MATH 640: Exam 2

Name:____

Instructions

1. Do not discuss this midterm with anyone other than Professor Meyer or the TA; and you may only ask clarifying questions.
2. Your responses to this exam *must* be typed.
3. This cover sheet *must* be the first page of your submission.
4. If you wish to cite a result or derivation from lecture, you may do so but be sure it is clearly cited (list the slide from the Note Set or the example) and relevant.
5. Please submit the exam to the assignment page on Canvas by **11:59pm on Sunday, May 2**.
6. Late submissions will be accepted up to 24 hours after the deadline, however they will be penalized: 4 points off for every six hours the submission is late.

Portion	Question	Points	Score
Theory	1	20	
	2	20	
Computing	1	20	
	2	20	
	3	20	
Total		100	

Theory

1. We wish to model a sequence of n iid strictly positive data points, s_i, using the inverse-Gaussian distribution which has the form

$$p(s_i) = \left(\frac{\lambda}{2\pi s_i^3}\right)^{1/2} \exp\left[-\frac{\lambda(s_i-\mu)^2}{2\mu^2 s_i}\right].$$

The inverse-Gaussian is parameterized by a scale, λ, and mean, μ. Use this to answer the following.

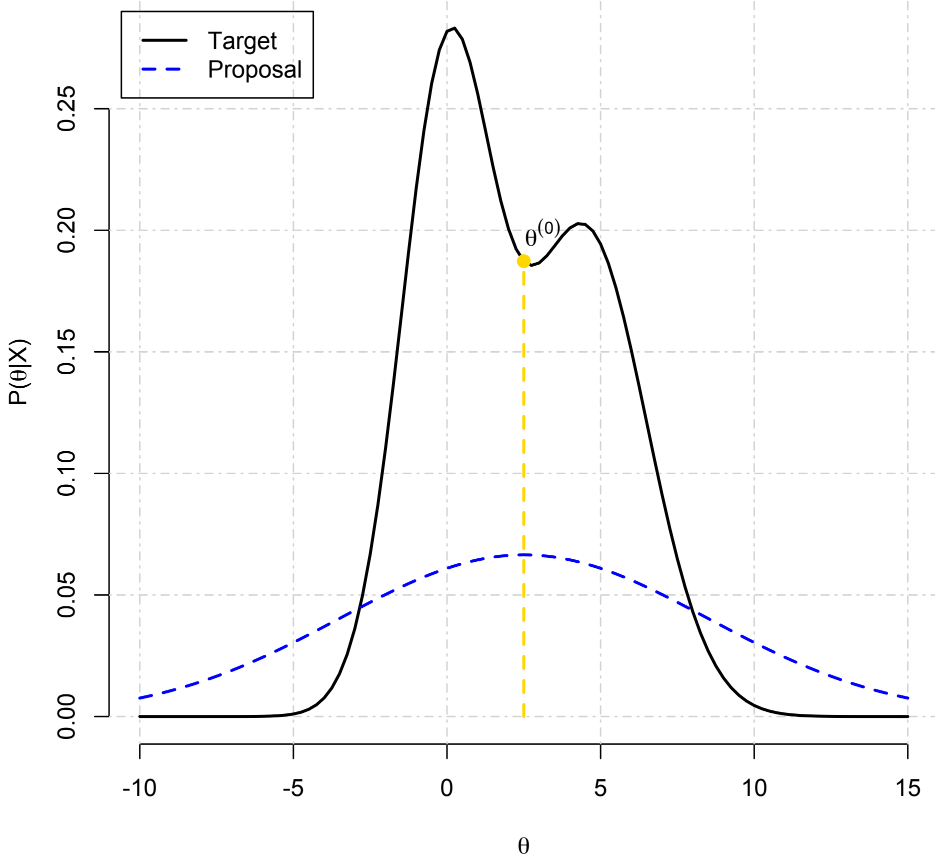
   (a) (10 points) Determine the likelihood and find a conditionally conjugate prior for λ. State the resulting conditional posterior distribution for λ.

   (b) (10 points) Assume a flat prior for $\pi(\mu) \propto 1$ (independent of λ) and define a transformation on μ such that the conditional is recognizable. State the conditional posterior distribution for μ.

