BE/Bi 103: Data Analysis in the Biological Sciences

Justin Bois

Caltech

Fall, 2016

5 Parallel tempering MCMC

In this lecture, we will discuss parallel tempering Markov chain Monte Carlo (PTMCMC). This technique allows for effective sampling of multimodal distributions and it avoids getting trapped on local maxima of the posterior. Perhaps even more importantly, it allows us to perform model selection.

5.1 The basic idea

Recall that the posterior distribution we seek to sample in the model selection problem is

$$P(\mathbf{a}_i \mid D, M_i, I) \propto P(\mathbf{a}_i \mid M_i, I) P(D \mid \mathbf{a}_i, M_i, I). \tag{5.1}$$

Now, we define

$$\pi(\mathbf{a}_i \mid D, M_i, \beta, I) = P(\mathbf{a}_i \mid M_i, I) \left[P(D \mid \mathbf{a}_i, M_i, I)\right]^\beta \tag{5.2}$$

$$= P(\mathbf{a}_i \mid M_i, I) \exp\left\{\beta \ln P(D \mid \mathbf{a}_i, M_i, I)\right\}. \tag{5.3}$$

Here, $\beta \in (0, 1]$ is an "inverse temperature" in analogy to statistical mechanics, where the quantity $-\ln P(D \mid \mathbf{a}_i, M_i, I)$ is an energy (so $P(D \mid \mathbf{a}_i, M_i, I)$ is analogous to a partition function).

If β is close to zero (the "high temperature" limit), we are just sampling the prior. If $\beta = 1$, we are sampling our target posterior, the so-called "cold distribution." So, lowering β has the effect of flattening the posterior distribution. Therefore, walkers at higher temperature (lower β) are not trapped at local maxima. By occasionally swapping walkers from adjacent temperatures, we can effectively sample a broader swath of parameter space.

In practice, we choose a set of β's with $\beta = \{\beta_0, \beta_1, \ldots, \beta_m\}$, with $\beta_{i+1} < \beta_i$ and $\beta_0 = 1$. We propose a swap roughly every n_s steps and accept it based on criteria that guarantees the posterior is a stationary distribution of the transition kernel. To do this in practice, we choose a uniform random number on $[0, 1]$ every iteration and propose a swap when this random number is less than $1/n_s$. When we do propose a swap, we randomly pick a temperature β_j from $\{\beta_1, \beta_2, \ldots \beta_m\}$. We then compute

$$r = \min\left(1, \frac{\pi(\mathbf{a}_{i,j} \mid D, M_i, \beta_{j-1}, I)}{\pi(\mathbf{a}_{i,j-1} \mid D, M_i, \beta_{j-1}, I)} \frac{\pi(\mathbf{a}_{i,j-1} \mid D, M_i, \beta_j, I)}{\pi(\mathbf{a}_{i,j} \mid D, M_i, \beta_j, I)}\right). \tag{5.4}$$

Here, we have defined $\mathbf{a}_{i,j}$ as the value of parameter i for a walker at temperature β_j. We then draw another uniform random number on $[0, 1]$ and accept the swap is that number if less than r.

This useful technique is implemented in `emcee.PTSampler`, which we will use in the next tutorial on model selection. Conveniently, it automatically chooses reasonable values of β and swapping rate, though you can specify these as well.

5.2 Model selection with PTMCMC

We will now do some clever ticks to see how we can use PTMCMC to do model selection without making the approximations we in the previous lecture. Recall the statement of Bayes's theorem for the model selection problem, equation (4.3).

$$P(M_i \mid D, I) = \frac{P(D \mid M_i, I)\, P(M_i \mid I)}{P(D \mid I)}. \tag{5.5}$$

The likelihood in the model selection problem is given by the evidence from the parameter estimation problem, as we derived in equation (4.5). Thus,

$$P(M_i \mid D, I) = \frac{P(M_i \mid I)}{P(D \mid I)} \left[\int \mathrm{d}\mathbf{a}_i\, P(\mathbf{a}_i \mid M_i, I)\, P(D \mid \mathbf{a}_i, M_i, I)\right]. \tag{5.6}$$

Now, we define a partition function

$$Z_i(\beta) = \int \mathrm{d}\mathbf{a}_i\, P(\mathbf{a}_i \mid M_i, I)\, \left[P(D \mid \mathbf{a}_i, M_i, I)\right]^\beta. \tag{5.7}$$

Our goal is to compute $Z_i(1)$, since this is exactly the integral in brackets in equation (5.6).

