Basic Probability

Difference in set-up between sets 1 and 2 notes compared to set 3 notes

- Practical Statistics:

  We are interested in characteristics of a population (such as average height μ or proportion, p, of all tigers who will switch to roaming only at night if humans move into their territory).

  Data is available; we make inference on the properties of the unknown population

- Mathematical Probability:

  We are now dealing with mathematical theory. In this scenario, we have complete information about a population in the form of a census and so know the values of all population parameters. Or we know all about all possible experimental outcomes.

  We use this information and probability theory (subject of this set of notes) to calculate the chance a random experiment has a particular outcome.

  Population is completely known; we study how the data would be like.

Sample space and random events

Random experiment – any situation involving **chance** that leads to results called **outcomes.**

Examples of a random experiment:

- In a group of 20 cancer patients, each is randomly assigning to 1 of 2 therapies.
- Selecting a simple random sample (SRS) from a population and doing a survey.
- Selecting at random, one individual from the population (a SRS of size 1)

Practical definition of Probability:

$$P(A) = \text{probability that A occurs} = \frac{\text{number of experiments th at A occurs}}{\text{total number of repeated experiments}}.$$

Therefore, in reality, probability can be interpreted as long-run frequency of repeated experiments.

But in this course we are interested in constructing a more general and abstract mathematical definition of probability space.

Sample space (S) – A set that contains all possible outcome of an experiment.

Examples

If the experiment consists of tossing one coin, the sample space contains two outcomes, heads and tails, thus, S={H,T}.

If the experiment consists of tossing THREE coins, the sample space contains all possible combinations of heads and tails, S={HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}.

If the experiment is to toss coins indefinite times until a Head appears, then the sample space is S={H, TH, TTH, TTTH, ….}

Consider an experiment where the observation is reaction time to a certain stimulus. Here, the sample space would consist of all positive numbers, that is, $S=(0, \infty)$.

The sample space can be classified into two type: countable and uncountable. If the elements of a sample space can be put into 1-to-1 correspondence with a subset of integers, the sample space is countable. Otherwise, it is uncountable.

Event – Any subset of S, including S itself and empty set.

Simple event – An event that consists of one and only one element of S.
Compound event – An event that consists of more than one element of S.

Example:

If the experiment consists of tossing THREE coins, the sample space contains all possible combinations of heads and tails, S={HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}.
Event A={HHH} means exactly three Heads occur.
Event A={HHH, HHT} means HHH or HHT occurs, in other words, the first two are both Heads.

Set operations

- The **complement** of event A,
  - Denoted A^c
  - The event that both A AND B occur

- The **intersection** of two events A and B,
  - Denoted $A \cap B$

- The event that both A AND B occur

- The **union** of two events A and B
  - Denoted $A \cup B$
  - The event that either event A OR B occurs

- **Mutually exclusive events –** Also called **disjoint events**
  - When the experiment is performed a single time, if the occurrence of one of event precludes the possibility of the occurrence of the other event then the events are called mutually exclusive.
  - $A \cap B$ = empty set.

Example: Paulos (1988) tells the story of a weather forecaster on American TV who reported that there was a *50% chance of rain on Saturday* and *a 50% chance of rain on Sunday,* from which he concluded that there was a *100% chance of rain on the weekend.*

- What is wrong with the weather forecaster's reasoning?

Basic Probability Rules

Probability Axiom

Probability P is any function from events to real values between 0 and 1, which satisfies that

- For any event A, $0 \leqslant P(A) \leqslant 1$.
- P(S) = 1.
- $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = \sum P(A_i)$, if A_1, A_2, A_3 are disjoints events.

Constructing probability while S is countable**:**

Denote $S=\{s_1, s_2, s_3, \ldots\}$, find any sequence of $p_1, p_2, p_3, \ldots$, such that $\sum p_i = 1$. Then define $P(A) = \sum_{\{i: s_i \in A\}} p_i$.

Probability equations

- $P(A^c) = 1 - P(A)$.
- $P(\varnothing) = 0$.
- $P(A) \leqslant 1$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- $P(A \cup B \cup C) =$
- If the events A and B are mutually exclusive then

  $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$.

