SOME EXERCISES IN BAYESIAN INFERENCE

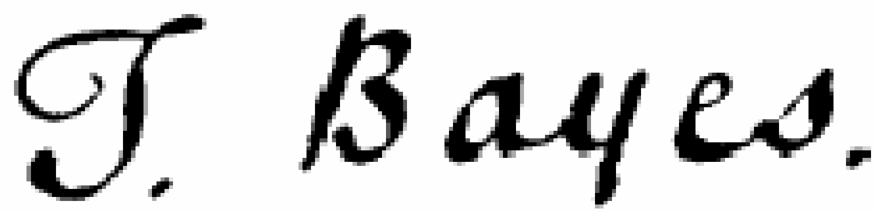
Borek Puza and Terry O’Neill

Australian National University
Canberra

Progic 2005

Thomas Bayes (1702-1761)

English Presbyterian minister and mathematician



Bayes' rule

prior probability of *A*

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

posterior probability of *A*

$$P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})$$

prior or unconditional probability of *B*

Example

1% of pop. has disease (*D*); rest is healthy (*H*)

90% of diseased persons test positive (+)

90% of healthy persons test negative (-)

Randomly selected person tests positive

Probability that person has disease is:

$$P(D \mid +) = \frac{P(D)P(+ \mid D)}{P(D)P(+ \mid D) + P(H)P(+ \mid H)}$$

$$= \frac{0.01 \times 0.9}{0.01 \times 0.9 + 0.99 \times 0.1} = \frac{0.009}{0.108} = \frac{1}{12}$$

Hypothetical population

1000

D 10 | H 990

D+ 9 | D- 1 | H+ 99 | H- 891

$\#(+) = 9 + 99 = 108$ $\#(D+) = 9$

$P(D|+) = \#(D+)/\#(+) = 9/108 = 1/12$

EXERCISE 1: TWO GUINEA PIGS

You have just met Ann, who has
2 baby guinea pigs born today

Each is equally likely to be a boy or girl

Find the probability p that both GP's are boys if:

(a) at least one is a boy
(b) the older one is a boy
(c) Ann tells you that the older one is a boy
(d) one was randomly picked & found to be a boy

(a) Sample space: $S = \{BB,BG,GB,GG\}$

At least one boy: $A = \{BB,BG,GB\}$

Two boys: BB

$p = P(BB \mid A) = 1/3$

Problem....

Suppose: $P(BB) = 1/6$ $P(BG) = 1/3$

$P(GB) = 1/3$ $P(GG) = 1/6$

Then: $P(B^*) = P(BB) + P(BG) = 1/2$

$P(G^*) = P(GB) + P(GG) = 1/2$

$P(^*B) = P(BB) + P(GB) = 1/2$

$P(^*G) = P(BG) + P(GG) = 1/2$

Thus each GP is equally likely to be a boy or a girl

But now....

$$p = P(BB \mid \overline{GG}) = \frac{P(BB)P(\overline{GG} \mid BB)}{P(\overline{GG})}$$

$$= \frac{(1/6)\times 1}{5/6} = \frac{1}{5}$$

If we assume BB, BG, GB, GG equally likely:

$$p = \frac{(1/4)\times 1}{3/4} = \frac{1}{3}$$ as before

(b) Older GP is a boy: B* = {BB, BG}

Both are boys: BB

So P(BB) = 1/2

Or....

$$p = P(BB \mid B^*) = \frac{P(BB)P(B^* \mid BB)}{P(B^*)} = \frac{(1/4)\times 1}{1/2} = \frac{1}{2}$$

NB: If P(BB) = 1/6, etc, then $p = \dfrac{(1/6)\times 1}{1/2} = \dfrac{1}{3}$

(c) Let T = “Ann tells you her older GP is a boy”

(Assume she’s not lying, has not erred, and

$$P(BB) = P(BG) = P(GB) = P(GG) = 1/4 \text{)}$$

Then

$$P(BB \mid T) = \frac{P(BB)P(T \mid BB)}{P(T)}$$

where

$$P(T) = P(BB)P(T \mid BB) + P(BG)P(T \mid BG) + P(GB)P(T \mid GB) + P(GG)P(T \mid GG)$$

$$P(BB \mid T) = \frac{P(T \mid BB)}{P(T \mid BB) + P(T \mid BG)} = \frac{1}{2}$$

But... we have assumed that

$$P(T \mid BB) = P(T \mid BG)$$

(not necessarily 1)

But is this assumption reasonable?

