Midterm Exam

INSTRUCTIONS: The exam is open book and open notes. You may not discuss the exam or work with anyone else. If you have questions, please consult one of the instructors.

You may use a computer for computation or to type your answers (if you so choose).

Your completed exam is due at the start of class (10:30 AM) on Wednesday, November 7, 2007.

1. An approximate model (ignoring discreteness) for z, the number of cigarettes smoked per day by a regular smoker, is a Pareto(k,θ) distribution:

$$p(z) = \frac{\theta k^{\theta}}{z^{(1+\theta)}},\ z > k$$

Suppose we collect data $\mathbf{z} = (z_1, \ldots, z_n)$ on a random sample of n regular smokers.

(a) Show that $x_i = \log\left(\frac{z_i}{k}\right)$ has an Exponential(θ) distribution.

(b) A conjugate $\text{Gamma}(a, b)$ prior distribution may be assumed for θ. Briefly explain what is meant by a conjugate prior and comment on the main advantages and disadvantages for choosing such a prior. Using this prior, derive the posterior distribution $p(\theta | x_1, \ldots, x_n)$.

(c) Show that the posterior in part (b) is equivalent to

$$p(\theta | z_1, \ldots, z_n) = \text{Gamma}\left[n + a, b + n\log(\overline{G}/k)\right],$$

where $\overline{G} = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\log(z_i)\right]$ is the geometric mean of the observations $z_1, \ldots, z_n$.

(d) Suppose that n = 30 individuals were sampled, with a geometric mean number of cigarettes smoked per day of $\overline{G} = 29.1$. Taking k = 20, a = 1, b = 1, write down the functional form of $p(\theta | z_1, \ldots, z_n)$ and obtain the value of $E(\theta | z_1, \ldots, z_n)$.

(e) A new treatment has been developed to help regular smokers give up smoking. A physician wants to decide whether to prescribe this new treatment (decision d = 1) or continue to prescribe an existing treatment (decision d = 2) to her patients. The loss function associated with each decision is a trade-off between the cost of the treatment and how effective it is at helping people to stop smoking, where the latter depends on the number of cigarettes currently smoked by the patient. Assuming that this trade-off can be represented by the following loss function,

$$L(d = 1, \theta) = 200 - \theta^5$$

$$L(d = 2, \theta) = 100 - \theta^4$$

use the posterior distribution for θ obtained above to determine which is the optimal decision for the physician to make. (You may find it helpful to note that $\Gamma(x) = (x-1)!$ for integer values of x).

2. A national airline operates a subsidiary budget airline. Over a 5-year period, it records the following data from the budget airline:

	Year, i					
	1	2	3	4	5	Total
Total number of flights, t_i (in 100's)	32	40	65	80	80	297
Number of flights delayed > 6 hours, y_i	24	31	48	52	58	213

These data may be modeled as $p(y_i|\theta) = \text{Poisson}(\theta t_i)$, where θ is the rate (per 100 flights) at which major delays (> 6 hours) occur.

(a) Briefly discuss the pros and cons of Jeffreys' prior for representing prior ignorance, and derive this prior for θ.

(b) Over the past 30 years, the average annual rate of major (> 6 hours) delays for flights operated by the parent airline was 0.6 per 100 flights, with standard deviation 0.05 per 100 flights.

i. Use this information to derive an informative $\text{Gamma}(\alpha, \beta)$ prior for θ, the delay rate for the budget airline.

ii. Using the fact that the Gamma prior is conjugate to the Poisson likelihood (or otherwise), write down an expression for the posterior mean of θ as a function of the data (y_i, t_i) and the prior parameters (α, β). Verify that this can be expressed as a weighted average of the data (MLE) and the prior mean:

$$E(\theta|y_1,\ldots,y_n,t_1,\ldots,t_n) = w\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n t_i} + (1-w)\frac{\alpha}{\beta}, \text{ where } w = \frac{\sum_{i=1}^n t_i}{\beta + \sum_{i=1}^n t_i}$$

(c) For each of the priors (Jeffreys' in part a) and (Gamma in part b, i), evaluate the posterior means of θ and explain how you would use a statistical package such as R to calculate the posterior probabilities that the delay rate for the budget airline is greater than the mean delay rate (0.6 per 100 flight) for the parent airline.

