STATS 270 Project

Egor Lappo

Metropolis-Hastings

This was the most straightforward approach. I chose the following proposal distributions. For τ, I chose Uniform$(0, 1)$; for μ_i, γ_i, I chose $N(\mu_i, 0.5)$ and $N(\gamma_i, 0.5)$ respectively; for σ^2 I chose $N(\sigma^2, 0.1)$ constrained to $[0, 10]$. These are reasonably efficient in their exploration of the probability space, and I've seen them being used in practice.

My program prints out the posterior means, together with a posterior confidence interval that ranges from 0.05th to 0.95th quantiles. Here is the output:

I also produced histograms for posterior distributions of the parameters.

Hamiltonian Monte-Carlo

I follow the method defined in lecture (and on Wikipedia) verbatim, by solving the Hamiltonian equations using the leapfrog integrator to generate a proposal. I chose the mass matrix to be the identity matrix. The other two parameters I needed to choose were the number of leapfrog steps L, and the time to integrate the trajectory for at each leapfrog step, Δt. If Δt is too big, the algorithm is imprecise, for example, the proposals for τ would be out of the range $[0, 1]$ and they will actualy be accepted often. If L and Δt are too small, there would be no benefit to HMC, the behavior would be as if I have MH sampling with proposals from the multivariate normal (leapfrog trajectories are approximately linear at short timescales).

After fine-tuning, I set $L = 5$, $\Delta t = 0.001$. Of course, in modern algorithms these values are adjusted as the sampling progresses, but since the posterior is not too complex in our case, these parameters work okay.

My program prints out the posterior means, together with a posterior confidence interval that ranges from 0.05th to 0.95th quantiles. Here is the output:

I also produced histograms for posterior distributions of the parameters.

Importance Sampling

In the importance sampling approach is the most free-form, since potentialy I can choose any trial density to generate samples from. However, the variance of samples may be unreasonably high. It may be good to use some information in the data to fine-tune the process. One suggestion was to sample from an "arbitrary" distribution to generate an intermediate posterior, and then sample more from this new distribution.

I, however, used an approach similar to what was described in class. I computed the sample means and variances in the first and second groups to obtain MLE estimates of μ_1, μ_2, γ_1, γ_2, and σ^2. Then, I selected the trial distributions "around" these values. In particular, I sampled μ_1 from $N(-1.5, 1.5)$, μ_2 from $N(-0.5, 1.5)$, γ_1 from $N(-0.3, 1.5)$, and γ_2 from $N(0.3, 1.5)$. I chose to set large variance here to be sure that I get a good sample even if my selection of the mean is poor. For τ, I sample from $\text{Uniform}(0, 1)$, and σ^2 is sampled from $\text{Exp}(12)$, since it has to be nonnegative.

Using IS, I compute posterior means of each of the parameters. Here is the output:

Gibbs Sampling

For Gibbs Sampling, I have used the "linear scan" version of the approach, in which I sample each parameter conditional on "current" values of other parameters in the same order in each iteration.

Discussion

Appendix

The distribution that I want to sample is

$$
\begin{aligned}
p(\theta \mid x) = {} & \frac{1}{\sigma^2} \frac{1}{(2\pi\sigma^2)^N} e^{-\frac{1}{2\sigma^2} \sum_{i_1} (x_{i_1 1} - \mu_1)^2} e^{-\frac{1}{2\sigma^2} \sum_{i_1} (x_{i_1 2} - \mu_2)^2} \\
& \times e^{-\frac{1}{2\sigma^2} \sum_{i_2} (x_{i_2 1} - \gamma_1)^2} e^{-\frac{1}{2\sigma^2} \sum_{i_2} (x_{i_2 2} - \gamma_2)^2} \\
& \times e^{-\frac{1}{2\sigma^2} \sum_{i_3} (x_{i_3 1} - \mu_1/2 - \gamma_1/2)^2} e^{-\frac{1}{2\sigma^2} \sum_{i_3} (x_{i_3 2} - \mu_1/2 - \gamma_2)^2} \\
& \times e^{-\frac{1}{2\sigma^2} \sum_{i_4} (x_{i_4 1} - \tau m u_1 - (1-\tau)\gamma_1)^2} e^{-\frac{1}{2\sigma^2} \sum_{i_4} (x_{i_4 2} - \tau \mu_1 - (1-\tau)\gamma_2)^2},
\end{aligned}
$$