Eg, suppose that BB is worth BIG $'s

Then maybe

$$P(T \mid BB) = 0.9$$

$$P(T \mid BG) = 0.2$$

In that case

$$P(BB \mid T) = \frac{0.9}{0.9 + 0.2} = \frac{9}{11}$$

(d) Let R = “A GP was picked randomly and found to be a boy”

(Assume the GP is a boy, ie no error, &

$$P(BB) = P(BG) = P(GB) = P(GG) = 1/4 \quad)$$

Then

$$p = P(BB \mid R) = \frac{P(BB)P(R \mid BB)}{P(R)} = \frac{(1/4)\times 1}{2/4} = \frac{1}{2}$$

But we should first ask:

With what probabilities was a GP going to be picked randomly & their sex revealed, given BB, BG and GB, respectively?

Eg:

$$P(BB \mid R) = \frac{P(R \mid BB)}{P(R \mid BB) + P(R \mid BG) + P(R \mid GB) + \cancel{P(R \mid GG)}}$$

$$= \frac{0.8}{0.8 + 0.5 + 0.5} = \frac{4}{9}$$

Moral

Just because something happened

(eg, Anne told you her oldest is a boy,

or a GP was picked randomly,

or even that you met Ann, etc)

does not mean that it HAD to happen

or even that it was going to happen with a FIXED probability

(eg regardless of BB, BG, etc)

EXERCISE 2 - THREE DOORS

On a game show you are shown 3 doors.

Behind one is a car; the others have goats

You pick door No. 1, and the host opens No. 3, which has a goat.

He then asks if you want to pick No. 2.

Find the pr. that the car is behind No. 2.

Let:

C = "Your initial guess is correct"
I = "Your initial guess is incorrect"
W = "You win the car by switching"

Then the required pr. is

$$p = P(W) = P(C)P(W|C) + P(I)P(W|I)$$

$$= (1/3)*1 + (2/3)*1$$

$$= 2/3$$

But this is ***wrong***

2/3 is the UNCONDITIONAL probability of you winning the car, as calculated BEFORE the game began

Whereas we want the CONDITIONAL probability, NOW, and given known events

Let:

1 = “Car is behind No. 1”

2 = “Car is behind No. 2”

3 = “Car is behind No. 3”

C = “You initially choose No. 1”

O = “Host opens No. 3 & gives you the option to switch to No. 2”

The unconditional pr. of known events is

$$P(CO) = P(1CO) + P(2CO) + \cancel{P(3CO)}$$

$$= P(1)P(C \mid 1)P(O \mid 1C) + P(2)P(C \mid 2)P(O \mid 2C)$$

$$= \left(\frac{1}{3}\right)\left(\frac{1}{3}\right)P(O \mid 1C) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right)(1) \qquad = \frac{1+q}{9}$$

where $\qquad q = P(O \mid 1C)$

(the pr. the host was going to do what he did given the car is behind No. 1 & given what you did)

Then: $$p = P(2 \mid CO) = \frac{P(2CO)}{P(CO)}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)(1)}{(1+q)/9}$$

$$= \frac{1}{1+q}$$

We have assumed:

(a) The car was definitely going to be hidden randomly: $P(1) = P(2) = 1/3$

(b) You were definitely going to pick a door randomly: $P(C|1) = P(C|2) = 1/3$

(c) The host was definitely going to open a goat door (other than the one picked by you) & give you the option to switch (implying $P(O|2C) = 1$)

But $q = P(O|1C)$ could be anything from 0 to 1.

So $p = 1/(1 + q)$ could be anything

from

$1/(1+1) = 1/2$ ($q=1$, host is 'malicious')

to

$1/(1+0) = 1$ ($q=0$, host is 'benevolent')

Eg: Host randomly picks a door to open:

$q=1/2 \Rightarrow p = 1/(1+1/2) = 2/3$

(Equality with 2/3 before is coincidental)

THE TWO MONTIES PROBLEM

Find the pr. the car is behind No. 2 if also:

(d) The host is one of two (M1 & M2) who take turns hosting on alternate nights

(e) If given a choice, M1 opens door with lowest number, & M2 flips a coin

(f) You randomly chose a night on which to play & have no other info re your host

If M1 is your host

$$q = P(O|1C) = 0 \quad \& \quad p = 1/(1 + 0) = 1$$

If M2 is your host

$$q = 1/2 \quad \& \quad p = 1/(1 + 1/2) = 2/3$$

So since M1 & M2 are equally likely to be your host,

$$p = 1*(1/2) + (2/3)*(1/2) = 5/6$$

But this is wrong

Although M1 & M2 were equally likely to be your host prior to the game, that is no longer true NOW

If given a choice, M2 was more likely to open No. 3 than M1

So the fact that No.3 WAS opened implies that M2 is now more likely to be your host

What exactly is the pr. that M2 is your host?

