Bayes Rule: How it can make you a better thinker

Giri Iyengar

Teaser

- You take a disease diagnostic test for cancer that is 99% sensitive & 99% specific*
- Test comes back positive... How likely are you to have cancer?
- SENSITIVITY: Probability that you'll test positive if you have the disease
- SPECIFICITY: Probability that you'll test negative if you don't have it
- *Doctor tells you that it is rare but fatal cancer. Its prevalence is 0.1% in the general population)

SCIENCE | The Odds, Continually Updated

The Odds, Continually Updated

By F. D. FLAM SEPT. 29, 2014

Bayesian statistics can help solve the Monty Hall problem of winning a car. Fred Westbrook Collection/GSN

21 March 2014 Last updated at 20:25 ET

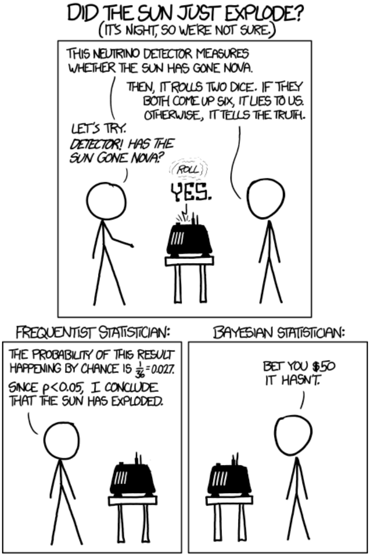
Share

MH370 Malaysia plane: How maths helped find an earlier crash

Members of the Brazilian Frigate Constituicao recovering debris in June 2009

Statisticians helped locate an Air France plane in 2011 which was missing for two years. Could mathematical techniques inspired by an 18th Century Presbyterian minister be used to locate the mysterious disappearance of Malaysia Airlines Flight MH370?

In June 2009, Air France flight 447 went missing flying from Rio de Janeiro in Brazil to Paris, France.


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

In praise of Bayes | The Economist

www.economist.com/node/382968

nePins ▾ Science Club ▾ Machine Lea...ng (Theory) Metamarkets ...rning Fails PHP on Trax Blog Saved Tabs

The Economist | World politics | Business & finance | Economics | Science & technology | Culture

In praise of Bayes

Bayesianism is a controversial but increasingly popular approach to statistics that offers many benefits—although not everyone is persuaded of its validity

Sep 28th 2000 | From the print edition

Like 40 Tweet 0

IT IS not often that a man born 300 years ago suddenly springs back to life. But that is what has happened to the Reverend Thomas Bayes, an 18th-century Presbyterian minister and mathematician—in spirit, at least, if not in body. Over the past decade the value of a statistical method outlined by Bayes in a paper first published in 1763 has become increasingly apparent and has resulted in a blossoming of "Bayesian" methods in scientific fields ranging from archaeology to computing. Bayes's fans have restored his tomb and posted pictures of it on the Internet, and a celebratory bash is planned for next year to mark the 300th anniversary of his birth. There is even a Bayes songbook—though, since Bayesians are an academic bunch, it is available only in the obscure file formats that are used for scientific papers.

Contingency Table

Disease	Total	Test +ve	Test -ve
Yes	100	99	1
No	99900	999	98901

Likelihood of you having the fatal cancer, after testing positive is $99/(99+999) \sim 9\%$

Back to the Economist article

"The essence of the Bayesian approach is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. In other words, it allows scientists to combine new data with their existing knowledge or expertise. The canonical example is to imagine that a precocious newborn observes his first sunset, and wonders whether the sun will rise again or not. He assigns equal prior probabilities to both possible outcomes, and represents this by placing one white and one black marble into a bag. The following day, when the sun rises, the child places another white marble in the bag. The probability that a marble plucked randomly from the bag will be white (ie, the child's degree of belief in future sunrises) has thus gone from a half to two-thirds. After sunrise the next day, the child adds another white marble, and the probability (and thus the degree of belief) goes from two-thirds to three-quarters. And so on. Gradually, the initial belief that the sun is just as likely as not to rise each morning is modified to become a near-certainty that the sun will always rise."

