STAT 530 - Bayesian Analysis (Term 2, 2009-10)

ASSIGNMENT 1

NOTE: More questions will be added as we cover material. Due Friday February 12.

1. [Jan. 12] Say that the statistical model is comprised of observing $Y \sim N(\theta, \sigma^2)$, where θ is unknown but σ is known. The standard conjugate prior for this situation is $\theta \sim N(\mu, \tau^2)$, involving two hyperparameters (μ, τ^2). Show that another conjugate prior, involving five hyperparameters, is a mixture of two normal distributions. That is, in technical terms, the prior density takes the form $p(\theta) = p\tau_1^{-1}\phi\{(\theta - \mu_1)/\tau_1\} + (1-p)\tau_2^{-1}\phi\{(\theta - \mu_2)/\tau_2\}$, where $\phi()$ is the standard normal density function. Give the form of the posterior distribution, i.e., how are the hyperparameters updated upon observation of $Y = y$. Can you give some intuition for how this update makes sense, particularly in terms of what happens to p. For a few choices of hyperparameters and y and σ, plot the prior and posterior densities together. Finally, explain briefy how the form of this prior to posterior update adapts to the setting of observing n independent and identically distributed observations from a normal distribution with unknown mean but known variance.

2. [Jan. 13] Choose a conjugate prior setting and investigate the performance of credible intervals (let's say at the 80% level) in this setting by doing the following.

(a) Demonstrate that for some combination of little information in the data (i.e., a small sample size) and non-negligible information in the prior (e.g., not a uniform distribution as the prior), the frequentist coverage will be wrong. More specifically, repeatedly generate datasets for a true parameter value in the tail of the prior, and verify that the proportion of 80% credible intervals containing this true value is less than the nominal 80%. Similarly, show that higher-than-nominal coverage obtains for a true value 'in the middle' of the prior.

(b) Now take the setting you chose in part (a), and investigate a different kind of coverage. Repeatedly sample a (parameter,dataset) pair by sampling a parameter value from the prior distribution and then sampling a dataset from the model using this parameter value. What do you find about the proportion of 80% credible intervals containing the parameter value in this kind of repeated sampling? (You could regard this kind of repeated sampling as corresponding to a series of studies of different phenomena, so that the true parameter value is different for each study.)

NOTE: When using simulation, please be mindful to control and/or report 'simulation error,' or 'simulation variability' if you prefer. Sometimes statisticians are quick to tell other kinds of scientists to control/report variability, but slow to do this in their own domain.

3. [Jan. 24] In class/text we compared the MSE of the sample mean to the MSE of the Bayesian point estimator, in the context of the two-parameter normal model with conjugate prior. Can you reformulate this comparison to more directly link up (i) the extent to which the true parameter values are in the middle or tail of the prior distribution with (ii) the extent to which the Bayesian estimator has better or worse MSE than the sample mean. There is quite a bit of flexibility in how you might approach this, particularly in terms of algebra versus simulation. Also bear in mind that the parameter vector is two-dimensional, so some sort of countour plot might help (see Fig 5.4 in the text for inspiration). Another thought is to go after how much weight the prior places on parameter values for which the Bayes MSE is the worst.

4. [Jan. 26] Text, Exercise 4.5.

5. [Jan. 29] Text, Exercise 6.2.

6. [Feb. 1] Consider the following situation. The observable data consists of a single observation $Y \sim N(\theta, 1)$. The investigator will use a prior distribution of the form $\theta \sim N(0, \tau^2)$, and estimate θ by $E(\theta|Y)$. What is the Bayes risk of this estimator when 'nature's prior' is $\theta \sim N(0, \kappa^2)$, i.e., what is the average mean-squared error of the investigator's estimator with respect to 'nature's' distribution over the parameter space? Fix κ^2 and plot the Bayes risk as the investigator's choice of τ^2 varies. Does the pattern you see conform with the basic tenets of decision theory? Is there any support for the notion that an 'overly-flat' prior is less damaging than an 'overly-concentrated' prior?

Last update Feb. 1, 2010. No further questions to be added.