PROBABILITY: AN INTERACTIVE REVIEW

CS1951A INTRO TO DATA SCIENCE

DAN POTTER

CONTENT CREDIT

Many of of these slides include content produced by

Jeremy Orloff

Jonathan Bloom

MIT Course Number

18.05

As Taught In Spring 2014

OUTLINE

1. **Probability Spaces & Probability Functions**
   1. Example: Rolling a Die
   2. Conditional Probability
   3. Independent Events
2. **Bayesian Statistics**

ALL SPORTS COMMENTARY

STATISTICS ≠ PROBABILITY THEORY

- **Probability theory:** mathematical theory that describes uncertainty.
  - Distributions/parameters are known
  - Consequences are derived
- **Statistics:** set of techniques for extracting useful information from data.
  - Distributions/parameters are generally unknown
    - Need to be ***estimated*** from the data
    - Consequences can then be derived
- Distributions/parameters provide a <u>formal</u> model

PROBABILITY SPACE

A probability space has three components:

(1) A sample space Ω that is the set of all possible outcomes of the random process modeled;

(2) A family of sets F representing the allowable events, where each set in F is a subset of the sample space Ω;

In a discrete probability space, we typically use the "power set of Ω" for F, written as $F = 2^{\Omega}$

(3) A probability function $P: F \to R$

PROBABILITY FUNCTION

A probability function is any $P: F \rightarrow R$ with the following properties:

① For any event *E in F*, $0 \leq P(E) \leq 1$

② $P(\Omega) = 1$

③ For any countable collection of pairwise disjoint events $\{E_1, E_2, E_{3,} \ldots\}$, P must be countably additive, i.e.,

$$P(\bigcup_{i=1} E_i) = \sum_{i=1} P(E_i)$$

Basically, P defines the probability of each event in F, and (3) means P for complicated events can be calculated using simpler events.

(3) also solves some limit problems when working with infinite sequences of events.

EXAMPLE: TOSSING A SINGLE COIN

$$\Omega = \{H, T\}$$

$$F = 2^{\Omega} = 2^{2} = 4\,Events$$

$$F = \{\{\ \}, \{H\}, \{T\}, \{H, T\}\}$$

$$\Pr(\{\ \}) = 0$$

$$\mathrm{P}(\{H\}) = 0.5$$

$$\mathrm{P}(\{T\}) = 0.5$$

$$\mathrm{P}(\{H, T\}) = P(\{H\} \bigcup \{T\}) = P(\{H\}) + P(\{T\}) = 1$$

EXAMPLE: ROLLING A FAIR DIE

$$\Omega = \{1,2,3,4,5,6\}$$

$$F = 2^{\Omega} = 2^{6}\, Events$$

$$\mathrm{P}(\{\ \}) = 0$$

$$\mathrm{P}(\{1\}) = \mathrm{P}(\{2\}) = \mathrm{P}(\{3\}) = \mathrm{P}(\{4\}) = \mathrm{P}(\{5\}) = \mathrm{P}(\{6\}) = \frac{1}{6}$$

$$\mathrm{P}(\{1,2\}) = \mathrm{P}(\{1,3\}) = \mathrm{P}(\{1,4\}) = \mathrm{P}(\{1,5\}) = \mathrm{P}(\{1,6\}) = \frac{2}{6}$$

...

EXPERIMENTS

- Experiment: a repeatable procedure
- Sample space: set of all possible outcomes Ω.
- Event: a subset of the sample space.
- Probability function, $P(\omega)$: gives the probability foreach outcome $\omega \in S$
    1. Probability is between 0 and 1
    2. Total probability of all possible outcomes is 1.

DISCRETE SAMPLE SPACE

Discrete = listable = countable

Examples:

$\{a, b, c, d\}$ (finite)

$\{0, 1, 2, \ldots\}$ (infinite)

EVENTS

Events are sets:

- Can describe in words
- Can describe in notation
- Can describe with Venn diagrams

Experiment: toss a coin 3 times.

Event:

You get 2 or more heads = $\{$ HHH, HHT, HTH, THH $\}$

CQ: Events, Sets and Words

Experiment: toss a coin 3 times.

