Vierfeldertafel, Baumdiagramm und der AIDS Test

Tests auf das Vorliegen einer Erkrankung haben natürlich eine „Fehlerwahrscheinlichkeit“ – sie können eine positive Testreaktion aufweisen, obwohl die Person gesund ist, oder keine positive Testreaktion zeigen, obwohl die Testperson krank ist!

GIB AIDS KEINE CHANCE

Bei den heutigen AIDS Schnelltests ist die Wahrscheinlichkeit für eine positive Testreaktion bei Infizierten 99,9%=0,999 (die sog. „Sensitivität“ des Tests). Nur 0,1%=0,001 der Infizierten werden also irrtümlich negativ getestet. Die Wahrscheinlichkeit, dass ein Nichtinfizierter tatsächlich auch negativ getestet wird, liegt immerhin bei 99,7% =0,997 (die sog. „Spezifität“ des Tests). Bei 0,3%=0,003 der Nichtinfizierten ist der Test also irrtümlich positiv. Man kann davon ausgehen, dass aktuell 0,1%=0,001 der deutschen Bevölkerung HIV infiziert sind.

Aufgaben

1. Stelle mit diesen Angaben das Baumdiagramm mit dem Merkmal Infektion/keine Infektion auf der **ersten** Stufe und Test (+ oder -) auf der **zweiten** Stufe auf!

Lösung: Es macht zuerst Sinn, sich klarzumachen, welche Bedeutung die Zahlen aus dem Text haben. Nämlich: $P_I(+)=0{,}999$ (sprich: „Wahrscheinlichkeit positiv getestet zu werden, **wenn** man infiziert ist), $P_I(-)=0{,}001$, $P_N(+)=0{,}003$ und $P_N(-)=0{,}997$. Zusätzlich: $P(I)=0{,}001$ und $P(N)=0{,}999$. Dies sind aber gerade die Teilwahrscheinlichkeiten aus dem Baumdiagramm (siehe Abb. 1).

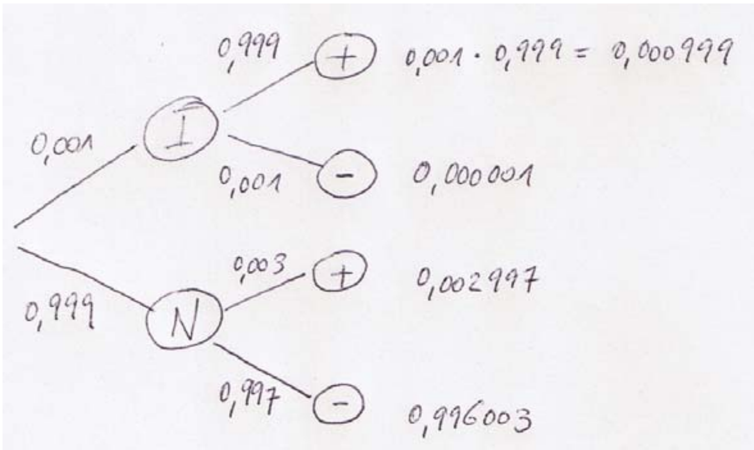


Abbildung 1 Baumdiagramm mit dem Merkmal Infiziert (I) und Nichtinfiziert (N) auf der ersten Stufe

2. Leite aus dem Baumdiagramm die Vierfeldertafel her. Verwende am besten die Dezimalpunktschreibweise (0,999 statt 99,9%). Runde die Zahlen nicht, denn einige Wahrscheinlichkeiten sind so klein, dass man tatsächlich 6 Nachkommastellen braucht! Vergiss die „Gesamt-Spalten“ nicht.

Lösung: Die Pfadwahrscheinlichkeiten (= Produkt der Teilwahrscheinlichkeiten) aus Auf. 1 sind die Einträge der Vierfeldertafel. Sprich: Die Wahrscheinlichkeit, dass man unter allen Menschen einen findet, auf den die beiden Merkmale zutreffen.

	+	-	gesamt
I	0,000999	0,000001	0,001
N	0,002997	0,996003	0,999
gesamt	0,003996	0,996004	1

3. Finde nun auch das „**zweite** Baumdiagramm“, d.h. das mit **vertauschten** Stufen!

Lösung: Zuerst müssen wir die Teilwahrscheinlichkeiten für die erste Stufe finden. Man kann sie aus der „Gesamt-Spalte“ für „+“ und „-“ (0,003996 und 0,996004) ablesen. Die Pfadwahrscheinlichkeiten kennt man ebenfalls schon – sie ändern sich nämlich nicht. Dann ist aber auch klar, was auf der zweiten Stufe steht: Einfach die Pfadwahrscheinlichkeiten durch die Wahrscheinlichkeit der ersten Stufe teilen (siehe Abb. 2).

