STAT40380/STAT40390/STAT40850
Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

March 2016

The likelihood principle

- There are other areas where Bayesian and frequentist statistical inference disagree
- The most substantial of these is the *likelihood principle*, which most frequentist hypothesis tests (and even some Bayesian methods) violate
- The principle states approximately that *all of the information from the data about our unknown parameters is stored in the likelihood function*
- Remember this is opposite to many other standard methods which use information which has not been observed to make inference about the unknown parameters
- The original result is due to Allan Birnbaum (1962), with more complete discussion in Berger and Wolpert (1988)

A quick exercise

Example

Suppose we observe a number of coin tosses such that we get 9 occurences of heads and 3 of tails. We might make two possible assumptions:

1. That the total number of coin tosses was fixed at $n = 12$, so if x represents the number of heads then $x \sim Bin(n, \theta)$
2. That we stopped when we reached 3 occurences of tails, so if x represents the number of heads then $x \sim NBin(k, 1 - \theta)$

Find the likelihoods and p-values under each of the models

Set-up

- Let x be a single data point (possibly a vector)
- Let θ be a parameter about which we are interested
- Let $p(x|\theta)$ be the pdf of x given the parameter θ
- Let $\tilde{x}$ be a random variable covering all the possible values of x that we might observe
- Define $E = \{\tilde{x}, \theta, p(x|\theta)\}$ to be an *experiment* that is carried out to observe x and learn about θ
- Define $Ev\{E, x, \theta\}$ to be the *evidence* provided about the value of θ after carrying out experiment E and observing x
- To a Bayenian statistician, Ev will normally be a posterior distribution
- To a frequentist, Ev might be made up of signiicance tests and confidence intervals

Example

Example

Suppose you are interested in the proportion θ of defective articles coming from a factory. A possible experiment E would consist of observing a fixed number n of articles chosen at random and observing the number x defective, so that $p(x|\theta)$ is a binomial density. If we give n a specific value (eg $n = 100$), then E is fully determined. If we then observe $x = 3$, then we denote our evidence as $Ev(E, 3, \theta)$, which would then contain our conclusions about the value of θ

The conditionality principle

- Idea: If you have decided which of two experiments you performed by tossing a coin, then once you tell me the *end result* of the experiment, my inference about the parameter θ will not be affected by the result of the coin toss
- Example: Suppose that a substance is sent to be analysed at one of two labs, Dublin or Belfast. The two labs seem equally well-equipped and a coin is tossed to decide which one whould be chosen. After the results come back, should the resulting report take into account the fact that the other laboratory could have been chosen? According to the conditionality principle the answer is no
- Define two experiments: $E_1(\tilde{y}, \theta, p(y|\theta))$ and $E_2(\tilde{z}, \theta, p(z|\theta))$
- Let $\tilde{k}$ be a random variable with $p(k = 1) = 0.5 = p(k = 2)$, independent of θ

The conditionality principle (cont)

- Define E^* as the *mixed* experiment $E^*(\tilde{x}, \theta, p(x|\theta))$ where we carry out E_1 if $k = 1$ and E_2 if $k = 2$.
- Now

$$x = \begin{cases} (1, y) & \text{if} \quad k = 1 \\ (2, z) & \text{if} \quad k = 2 \end{cases}$$

  and

$$p(x|\theta) = \begin{cases} \frac{1}{2} p(y|\theta) & \text{if} \quad k = 1 \\ \frac{1}{2} p(z|\theta) & \text{if} \quad k = 2 \end{cases}$$

- *The weak conditionality principle* states

$$Ev(E^*, x, \theta) = \begin{cases} Ev(E_1, y, \theta) & \text{if} \quad k = 1 \\ Ev(E_2, z, \theta) & \text{if} \quad k = 2 \end{cases}$$

  that is, our inference on θ depends only on the experiment actually performed

The sufficiency principle

- We have seen the sufficiency principle before in earlier lectures
- The *sufficiency principle* states that: if $t(x)$ is sufficient for θ given x, then any inference we make about θ is based on the value of t. Once we know t we hav no further need for the value of x.
- More formally, for an experiment $E(\tilde{x}, \theta, p(x|\theta))$, if $t(x_1) = t(x_2)$, then:

$$Ev(E, x_1, \theta) = Ev(E, x_2, \theta)$$

- The sufficiency principle is not a controversial aspect of the likelihood principle

The likelihood principle

Likelihood principle

Consider two experiments $E_1(\tilde{y}, \theta, p(y|\theta))$ and $E_2(\tilde{z}, \theta, p(z|\theta))$. If we observe y^* and z^*, such that:

$$p(y^*|\theta) = c \times p(z^*|theta)$$

where c is a constant, then:

$$Ev(E_1, y^*, \theta) = Ev(E_2, z^*, \theta)$$

ie if the likelihood contains the same information for each experiment then our inferenced about θ will be identical

- Birnbaum proved that the likelihood principle is equivalent to the suffucuency and conditionality principles
- Beware: Jeffreys' priors violate the likelihood principle

The stopping time principle

- One of the key consequences of the likelihood principle is that the rules used to stop an experiment are irrelevant in drawing inferences about the parameter θ
- Examples:
  1. Fixing the number of trials n and then running n samples
  2. Stopping after the first k successes (or failures)
  3. Stopping after the first k successes (or failures) or after n trials, whichever happens first
  4. Stopping after an observation is a fixed distance away from the mean
- If $x_1, \ldots, x_s$ are our data and s is a stopping rule, then:

$$p(\mathbf{x}|\theta) = p(x_1|\theta) \times \cdots \times p(x_s|\theta)$$

  does not depend on the assumptions behind s
- This is because, according to the likelihood principle, only the likelihood determines our inference about θ

More on stopping times

- A frequentist statistician is supposed to choose the stopping rule before the experiment and then follow it exactly
- For a Bayesian statistician, the stopping rule is irrelevant and of no consequence to the posterior distribution
- A quote from Savage (1962)

  "I learned the stopping rule principle in . . . the summer of 1952. Frankly I thought is a scandal that anyone in the profession could advance an idea so **patently wrong**, *even as today I can scarcely believe some people can resist an idea so* **patently right**"