Petter Mostad
Mathematical Statistics
Chalmers

MSA100/MVE185 Computer Intensive Statistical Methods

TRIALEXAM October 2008
Allowed to use during the exam: Pocket calculator, books, copies, and notes.

1. Assume we make three observations of an unknown quantity which has as prior a normal distribution with expectation 4 and standard deviation 1. We use an unbiased measuring instrument with normally distributed error with standard deviation 0.5. The results are 2, 3, 4. What is the posterior distribution for the unknown quantity?

2. Explain briefly the differences between Importance Sampling and Sampling Importance Resampling. In what situations would you use one or the other?

3. Assume you want to simulate from a probability distribution on $\mathbb{R}^2$ whose density is proportional to a function $g(\theta) = g(\theta_1, \theta_2)$. You have decided to use the Metropolis-Hastings algorithm, and you want to use a proposal function that proposes a new value $\lambda \in \mathbb{R}^2$ given the old value $\theta \in \mathbb{R}^2$ as follows:

   With probability 0.8, λ is equal to θ plus (u_1, u_2), where u_1 and u_2 are independently uniformly distributed on $[-1, 1]$. With probability 0.2, λ is drawn directly from the binormal distribution with expectation zero and covariance matrix

   $$\Sigma = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

   Write the expression (involving the function g and the old value θ) for the probability of accepting a new proposed value λ.

4. Prove that, when the test statistics are independent, the Sidák method for multiple testing adjustment controls for the Family Wise Error Rate. (See the lecture notes for the necessary definitions).

5. We consider ways to analyse data in a contingency table, looking at a table with two rows and three columns. We denote the observed count for the i'th row and j'th column by y_{ij}, and we use standard notation for sums of rows and columns, as follows:

y_{11}	y_{12}	y_{13}	$y_{1\cdot}$
y_{21}	y_{22}	y_{23}	$y_{2\cdot}$
$y_{\cdot 1}$	$y_{\cdot 2}$	$y_{\cdot 3}$	$y_{\cdot\cdot}$

   We assume that $y = (y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23})$ is multinomially distributed with probabilities $p = (p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23})$, where $\sum p_{ij} = 1$. The multinomial with probabilities $q = (q_1, \ldots, q_k)$ is given by

   $$\pi(z_1, \ldots, z_k \mid q_1, \ldots, q_k) = \frac{\Gamma\left(\sum z_i + 1\right)}{\Gamma(z_1 + 1)\Gamma(z_2 + 1)\ldots\Gamma(z_k + 1)} q_1^{z_1} \ldots q_k^{z_k}.$$

   The prior distribution for p is specified in various ways in the models below.

(a) In the model M_D, we assume that p has a Dirichlet$(1, 1, 1, 1, 1, 1)$ distribution, where the Dirichlet$(\alpha_1, \ldots, \alpha_k)$ distribution is given by

$$\pi(p_1, \ldots, p_k \mid \alpha_1, \ldots, \alpha_k) = \frac{\Gamma\left(\sum_i^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \ldots p_k^{\alpha_k - 1}.$$

Find the posterior $\pi(p \mid y)$ and the prior predictive distribution $\pi(y)$ assuming the model M_D.

(b) In the model M_I, we assume that $p_{ij} = r_i s_j$, for $i = 1, 2$ and $j = 1, 2, 3$, where the parameter $\theta = (r_1, r_2, s_1, s_2, s_3)$ has a prior distribution that is a product of a Dirichlet$(1, 1)$ and a Dirichlet$(1, 1, 1)$ distribution, so that $(r_1, r_2) \sim$ Dirichlet$(1, 1)$ and $(s_1, s_2, s_3) \sim$ Dirichlet$(1, 1, 1)$. Find the posterior $\pi(\theta \mid y)$ and the prior predictive distribution $\pi(y)$ assuming the model M_I.

(c) Find the Bayes Factor for making a choice between the two models.

(d) What is the value of the Bayes Factor with the following data

0	1	0
1	0	2