MATH 640: Bayesian Statistics

Homework 4, due Sunday, March 17

Please submit a PDF or .doc version of your homework to Canvas by 11:59pm on the due date. Please type *all* responses. You are encouraged to use `R` for all calculations.

Theoretical Exercises

1. Suppose $x_i \sim Poisson(\lambda)$ for $i = 1, \ldots, n$. We wish to build a hierarchical model for our Poisson data. Begin by using the conjugate prior for $\lambda \sim Gamma(\alpha, \beta)$ where α and β are unknown. To account for this uncertainty, assume the conjugate prior for α and β which, if you recall from Homework 3, is

$$\pi(\alpha, \beta) \propto \frac{\beta^{\alpha s}}{\Gamma(\alpha)^r} p^{\alpha-1} e^{-\beta q}.$$

   Determine the full posterior, the conditional posterior of $\lambda|\alpha, \beta, X$, and the marginal posterior of $\alpha, \beta|X$. If any of these are recognizable, be sure to name them.

2. Consider a hierarchical model for the univariate normal model with unknown mean and known variance. Let $y_i \sim N(\mu, \sigma_0^2)$ for $i = 1, \ldots, n$. Place a conjugate prior for μ with unknown mean θ and known variance τ_0^2, thus $\mu|\theta \sim N(\theta, \tau_0^2)$ and let the hyperprior be $\theta \sim N(0, \gamma_0^2)$. Find the full posterior, the conditional posterior of $\mu|\theta, y$, and the marginal posterior of $\theta|y$.

3. Consider a multi-center study where we're interested in a multinomial variable, i.e. $\mathbf{z}_j \sim Multinom(n_j; \theta_{1j}, \theta_{2j}, \theta_{3j}, \theta_{4j})$ where $\mathbf{z}_j = [\ z_{1j} \quad z_{2j} \quad z_{3j} \quad z_{4j}\ ]'$, the counts for each level at center $j = 1, \ldots, J$. Let $\boldsymbol{\theta}_j = [\ \theta_{1j} \quad \theta_{2j} \quad \theta_{3j} \quad \theta_{4j}\ ]'$ and assume that the $\boldsymbol{\theta}_j$ are iid samples from a common Dirichlet distribution, thus $\boldsymbol{\theta}_j \sim Dir(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Further, let the hyperprior on $\boldsymbol{\alpha} = [\ \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4\ ]'$ be a constant, i.e. $\pi(\boldsymbol{\alpha}) \propto 1$. Further, let $\boldsymbol{\Theta} = [\ \boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad \ldots \quad \boldsymbol{\theta}_J\ ]$ and let $\mathbf{Z} = [\ \mathbf{z}_1 \quad \mathbf{z}_2 \quad \cdots \quad \mathbf{z}_J\ ]$. Find $\mathcal{L}(\mathbf{Z}|\boldsymbol{\Theta})$, $\pi(\boldsymbol{\Theta}|\boldsymbol{\alpha})$, as well as the full posterior, conditional posterior of $\boldsymbol{\Theta}|\boldsymbol{\alpha}, \mathbf{Z}$, and the marginal posterior of $\boldsymbol{\alpha}|\mathbf{Z}$.

Analysis Exercises

1. Consider the intercept-only logistic regression model where $y_i \sim Bern(\theta_i)$ for $\text{logit}(\theta_i) = \beta$. The likelihood then has the form

$$\mathcal{L}(y_i|\beta) \propto \exp\left\{\sum_{i=1}^{n} \left[y_i\beta - \log\left(1 + e^{\beta}\right)\right]\right\}.$$

   Using the flat prior $\pi(\beta) \propto 1$, fully derive out a normal approximation to the posterior of $P(\beta|y_i)$. Using data on whether or not forest fires between January 2000 and December 2003 from the Montesinho natural park the northeast region of Portugal consumed at least 1 hectare of land, `forestfire.txt`, find 10000 posterior samples for β. Assess the convergence of your samples both visually and using Geweke's diagnostic. Interpret your results on a meaningful scale. Next find Jeffreys' prior for β and use it to derive another normal approximation to the corresponding posterior. Apply this model to the same dataset, assess for convergence as before. Based on the inferential and convergence results of your two approaches, is the model sensitive to the choice of non-informative priors? Set the seed of 821 for each run.

2. Using the daily Capital Bike Share data, `bikeshare.txt`, fit a Poisson regression model first with the casual users count as the outcome and then with the registered users count as the outcome. Using Bayesian inference, determine the best set of predictors for each outcome. For each final model, check the convergence of all model parameters using visual inspection as well as Geweke's diagnostic. Also provide a brief interpretation of each parameter. Do the models differ substantially from each other? Data variable names are provided alongside this assignment, `bikeshareDataDictionary.rtf`. For each model, set the seed to 1959.