Bayesian Theory and Computation, Problem Set 2

April 20, 2021 Due 05/04/2021

Problem 1.

In a linear regression model the n-vector of responses y has distribution $y|\beta \sim \mathcal{N}(X\beta, I_n)$, with mean response vector $\mu = \mathbb{E}(y|\beta) = X\beta$ and identity variance matrix, where X is the $n \times p$ design matrix of rank p and β is the p-vector of regression coefficients. Suppose that the prior for β is $\beta \sim \mathcal{N}(0, g^{-1}(X^T X)^{-1})$ for some number $g > 0$ (this is known as g-prior).

(1) What is the posterior distribution of $\beta|y$?

(2) Show that the posterior mean $\mathbb{E}(\beta|y)$ can be expressed as a function of $\hat{\beta}$, the usual MLE of β.

(3) What is the posterior mean and posterior variance matrix of μ?

(4) Consider the special case of an orthogonal design, so that $X^T X = I_p$. Denote by μ_i the ith element of μ. Under the posterior $p(\mu|y)$, are μ_j and μ_k independent for $j \neq k$?

Problem 2.

The following table gives the worldwide number of fatal accidents and deaths on scheduled airline flights per year over a ten year period (from Table 2.2 in Gelman, et al.). Death rate is passenger deaths per 100 million passengers miles.

Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.13
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

Assume that the number of fatal accidents in each year, Y_i, follow independent Poisson distribution with intensity linear in the number of passenger miles flown, X_i. In other words, $Y_i \sim \text{Poisson}(\lambda_i)$ with $\lambda_i = \alpha + \beta X_i$. (You can approximate the number of passenger miles flown by dividing the appropriate columns of the table and by ignoring round-off errors.) Consider a Bayesian solution to this Poisson regression problem, using constant priors on α, β subject to the natural constraint that $\alpha + \beta X_i \geq 0$ for all i.

(1) Choose a suitable fine grid, and evaluate the posterior means of α and β by writing these as suitable integrals and then approximating such integrals with finite sums. Note that the joint posterior is not normalized.

(2) Compute the MLE and observed Fisher information matrix. Based on your computation, construct a normal approximation to the posterior distribution. Report the posterior means and 95% credible intervals for α and β based on this normal approximation. Compare the posterior mean estimates with those in (1).

(3) The normal has thin tails. Consider a bivariate t distribution with four degrees of freedom to approximate the posterior. Generate 10000 draws from this bivariate t distribution and report posterior means and 95% credible intervals based on these draws. Compare with (2).

(4) We can refine the approximation in (3) by using importance sampling. Describe how importance sampling works (using the bivariate t as the proposal density) in this context. Compare the estimated posterior means and 95% intervals with those in (2) and (3).

Problem 3.

Consider the following model

$$z_i|\mu_i \sim \mathcal{N}(\mu_i, 1), \quad \mu_i \sim \mathcal{N}(0, A), \quad i = 1, \ldots, N$$

The *James-Stein estimator* is defined to be

$$\hat{\mu}^{\mathrm{JS}} = \left(1 - \frac{N-2}{S}\right) z, \quad S = \|z\|^2 = \sum_{i=1}^{N} z_i^2$$

(1) What is the Bayes estimator $\hat{\mu}^{\mathrm{Bayes}}$ for $\mu = (\mu_1, \ldots, \mu_N)$ when the square error lost is used? And what is the MLE $\hat{\mu}^{\mathrm{MLE}}$?

(2) Compute the Bayes risk for $\hat{\mu}^{\mathrm{JS}}$, $\hat{\mu}^{\mathrm{Bayes}}$ and $\hat{\mu}^{\mathrm{MLE}}$.

(3) Simulate data sets with different μ and N. Estimate the frequentist risk of $\hat{\mu}^{\mathrm{JS}}$

$$\mathbb{E}_\mu \|\hat{\mu}^{\mathrm{JS}} - \mu\|^2$$

How does it compare with $\mathbb{E}_\mu \|\hat{\mu}^{\mathrm{MLE}} - \mu\|^2$? Report your findings.

Problem 4.

Consider the following nonlinear reference model

$$
\begin{aligned}
X_t &= \frac{1}{2}X_{t-1} + \frac{25X_{t-1}}{1+X_{t-1}^2} + 8\cos(1.2t) + \xi_t \\
Y_t &= \frac{X_t^2}{20} + \eta_t
\end{aligned}
\tag{1}
$$

where $X_0 \sim \mathcal{N}(0,5)$, ξ_t and η_t are mutually independent white Gaussian noises, $\xi_t \sim \mathcal{N}(0,\sigma_\xi^2)$ and $\eta_t \sim \mathcal{N}(0,\sigma_\eta^2)$ with $\sigma_\xi^2 = 10$ and $\sigma_\eta^2 = 1$. Since the observation Y_t is no longer a linear function of X_t, it is not possible to evaluate analytically $p(Y_t|X_{t-1})$ or to sample simply from $p(X_t|X_{t-1},Y_t)$. We propose to approximate it with a locally linearized observation equation.

(1) Let $f(X_{t-1}) = \frac{1}{2}X_{t-1} + \frac{25X_{t-1}}{1+X_{t-1}^2} + 8\cos(1.2t)$. Consider a locally linear approximation of Y_t at $f(X_{t-1})$ as follows

$$
Y_t \approx \frac{f^2(X_{t-1})}{20} + \frac{f(X_{t-1})}{10}(X_t - f(X_{t-1})) + \eta_t
$$

Derive the approximate locally optimal importance distribution $p(X_t|X_{t-1},Y_t)$ using the above linear approximation.

(2) Simulate a data set from the model (1) with $t = 0,\ldots,100$. Implement sequential importance sampling with $N = 100$ particles using the importance distribution derived in (1). Report your estimate of the marginal posterior $p(x_{100}|y_{\leq 100})$ as a histogram. Plot the ESS of the importance weights as a function of time t.

(3) Run another sequential importance sampling with $N = 100$ particles using the naive importance distribution $p(X_t|X_{t-1})$. Report the results and compare them with (2).