Example of finding probabilities:

Booth et al (2003, Journal of the American Medical Association, 289, 2801-9) report the results of a study conducted in 2003, involving 144 adult patients admitted to one of 10 hospitals in the greater Toronto area. The patients had a diagnosis of suspected or probable severe acute respiratory syndrome (SARS). The earliest symptoms of SARS are shown in the table below. Suppose that a patient is selected at random from the 144 patients in the study.

Note: P(randomly chosen patient falls into a category) = proportion of patients in that category.

Symptom	% of patients
Fever without prodome or cough	29%
Fever with just prodome	23%
Fever with just cough	11%
Fever with both prodome and cough	11%
Cough alone	9%
Prodome with just cough	4%
Prodome alone	13%

- The probability that the patient reported a cough
- The probability that the patient reported none of the three symptoms.

Counting

Counting technique is used to probability calculation when the sample space is finite with equal probability, i.e., $S=\{s_1,\ldots,s_N\}$ and $P(\{s_i\})=1/N$, then

$$P(A)=\frac{N(A)}{N},$$

where N(A) is the number of elements in the event A.

How to count N(A) and N

Proposition: If a job consists of k separate tasks, the i^{th} of which can be done in n_i ways for i=1,…,k, then the entire job can be done in $\prod n_i$ ways.

Permutation: An ordered subset is called a permutation. The number of permutations of size k that can be formed from the n individuals will be denoted by $P_{n,k}=n(n-1)...(n-k+1)=n!/(n-k)!$.

Proposition: An unordered subset is called a combination. The number of permutations of size k that can be formed from the n individuals will be denoted by $C_{n,k} = \binom{n}{k}$ =n!/[(n-k)!k!].

Examples:

A particular iPod playlist contains 100 songs, 10 of which are by the Beatles. Suppose the shuffle feature is used to play the songs in random order. What is the probability that the first Beatles song heard is the fifth song played?

P(the first Beatles song heard is the fifth song played)

$$= \frac{90*89*88*87*10}{100*99*98*97*96} = \frac{\binom{95}{9}}{\binom{100}{10}} .$$

Consider choosing a five-card poker hand from a standard deck of 52 playing cards. What is the probability of having four aces? If we specify that four of the cards are aces, then there are 48 different ways of specifying the fifth card. Thus,

P(four aces)= $\dfrac{48}{\binom{52}{5}}$;

P(four of a kind)= $\dfrac{13*48}{\binom{52}{5}}$;

P(exactly one pair)= $\dfrac{13*\binom{4}{2}\binom{12}{3}4^3}{\binom{52}{5}}$.

<u>Condition probability</u>

Motivating example: Suppose you want to calculate the probability a randomly chosen person is colorblind. In the US, 4.25% of the population is colorblind.

- P(colorblind) = 0.0425

Now, suppose you have additional information. Namely, the person to be chosen is male – that means, we are restricting ourselves to a subset of the population.

- Since 8% of all men are colorblind, the probability has changed to .08

When we have additional information, we must *"condition"* our probability based on this additional information, or based on this sub-population. The probability adjusted for the additional information is called a <u>conditional</u> probability.

Notation:

A = event of interest. In our example, A = *colorblind*

B = event that is known to occur. In our example, B = *male*

- In symbols this adjusted probability is written **P(A|B).** The vertical line in the formula immediately before B designates conditional probability.
    - For our example, we'd write P(colorblind | male) = .08
- Wording: P(A|B) is called *the conditional probability of A given that B has occurred*.
    - For our example, we'd say the probability a randomly chosen person is colorblind given the person chosen is male is .08.

Consider an equal probability case,

$$P(A|B) = \frac{N(A \cap B)}{N(B)} = \frac{N(A \cap B)/N}{N(B)/N} \quad .$$

Example: Number of accidents in which they were involved in the last year, giving the results in the table below. A driver is selected at random from the 1000 drivers. Find the following probabilities:

	Age < 30	Age 30 or older	Totals
At most one accident	130	170	300
More than one accident	470	230	700
Totals	600	400	1000

- The probability that the driver has had more than one accident and is aged less than 30
- The probability that the driver is aged less than 30
- A driver is selected at random from the 600 drivers aged less than 30; what is the probability that the driver has had more than one accident?

General Formula: When P(B) > 0, the conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad .$$

All the probability axioms hold for the conditional probability.

