Bayesian Statistics II: Techniques and Models

2 Markov chain Monte Carlo (MCMC)

2.1 Metropolis-Hastings

Metropolis-Hastings is an algorithm that allows us to sample from a generic probability distribution (which we'll call our target distribution), even if we don't know the normalizing constant. To do this, we construct and sample from a Markov chain whose stationary distribution is the target distribution that we're looking for.

Let the target distribution be $p(\theta)$, which we only know up to a normalising constant, i.e.

$$p(\theta) \propto p^*(\theta)$$

The Metropolis-Hastings is as follows:

Result: Write here the result
Select initial value θ_0
for $i = 1, \ldots, m$ **do**
  Draw candidate $\theta^* \sim q(\theta^*|\theta_{i-1})$;
  Compute $\alpha = \frac{p^*(\theta^*)/q(\theta^*|\theta_{i-1})}{p^*(\theta_{i-1})/q(\theta_{i-1}|\theta^*)} = \frac{p^*(\theta^*)q(\theta_{i-1}|\theta^*)}{p^*(\theta_{i-1})q(\theta^*|\theta_{i-1})}$; // $\alpha \geq 0$
  if $\alpha \geq 1$ **then**
    Accept θ^* and set $\theta_i \leftarrow \theta^*$; // move to θ^*
  else
    Accept θ^* and set $\theta_i \leftarrow \theta^*$ with probability α ; // move to θ^*
    Reject θ^* and set $\theta_i \leftarrow \theta_{i-1}$ with probability $1 - \alpha$; // stay where we are
  end
end

Note that the new value θ_i only depends on the current value θ_{i-1}, hence it is a Markov chain.

Further, the candidate generating distribution $q(\theta^*|\theta_{i-1})$ is an important choice. One choice is using the same distribution, independent of the value of θ, i.e. $q(\theta^*|\theta_{i-1}) = q(\theta^*)$. In order for this to be effective, $q(\theta^*) \approx p(\theta)$. Another choice for the candidate generating distribution is one which centres its mean around the previous value of θ, θ_{i-1}, which is known as *random walk* Metropolis-Hastings. An example would be $q(\theta^*|\theta_{i-1}) = \mathcal{N}(\theta_{i-1}, 1)$. Choosing a normal distribution here has the added advantage of simplifying the ratio for α, since the normal is symmetric about its mean.

2.2 Gibbs sampling

2.2.1 Multiple parameter sampling and full conditional distributions

Suppose we're interested in the posterior distribution of multiple parameters and that posterior distribution doesn't have a standard form. One option is to perform Metropolis Hastings by sampling candidates for all the parameters at once and accepting or rejecting all of those candidates together. A simpler alternative is to sample fewer parameters at a time.

Suppose we know the posterior distribution over two parameters up to a normalising constant.

$$p(\theta, \phi|y) \propto g(\theta, \phi)$$

Result: S-many samples of $(\theta_1, ..., \theta_K)$
Select initial value $\theta_1^{(0)}, ..., \theta_K^{(0)}$
for s = *1, ..., S* **do**
  for k = *1, ..., K* **do**
    Draw $\theta_k^{(s)} \sim p(\theta_k|\theta_1^{(s)}, ..., \theta_{k-1}^{(s)}, \theta_{k+1}^{(s-1)}, ..., \theta_K^{(s-1)}, y)$;   // draw θ_k given the others
  end
end

2.2.2 Conditionally conjugate prior example with Normal likelihood

2.2.3 Computing example with Normal likelihood

2.3 Assessing convergence

3 Common statistical models

4 Count data and hierarchical modeling