Which of following equals the event "exactly two heads"?

$$
\begin{aligned}
A &= \{THH, HTH, HHT, HHH\} \\
B &= \{THH, HTH, HHT\} \\
C &= \{HTH, THH\}
\end{aligned}
$$

(1) A (2) B (3) C (4) A or B

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

Which of following equals the event "exactly two heads"?

$$A = \{THH, HTH, HHT, HHH\}$$
$$B = \{THH, HTH, HHT\}$$
$$C = \{HTH, THH\}$$

(1) A (2) B (3) C (4) A or B

Answer: 2) B, The event "exactly two heads" determines a unique subset, containing all outcomes that have exactly two heads.

CQ: Events, Sets and Words

Experiment: toss a coin 3 times.

Which of the following describes the event $\{THH, HTH, HHT\}$?

(1) "exactly one head"
(2) "exactly one tail"
(3) "at most one tail"
(4) none of the above

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

Which of the following describes the event $\{THH, HTH, HHT\}$?

(1) "exactly one head"
(2) "exactly one tail"
(3) "at most one tail"
(4) none of the above

answer: (2) "exactly one tail"
Notice that the same event $E \subset \Omega$ may be described in words in multiple ways ("exactly 2 heads" and "exactly 1 tail").

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

The events "exactly 2 heads" and "exactly 2 tails" are disjoint.

(1) True (2) False

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

The events "exactly 2 heads" and "exactly 2 tails" are disjoint.

(1) True (2) False

answer: True
$\{THH, HTH, HHT\} \cap \{TTH, THT, HTT\} = \emptyset$.

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

The event "at least 2 heads" implies the event "exactly two heads".

(1) True (2) False

CQ: EVENTS, SETS AND WORDS

Experiment: toss a coin 3 times.

The event “at least 2 heads” implies the event “exactly two heads”.

(1) True (2) False

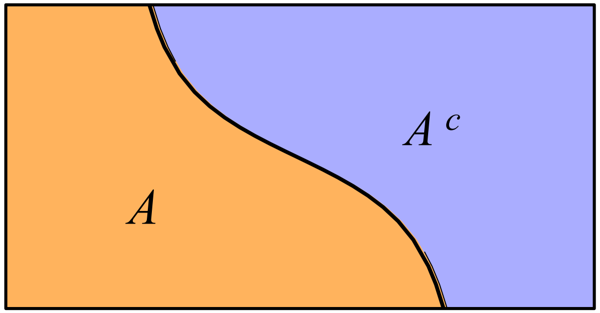
False. It’s the other way around: “Exactly 2 heads” => “At least two heads”, i.e. “Exactly two heads” = {THH,HTH,HHT} $\subset$ “At least two heads” = {THH,HTH,HHT,HHH},
so whenever the event “Exactly two heads occurs” the event “At least two heads occurs” occurs.

PROBABILITY AND SET OPERATIONS ON EVENTS

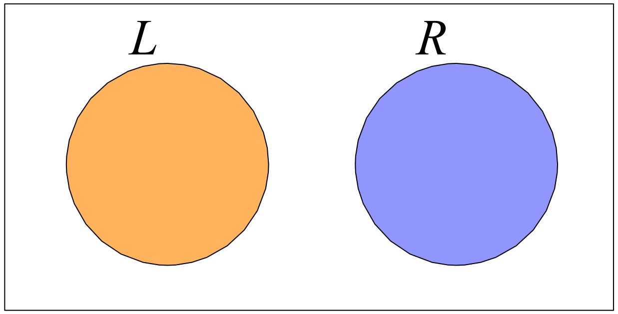
Rule 1. Complements

Rule 2. Disjoint events

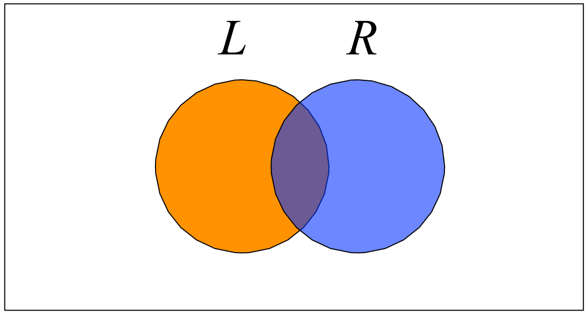
Rule 3. Inclusion-exclusion principle



$\Omega = A \cup A^c$, no overlap



$L \cup R$, no overlap



$L \cup R$, overlap $= L \cap R$

CONCEPT QUESTION

- Class has 50 students
- 20 male (M), 25 brown-eyed (B)

For a randomly chosen student what is the range of possible values for $p = P(M \cup B)$?

