Evidence Synthesis Begins Here

Tony O'Hagan

Outline of the course

- Three sessions

1. Evidence Synthesis Begins Here
   - Bayes' theorem – synthesising prior information and data
   - Elicitation – formulating prior information
   - Sequential updating – synthesising prior and multiple data items
2. Pooling and Meta-analysis
   - Pooling – synthesising prior judgements from multiple experts
   - Meta-analysis – synthesising related data items
3. Model-based Synthesis
   - Modelling – synthesising evidence from disparate sources
   - Propagating uncertainty – computing the synthesis

Bayes' theorem

Synthesising prior information and data

Bayesian statistics

- This course will adopt the Bayesian paradigm for statistics
  - Some kinds of evidence synthesis can be handled within the traditional frequentist framework
    - But if you want to learn about that you will need another teacher!
  - The Bayesian approach has a number of advantages
    - It is more general and flexible
    - It provides results that are more natural and interpretable
- This session begins with Bayes' theorem
  - The simplest and most fundamental kind of evidence synthesis
  - Some of you may not be familiar with Bayesian statistics
    - Others may need some more persuading of its benefits!

A little history

- Thomas Bayes' famous paper of 1763 presented posthumously to the Royal Society by his friend Richard Price
  - Widely adopted by others (Laplace etc.)
  - Solved the 'inverse probability' problem
  - Fell out of favour, largely for want of modern ideas of probability

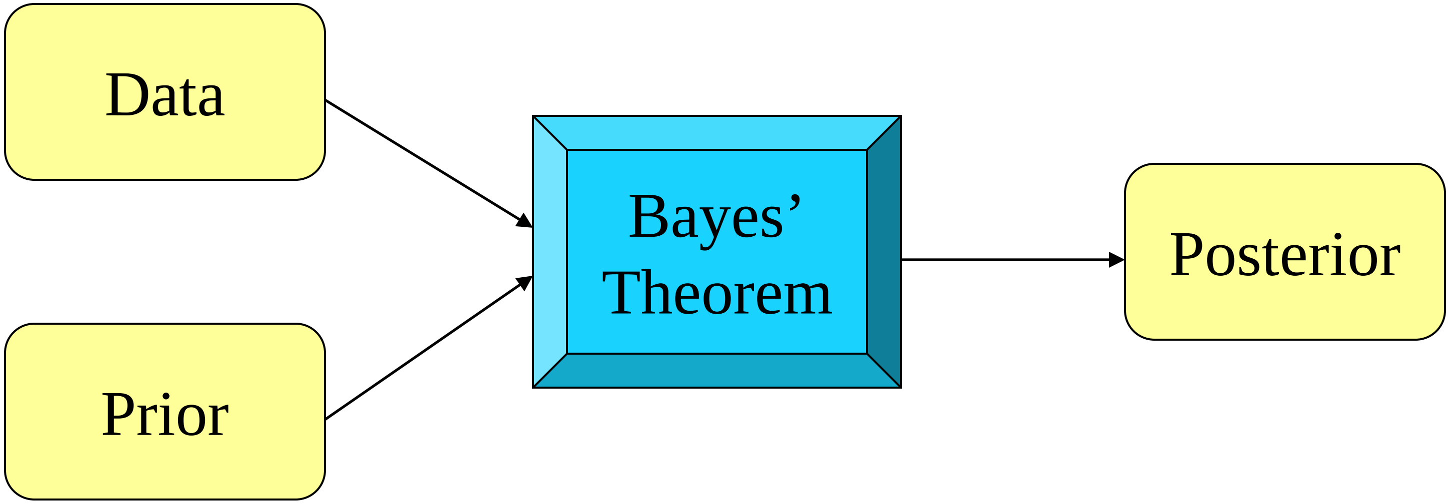
- Around 1900 mathematicians were rebuilding foundations
  - Mathematical theory of inference, starting with Karl Pearson
  - Fully developed by Egon Pearson and Ronald Fisher in 1920s/30s
  - Later dubbed 'frequentist'
- Revival of Bayesian methods in 1950s
  - Jimmy Savage and Dennis Lindley

The Bayesian method

In Bayesian statistics we

- Create a model to link data to parameters
  - This step is also a feature of frequentist statistics
- Formulate prior information about parameters
  - This one is unique to the Bayesian approach
- Combine the two sources of information using Bayes' theorem
  - Evidence synthesis
- Use the resulting *posterior* distribution to derive inferences about parameters

The Bayesian method (cont.)



