Probability Part I: Fundamentals

Lecture 7 of *Core Mathematics* in the MPH and MSc in Epidemiology Courses, Imperial College London

Chris Wymant

November 2015

The Relevance

- You have an idea about a disease, e.g. that something might help or hinder its spread. You collect relevant data. Do these data support the idea? Let the 'p-value' p be the probability that these data[1] could have arisen if your idea is *not* true – the 'null hypothesis'. If p is large, these data do not support your idea. The smaller p is, all else being equal, the more convincing your idea is.
- Optimal allocation of public health resources may well involve esimating how likely different scenarios are.
- Inferring the best parameter values for a model usually involves calculating how likely it would be to obtain the known data as a function of those parameters ('the likelihood' function).
- ...

[1] these exact data *or* different data that would be even less compatible with the null hypothesis.

What is Probability?

Notation: $P(A)$ is

- the probability that A is true, or
- the probability that A will happen, or
- the probability of A,

depending on how exactly one phrases what it is that A represents.

A meaningless definition, appealing to intuition: probability is a measure of how likely (i.e. how probable) something is.

The *frequentist* definition of probability, where N is the total number of trials or times an experiment is repeated, is

$$P(A) = \text{the large } N \text{ limit of } \big(\text{the number of times } A \text{ occurs } / \ N\big)$$

$0 \leq P(A) \leq 1$ always: 0 means impossible, 1 means certain.

For probabilities, the percent symbol simply represents the number 0.01; e.g. $5\% = 5 \times 0.01 = 0.05$.

Independency, Mutual Exclusivity I

Two events being *independent* means they are not connected / do not affect each other. If A and B are possible outcomes from independent events, then $P(A \text{ and } B) = P(A)P(B)$.

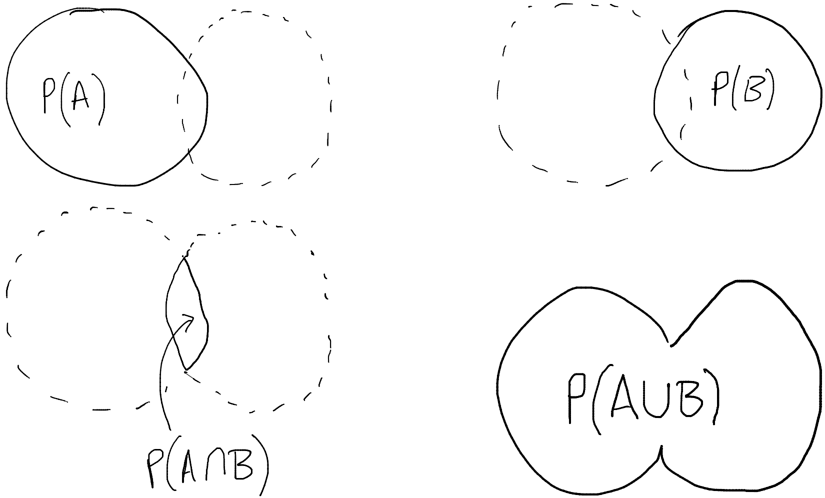
Two outcomes being *mutually exclusive* means that if one occurs, the other cannot: the probability of both is zero.
A and B are mutually exclusive $\iff P(A \text{ and } B) = 0$.

Avoiding double counting: $P(A)$ includes the probability $P(A \text{ and } B)$. $P(B)$ also includes the probability $P(A \text{ and } B)$.

$$\therefore \; P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B), \qquad \text{always.}$$

The subtracted term on the right-hand side vanishes if and only if A and B are mutually exclusive.

Venn Diagrams



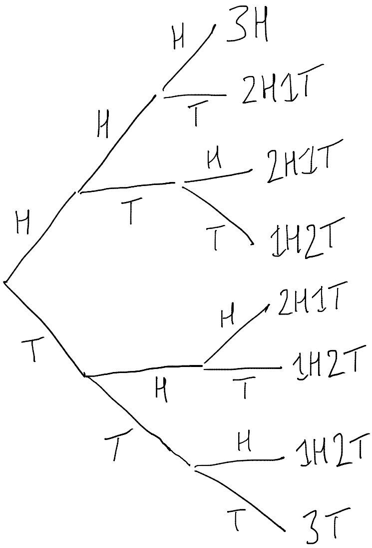
Illustrating why $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
($\cup$ means 'or'; $\cap$ means 'and')

Independency, Mutual Exclusivity II

Briefly discuss which of the following are independent, which are mutually exclusive, and which are neither.

1. Person A has flu. Person B has flu. (They live in the same house.)
2. Person A has flu. Person B has flu. (They live in different countries.)
3. Lab A gives a false positive for the blood test. Lab B gives a false positive for the blood test.
4. The patient's age is zero to eighteen. The patient's age is eighteen or older.
5. The patient is male. The patient is female.
6. Study A concludes something. Study B concludes the same thing. (They used the same patients.)
7. The outcome of my first simulation. The outcome of my second simulation. (The simulation is stochastic, and used the same input and *seed*.)

