Dipl.-Kfm. Jörg Petermann Unternehmensberatung

Benutzerhinweise für den PU Bayes-Korrektor V1.0

Trier, 20. Juli 2007

1 Bayes' Theorem

Die Wahrscheinlichkeitsrechnung hält immer wieder Überraschungen bereit. Eine davon hat Thomas Bayes (Engl. Mathematiker, ca. 1702 – 1761) beschrieben.

Sein Theorem gestattet, zunächst angenommene Wahrscheinlichkeiten im Licht einer neuen Information zu revidieren.

Angenommen, ich habe eine Annahme (oder sogar das Wissen) um das Auftreten eines bestimmten Ereignisses.

So sei die **Wahrscheinlichkeit p** (von probability), daß jemand AIDS habe in Deutschland

p(AIDS-infiziert) = 49.000 [Infizierte] / 82.000.000 [Bürger] = 0,000 597 561.

Die Wahrscheinlichkeit, daß ein willkürlich ausgewählter Bürger AIDS hat, beträgt demnach also ca. p(AIDS-infiziert) = 0,06%.

Da ich diese Wahrscheinlichkeit im Vorhinein kenne, nennt man sie **Apriori-Wahrscheinlichkeit**. (Korrekt eigentlich a priori von lat. „vom früheren her“).

Entscheidungstheoretisch beschreibt diese Apriori-Wahrscheinlichkeit einen **Umweltzustand**. In der Entscheidungstheorie werden Umweltzustände mit „s“ für „Szenario“ benannt.

Wir können also sagen:

p(s1) [Bürger AIDS-infiziert] = 0,000 597 561

p(s2) [Bürger nicht AIDS-infiziert] = 1 – 0,000 597 561 = **0,999 402 439**

Nun schicken wir einen willkürlich ausgewählten Bürger zum AIDS-Test. Er hat keinen Grund anzunehmen, daß die Wahrscheinlichkeit, daß er AIDS hätte, für ihn gesteigert wäre, etwa weil er gebrauchte Spritzen benutzt hätte o.ä.

Wir fragen den Arzt, der den Test machen soll, wie zuverlässig dieser Test sei.

Dieser sagt: *„Unter all jenen, die AIDS haben, schlägt der Test bei 99,5% aller untersuchten positiv an.“*

Folglich werden 0,5% der AIDS-infizierten durch diesen Test nicht entdeckt.

Außerdem habe der Test eine gewisse Irrtumswahrscheinlichkeit: *„Unter den NICHT mit AIDS infizierten werden 1% irrtümlich als positiv ausgewiesen. Die restlichen 99% würden dagegen korrekt als negativ diagnostiziert.“*

Die Beobachtung, die wir machen, wollen wir mit **y** beschreiben. Wir können sagen:

y1: Der Test ist positiv

y2: Der Test ist negativ

Aus der Beschreibung des Arztes können wir folgende Wahrscheinlichkeiten formulieren:

- p (**Test positiv** unter der Bedingung **AIDS-infiziert**) = 99,5%
- p (**Test negativ** unter der Bedingung **AIDS-infiziert**) = 0,5%
- p (**Test positiv** unter der Bedingung **NICHT AIDS-infiziert**) = 1%
- p (**Test negativ** unter der Bedingung **NICHT AIDS-infiziert**) = 99%

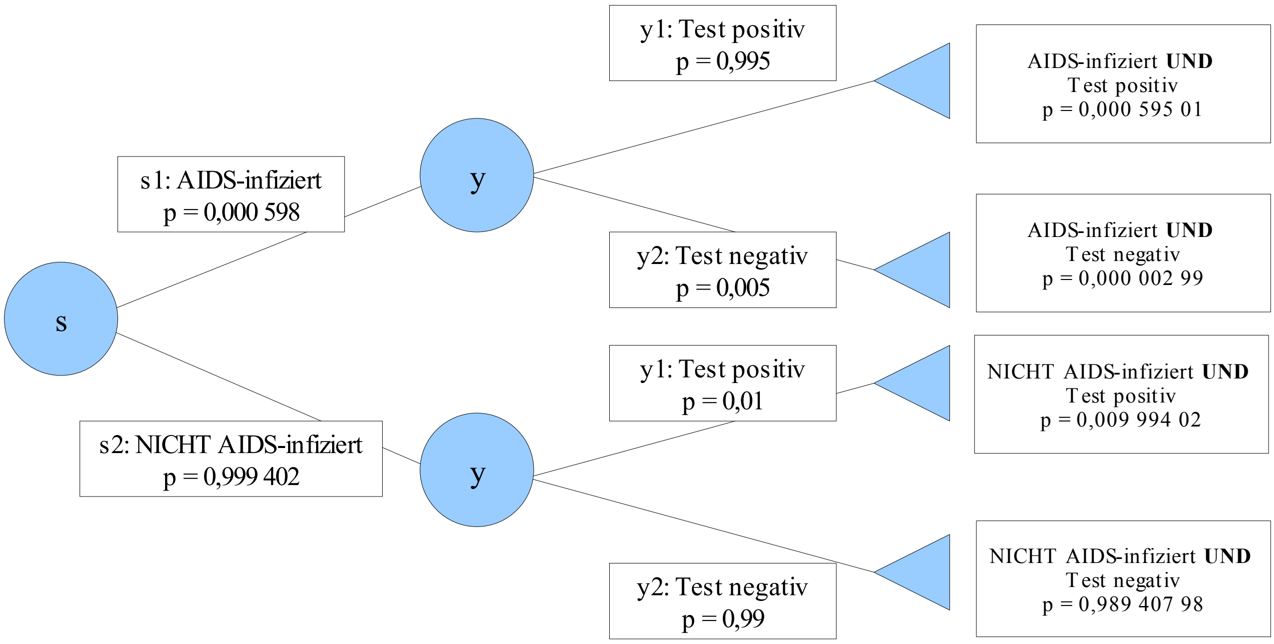
Mathematisch: (Der Strich „|" liest sich „unter der Bedingung")

