APTS module *Statistical Inference*

December 2007

Assessment material

The work provided here is intended to take students up to half a week to complete. Students should talk to their supervisors to find out whether or not their department requires this work as part of any formal accreditation process (APTS itself has no resources to assess or certify students). It is anticipated that departments will decide on the appropriate *level* of assessment locally, and may choose to drop some (or indeed all) of the items, accordingly.

Students should do both A and B. Part B, at least, is likely to require access to library and/or computer resources.

Part A

1. [From lecture 1.] Suppose that s^2 is the residual mean square with $d_{\rm res}$ degrees of freedom in a normal theory linear model and σ^2 is the true variance. Suppose that it is decided to base inference about σ^2, whether Bayesian or frequentist, solely on s^2. You may assume that the random variable S^2 is such that $d_{\rm res}S^2/\sigma^2$ is distributed as chi-squared with $d_{\rm res}$ degrees of freedom.
   (i) What is the 95 per cent upper confidence limit for σ? (ii) For large d the chi-squared distribution with d degrees of freedom is approximately normal with mean d and variance $2d$. How large would $d_{\rm res}$ have to be for the 95 percent upper limit to be $1.2s_{\rm res}$? (iii) What is the conjugate prior in a Bayesian analysis? When, if ever, would posterior and confidence limits agree?

2. [From lecture 2. If you get stuck, any good book on the analysis of binary/categorical data should have some discussion of this.] For the binary matched pairs model, derive the conditional binomial distribution for inference on the common log odds ratio ψ. Discuss whether it is reasonable to discard all the data from 'non-mixed' pairs.

3. [From lecture 4.] Let $Y_1, \ldots, Y_n$ have independent Poisson distributions with mean μ. Obtain the maximum likelihood estimate of μ and its variance,

   (a) from first principles;

   (b) by the general results of asymptotic theory.

   Suppose now that it is observed only whether each observation is zero or non-zero.

   (c) What now are the maximum likelihood estimate of μ and its asymptotic variance?

   (d) At what value of μ is the ratio of the latter to the former variance minimized?

   (e) In what practical context might these results be relevant?

4. [The 'extra' exercise from end of lecture 4.] Suppose that $Y_1, \ldots, Y_n$ are independent, with $Y_i \sim N(\lambda + \psi x_i, \sigma^2)$ and σ^2 known.

   (a) Calculate the expected information matrix $i(\psi, \lambda)$, and relate this to what you know about least squares.

   (b) Find a new parameterization, (ψ, τ) say, in which τ is orthogonal to ψ.

Part B

One of these two:

1. Write a short summary (2 pages or so) of the uses and limitations of *one* of the following:

   (a) 'non-informative' prior distributions;

   (b) orthogonal parameters;

   (c) empirical Bayes methods.

2. [From lecture 6.] Let $Y_1, \ldots, Y_n$ be independently binomially distributed each corresponding to ν trials with probability of success θ. Both ν and θ are unknown. Construct simple (inefficient) estimates of the parameters. When would you expect the maximum likelihood estimate of ν to be at infinity? Set up a Bayesian formulation.

   [HINT: For the simple estimates, think of two mathematical properties specifying aspects of the binomial distribution, equate these to the corresponding features of the data and solve for an estimate of ν. Are there circumstances in which the estimate is infinite or undefined? Why is this? Suggest a combination of parameter values for which such anomalies are quite likely and simulate say 10 realizations and look at the corresponding likelihoods. When interesting parameter combinations have been found make a more detailed study.]