Joint, Conditional, & Marginal Probabilities

Statistics 110

Summer 2006

Joint, Conditional, & Marginal Probabilities

The three axioms for probability don't discuss how to create probabilities for combined events such as $P[A \cap B]$ or for the likelihood of an event A given that you know event B occurs.

Example:

Let A be the event it rains today and B be the event that it rains tomorrow. Does knowing about whether it rains today change our belief that it will rain tomorrow. That is, is $P[B]$, the probability that it rains tomorrow ignoring information on whether it rains today, different from $P[B|A]$, the probability that it rains tomorrow given that it rains today.

$P[B|A]$ is known as the conditional probability of B given A.

It is quite likely that $P[B]$ and $P[B|A]$ are different.

Example: ELISA (Enzyme-Linked Immunosorbent Assay) test for HIV

ELISA is a common screening test for HIV. However it is not perfect as

$$P[+\text{test}|\text{HIV}] = 0.98$$
$$P[-\text{test}|\text{Not HIV}] = 0.93$$

So for people with HIV infections, 98% of them have positive tests (sensitivity), whereas people without HIV infections, 93% of them have negative tests (specificity).

These give the two error rates

$$P[-\text{test}|\text{HIV}] = 0.02 = 1 - P[+\text{test}|\text{HIV}]$$
$$P[+\text{test}|\text{Not HIV}] = 0.07 = 1 - P[-\text{test}|\text{Not HIV}]$$

(Note: the complement rule holds for conditional probabilities)

When this test was evaluated in the early 90s, for a randomly selected American, $P[\text{HIV}] = 0.01$

The question of real interest is what are

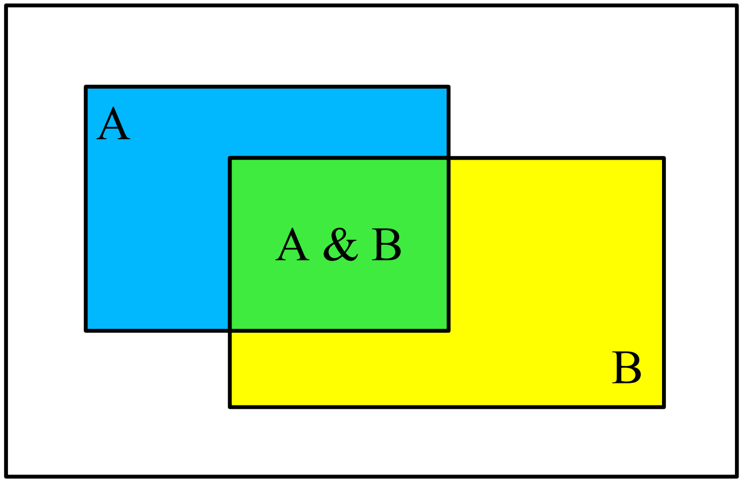
$$P[\text{HIV}|+\text{test}]$$
$$P[\text{Not HIV}|-\text{test}]$$

To figure these out, we need a bit more information.

Definition:

Let A and B be two events with $P[B] > 0$. The conditional probability of A given B is defined to be

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$



One way to think about this is that if we are told that event B occurs, the sample space of interest is now B instead of Ω and conditional probability is a probability measure on B.

Since conditional probability is just ordinary probability on a reduced sample space, the usual axioms hold. e.g.

- $P[B|B] = 1$ (Boundedness)
- $P[A|B] \geq 0$ (Positivity)
- If A_1 and A_2 are disjoint, then $P[A_1 \cup A_2|B] = P[A_1|B] + P[A_2|B]$ (Additivity)

Also the usual theorems also hold. For example

- $P[A^c|B] = 1 - P[A|B]$
- $P[A \cup B|C] = P[A|C] + P[B|C] - P[A \cap B|C]$
- $P[A \cap B|C] \geq P[A|C] + P[B|C] - 1$

We can use the definition of conditional probability to get

Multiplication Rule:

Let A and B be events. Then

$$P[A \cap B] = P[A|B]P[B]$$

Also the relationship also holds with the other ordering, i.e.

$$P[A \cap B] = P[B|A]P[A]$$

Note that $P[A \cap B]$ is sometimes known as the joint probability of A and B.

