enthalten

Dann ist P ein Wahrscheinlichkeitsmaß auf Ω, und (Ω, P) ist ein wahrscheinlichkeitstheoretisches Modell des Zufallsexperiments.

Rechenregeln

- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B)$, wenn $A \cap B = \emptyset$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subset B$, dann folgt $P(A) \leq P(B)$

Die Laplace-Verteilung

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

$$\begin{aligned} A &= \text{“beide gehören zu } S_1\text{”} \\ B &= \text{“beide gehören zu } S_2\text{”} \end{aligned}$$

Wie oben

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

Daher

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

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 beträgt

$$\frac{250\,000 + 22\,500}{422\,500} = 0.64$$

Unabhängigkeit

Stochastische Unabhängigkeit

- Zwei Ereignisse A und B heißen *(stochastisch) unabhängig*, wenn
$$P(A \cap B) = P(A) \cdot P(B)$$
- Diese Formel ist die *Produktformel für unabhängige Ereignisse*
- 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 Zahlenlotto

- Beim Zahlenlotto ist die Gewinnwahrscheinlichkeit 1 : 14 000 000 für den Hauptgewinn
- Wie viele Spiele sind nötig für eine 40%-Chance auf den Hauptgewinn?
- Misserfolgswahrscheinlichkeit $q = 0.999\,999\,929$
- Wahrscheinlichkeit von n Misserfolgen ist $q^n = 0.60$
- Logarithmiere: $n \cdot \ln(q) = \ln(0.60)$
- Lösung

$$n = \frac{\ln(0.60)}{\ln(q)} = 7\,200\,000$$

- Wünscht man eine 80% Chance auf den Hauptgewinn, dann

$$n = \frac{\ln(0.20)}{\ln(q)} = 23\,000\,000$$

Bedingte Wahrscheinlichkeit

Nutzung von Zusatzinfo

- Die bedingte Wahrscheinlichkeit ist eine Wahrscheinlichkeit unter Berücksichtigung von Zusatzinformationen
- Beispielsweise ist für einen 50-jährigen die Wahrscheinlichkeit, 80 Jahre zu werden, (etwas) höher als für ein Neugeborenes
- Allgemein wird mit $P(A|B)$ die Wahrscheinlichkeit von A bezeichnet, wenn bereits bekannt ist, dass B eingetreten ist

Bedingte Wahrscheinlichkeit

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

bezeichnet man als *bedingte Wahrscheinlichkeit* von A unter der *Hypothese* B

- Die Hypothese B ist also vorausgesetzt (im Beispiel ist jemand bereits 50 Jahre alt geworden)
- Das Ereignis A ist das Ereignis, dessen Wahrscheinlichkeit interessiert (im Beispiel ist A das Ereignis, älter als 80 zu werden)
- Man bezeichnet $P(A)$ auch als *totale* Wahrscheinlichkeit, wenn man den Unterschied zu einer bedingten Wahrscheinlichkeit verdeutlichen will

Rechenregeln

- $P(A|B)$ ist eine Wahrscheinlichkeit für A, erfüllt also die Rechenregeln für Wahrscheinlichkeiten
- die wichtigste ist die Regel für die Wahrscheinlichkeit des Komplementärereignisses

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

Produktformel

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Insbesondere sind A und B genau dann unabhängig, wenn

$$P(A|B) = P(A)$$

Heuristische Begründung der Formel

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

- unter der Hypothese B ist B sicher, also $P(B|B) = 1$; daher wird durch $P(B)$ geteilt
- unter der Hypothese B sind diejenigen Elementarereignisse von A, die nicht in B liegen, irrelevant; daher steht im Zähler $P(A \cap B)$ und nicht $P(A)$

Begriffsklärung

Ein Spam-Filter unterscheidet zwischen Spam und erwünschter Mail. Dabei kommen Fehler vor. Für eine zufällig ausgewählte Mail interessieren zwei Ereignisse

$$A : \text{“es handelt sich um Spam”}$$
$$B : \text{“der Filter hält sie für Spam”}$$

Dann ist $P(B|A)$ die Wahrscheinlichkeit, dass Spam in den Spam-Ordner gelegt wird
und $P(A|B)$ ist die Wahrscheinlichkeit, dass im Spam-Ordner aufgefundene Mail tatsächlich Spam ist.

Beispiel zur bedingten Wahrscheinlichkeit

- Sei E_{m50} das Ereignis, dass ein männliches Neugeborenes ein Alter von mindestens 50 Jahren erreichen wird; laut (österreichischer) Sterbetafel ist $P(E_{m50}) = 0.919$
- für 80 Jahre $P(E_{m80}) = 0.365$
- für weibliche Neugeborene $P(E_{w50}) = 0.958$ und $P(E_{w80}) = 0.566$
- dann wegen $E_{m50} \cap E_{m80} = E_{m80}$

$$P(E_{m80}|E_{m50}) = \frac{0.365}{0.919} = 0.397$$

- und

$$P(E_{w80}|E_{w50}) = \frac{0.566}{0.958} = 0.591$$

Wahrscheinlichkeitsbäume

$$
\begin{tikzcd}
 & & & {} \arrow[dll, "P(A)"'] \arrow[drr, "P(A^c)"] & & & \\
 & A \arrow[dl, "P(B|A)"'] \arrow[dr, "P(B^c|A)"] & & & & A^c \arrow[dl, "P(B|A^c)"'] \arrow[dr, "P(B^c|A^c)"] & \\
B & & B^c & & B & & B^c
\end{tikzcd}
$$

Juwelenwespen

Juwelenwespen legen Eier in Puppen anderer Insekten. Das Geschlechterverhältnis ist

- 95% weibliche und 5% männliche Eier, wenn die Puppe noch nicht von einer anderen Juwelenwespe infiziert wurde
- 10% weibliche und 90% männliche Eier andernfalls

Wahrscheinlichkeitsbaum für die Juwelenwespe

30% aller Puppen seien beim Anflug einer Wespe bereits einmal infiziert, aber keine Puppe zweimal