(a) $p \leq .4$
(b) $.4 \leq p \leq .5$
(c) $.4 \leq p \leq .9$
(d) $.5 \leq p \leq .9$
(e) $.5 \leq p$

CONCEPT QUESTION

- Class has 50 students
- 20 male (M), 25 brown-eyed (B)

For a randomly chosen student what is the range of possible values for $p = P(M \cup B)$?

(d) $.5 \leq p \leq .9$

The easy way to answer this is that A ∪ B has a minumum of 25 members (when all males are brown-eyed) and a maximum of 45 members (when no males have brown-eyes). So, the probability ranges from .5 to .9

Thinking about it in terms of the inclusion-exclusion principle we have
$P(M \cup B) = P(M) + P(B) - P(M \cap B) = .9 - P(M \cap B)$.
So the maximum possible value of $P(M \cup B)$ happens if M and B are disjoint, so $P(M \cap B) = 0$. The minimum happens when $M \subset B$, so $P(M \cap B) = P(M) = .4$.

CONCEPT QUESTION

Lucky Larry has a coin that you're quite sure is not fair.

- He will flip the coin twice
- It's your job to bet whether the outcomes will be the same (HH, TT) or different (HT, TH).

Which should you choose?

1. Same
2. Different
3. It doesn't matter, same and different are equally likely

Concept Question

Lucky Larry has a coin that you're quite sure is not fair.

- He will flip the coin twice
- It's your job to bet whether the outcomes will be the same (HH, TT) or different (HT, TH).

Which should you choose?

1. Same 2. Different 3. Doesn't matter

HINT: Let p be the probability of heads and use probability to answer the question.

(If you don't see the symbolic algebra try p = .2, p=.5)

Algebraic Answer: 1. Same (same is more likely than different)

The key bit of arithmetic is that if a != b then
(a—b)^2 >0 ⇔ a2 +b2 >2ab.
To keep the notation cleaner, let's use P(T) = (1 — p) = q. Since the flips are independent (we'll discuss this next week) the probabilities multiply. This gives the following 2 × 2 table.

p^2 pq
qp q^2

So, P(same) = p^2 + q^2 and P(diff) = 2pq. Since the coin is unfair we know p != q. Now we use our key bit of arithmetic to say p^2 + q^2 > 2pq ⇒ P(same) > P(different).

COMPUTING CONDITIONAL PROBABILITY

The **conditional probability** that event A occurs given that event B occurs is:

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

The conditional probability is only well-defined if $\Pr(B) > 0$

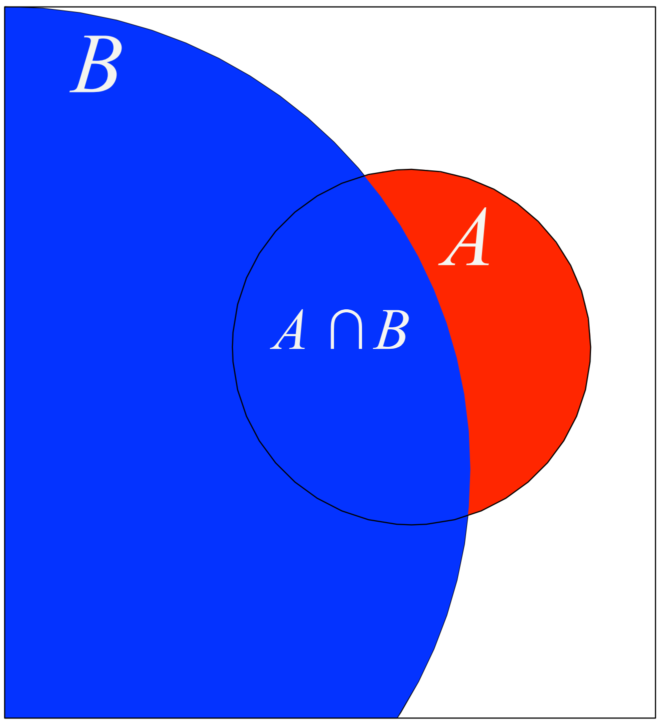
By conditioning on B we restrict the sample space to set B.

