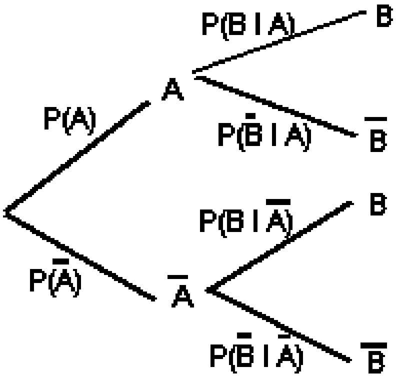
Prof. Dr. Dörte Haftendorn Mai 02



Es geht hier um den Zusammenhang von Baumdiagrammen und den in Statistik-Büchern gepriesenen "bedingten Wahrscheinlichkeiten", deren Behandlung von der "Bayes-Form" gekrönt wird. Als weiteres praktisches Werkzeug werden Vierfeldertafen gezeigt.

Bezeichnungen: $A, B, ...$ seien Ereignisse, $\overline{A}, \overline{B}, ...$ seien die Gegenereignisse.

$P(A) =$ Wahrscheinlichkeit, dass *A* (überhaupt) eintrifft, ohne weitere Bedingungen, "**a-priori**".

$P_A(B) = P(B|A) =$ Wahrscheinlichkeit, dass B eintritt, wenn *A* schon eingetreten ist, "**a-posteriori**". Diese Wahrscheinlichkeiten stehen natürlicherweise an den Ästen der 2. Stufe.

1. Pfadregel Das Ereignis "A und B" tritt mit der Wahrscheinlichkeit ein, die man durch Multiplizieren längs des Pfades Start-A-B erhält, in Formeln

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

Baumdiagramme haben keine Lobby, aber sie helfen sehr beim Verstehen.
Sofort überzeugend sind die 4-Felder-Tafeln, sie erledigen nicht alles, aber sie kommen doch recht selten vor.
Schick sehen dagegen die **Bayes-Formeln** aus, so richtig schön unverständlich.
Hier soll gezeigt werden, dass alle drei Werkzeuge sich gegenseitig erklären können, dass man insbesondere im Schulzusammenhang das meist unverstandene Anwenden der Bayes-Formeln wohl besser vermeidet. Wenn man sie denn aber nehmen will, kann man sie wie hier gezeigt erklären.
Oft kann man Wahrscheinlichkeiten besser verstehen, wenn man mit (fiktiven) Anzahlen denkt. Eine solche Gesamtzahl sei n. Dann ist $n\ P(A) =$ Anzahl der Fälle, in denen A insgesamt eintritt.
Parallel zur Theorie sei hier ein **Beispiel** betrachtet:

Aufgabe:

Die **Studis in Jeriwan** hören eine Statistikvorlesung, aber nur 10% machen die Übungsaufgaben. Von diesen Fleißigen fallen nur 5% durch die Klausur, während von je 50 Durchgefallenen nur einer zu diesen Fleißigen gehört.
Frage zur Aufgabe: Mit welcher Wahrscheinlichkeit fällt ein fauler Studi durch?
A= Studi macht Aufgaben, *B*= Studi besteht die Klausur,

	B	$\overline{B}$	
A	95	5	100
$\overline{A}$			900
			1000

Zunächst kann man mit fiktiven n=1000 die ersten beiden Angaben unterbringen (schwarz, 1000,100,5).

	B	$\overline{B}$	
A	95	5	100
$\overline{A}$	655	245	900
	750	250	1000

Zu den 5 durchgefallenen Fleißigen gehören dann nach der letzten Angabe 5 mal 50 =250 Durchgefallene überhaupt. (Lila, 250).

Jetzt lassen sich leicht die fehlenden Zahlen ergänzen (braun, 245, 750,655).

Mit einer Vierfeldertafel (nach Fisher) kann man auf einleuchtende Weise das Zusammenwirken zweier Ereignisse A und B betrachten. Bei Verwendung absoluter Zahlen müssen alle Zeilen- und Spaltensummen stimmen.
Das ist bei der Tafel mit den Prozentwerten auch so, aber man hat mehr Mühe, die Eintrage zu bestimmen. Man kann sie aus der absoluten Tafel aber erstellen und umgekehrt.