(d) Imagine that you carried out the above analysis after being commissioned by the airline to analyze the performance of its budget subsidiary. One of the airline managers knows a little about statistics and argues that you should have carried out a frequentist (non-Bayesian) analysis of the data. He argues that if you assume a non-informative prior, then there is no need to bother with a Bayesian approach, since the inference is just based on the data. He also argues that if you assume an informative prior, then why bother to collect new data since the inference is dominated by the prior? Briefly discuss how you would counter these arguments (you may find it helpful to use some of your results from (a) – (c) to illustrate your points).

3. Consider the following Bayesian forecasting model for sequential data: Assume that at each discrete time step $t \in \{1,2,\ldots\}$ a scalar observation y_t is made in the presence of additive noise v_t:

$$y_t = x_t + v_t, \qquad v_t \sim \text{Normal}\left(0, \tau_v^{-1}\right).$$

The underlying process of interest, x, is believed to evolve according to

$$x_t = x_{t-1} + u_t, \qquad u_t \sim \text{Normal}\left(0, \tau_u^{-1}\tau_v^{-1}\right).$$

We assume that the precision τ_u is known, the observation precision τ_v is unknown, and u and v are conditionally independent given τ_v.

(a) Suppose you have some prior knowledge about the initial value x_0 of the process x, conditional on τ_v. Verify that $p(x_1 | x_0, \tau_v)$ belongs to the one-parameter exponential family. Use this result to show that the appropriate conjugate form for $p(x_0 | \tau_v)$ is a Normal distribution and state its parameters.

(b) Assuming that $p(x_0 | \tau_v)$ has mean μ_0 and precision $(\phi_0 \tau_v)$, where μ_0 and ϕ_0 are known constants, state the form of the distribution $p(x_1 | \tau_v)$ and its parameters. (You might or might not find it helpful to use properties of conditional expectations and variances).

(c) Comment on the choice of a $\text{Gamma}(\alpha_v, \beta_v)$ distribution for τ_v as an expression of prior uncertainty. In your answer, refer both to its functional form relative to the likelihood $p(y_t | x_t, \tau_v)$ and the influence of hyperparameter values α_v and β_v on posterior estimation of τ_v as the number of observations increases. What prior is obtained in the limit as $\alpha_v, \beta_v \to 0$?

(d) Assume that the first data point y_1 has been observed. Outline how, given the distributions specified above, you would calculate the posterior distribution $p(x_1 | y_1)$ up to a constant of proportionality. (You need not compute any resultant integrals.) What has become of the precision term τ_v?

(e) Suppose $\tau_u \to \infty$, so that x_t is constant for all $t \in \{1,2,\ldots\}$, and hence inference need only concern x_1. How can the expression $p(x_1 | y_1)$ be used to update the posterior in light of a new observation y_2? Write down a general recursive expression for updating the posterior distribution of x_1, given data $(y_1, y_2, \ldots, y_n)$.

4. This table shows the data from the bioassay example we discussed in class on Friday, October 26. Briefly, the table lists the number of animals that died after being exposed to some chemical. The logarithm of the drug dose is also in the table.

Log Dose (log g/ml)	No. Animals(n_i)	No. Deaths(y_i)
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5

In class, we fit a logistic regression model to relate the risk of death with log dose. An alternative model for such data analysis is the probit model. Probit models relate the probability of the event (death) with the cumulative distribution function of a normal distribution. That is,

$$\Pr\left(Y_i = 1 \middle| X_i = x_i\right) = \int_{-\infty}^{\alpha+\beta x_i} \varphi(u)\,du = \Phi(\alpha + \beta x_i)$$

is the probit model. (Probit models are often close to logistic models.)

We have historical data that allows us to carry out inference with a mildly informative prior. The prior information characterizes our uncertainty about the parameter vector as follows.

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \sim N\left[\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix}, \Sigma_0\right], \quad \begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix} = \begin{pmatrix}-17.31\\ 2.57\end{pmatrix} \quad \text{and} \quad \Sigma_0 = \begin{bmatrix}1053.72 & -156.45\\ -156.45 & 23.24\end{bmatrix}$$

a) Generate a random sample of 1,000 from the joint posterior distribution of α and β.

b) Plot the marginal posterior densities of α and β.

c) Plot the posterior density of the log(LD_{50}). The LD_{50} is the dose that is lethal to 50% of the animals. (Note: if the parameter β is negative, the LD_{50} takes on a different meaning, so be careful in your analysis.)

d) Plot the posterior density of the LD_{50} (<u>not</u> the logarithm of the LD_{50}).