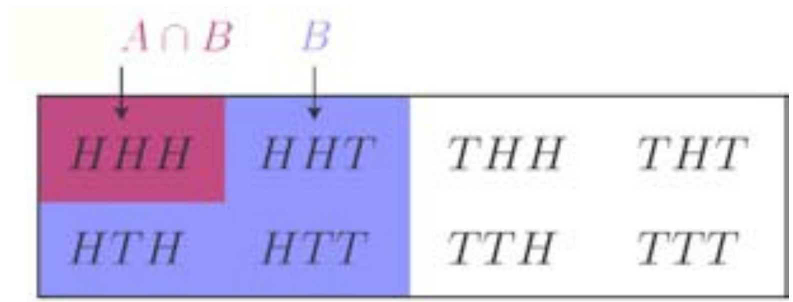
Thus we are interested in $\Pr(A \cap B)$ normalized by $\Pr(B)$.

CONDITIONAL PROBABILITY

'the probability of A given B'.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$





Conditional probability: Abstractly and for coin example

CONCEPT QUESTION

Toss a coin 4 times.
Let A = 'at least three heads'
Let B = 'first toss is tails'

1. What is $P(A|B)$?

a) 1/16 b) 1/8 c) 1/4 d) 1/5

2. What is $P(B|A)$?

a) 1/16 b) 1/8 c) 1/4 d) 1/5

CONCEPT QUESTION

Toss a coin 4 times.

Let A = 'at least three heads'

Let B = 'first toss is tails'

1. What is $P(A|B)$?
2. What is $P(B|A)$?

answer: 1. (b) 1/8. 2. (d) 1/5.
Counting we have |A ∩ B| = 1, |B| = 8, |A| = 5. Since all sequences are equally likely
P(A|B)= P(A ∩ B) = |A ∩ B|/|B|=1/8 P(B|A)= |B ∩ A|/|A|=1/5

DISCUSSION QUESTION

*"Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure and a passion for detail."**

What is the probability that Steve is a librarian?
What is the probability that Steve is a farmer?

*From *Judgment under uncertainty: heuristics and biases* by Tversky and Kahneman.

DISCUSSION QUESTION

*"Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure and a passion for detail."**

Discussion: Most people say that it is more likely that Steve is a librarian than a farmer. Almost all people fail to consider that for every male librarian in the United States, there are more than fifty male farmers. When this is explained, most people who chose librarian switch their Solution to farmer.
This illustrates how people often substitute representativeness for likelihood. The fact that a librarian may be likely to have the above personality traits does not mean that someone with these traits is likely to be a librarian.

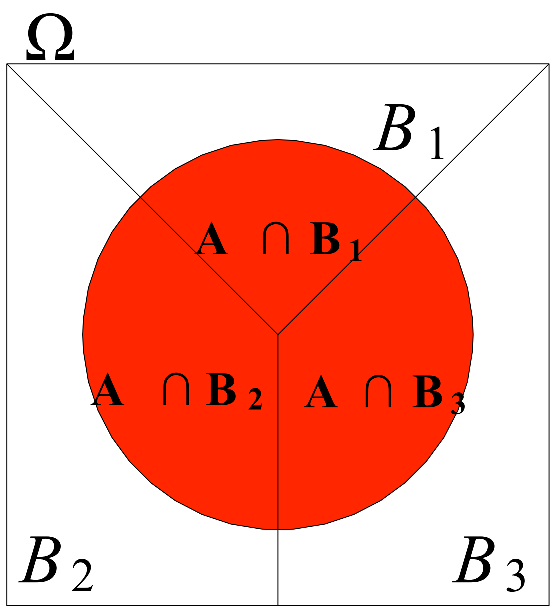
If the gender is changed to female, and now it was the case that there are an equal number of female librarians to female farmers, how would your answer change?

Multiplication Rule, Law of Total Probability

MR: $P(A \cap B) = P(A|B) \cdot P(B)$.

LoTP: If B_1, B_2, B_3 *partition* Ω then

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$
$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$



INDEPENDENCE

Events A and B are independent if the probability that one occurred is not affected by knowledge that the other occurred.

A and B are independent $\Leftrightarrow \; P(A \cap B) = P(A)P(B)$

$$\Leftrightarrow \; P(A \mid B) = P(A)$$

$$\Leftrightarrow \; P(B \mid A) = P(B)$$

Prove (or disprove) one of these expressions to show independence (or lack thereof).

CONCEPT QUESTION: INDEPENDENCE

Roll two dice.

A = 'first die is 3'

B = 'sum is 6'

A and B are independent

a) True   b) False

CONCEPT QUESTION: INDEPENDENCE

Roll two dice.

