Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 1: PROBABILITY THEORY

Key points about probability

1. Rules → data: Given the rules, describe the likelihoods of various events occurring.
2. Probability is about prediction — looking forward.

Key points about statistics

1. Rules ← data: Given only the data, try to guess what the rules were. That is, some probability model controlled what data came out, and the best we can do is guess — or approximate — what that model was. We might guess wrong; we might refine our guess as we get more data.
2. Statistics is about making inference — looking backward.

Classical or a priori probability

- The classical definition of probability is prompted by the close association between the theory of probability of the early ages and games of chance.

  Classical probability: If a random experiment can result in n mutually exclusive and equally likely outcomes and if n_A of these outcomes have an attribute A, then the probability of A is the fraction n_A/n.

- In this context

  An event: a possible outcome or set of possible outcomes of an experiment or observation.

Axiomatic definition of probability

- For S_e an algebra of events, a probability function P(.) is a set function with domain S_e and counterdomain the interval [0,1], which satisfies the following axiom:
  - Axiom 1: $P(A) \geq 0$ (nonnegative), for every event A
  - Axiom 2: $P(S_e) = 1$ (normed)
  - Axiom 3: For a countable number of mutually exclusive events $A_1, A_2, ...$ in S_e, and if the union of these events is itself an event,

$$P(A_1 \cup A_2 \cup ...) = P(\sum_j A_j) = \sum_j P(A_j) \quad \text{(additive)}$$

The 4-step method to solve practical problems

1. Find the sample space
2. Define the events of interest
3. Assign outcome probabilities
4. Compute event probabilities

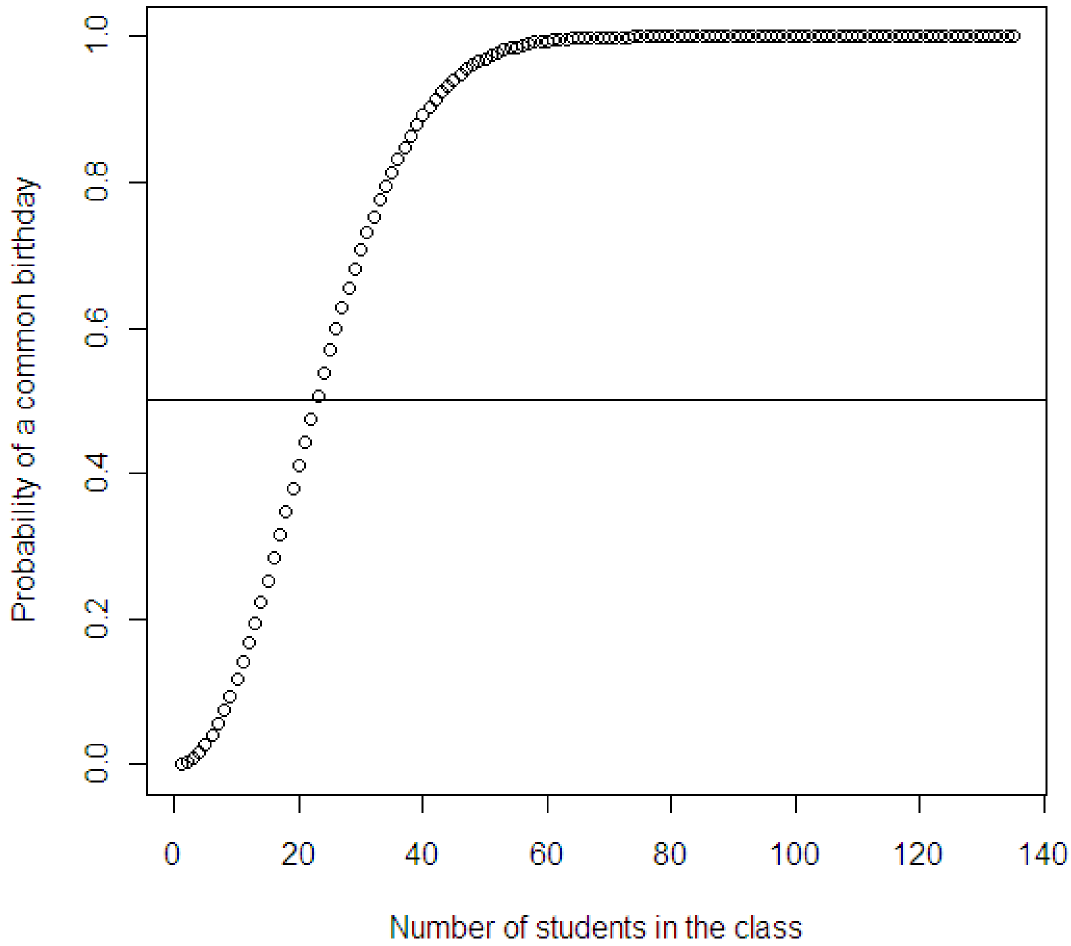
The birthday problem can be solved that way (see class 1 notes)

