Data provided:
"Statistics Tables" by H.R. Neave

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2008-2009**

Statistical Reasoning **2 hours**

RESTRICTED OPEN BOOK EXAMINATION.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator that conforms to University regulations.
*All answers will be marked but credit will be given for only the best **THREE** answers.*
All questions carry equal marks. Total marks 90.

1 A botanist is interested in the effectiveness of a weak concentration of a weedkiller. Several plots are prepared containing weeds. The ith plot contains n_i weeds. The effectiveness will be measured by the probability that a weed is killed three days after an application of the weedkiller. Suppose that this probability is θ. The numbers of dead weeds in each plot are independent variables with the number in the ith pot having the $\text{Bi}(n_i, \theta)$ distribution. Suppose that r_i dead weeds are found in the ith pot. Let $N = \sum_i n_i$ and $R = \sum_i r_i$.

(a) The botanist's beliefs about θ are that its distribution is unimodal and symmetrical and that its variance is 0.05. The statistician suggests that these beliefs should be represented by a $\text{Be}(2, 2)$. Give a detailed justification for the statistician's proposal for the prior. ***(9 marks)***

(b) Show, with the necessary justification, that the posterior will be $\text{Be}(2 + R, 2 + N - R)$. ***(7 marks)***

(c) Give (in terms on N and R) the posterior mean and variance here. ***(2 marks)***

(d) The following data have been obtained:

i	1	2	3	4	5	6
n_i	9	10	9	11	10	11
r_i	4	4	5	3	5	4

Compare the prior and posterior mean. ***(2 marks)***

(e) If $\Theta \sim \text{Be}(a, b)$ (for $b \geqslant 2$) find the mean and variance of $\Theta/(1 - \Theta)$.
[Recall that $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ and $\Gamma(a + 1) = a\Gamma(a)$.] ***(7 marks)***

(f) Suppose the botanist reveals that her real interest lies in the odds of a weed dying, that is in $\theta/(1 - \theta)$. Compare the prior and posterior mean of this. Comment on its prior variance. ***(3 marks)***

2 The variables X_i for $i = 1, 2, \ldots, n$ are independent and identically distributed, each with the distribution $\text{Ga}(1/2, \lambda/2)$. The prior on λ is $\text{Ga}(6, 3)$.

(a) Show that the mode of $\text{Ga}(a, b)$ is $(a-1)/b$. ***(4 marks)***

(b) Derive the maximum likelihood estimator of λ. ***(6 marks)***

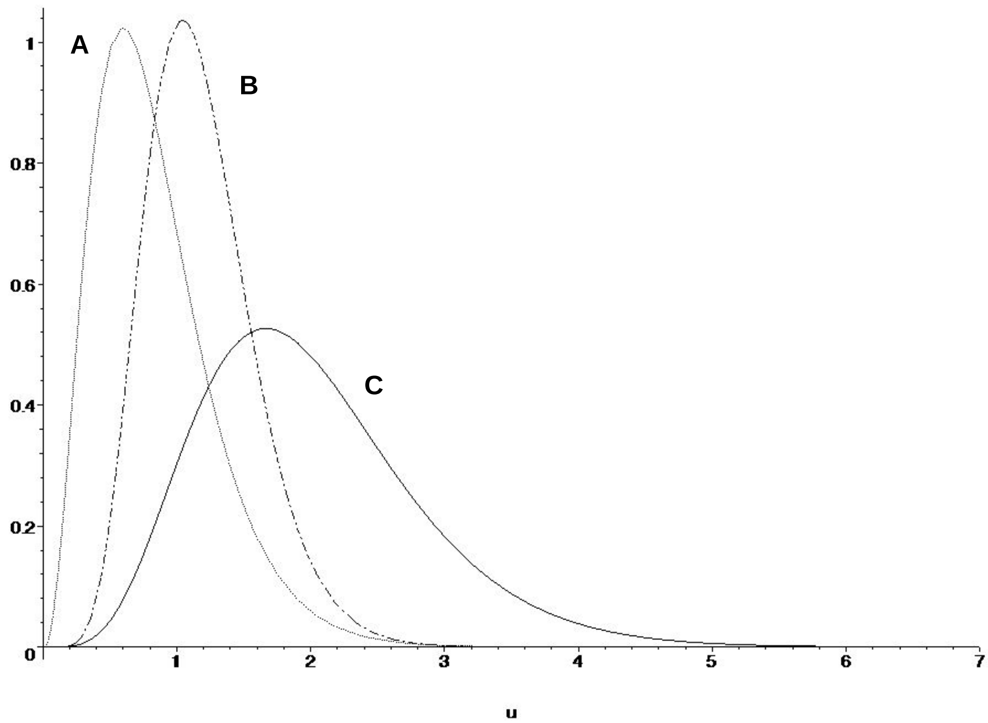
(c) Obtain the posterior distribution in terms of the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. ***(4 marks)***

