Friday

Applied Statistics

Lectures 2 & 3

David Barton & Sabine Hauert

Department of Engineering Mathematics

Outline

Last lecture

- Why statistics?
- Introduction to design of experiments
- Parametric versus non-parametric models

This lecture

- Basics of probability
- Statistical inference
- Introduction to hypothesis testing

OpenIntro Statistics

Chapters 2–4, particularly §4.3

Probability

Probability (Wikipedia)

Probability describes the plausibility of (random) observed data — assumed to be described by a statistical model, and specified parameter value(s) — without reference to any observed data

forecasting

Different (practical) interpretations of probability

- Classical — often connected with gambling, dice, cards, etc
- Frequentist — use a large number of trials to estimate probabilities
- Bayesian — "reasonable expectation" or "belief" (subjective)

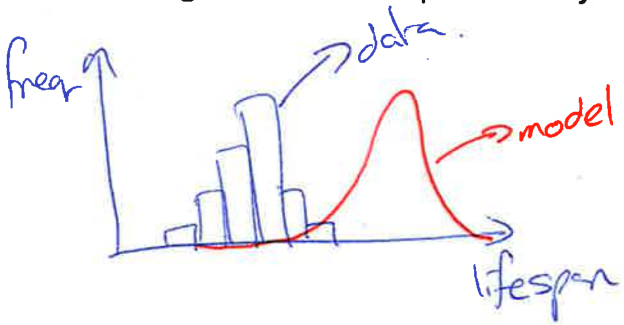
No interpretation fits all physical scenarios!

Likelihood

Likelihood (Wikipedia)

Likelihood describes the plausibility, given specific observed data, of a parameter value of the statistical model which is assumed to describe that data

- Statistics mostly deals with likelihoods
- Informally probabilities and likelihoods are used interchangeably
- Maximum likelihood for finding parameters of statistical models
- Histograms versus probability density functions



Statistical inference

Definition from Wikipedia

Statistical inference makes propositions about a population, using data drawn from the population with some form of sampling. Given a hypothesis about a population, for which we wish to draw inferences, statistical inference consists of (first) selecting a statistical model of the process that generates the data and (second) deducing propositions from the model.

Interview selection

A leading FTSE 100 company claims to have a policy of equal opportunity for both men and women on its graduate training scheme. However, upon auditing it was found that during their last recruitment round they appointed 19 women and only 1 man despite roughly equal numbers of men and women applying.

Is the selection process fair to both men and women?

↳ where is the dividing line?
↳ 15 women – 5 men?

(What about the design of the experiment?!)

↳ Bad!

Construct the hypothesis

First decide on a *default position* (e.g., innocent until proven guilty) — this is the null hypothesis

↳ what do we think is true in the absense of any information? (Bayesian: prior belief)

Null hypothesis H_0 the interview selection process is fair to both men and women.

↳ Forms the basis of a statistical model.

Alternative hypothesis H_1 the interview selection process is biased.

↳ opposite H_0

Interview selection — a statistical model

What is the individual/population/sample?

individual: outcome of a recruitment round.
population: all recruitment rounds
~~sing~~ sample: single individual in my sample.

What does *fair* actually mean?

Equally probable to select a man or a woman.

What is a statistical model that could generate data for an individual observation?

X: number of women recruited

$X \sim \text{Bin}(20, 0.5)$

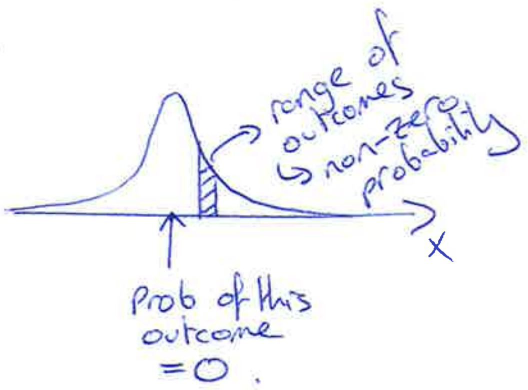
trials   prob. of success

generates probabilities for each outcome.