Bayes Rule

- Rule for incorporating new evidence into your belief about events
- Mathematically:
- $P(D|e) = P(e|D)P(D)/P(e)$ where $P(D)$ is the probability of having cancer.
- $P(e|D)$ = probability of testing +ve when you have the disease
- $P(e)$ = probability of testing +ve whether you have the disease or not.
- Using Total Probability Rule, it is $P(e|D) + P(e|!D)$

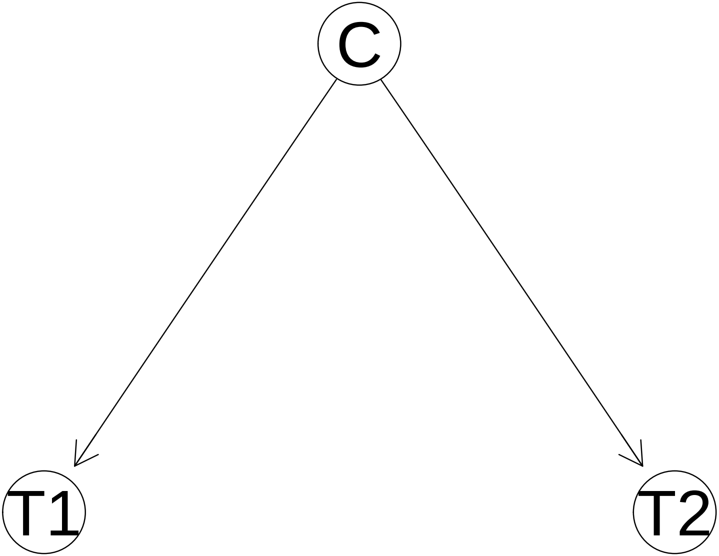
Bayesian Networks

- A type of model that provides you with a way of encoding relationships between random variables of interest (e.g. Cancer & various tests for detecting it)
- The structure of the network suggests a set of independence and dependence relationships between random variables

A Simple two-test Cancer network

```
#source("http://bioconductor.org/biocLite.R")
#biocLite(c("gRain","Rgraphviz"))
suppressWarnings(suppressMessages(library(Rgraphviz)))
suppressWarnings(suppressMessages(library(gRain)))
yn <- c("yes","no")
# specify the Conditional Probability Tables
node.C  <- cptable(~ C, values=c(1, 99), levels=yn)
node.T1 <- cptable(~ T1 + C, values=c(9,1,2,8), levels=yn)
node.T2 <- cptable(~ T2 + C, values=c(9,1,2,8), levels=yn)
plist <- compileCPT(list(node.C, node.T1, node.T2))
bn.cancer <- grain(plist)
```

$$P(C, T1, T2) = P(C) \times P(T1|C) \times P(T2|C)$$


C
T1
T2

Compute some probabilities

```
querygrain(bn.cancer, nodes=c("C","T1"), type="joint")
```

```
##       T1
## C         yes    no
##   yes 0.009 0.001
##   no  0.198 0.792
```

Provide some evidence from the tests

```
bn.cancer.1 <- setFinding(bn.cancer,nodes=c("T1"), states=c("y
querygrain(bn.cancer.1, nodes=c("C"))
```

```
## $C
## C
##        yes         no
## 0.04347826 0.95652174
```

```
bn.cancer.2 <- setFinding(bn.cancer.1,nodes=c("T2"), states=c(
querygrain(bn.cancer.2, nodes=c("C"))
```

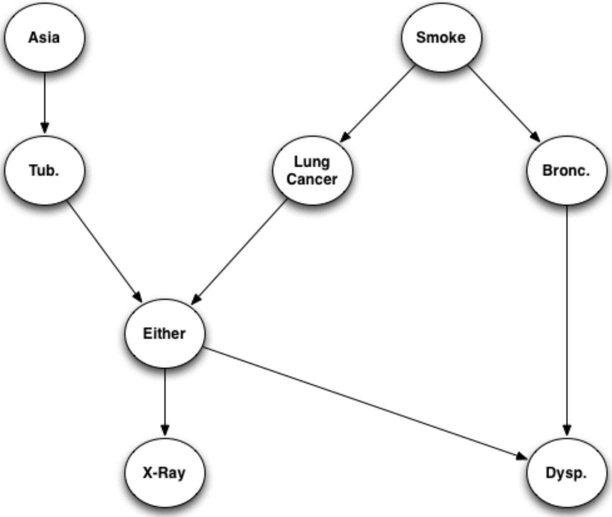
```
## $C
## C
##       yes        no
## 0.1698113 0.8301887
```

A more complex example

Lauritzen and Spiegelhalter chest clinic example

"Shortness–of–breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X–ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea."