2. Let $p(\theta|X)$ denote a target density of interest and let $J\left(\theta^*|\theta^{(b-1)}\right)$ be a jumping density. Use this to answer the following.

   (a) (10 points) Compare the Metropolis Algorithm to the Rejection Sampler. Under what conditions is it appropriate to use each?

   (b) (10 points) Using the illustration below, describe how the Metropolis Algorithm draws works given the sampled proposal values, in order, $\theta^* = -10, 5, 0, 10$. That is, -10 is the first proposed value, 5 is second, and so on. If the starting value is $\theta^{(0)} = 2.5$, state the possible values of $\theta^{(b)}$ in each of the first four steps, $b = 1, 2, 3, 4$. Which combination of draws is most likely? Explain.



Note: A pdf file of this figure is included alongside the exam.

Computing

1. (20 points) The REIGN dataset calculates the monthly risk of a coup occurring for each country in the world. We are interested in modeling the log transformed risk of coup from the month of December in the year 1980. The data is roughly symmetric and unimodal, however it is unclear whether it is normal. One model we can use to help determine if it is normal is the generalized normal which has as its pdf

$$p(z_i) = \frac{\beta}{2\Gamma(1/\beta)} \exp\left(-|z_i|^\beta\right). \tag{1}$$

If $\beta \approx 2$, the model suggests a Gaussian likelihood and if $\beta \approx 1$, it suggests the Laplacian. Thus, we can use the generalized normal model as one a way to determine which of these two likelihoods is a better fit for the data. (**Hint 1**: you will write two samplers, one to determine the form of the likelihood using (1) and then one to draw samples from your chosen likelihood, Laplace or Normal.)

Derive the likelihood and posterior assuming a flat prior for β, $\pi(\beta) \propto 1$. Then implement a sampler of your choosing to generate posterior samples for β. The data for this model must be standardized first (i.e. (1) assumes the center of the data is 0 and scale is 1, so if x_i is log-risk of coup, then the standardized data is $z_i = [x_i - \bar{x}]/\text{sd}[x]$). Generate 8,000 retained samples (i.e. post-burnin, thinned samples) with your sampler, setting the first seed to 37 (see Hint 2 for more detail). Based on your samples of β, determine which of the two likelihoods, Laplace or Normal, matches the data best (you may need to round). (**Hint 2**: You may use ether a rejection sampler or M-H to draw values of β for the first sampler. If you chose to use a Metropolis or Metropolis-Hastings, run four separate chains each resulting in 2000 post-burnin and thinned samples by setting the first starting value to $\beta^{(1)} = 1.25$, the second starting value and seed to $\beta^{(1)} = 1$ and 181, the third starting value and seed to $\beta^{(1)} = 1.5$ and 142, and the fourth starting value and seed to $\beta^{(1)} = 2$ and 1004. Check convergence using all post-burnin, *un*thinned chains. Then combine across chains to obtain 8,000 samples for generating posterior summaries of β.)

Next, return to the original data, i.e. unstandardized data x_i, and model it with your chosen likelihood, Laplace or Normal. In either case, derive the posterior and full conditionals. Then use a Gibbs Sampler to draw posterior estimates. Summarize the results for all model parameters in the usual fashion. Data with the log-transformed risk of coup in each country during the month of December, 1980 can be found in the file `coup1280.txt`. (**Hint 3**: The Gibbs Sampler should not include an M-H step, regardless of chosen likelihood. **Hint 4**: the `R` package `statmod` contains a inverse-Gaussian sampler, `rinvgauss`.)

2. (20 points) The Veteran's Administration conducted a study of time to death in veterans with various types of lung cancer. In total, 137 veterans were part of the study with lung cancer types including small cell, adenocarcinoma, squamous cell, and large cell. We are interested in building a parametric model for the survivor function of the 27 vets with large cell lung cancer. The Weibull distribution is commonly used as a parametric model for survivor functions. The form of the Weibull distribution is

$$p(t_i) = \frac{\theta}{\lambda^\theta} t_i^{\theta-1} \exp\left[-\left(\frac{t_i}{\lambda}\right)^\theta\right] \text{ for } t_i > 0 \text{ and } \lambda, \theta > 0.$$

