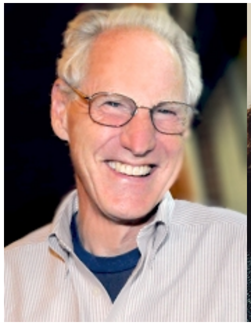
STATISTICAL METHODS FOR THE PHYSICAL SCIENCES

Week 2: Probability theory

The Monty Hall problem

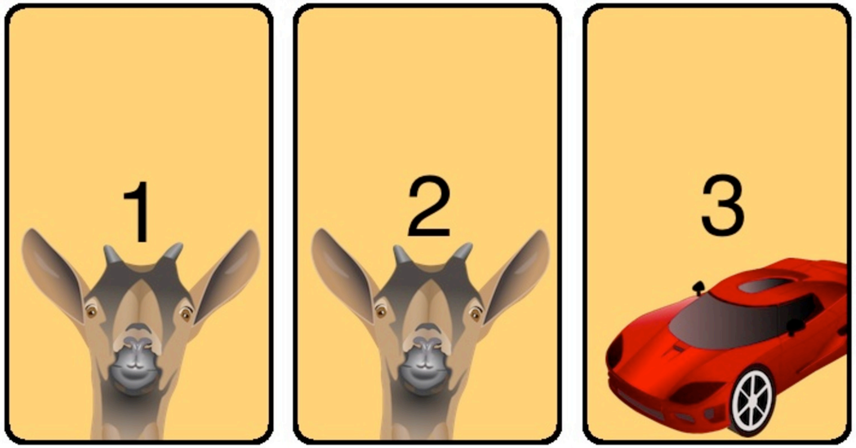
- Contestants on a game show have to pick one of three doors to win a prize. Behind one is a car (the prize). Behind the other two are goats…
- The contestant chooses a door. Then the host opens one of the remaining two doors, which must hide a goat.
- The contestant now gets to choose – to stay with their choice or switch to the other door…



Steve Selvin

Marilyn vos Savant

Monty Hall



Is it better to stay or switch?

Probability theory: elements, outcomes, events

- *Sample space* (denoted Ω) contains the set of all possible outcomes (*elements* or *elementary outcomes*)
  - E.g. tossing a coin the sample space of elementary outcomes is:
    $\Omega = \{H, T\}$
  - Tossing a coin twice we have: $\Omega = \{HH, HT, TH, TT\}$
- A set containing one or more outcomes is called an *event*: an event occurs if the outcome is in that set (the event can also be a set of outcomes).
- E.g. what is the sample space of the Geiger-Rutherford experiment?

Combining events with set theory & Venn diagrams

Set notation for events:

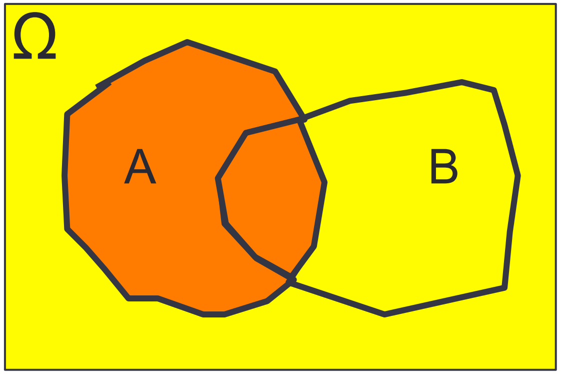
A = an event B = another event Ω = sample space $\varnothing$ = null (empty) set

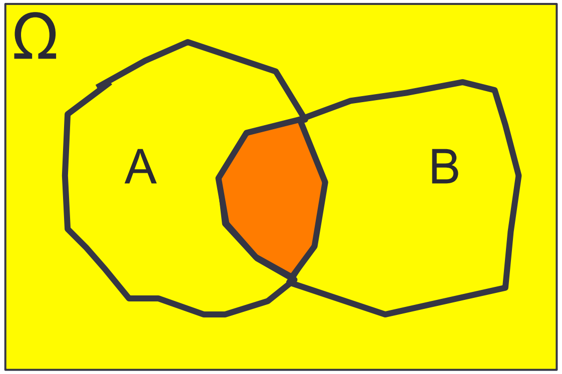
$A \cap B$ = 'A and B' (intersection of sets)

