Applied Bayesian Statistics

David Spiegelhalter
Statistical Laboratory, Cambridge

with thanks to:
Nicky Best
Dave Lunn
Andrew Thomas

Course: Lent Term 2010

©*University of Cambridge 2010*

Course Summary (provisional)

Jan 18	Lect 1:	Probability and Bayes theorem for discrete observables
Jan 20	Lect 2:	Bayesian inference: conjugate analysis
Jan 25	Lect 3:	Univariate conjugate analysis: predictions and priors
Jan 27	Prac 1:	Exact Bayesian analysis using First Bayes
Feb 1	Lect 4:	Graphical models and Monte Carlo methods - WinBUGS
Feb 3	Prac 2:	Monte Carlo analysis using WinBUGS
Feb 8	Lect 5:	MCMC methods: univariate
Feb 10	Prac 3:	MCMC analysis using WinBUGS
Feb 15	Lect 6:	Prior distributions: univariate
Feb 17	Lect 7:	Multivariate distributions
Feb 22	Lect 8:	Regression models
Feb 24	Prac 4:	Modelling in WinBUGS
Mar 1	Lect 9:	Prediction, ranking and relation to classical methods
Mar 4	Lect 10:	Model criticism and comparison [NOTE DATE CHANGE]
Mar 8	Prac 5:	More MCMC
Mar 10	Lect 11:	Hierarchical models

+ 3 exercise sheets

Learning Objectives

- Understanding of principles underlying basic Bayesian modelling (although fairly informal mathematically)
- Understanding how to use Bayesian analysis for making inference about real-world problems
- Insight into computational techniques used for Bayesian analysis
- Appreciation of the need for sensitivity analysis, model checking and comparison, and the potential dangers of Bayesian methods

Material mainly taken from forthcoming book

Bayesian Analysis using the BUGS language: a Practical Introduction
D Spiegelhalter, N Best, D Lunn, A Thomas
Chapman and Hall, 2010

Other good sources:

- *Bayesian Data Analysis; 2nd Ed* by Andrew Gelman, John Carlin, Hal Stern, and Don Rubin. Chapman and Hall (2004)
- *Bayesian Methods for Data Analysis* by Brad Carlin and Tom Louis, Chapman and Hall (2008)
- *Bayesian Modelling using WinBUGS* by Ioannis Ntzoufras, Wiley (2009)

Background reading in *Introduction to Bayesian Methods in Healthcare Evaluation:* Spiegelhalter, Abrams, Myles. Wiley (2004)

Plus numerous websites.

Past papers for 2008 and 2009 available on -

`http://www.maths.cam.ac.uk/teaching/pastpapers/2009/Part_3/index.html`

Lecture 1.
Probability and Bayes theorem for discrete observables

Summary

1. Basic probability for discrete random variables
2. But what do we mean by 'probability'?
3. Bayes theorem for discrete variables
4. Diagnostic testing
5. The three coins
6. Use of likelihood ratios
7. Applications in forensic science and legal reasoning

Reference:

Probability and Proof - downloadable from Prof Dawid's website
`www.statslab.cam.ac.uk/~apd/`

Probability (informal)

1. **Bounds:**
$$0 \leq p(a|H) \leq 1,$$
where a is an event, $p(a|H) = 0$ if a is impossible and $p(a|H) = 1$ if a is certain in the context H.

2. **Addition rule:** If a and b are mutually exclusive (*i.e.* one at most can occur)
$$p(a \bigcup b|H) = p(a|H) + p(b|H).$$

3. **Multiplication rule:** For any events a and b,
$$p(a \bigcap b|H) = p(a|b, H)p(b|H).$$
We say that a and b are independent if $p(a \text{ and } b|H) = p(a|H)p(b|H)$ or equivalently $p(a|b, H) = p(a|H)$

But what do we mean by 'probability'? (Dawid)

1. *Statistical Probability*: proportions of a sampled population
2. *Classical Probability*: symmetry of mechanism (eg dice)
3. *Empirical Probability*: long-run proportion of occurrence - *frequentist*
4. *Metaphysical Probability*: proportion of possible future worlds
5. *Subjective Probability*: willingness-to-bet / degree-of-belief
6. *Logical Probability*: degree of implication

General agreement on mathematics, but not on interpretation.

In this course we shall take the *subjective* view.

Subjectivity and context

- All probabilities are *conditional on context* H
- They are Your probabilities *for* an event, not a property *of* the event
- Probabilities are therefore *subjective* and can be given for unique events, *e.g.* the probability of *aliens openly visiting earth in the next 10 years*
- They express Your relationship to the event - different stakeholders will have different information and different probabilities
- Vital that probabilities obey the Rules! *i.e.* they *cohere*
- Can derive 'axioms' by principles of rational behaviour: such as avoiding certain losses when betting
- Other frameworks exist for deriving probability 'axioms' from more basic principles

Chance or ignorance?

