2.3 Bedingte Wahrscheinlichkeit

Definition: Es sei P ein Wahrscheinlichkeitsmaß über der Menge Ω. Sind $A, B \subseteq \Omega$ Ereignisse und ist $P(B) \neq 0$, dann heißt $P(A|B) := \frac{P(A \cap B)}{P(B)}$ die bedingte Wahrscheinlichkeit von A unter der Bedingung B. In Lehrbüchern wird $P(A|B)$ auch oft mit $P_B(A)$ bezeichnet.

Beispiel: In einem Studiengang gibt es 360 Studienplätze. Darauf bewerben sich 100 Frauen und 400 Männer. Alle Studienplätze werden vergeben. Es werden 80 Frauen zugelassen. Diese Daten könnte man in eine **Vierfeldertafel** eintragen

	A	$\bar{A}$	
B	$\lvert A \cap B \rvert$	$\lvert \bar{A} \cap B \rvert$	$\lvert B \rvert$
$\bar{B}$	$\lvert A \cap \bar{B} \rvert$	$\lvert \bar{A} \cap \bar{B} \rvert$	$\lvert \bar{B} \rvert$
	$\lvert A \rvert$	$\lvert \bar{A} \rvert$	$\lvert \Omega \rvert$

Das wäre hier (Z = zugelassen, F = Frau, M = Mann, grün markiert sind die gegebenen Daten; die restlichen Zahlen können daraus ermittelt werden: In der letzten Spalte bzw. Zeile stehen jeweils die Summen)

	Z	$\bar{Z}$	Bewerbungen
F	80	20	100
M	280	120	400
Bewerbungen	360	140	500

Die Zulassungsquote für Frauen beträgt also $\frac{80}{100} = 80\%$, diejenige für Männer $\frac{280}{400} = 70\%$, insgesamt $\frac{360}{500} = 72\%$

Ebenso benutzt man Vierfeldertafeln entsprechend für Wahrscheinlichkeiten:

	A	$\bar{A}$	
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

In unserem Beispiel könnte man die relativen Häufigkeiten nehmen, indem wir alle Zahlen durch $|\Omega| = 500$ dividieren

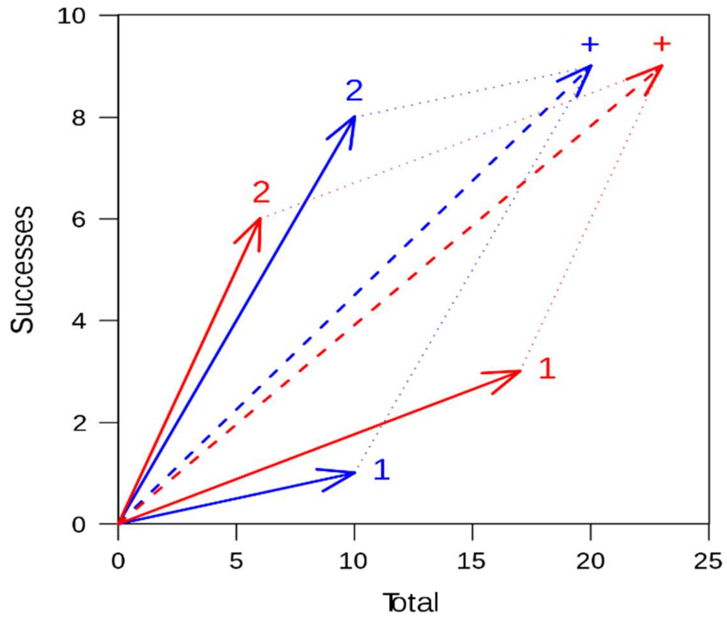
	Z	$\bar{Z}$	
F	0,16	0,04	0,2
M	0,56	0,24	0,8
	0,72	0,28	1

(Hier käme man übrigens schon mit drei gegebenen Zahlen aus.)

Jetzt lassen sich die Zulassungsquoten auch als bedingte Wahrscheinlichkeiten interpretieren. Z.B. ist sie bei Männern
$\frac{280}{400} = \frac{280/500}{400/500} = \frac{P(M \cap Z)}{P(M)} = \frac{0{,}56}{0{,}8} = 70\%.$
Das entspräche dann der Wahrscheinlichkeit, zugelassen zu werden – gesetzt den Fall, man ist männlichen Geschlechts.