Example: A news magazine publishes three columns entitled “Art” (A), “Books” (B), and “Cinema” (C). Reading habits of a randomly selected reader with respect to these columns are

Read regularly	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
Probability	.14	.23	.37	.08	.09	.13	.05

$P(A|B)$ =

$P(A|B \cup C)$ =

$P(A|\text{reads at least one})$ =

$P(A \cup B|C)$ =

Multiplication rule

Use conditional probability to calculate the probability of intersection:

$$P(A \cup B) = P(A|B) * P(B) = P(B|A) * P(A)$$

Example using the multiplication rule:

An individual’s blood type is described by the ABO system and the Rhesus factor.

In the American population 16% of individuals have a negative Rhesus factor (Rh−), and 43.75% of those with Rh− are blood type O.

- What percent of Americans are Rh− *and* have the blood type O?
- **Solution:** Use the general multiplication rule.

  You are given the following probabilities: P(Rh–) = 0.16 and P(O| Rh–) = 0.437

  By the multiplication rule: P(Rh– and O) = P(Rh–) * P(O| Rh–) = (0.16)(0.4375) = 0.07.

Example using the multiplication rule and tree diagram:

Intense or repetitive sun exposure can lead to skin cancer. Patterns of sun exposure, however, differ between men and women.

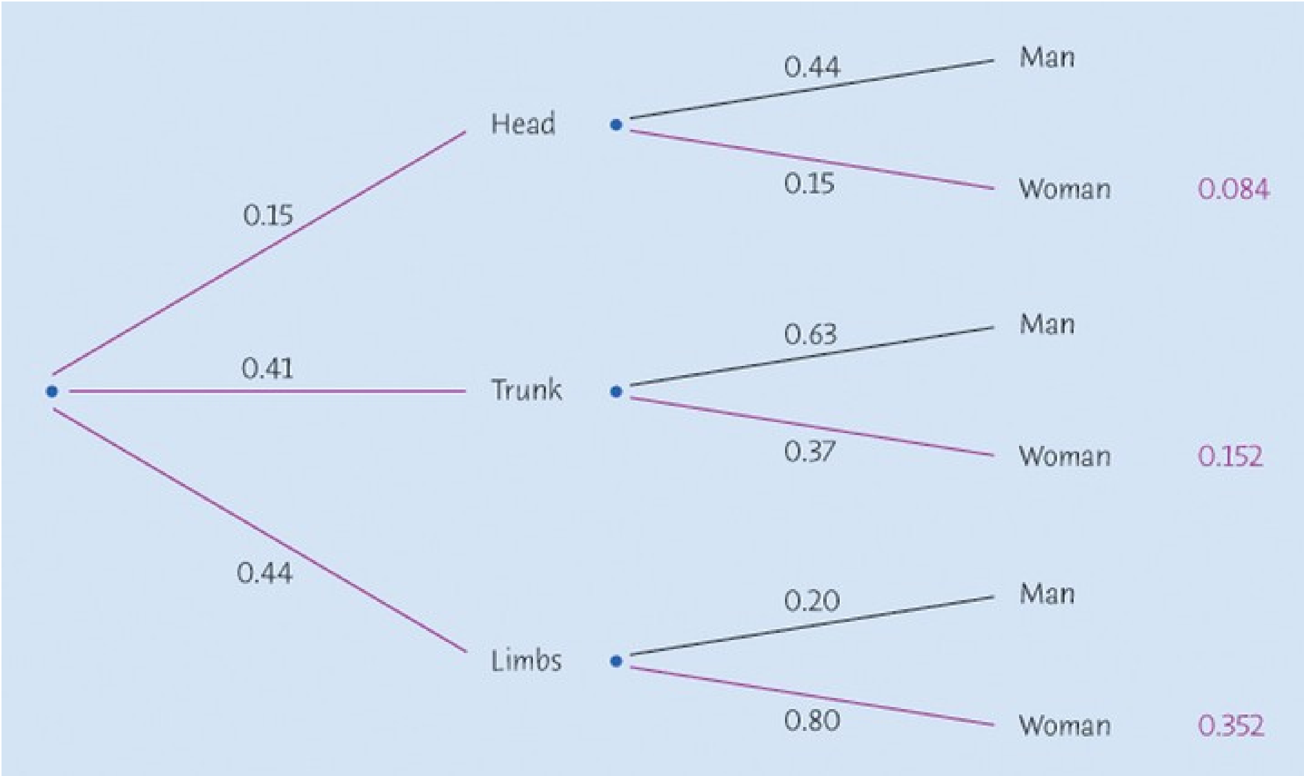
A study of cutaneous malignant myeloma in the Italian population found that 15% of skin cancers are located on the head and neck, another 41% on the trunk, and the remaining 44% on the limbs.