$A \cup B$ = 'A or B' (union of sets)

A^C = complement of A ('not A')

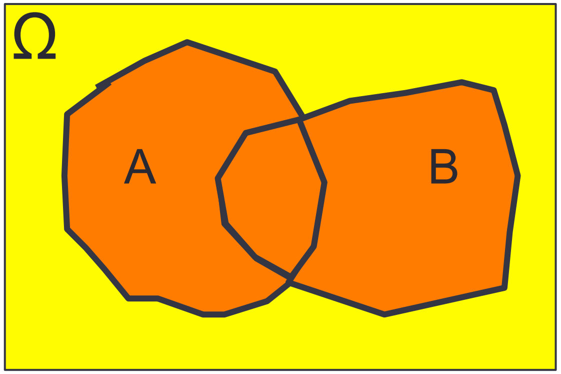
A

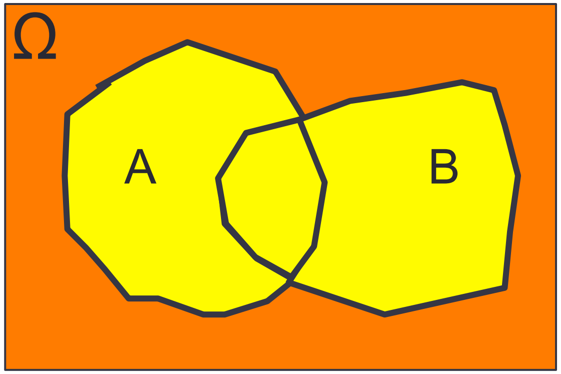




$A \cap B$

$A \cup B$





$(A \cup B)^C$

Probability with set notation

The probability of an event A: $\Pr(A) = x \text{ where } 0 \leq x \leq 1$

So we can also define: $\Pr(\varnothing) = 0 \qquad \Pr(\Omega) = 1$

$$\Pr(A \cup A^C) = 1 \quad \Pr(A \cap A^C) = 0$$

Two events, A & B are *mutually exclusive* if:

$$A \cap B = \varnothing$$

If A & B are mutually exclusive then:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

Since: $A \cap A^C = \varnothing$ we can infer that:

$$\Pr(A) = 1 - \Pr(A^C)$$

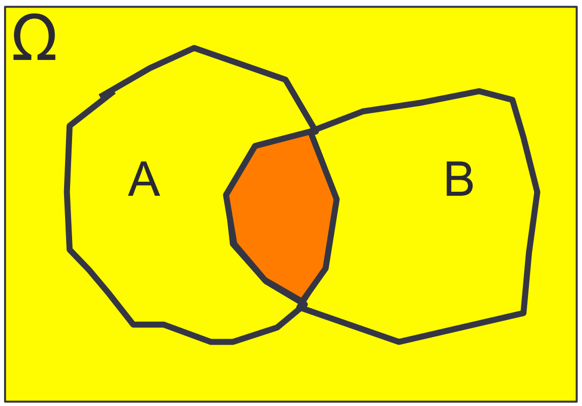
Note: in Venn diagrams, probability can be equated to the ***area*** of each region

Conditional probability

Probability of event A *conditional* on event B ('probability of A given B'):

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Another way to look at it: what fraction of B is the filled area?



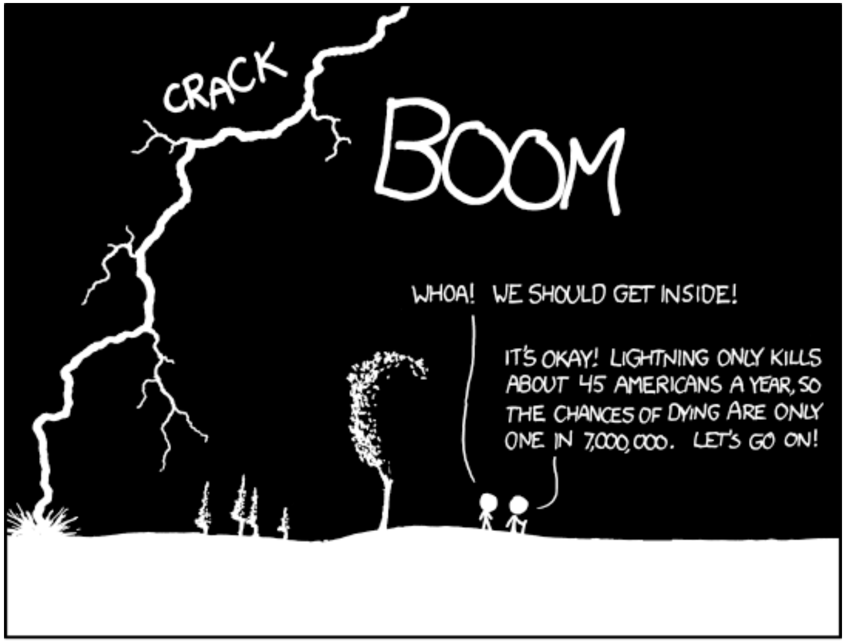
Also it follows that:

$$\Pr(A|A) = 1 \quad \Pr(A|A^C) = 0$$

Note that generally:

$$\Pr(A|B) \neq \Pr(B|A)$$

e.g. the 'prosecutors fallacy' is an example of people not knowing this rule!



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

The Lucia de Berk case

- Paediatric nurse sentenced to life imprisonment for 7 murders and 3 attempted murders of patients under her care.
- Guilty verdict depended heavily on a statistical estimate of the probability that one nurse's work times would coincide with so many deaths/resuscitations by chance: 1 in 342 million!
- But: probability found by multiplying each probability of one coincidence together. This is a probability for a single trial – i.e. one person. But many Paediatric nurses work in the Netherlands…
- Also bias led to inflated number of events associated with her once she had already been accused.
- Correct question: what is the probability that she is innocent given she was associated with the deaths. Statisticians calculated actual probability at 1 in 25!
- Retried and found innocent in 2010 after a campaign by Dutch journalists and scientists.

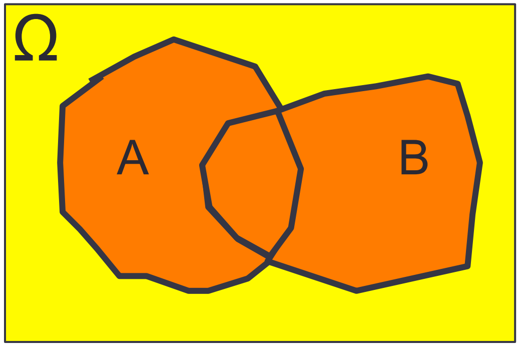
Rules of probability calculus

- Convexity rule:

$$0 \geq \Pr(A|B) \leq 1 \text{ and } \Pr(A|A) = 1$$

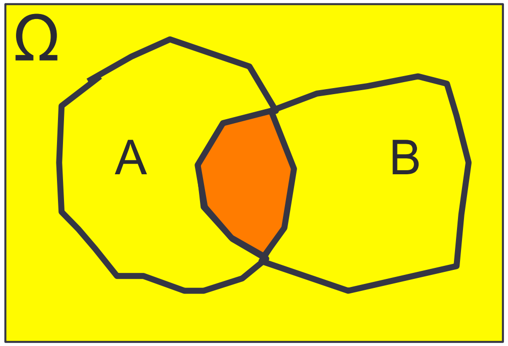
- Addition rule:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$



- Multiplication rule:

$$\Pr(A \cap B) = \Pr(A|B)\,\Pr(B)$$



Events A and B are *independent* if:

$$\Pr(A|B) = \Pr(A)$$

$$\Rightarrow \Pr(A \cap B) = \Pr(A)\,\Pr(B)$$

What would a Venn diagram representation of independent events look like?