Illustrating All Possibilities: Trees



Toss a fair coin three times. Recalling that the probabilities of independent outcomes (subsequent tosses) are multiplied, and those of mutually exclusive outcomes (i.e. distinct possibilities) are added, the probability of getting

- 3 heads is $\frac{1}{8}$,
- 2 heads is $\frac{3}{8}$,
- 1 head is $\frac{3}{8}$,
- 0 heads is $\frac{1}{8}$.

Collective Exhaustivity

A set of outcomes are *collectively exhausive* if at least one of them one must occur. e.g. here are some possibilities when rolling a die once:

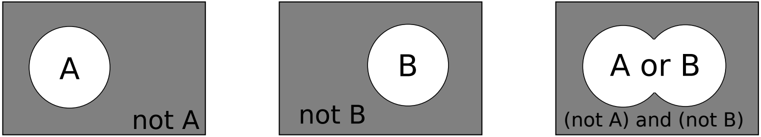
outcome A	outcome B	mutually exclusive	collectively exhaustive
1 or 2	2 or 3	no	no
1	2	yes	no
more than 3	less than 5	no	yes
odd	even	yes	yes

Probabilities Sum to One

Considering all possible distinct outcomes[2] and summing their probabilities gives 1.

$$
\begin{aligned}
P(A) &= 1 - P(\text{not } A) \\
P(A \text{ or } B) &= 1 - P(\text{`not } A\text{' and `not } B\text{'}) \\
P(A \text{ or } B \text{ or } C) &= 1 - P(\text{`not } A\text{' and `not } B\text{' and `not } C\text{'})
\end{aligned}
$$

The right-hand sides will often become easier to evaluate than explicit calculations of the left-hand sides as we consider more possibilities.

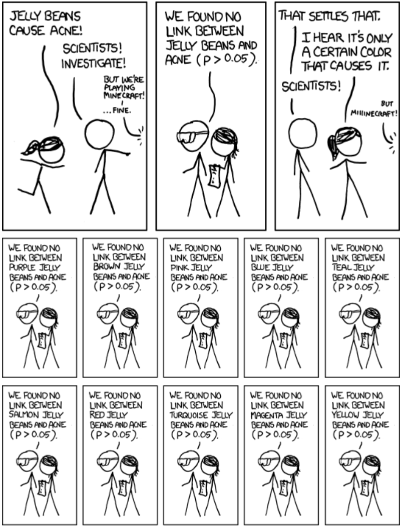


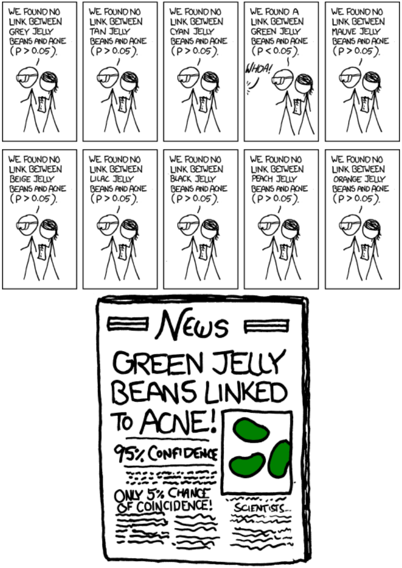
[2] i.e. finding a set of mutually exclusive, collectively exhaustive outcomes

Some Exercises

1. What is the probability that, rolling a die three times, I get at least one six? How about if I roll it N times?
2. When treating livestock / poultry / farmed fish with antibiotics, the probability of a resistant and readily transmissible strain arising in an individual animal is p. What is the probability of such a strain arising when N animals are treated?
3. What is the probability that, testing N ideas all of which are actually false (e.g. associations between disease and something that really has no effect, in unbiased data), I find at least one p-value less than 0.05?
4. The records for N patients were moved to a new database which, due to an error, does not allow February 29th as a valid birthday: the data of such patients (if any), would have been lost. What is the probability that data loss occurred?

In each case, what assumptions did you make to find your answer?





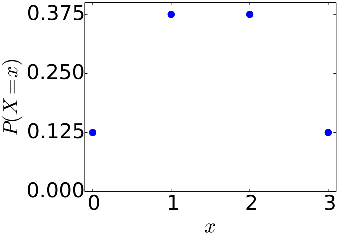
http://xkcd.com/882/

The Concept of a Discrete Probability Distribution

Wikipedia: a *random variable* is "a variable whose value is subject to variations due to chance". (Scientific measurements often are.)