A = 'first die is 3'
B = 'sum is 6'

A and B are independent

a) True b) False

answer: False. P(B) = 5/36, P(B|A) = 1/6. Not equal ⇒ P(B)!=P(B|A) not independent.
Calculation B = |{(1+5, 2+4, 3+3, 4+2, 5+1}|=5 => P(B) = 5/36, |BA| = |{3+3}|=1, |A| = |{3}|=1

Can also show P(A|B) != P(A) or P(A ∩ B) != P(A)P(B).

CONCEPT QUESTION: INDEPENDENCE

Roll two dice.

A = 'first die is 3'
B = 'second die is 6'

A and B are independent

a) True b) False

CONCEPT QUESTION: INDEPENDENCE

Roll two dice.

A = 'first die is 3'

B = 'second die is 6'

A and B are independent

a) True b) False

Answer: True, 1/6 * 1/6 = 1/36

K-WISE INDEPENDENCE

More generally, events $E_1, E_2, \ldots, E_k$ are **mutually independent** if and only if for any subset $I \subseteq \{1,2,\ldots,k\}$

$$\boxed{\Pr \bigcap_{i \in I} E_i = \prod_{i \in I} \Pr(E_i)}$$

DISCUSSION QUESTION: INDEPENDENCE

Toss fair coin 10 times, each time it comes up heads.

What is the probability the next toss will be tails?

DISCUSSION QUESTION: INDEPENDENCE

Toss fair coin 10 times, each time it comes up heads.

What is the probability the next toss will be tails?

½ , we assume the tosses are independent (fair coin), so the prior observations will have no effect on the future, i.e.,
$P(T \mid HHHHHHHHHH) = P(T)$

EXAMPLE

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$$BB = \text{Two Boys}$$

$$B = \text{At least one kid is a boy}$$

What do we need to compute?

EXAMPLE

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$$BB = \text{Two Boys}$$

$$B = \text{At least one kid is a boy}$$

What do we need to compute?

BB = Two Boys, B= At least one kid is a boy, direct calculation:

P(BB | B) = P(BB & B) / P(B)

Or use Bayes Rule:

P(BB | B) = P(B | BB) · P(BB) / P(B) = 1 · 1/4 / 3/4 = 1/3

EXAMPLE

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

This is equivalent to asking what is P(BB|B)

$$BB = \text{Two Boys}, \text{GG} = \text{Two Girls}$$

$$B = \text{At least one kid is a boy}$$

$$P(BB) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$$

$$P(B) = 1 - P(GG) = 1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

$$P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Direct Calculation

BAYESIAN STATISTICS

BOARD QUESTION: EVIL SQUIRRELS

Of the one million squirrels on Brown's campus most are good-natured. But one hundred of them are pure evil! An enterprising CS student develops an "Evil Squirrel Alarm" which she offers to sell to Brown for a passing grade. Brown decides to test the reliability of the alarm by conducting trials.

EVIL SQUIRRELS CONTINUED

- When presented with an evil squirrel, the alarm goes off 99% of the time.
- When presented with a good-natured squirrel, the alarm goes off 1% of the time.

a) If a squirrel sets off the alarm, what is the probability that it is evil?

b) Should Brown co-opt the patent rights and employ the system?

ANALYSIS

	Evil	Nice	
Alarm	99	9999	10098
No alarm	1	989901	989902
	100	999900	1000000

Summary:

Probability a random test is correct $= \frac{99+989901}{1000000} = .99$

Probability a positive test is correct $= \frac{99}{10098} \approx .01$

These probabilities are not the same!

Bayes Theorem

Also called Bayes Rule and Bayes Formula.

Allows you to find $P(A|B)$ from $P(B|A)$, i.e. to 'invert' conditional probabilities.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

BAYES' LAW

$P(D|H)$: **Likelihood**
Probability of collecting this data when our hypothesis is true

$P(H)$: **Prior**
The probability of the hypothesis being true before collecting data

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

$P(H|D)$: **Posterior**
The probability of our hypothesis being true given the data collected

$P(D)$: **Marginal**
What is the probability of collecting this data under all possible hypotheses?