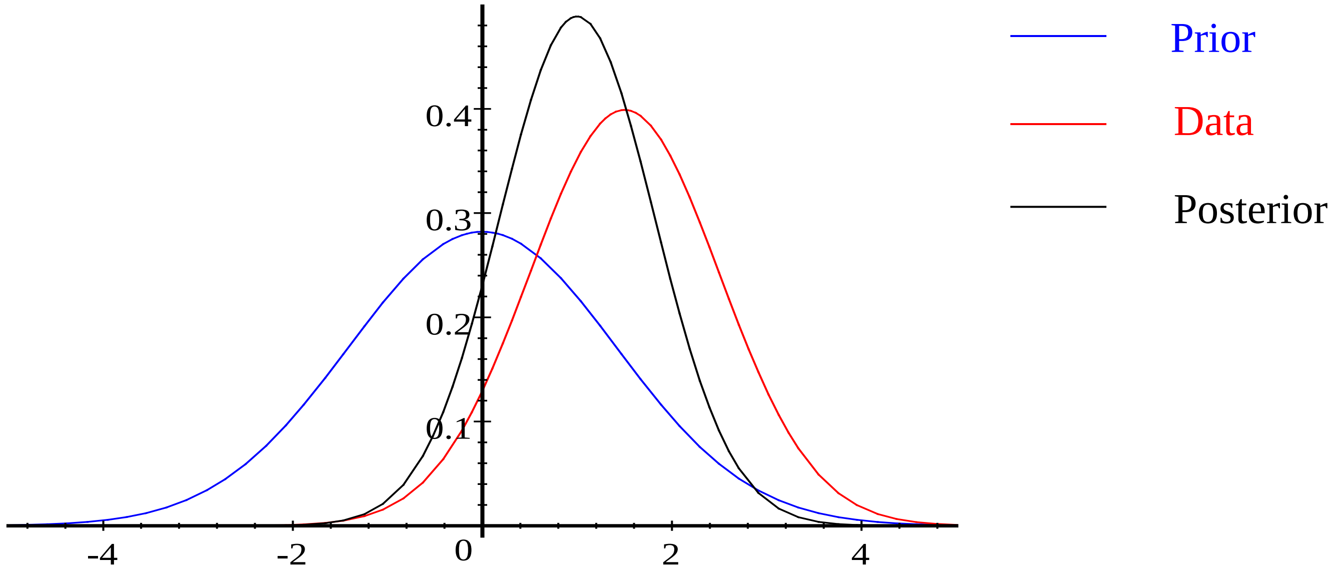
Bayes' theorem

- Bayes' theorem has a simple form:

$$p(\theta \mid x) \propto p(\theta) \times p(x \mid \theta)$$

  - Where θ is the unknown quantity of interest, x is the data
  - The first term on the RHS, $p(\theta)$, is the prior density for θ
    - Expressing the prior information
  - The second term on the RHS, $p(x \mid \theta)$, is the likelihood
    - Specifying the information in the data, as in frequentist statistics
  - The LHS, $p(\theta \mid x)$, is the posterior density for θ
    - Expressing what is known about θ after synthesising the two sources
  - The symbol $\propto$ is read as "is proportional to"
    - Meaning the RHS needs to be scaled to make it integrate to one
    - As any proper density function such as $p(\theta \mid x)$ must

The Bayesian method (cont.)