Back to ELISA example

To get $P[\text{HIV}|+\text{test}]$ and $P[\text{Not HIV}|-\text{test}]$ we need the following quantities

$$P[\text{HIV} \cap +\text{test}], P[\text{Not HIV} \cap -\text{test}], P[+\text{test}], P[-\text{test}]$$

The joint probabilities can be calculated using the multiplication rule

$$
\begin{aligned}
P[\text{HIV} \cap +\text{test}] &= P[\text{HIV}]P[+\text{test}|\text{HIV}] = 0.01 \times 0.98 = 0.0098 \\
P[\text{HIV} \cap -\text{test}] &= P[\text{HIV}]P[-\text{test}|\text{HIV}] = 0.01 \times 0.02 = 0.0002 \\
P[\text{Not HIV} \cap +\text{test}] &= P[\text{Not HIV}]P[+\text{test}|\text{Not HIV}] \\
&= 0.99 \times 0.07 = 0.0693 \\
P[\text{Not HIV} \cap -\text{test}] &= P[\text{Not HIV}]P[-\text{test}|\text{Not HIV}] \\
&= 0.99 \times 0.93 = 0.9207
\end{aligned}
$$

The marginal probabilities on test status are

$$
\begin{aligned}
P[+\text{test}] &= P[\text{HIV} \cap +\text{test}] + P[\text{Not HIV} \cap +\text{test}] \\
&= 0.0098 + 0.0693 = 0.0791 \\
P[-\text{test}] &= P[\text{HIV} \cap -\text{test}] + P[\text{Not HIV} \cap -\text{test}] \\
&= 0.0002 + 0.9207 = 0.9209
\end{aligned}
$$

It is often convenient to display the joint and marginal probabilities in a 2 way table as follows

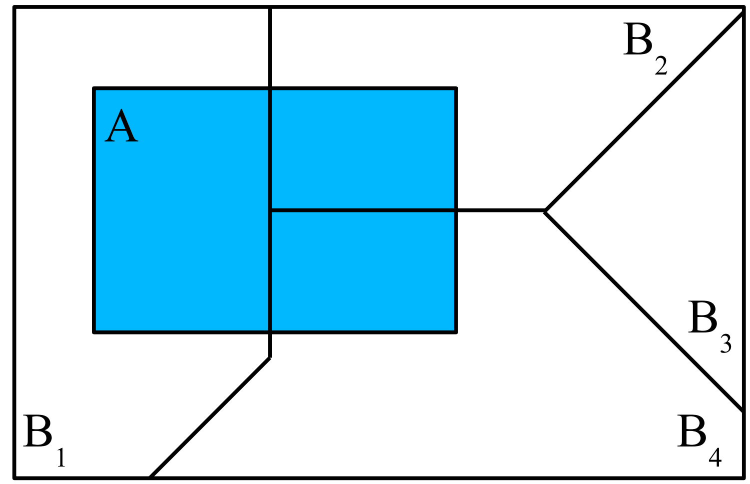
	HIV	Not HIV	
+Test	0.0098	0.0693	0.0791
−Test	0.0002	0.9207	0.9209
	0.0100	0.9900	1.0000

Note that the calculation of the test status probabilities is an example of the Law of Total Probability

Law of Total Probability:
Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j$ for $i \neq j$ with $P[B_i] > 0$ for all i. Then for any event A,

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i]$$

Proof



$$
\begin{aligned}
P[A] &= P[A \cap \Omega] \\
&= P\left[A \cap \left(\bigcup_{i=1}^{n} B_i\right)\right] \\
&= P\left[\bigcup_{i=1}^{n} (A \cap B_i)\right]
\end{aligned}
$$

Since the events $A \cap B_i$ are disjoint

$$
\begin{aligned}
P\left[\bigcup_{i=1}^{n} (A \cap B_i)\right] &= \sum_{i=1}^{n} P[A \cap B_i] \\
&= \sum_{i=1}^{n} P[A|B_i]P[B_i]
\end{aligned}
$$

Note: If a set of events B_i satisfy the conditions above, they are said to form a partition of the sample space.