Extending the rules of probability calculus: 'Extension of the conversation'

What can we say about the probability of event *B* given that some event *A* has occurred?

$$\Pr(B) = \Pr\left((B \cap A) \cup (B \cap A^C)\right)$$

$$= \Pr(B \cap A) + \Pr(B \cap A^C)$$ (apply addition rule)

$$= \Pr(B|A)\Pr(A) + \Pr(B|A^C)\Pr(A^C)$$ (apply multiplication rule)

Extending the rules of probability calculus: Law of total probability

Now we 'extend the conversation' even more and consider the probability of *B* given an exhaustive set of all possible mutually exclusive events: $\Omega = \{A_1, A_2, \ldots\ A_n\}$

$$
\begin{aligned}
&\Pr(B) = \Pr(B \cap \Omega) \\
&= \Pr\left((B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)\right) \\
&= \Pr(B \cap A_1) + \Pr(B \cap A_2) + \ldots + \Pr(B \cap A_n) \\
&= \sum_{i=1}^{n} \Pr(B \cap A_i) \\
&= \sum_{i=1}^{n} \Pr(B|A_i)\Pr(A_i)
\end{aligned}
$$

(apply addition rule)

(apply multiplication rule)

This summation to eliminate conditionals is called marginalisation. We can say that we obtain the marginal distribution of *B* (the marginal variable) by *marginalising* over *A* (*A* is *'marginalised out'*)

Extending the rules of probability calculus: Bayes theorem

Rev. Thomas Bayes

Pierre-Simon Laplace

$$\Pr(A \cap B) = \Pr(B \cap A) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

$$\Rightarrow \Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \qquad (\text{if } \Pr(B) \neq 0)$$

We can generalise from two events to a set of exclusive exhaustive events: $\{A_1, A_2, \ldots, A_n\}$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B)} = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{i=1}^{n}\Pr(B|A_i)\Pr(A_i)}$$

Simple applications of Bayes' theorem 1

A disease is prevalent in the population, so that 1% of people have the disease. An effective test exists for the disease – when someone has it, it is detected 99% of the time. The test sometimes also gives a 'false positive', a detection when someone doesn't have the disease. It does this 2% of the time.

The question is: what are my chances of having the disease if the test comes up positive?

Prevalence of disease: $$\Pr(D) = 0.01, \quad \Pr(D^C) = 0.99$$

True positive rate 99%: $$\Pr(+|D) = 0.99, \quad \Pr(-|D) = 0.01$$

False positive rate 2%: $$\Pr(+|D^C) = 0.02, \quad \Pr(-|D^C) = 0.98$$

Apply Bayes theorem: $$\Pr(D|+) = \frac{\Pr(+|D)\Pr(D)}{\Pr(+)} = \frac{0.99 \times 0.01}{0.0297} = \frac{1}{3}$$

Where we must marginalise over the probability of positive detection

$$\begin{aligned}\Pr(+) &= \Pr(+|D)\Pr(D) + \Pr(+|D^C)\Pr(D^C) \\ &= 0.99 \times 0.01 + 0.02 \times 0.99 = 0.0297\end{aligned}$$

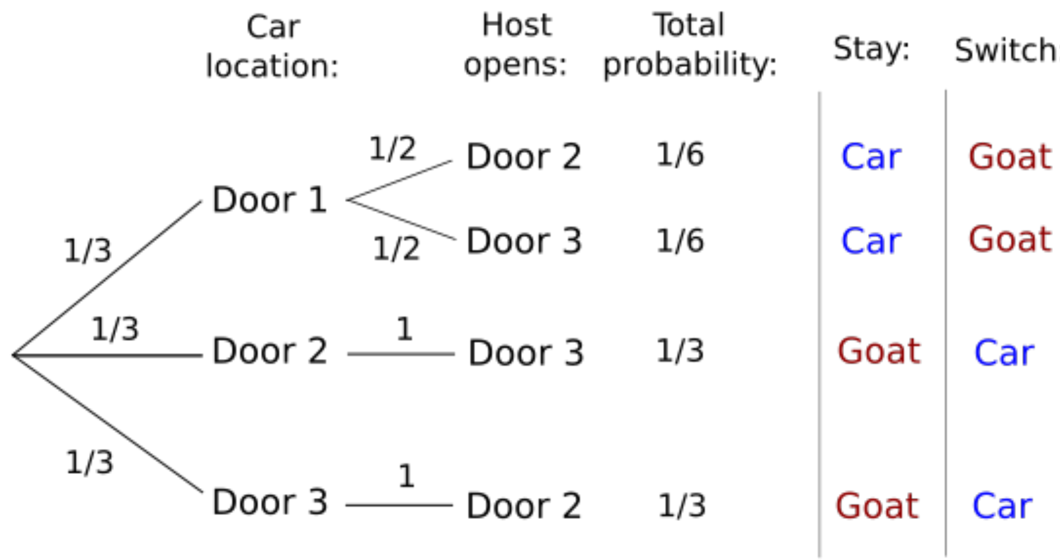
Simple applications of Bayes' theorem 2

- A particle detector is designed to detect muons, with probability 99%.
- But sometimes it records a pion as a muon, with probability 2%.
- If we place it in a beam consisting of 99% pions and 1% muons, what is the chance that the muon we detect is actually a pion?