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

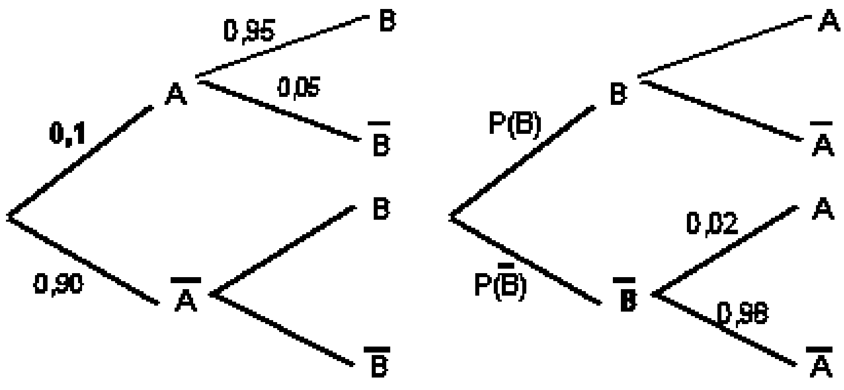
Nun kann man alle Verhältnisse berechnen:
$P(A)=\frac{100}{1000}=0{,}1=10\%$, $P(B|\overline{A})=\frac{655}{900}=0{,}728=72{,}8\%$ =W. ,dass ein Fauler besteht,
$P(A|B)=\frac{95}{750}=0{,}127=12{,}7\%$ =W, dass einer der bestanden hat, tatsächlich vorher Aufgaben gemacht hat.
Da nun unter denen, die bestanden haben, der Anteil der Fleißigen höher ist als unter den Studis insgesamt, kann man sagen, dass sich der Fleiß lohnt,
allgemein: die **Merkmale A und B sind "abhängig".**

$\boxed{P(A|B)=\dfrac{P(A\cap B)}{P(B)}}$ entspricht der letzten Berechnung,
und das ist schon die **1. Formel von Bayes**.

Aus der 4-Felder-Tafel kann man also alle vier bedingten Wahrscheinlichkeiten ausrechnen.
Wenn in den Spalten (oder den Zeilen) gleiche Verhältnisse vorliegen,
dann sind **A und B unabhängig**.
Also: wenn $P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{P(A)}{1}=P(A)$ und damit $\boldsymbol{P(A\cap B)=P(A)\cdot P(B)}$,

dann sind A und B unabhängig.

Behandlung der Aufgabe mit Baumdiagrammen



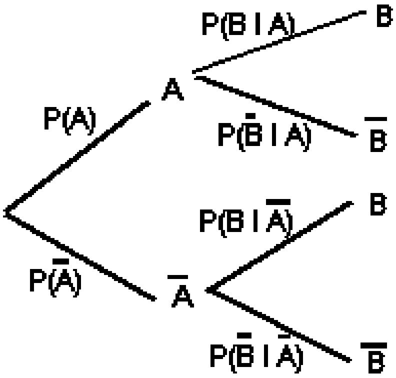
Diese beiden Baumdiagramme sind mit den Angaben der Aufgabe beschriftet.

Die 1. Pfadregel berechnet die W. der Und-Ereignisse.
Die Pfad-Wahrscheinlichkeit $P(A\cap\overline{B})$ muss in beiden Bäumen denselben Wert ergeben:

$P(A\cap\overline{B})=0{,}1\cdot 0{,}05=P(\overline{B})\cdot 0{,}02$ also folgt $P(\overline{B})=\dfrac{0{,}1\cdot 0{,}05}{0{,}02}=0{,}25$

Wahrscheinlichkeit, dass ein beliebiger Studi durchfällt.
Ebenso lassen sich jetzt alle anderen Wahrscheinlichkeiten berechnen.
Frage zur Aufgabe: Mit welcher Wahrscheinlichkeit fällt ein fauler Studi durch?
$0{,}9\cdot P(\overline{B}|\overline{A})=0{,}25\cdot 0{,}98$ Also $P(\overline{B}|\overline{A})=\frac{0{,}25\cdot 0{,}98}{0{,}90}=0{,}272=27{,}2\%$.
Da ein fleißiger Studi nur mit 5% Wahrscheinlichkeit durchfällt, ist das Risiko für Faule mehr als fünfmal so hoch.
Die allgemeinere Überlegung folgt auf der nächsten Seite.