Now, we're going to do a usual trick in statistical mechanics: we will differentiate the log of the partition function (analogous to the derivative of a free energy).

$$\begin{aligned}
\frac{\partial}{\partial\beta} \ln Z_i(\beta) &= \frac{1}{Z_i(\beta)} \frac{\partial Z_i}{\partial\beta} \\
&= \frac{1}{Z_i(\beta)} \int \mathrm{d}\mathbf{a}_i\, \frac{\partial}{\partial\beta} \exp\left\{\ln P(\mathbf{a}_i \mid M_i, I) + \beta \ln P(D \mid \mathbf{a}_i, M_i, I)\right\} \\
&= \frac{1}{Z_i(\beta)} \int \mathrm{d}\mathbf{a}_i\, \ln P(D \mid \mathbf{a}_i, M_i, I)\, \exp\left\{\ln P(\mathbf{a}_i \mid M_i, I) + \beta \ln P(D \mid \mathbf{a}_i, M_i, I)\right\} \\
&= \frac{1}{Z_i(\beta)} \int \mathrm{d}\mathbf{a}_i\, \ln P(D \mid \mathbf{a}_i, M_i, I)\, P(\mathbf{a}_i \mid M_i, I)\, \left[P(D \mid \mathbf{a}_i, M_i, I)\right]^\beta \\
&= \left\langle \ln P(D \mid \mathbf{a}_i, M_i, I)\right\rangle_\beta,
\end{aligned} \tag{5.8}$$

where the averaging is done over the distribution $\pi(\mathbf{a}_i \mid D, M_i, \beta, I)$, and the subscript β indicates that the averaging is done for a specific value of β. We can integrate both sizes of this equation to give

$$\int_0^1 \mathrm{d}\beta \, \frac{\partial}{\partial \beta} \ln Z_i(\beta) = \ln Z_i(1) - \ln Z_i(0) = \int_0^1 \mathrm{d}\beta \, \left\langle \ln P(D \mid \mathbf{a}_i, M_i, I) \right\rangle_\beta . \tag{5.9}$$

Now, if the prior is normalized, as it should be,

$$Z_i(0) = \int \mathrm{d}\mathbf{a}_i \, P(\mathbf{a}_i \mid M_i, I) = 1, \tag{5.10}$$

which means $\ln Z_i(0) = 0$. Thus, we get

$$\ln Z_i(1) = \int \mathrm{d}\mathbf{a}_i \, P(D \mid \mathbf{a}_i, M_i, I) \, P(\mathbf{a}_i \mid M_i, I) = \int_0^1 \mathrm{d}\beta \, \left\langle \ln P(D \mid \mathbf{a}_i, M_i, I) \right\rangle_\beta . \tag{5.11}$$

Fortunately, we have done MCMC, so we can easily compute the integrand for each β from our samples.

$$\left\langle \ln P(D \mid \mathbf{a}_i, M_i, I) \right\rangle_\beta = \frac{1}{n_\text{samples}} \sum_\text{samples} \ln P(D \mid \mathbf{a}_i, M_i, \beta, I). \tag{5.12}$$

Since we had to compute the log likelihood for every step, we have all we need. We then perform numerical quadrature across the values of β that we sampled to get the integral. We therefore can compute the odds ratio of two models M_i and M_j as

$$O_{ij} = \frac{P(M_i \mid I)}{P(M_j \mid I)} \, \frac{Z_i(1)}{Z_j(1)} = \frac{P(M_i \mid I)}{P(M_j \mid I)} \, \exp\left\{ \frac{\int_0^1 \mathrm{d}\beta \, \left\langle \ln P(D \mid \mathbf{a}_i, M_i, I) \right\rangle_\beta}{\int_0^1 \mathrm{d}\beta \, \left\langle \ln P(D \mid \mathbf{a}_j, M_j, I) \right\rangle_\beta} \right\}, \tag{5.13}$$

where the last ratio is via numerical quadrature on results computed directly from our PTMCMC traces using equation (5.12). We can get $\ln Z_i(1)$ using the built-in `thermodynamic_integration_log_evidence()` method of an `emcee.PTSampler` instance. Note that we have made no approximations at all in the model. The calculation is only approximate to the extent that the PTMCMC sampler takes a finite number of samples and numerical quadrature is not exact.