- What happens if we require the particle to trigger two independent detectors, to count as a 'detection'?

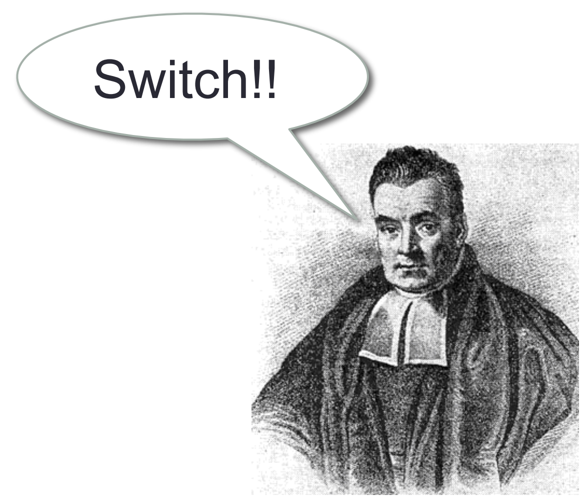
Back to the Monty Hall problem

We choose door 1. What is the probability that door 1 has the car given that Monty shows us that Door 2 has a goat?



$$\Pr(C_1|G_2) = \frac{\Pr(G_2|C_1)\Pr(C_1)}{\sum_{i=1}^{3}\Pr(G_2|C_i)\Pr(C_i)}$$

$$= \frac{1/2 \times 1/3}{1/2 \times 1/3 + 0 \times 1/3 + 1 \times 1/3} = \frac{1}{3}$$



The Bayesian approach to hypothesis testing

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B)} = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{i=1}^{n}\Pr(B|A_i)\Pr(A_i)}$$

We can generalise Bayes' theorem, by defining the events corresponding to the data (the *evidence*, *E*) and the hypothesis (*H*):

$$\Pr(H|E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}$$

Bayesian vs. Frequentist reasoning

Adapted from J. VanderPlas “Frequentism and Bayesianism: A Python-driven Primer
http://arxiv.org/pdf/1411.5018v1.pdf

- ***Frequentism***: probability is a limiting case for repeated measurements; probabilities are fundamentally related to frequencies of events (often some imaginary repeated measurements).
- But how can we talk about the probability of the ‘true’ mass of Saturn (there is only one Saturn!)?

- **Bayesianism**: Probability extended to cover *degrees of certainty about statements*.
- For a Bayesian, probabilities are fundamentally related to their own (prior) knowledge about an event

Frequentist vs. Bayesian approach to measurement

Consider a ‘true’ quantity *F* which we want to measure (e.g. the flux of a star, the rest-mass of a particle). We make a set (the data, *D*) of independent measurements of this quantity, which give values F_i and associated errors e_i:

$$D = \{F_i, e_i\}$$

A frequentist approach would be to pose a hypothesis (i.e. what is the true value of *F*?) and ask the question: *what is the probability that I would have obtained this particular data, given my hypothetical value for F?*

$$\mathrm{Pr}(D|F)$$

A Bayesian approach is to ask: *what is the probability of my hypothetical value for F being true, given the data I have?*

$$\mathrm{Pr}(F|D)$$

Application of Bayes' theorem

The **likelihood** (of the data given the model parameters), same as Frequentist approach