where $\theta = (\sigma^2, \tau, \mu_1, \mu 2, \gamma_1, \gamma_2)$ are the parameters, N is the total number of observations, and i_k indexes data points in kth group.

For the Hamiltonian Monte Carlo, I set $U(\theta) = -\log p(\theta \mid x)$, so that

$$U(\theta) = N \log 2\pi + (N+1) \log \sigma^2$$

$$+ \frac{1}{2\sigma^2} \left[\sum_{i_1} (x_{i_1 1} - \mu_1)^2 + \sum_{i_1} (x_{i_1 2} - \mu_2)^2 \right.$$

$$+ (x_{i_2 1} - \gamma_1)^2 + (x_{i_2 2} - \gamma_2)^2$$

$$+ (x_{i_3 1} - \mu_1/2 - \gamma_1/2)^2 + (x_{i_3 2} - \mu_1/2 - \gamma_2)^2$$

$$\left. + (x_{i_4 1} - \tau m u_1 - (1-\tau)\gamma_1)^2 + (x_{i_4 2} - \tau \mu_1 - (1-\tau)\gamma_2)^2 \right].$$

I compute the partial derivatives of U.

$$\frac{\partial U}{\partial \sigma^2} = \frac{N+1}{\sigma^2} - \sum_{i_1} \frac{(x_{i_1 1} - \mu_1)^2}{2(\sigma^2)^2} - \sum_{i_1} \frac{(x_{i_1 2} - \mu_2)^2}{2(\sigma^2)^2}$$

$$- \sum_{i_2} \frac{(x_{i_2 1} - \gamma_1)^2}{2(\sigma^2)^2} - \sum_{i_2} \frac{(x_{i_2 2} - \gamma_2)^2}{2(\sigma^2)^2}$$

$$- \sum_{i_3} \frac{(x_{i_3 1} - \mu_1/2 - \gamma_1/2)^2}{2(\sigma^2)^2} - \sum_{i_3} \frac{(x_{i_3 2} - \mu_2/2 - \gamma_2/2)^2}{2(\sigma^2)^2}$$

$$- \sum_{i_4} \frac{(x_{i_4 1} - \tau\mu_1 - (1-\tau)\gamma_1)^2}{2(\sigma^2)^2} - \sum_{i_4} \frac{(x_{i_4 2} - \tau\mu_2 - (1-\tau)\gamma_2)^2}{2(\sigma^2)^2}$$

$$\frac{\partial U}{\partial \tau} = \frac{1}{\sigma^2} \sum_{i_4} [(\gamma_1 - \mu_1)(x_{i_4 1} - \tau\mu_1 - (1-\tau)\gamma_1) + (\gamma_2 - \mu_2)(x_{i_4 2} - \tau\mu_2 - (1-\tau)\gamma_2)]$$

$$\frac{\partial U}{\partial \mu_k} = -\sum_{i_1} \frac{(x_{i_1 k} - \mu_k)}{\sigma^2}$$

$$- \sum_{i_3} \frac{(x_{i_3 k} - \mu_k/2 - \gamma_k/2)}{2\sigma^2}$$

$$- \tau \sum_{i_4} \frac{(x_{i_4 k} - \tau\mu_k - (1-\tau)\gamma_k)}{\sigma^2}$$

$$\frac{\partial U}{\partial \gamma_k} = -\sum_{i_2} \frac{(x_{i_2 k} - \gamma_k)}{\sigma^2}$$

$$-\sum_{i_3} \frac{(x_{i_3 k} - \mu_k/2 - \gamma_k/2)}{2\sigma^2}$$

$$-(1-\tau)\sum_{i_4} \frac{(x_{i_4 k} - \tau\mu_k - (1-\tau)\gamma_k)}{\sigma^2}$$

Then, I use this in the leapfrog integrator.