Visualising a statistical model

# Women	Probability
0	9.5×10^{-7}
1	1.9×10^{-5}
2	1.8×10^{-4}
3	1.1×10^{-3}
4	4.6×10^{-3}
5	0.015
6	0.037
7	0.074
8	0.12
9	0.16
10	0.18 ← most probable outcome

# Women	Probability
11	0.16
12	0.12
13	0.074
14	0.037
15	0.015
16	4.6×10^{-3}
17	1.1×10^{-3}
18	1.8×10^{-4}
19	1.9×10^{-5}
20	9.5×10^{-7}

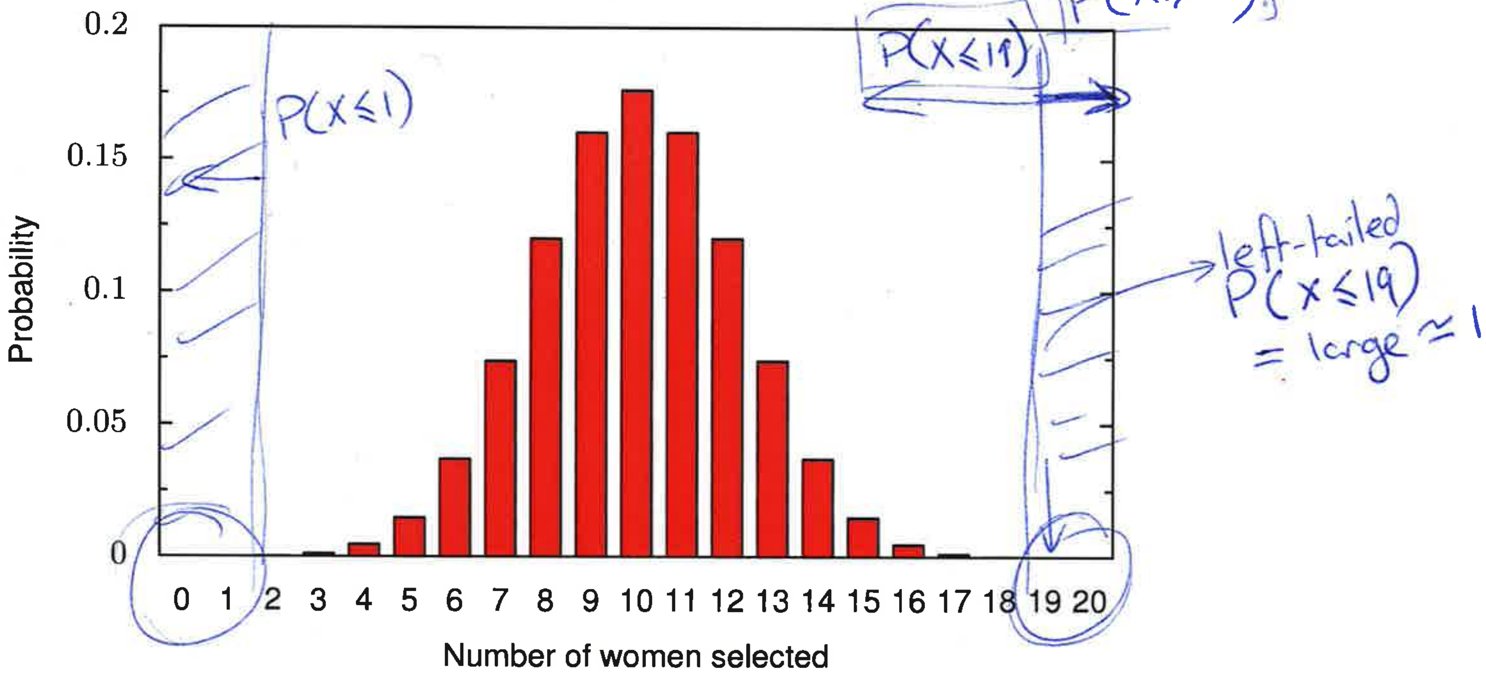


Consider probability of the outcome or something even more extreme → $1.9 \times 10^{-5} + 9.5 \times 10^{-7}$ = small!

Visualising a statistical model

small.
$P(X \geq 19)$

large
$P(X \leq 19)$



Statistical significance

Testing the hypothesis involves calculating the probability of the sample occurring *assuming that the null hypothesis is correct*

When is the measurement (or a measurement even more extreme) so improbable that we say the null hypothesis is incorrect? *This is the (statistical) significance level.*

