Likelihood and Probability in Scientific Inference

UCL Graduate School: Graduate Skills Course

Your hosts for today:

James Mallet, Professor of Biological Diversity
Ziheng Yang, Professor of Statistical Genetics
http://abacus.gene.ucl.ac.uk/
http://abacus.gene.ucl.ac.uk/jim/
(Department of Biology, UCL)

What we will cover:

- The basis of inference in science.
- Compare the "frequentist" approach we usually learn, with today's widely used alternatives in statistical inference.
  - Particularly the use of "likelihood" and Bayesian probability.
- We will hopefully empower to develop your own analyses, using simple examples.

What we will not cover:

Not suitable for people already well-versed in statistics. They'll already know most of this!

Not suitable for people who've no idea about statistics. At least GCSE knowledge required.

We won't have time to teach you all you need to know to analyse your data.

We won't have time to go into very complicated examples.

Instead, we hope

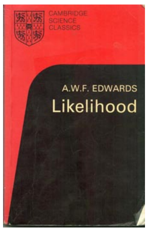
You begin to develop a healthy disrespect for most "off-the-shelf" methods. (But you will probably still use them).

You start to form your own ideas of how statistics and scientific inference are related (a philosophy of science topic).

That your interest in likelihood and Bayesian analysis is piqued, and you might be motivated to do further reading.

You become empowered to perform simple statistical analyses, using Excel and Excel's Solver "add-in". + a little programming ⇒ you can analyse much more difficult problems.

My main source



Anthony W. Edwards (1972); reprinted 1992: *Likelihood*. Cambridge UP

see also more in-depth:
Yudi Pawitan (2001). *In all Likelihood. Statistical Modelling and Inference using Likelihood*. Oxford UP

Overview

- What is scientific inference?
- Three philosophies of statistical inference:
  - Frequentist (probability in the long run)
  - Likelihood (likelihood)
  - Bayesian (posterior probability)
- Common Ground: Opposing philosophies agree (approximately) on many problems

- Discussion
- Exercises, example of ABO bloodgroups
- Ziheng's talk: when philosophies conflict ...

Scientific Inference

- What is scientific inference?
- Three philosophies of statistical inference:
  - Frequentist (probability in the long run)
  - Likelihood (likelihood measures strength)
  - Bayesian (posterior probability " ")
- Common ground: Opposing philosophies agree (approximately), in many problems.

The nature of scientific inference

"I'm sure this is true"
"I'm pretty sure" "I'm not sure"
"It is likely that..."
"This seems most probable to me"

All of inference about the world is likely to be based on probability; it's statistical.

(Except divine revelation!)

Models and hypotheses

Science is about trying to find "predictability" or "regularities" in nature, which we can use.

For some reason, this usually seems to work ...

Models and *hypotheses* allow prediction. We test them by analysing something about their "likelihood" or "probability"

Models and hypotheses in statistical inference

Models are assumed to be true for the purposes of the particular test or problem
e.g. we assume height in humans to be normally distributed.

Hypotheses are "parameters" that are the focus of interest in estimation
e.g. mean and variance of height humans.

Data is typically discrete

... Counts of things
... Measurements to nearest mm, 0.1ºC
Data is also finite

Models, hypotheses can be discrete too, *or continuous. Models and hypotheses may be finite, or infinite in scope.*

A good method of inference should take this discreteness of data into account when we analyse the data. Many analyses, particularly frequentist, don't!

For example, milk fat

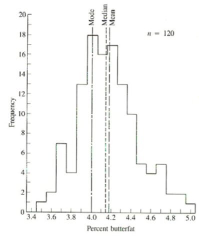


Figure 4.1 An asymmetrical frequency distribution (skewed to the right) showing location of the mean, median, and mode. Percent butterfat in 120 samples of milk (from a Canadian cattle breeders' record book).

From Sokal & Rohlf 1981, Biometry, p. 47

Null hypotheses in statistics

We are often taught in biology a simplistic kind of "Popperian" approach to science, to falsify simple hypotheses. We then try to test the null hypothesis!

(Zero-dimensional statistics, if you like; only one hypothesis can be excluded).

In this view, estimation (e.g. mean, variance) is like natural history, not good science.

Physics-envy?

Estimation is primary

Edwards argues that we should turn this argument on its head.

Estimation of a distribution or model can lead to testing of an infinitude of hypotheses, *including* the null hypothesis.

Uses full dimensionality of the problem: ≥1 – n-dimensional statistical analyses.

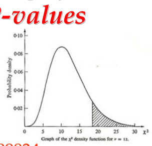
More powerful!