Für Frauen ist die Zulassungsquote
$\frac{80}{100} = \frac{80/500}{100/500} = \frac{P(W \cap Z)}{P(W)} = \frac{0{,}16}{0{,}2} = 80\%.$
Die allgemeine Zulassungsquote liegt bei $\frac{360}{500} = 72\%.$

Das Simpson-Paradoxon



Wir stellen nun die Zulassungszahlen von zwei Studiengängen nebeneinander

I	Z	$\bar{Z}$	Bewerbungen	
F	80	20	100	80%
M	280	120	400	70%
Bewerbungen	360	140	500	72%

II	Z	$\bar{Z}$	Bewerbungen	
F	160	240	400	40%
M	20	80	100	20%
Bewerbungen	180	320	500	36%

I+II	Z	$\bar{Z}$	Bewerbungen	
F	240	260	500	48%
M	300	200	500	60%
Bewerbungen	540	460	1000	54%

Das führt zu dem folgenden paradoxen Ergebnis: Sowohl im Studiengang I als auch im Studiengang II liegt die Zulassungsquote bei den Frauen (80% und 40%) **höher** als bei den Männern (70% und 20%), also ist in beiden Fällen $P(Z|F) > P(Z|M)$. Für die gesamte Fakultät (I+II) gilt umgekehrt: Die Zulassungsquote (insgesamt 54%) ist bei den Frauen mit nur 48% $= P(Z|F)$ deutlich **niedriger** als die der Männer mit 60%$= P(Z|M)$.

Zur weiteren Diskussion siehe z.B. Arens Kap. 37.3 oder https://m.spiegel.de/wissenschaft/mensch/simpsons-paradoxon-diese-statistik-kann-nicht-stimmen-oder-doch-a-1068204.html

Bayes: „Schließen von der Wirkung auf die Ursache“

Die Definition der bedingten Wahrscheinlichkeit besagt $P(A|B) = \frac{P(A \cap B)}{P(B)}$ und $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Aus der zweiten Formel erhalten wir $P(A \cap B) = P(B|A) \cdot P(A)$. Eingesetzt in erste Formel ergibt das die sogenannte

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

Zunächst nicht in der vollen Allgemeinheit sehen wir uns noch zwei Formeln an:

Satz von der totalen Wahrscheinlichkeit $P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})$

Beweis → Vorlesung

Satz von Bayes $P(\, A \mid B\,) = \frac{P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})} \cdot P(A)$

Die dritte Formel ist nur eine Kombination der ersten beiden Formeln.

Beispiel:

Den praktischen Nutzen wollen wir an einem Beispiel aus der Medizin erläutern (nach Arens Kap. 37.2): Ein Patient wird untersucht. Das Ereignis A soll sein, dass er die Krankheit A hat, das Ereignis B, dass er das Symptom B zeigt (z.B. dass ein bestimmter Test positiv verläuft). Statt $A, \bar{A}, B, \bar{B}$ verwendet man hier gerne die Bezeichnungen A^+, A^-, B^+, B^- (z.B. A^+ = HIV-positiv, B^+ = Bluttest positiv).

Die **Qualität** eines diagnostischen Tests wird in der medizinischen Statistik durch die folgenden Maßzahlen bestimmt:

Sensitivität $P(B^+ \mid A^+)$ **Spezifität** $P(B^- \mid A^-)$

$P(B^+ \mid A^+)$ gibt an, mit welcher Wahrscheinlichkeit der Test bei Vorliegen der Krankheit Alarm gibt, und $P(B^- \mid A^-)$ gibt an, mit welcher Wahrscheinlichkeit der Test bei Nicht-Vorliegen der Krankheit Entwarnung gibt. Dann wird noch benötigt die sogenannte

Prävalenz $P(A^+)$

Bei Vorliegen dieser drei Größen (das wäre die Aufgabe der Statistik) lässt sich dann, je nach dem, ob der Test positiv oder negativ (Bedingung B^+ oder B^-) verlaufen ist, nach Bayes auf die Wahrscheinlichkeiten $P(A^+ \mid B^+)$ bzw. $P(A^- \mid B^-)$ schliessen:

$$P(\, A^+ \mid B^+\,) = \frac{P(B^+|A^+)}{P(B^+)} \cdot P(A^+)$$

Dabei ist $P(B^+) = P(A^+) \cdot P(B^+ \mid A^+) + P(A^-) \cdot P(B^+ \mid A^-)$
mit $P(A^-) = 1 - P(A^+)$ und $P(\, B^+ \mid A^-\,) = 1 - P(\, A^- \mid B^-\,)$

bzw.