- It can be shown that there is an asymptotically tight solution to the birthday problem:

$$\Pr(\text{all } m \text{ birthdays different}) \sim e^{-\frac{m(m-1)}{2N}}$$

- For which m is this solution = ½? ***"the birthday principle"***

$$m \approx \sqrt{(2\ln 2)N} \approx 1.18\sqrt{N}.$$


Probability of a common birthday
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140
Number of students in the class

But it is not always that easy to delineate the sample space…

“Is 1 + 1 = 1 in probability theory?

Consider tossing a coin and throwing a dice.

Let the set of all possible outcomes for the coin be C. which implies p(C) = 1.

Let the set of all possible outcomes for the dice be D, which implies p(D) = 1.

Now p(C∪D) which is the probability that either the events D or C occur is also 1.

C and D are disjoint sets and therefore p(C∪D)= p(C) + P(D) which implies 1 = 1 + 1.

Help! “

The answer lies in a proper delineation of the sample space for this problem. If you throw either dice or a coin but you do not know (or do not specify) which, then the sample space is

{H, T, 1, 2, 3, 4, 5, 6}

so that P(C) = 1 and P(D) = 1 are both false.

If you throw both a dice and a coin then the sample space is

{H, T} × {1, 2, 3, 4, 5, 6}

in which case the events C and D are simply not defined.

If you just throw a coin then certainly P(C) = 1. If, in another experiment, you throw a dice then, too, P(D) = 1. But in this case the event CUD is undefined because the events C and D do not belong to the same space.

(http://www.cut-the-knot.org/)

Corresponding statements in set theory and probability: easy to remember probability “rules” and definitions

Set theory	Probability theory
Space, S	Sample space, sure event
Empty set, $\emptyset$	Impossible event
Elements $a, b, \ldots$	Sample points a, b, ... (or simple events)
Sets $A, B, \ldots$	Events A, B, ...
A	Event A occurs
$\overline{A}$	Event A does not occur
$A \cup B$	At least one of A and B occurs
AB	Both A and B occur
$A \subset B$	A is a subevent of B (i.e. the occurrence of A necessarily implies the occurrence of B)
$AB = \emptyset$	A and B are mutually exclusive (i.e. they cannot occur simultaneously)

Rules of probability using set theory

Special addition rule

- If $(A \cap B) = 0$, the events are mutually exclusive, so

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

- In general, if several events $A_1, A_2, ..., A_k$ are mutually exclusive (i.e., at most one of them can happen in a single experiment), then

$$P(A_1 \,\text{or}\, A_2 \,\text{or} \ldots \text{or}\, A_k) = P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + \ldots + P(A_k) = \sum_k P(A_k)$$

Boole's inequality for events $A_1, A_2, ..., A_n$

$$P[A_1 \cup A_2 \cup \cdots \cup A_n] \leq P[A_1] + P[A_2] + \cdots + P[A_n].$$

PROOF $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 A_2] \leq P[A_1] + P[A_2].$