Moreover, 44% of the individuals with a skin cancer on the head are men, as are 63% of those with a skin cancer on the trunk and only 20% of those with a skin cancer on the limbs are men.

- We are told the following probabilities

$$P(\text{head}) = 0.15 \qquad P(\text{man} \mid \text{head}) = 0.44$$
$$P(\text{trunk}) = 0.41 \qquad P(\text{man} \mid \text{trunk}) = 0.63$$
$$P(\text{limbs}) = 0.44 \qquad P(\text{man} \mid \text{limbs}) = 0.20$$

.56



- What percent of all individuals with skin cancer are women?

- What percent of women with skin cancer have the cancer on the limbs?

$$P(limbs|woman)=\frac{P(limbs\cap woman)}{P(woman)}=\frac{0.352}{0.084+0.152+0.352}=0.599$$

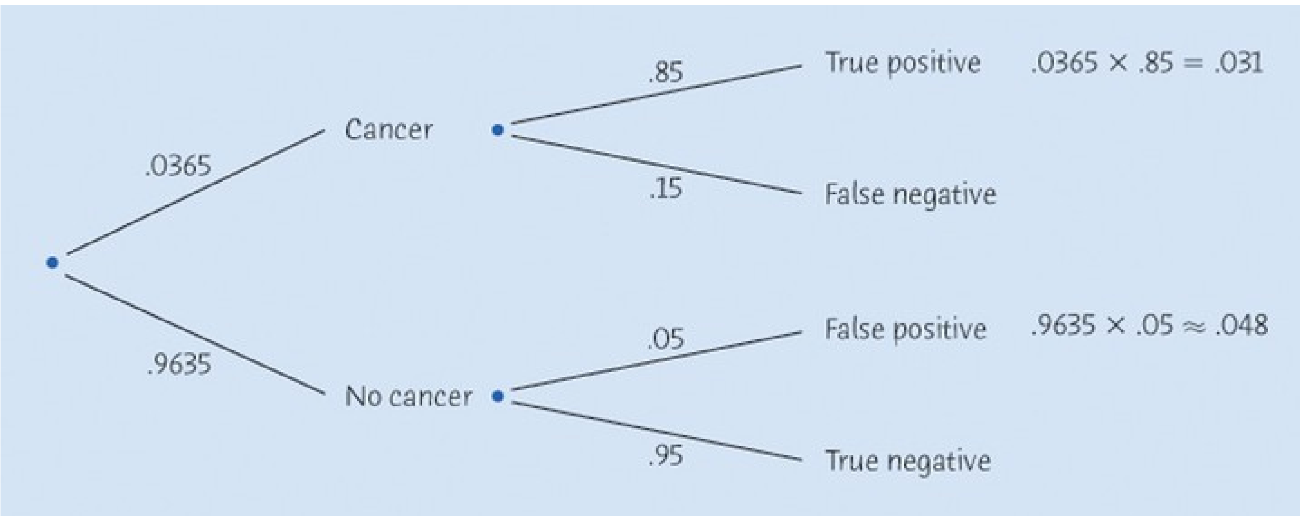
Bayes' Formula

- Used when: Know ***P*(*B*|*A*)** but want ***P*(*A*|*B*).**
- For example, diagnostic tests provide *P*(Test positive | disease) but we are interested in *P*(disease | Test positive)
- Bayes' Theorem provides a method for finding *P*(*A*|*B*) from *P*(*B*|*A*)
- Suppose that $A_1, A_2, \ldots, A_k$ are disjoint events whose probabilities are not 0 and add to exactly 1. That is, any outcome has to be exactly in one of these events. Then if *B* is any other event whose probability is not 0 or 1,

$$P(A_i|B)=\frac{P(A_i\cap B)}{P(B)}=\frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1)+\ldots+P(B|A_k)P(A_k)}$$