Bayes' theorem in action

Posterior ∝ Prior × Likelihood

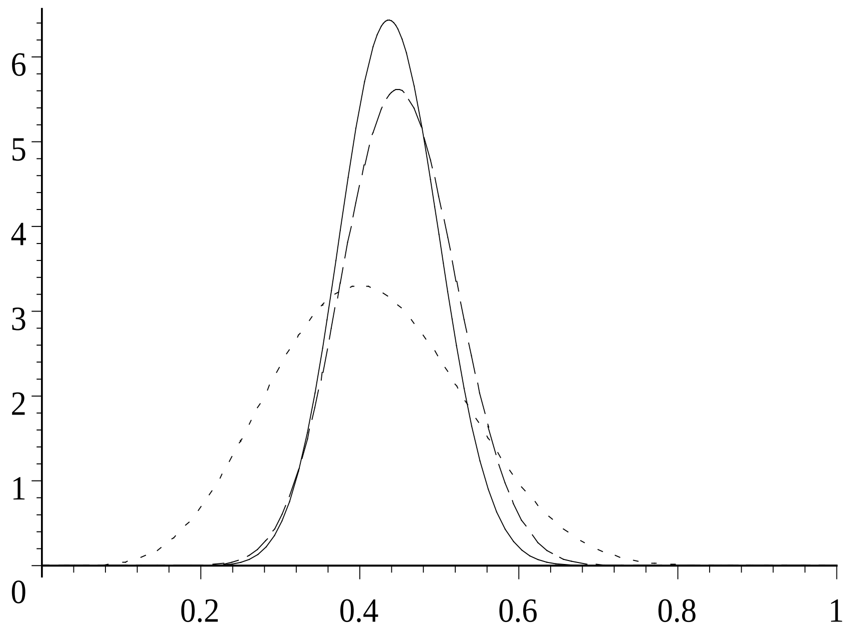
Example – prior distribution

- A new treatment protocol is proposed for treating a certain form of cancer
- With current treatment, 40% of patients survive 6 months
- Prior information suggests the new protocol might be expected to improve survival slightly
  - The doctor gives a prior estimate of 45%
  - She expresses her uncertainty in a standard deviation of 7%
- The prior information is represented in probabilistic form
  - Formally, a Beta (22.28, 27.23) prior distribution is selected to give the required mean and standard deviation
  - Prior probability that new treatment is superior to old is 0.76

Example – first year's data

- After one year of using the new treatment protocol, 15 patients have been observed for 6 months, and 6 have survived (i.e. 40%)
- The posterior distribution is Beta (28.28, 36.23)
  - The mean is 0.438 (43.8%), which is a compromise between the prior estimate of 45% and the data estimate of 40%.
  - It is closer to the prior because the prior information is more informative than this small sample.
  - Prior information leads to a higher estimate than the frequentist analysis would have done on the basis of the data alone.
  - There is now only a 0.73 probability that the new treatment is superior

Example – first year triplot



Triplot: First year's data.

Dashed line is the prior.

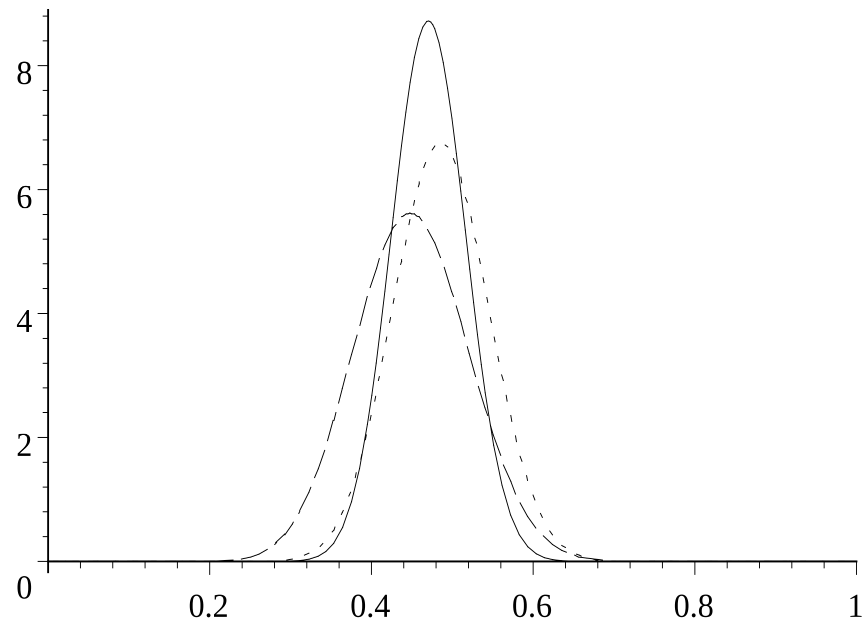
Dotted is the likelihood.

Solid line is the posterior.

Example – more data

- After two years, a total of 70 patients have been followed up for 6 months, and we now have 34 survivors (i.e. 48.6%)
- The posterior is now Beta (56.28, 63.23).
  - Mean is 0.471 (47.1%).
  - Still compromises between the prior mean (45%) and data estimate (48.6%), but is now closer to the data because they are now more informative than the prior
  - The prior is now exerting a moderating influence on the data!
  - The posterior probability that the new treatment is more effective than the old is 0.941

Example – new triplot



Triplot: Two years' data.

Dashed line is the prior.

