STK 4021 - Applied Bayesian Analysis and Numerical Methods
Thordis L. Thorarinsdottir
Fall 2014

Problem Set 2

Problem 1 (Galenshore distribution). An unknown quantity Y has a Galenshore(a, θ) distribution if its density is given by

$$p(y|\theta) = \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for $y > 0$, $\theta > 0$, and $a > 0$. Assume for now that a is known. For this density,

$$\mathbb{E}[Y|\theta] = \frac{\Gamma(a + 1/2)}{\theta\Gamma(a)}, \quad \mathbb{E}[Y^2|\theta] = \frac{a}{\theta^2}.$$

(a) Identify a class of conjugate prior densities for θ. Plot a few members of this class of densities.

(b) Let $Y_1, \ldots, Y_n \sim$ Galenshore(a, θ) be conditionally i.i.d. Find the posterior distribution of θ given $\mathcal{D} = \{Y_1, \ldots, Y_n\}$, using a prior from your conjugate class.

(c) Write down $p(\theta_a|\mathcal{D})/p(\theta_b|\mathcal{D})$ and simplify. Identify a sufficient statistic.

(d) Determine $\mathbb{E}[\theta|\mathcal{D}]$.

(e) Determine the form of the posterior predictive density $p(\tilde{y}|\mathcal{D})$.

Problem 2 (unit information prior). Let $X_1, \ldots, X_n \sim p(x|\theta)$ be conditionally i.i.d. For observations $\mathcal{D} = (x_1, \ldots, x_n)$, the log likelihood is given by $l(\theta|\mathcal{D}) = \sum \log p(x_i|\theta)$, and we denote by $\hat{\theta}$ the maximum likelihood estimate (MLE). The Fisher information, $J(\theta) = -\partial^2 l(\theta|\mathcal{D})/\partial\theta^2$, describes the precision of the MLE $\hat{\theta}$. For situations in which it is difficult to quantify prior information in terms of a probability distribution, some have suggested that the "prior" distribution be based on the likelihood, for example, by centering the prior distribution around the MLE $\hat{\theta}$. To deal with the fact that the MLE is not really prior information, the curvature of the prior is chosen so that it has only "one nth" as much information as the likelihood, so that $-\partial^2 \log p(\theta)/\partial\theta^2 = J(\theta)/n$. Such a prior is called *unit information prior*, as it has as much information as the average amount of information from a single observation. The unit information prior is not really a prior distribution, as it is computed from the observed data. However, is can be roughly viewed as the prior information of someone with weak but accurate prior information.

(a) Let $X_1, \ldots, X_n \sim$ Bernoulli(θ) be conditionally i.i.d. Obtain the MLE $\hat{\theta}$ and $J(\hat{\theta})/n$.

(b) Find a probability density $p_U(\theta)$ such that $\log p_U(\theta) = l(\theta|\mathcal{D})/n + c$, where c is a constant that does not depend on θ. Compute the information $-\partial^2 \log p(\theta)/\partial\theta^2$ of this density.

(c) Obtain a probability density for θ that is proportional to $p_U(\theta) \times p(\mathcal{D}|\theta)$. Can this be considered a posterior distribution for θ?

Problem 3 (Poisson population comparison). Let θ_A and θ_B be the average number of children of men in their 30s with and without bachelor's degrees, respectively. Such data is given in the files `menchild30bach.dat` and `menchild30nobach.dat` which are available on the course website. We'll assume Poisson sampling model for the two groups, with the parameterization $\theta_A = \theta$ and $\theta_B = \theta \times \gamma$. In this parameterization, γ represents the relative rate θ_B/θ_A. Let $\theta \sim \Gamma(a_\theta, b_\theta)$ and let $\gamma \sim \Gamma(a_\gamma, b_\gamma)$.

(a) Obtain the form of the full conditional distribution of θ given $\mathcal{D}_A$, $\mathcal{D}_B$, and γ.

(b) Obtain the form of the full conditional distribution of γ given $\mathcal{D}_A$, $\mathcal{D}_B$, and θ.

(c) Set $a_\theta = 2$ and $b_\theta = 1$. Let $a_\gamma = b_\gamma \in \{8, 16, 32, 64, 128\}$. For each of these five values, run a Gibbs sampler of at least 5,000 iterations and obtain $\mathbb{E}[\theta_B - \theta_A|\mathcal{D}_A, \mathcal{D}_B]$. Describe the effects of the prior distribution for γ on the results.

Solutions will be discussed in class on September 12.