Prof. Dr. Dörte Haftendorn Mai 02



$P(A \cap B) = P(A) \cdot P(B|A)$ 1. Pfadregel

Die Wahrscheinlichkeit eine Pfades errechnet man als das Produkt der Ast-Wahrscheinlichkeiten.

Gleichung von Bayes $P(B|A) = \frac{P(A \cap B)}{P(A)}$, eine Umformung davon.

2. Pfadregel Tragen mehrere Pfade zu dem betrachteten Ereignis bei, so muss man die Pfadwahrscheinlichkeiten addieren.

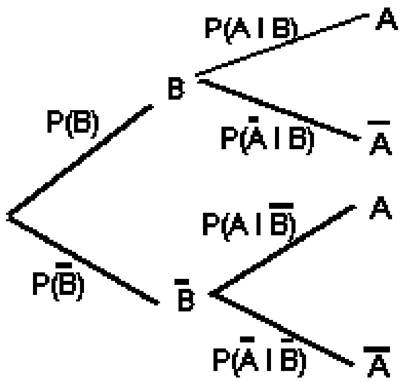
$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A})$

Allgemein, wenn n Ereignisse A_i die erste Stufe bilden:

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

und dieses ist gleichzeitig der

Satz von der totalen Wahrscheinlichkeit



Wieder gebe es n Ereignisse A_i, die zusammen alle Fälle dieser Stufe des mehrstufigen Zufallsversuches ausmachen, vornehm: $\Omega = \bigcup_{i=1}^{n} A_i$ (Bei A und $\overline{A}$ ist das von selbst erfüllt.)

Leitfrage: **Das Ereignis B kann die Ursachen $A_1, A_2, ...A_n$ haben. Nun ist B eingetreten. Mit welcher Wahrscheinlichkeit war dann das spezielle A_k die Ursache?** $P(A_k|B) = ?$

Wie schon auf Seite 1 im Beispiel kann man $P(A_k \cap B)$ auch noch mit dem umgekehrten Baum berechnen.

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

Division durch $P(B)$ ergibt mit der 2. Pfadregel (= mit dem Satz von der totalen Wahrscheinlichkeit):

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$ den Satz von Bayes

	B	$\overline{B}$	
A_1	$n\,P(A_1 \cap B)$	$n\,P(A_1 \cap \overline{B})$	$n\,P(A_1)$
A_2	$n\,P(A_2 \cap B)$	$n\,P(A_2 \cap \overline{B})$	$nP(A_2)$
A_3	$n\,P(A_3 \cap B)$	$n\,P(A_3 \cap \overline{B})$	$n\,P(A_3)$
	$n\,P(B)$	$n\,P(\overline{B})$	n

Also: Mache Baumdiagramme oder Mehr-Felder-Tafeln und denk' damit nach.
Am weitesten reichen Baumdiagramme, denn sie können leicht auch mehr als zwei Stufen haben.
Vermeide das Augenpulver der Bayes-Formeln (besonders wenn es um Lernen und Verstehen geht)!

Die bedingten Wahrscheinlichkeiten ergeben sich als Anteile

$P(B|A_k) = \frac{P(A_k \cap B)}{P(A_k)}$ für die Zeilen und $P(A_k|B) = \frac{P(A_k \cap B)}{P(B)}$ für die Spalten.

Dabei müssen natürlich die Zeilen- und Spaltensummen stimmen. Diese Tatsache drückt sich dann gerade in der Gleichung $n\,P(B) = n(\sum_{i=1}^{n} P(A_i \cap B) = n \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)$ aus und das ist (nach Kürzung von n) nichts anderes als der **"Satz von der totalen Wahrscheinlichkeit".**