The unconditional pr. of known events is (was)

$$P(CO) = P(q=0)P(CO \mid q=0) + P(q=1/2)P(CO \mid q=1/2)$$

$$= \frac{1}{2}\left(\frac{1+0}{9}\right) + \frac{1}{2}\left(\frac{1+1/2}{9}\right) = \frac{5}{36}$$

So now the pr. that M2 is your host equals:

$$P(q=1/2 \mid CO) = \frac{P(q=1/2)P(CO \mid q=1/2)}{P(CO)}$$

$$= \frac{(1/2)(1+1/2)/9}{5/36} = \frac{3}{5}$$

So: $P(p=2/3|CO) = 3/5$ (M2: $q = 1/2$)

$P(p=1|CO) = 2/5$ (M1: $q = 0$)

So posterior pr. that the car is behind No. 2 is

$$E(p|CO) = 1*P(p=1|CO) + (2/3)*P(p=2/3|CO)$$

$$= 1*(2/5) + (2/3)*(3/5) = 4/5$$

Earlier mistake was to not condition on CO:

$$Ep = 1*P(p=1) + (2/3)*P(p=2/3)$$
$$= 1*(1/2) + (2/3)*(1/2) = 5/6$$

Numerical illustration

18000 hypothetical games on subsequent nights, M1 & M2 alternate

9000 hosted by M1

– opens door with lowest number (q=0)

9000 hosted by M2

– mentally flips a coin (q=1/2)