The three philosophies

- What is scientific inference?
- Three philosophies of statistical inference:
  - Frequentist (probability in the long run)
  - Likelihood (likelihood)
  - Bayesian (posterior probability)
- Common ground: Opposing philosophies agree (approximately), in many problems.

1. Frequentist, significance testing, P-values

Perfected in 1920s
(Pearson, Fisher et al.)
e.g. χ^2 test, or *t*-test
χ^2= 5.28, d.f. = 1;
or t=3.92, d.f.=10
We find P<0.05, or P=0.009834



This is "tail probability" or "probability in the long run" of getting results at least as extreme as the data under the null hypothesis

Philosophical problems with frequentist approach

We only have one set of data; seems to imagine the experiment done a very large number of times

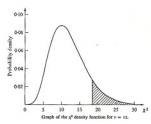
Often tend to assume the data come from a continuous distribution;
e.g. χ^2 tests on count data, $\Sigma(O-E)^2/E$

Encourages testing of null hypothesis

P - values

P-values are "tail probabilities"

"What the use of *P* implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred" Jeffreys 1961



Alternatives to frequentism

- Frequentism: "Probability in the long run"
- Two alternative measures of support:
  - Bayesian Probability (Thomas Bayes 1763, Marquis de Laplace 1820)
    "The probability of a hypothesis given the data"
  - Likelihood (RA Fisher 1920s, Edwards 1972)
    "The probability of the data given a hypothesis"
    (can be viewed as a simplified form of Bayesian probability)

2. Likelihood

The *likelihood* of a hypothesis (H) after doing an experiment or gathering data (D) is the *probability of the data given the hypothesis*

$$L(H \mid D) = P(D \mid H)$$

Probabilities add to 1 for each hypothesis (by definition), but do not add to 1 across *different* hypotheses – hence "Likelihood"

The Law of Likelihood

"Within the framework of a statistical model, a particular set of data supports one statistical hypothesis better than another if the likelihood of the first hypothesis on the data exceeds the likelihood of the second hypothesis"

$$Likelihood\ Ratio = \frac{P(D \mid H_1)}{P(D \mid H_2)}$$

Support

Support is defined as the natural logarithm of the likelihood ratio

$$Support = \log_e \frac{P(D \mid H_1)}{P(D \mid H_2)}$$

$$= \log_e P(D \mid H_1) - \log_e P(D \mid H_2)$$

Example: binomial distribution

Supposing we are interested in estimating the allele frequency of a gene in a sample:

A	a	Total alleles
2	8	10
i	*(n-i)*	*n*

This is a problem that is well suited to the binomial theorem:

$$P(D|H_j) = \binom{n}{i} p^i (1-p)^{(n-i)} = \frac{n!}{i!(n-i)!} p^i (1-p)^{(n-i)}$$

A common frequentist approach:

Sample mean $p^* = 2/10 = 0.2$

Sample variance, $s_p^2 = p^* q^* / n = 0.2 \times 0.8/10 = 0.016$

Standard deviation of mean, $s_p = \sqrt{0.016} = 0.126$

95% conf. limits of mean = $p^* \pm t_{9,0.05} s_p$

= 0.2 ± 2.262 x 0.126

= (-0.085, +0.485)

Note the NEGATIVE lower limit!

Likelihood approach

To get the support for two hypotheses, we need to calculate:

$$Support = \log_e \frac{P(D \mid H_1)}{P(D \mid H_2)}$$

Note! The binomial coefficient depends only on the data (D), not on the hypothesis (H)

$$P(D|H_j) = \binom{n}{i} p^i (1-p)^{(n-i)} = \frac{n!}{i!(n-i)!} p^i (1-p)^{(n-i)}$$

Binomial coeff. cancels! No need to calculate the tedious constant! Just need the $p^i(1-p)^{(n-i)}$ terms

Likelihood & the binomial

Binomial probability			sample size		"successes"		
using likelihood			n=	10	i=	2	
	Likelihood/B		ln likelihood		ln likelihood ratio		
Hj = p	p^i(1-p)^(n-i)						
0	0		#NUM!	(impossible)	#NUM!	(->minus infinity)	
0.001	1.002E-06		-13.81351		-8.36546		
0.01	9.22745E-05		-9.290743		-4.19635		
0.05	0.001658551		-6.401811		-1.39779		
0.1	0.004304672		-5.448054		-0.44403		
0.15	0.006131037		-5.094391		-0.09037		
0.2	0.006710886		-5.004024	*=max (i=2)!	0		
0.25	0.006257057		-5.074045		-0.07002		
0.3	0.005188321		-5.261345		-0.25732		
0.35	0.003903399		-5.545908		-0.54188		
0.4	0.002687386		-5.919186		-0.91516		
0.45	0.001695612		-6.379711		-1.37569		
0.5	0.000976563		-6.931472		-1.92745		
0.55	0.000508658		-7.583736		-2.57971		
0.6	0.00023593		-8.351977		-3.34795		
0.65	9.51417E-05		-9.260143		-4.25612		
0.7	3.21489E-05		-10.34513		-5.34111		