Dotted is the likelihood.

Solid line is the posterior.

First Bayes

- Reasonably user-friendly, software exists that allows you to explore simple examples like this
  - Including drawing triplots

- First Bayes can be downloaded from

  http://www.firstbayes.co.uk

  and is free for non-commercial use

Frequentist versus Bayesian Statistics

Bayesian versus frequentist

- The debate between these two philosophies has been running for decades
  - Sometimes very heatedly!
- The differences run deep
  - The nature of probability
  - The nature of parameters
  - The meaning of inference

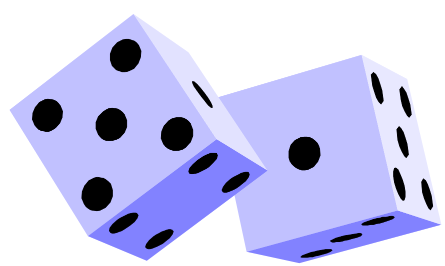
The nature of probability

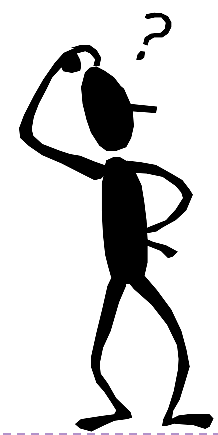
Frequentist

- Probability is a limiting long-run frequency
- It therefore applies only to events that are (at least in principle) repeatable

Bayesian

- Probability measures a personal degree of belief
- It applies to any event or proposition about which we are uncertain





The nature of parameters

Frequentist

- Parameters are not repeatable random things
- They are therefore not random variables, but fixed (unknown) quantities

Bayesian

- Parameters are unknown
- They are therefore random variables

The use of prior information

Frequentist

- Prior information is in principle subjective
- It therefore has no place in science
- In particular, prior distributions do not exist

Bayesian

- Prior information is in principle subjective
- But so is science
- Prior distributions can and should be as objective as possible
- Don't throw away information

The nature of inferences

Frequentist

- Unbiased estimators, significance tests, confidence intervals
- Justified through long-run repetition
- Do not (although they appear to) make statements about parameters

Bayesian

- Point estimates, comparing hypotheses, interval estimates, and much more
- Make direct probability statements about parameters

The nature of inferences (cont.)

Frequentist

- We reject this hypothesis at the 5% level of significance
- In 5% of samples where the hypothesis is true we will reject it (but nothing stated about this sample)

Bayesian

- The probability that this hypothesis is true is 0.05
- The statement applies on the basis of this sample (as a personal degree of belief)

The nature of inferences (cont.)

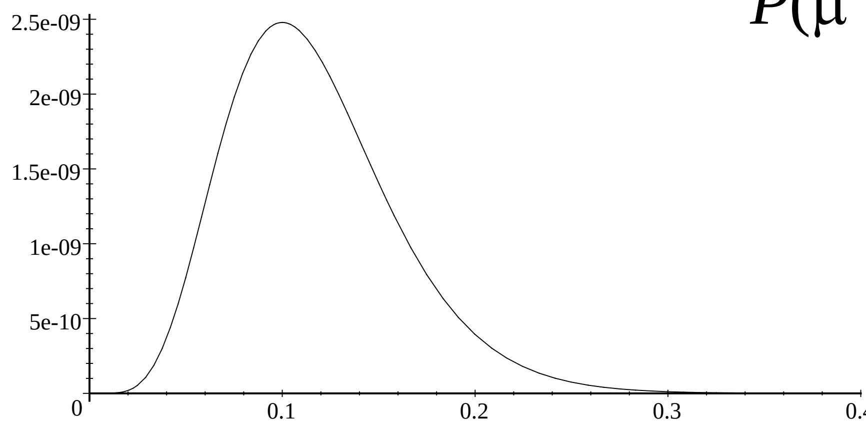
Frequentist

- Everything must be expressed as an (unbiased) estimator, significance test or confidence interval

Bayesian

- Many other kinds of statements about parameters can be made
- Of particular value is simply to plot the posterior density