- p(y1|s1) = 0,995
- p(y2|s1) = 0,005
- p(y1|s2) = 0,01
- p(y2|s2) = 0,99

Als Matrix:

	y1: Test positiv	y2: Test negativ	Summe
s1: AIDS-infiziert (p = 0,06%)	99,5%	0,5%	100%
s2: NICHT AIDS-infiziert (p = 99,94%)	1%	99%	100%

Als Entscheidungsbaum:



Der Test des von uns willkürlich ausgewählten Bürgers ist positiv! Dieser gerät sofort in Panik. *„Mit 99,5% Wahrscheinlichkeit habe ich AIDS! Das ist ja so gut wie sicher!“*

„Ganz ruhig“, sagen Sie. Schließlich haben sie den Bayes-Korrektor.

Wo steckt der Fehler des Bürgers?

Er hat sich von der Treffsicherheit des Tests beeindrucken lassen, ohne die Apriori-Wahrscheinlichkeit zu bedenken. Denn die Voraussetzung dafür, daß der Wert von 99,5% stimmt ist (wie man im Entscheidungsbaum sieht), daß der Bürger bereits **definitiv mit AIDS infiziert** ist. Aber genau das wissen wir ja noch gar nicht. Und die Wahrscheinlichkeit dafür ist mit < 0,06% sehr gering.

Uns interessiert etwas völlig anderes:

Nicht:

Wie hoch ist die Wahrscheinlichkeit, **daß der Test positiv ist, wenn jemand AIDS hat**,

Sondern genau umgekehrt:

Wie hoch ist die Wahrscheinlichkeit, **daß jemand AIDS hat, wenn der Test positiv ist**.

Dazu müssen wir wissen, wie hoch überhaupt die Wahrscheinlichkeit ist, daß der Test positiv ausfällt. Dies entspricht der Summe der gemeinsamen Wahrscheinlichkeiten für alle Fälle, in denen der Test positiv ist, egal ob zutreffend oder nicht.

Zwei Fälle kommen in Frage:

	Test positiv UND AIDS-infiziert	p =	0,000 595 01
+	Test positiv UND NICHT AIDS-infiziert	p =	0,009 994 02
=	Test positiv	Σp =	0,010 589 03