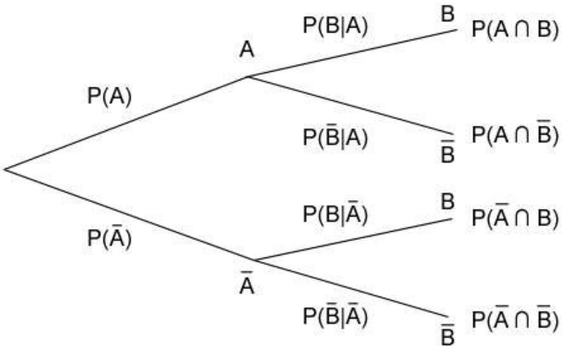
$$P(\, A^- \mid B^-\,) = \frac{P(B^-|A^-)}{P(B^-)} \cdot P(A^-) \qquad \text{mit } P(B^-) = 1 - P(B^+).$$

Zur Kritik des *Eliza*-Aidstest 1989 standen in der ZEIT die Zahlen
$P(\, B^+ \mid A^+\,) = 0{,}999$; $P(\, B^- \mid A^-\,) = 0{,}995$ und $P(A^+) = 0{,}001$
Einsetzen in die obigen Formeln liefert dann $P(\, A^+ \mid B^+\,) = 0{,}1667$. Dies bedeutet: Ist der Testbefund positiv, dann ist die Wahrscheinlichkeit für HIV-positiv nicht einmal 17%. Dies war der Grund, warum sich die ZEIT gegen die geplante Massenanwendung dieses Tests wandte: Rund 83% aller positiv getesteten Personen wären grundlos in Verzweiflung gestürzt worden.
Ebenso folgt aus den gegebenen Zahlen, dass $P(\, A^- \mid B^-\,) = 0{,}999999$ ist. Dazu wäre z.B. das Testen von Blutkonserven eine sinnvolle Anwendung: Ist der Befund positiv, wird die Konserve vernichtet. Nur wenn der Befund negativ ist, kann die Konserve verwendet werden. Zwar werden rund 83% der vernichteten Konserven fälschlich vernichtet, aber dafür hat der Test nur gute Konserven herausgefiltert.

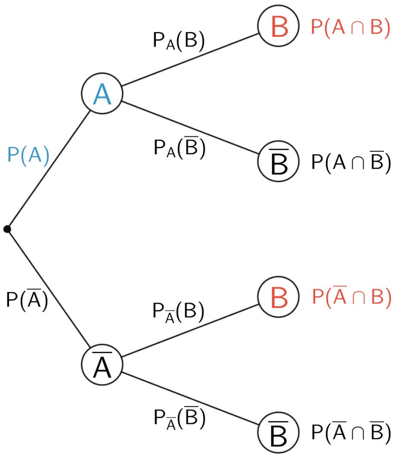
Baumdiagramm versus Vierfeldertafel

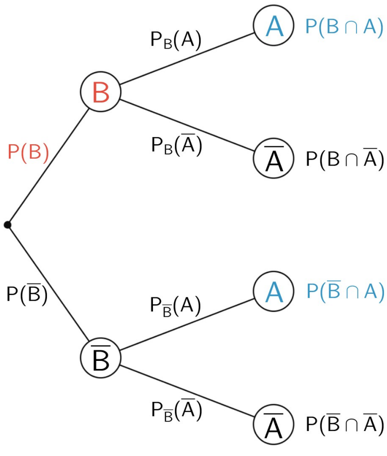
$P(B|A) = \frac{P(A\cap B)}{P(A)}$; $P(A|B) = \frac{P(A\cap B)}{P(B)}$

	A	$\bar{A}$	
B	$P(A\cap B)$	$P(\bar{A}\cap B)$	$P(B)$
$\bar{B}$	$P(A\cap \bar{B})$	$P(\bar{A}\cap \bar{B})$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1



Hier zur Abwechslung die Schreibeise $P_A(B)$ für $P(B|A)$





$P(B) = P(A\cap B) + P(\bar{A}\cap B)$ $\qquad$ $P(A) = P(B\cap A) + P(\bar{B}\cap A)$

https://www.mathelike.de/abi-check-mathe-abi-skript-bayern/3-stochastik/3-1-wahrscheinlichkeitsrechnung/3-1-4-baumdiagramm-und-vierfeldertafel.html

