

Discussion — Monday, March 30

MCMC Summary

Markov chain Monte Carlo (MCMC) is a general computational method for taking a (dependent) random sample from an arbitrary distribution h defined on a space Ω which need only be known up to a normalizing constant. (Specifically, $h(\theta) \geq 0$ for all $\theta \in \Omega$ and $0 < \int_{\theta \in \Omega} h(\theta) d\theta < \infty$, and the value of the integral is unknown.) Once a sample is obtained, any characteristics of the distribution can be estimated by computing equivalent quantities from the sample. For example, the sample mean will estimate the mean of the distribution, the proportion of sample points in a region estimates the probability of that region, and so on. MCMC is especially useful in Bayesian inference where the posterior density $\pi(\theta | s) = \pi(\theta)f_{\theta}(s)/m(s)$, where in many problems $h(\theta | s) = \pi(\theta)f_{\theta}(s)$ can be computed easily, but $m(s)$ cannot.

The version of MCMC we discuss here is called the Metropolis-Hastings algorithm. The general idea is to modify a Markov chain q defined on Ω , the space of the possible values of θ , by rejecting some proposed transitions and resampling the current value of θ . When this is done properly, the stationary distribution of the modified Markov chain is proportional to the target distribution h . Here is the algorithm for selecting a sample $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$.

(Metropolis-Hastings)

1. Choose a value $\theta^{(0)} \in \Omega$ for which $h(\theta) > 0$.
2. Let $i = 0$, $B > 0$ be some large number (the MCMC sample size).
3. Propose random value $\theta^* \sim q(\cdot | \theta^{(i)})$.
4. Let $\alpha = \min \left\{ 1, \frac{h(\theta^*)}{h(\theta^{(i)})} \times \frac{q(\theta^{(i)} | \theta^*)}{q(\theta^* | \theta^{(i)})} \right\}$ be the *acceptance probability*.
5. Generate $U \sim \text{Uniform}(0,1)$
6. If $U < \alpha$, then set $\theta^{(i+1)} = \theta^*$ (*the proposal is accepted*). Otherwise, set $\theta^{(i+1)} = \theta^{(i)}$ (*the proposal is rejected*).
7. Set $i = i + 1$.
8. If $i < B$, go back to step 3. Otherwise, end.

Notice that the specification of q is essentially arbitrary. The only real requirement is *irreducibility* — it must be possible to use q to get from any point in Ω to any other in a finite number of steps.

Here $q(\cdot | \theta)$ is a probability density on Ω indexed by the value of θ . The collection of distributions $\{q(\cdot | \theta)\}$ determine a Markov chain on Ω . If Ω is a discrete space, then $\{q(\cdot | \theta)\}$ can be represented by a Markov transition matrix, but the idea generalizes to general state spaces.

Problems

In a homework problem, you found that when $\theta \sim \text{Gamma}(\alpha_0, \beta_0)$, and $s = (x_1, \dots, x_n) \mid \theta \sim \text{i.i.d. Poisson}(\theta)$, then $\theta \mid s \sim \text{Gamma}(\alpha_0 + \sum_{i=1}^n x_i, \beta_0 + n)$. But what if we wanted to use a non-canonical prior distribution, such as $\pi(\theta) = (1/M)$ for $0 < \theta < M$. In this case, the posterior density is

$$\pi(\theta \mid s) = \frac{(1/M)e^{-n\theta}\theta^{\sum_{i=1}^n x_i} / \prod_{i=1}^n x_i!}{\int_0^M (1/M)e^{-nw}w^{\sum_{i=1}^n x_i} / \prod_{i=1}^n x_i! dw} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\int_0^M e^{-nw}w^{\sum_{i=1}^n x_i} dw}$$

and the bottom integral has no simple solution.

To make this problem concrete, assume $M = 20$, $n = 50$, and $\bar{x} = 7.86$.

1. Calculate the unnormalized density on a fine grid and estimate the normalizing constant numerically.
2. Develop an MCMC approach using the proposal density $\theta^* \mid \theta \sim \text{Uniform}(\theta - c, \theta + c)$ where $c > 0$ is a fixed *tuning parameter*. Select a value of c that works well.
3. Develop an MCMC approach using the proposal density $\theta^* \mid \theta \sim \text{Normal}(\theta, c^2)$ where $c > 0$ is a fixed *tuning parameter*. Select a value of c that works well.
4. Develop an MCMC approach using the proposal density $\theta^*/20 \mid \theta \sim \text{Beta}(c\theta, c(1 - \theta))$ where $c > 0$ is a fixed *tuning parameter*. Select a value of c that works well.
5. For each approach, estimate the posterior mean, estimate $P(\theta > 8.5)$, and find a 90% credible region for θ .