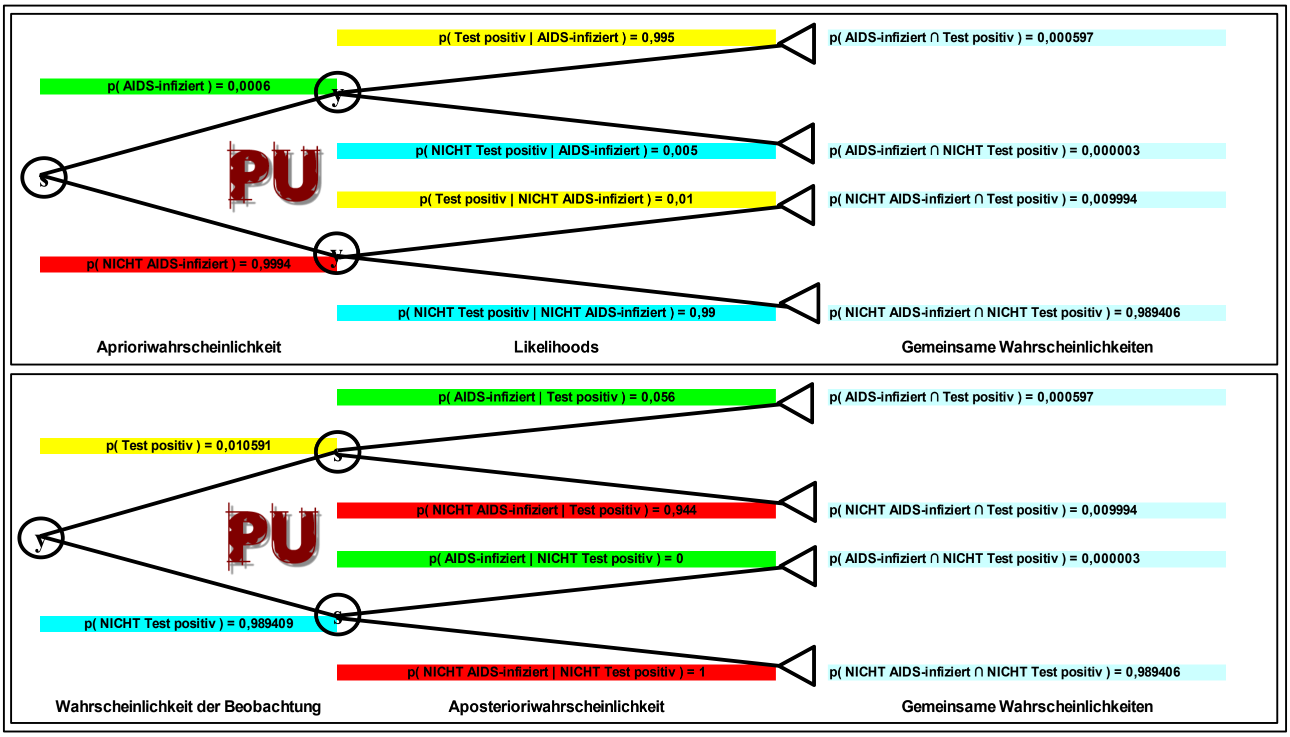
Um nun zu ermitteln, wie groß die Wahrscheinlichkeit dafür ist, daß unser Bürger tatsächlich AIDS-infiziert ist, müssen wir die Wahrscheinlichkeit für das Szenario:

Test positiv UND AIDS-infiziert durch die Summe der Wahrscheinlichkeiten dafür teilen, daß der Test positiv ausfällt, also:

	Test positiv UND AIDS-infiziert	p =	0,000 595 01
/	Test positiv (s.o.)	p =	0,010 589 03
=	p (AIDS-infiziert | Test positiv)	p =	0,056 191 17

Wir können den Bürger also beruhigen. Die Wahrscheinlichkeit, daß er tatsächlich AIDS-infiziert ist, liegt deutlich unter 6%.

Überprüfen Sie es mit dem Bayes-Korrektor!



2 Mathematisches

Seien die Apriori-Wahrscheinlichkeiten $p(s_i)$ für mehrere Umweltzustände $s \in S$ bekannt.

Erhalten wir nun eine Information $y \in Y$.

Es seien ferner die bedingten Wahrscheinlichkeiten („Likelihoods") für das Beobachten von y_j bekannt, wenn der Umweltzustand s_i der wahre Umweltzustand ist, $p(y_j|s_i)$.

Dann lassen sich die gemeinsamen Wahrscheinlichkeiten errechnen aus:

$p(y_j \cap s_i) = p(s_i) * p(y_j|s_i)$

Durch Division der gemeinsamen Wahrscheinlichkeit $p(y_j \cap s_i)$ durch die Wahrscheinlichkeit des Auftretens von y_j können wir $p(s_i|y_j)$ ermitteln.

$p(y_j \cap s_i) / p(y_j) = p(s_i |y_j)$

$p(y_j)$ ergibt sich aus der Summe aller Fälle, in denen y_j beobachtet wird, $p(y_j) = \Sigma_i\, p(y_j \cap s_i)$.

$$p(s_i|y_j) = \frac{p(s_i \cap y_j)}{\sum_i (y_i)} = \frac{p(s_i) \cdot p(y_j|s_i)}{\sum_i p(s_i) \cdot p(y_j|s_i)}$$

3 Weitere Anwendungsfälle

3.1 Lügendetektor

Der Täter T wird verdächtigt, ein Verbrechen begangen zu haben. Die Apriori-Wahrscheinlichkeit, daß er es war, schätzt der ermittelnde Beamte auf 0,7 (=70%).

Nun wird ein Lügendetektor-Test durchgeführt. Es gelten die Likelihoods:

p(Test positiv | Schuldig) = 0,9	p(Test positiv | Unschuldig) = 0,2
p(Test negativ | Schuldig) = 0,1	p(Test negativ | Unschuldig) = 0,8

Die Wahrscheinlichkeit, daß er schuldig ist, wenn der Test positiv anschlägt, beträgt 91,3%.