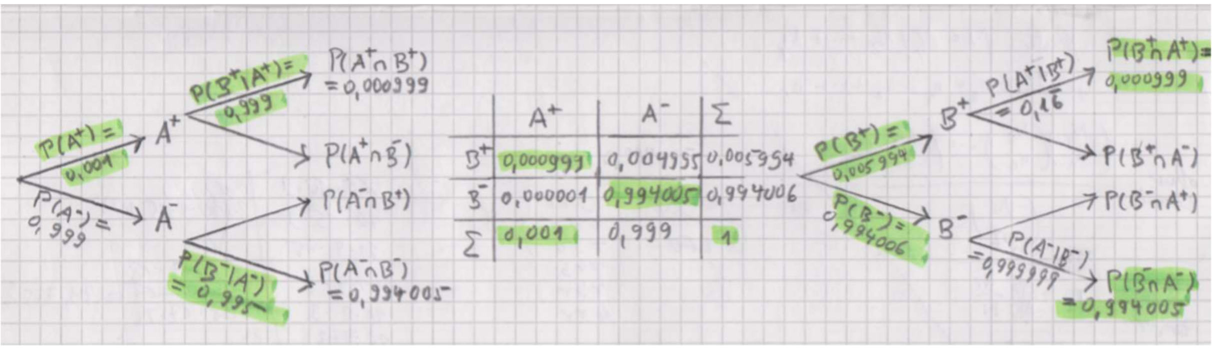
Verzweigungsregel: Die Summe der Wahrscheinlichkeiten an den Ästen, die von einem Knoten ausgehen, ist gleich eins. Beispiel: $P(B|A) + P(\bar{B}|A) = 1$

Erste Pfadregel: Die Wahrscheinlichkeit am Ende eines Pfades ist gleich dem Produkt der Wahrscheinlichkeiten längs dieses Pfades. Beispiel: $P(A\cap B) = P(A)\cdot P(B|A)$

Zweite Pfadregel: Die Wahrscheinlichkeit eines Ereignisses ist gleich der Summe der Wahrscheinlichkeiten aller Pfade, die zu diesem Ereignis führen. Das ist wieder der Satz von der totalen Wahrscheinlichkeit:
$P(B) = P(A\cap B) + P(\bar{A}\cap B) = P(A)\cdot P(B|A) + P(\bar{A})\cdot P(B|\bar{A})$
$P(A) = P(B\cap A) + P(\bar{B}\cap A) = P(B)\cdot P(A|B) + P(\bar{B})\cdot P(A|\bar{B})$

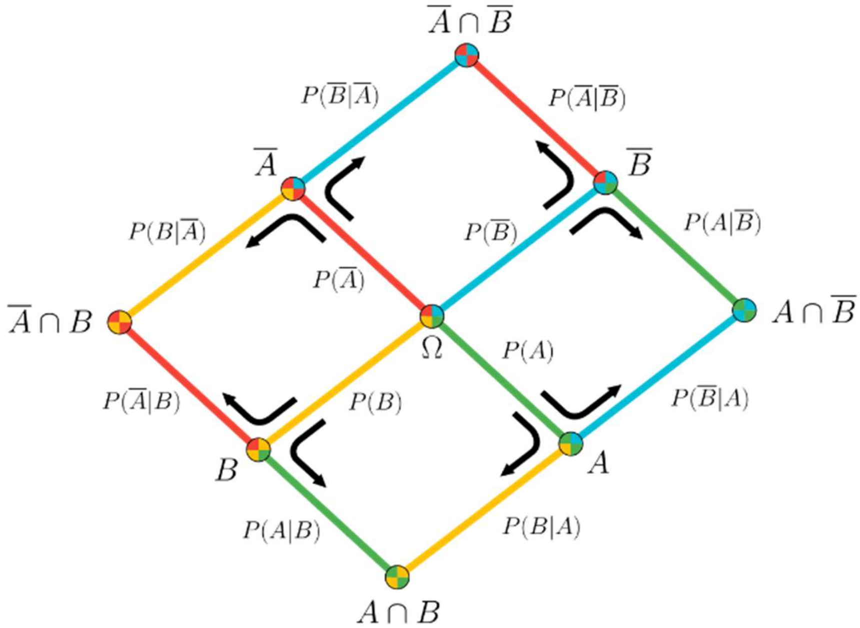
Ob man lieber mit Vierfeldertafeln oder mit Baumdiagrammen arbeitet oder beides kombiniert, ist letztlich Geschmackssache. Dazu noch einmal das vorige Beispiel:

$P(B^+ \mid A^+) = 0{,}999$; $P(B^- \mid A^-) = 0{,}995$ und $P(A^+) = 0{,}001$



Die an der jeweiligen Stelle bereits bekannten Zahlen sind grün markiert.
Ergebnis: $P(A^+ \mid B^+) = 0{,}1667$; $P(A^- \mid B^-) = 0{,}999999$

Noch ein schönes Bild zur ersten Pfadregel



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

https://www.bigdata-insider.de/so-verfeinert-das-bayes-theorem-spam-filter-und-mehr-a-720429/

Der **Satz von der totalen Wahrscheinlichkeit** und entsprechend der **Satz von Bayes** lassen sich ohne weitern Aufwand auf mehrere Ereignisse erweitern:

Ist Sind $A_1, A_2, \ldots, A_n$ paarweise disjunkt und ist $P(B) \neq 0$, so gilt

$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B \mid A_i)$

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B \mid A_i)} = \frac{P(A_k) \cdot P(B|A_k)}{P(B)}$$

Dazu nochmal ein

Beispiel aus der Medizin (https://www.mathe-online.at/mathint/wstat1/i.html):

Ein medizinisches Symptom S kann von zwei bekannten Krankheiten A und B hervorgerufen werden (A ist selten und gefährlich, B ist häufig und harmlos), kann aber auch bei gesunden Menschen (C) auftreten. Wenn das Symptom bei jemandem auftritt, möchte man wissen, mit welcher Wahrscheinlichkeit er welche Krankheit hat bzw. gesund ist. Epidemiologische und medizinische Untersuchungen studieren normalerweise nicht direkt diese Frage, sondern die Häufigkeit von Symptomen bei gegebenen Krankheiten

Kategorie	Auftreten von S mit Wahrscheinlichkeit
A	0.5
B	0.2
C	0.1

und die Erkrankungswahrscheinlichkeiten (A-priori-Wahrscheinlichkeiten)

Kategorie	Erkrankung erfolgt mit Wahrscheinlichkeit
A	0.01
B	0.15

Frage: Wie groß ist unter Zugrundelegung dieser Daten die Wahrscheinlichkeit, dass ein Mensch, der Symptom S ausbildet, an A bzw. B erkrankt ist bzw. gesund ist?

Um obige Formeln anwenden zu können, sehen wir zuerst, dass S die Rolle von B (in der Formel) und A, B, C die Rolle von A_1, A_2, A_3 (in der Formel) spielen. Gegeben sind also
$P(B|A_1) = 0{,}5; P(B|A_2) = 0{,}2; P(B|A_3) = 0{,}1$.
Und weiter $P(A_1) = 0{,}01; P(A_2) = 0{,}15$ und folglich $P(A_3) = 0{,}84$. Somit erhalten wir

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3) = 0{,}01 \cdot 0{,}5 + 0{,}15 \cdot 0{,}2 + 0{,}84 \cdot 0{,}1 = 0{,}119$$

$P(A_1|B) = \frac{P(A_1) \cdot P(B|A_1)}{P(B)} = \frac{0{,}01 \cdot 0{,}5}{0{,}119} = 4{,}2\%$ (= Wahrscheinlichkeit für die seltene und gefährliche Krankheit A beim Auftreten von Symptom S)

$P(A_2|B) = \frac{P(A_2) \cdot P(B|A_2)}{P(B)} = \frac{0{,}15 \cdot 0{,}2}{0{,}119} = 25{,}2\%$ und folglich $P(A_3|B) = 70{,}6\%$
(25,2% = Wahrscheinlichkeit für die häufige und harmlose Krankheit B beim Auftreten von Symptom S)

Ein Patient, bei dem das Symptom aufgetreten ist, braucht sich also keine übermäßigen Sorgen zu machen (obwohl eine Abklärung sinnvoll ist, denn von 1000 Patienten, bei denen das Symptom auftritt, werden ungefähr 40 tatsächlich an der schwereren Krankheit A erkrankt sein).

Bemerkungen:

- Vergleichen Sie den relativ kleinen Wert $P(A|S) \approx 0.042$ mit dem eher großen Wert $P(S|A) = 0.5$ aus der ersten Tabelle (der auf einen Menschen mit Symptom S einen bedrohlichen Eindruck machen mag). Wie ist dieser Unterschied zu erklären? Wie kommt er in der obigen Rechnung zustande?