One way to think of this theorem is to find the probability of A, we can sum the conditional probabilities of A given B_i, weighted by $P[B_i]$.

Now we can get the two desired conditional probabilities.

$$P[\text{HIV}|+\text{test}] = \frac{P[\text{HIV} \cap +\text{test}]}{P[+\text{test}]} = \frac{0.0098}{0.0791} = 0.124$$

$$P[\text{Not HIV}|-\text{test}] = \frac{P[\text{Not HIV} \cap -\text{test}]}{P[-\text{test}]} = \frac{0.9207}{0.9209} = 0.99978$$

These numbers may appear surprising. What is happening here is that most of the people that have positive test are actually uninfected and they are swamping out the the people that actually are infected.

One key thing to remember that $P[A|B]$ and $P[B|A]$ are completely different things. In the example $P[\text{HIV}|+\text{test}]$ and $P[+\text{test}|\text{HIV}]$ are describing two completely different concepts.

The above calculation are an example of Bayes Rule.

Bayes Rule:

Let A and $B_1, B_2, \ldots, B_n$ be events where B_i are disjoint, $\bigcup_{i=1}^n B_i = \Omega$, and $P[B_i] > 0$ for all i. Then

$$P[B_j|A] = \frac{P[A|B_j]P[B_j]}{\sum_{i=1}^n P[A|B_i]P[B_i]}$$

Proof. The numerator is just $P[A \cap B_j]$ by the multiplication rule and the denominator is $P[A]$ by the law of total probability. Now just apply the definition of conditional probability. $\square$

One way of thinking of Bayes theorem is that is allows the direction of conditioning to be switched. In the ELISA example, it allowed switching from conditioning on disease status to conditioning to on test status.

While it is possible to directly apply Bayes theorem, it is usually safer, particularly early on, to apply the definition of conditional probability and calculate the necessary pieces separately, as I did in the ELISA example.

There is another way of looking at Bayes theorem. Instead of probabilities, we can look at odds. An equivalent statement is

$$\underbrace{\frac{P[B_i|A]}{P[B_j|A]}}_{\text{Posterior Odds}} = \underbrace{\frac{P[A|B_i]}{P[A|B_j]}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{P[B_i]}{P[B_j]}}_{\text{Prior Odds}}$$

This approach is useful when $B_1, B_2, \ldots, B_n$ is a set of competing hypotheses and A is information (data) that we want to use to try to help pick the correct hypothesis.

This suggests another way of thinking of Bayes theorem. It tells us how to update probabilities in the presence of new evidence.

Who is this man?

This is reportedly the only known picture of the Reverend Thomas Bayes, F.R.S. — 1701? - 1761.

