Homework 4

ISyE 6420
Fall 2019

> Due October 13, 2019, 11:55pm. HW4 is not time limited except the due date. Late submissions will not be accepted.
>
> Use of all available electronic and printed resources is allowed except direct communication that violates Georgia Tech Academic Integrity Rules.

1. Metropolis: The Bounded Normal Mean. Suppose that we have information that the normal mean θ is bounded between $-m$ and m, for some known number m. In this case it is natural to elicit a prior on θ with the support on interval $[-m, m]$.

A prior with interesting theoretical properties supported on $[-m, m]$ is Bickel-Levit prior[1],

$$\pi(\theta) = \frac{1}{m}\cos^2\left(\frac{\pi\theta}{2m}\right), \quad -m \le \theta \le m.$$

Assume that a sample $[-2, -3, 4, -7, 0, 4]$ is observed from normal distribution

$$f(y|\theta) \propto \sqrt{\tau}\exp\left\{-\frac{\tau}{2}(y-\theta)^2\right\},$$

with a known precision $\tau = 1/4$. Assume also that the prior on θ is Bickel-Levit, with $m = 2$.

This combination likelihood/prior does not result in an explicit posterior (in terms of elementary functions). Construct a Metropolis algorithm that will sample from the posterior of θ.

(a) Simulate 10,000 observations from the posterior, after discarding first 500 observations (burn-in), and plot the histogram of the posterior.

(b) Find Bayes estimator of θ, and 95% equitailed Credible Set based on the simulated observations.

Suggestions:

(i) Take uniform distribution on $[-m, m]$ as a proposal distribution since it is easy to sample from. This is an independence proposal, the proposed θ' does not depend on the current value of the chain, θ.

---

[1] When m is large, this prior is an approximation of the least favorable distribution in a Bayes-Minimax problem with class of priors limited to symmetric and unimodal distributions.

(ii) You will need to calculate $\sum_{i=1}^{n}(y_i - \theta)^2$ for current θ and $\sum_{i=1}^{n}(y_i - \theta')^2$ for the proposed θ', prior to calculating the Metropolis ratio.

Gibbs Sampler and High/Low Protein Diet in Rats. Armitage and Berry (1994, p. 111)[2] report data on the weight gain of 19 female rats between 28 and 84 days after birth. The rats were placed in randomized manner on diets with high (12 animals) and low (7 animals) protein content.

High protein	Low protein
134	70
146	118
104	101
119	85
124	107
161	132
107	94
83	
113	
129	
97	
123	

We want to test the hypothesis on dietary effect: Did a low protein diet result in a significantly lower weight gain?

The classical t test against the one sided alternative will be significant at 5% significance level, but we will not go in there. We will do the test Bayesian way using Gibbs sampler.

Assume that high-protein diet measurements $y_{1i}, i = 1, \dots, 12$ are coming from normal distribution $\mathcal{N}(\theta_1, 1/\tau_1)$, where τ_1 is the precision parameter,

$$f(y_{1i}|\theta_1, \tau_1) \propto \tau_1^{1/2} \exp\left\{-\frac{\tau_1}{2}(y_{1i} - \theta_1)^2\right\}, \; i = 1, \dots, 12.$$

The low-protein diet measurements $y_{2i}, \; i = 1, \dots, 7$ are coming from normal distribution $\mathcal{N}(\theta_2, 1/\tau_2)$,

$$f(y_{2i}|\theta_2, \tau_2) \propto \tau_2^{1/2} \exp\left\{-\frac{\tau_2}{2}(y_{2i} - \theta_2)^2\right\}, \; i = 1, \dots, 7.$$

Assume that θ_1 and θ_2 have normal priors $\mathcal{N}(\theta_{10}, 1/\tau_{10})$ and $\mathcal{N}(\theta_{20}, 1/\tau_{20})$, respectively. Take prior means as $\theta_{10} = \theta_{20} = 110$ (apriori no preference) and precisions as $\tau_{10} = \tau_{20} = 1/100$.

Assume that τ_1 and τ_2 have the gamma $\mathcal{G}a(a_1, b_1)$ and $\mathcal{G}a(a_2, b_2)$ priors with shapes $a_1 = a_2 = 0.01$ and rates $b_1 = b_2 = 4$.

(a) Construct Gibbs sampler that will sample $\theta_1, \tau_1, \theta_2$, and τ_2 from their posteriors.

[2] Armitage, P. and Berry, G. (1994). Statistical Methods in Medical Research (3rd edition). Blackwell

(b) Find sample differences $\theta_1 - \theta_2$. Proportion of positive differences approximates the posterior probability of hypothesis $H_0 : \theta_1 > \theta_2$. What is this proportion if the number of simulations is 10,000, with burn-in of 500?

(c) Using sample quantiles find the 95% equitailed credible set for $\theta_1 - \theta_2$. Does this set contain 0?

Hint: No WinBUGS should be used (except maybe to check your results). Use Octave (MATLAB), or R, or Python here. You may want to consult Handout `GIBBS.pdf` from the course web repository.