The proof is completed using mathematical induction. ////

Multiplication Rule

We can re-arrange the definition of the conditional probability

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

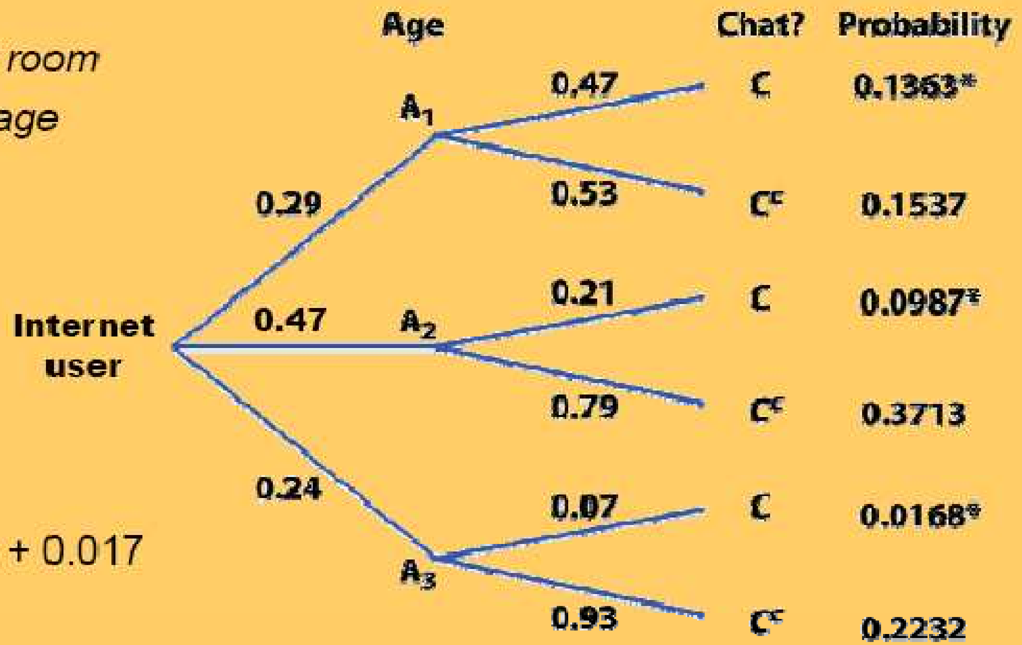
to obtain equivalent expressions for $P(A \text{and } B)$:

$$P(A \cap B) = \begin{cases} P(A|B)P(B) \\ P(B|A)P(A) \end{cases}$$

You can often think of $P(A \text{ and } B)$ as being the probability of first getting A with probability $P(A)$, and then getting B with probability $P(B|A)$. This is the same as first getting B with probability $P(B)$ and then getting A with probability $P(A|B)$.

Conditional probabilities can get complex, and it is often a good strategy to build a **probability tree** that represents all possible outcomes graphically and assigns conditional probabilities to subsets of events.

Tree diagram for chat room habits for three adult age groups



P(chatting) $= 0.136 + 0.099 + 0.017$

$= 0.252$

About 25% of all adult Internet users visit chat rooms.

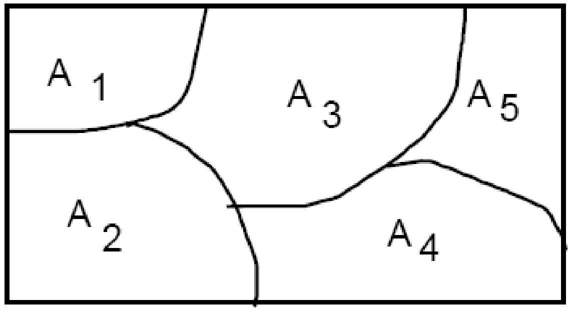
Special Multiplication Rule

If two events *A* and *B* are *independent* then P(A| B) = P(A) and P(B| A) = P(B): knowing that *A* has occurred does not affect the probability that *B* has occurred and vice versa. In that case

$P(A \text{ and } B) = P(A \cap B) = P(A)\, P(B)$

Probabilities for any number of independent events can be multiplied to get the joint probability. For example if you toss a fair coin twice, the outcome of the first throw shouldn't affect the outcome of the second throw, so the throws are independent.