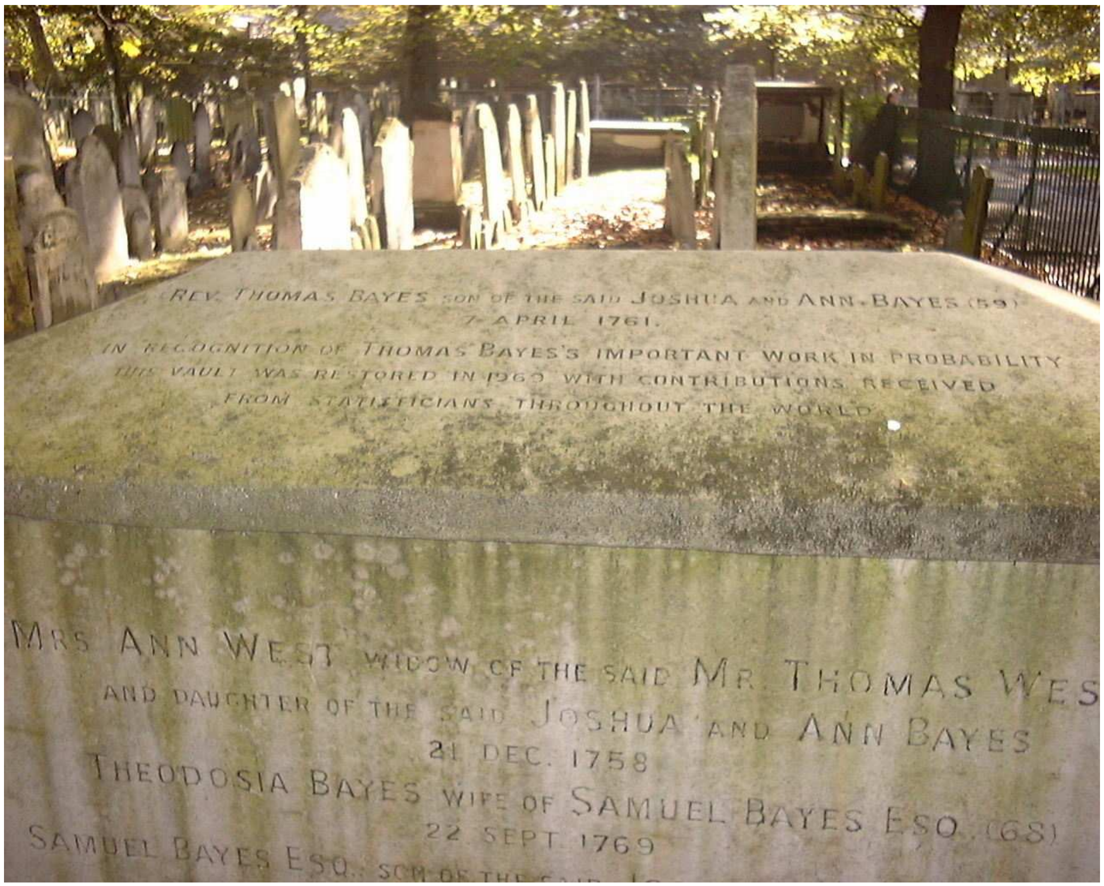
This picture was taken from the 1936 History of Life Insurance (by Terence O'Donnell, American Conservation Co., Chicago). As no source is given, the authenticity of this portrait is open to question.

So what is the probability that this is actually Reverend Thomas Bayes?

However there is some additional information. How does this change our belief about who this is?

Many details about Bayes are sketchy. Much of his work was unpublished and what was often was anonymous. According to Steve Stigler

> The date of his birth is not known: Bayes's posterior is better known than his prior.



Actually his date of death isn't that well known either. Its general considered to be April 7, 1761 however it has also be reported as April 17th of the same year.

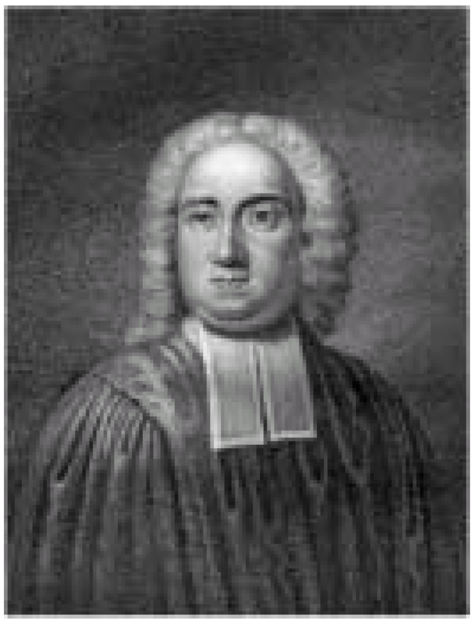
There is some additional information we can get from this picture to help use decide whether this is Bayes or not.