Chest Clinic example as a Bayesian Network



Specify the conditional probabilities

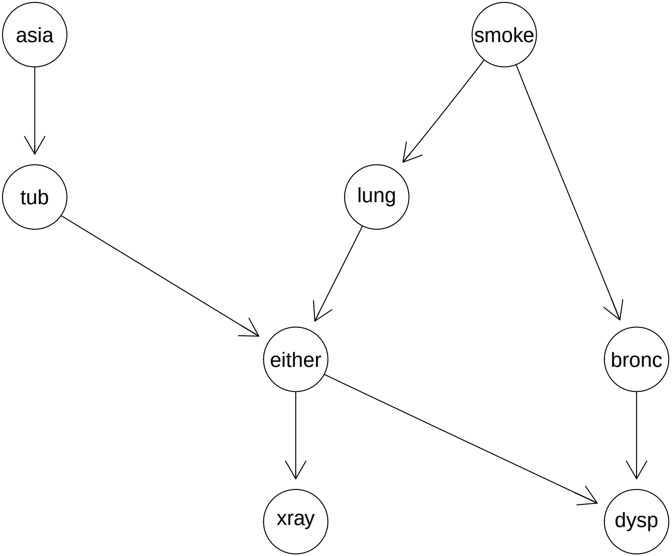
```
yn <- c("yes","no")
a <- cptable(~asia, values=c(1,99),levels=yn)
t.a <- cptable(~tub|asia, values=c(5,95,1,99),levels=yn)
s <- cptable(~smoke, values=c(5,5), levels=yn)
l.s <- cptable(~lung|smoke, values=c(1,9,1,99), levels=yn)
b.s <- cptable(~bronc|smoke, values=c(6,4,3,7), levels=yn)
e.lt <- ortable(~either|lung:tub, levels=yn)
x.e <- cptable(~xray|either, values=c(98,2,5,95), levels=yn)
d.be <- cptable(~dysp|bronc:either, values=c(9,1,7,3,8,2,1,9),
```

Compile the list of conditional probabilities and build the network

```
plist <- compileCPT(list(a, t.a, s, l.s, b.s, e.lt, x.e, d.be
plist
```

```
## CPTspec with probabilities:
##  P( asia )
##  P( tub | asia )
##  P( smoke )
##  P( lung | smoke )
##  P( bronc | smoke )
##  P( either | lung tub )
##  P( xray | either )
##  P( dysp | bronc either )
```

Chest Network



Query the marginal probabilities of various diseases

```
(querygrain(chest_n, nodes=c("lung","bronc","tub"), type="marg
```

```
## $tub
## tub
##    yes     no
## 0.0104 0.9896
##
## $lung
## lung
##   yes    no
## 0.055 0.945
##
## $bronc
## bronc
##  yes   no
## 0.45 0.55
```

Provide some evidence and see how things change

- The patient visits the ER complaining of shortness of breath

```r
chest_n.1 <- setFinding(chest_n,nodes=c("dysp"), states=c("yes
(querygrain(chest_n.1, nodes=c("lung","bronc","tub"), type="ma
```

```
## $tub
## tub
##        yes         no
## 0.01884531 0.98115469
##
## $lung
## lung
##       yes        no
## 0.1027592 0.8972408
##
## $bronc
## bronc
##       yes        no
## 0.8339673 0.1660327
```

Add in more pieces of evidence

- The nurse finds out that the patient returned from Asia two weeks ago
- The chest X-Ray is not clean

```r
chest_n.2 <- setFinding(chest_n.1,nodes=c("asia","xray"), stat
(querygrain(chest_n.2, nodes=c("lung","bronc","tub"), type="ma
```

```
## $tub
## tub
##       yes        no
## 0.3917117 0.6082883
##
## $lung
## lung
##       yes        no
## 0.4442705 0.5557295
##
## $bronc
## bronc
##       yes        no
## 0.6288218 0.3711782
```