Bayes' formula is a mathematical representation of tree diagram method

- Breast cancer occurs most frequently among older women. Of all age groups, women in their 60s have the highest rate of breast cancer. The National Cancer Institute (NCI) compiles U.S. epidemiology data for a number of different cancers. The NCI estimates that 3.65% of women in their 60s get breast cancer.
- Mammograms are X-ray images of the breast used to detect breast cancer. A mammogram can typically identify correctly 85% of cancer cases and 95% of cases without cancer.
- If a woman in her 60s gets a positive mammogram, what is the probability that she indeed has breast cancer?
- From the above we know the following probabilities:



Solve it by definition of condition probability $P(cancer|\text{positive test}) = \dfrac{P(cancer \cap \text{positive test})}{P(\text{positive test})}$

Solve it by Bayes' formula:

$$\begin{aligned}P(\text{cancer} \mid \text{test+}) &= \frac{P(\text{test+} \mid \text{cancer})P(\text{cancer})}{P(\text{test+} \mid \text{cancer})P(\text{cancer}) + P(\text{test+} \mid \text{no cancer})P(\text{no cancer})} \\ &= \frac{(0.85)(0.0365)}{(0.85)(0.0365) + (0.05)(0.9635)} \\ &= \frac{0.031}{0.031 + 0.048} = 0.392\end{aligned}$$

Diagnostic tests in medicine

- The performance of a diagnostic test for a given disease can be defined by two properties.

  1) The first property is the test's ability to appropriately give a positive result when a person tested has the disease. This is the test's **sensitivity**, which is *P*(positive test | disease).

  2) The second property is the test's ability to come up negative when a person tested doesn't have the disease. This is the test's **specificity**, which is *P*(negative test | no disease).

		True health status	
		Disease	No disease
Test result	Positive		
	Negative		

(a)

Independence

- Two events *A* and *B* that both have positive probability are **independent** if

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

- If A and B are independent, then A^c and B are also independent, since

$$P(A^c|B) = 1 - P(A|B) = 1 - P(A) = P(A^c)$$

- Because of the multiplication rule, this also implies that

$$P(A \cap B) = P(A) * P(B)$$

- Events $A_1, A_2, \ldots, A_n$ are called **mutually independent**, if

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) * \ldots * P(A_{i_k})$$

for any subest indices $i_1, i_2, \ldots, i_k$.

Example: Recall that 10000 drivers were questioned and classified according to their age and the number of accidents in which they were involved in the last year, giving the results in the table below.

	Age < 30	Age 30 or older	Totals
At most one accident	130	170	300
More than one accident	470	230	700
Totals	600	400	1000

How can we use this to tell if ***Age*** and ***More than One Accident*** are independent?

Example: Each day, Monday through Friday, a batch of components sent by a first supplier arrives at a certain inspection facility. Two days a week, a batch also arrives from a second supplier. Eighty percent of all supplier 1's batches pass inspection, and 90% of supplier 2's do likewise. What is the probability that, on a randomly selected day, two batches pass inspection? We will answer this assuming that on days when two batches are tested, whether the first batch passes is independent of whether the second batch does so.

Example: If we toss 6 dices, what is probability that the largest value is smaller than 4. What is probability that the smallest value is smaller than 4.

Example: Let an experiment consists of tossing two dice. Define events A = {double appear}, B = {the sum is between 7 and 10} and C = {the sum is 2, 7 or 8}. Are A, B and C mutually independent?

Example: Textbook example 2.36

Example from Rosner, *Fundamentals of Biostatistics,* 7th ed. Page 43.

Suppose we are conducting a hypertension-screening program in the home. We want to compute the probability that both mother and father are hypertensive. Suppose P(mom hypertensive) = 0.10 and P(dad hypertensive) = 0.20. Then, **if** the mom and dad are hypertensive *independently* of one another, P(both are hypertensive) = P(mom hypertensive) P(dad hypertensive) = (.10)(.20) = 0.02.

One way to interpret this example is to assume that the hypertensive status of the mom doesn't depend at all on the hypertensive status of the dad. If this is true, then 10% of all households where the father is hypertensive, the mom is also hypertensive. This is an application of the formula at the top of this page. [P(mom hypertensive | dad hypertensive) = P(mom hypertensive) when events are independent.]

We would expect these two events (mom hypertensive, dad hypertensive) to be independent if the primary determinant of elevated blood pressure is genetic.

However, if the primary determinant of elevated blood pressure is not genetic but instead was to something environmental, then we would expect the mother would be more likely to have elevated blood pressure if the father had elevated blood pressure than if the father did not have elevated blood pressure. The reason for this is that both husband and wife will be subject to the same stresses like money issues, poor health, etc.