1. The caption under the photo in O'Donnell's book was "Rev. T. Bayes: Improver of the Columnar Method developed by Barrett".

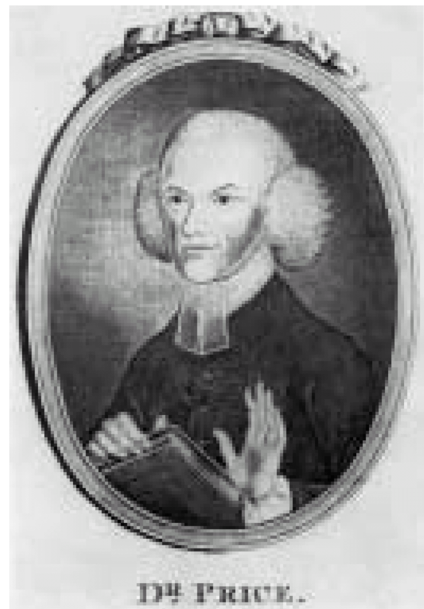
   There are some problems with this claim. First Barrett was born in 1752 and would have been about 9 years old when Bayes died. In addition, the method that Bayes allegedly improved was apparently developed between 1788 and 1811 and read to the Royal Society in 1812, long after Bayes' death.

   So there is a problem here, but whether it is a problem with just the caption or the picture as well isn't clear.

2. Bayes was a Nonconformist (Presbyterian) Minister. Does the clothing in the picture match that of a Nonconformist Minister in the 1740's and 1750's. The picture has been compared to three other Ministers, Joshua Bayes, Bayes' father, Richard Price (portrait dated 1776), the person who read Bayes' paper to the Royal Society, and Philip Doddridge, a friend of Bayes' brother-in-law.



Joshua Bayes
(1671-1746)



Richard Price
(1723-1791)

Philip Doddridge
(1702-1751)

Two things stand out in the comparisons.

(a) No wig. It is likely that Bayes should have been wearing a wig similar to Doddridge's, which was going out of fashion in the 1740's or similar to Price's, which was coming into style at the time.

(b) Bayes appears to be wearing a clerical gown like his father or a larger frock coat with a high collar. On viewing the other two pictures, we can see that the gown is not in the style for Bayes' generation and the frock coat with a large collar is definitely anachronistic.

(Interpretation of David Bellhouse from IMS Bulletin **17**, No. 1, page 49)

Question: How to we incorporate this information to adjust our probability that this is actually a picture of Bayes?

Answer: $P[\text{This is Bayes}|\text{Data}]$ which can be determined by Bayes' Theorem.

$$P[B|\text{Data}] = \frac{P[B]P[\text{Data}|B]}{P[B]P[\text{Data}|B] + P[B^c]P[\text{Data}|B^c]}$$

Note that this is not easy to do as assigning probabilities here is difficult.

For an example of how this can be done, see the paper (available on the course web site)

Stigler SM (1983). Who Discovered Bayes's Theorem. *American Statistician* **37**: 290-296.

In this paper, Stigler examines whether Bayes was the first person to discover what is now known as Bayes' Theorem. There is evidence that the result was known in 1749, 12 years before Bayes' death and 15 years before Bayes' paper was published. In Stigler's analysis

$$P[\text{Bayes 1st discovered} \mid \text{Data}] = 0.25$$
$$P[\text{Saunderson 1st discovered} \mid \text{Data}] = 0.75$$

Note that Bayes actually proved a special case of Bayes' Theorem involving inference on a Bernoulli success probability.

Monty Hall Problem

There are three doors. One has a car behind it and the other two have farm animals behind them. You pick a door, then Monty will have the lovely Carol Merrill open another door and show you some farm animals and allow you to switch. You then win whatever is behind your final door choice.