A *discrete* random variable is one whose values are discrete, i.e. they can only be certain specific values, such as heads or tails, or whole numbers. We talk about the probability of a discrete random variable, X say, being equal to some particular value, x say. The function that quantifies this probability for all possible values of x is called the probability distribution. Using the previous example we could say X is the number of heads in three coin tosses:

$$P(X = x) = \begin{cases} \frac{1}{8}, & \text{if } x = 0, \\ \frac{3}{8}, & \text{if } x = 1, \\ \frac{3}{8}, & \text{if } x = 2, \\ \frac{1}{8}, & \text{if } x = 3, \end{cases}$$



Binomial Probability I

Imagine that in one trial/event/experiment, there are two possible outcomes: one with probability p (nominally 'success') and the other with probability $q = 1 - p$ (nominally 'failure'). How probable is it that in N independent repetitions of the trial, there are m successes? (And by implication, $N - m$ failures.)

One way this can happen is getting m successes in a row, each with probability p, then $N - m$ failures in a row, each with probability q. Multiplying these probabilities for independent events gives $p^m q^{N-m}$. The m successes could come anywhere in the N trials; each ordering is a distinct possibility, so we add their probabilities together. Each has the same probability ($p^m q^{N-m}$), so summing them is achieved by multiplying that probability by how many there are, i.e the number of ways one can choose m from N . . .

Binomial Probability II

… which is defined to be ${}^{N}C_m$. The answer is then

$${}^{N}C_m \, p^m q^{N-m}$$

The binomial coefficient is

$${}^{N}C_m = \binom{N}{m} = \frac{N!}{m!(N-m)!},$$

or just use your calculator / Excel / R etc.
For example, ${}^{4}C_2$ is the number of ways to choose two things from four: I can choose 1 and 2, or 1 and 3, or 1 and 4, or 2 and 3, or 2 and 4, or 3 and 4. Six ways to choose: ${}^{4}C_2 = 6$.
Special values: ${}^{N}C_0 = {}^{N}C_N = 1$, and ${}^{N}C_1 = {}^{N}C_{N-1} = N$. Why?

Exercise: the case fatality ratio of a deadly disease is roughly two-thirds. What is the probability that, of three patients, (a) exactly one dies? (b) at least one dies?

Poissonian Probability

If the expected number of events in a given time interval is λ (i.e. λ is the rate multiplied by the length of the time interval), and each event occurs independently of the time since the last event,

$$P(\text{there are exactly } k \text{ events}) = \frac{\lambda^k e^{-\lambda}}{k!},$$

the *Poisson distribution*[3]. Note that the variable for binomial probability was confined to a finite range: $k \in \{0, 1, 2, \ldots N\}$, whereas here the variable has no upper bound: $k \in \{0, 1, 2, \ldots \infty\}$. In both cases, however, summing the probabilities over all possible values of k gives 1.

Exercise: an infected person infects others at a rate $0.012\,\text{hour}^{-1}$. Assume an infectious period of one week. What is the probability that he/she infects (a) 2 or 3 people? (b) at least one person?

[3] The proof of this is the *Poisson limit theorem*.

The Offspring Distribution

The *offspring distribution* is the discrete probability distribution for the total number of people one person infects in the full course of his/her illness, which can take the values $0, 1, 2, \ldots \infty$
e.g. in more or less the simplest model one can have – a single state of infectiousness during which one infects others a constant rate β, with a constant hazard for removal from the state (by recovery or death) μ – the offspring distribution is

$$P(\text{infecting } n \text{ people}) = \frac{\alpha}{\alpha + \beta}\left(\frac{\beta}{\alpha + \beta}\right)^{n}$$

This is a *geometric distribution*, which means increasing *n* by 1 decreases the probability by a constant factor regardless of the value of *n*. Note that increasing both α and β by the same factor does not change this distribution.

Next week:

- Continuous probability (everything today was discrete!)
- Hazards and risks
- The mean, variance, and median
- Conditional probability
- An introduction to Bayes' Theorem

Extra Slides

Critically Appraising Statements of Probability

Discuss the meaning of:

- ‘quite possible’,
- ‘very certain’,
- ‘totally impossible’.

Explain the difference between the UK’s terror threat levels one and two:

- low - an attack is unlikely
- moderate - an attack is possible but not likely

(Source: `https://www.gov.uk/terrorism-national-emergency/terrorism-threat-levels`)

Binomial Probability Again

Recall that p is the probability of one outcome, and $q = 1 - p$ is the outcome of the other.

$$\begin{aligned} 1 &= P(\text{one outcome or the other in event 1}) \times \\ &\qquad P(\text{one outcome or the other in event 2}) \ldots \\ &= \underbrace{(p+q)}_{\text{event 1}} \underbrace{(p+q)}_{\text{event 2}} \ldots \underbrace{(p+q)}_{\text{event N}} \\ &= (p+q)^N \\ &= \sum_{k=0}^{N} \underbrace{{}^N C_k \, p^k q^{N-k}}_{\text{this is } P(\text{the one outcome occurs } k \text{ times})} \end{aligned}$$

The final step is called *the Binomial Theorem*: when we multiply together all the brackets (i.e. expand $(p+q)^N$), the number of times we get the term $p^k q^{N-k}$ is ${}^N C_k \, p^k q^{N-k}$.
The final line implies that the probability of the one outcome occuring 0, or 1, or… or N times is 1. That's reassuring.