- Wenn sich herausstellt, dass der Patient einer für eine Erkrankung an A anfälligen Risikogruppe angehört, so wirkt sich das auf die für ihn zutreffenden Apriori-Wahrscheinlichkeiten aus. Das ist eine gute Illustration der Tatsache, dass Wahrscheinlichkeiten unseren Kenntnisstand ausdrücken und sich mit diesem ändern können.
  Der Satz von Bayes eignet sich gut zur Anwendung in derartigen Domänen "unsicheren" Wissens. Ein weiteres − modernes − Anwendungsgebiet sind **Spam-Filter**, d.h. Computerprogramme, die unter den eingehenden E-mails eines Benutzers die Spam-Mails herausfinden sollen. Neben den Wahrscheinlichkeiten, dass Mails mit gewissen Charakteristika als Spam zu klassifizieren sind, werden die A-priori-Wahrscheinlichkeiten, dass der Benutzer solche Mails überhaupt bekommt, berücksichtigt. "Bayessche Spam-Filter" werten beispielsweise Worthäufigkeiten in bereits vom Benutzer erhaltenen und

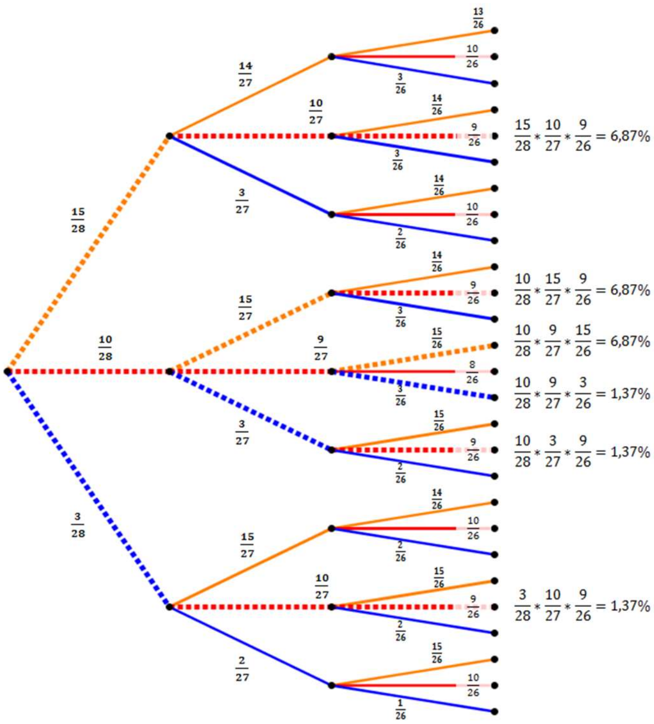
klassifizierten E-mails aus und sind damit "selbstlernende" Programme, die die verwendeten Apriori-Wahrscheinlichkeiten ständig verbessern.

Bemerkung:

Den allgemeinen **Satz von der totalen Wahrscheinlichkeit**, $P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|\, A_i)$, kann man auch im Sinne der **zweiten Pfadregel** (Die Wahrscheinlichkeit eines Ereignisses ist gleich der Summe der Wahrscheinlichkeiten aller Pfade, die zu diesem Ereignis führen.) deuten. Im entsprechenden Baumdiagramm gibt es dann auch $n -$ fache Verzweigungen an einem Knoten.
Auch die **Verzweigungsregel** (Die Summe der Wahrscheinlichkeiten an den Ästen, die von einem Knoten ausgehen, ist gleich eins.) behält ihre Gültigkeit: $\sum_{i=1}^{n} P(A_i|B) = 1$.
Auch für die **erste Pfadregel** (Die Wahrscheinlichkeit am Ende eines Pfades ist gleich dem Produkt der Wahrscheinlichkeiten längs dieses Pfades.) gibt es eine Verallgemeinerung für längere Pfade:

$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|\, A_1) \cdot P(A_3|\, A_1 \cap A_2) \cdot \ldots \cdot P(A_n|\, A_1 \cap A_2 \cap \ldots \cap A_{n-1})$ oder kurz $P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P\left(A_i|\bigcap_{j=1}^{i-1} A_j\right)$, wobei der erste Faktor als $P(A_1)$ zu lesen ist.