We can think of two broad types (at least) of uncertainty

1. *Aleatory*: essentially unpredictable, chance
2. *Epistemic*: due to lack of knowledge

So what type of uncertainty is involved in deciding -

1. Which numbers to choose for the lottery?
2. Which lottery scratch-card to buy?
3. Is Bin Laden alive?
4. Which card will be on top when I shuffle?
5. What card will I get dealt next in poker?
6. Am *I* (DJS) going to survive another ten years?

From the subjectivist viewpoint, no need to worry about this distinction, it's all just uncertainty

Bayes theorem

Since $p(b \bigcap a) = p(a \bigcap b)$, Rule 3 implies that $p(b|a)p(a) = p(a|b)p(b)$, or equivalently

$$p(b|a) = \frac{p(a|b)}{p(a)} \times p(b).$$

This is Bayes theorem

An initial probability $p(b)$ is changed into a conditional probability $p(b|a)$ when taking into account the event a occurring

Provides a formal mechanism for learning from experience

Odds form for Bayes theorem

Odds = $P/(1-P)$

e.g. P=0.8, odds = 0.8/0.2 = 4

Let $\bar{b}$ = 'not b', so that $p(\bar{b}) = 1 - p(b)$

Since

$$p(\bar{b}|a) = p(a|\bar{b}) \times p(\bar{b})/p(a)$$

we get

$$\frac{p(b|a)}{p(\bar{b}|a)} = \frac{p(a|b)}{p(a|\bar{b})} \times \frac{p(b)}{p(\bar{b})}$$

Bayes theorem for two hypotheses For a null hypothesis H_0, H_1 = 'not H_0'

$$\frac{p(H_0|y)}{p(H_1|y)} = \frac{p(y|H_0)}{p(y|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

posterior odds = likelihood ratio $\times$ *prior odds.*

By taking logarithms we also note that

log(posterior odds) = log(likelihood ratio) $+$ *log(prior odds).*

where the likelihood ratio is also known as the 'Bayes factor', and the log(likelihood ratio) has also been termed the 'weight of evidence'

(Used by Alan Turing when using these techniques for breaking the Enigma codes at Bletchley Park)

Diagnosis: Bayes theorem in diagnostic testing A new home HIV test is claimed to have "95% sensitivity and 98% specificity",

To be used in a population with an HIV prevalence of 1/1000

Expected status of 100000 tested individuals in population:

	HIV -	HIV +	
Test -	97902	5	97907
Test +	1998	95	2093
	99900	100	100000

Thus of 2093 who test positive (*i.e.* have observation y), only 95 are truly HIV positive, giving a 'predictive value positive' of only 95/2093 = 4.5%.

Using Bayes theorem Let H_0 be the hypothesis that the individual is truly HIV positive, and y be the observation that they test positive.

"95% sensitivity" means $p(y|H_0) = 0.95$. and "98% specificity" means $p(y|H_1) = 0.02$

Prior odds $p(H_0)/p(H_1)$ is 1/999

Likelihood ratio $p(y|H_0)/p(y|H_1)$ is $0.95/0.02 = 95/2 = 47.5$

The posterior odds is $(95/2) \times 1/999 = 95/1998$

Odds correspond to a posterior probability $p(H_0|y) = 95/(95 + 1998) = 0.045$, as found directly from the table.

The crucial finding is that over 95% of those testing positive will, in fact, not have HIV.

Likelihood ratios and Bayes factors

Bayes factor range	Strength of evidence in favour of H_0 and against H_1
> 100	Decisive
32 to 100	Very strong
10 to 32	Strong
3.2 to 10	Substantial
1 to 3.2	'Not worth more than a bare mention'

	Strength of evidence against H_0 and in favour of H_1
1 to 1/3.2	'Not worth more than a bare mention'
1/3.2 to 1/10	Substantial
1/10 to 1/32	Strong
1/32 to 1/100	Very strong
$< 1/100$	Decisive

Calibration of Bayes factor (likelihood ratio) provided by Harold Jeffreys (1939, 1961)

The three coins

I have 3 coins:

1. One with two heads
2. One with two tails
3. One with a head and a tail

I pick a coin at random and flip it.

It comes up heads. What is the chance that the other side is also a head?