The law of total probability: relating the prob of an event to cond probs



If $A_1, A_2, \ldots, A_k$ form a partition (a mutually exclusive list of all possible outcomes) and B is any event then

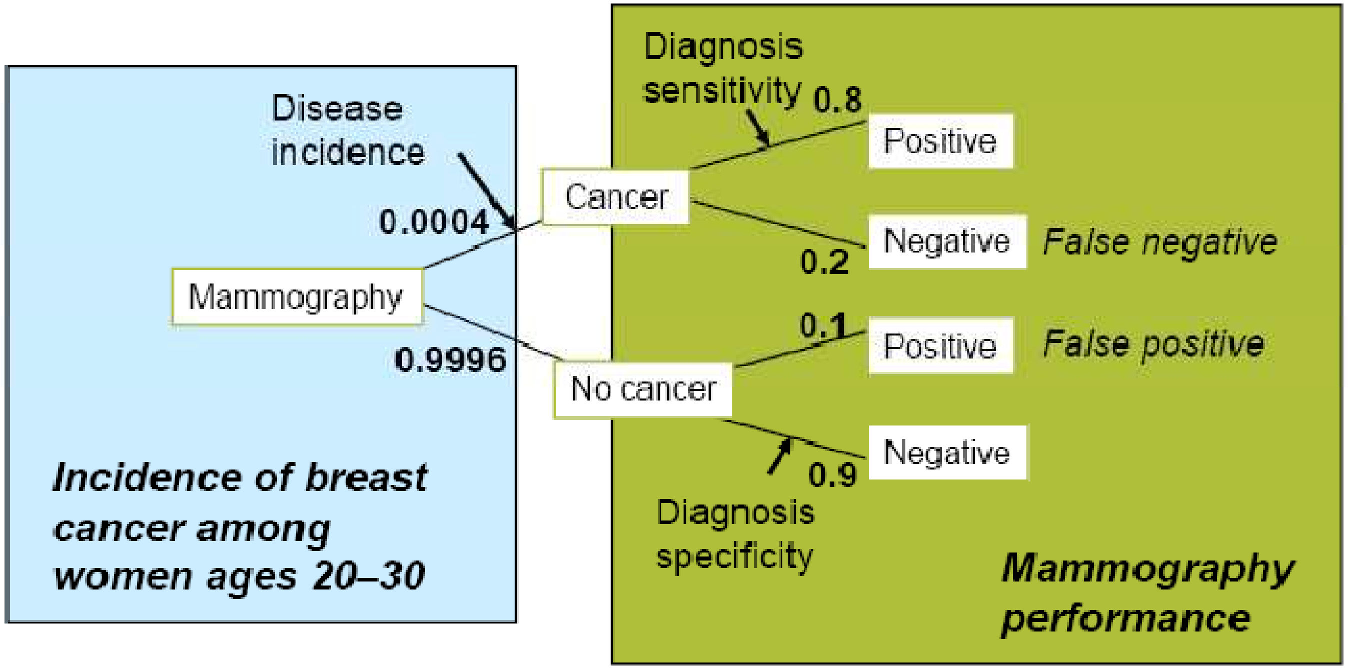
$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k) = \sum_k P(B|A_k)P(A_k)$$