Determine the likelihood and, using the non-informative joint prior of $\pi(\lambda, \theta) \propto \left(\lambda^\theta\right)^{-1}$, find the posterior and full conditionals. Write a Gibbs-MH sampler to implement your model taking $B = 50000$ total samples and, post-burnin, determine an appropriate level of thinning. For the first run, use the starting values of $\theta^{(1)} = 0.1$ and $\lambda^{(1)} = 1$ and set the seed to 121—use this run to tune your acceptance rate. Repeat this step three more times using the starting values of $\theta^{(1)} = 0.2$ and $\lambda^{(1)} = 3$ with seed 75, $\theta^{(1)} = 0.15$ and $\lambda^{(1)} = 0.5$ with seed 340, and $\theta^{(1)} = 0.05$ and $\lambda^{(1)} = 1.5$ with seed 19. Using the full post-burnin chains from each of the four runs, assess convergence for all model parameters with the Gelman-Rubin diagnostic. Then thin the chains based on your selected level of thinning and combine to determine posterior summaries of λ and θ. (**Hint 1**: one of the full conditionals will be recognizable but requires a transformation, the other full conditional will require an M-H step. Use a *Gamma* for your proposal, vary the proposal parameters in increments of 0.1 from $\alpha = 4$ to 5 and $\beta = 3$ to 4 to obtain an ideal acceptance rate.)

Once you have the posterior samples, estimate the survivor function using the following relationship:

$$S(t) = 1 - F(t) = 1 - \int_0^t p(a)da.$$

First, find the CDF of the Weibull. Next, using your posterior samples of θ and λ, generate the posterior distribution of survivor functions. Since $S(t)$ is a function of t, you will need to assume a grid of possible survival times starting. Set your grid to begin at 1 and go to the largest observed survival time in the data by 1. The survivor function, $S(t)$, is a function of your posterior samples and your grid t, no new model needs to be fit. Summarize your results graphically with a plot of the median survivor function and credible interval for the survivor function. The data for this problem is in the file `valc.txt`. Survival is measured in days in this dataset. (**Hint 2:** The survivor function is a function of t and will be a curve when plotted. To plot it, you must construct, and evaluate it on, a grid starting at 1 and going to the max of the data. The `R` function `seq()` is useful for constructing grids. **Hint 3:** The credible intervals are also functions of t and will also be curves when plotted.)

3. (20 points) Data on the physiochemical properties of Vinho Verde, a Portuguese wine that is typically white, is in the file `vinhoverde.txt`. It contains data on a wide variety of properties recorded on nearly 5,000 different bottles. We want to build a logistic regression model to predict whether or not the wine was rated as a quality wine (`quality`). The covariates of interest are pH level (`pH`), sulphate level (`sulphates`), and alcohol by volume (`alcohol`). Using the M-H sampler for GLMs, first run a model using the flat prior, $\pi(\boldsymbol{\beta}) \propto 1$. Set your seed to 870 and retain 4000 total samples, after burnin and thinning. Generate 90% credible intervals, the posterior probability that each coefficient is larger than 0, and the WAIC.

   Since the logistic model estimates coefficients on the log odds scale, one could argue that coefficients in excess of nine (in absolute value) are highly unlikely (9 would correspond to an odds ratio > 8000 and -9 to an odds ratio < 0.0002). Rerun the model using a multivariate normal prior on $\boldsymbol{\beta}$ with mean equal to $\mathbf{0}$ and covariance equal to a 4×4 identity matrix multiplied by 9 (a variance of 9 for each coefficient means a standard deviation of 3). Set your seed to 1908 and obtain 4000 total samples, after burnin and thinning. Generate 90% credible intervals, the posterior probability that each coefficient is larger than 0, and the WAIC.

   Compare and contrast the posterior results from each model. Do the coefficients change and if so, how do they change? Which model generates tighter credible intervals and smaller posterior probabilities? Which model has the smaller WAIC? Using these results, suggest a preferred model: using the flat prior or using the multivariate normal prior. (**Hint 1**: the M-H sampler from class may need to be altered to avoid computational underflows, consider building posterior on log scale. **Hint 2**: the multivariate normal prior need only be specified up to a constant of proportionality.)