For Gibbs sampling, I compute the conditional distributions. The easiest one is the distribution for σ^2. Since the numerator in the exponential is "fixed" (conditioned upon), I know that

$$p(\sigma^2 \mid \tau, \mu_i, \gamma_i, x) \propto \text{Inv-}\chi^2(\nu),$$

where $\nu = 2N$.

For μ_k, I am working with the density

$$p(\mu_k \mid \tau, \sigma^2, \ldots, \ x) \propto \exp\frac{-1}{2\sigma^2}\left[\sum(x_{i_1 k} - \mu_k)^2 + \sum(x_{i_3 k} - \mu_k/2 - \gamma_k/2)^2 + \sum(x_{i_4 k} - \tau\mu_k - (1-\tau)\gamma_k)^2\right]$$

$$\propto \exp\frac{-1}{2\sigma^2}\left[n_1(\bar{x}_{1k} - \mu_k)^2 + n_3(\bar{x}_{3k} - \mu_k/2 - \gamma_k/2)^2 + n_4(\bar{x}_{4k} - \tau\mu_k - (1-\tau)\gamma_k)^2\right]$$

$$\sim N\left(\frac{n_1\bar{x}_{1k} + n_3(2\bar{x}_{3k} - \gamma_k)/4 + \tau n_4(\bar{x}_{4k} - (1-\tau)\gamma_k)}{n_1 + n_3/4 + \tau^2 n_4}, \frac{\sigma^2}{n_1 + n_3/4 + \tau^2 n_4}\right)$$

where n_j is the size of group j, and $\bar{x}_{jk}$ is the mean of kth variable in group j. In the calculation I omit the constant factors, which in our care are summands in the exponential that do not depend on μ_k. The general fact that I used is that I can complete the square like

$$\exp\sum a_i(b_i - \mu)^2 \rightarrow \exp\left[\left(\sum a_i\right)\left(\mu - \frac{\sum a_i b_i}{\sum a_i}\right)^2\right].$$

The derivation is almost equivalent for γ_i, but there instead I get

$$p(\gamma_k \mid \tau, \sigma^2, \ldots, x)$$
$$\sim N\left(\frac{n_2\bar{x}_2 k + n_3(2\bar{x}_{3k} - \mu_k)/4 + (1-\tau)n_4[\bar{x}_{4k} - \tau\mu_k]}{n_2 + n_3/4 + (1-\tau)^2 n_4}, \frac{\sigma^2}{n_2 + n_3/4 + (1-\tau)^2 n_4}\right).$$

Finally, for τ I get

$$
\begin{aligned}
p(\tau \mid \sigma^2, \ldots, x) &\propto \exp \frac{-1}{2\sigma^2} \left[\sum_{i_4, k} (x_{i_4 k} - \tau \mu_k - (1-\tau)\gamma_k)^2 \right] \\
&\propto \exp \frac{-1}{2(\sigma^2/n_4)} \left[\sum_k (\bar{x}_{4k} - \tau \mu_k - (1-\tau)\gamma_k)^2 \right] \\
&\sim N \left(\frac{n_4(\mu_1 - \gamma_2)(\bar{x}_{41} - \gamma_1) + n_4(\mu_2 - \gamma_2)(\bar{x}_{42} - \gamma_2)}{n_4(\mu_1 - \gamma_2)^2 + n_4(\mu_2 - \gamma_2)^2}, \frac{\sigma^2}{n_4(\mu_1 - \gamma_2)^2 + n_4(\mu_2 - \gamma_2)^2} \right)
\end{aligned}
$$

however, I have to constrain this distribution to be within $[0, 1]$ by rejecting samples outside the range.