↳ 5% is very common

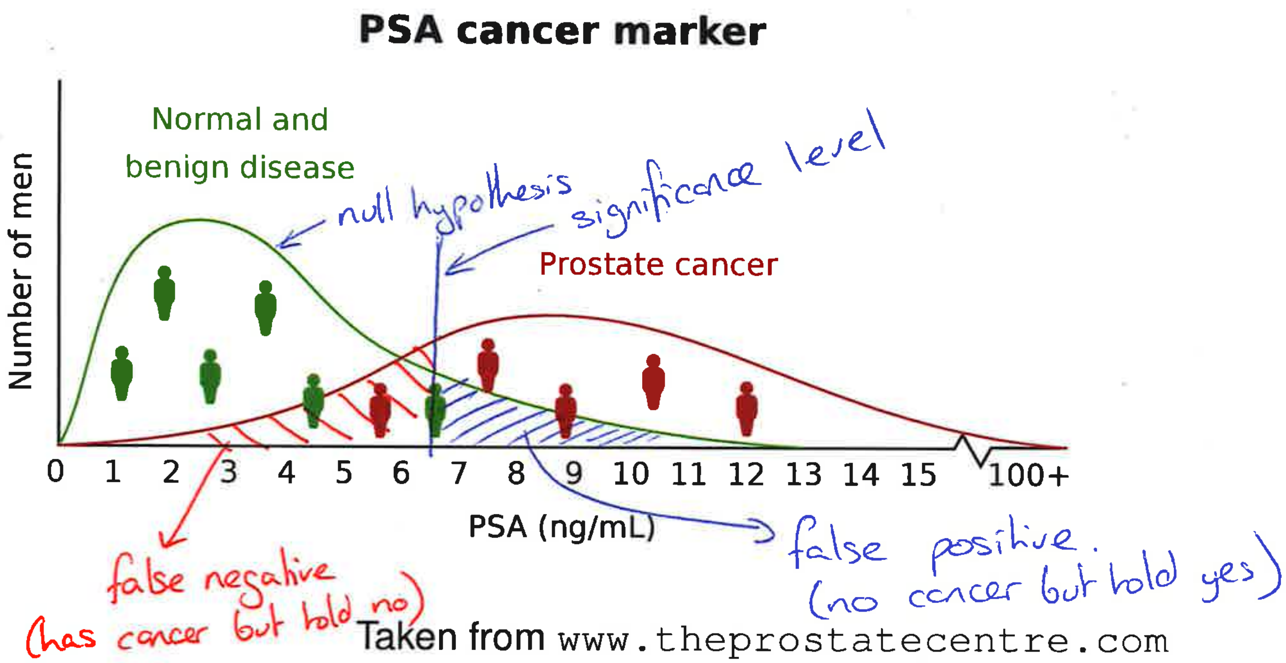
Criminal trials in the UK?

↳ Innocent until proven guilty.

Scientific discoveries (e.g., Higgs Boson)?

↳ Physics : 5σ $\sim 10^{-7}$

Where to draw the line?



Taken from www.theprostatecentre.com

Errors in hypothesis testing

Since hypothesis testing is a probabilistic technique, invariably there can be errors. There are two main types of error to consider.

Type 1 (false positive) These errors occur when you incorrectly reject the null hypothesis, i.e., you think a change has occurred when it hasn't. Since the null hypothesis was actually correct we know that these types of error occur at the same level as the significance level chosen (e.g., 5%).

Type 2 (false negative) These errors occur when you incorrectly accept the null hypothesis, i.e., you think a change has not occurred when it has. Since the null hypothesis was incorrect, we don't know the distribution of errors or the percentage of errors.

Decreasing type 1 errors will almost certainly increase type 2 errors!

Exercises

For each of the following, decide what the null hypothesis H_0 and the alternative hypothesis H_1 should be.

What would be a type 1 error and a type 2 error in each case?

1. Drinking coffee will make us live longer

2. The medical treatment improved the patient's condition

3. Product A is more efficient than product B

Testing the hypothesis — p-value

↳ Probability of seeing the data or something more extreme.

To test the hypothesis, calculate the *p-value*

- $p = 2\min(P(X \geqslant x), P(X \leqslant x))$ — two-tailed (two-sided) test → unfair by biasing → women or by biasing → men.
- → $p = P(X \leqslant x)$ — left-tailed (one-sided) test → don't care about being biased towards women but do care about being biased towards men.
- $p = P(X \geqslant x)$ — right-tailed (one-sided) test → don't care about being biased towards men but do care about being biased towards women.

By default use a two-tailed test unless you're sure that you only care about differences in one direction.

Interview selection — testing the hypothesis

Use a statistical significance level of 5% (i.e. 0.05)

For the interview selection we have

$$X \sim \mathrm{Bin}(n, p), \quad P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

with $n = 20$, $p = 0.5$. Binomial coefficient is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Need to calculate

$$p = \underline{2\min(P(X \leqslant \overset{19}{1}), P(X \geqslant \overset{19}{1}))} = 4.0 \times 10^{-5} \quad < 5\%$$

$\therefore$ Reject null hypothesis.
Accept the alternative hypothesis.

Interview selection — exercise

A morally dubious colleague is interviewing and they want to know how biased their selection can be (women:men) without coming under suspicion. They are told to select 5 people in total.

Statistical assumptions — independence

The statistical model for interview selection like many statistical models assumes independence (where?)

What if 16 of the 20 people selected in the last recruitment round had graduated from Mechanical Engineering at Bristol?

Has anything changed?

Quote of the day

Mark Twain

It ain't what you don't know that gets you into trouble.
It's what you know for sure that just ain't so.

Exercises

- 4.17–4.20 from OpenIntro Statistics