M2: 9000

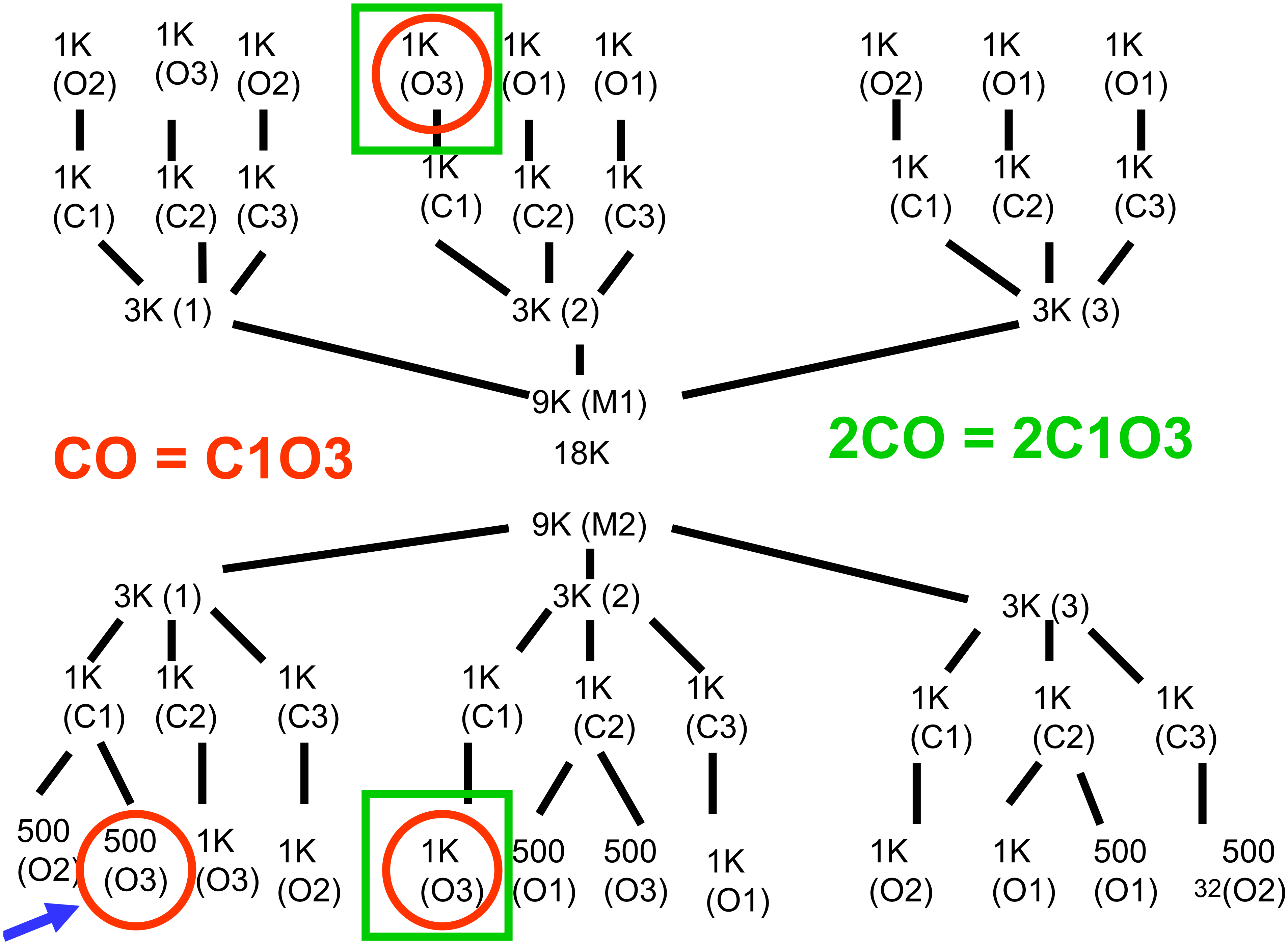
Car behind No. 1: 3000

Player picks No. 1: 1000

Host opens No. 3: 500

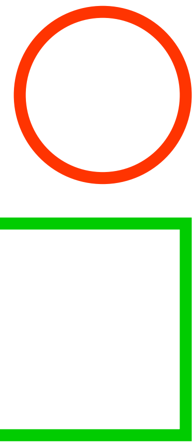
Host opens No. 2: 500

These are just 2 branches of a tree with
2*3*3 + 3 = 21 branches



1K (O2)
1K (O3)
1K (O2)
1K (O3)
1K (O1)
1K (O1)
1K (O2)
1K (O1)
1K (O1)
1K (C1)
1K (C2)
1K (C3)
1K (C1)
1K (C2)
1K (C3)
1K (C1)
1K (C2)
1K (C3)
3K (1)
3K (2)
3K (3)
9K (M1)
CO = C1O3
18K
2CO = 2C1O3
9K (M2)
3K (1)
3K (2)
3K (3)
1K (C1)
1K (C2)
1K (C3)
1K (C1)
1K (C2)
1K (C3)
1K (C1)
1K (C2)
1K (C3)
500 (O2)
500 (O3)
1K (O3)
1K (O2)
1K (O3)
500 (O1)
500 (O3)
1K (O1)
1K (O2)
1K (O1)
500 (O1)
500 (O2)

We find:



#(CO) = 1000 + 500 + 1000 = 2500

#(2CO) = 1000 + 1000 = 2000

So P(2|CO) = P(2CO)/P(CO)
= #(2CO)/#(CO)
= 2000/2500
= 4/5, as before

Some statements & their meaning

$$P(p=2/3) = 1/2$$

The prior pr. that p is 2/3 equals 1/2

Before the game there is a 50% pr. your host will be M2. In that case (only) if you pick No. 1 & the host opens No. 3, there is a 2/3 chance the car is behind No. 2

$P(p=2/3|CO) = 3/5$

The posterior pr. that p is 2/3 equals 3/5

If you picked No. 1 & the host opened No. 3, there is a 60% chance that the host is M2.

In that case (only) there is a 2/3 chance the car is behind No. 2.

In the presence of a prior, the required pr. is

$$P(2|CO) = E\{ P(2|CO,q) \mid CO \} = E(p|CO)$$

In the absence of a prior (original problem), the required pr. is

$$p = P(2|CO,q) = 1/(1 + q)$$

where q is an unknown constant

The maximum likelihood estimate of p is 1/2

The likelihood function is the pr. of known events as a function of unknown parameters:

$$L(q) = P(CO|q) = (1 + q)/9$$

$L(q)$ has max at $q = 1$ (No. 3 was opened)

So the MLE of $p = 1/(1 + q)$ is $1/(1 + 1) = 1/2$

Lends support to popular idea that whether or not you switch makes no difference!

The method of moments estimate of p is 1/2

Let $U = I(CO)$ (= 1 if CO, & = 0 o/w)

Then $EU = P(CO|q) = (1 + q)/9$

Also, $u = 1$ (since CO actually occurred)

Equate $u = EU$

Get $1 = (1 + q)/9$

Solution is $q = 8$

Closest possible value of q is 1

Corresponding value of p is $1/(1 + 1) = 1/2$

Another problem

Suppose q ~ U(0,1) (a priori ignorance)
Then

$$P(CO) = \int P(CO \mid q) f(q) dq = \int_0^1 \frac{1+q}{9} \times 1 dq = \frac{1}{6}$$

$$f(q \mid CO) = \frac{f(q) P(CO \mid q)}{P(CO)} = \frac{1 \times (1+q)/9}{1/6} = \frac{2}{3}(1+q)$$

$$E(p \mid CO) = \int p f(q \mid CO) dq = \int_0^1 \frac{1}{1+q} \times \frac{2}{3}(1+q) dq = \frac{2}{3}$$

In a 1992 paper on the Monty Hall problem:

The pr. the car is behind No. was calculated as:

$$\int_0^1 \frac{1}{1+q} \times 1 dq = \log 2 = 0.963$$

But this is wrong, because it is E(p) and not the required E(p|CO) = 2/3.

f(q) = 1 is used instead of f(q|CO) = 2(1+q)/3

This error poignantly reinforces the sentiment in the abstract of that paper:

"The solution and failed attempts at solution [of the Monty Hall problem] are rich in their lessons in thinking about conditional probability."

THANK YOU