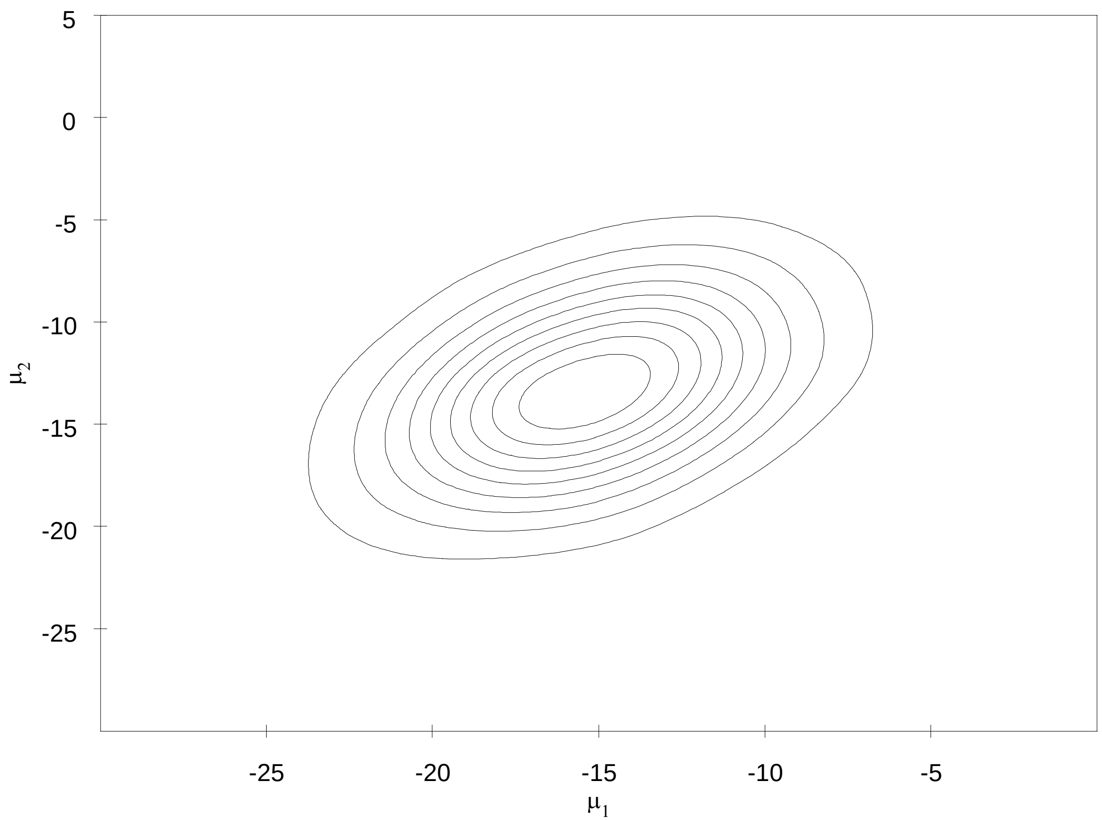
The nature of inferences (cont.)



$P(\mu > 0.2) = 0.07$

Posterior density of parameter μ

The nature of inferences (cont.)



Contours of posterior density of parameters μ_1 and μ_2

Elicitation

Formulating evidence from an expert

Formulating prior distributions

- We need to specify prior information about the parameters
  - Conventional weak prior distributions
  - Genuinely informative prior distributions
- Weak prior distributions
  - Generally very wide and flat densities, very large variances
  - Many different names
    - Non-informative, diffuse, vague, ignorance, reference, 'objective', default, flat
  - Different principles lead to different weak priors
    - Choice should not matter if other evidence strong enough
    - Don't believe any claims for a particular choice to be the unique logical or objective weak prior
- Informative prior distributions based on expert judgement

What is elicitation?

- The process of
  - representing the knowledge
  - of one or more persons (experts)
  - concerning an uncertain quantity
  - as a probability distribution for that quantity

- Typically conducted as a dialogue between
  - the experts – who have substantive knowledge about the quantity (or quantities) of interest – and
  - a facilitator – who has expertise in the process of elicitation

The facilitator

- The role of facilitator is important and demanding
  - Need to understand various psychological pitfalls
  - Need to manage the interaction with the expert(s)
  - Need to prepare thoroughly and work to a well-designed plan
  - Need to document the process

The SHELF system

- SHELF is a package of documents and simple software to aid elicitation
  - General advice on conducting the elicitation for would-be facilitators
  - Templates for recording the elicitation
    - Suitable for several different basic methods
  - Annotated versions of the templates with detailed guidance
  - Some R functions for fitting distributions and providing feedback
- SHELF is freely available and we welcome comments and suggestions for additions
  - Developed by Jeremy Oakley and myself
    - R functions by Jeremy
  - http://tonyohagan.co.uk/shelf
  - Version 2.0 released September 2010

