Petter Mostad
Mathematical Statistics
Chalmers

MSA100/MVE185 Computer Intensive Statistical Methods

Exam 25 October 2008
Examiner: Petter Mostad, phone 0707163235, visits the exam at 9.30 and at 11.30.
Allowed to use during the exam: Pocket calculator, books, copies, and notes.

1. Assume y has a Negative Binomial distribution with parameters p and r, where $r > 0$ is a given number of successful trials, and p with $0 < p < 1$ is the unknown probability of success in each trial, so that

$$\pi(y \mid p) = \binom{y + r - 1}{y} p^r (1-p)^y = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)} p^r (1-p)^y.$$

Assume also that p has as a prior a Beta distribution with parameters α and β, so that

$$\pi(p \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}.$$

(a) Find the posterior distribution $\pi(p \mid y)$.

(b) Find the prior predictive distribution $\pi(y)$.

(c) Suppose Karl is doing a lab experiment which he supposes has constant probability of success p, with independent results in each trial. He needs 10 successful experiments, and to achieve that, he has had to try 36 times in total. Using the improper prior

$$\pi(p) = \frac{1}{p(1-p)},$$

what is the posterior distribution for p? What is the expectation of that posterior?

(d) After the 36 trials above, Karl learns that he will need to perform an additional n successful experiments. Write down the probability distribution for the number of unsuccessful experiments he will have to endure to reach this goal.

2. Assume you have computed hypothesis tests for each of four null hypotheses H_{01}, H_{02}, H_{03}, and H_{04}, and that the resulting p-values were

H_{01} :	0.071
H_{02} :	0.002
H_{03} :	0.064
H_{04} :	0.027.

(a) If you want to guarantee that the Family Wise Error Rate is limited by 10%, which hypotheses can you reject?

(b) If you know that all the test statistics are independent, which hypotheses can you reject, still guaranteeing that the Family Wise Error Rate is limited by 10%?

3. Assume a probability density function on the whole real line is defined by

$$f(x) = C\left(1 + x^2 + \log(1 + x^2)\right)^{-2},$$

where C is some constant so that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

(a) Find the mode of the density function, and the parameters of the normal distribution approximating it, with expectation at this mode.

(b) Use the approximation above to compute an estimate for the constant C.

4. Anna is investigating the time to failure of n mechanical toys, handmade at a particular factory. She wants to model y_i, the time to failure for toy i, $i = 1, \dots, n$, with an exponential distribution with rate λ_i, so that

$$\pi(y_i \mid \lambda_i) = \lambda_i \exp(-\lambda_i y_i),$$

where she assumes the λ_i may be different for each toy. She assumes these rates λ_i come from a common Gamma distribution with parameters α and β, so that

$$\pi(\lambda_i \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} \exp(-\beta\lambda_i).$$

(a) Find $\pi(\lambda_i \mid y_i, \alpha, \beta)$, the posterior for the failure rate, given α and β.

(b) Find $\pi(y_i \mid \alpha, \beta)$, the probability distribution for the time to failure of the i'th toy, given α and β.

(c) Assume that Anna uses the improper prior

$$\pi(\alpha, \beta) = \frac{1}{\beta(\alpha + 1)^2}.$$

Write down an expression for a function that is equal to $\log(\pi(\alpha, \beta \mid y_1, \dots, y_n)) + C$, the logarithm of the posterior for alpha and beta, given observed failure times $y_1, \dots, y_n$, plus an unknown constant C.

(d) Explain how the above computations may be used in an algorithm to generate a matrix A where each row $(\alpha, \beta, \lambda_1, \dots, \lambda_n)$ represents a sample from the posterior $\pi(\alpha, \beta, \lambda_1, \dots, \lambda_n \mid y_1, \dots, y_n)$. Explain each part of the algorithm.

(e) Explain how you can use such a matrix A, with sufficiently many rows, to find an approximate answer to the following question: What is the probability that the expected lifetime of the toy that lasted the longest was more than twice the expected lifetime of the toy that lasted the shortest time?

5. Define a probability density on real parameters $\theta_1 > 0$ and $\theta_2 > 0$ by

$$g(\theta_1, \theta_2) = C \exp\left(-\theta_1^2\theta_2 + \theta_1 \log\theta_2\right),$$

where C is a constant so that the density has total integral 1.

(a) Assuming that you can identify the conditional distributions $\pi(\theta_1 \mid \theta_2)$ and $\pi(\theta_2 \mid \theta_1)$, can you suggest a simulation method for obtaining a sample from the distribution defined by g?

(b) What is the distribution $\pi(\theta_1 \mid \theta_2)$?

(c) What is the distribution $\pi(\theta_2 \mid \theta_1)$?