4. Vergleiche die Wahrscheinlichkeit, tatsächlich infiziert zu sein, wenn man positiv getestet wurde, mit der „Umgekehrten“, d.h. positiv getestet zu werden, wenn man infiziert ist.

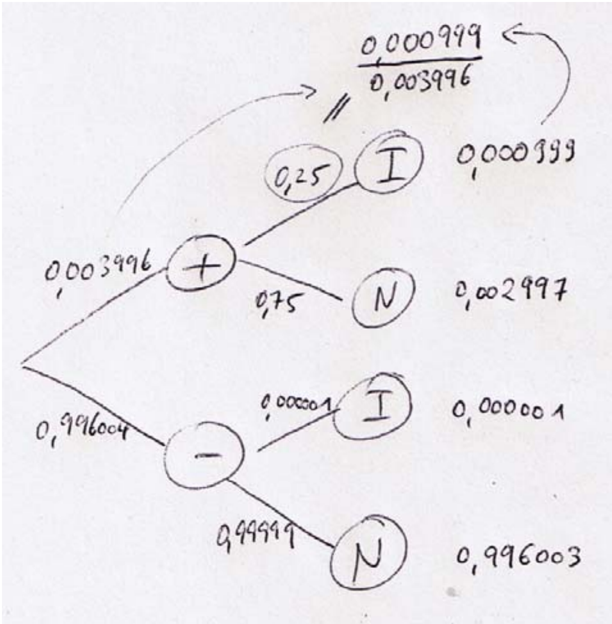


Abbildung 2 Baumdiagramm mit dem Testergebnis auf der ersten Stufe

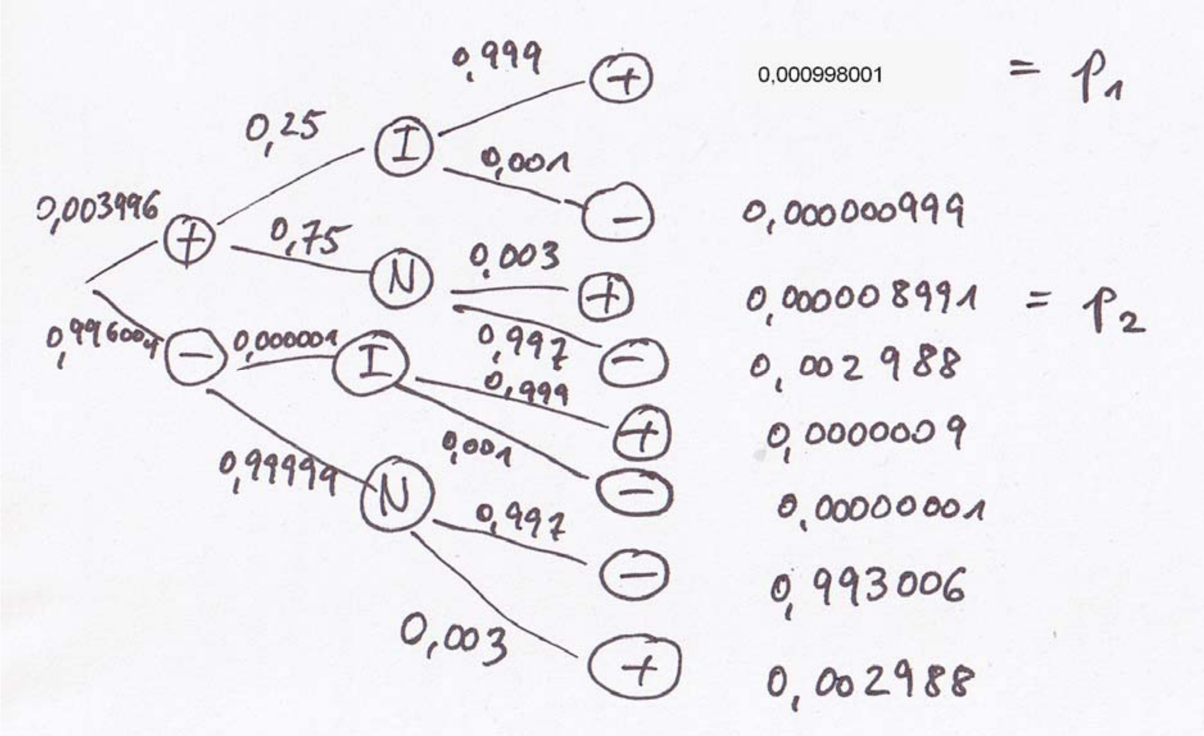
Lösung: Aus dem obigen Baum liest man ab: $P_+(I)=0,25=25\%$. In Worten: die Wahrscheinlichkeit, infiziert zu sein, **wenn** man positiv getestet wurde, *liegt nur bei 25%!!!* Die „umgekehrte“ Wahrscheinlichkeit stand schon im Text: $P_I(+)=0,999$ (sprich: „Wahrscheinlichkeit positiv getestet zu werden, **wenn** man infiziert ist). Diese beiden ähnlich klingenden Aussagen dürfen also nicht verwechselt werden.