(d) Suppose that $n = 5$ and the observations are: 0.30, 5.18, 0.14, 1.47, 1.29.

(i) Give the maximum likelihood estimate. ***(2 marks)***

(ii) With the prior specified above, find the posterior mode. ***(2 marks)***

(e) The prior, likelihood and posterior give the following triplot.



Identify which curve is the likelihood, which the prior and which the posterior. Justify your answer. ***(6 marks)***

(f) What does the triplot show about how the data modifies beliefs concerning the hypothesis $H : \lambda \geqslant 2.5$? ***(6 marks)***

3 The variables X_i for $i = 1, 2, \ldots, n$ are independent and X_i has the Poisson distribution with mean $e^{\lambda} t_i$, where λ is unknown and the t_i are known.

(a) Show that

$$L(\lambda; \mathbf{x}) \propto \exp\left(\lambda \sum x_i - e^{\lambda} \sum t_i\right).$$

(5 marks)

(b) Find the maximum likelihood estimator of λ. ***(2 marks)***

(c) Show that there is a uniformly most powerful (UMP) test of $H_0 : \lambda = -1$ against $H_1 : \lambda < -1$ and that, for suitable c, its form is 'Reject H_0 when $\sum X_i < c$'. ***(8 marks)***

(d) The following data has been obtained:

i	1	2	3	4	5	6
t_i	17	9	30	11	4	6
x_i	5	2	7	3	1	1

(i) Give the maximum likelihood estimate of λ. ***(1 mark)***

(ii) Does $\lambda = -1.5$ fall in the 2-unit likelihood region? ***(7 marks)***

(iii) Use that $\sum X_i$ is approximately normal with mean and variance both equal to $e^{\lambda}(\sum t_i)$ to perform an approximate test of H_0 against H_1 using a test of size 0.05. ***(7 marks)***

4 The random vector (X, Y) has the joint density

$$\frac{\sqrt{1-\rho^2}}{2\pi} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2}\right)$$

which depends on the unknown parameter ρ, which is the correlation between X and Y. Suppose $(X_1, Y_1) \dots (X_n, Y_n)$ are n independent vectors from this distribution.

(a) Show that (up to an additive factor that does not depend on ρ) the log-likelihood is

$$\frac{n}{2} \ln(1-\rho^2) + \rho n S$$

where $S = \frac{1}{n} \sum X_i Y_i$. ***(5 marks)***

(b) Show that the maximum likelihood estimator, $\hat{\rho}$, satisfies

$$\hat{\rho}^2 + \frac{\hat{\rho}}{S} - 1 = 0.$$

Show that there is just one possible value for $\hat{\rho}$ in $[-1, 1]$. ***(8 marks)***

(c) Show that the asymptotic variance of the maximum likelihood estimator (for large n) is

$$\frac{(1-\rho^2)^2}{(1+\rho^2)n}.$$

(5 marks)

(d) Suppose that $n = 300$ and $\sum x_i y_i = 200$. Give an approximate 95% confidence interval for ρ. ***(12 marks)***

End of Question Paper