Proof: This follows since

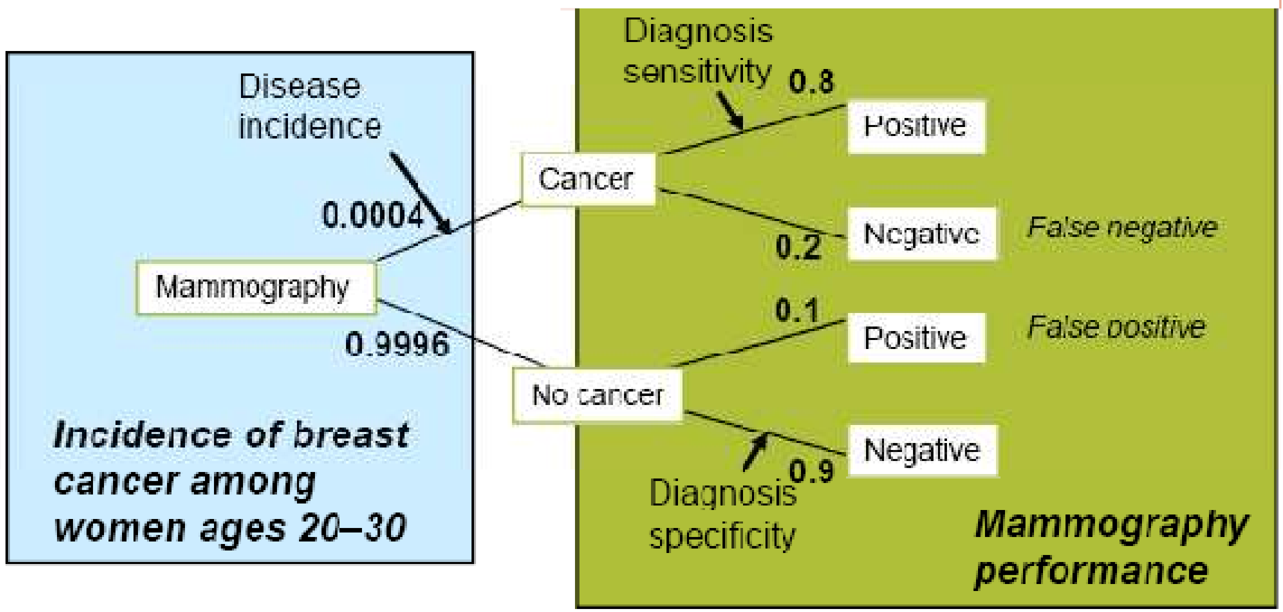
$$\begin{aligned} \mathrm{P}(B) &= \mathrm{P}(B \mid A_1)\mathrm{P}(A_1) + \mathrm{P}(B \mid A_2)\mathrm{P}(A_2) + \ldots + \mathrm{P}(B \mid A_k)\mathrm{P}(A_k) \\ &= \mathrm{P}(B \cap A_1) + \mathrm{P}(B \cap A_2) + \ldots + \mathrm{P}(B \cap A_k) \\ &= \mathrm{P}(B \cap A_1 \text{ or } B \cap A_2 \text{ or.. + or } B \cap A_k) \\ &= \mathrm{P}(B \cap (A_1 \text{ or } A_2 \text{ or } A_k)) \\ &= \mathrm{P}(B) \end{aligned}$$

Breast cancer screening

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?



She could either have a positive test and have breast cancer or have a positive test but not have cancer (false positive).



Possible outcomes given the positive diagnosis: positive test and breast cancer or positive test but no cancer (false positive).

$$P(cancer \mid pos) = \frac{P(cancer\ and\ pos)}{P(cancer\ and\ pos) + P(nocancer\ and\ pos)}$$

$$= \frac{0.0004 * 0.8}{0.0004 * 0.8 + 0.9996 * 0.1} \approx 0.3\%$$

This value is called the positive predictive value, or *PV+*. It is an important piece of information but, unfortunately, is rarely communicated to patients.

Bayes' Theorem

The multiplication rule gives $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.
Bayes' theorem follows by diving through by $P(B)$ (assuming $P(B) > 0$):

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

This is an incredibly simple, useful and important result. If you have a model that tells you how likely X is given Y, Bayes' theorem allows you to calculate the probability of Y if you observe X. This is the key to learning about your model from statistical data.

Note: often the Total Probability rule is often used to evaluate P(B):

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

Principle of proportionality

- This is an immediate consequence of Bayes' Theorem.
- If various alternatives are equally likely, and then some event is observed, the updated probabilities for the alternatives are proportional to the probabilities that the observed event would have occurred under those alternatives.