You choose door 1 and then Monty opens door 2 and shows you the farm animals. Should you switch to door 3?

Answer: It depends

Three competing hypotheses D_1, D_2, and D_3 where

$$D_i = \{\text{Car is behind door } i\}$$

What should our observation A be?

$A = \{$Door 2 is empty$\}$?

or $A = \{$The opened door is empty$\}$?

or something else?

We want to condition on all the available information, implying we should use

$A = \{$After door 1 was selected, Monty chose door 2 to be opened$\}$

Prior probabilities on car location: $P[D_i] = \frac{1}{3}, i = 1, 2, 3$

Likelihoods:

$$P[A|D_1] = \frac{1}{2} \qquad (*)$$

$$P[A|D_2] = 0$$

$$P[A|D_3] = 1$$

$$P[A] = \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2}$$

Posterior probabilities on car location:

$$P[D_1|A] = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3} \qquad \text{No change!}$$

$$P[D_2|A] = \frac{0 \times \frac{1}{3}}{\frac{1}{2}} = 0$$

$$P[D_3|A] = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \qquad \text{Bigger!}$$

If you are willing to assume that when two empty doors are available, Monty will randomly choose one of them to open (with equal probability) (assumption *), then you should switch. You'll win the car $\frac{2}{3}$ of the time.

Now instead, assume Monty opens the door based on the assumption

$$P[A|D_1] = 1 \qquad\qquad (**)$$

i.e. Monty will always choose door 2 when both doors 2 and 3 have animals behind them. (The other two are the same.) Now

$$P[A] = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{2}{3}$$

Now the posterior probabilities are

$$P[D_1|A] = \frac{1 \times \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

$$P[D_2|A] = 0$$

$$P[D_3|A] = \frac{1 \times \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

So in this situation, switching doesn't help (doesn't hurt either).

Note: This is an extremely subtle problem that people have discussed for years (go do a Google search on Monty Hall Problem). Some of the discussion goes back before the show Let Make a Deal ever showed up on TV. The solution depends on precise statements about how doors are chosen to be opened. Changing the assumptions can lead to situations that changing can't help and I believe there are situations where changing can actually be worse. They can also lead to situations where the are advantages to picking certain doors initially.

General Multiplication Rule

The multiplication rule discussed earlier can be extended to an arbitrary number of events as follows

$$P[A_1 \cap A_2 \cap \ldots \cap A_n] = \\ P[A_1] \times P[A_2|A_1] \times P[A_3|A_1A_2] \times \ldots \times P[A_n|A_1A_2 \ldots A_{n-1}]$$

This rule can be used to build complicated probability models.

Example: Jury selection

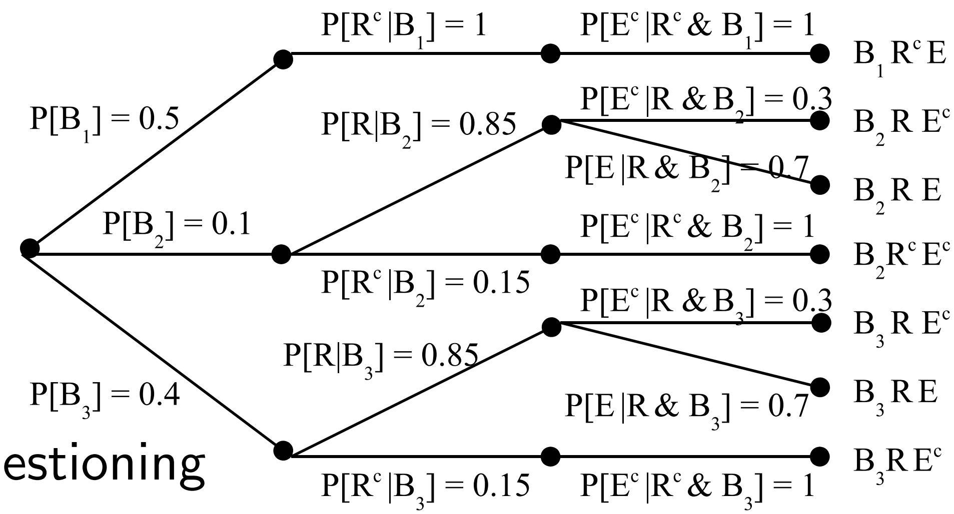
When a jury is selected for a trial, it is possible that a prospective can be excused from duty for cause. The following model describes a possible situation.