Beispiel (https://www.mathe-lerntipps.de/mehrstufige-zufallsexperimente/):
In einem Krug sind 15 orange 10 rote und 3 blaue Kugeln. Wir ziehen drei-mal hintereinander. Die Kugeln die wir ziehen behalten wir und legen sie nicht in den Krug zurück. Wie groß ist die Wahrscheinlichkeit, dass wir genau zwei rote Kugeln ziehen? (Die dritte Kugel kann also entweder blau oder orange sein)



Die 6 Pfade, welche zum gewünschten Ergebnis führen, sind gestrichelt. Die Wahrscheinlichkeiten jeweils am Ende dieser Pfade werden mit der ersten Pfadregel ermittelt und dann aufsummiert (zweite Pfadregel).
Ergebnis: 24,72%

Bemerkung:

Bei solchen „Baum-Aufgaben“ zeichnet man oft von vorneherein nur die relevanten Pfade ein, im obigen Beispiel die 6 gestrichelten. Für solche **reduzierten** Bäume gilt dann natürlich die Verzweigungsregel nicht mehr.

Generell ist zu diesem Aufgabentyp folgendes zu sagen: Immer, wenn es sich, wie hier, um Stichproben ohne Zurücklegen handelt, bei denen die Grundgesamtheit sich zerlegen lässt in paarweise disjunkte Teilmengen $\Omega = \bigcup_{i=1}^{r} \Omega_i$, $|\Omega| = N$, $|\Omega_i| = N_i$, also $\sum_{i=1}^{r} N_i = N$, und n Exemplare **ohne Zurücklegen** gezogen werden, dann hat das Ereignis, dass davon jeweils n_i Exemplare zu A_i gehören (also $\sum_{i=1}^{r} n_i = n$), die Wahrscheinlichkeit

$$\frac{\binom{N_1}{n_1} \cdot \binom{N_2}{n_2} \cdot \dots \cdot \binom{N_r}{n_r}}{\binom{N}{n}} .$$

Daraus bekommen wir als Spezialfall wieder das Laplace-Experiment A aus Kapitel 2.2 ($r = 2$). Damit können wir uns die vielen großen und bunten Bäume, wie sie so zahlreich im Internet zu finden sind, sparen (allenfalls taugen Sie als Erläuterung mancher Herleitung dieser Formel).
In unserem Beispiel wäre $r = 3$, $N_1 = 15$, $N_2 = 10$, $N_3 = 3$, und das untersuchte Ereignis ($n = 3$) könnte man zerlegen in zwei disjunkte Ereignisse:
Einmal mit $n_1 = 1, n_2 = 2, n_3 = 0$, und einmal mit $n_1 = 0, n_2 = 2, n_3 = 1$. Dann ergibt sich insgesamt

$$\frac{\binom{15}{1} \cdot \binom{10}{2} \cdot \binom{3}{0}}{\binom{28}{3}} + \frac{\binom{15}{0} \cdot \binom{10}{2} \cdot \binom{3}{1}}{\binom{28}{3}} = (15 + 3) \cdot \frac{10 \cdot 9}{2} \cdot \frac{3 \cdot 2}{28 \cdot 27 \cdot 26} = \frac{45}{182} \approx 27{,}725\%$$

Definition: Zwei Ereignisse A und B heißen **stochastisch unabhängig** genau dann, wenn
$P(A \cap B) = P(A) \cdot P(B)$

Bemerkungen

Die in der Definition gegebene Gleichung ist vielleicht die handlichste zum Nachprüfen; anschaulicher wird es mit dem Begriff der bedingten Wahrscheinlichkeit: Es gilt

$$\begin{aligned} & P(A \cap B) = P(A) \cdot P(B) \\ \Leftrightarrow \quad & P(A|B) = P(A) \\ \Leftrightarrow \quad & P(A\,|\,\bar{B}) = P(A) \\ \Leftrightarrow \quad & P(B|A) = P(B) \\ \Leftrightarrow \quad & P(B\,|\,\bar{A}) = P(B) \end{aligned}$$

Beweis → Vorlesung.
Genauso einfach ist (vgl. Übungsblatt)

$$P(A \cap B) = P(A) \cdot P(B) \quad \Leftrightarrow \quad P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B) \quad \Leftrightarrow \quad P(A \cap \bar{B}) = P(A) \cdot P(\bar{B})$$
$$\Leftrightarrow \quad P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B})$$

Sowohl bei einer Vierfeldertafel, als auch bei einem Baumdiagramm lässt sich leicht die Unabhängigkeit von Ereignissen überprüfen (→ Übunsblatt)

Definition: Die Ereignisse $A_1, A_2, \dots, A_n$ heißen **total unabhängig** genau dann, wenn für jede Teilmenge $T \subseteq [1{:}n]$, die mehr als ein Element enthält, gilt $P(\bigcap_{i \in T} A_i) = \prod_{i \in T} P(A_i)$

Bemerkung: Auf dem Übungsblatt finden Sie ein Beispiel dafür, dass im Allgemeinen paarweise Unabhängigkeit nicht hinreichend ist für totale Unabhängigkeit.