Homework: look at all the sites on the Monty Hall problem

Bayes' theorem for multiple hypotheses

If $\{a_k\}$ is a finite set of mutually exclusive and exhaustive events:

$p(\bigcup_k a_k) = \sum_k p(a_k) = 1$

then

$$p(a_j|b) = \frac{p(b|a_j)p(a_j)}{\sum_k p(b|a_k)p(a_k)}.$$

Bayesian reasoning in computer-aided diagnosis

1. Probabilities of diseases change as symptoms taken into account
2. Standard assumption is that symptoms are independent within each disease class ('Naive' or 'idiot's Bayes')
3. Makes sequential updating using Bayes theorem straightforward
4. Used in artificial intelligence / expert systems
5. Also basic model for Spam filters etc- see 'Naive Bayes classifier' and 'Bayesian spam filtering' on Wikipedia for background
6. Extends to general 'graphical models' (see later)

Conditional independence

If D be the unknown 'disease', comprising a set of mutually exclusive and exhaustive $d_k, k = 1, .., K$, each with prior probability $p(d_k)$

$S_1, S_2, .., S_J$ is a set of J questions that are going to be asked the patient

If the responses are $\underline{s} = s_1, s_2, ..., s_J$, then the posterior distribution over the possible diseases is

$$p(d_i|\underline{s}) = \frac{p(\underline{s}|d_i)p(d_i)}{\sum_k p(\underline{s}|d_k)p(d_k)}.$$

If we assume the 'symptoms' are conditionally independent given each possible disease, then $p(\underline{s}|d_i) = \prod_j p(s_j|d_i)$, and so

$$p(d_i|\underline{s}) = \frac{\prod_j p(s_j|d_i)p(d_i)}{\sum_k \prod_j p(s_j|d_i)p(d_k)}.$$

Only need specify conditional probabilities $p(s_j|d_i)$ and prior probabilities $p(d_i)$

'Sequential updating': posterior following s_1 becomes prior for s_2 etc

log-Bayes factor between two diseases d_1 and d_2 is

$$\log \frac{p(d_1|\underline{s})}{p(d_2|\underline{s})} = \sum_j \log \frac{p(s_j|d_1)}{p(s_j|d_2)} + \log \frac{p(d_1)}{p(d_2)}.$$

Bayesian evidential reasoning

- Increasing attention to using Bayesian analysis in forensic science and general evidential reasoning
- Dawid is international expert
- Many fascinating case studies: OJ Simpson, Sally Clark, Turin Shroud, Jesus' tomb etc
- Controversial in a formal legal context
- Can be very difficult, easy to make mistakes!
- Important ideas: *prosecutor's fallacy* and *defence fallacy*

Regina vs Denis John Adams

- Rape in Hemel Hempstead in 1991, DNA from semen stored on database
- Adams arrested in 1993 for another offence, his DNA matched sample
- Adams tried in 1995, claimed misidentification, DNA only direct evidence
- Analysis below is purely illustrative (Dawid, 2005)

Suppose DNA were only evidence

- Let g be 'guilty', $\overline{g}$ = 'not guilty'
- Let m be the evidence of the DNA match
- Prosecution's forensic evidence said 1 in 200 million chance of a chance DNA match, defence said lower;
- we shall take $p(m|\overline{g}) = 1/2,000,000$, $p(m|g) = 1$
- *Prosecutor's Fallacy*: therefore there is 1/2,000,000 chance he is not guilty!
- Confuses $p(m|\overline{g})$ with $p(\overline{g}|m)$
- When would this be appropriate?

Defence's Fallacy

- there are 200,000 possible attackers, therefore the prior probability $p(g) = 1/200,000$
- the posterior odds on guilt

$$\frac{p(g|m)}{p(\overline{g}|m)} = \frac{p(m|g)}{p(m|\overline{g})} \times \frac{p(g)}{p(\overline{g})} = 2,000,000 \times \frac{1}{200,000} = 10$$

- so only 91% probability he is guilty - not 'beyond reasonable doubt'
- Can 'explain' this by saying that we would expect 200,000 / 2,000,000 = 0.1 people to match by chance alone.
- When would this be appropriate?

Other evidence

- Victim did not pick out Adams in identification parade, and even said he did not look like her attacker!
- Adam's girlfriend provided alibi (unchallenged)
- Let b_1 = identification evidence, b_2 = alibi evidence
- Assume these are conditionally independent
- Assume $p(b_1|g) = 10\%, p(b_1|\overline{g}) = 90\%$, so likelihood ratio = 1/9
- Assume $p(b_2|g) = 25\%, p(b_2|\overline{g}) = 50\%$, so likelihood ratio = 1/2
- Assume prior $p(g) = 1/200,000$
- Then based on b_1, b_2, odds on guilt are 1 in 3.6 million
- Now with DNA evidence, odds on guilt are 1 in 1.8 = 5/9, or a probability of 36% of guilt

What actually happened

- Jury were led through a Bayesian analysis asked to provide their own probabilities
- Declared Adams guilty, but went to Appeal
- Appeal allowed and retrial ordered.
- Jury again presented with Bayesian arguements, again found Adams guilty, and again went to Appeal
- Went to Appeal Court, who strongly objected to Bayesian reasoning being used

what the Appeal Court thought of Bayes theorem

- "*it trespasses on an area peculiarly and exclusively within the province of the jury, namely the way in which they evaluate the relationship between one piece of evidence and another*"
- "*to introduce Bayes's theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity*"
- The task of the jury was said to be to "*evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them*".
- Appeal dismissed, but it was revealed Adams had a brother whose DNA had not been tested