- Bias of randomly selected juror

- B_1: Unbiased
- B_2: Biased against the prosecution
- B_3: Biased against the defence

- R: Existing bias revealed during questioning

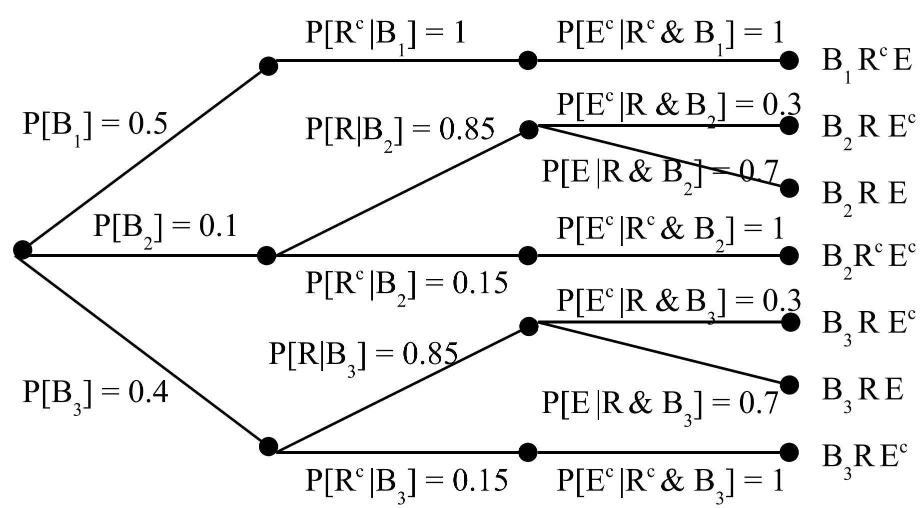
- E: Juror excused for cause



The probability of any combination of these three factors can be determined by multiplying the correct conditional probabilities. The probabilities for all twelve possibilities are

	R		R^c	
	E	E^c	E	E^c
B_1	0	0	0	0.5000
B_2	0.0595	0.0255	0	0.0150
B_3	0.2380	0.1020	0	0.0600

For example $P[B_2 \cap R \cap E^c] = 0.1 \times 0.85 \times 0.3 = 0.0255.$



From this we can get the following probabilities

$$P[E] = 0.2975 \quad P[E^c] = 0.7025$$
$$P[R] = 0.4250 \quad P[R^c] = 0.5750$$

$$P[B_1 \cap E] = 0 \qquad P[B_2 \cap E] = 0.0595 \quad P[B_3 \cap E] = 0.238$$
$$P[B_1 \cap E^c] = 0.5 \quad P[B_2 \cap E^c] = 0.0405 \quad P[B_3 \cap E^c] = 0.162$$

From these we can get the probabilities of bias status given that a person was not excused for cause from the jury.

$$P[B_1|E^c] = \frac{0.5}{0.7025} = 0.7117 \quad P[B_1|E] = \frac{0}{0.2975} = 0$$

$$P[B_2|E^c] = \frac{0.0405}{0.7025} = 0.0641 \quad P[B_2|E] = \frac{0.0595}{0.2975} = 0.2$$

$$P[B_3|E^c] = \frac{0.1620}{0.7025} = 0.2563 \quad P[B_3|E] = \frac{0.238}{0.2975} = 0.8$$