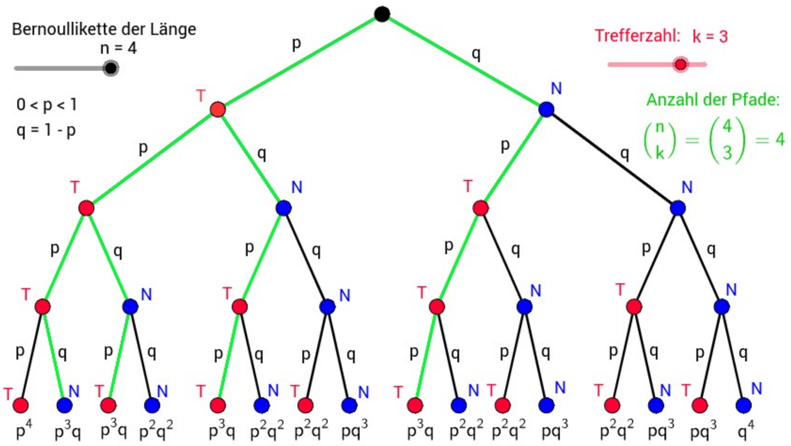
Voriges **Beispiel**, diesmal **mit** Zurücklegen:
Jetzt stehen an den Kanten (egal auf welcher Stufe) immer nur die Zahlen $\frac{15}{28}$ (bei rot), $\frac{10}{28}$ (bei orange) und $\frac{3}{28}$ (bei blau), denn weil zurückgelegt wird (→ Unabhängigkeit), wird nicht „runtergezählt". Ansonsten hat der Baum wieder das gleiche Aussehen. Auch die Pfadregeln gelten genauso. Am Ende der Pfade steht dann dreimal $\frac{10}{28} \cdot \frac{10}{28} \cdot \frac{15}{28} \approx 6{,}83\%$ und dreimal $\frac{10}{28} \cdot \frac{10}{28} \cdot \frac{3}{28} \approx 1{,}37\%$, was insgesamt eine Wahrscheinlichkeit von 24,6% ergibt.

Aber auch hier können wir uns den großen bunten Baum sparen: Wieder gehen wir aus von einer Zerlegung der Grundgesamtheit in paarweise disjunkte Teilmengen $\Omega = \bigcup_{i=1}^{r} \Omega_i$, $|\Omega| = N$, $|\Omega_i| = N_i$, also $\sum_{i=1}^{r} N_i = N$, und n Exemplare werden **mit Zurücklegen** gezogen. Dann hat das Ereignis, dass davon jeweils n_i Exemplare zu Ω_i gehören (also $\sum_{i=1}^{r} n_i = n$), die Wahrscheinlichkeit

$$\frac{n!}{n_1!n_2!\ldots n_r!} \cdot {p_1}^{n_1} \cdot {p_2}^{n_2} \cdot \ldots \cdot {p_l}^{n_r} ,$$

wobei $p_i := \frac{N_i}{N}$ für alle i

Bemerkung: Ausdrücke $\frac{n!}{n_1!n_2!\ldots n_r!}$ heißen auch **Multinomialkoeffizienten**, und lassen sich kombinatorisch deuten als die Anzahl aller Wörter der Länge n über dem Alphabet $\{A_1, A_2, \ldots, A_r\}$, wobei jeweils n_i mal der Buchstabe A_i vorkommt ($i = 1, \ldots, r$). Im Baum entspricht jeder relevante Weg genau einem solchen Wort und umgekehrt.
Als Spezialfall bekommen wir wieder das Laplace-Experiment B aus Kapitel 2.2 ($r = 2$). Solche Bernoulli-Ketten werden deshalb auch gerne an „Binärbäumen“ erklärt (https://www.geogebra.org/m/QmYFc74R):



Der Name Multinomialkoeffizienten kommt übrigens von der sogenannten **Multinomialformel**
$(\sum_{i=1}^{r} x_i)^n = \sum \frac{n!}{n_1!n_2!\ldots n_r!} \cdot {x_1}^{n_1} \cdot {x_2}^{n_2} \cdot \ldots \cdot {x_l}^{n_r}$, wobei sich die Summe über alle $(n_1, n_2, \ldots, n_r)$ erstreckt, für die $\sum_{i=1}^{r} n_i = n$ ist.