Kann man dieses verblüffende Ergebnis eigentlich anschaulich verstehen? Warum ist das Testergebnis so wenig zuverlässig? Die Antwort ist eigentlich einfach: der Test macht zwar nur in 3 Promille der Fälle den Fehler, einen Nichtinfizierten positiv zu testen, aber es gibt (zum Glück) so viele Nichtinfizierte, dass diese kleine Wahrscheinlichkeit trotzdem zu einem nennenswerten Effekt führt.

Zusatzaufgabe: Bei einem einmaligen Test sind also 3 von 4 Ergebnissen falsch! Das klingt etwas entmutigend. Zum Glück ist bei einem zweiten Test die Verlässlichkeit deutlich höher. Die „wirklich“ infizierten werden wieder positiv getestet werden, die anderen kein (bzw. genauer: „kaum ein“) zweites Mal „Pech“ haben. Konkret: **Mit welcher Wahrscheinlichkeit ist man infiziert, wenn man zwei Mal positiv getestet wurde?**

1. Lösung: Stellen wir uns 1000 Personen mit einem positiven Test vor. Im Schnitt ist er bei 750 irrtümlich. Also gilt bei einem zweiten Test: $0,003\cdot 750=2,25$ Personen werden erneut irrtümlich positiv getestet. Außerdem: $250 \cdot 0,999=249,75$ werden erneut korrekt positiv getestet. Insgesamt haben wir also $249,75+2,25=252$ positive Tests. Macht einen Bruchteil von $249,75/252\approx 0,991$ von **Infizierten**, die zweimal positiv getestet wurden und einen Bruchteil von $2,25/252\approx 0,009$ von **Nichtinfizierten**, die zwei Mal positiv gestestet werden. Symbolisch $P_{++}(I)=0,991$ bzw. $P_{++}(N)=0,009$. Das ist erfreulich (zur Erinnerung: $P_+(N)=0,75$!). Der zweite Test drückt die Irrtumswahrscheinlichkeit also von 75% auf unter 1 Prozent (Ein dritter Test würde das Ergebnis schließlich über jeden vernünftigen Zweifel erhaben machen).

2. Lösung: Die Zusatzaufgabe haben wir nun „ohne ein Baumdiagramm" gelöst. Natürlich kann man das auch mit dieser Methode machen. Dazu muss der Baum aus Abbildung 2 einfach weitergezeichnet werden. Siehe die folgende Abbildung. An die dritte Stufe trägt man einfach die Teilwahrscheinlichkeiten aus dem Text (bzw. der Abb. 1) ein. Man kennt schließlich die Wahrscheinlichkeit, mit der ein (Nicht-)Infizierter positiv(oder negativ) getestet wird:



Wie berechnet man damit die gesuchte Irrtumswahrscheinlichkeit? Nun, man erkennt, dass es zwei Pfade gibt, bei denen es beide Male zu positiven Tests kommt. Bei p_1 ist ein Infizierter betroffen und bei p_2 ein Nichtinfizierter. Insgesamt also bei einem Bruchteil von $p_1+p_2\approx 0{,}001006992$. Mit $p_1/(p_1+p_2)\approx 0{,}991=99{,}1\%$ berechnet man also den Bruchteil der Infizierten, die doppelt positiv getestet wurden. Mit $p_2/(p_1+p_2)\approx 0{,}009=0{,}9\%$ den Bruchteil der Nichtinfizierten mit doppelt positiven Tests. Hinweis: Bei dieser Rechnerei muss man viele Nachkommastellen mitnehmen, obwohl die Genauigkeit der Ausgangsangaben natürlich gar nicht so hoch ist…