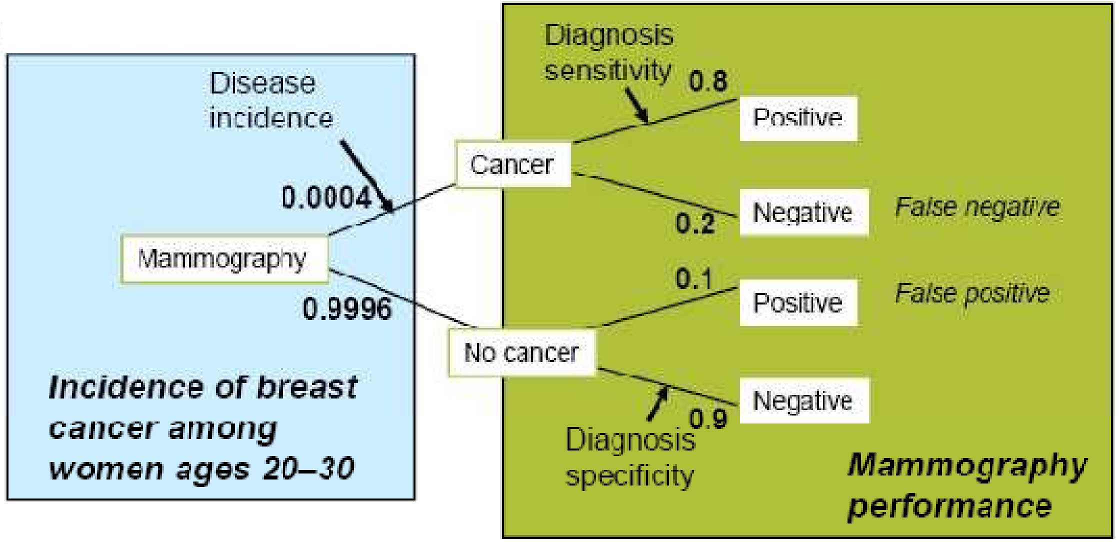
The formal derivation is simple. Assume

(*) $P(A_1) = P(A_2) = ... = P(A_n) > 0$ and $P(B) > 0$.

Then $P(A_m|B) = K\,P(B|A_m)$, for all $m = 1, 2, ..., n$, where $K > 0$ does not depend on m.

Breast cancer screening example: application of Bayes' theorem

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?



This time, we use Bayes's rule: $P(A_i \mid C) = \dfrac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + P(C \mid A_2)P(A_2) + \cdots + P(A_k)P(C \mid A_k)}$

A1 is cancer, A2 is no cancer, *C* is a positive test result.

$$P(cancer \mid pos) = \frac{P(pos \mid cancer)P(cancer)}{P(pos \mid cancer)P(cancer) + P(pos \mid nocancer)P(nocancer)}$$

$$= \frac{0.8 * 0.0004}{0.8 * 0.0004 + 0.1 * 0.9996} \approx 0.3\%$$

Bayesian odds

- On occasion when there are two events, say *A* and *B*, whose comparative **posterior probabilities** are of interest, it may be more advantageous to consider the ratios, i.e.:

$$\frac{p(A|C)}{p(B|C)} = \frac{p(C|A)}{p(C|B)} \cdot \frac{p(A)}{p(B)}.$$

- Ward Edwards gives a simple example where the latter formula comes in handy:
  There are two bags, one containing 700 red and 300 blue chips, the other containing 300 red and 700 blue chips. Flip a fair coin to determine which one of the bags to use. Chips are drawn with replacement. In 12 samples, 8 red and 4 blue chips showed up. What is the probability that it was the predominantly red bag?

.

Solution:

Author Edwards writes

Clearly the sought probability is higher than 0.5.

Is it?

Let *A* be the event of selecting the first bag. Let *B* be the event of selecting the second bag. Finally, let C be the result of the experiment, i.e., drawing 8 red and 4 blue chips from the selected bag. Clearly,

$$p(C|A) = (\frac{7}{10})^8 (\frac{3}{10})^4$$
$$p(C|B) = (\frac{7}{10})^4 (\frac{3}{10})^8$$

so that $\frac{p(C|A)}{p(C|B)} = \left(\frac{7}{3}\right)^4 \approx 29.642$.