The Law and Probability

The Collins Case

In 1964, an elderly woman was mugged in a lane way. Shortly afterward, a witness saw a blonde girl --- her ponytail flying --- run out of the lane way and get into a yellow car driven by a bearded black man with a moustache, and speed away together.

Eventually, the police arrested Janet and Malcolm Collins, a married couple who matched the description given above and who also owned a yellow car.

What are the chances that a randomly selected couple meets the description?

The prosecuting lawyer invoked the laws of probability. The following probabilities were put to the jury:

- *P*(Woman having blonde hair) = 1/3
- *P*(Woman having a ponytail) = 0.1
- *P*(Being an interracial couple) = 0.001
- *P*(Driving a yellow car) = 0.1
- *P*(Man having a moustache) = 0.25
- *P*(Black man wearing a beard) = 0.1

The prosecutor multiplied the numbers using the multiplication rule for <u>independent events</u> to produce a joint probability of

$$1/3 \times 0.1 \times 0.001 \times 0.1 \times 0.25 \times 0.1 = 0.000000833$$

There seems little doubt that Janet and Malcolm Collins are guilty, right?

The jury found them guilty beyond a reasonable doubt. Malcolm Collins received a sentence of “one year to life”; Janet Collins received a sentence of “not less than one year”.

Yet, a few years later, the California Supreme Court overturned the verdict.
There was no evidence, it said, that the individual probability values were even roughly accurate. Most important, they were shown to be not independent of one another, as they must be to satisfy the probability law that was applied. The court found that a black man having a beard is not independent of his being an inter-racial marriage. Indeed, the court heard evidence that there was ***a 41 percent chance that at least one other couple in the area might have had the admittedly unusual characteristics listed above***.

The Sally Clark Case

Sally Clark, a lawyer in the UK, was convicted and jailed for life in November 1999 of murdering two of her three sons (her 11 week old son, Christopher, died in December 1996 while her eight week old, Harry, died in January 1998). Sally Clark spent more than three years in prison before being cleared by the Court of Appeal, at the second appeal in January 2003. Sally Clark died on March 15, 2007, aged 42.

Professor Roy Meadow testified at the first trial that the chance of two children from the same family suffering Sudden Infant Death Syndrome (SIDS) was 1 in 73 million, which was obtained by multiplying 1/8500 (the probability of a single incidence of SIDS) by 1/8500. Professor Meadow also concluded that having two SIDS deaths in a single family is expected to happen in England about once every 100 years.

The Clark case has been the subject of much scholarly debate. For example, the *British Medical Journal* published an editorial by Stephen Watkins in January 2000 entitled ‘Convicted by mathematical error?’ Watkins argues:

> *“Cot deaths are not random events [i.e. independent]. There have been several studies of recurrence [i.e. two cot deaths]. At least one study did show no increase in recurrence rates, several others showed recurrence rates about five times the general rate, implying recurrence somewhere in England about once every year and a half [and not once every 100 years as claimed at the original Clark trial].”*

In July 2005, the General Medical Council struck off pediatrician Professor Sir Roy Meadow after his "misleading" evidence in the Sally Clark case. The GMC found that Sir Roy had been found guilty of serious professional misconduct. Sir Roy had stood by his evidence, but admitted his use of statistics at Mrs Clark's 1999 trial was "insensitive". The GMC found that Sir Roy's conduct had been "fundamentally unacceptable". (Source: http://news.bbc.co.uk/2/hi/health/4685511.stm)

Prosecutor’s Fallacy

We discussed court cases which involved probability. In each case, the prosecution misused the multiplication rule to calculate the probability that the defendant matched observed characteristics, under the assumption that they were innocent. We define the events:

- E: evidence against the defendant
- I: defendant is innocent.

The prosecution is both cases tried to calculate $P(E|\ I)$, the probability of observing the evidence against the defendant if the defendant is innocent. In other words, *if* the person is innocent, what is the probability that *by chance* these events could have occurred. However, juries often $P(I\ |E)$ = given the evidence, the probability the defendant is innocent with $P(E|\ I)$.

Confusing $P(E|\ I)$ and $P(I\ |E)$ is known as the **prosecutor's fallacy**. As we have seen, reversing the direction of a conditional probability can produce a vastly different answer.