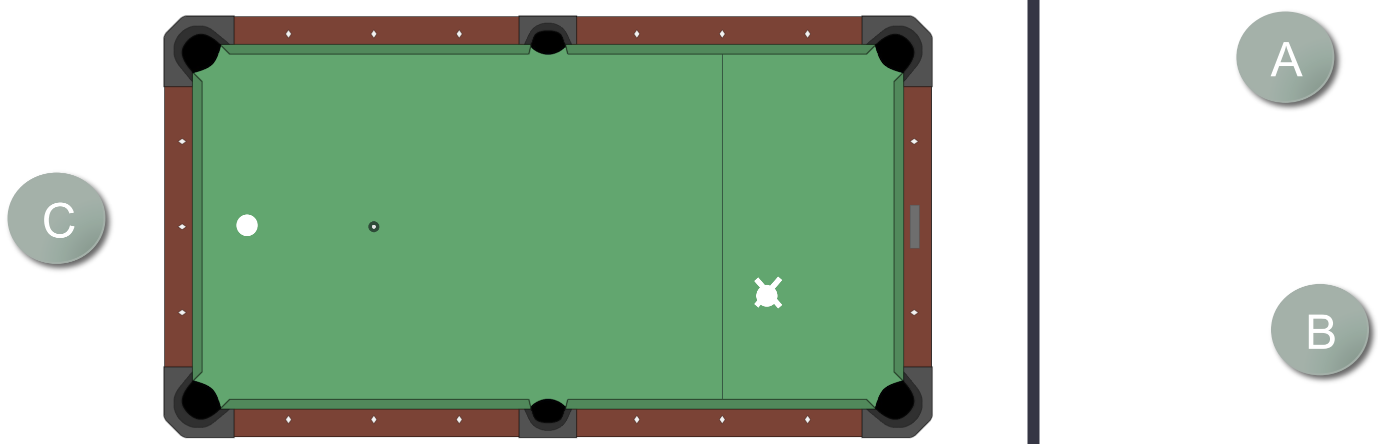
The **model prior** (encodes what we knew about the model before considering the data *D*)

$$\Pr(F|D) = \frac{\Pr(D|F)\Pr(F)}{\Pr(D)}$$

The **posterior**: probability of model parameters given the data

The **model evidence** – overall probability of obtaining *D*, marginalised over all model parameters (essentially a normalisation term)

Bayes' billiards game



Carol rolls a billiard ball down the table, marking where it stops. Then she starts rolling balls down the table. If the ball lands to the left of the mark, Alice gets a point, to the right and Bob gets a point. First to 6 points wins.

After some time, Alice has 5 points and Bob has 3. What is the probability that Bob wins the game (Pr(B))?
A Frequentist approach would be to estimate the most likely probability that Alice gets a point, based on the result from 8 rolls:

$$\hat{p} = 5/8 \qquad \Pr(B) = (1 - \hat{p})^3 = 0.053$$

Bayes' billiards game: Bayesian approach

We want to know the probability that Bob wins given the data we have in hand, this can be thought of as a 'marginal probability':

$$\Pr(B|D) \equiv \int_{-\infty}^{\infty} \Pr(B, p|D)\mathrm{d}p$$

Now apply Conditional Probability definition:

$$\Pr(B|D) = \int \Pr(B|p, D)\Pr(p|D)\mathrm{d}p$$

Now apply Bayes rule:

$$\Pr(B|D) = \int \Pr(B|p, D)\frac{\Pr(D|p)\Pr(p)}{\Pr(D)}\mathrm{d}p$$

And we can separate and expand Pr(*D*) since it is not itself a function of *p*:

$$\Pr(B|D) = \frac{\int \Pr(B|p, D)\Pr(D|p)\Pr(p)\mathrm{d}p}{\int \Pr(D|p)\Pr(p)\mathrm{d}p}$$

Bayes' billiards game: Bayesian approach

We need to solve:

$$\Pr(B|D) = \frac{\int \Pr(B|p,D)\Pr(D|p)\Pr(p)\mathrm{d}p}{\int \Pr(D|p)\Pr(p)\mathrm{d}p}$$

This term is easy: $\Pr(B|p,D) = (1-p)^3$

Here we need to account for the number of permutations/ combinations, the constant doesn't matter (it cancels from top and bottom):

$$\Pr(D|p) \propto p^5(1-p)^3$$

We can assume a uniform prior, i.e. probability distribution of *p* is constant between 0 and 1:

$\Pr(p) = constant$ which also cancels…

$$\Pr(B|D) = \frac{\int_0^1 (1-p)^6 p^5 \mathrm{d}p}{\int_0^1 (1-p)^3 p^5 \mathrm{d}p} \rightarrow \Pr(B|D) = 0.091$$

i.e. a different result (which simulations show is correct…)