Evil Squirrels Bayes Solution

H = Evil

D = Alarm

	Evil	Nice	
Alarm	99	9999	10098
No alarm	1	989901	989902
	100	999900	1000000

answer: a) Let E be the event that a squirrel is evil. Let A be the event that the alarm goes off. By Bayes Theorem, we have:

$$\begin{aligned} P(E \mid A) &= \frac{P(A \mid E)P(E)}{P(A \mid E)P(E) + P(A \mid E^c)P(E^c)} \\ &= \frac{.99 \frac{100}{1000000}}{.99 \frac{100}{1000000} + .01 \frac{999900}{1000000}} \\ &\approx .01. \end{aligned}$$

b) No. The alarm would be more trouble than its worth, since for every true positive there are about 99 false positives.

BAYES THEOREM
(ALTERNATE FORM)

Let $E_1, E_2, \ldots E_n$ be **mutually disjoint** events in a sample space Ω, and $\bigcup_{i=1}^{n} E_i = \Omega$

Then:

$$\boxed{\Pr(E_j \mid B) = \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_{i=1}^{n} \Pr(B \mid E_i)\Pr(E_i)}}$$

$$\text{Conditional Probability: } \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\text{Law of Total Probability: } \Pr(B) = \sum_{j=1}^{n} \Pr(B \mid E_j)\Pr(E_j)$$

APPLICATION: FINDING A BIASED COIN

- We are given three coins. 2 coins are fair, and the 3rd is biased (landing heads with probability 2/3)
- **We need to identify the the biased coin**
- We flip each of the coins. The first and second come up heads, and the third comes up tails
- **What is the probability that the first coin was the biased one?**

APPLICATION: FINDING A BIASED COIN

Let E_i be the event that the *ith* coin flip is the biased one and let B be the event that the three coin flips came up HEADS, HEADS, and TAILS.

Before we flip the coins we have $\Pr(E_i) = 1/3$ for $i=1,\ldots,3$, thus

$$\mathrm{P}(B|E_1) = \mathrm{P}(B|E_2) = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6}$$

and

$$\Pr(B|E_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{12}$$

Applying Bayes we have

$$\Pr(E_1|B) = \frac{\mathrm{P}(B|E_1)\mathrm{P}(E_1)}{\sum_{i=1}^{3} \Pr(B|E_i)\Pr(E_i)} = \frac{2}{5}$$

The outcome HHT increases the probability that the first coin is the biased one from 1/3 to 2/5.

EXAMPLE REVISITED

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$$BB = \text{Two Boys, GG} = \text{Two girls}$$

$$B = \text{At least one kid is a boy}$$

$$P(BB) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$P(B) = 1 - P(GG) = 1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

$$P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Direct Calculation

$$\boxed{P(BB|B) = \frac{P(B|B\,B) * P(BB)}{P(B)} = \frac{1 * 1/4}{3/4} = \frac{1}{3}}$$

Bayes Calculation

IN CLASS: DRUG TEST

- 0.4% of Rhode Island PhDs use Marijuana*
- Drug Test: The test will produce 99% true positive results for drug users and 99% true negative results for non-drug users.
- If a randomly selected individual is tested positive, what is the probability he or she is a user?

$$P(User|+) = \frac{P(+|User)P(User)}{P(+)}$$

$$= \frac{P(+|User)P(User)}{P(+|User)P(User) + P(+|!User)P(!User)}$$

$$= \frac{0.99 \times 0.004}{0.99 \times 0.004 + 0.01 \times 0.996}$$

$$= 28.4\%$$

SPAM FILTERING WITH NAÏVE BAYES

$$P(spam|words) = \frac{P(spam)P(words|spam)}{P(words)}$$

$$P(spam|viagra, rich, ..., friend) = \frac{P(spam)P(viagra, rich, ..., friend|spam)}{P(viagra, rich, ..., friend)}$$

$$P(spam|words) \approx \frac{P(spam)P(viagra|spam)P(rich|spam)\ldots P(friend|spam)}{P(viagra, rich, ..., friend)}$$

Washington Post

Annual physical exam is probably unnecessary if you're generally healthy

> *For patients, the negatives include time away from work and possibly unnecessary tests. "Getting a simple urinalysis could lead to a false positive, which could trigger a cascade of even more tests, only to discover in the end that you had nothing wrong with you." Mehrotra says.*

http://www.washingtonpost.com/national/health-science/annual-physical-exam-is-probably-unnecessary-if-youre-generally-healthy/2013/02/08/2c1e326a-5f2b-11e2-a389-ee565c81c565_story.html