Contents

- SHELF Overview_v2.0
- SHELF Pre-elicitation Briefing
- SHELF Pre-elicitation Form + version with notes
- SHELF 1 (Context) + version with notes
- SHELF 2 (Distribution) Q + version with notes
- SHELF 2 (Distribution) QP + version with notes
- SHELF 2 (Distribution) R + version with notes
- SHELF 2 (Distribution) RP + version with notes
- SHELF 2 (Distribution) T + version with notes
- SHELF 2 (Distribution) TP + version with notes
- SHELF2 Distribution fitting instructions
- shelf2.R

ELICITATION RECORD – Part 2 – Distribution

Tertile Method

Elicitation title	
Session	
Date	
Quantity	
Start time	

Definition	
Evidence	
Plausible range	
Median	
Upper and lower tertiles	
Fitting	
Group elicitation	
Fitting and feedback	
Chosen distribution	
Discussion	

Median	The next steps should be done by each expert **separately**, without discussion. Each expert should specify their median value for X. This is a value such that they think 'X lies below the median' and 'X lies above the median' are equally likely propositions. Formally, if M is the median, then $P(X < M) = 0.5$. [The facilitator should instruct the experts to write down their own median values, but not to reveal them yet. Nothing should be written in this field until after the upper and lower tertiles have also been elicited.] *The judgement of equal probability is generally found to be simple for experts, and is not subject to systematic biases.*
Upper and lower tertiles	Each expert should now specify their upper and lower tertiles by considering the range from L to U and dividing it into three equally likely intervals. Formally, if T1 is the lower tertile and T2 is the upper tertile, then $P(L < X < T1) = P(T1 < X < T2) = P(T2 < X < U) = 0.33$. Before deciding definitely on these values, experts should be asked to check not only that they regard each of the three ranges (L to T1, T1 to T2 and T2 to U) as equally likely, but that the ranges T1 to M and M to T2 are also equally likely.. [When asking for the tertiles, the facilitator should point out that generally experts would feel that values of X close to M are more probable than values close to L or U, and so the interval

Sequential updating

Synthesising multiple items of data

Bayes' theorem is naturally sequential

- Remember the cancer treatment example
  - First we updated using the first year's data
    - Prior Beta (22.28, 27.23), 6 survivors out of 15
    - Posterior Beta (28.28, 36.23)
  - We then updated using two years' data
    - 34 survivors out of 70
    - Posterior Beta (56.28, 63.23)
  - We could have done this in two ways
    - Adding all two years' data to the prior
    - Adding the second year's data to the first year's posterior
- This works quite generally
  - "Today's posterior is tomorrow's prior"

Sequential updating

- Suppose we have data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
  - Bayes' theorem: $p(\theta \mid \mathbf{x}) \propto p(\theta) \times p(\mathbf{x} \mid \theta)$
  - Likelihood factorises:

    $$p(\mathbf{x} \mid \theta) = p(x_1 \mid \theta) \times p(x_2 \mid x_1, \theta) \times \ldots \times p(x_n \mid x_1, x_2, \ldots, x_{n-1}, \theta)$$
  - Often this simplifies to

    $$p(\mathbf{x} \mid \theta) = p(x_1 \mid \theta) \times p(x_2 \mid \theta) \times \ldots \times p(x_n \mid \theta)$$
- We therefore get

  $$p(\theta \mid \mathbf{x}) \propto [[p(\theta) \times p(x_1 \mid \theta)] \times p(x_2 \mid x_1, \theta)] \times \ldots$$
  $$\propto [[p(\theta \mid x_1) \times p(x_2 \mid x_1, \theta)] \times p(x_3 \mid x_1, x_2, \theta)] \times \ldots$$
  - Successive items of data are employed to update the synthesis
    - Using the previous posterior as prior

Summary of session 1

- Bayes' theorem has a natural evidence synthesis interpretation
  - Synthesising prior information and data
- Prior information can be formulated rigorously using elicitation
  - Often resulting in valuable additional information
  - SHELF system provides a starting point for would-be facilitators
- Where we have multiple data items Bayes' theorem has a nice sequential updating property
  - Evidence can be synthesised in one block or one item or a time
    - Or even in handy-sized blocks
- The kinds of evidence synthesis considered so far all deal with direct evidence
  - All items are expressed as relating to the same θ