Bayesian Theory and Computation, Problem Set 1

March 05, 2025 Due 03/19/2025

Problem 1.

Paleobotanists estimate the moment in the remote past when a given species became extinct by taking cylindrical, vertical core samples well below the earth's surface and looking for the last occurrence of the species in the fossil record, measured in meters above the point P at which the species was known to have first emerged. Letting $\{y_i, i = 1, \ldots, n\}$ denote a sample of such distances above P at a random set of locations, the model

$$y_i|\theta \sim \text{Uniform}(0, \theta)$$

emerges from simple and plausible assumptions. In this model the unknown $\theta > 0$ can be used, through carbon dating, to estimate the species extinction time. This problem is about Bayesian inference for θ, and it will be seen that some of our usual intuitions do not quite hold in this case.

(1) Show that the likelihood may be written as

$$\ell(\theta|y) = \theta^{-n} I(\theta \geq \max(y_1, \ldots, y_n)),$$

where $I(\cdot)$ is the indicator function. [5pts]

(2) The Pareto distribution $\theta \sim \text{Pareto}(\alpha, \beta)$ has density:

$$p(\theta) = \begin{cases} \alpha\beta^\alpha\theta^{-(\alpha+1)} & \text{if } \theta \geq \beta \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha, \beta > 0$. With the likelihood viewed as a constant multiple of a density for θ, show that the likelihood corresponds to the Pareto$(n-1, m)$ distribution. Now let the prior for θ be taken to be Pareto(α, β) and derive the posterior distribution $p(\theta|y)$. Is the Pareto conjugate to the uniform? [5pts]

(3) In an experiment conducted in the Antarctic in the 1980's to study a particular species of fossil ammonite, the following was a linearly rescaled version of the data obtained, in ascending order: $y = (0.4, 1.0, 1.5, 1.7, 2.0, 2.1, 3.1, 3.7, 4.3, 4.9)$. Prior information equivalent to a Pareto prior with $(\alpha, \beta) = (2.5, 4)$ was available. Plot the prior, likelihood, and posterior distributions arising from this data set on the same graph, and briefly discuss what this picture implies about the updating of information from prior to posterior in this case. [5pts]

(4) Make a table summarizing the mean and standard deviation for the prior, likelihood and posterior distributions, using the (α, β) choices and the data in part (3) above. In Bayesian updating the posterior mean is often a weighted average of the prior mean and the likelihood mean (with positive weights), and the posterior standard deviation

is typically smaller than either the prior or likelihood standard deviations. Are each of these behaviors true in this case? Explain briefly. [10pts]

Problem 2.
Consider the following Markov chain taking values on the nonnegative integers. The one-step transition probabilities are given by $(i, j \geq 0)$

$$\Pr(X_{t+1} = j | X_t = i) \equiv p_{ij} = \begin{cases} \frac{1}{2j} & j = i+1, i \geq 0 \\ \frac{j}{2(j+1)} & j = i \geq 1 \\ \frac{1}{2} & j = i = 0 \\ \frac{1}{2} & j = i-1, i \geq 1 \\ 0 & |j-i| > 1 \end{cases}$$

(1) Verify that these transition probabilities are legitimate. [5pts]

(2) Argue that this Markov chain is irreducible. [5pts]

(3) Is this chain aperiodic? Why or why not? [5pts]

(4) Verify that the Poisson(1) distribution is a stationary distribution. [5pts]

(5) Starting with $X_0 = 1$, simulate this Markov chain for $10,000$ iterations. Compute the mean and variance of these $10,000$ draws and compare with the theoretical mean and variance of the stationary distribution. [10pts]

Problem 3.
Consider the following model

$$\begin{aligned} \phi_1 | \mu, \sigma^2 &\sim \mathcal{N}(\mu, \sigma^2/(1-\rho^2)), \\ \phi_{j+1} | \phi_1, \ldots, \phi_j, \mu, \sigma^2 &\sim \mathcal{N}(\mu + \rho(\phi_j - \mu), \sigma^2), \quad j = 1, \ldots, J-1 \\ y_j | \phi_1, \ldots, \phi_J, \tau^2 &\sim \mathcal{N}(\phi_j, \tau^2), \quad j = 1, \ldots, J, \end{aligned}$$

where μ, σ^2, τ^2 are unknown parameters of interest, ρ is a known constant between -1 and 1, $\phi = (\phi_1, \ldots, \phi_J)$ is a sequence of latent variables, and $y = (y_1, \ldots, y_J)$ is the observed data. (This is known as a *Gaussian state-space model.*)

(1) Assume the prior $p(\mu, \sigma^2, \tau^2) \propto \tau^{-4} e^{-1/(2\tau^2)}$. That is, the prior on (μ, σ^2) is a non-informative constant prior, and independently, the prior on $1/\tau^2$ is an exponential distribution (check this). Write down the joint posterior of all parameters and latent variables given the observed data y. [5pts]

(2) Derive the full conditionals. That is, compute the distribution of each of $\mu, \sigma^2, \tau^2, \phi_1, \ldots, \phi_J$, given all others and the observed data y. [15pts]

(3) Implement a Gibbs sampler for simulating from the joint posterior. [10pts]

(4) Simulate a data set with $J = 100, \rho = 0.8, \mu = 0, \tau^2 = \sigma^2 = 1$. Use you Gibbs sampler to simulate from the posterior. Comment on its convergence after examining

the trajectories of the draws of μ, σ^2, τ^2 and the autocorrelations. [10pts]

(5) Based on the Gibbs draws, plot the marginal posteriors of μ, σ^2 and τ^2 and compare with the true values. Do the true values lie within or outside of their respective 95% credible intervals? [5pts]