Now, $p(A)=p(B)=0.5$, implying that

$$\frac{p(A|C)}{p(B|C)} = \frac{p(C|A)}{p(C|B)} \times 1 = 29.642$$

From $p(A|C)+p(B|C)=1$, it then follows that

$$\frac{p(A|C)}{1-p(A|C)} = 29.642$$ [this is an odds!!!]

and

$$p(A|C) \approx \frac{29.642}{1+29.642} = \frac{29.642}{30.642} \approx 0.967$$

(http://www.cut-the-knot.org/)

Odds

- Note that by our assumption of equal probabilities for the events A and B,

$$\frac{p(A|C)}{p(B|C)} = \frac{p(A|C)}{1 - p(A|C)}$$

  and is therefore a genuine *odds*.

- The experts on this issue live just south of here in a town called Peculiar, Missouri. The sign just outside city limits reads "Welcome to Peculiar, where the odds are with you." ☺ ☺ ☺

- Odds are just an alternative way of expressing the likelihood of an event such as catching the flu. Probability is the expected number of flu patients divided by the total number of patients. Odds would be the expected number of flu patients divided by the expected number of non-flu patients.

- During the flu season, you might see ten patients in a day. One would have the flu and the other nine would have something else.
  - So the probability of the flu in your patient pool would be one out of ten.
  - The odds would be one to nine.
- It's easy to convert a probability into an odds. Simply take the probability and divide it by one minus the probability:

odds = probability / (1-probability)

- If you know the odds in favor of an event, the probability is just the odds divided by one plus the odds.

probability = odds / (1+odds)

- You should get comfortable with converting probabilities to odds and vice versa. Both are useful depending on the situation.

3.4 A posteriori or frequency probability

Limitations of the classical definition: how to assign numbers to "probabilities of events"

Classical probability: If a random experiment can result in n mutually exclusive and equally likely outcomes and if n_A of these outcomes have an attribute A, then the probability of A is the fraction n_A / n.

- Limitation 1: The definition of probability must be modified somehow when the total number of possible outcomes is infinite

  - What is the probability that an integer drawn at random from the positive integers be even? Start with the first 2*N* integers… Your answer would be N/2N = ½
  - Can you make this argument under all circumstances?

    - Natural ordering: 1,2,3,4,5,6,… → 1/2
    - Different ordering 1,3,   2; 5,7,   4; 9,11,   6;… (first pair of odd integers, first even, etc) → 1/3
    - Oscillating sequence of integers → never attains definite value

- Limitation 2: Suppose that we toss a coin known to be biased in favor of heads (it is bent so that a head is more likely to appear than a tail).

    - What is the probability of a head?
        - The classical definition leaves us completely helpless…

- Limitation 3: Suppose notions of symmetry and equally likely do not apply?
    - What is the probability that a female will die before the age of 60?
    - What is the probability that a cookie bought at a certain bakery will have less than 3 raisins in it?
    - What is the probability that my boy friend truly loves me?

How to deal with these limitations?

We assume that a series of observations (or experiments) can be made under quite uniform conditions:

- An observation of a random experiment is made
- Then the experiment is repeated under similar conditions, and another observation is taken
- This is repeated many times, and while conditions are similar each time, there is an uncontrollable variation which is haphazard or random so that the observations are individually unpredictable.
- In many cases the observations will fall into certain classes wherein the relative frequencies are quite stable. [Under stable or statistical regularity conditions, it is expected that this ratio will tend to a unique limit as the number of experiments becomes large.]

A posteriori definition of probability

Frequency probability: Assuming that a random experiment is performed a large number of times, say n, then for any event A let n_A be the number of occurrences of A in the n trials and define the ratio n_A/n as the relative frequency of A. The limiting value of the relative frequency is a probability measure of A.