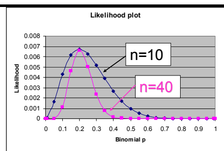
Likelihood & the binomial

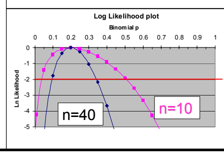
The support curve gives a measure of belief in the continuously variable hypotheses

Edwards: 2 units below the can be viewed as "support limits"

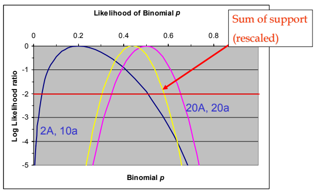
(equivalent to approx 2 standard deviations in the frequentist approach)

$\log_e LR=2$ implies $LR=e^2$, the best is 7.4x as good





Sum of support from different experiments



Support provides a way to adjudicate between data from different experiments

3. Bayes' Theorem

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

Named after its inventor, Thomas Bayes in 18th Century England. Led by Bayes and Laplace, the theorem and "Bayesian Probability" has come to be used in a system of inference …

$$P(H \mid D) = k.P(D \mid H)P(H)$$

Posterior Probability — Likelihood — Prior Probability

Bayes' Theorem as a means of inference

$$\frac{P(H_1 \mid D)}{P(H_2 \mid D)} = \frac{k.P(D \mid H_1)P(H_1)}{k.P(D \mid H_2)P(H_2)}$$

If the prior is "uniform", $P(H_1)=P(H_2)$

$$\frac{P(H_1 \mid D)}{P(H_2 \mid D)} = \frac{P(D \mid H_1)}{P(D \mid H_2)}$$

The ratio of posterior probabilities collapses to … a likelihood ratio!

Common ground

- What is scientific inference?
- Three philosophies of statistical inference:
  - Frequentist (probability in the long run)
  - Likelihood (likelihood)
  - Bayesian (posterior probability)
- Common ground: Opposing philosophies agree (approximately), in many problems.

Opposing philosophies

Important to realize there isn't just one way of doing statistics. *For me:*

Edwards' argument for likelihood as *the* means of inference seems powerful. Probability of the data given the hypothesis is a good measure.

Bayesian difficulties: "probability of a hypothesis" without data (the prior probability)

Frequentist difficulties: *P*-values: probability based on events that haven't happened

In practice

In practice, in most applications, all three approaches tend to support similar hypotheses.

Edwards shows that significance tests are justifiable by appealing to likelihood ratios – tail probability low when likelihood ratio (itself often proportional to relative Bayesian probability) is high.

In very complex estimation problems (e.g. GLM), where we test for "significance" of ν extra parameters, we use the chi-square approximation:

$2\log_e LR$="deviance"$\approx \chi^2_\nu$

This interpretation employs a frequentist approach.

Conclusion
Utility of likelihood

Estimation and hypothesis testing of complex problems today almost always use likelihood or Bayesian methods, often using MCMC optimization, for example:

Generalized Linear Models, Deviance

Phylogeny estimation, molecular clock estimation

Linkage mapping, QTL analysis in human genetics

High energy physics experiments

At the very least, these methods *enable more complex problems to be analysed*. At best, they may provide an *improved philosophical basis for inference.*

Excel exercise with "Solver"

go to www.ucl.ac.uk/~ucbhdjm/bin/

open the ABO_Student.xls file

follow instructions

Relationship of likelihood ratio to frequentism

In large samples,

$$G = 2\{\log_e P(D \mid H_1) - \log_e P(D \mid H_2)\}$$

converges to a χ^2 distribution with the numbers of degrees of freedom given by the numbers of free parameters.

For a test of null hypothesis H_0 vs. max. likelihood hypothesis H_1: *P* can be calculated from the integral of the χ^2 probability density function.

Also, note that with a support value ($\Delta \ln L$) of 2.0, $G = 4.0 \approx 1.96^2 = 3.84$, i.e. the value of χ^2 which is "significant" at P=0.05 with 1 degree of freedom.