Chapter 11: Sampling Methods

Lei Tang

Department of CSE
Arizona State University

January 7, 2008

Slice Sampling

Metropolis algorithm is sensitive to step size:

- Too samll $\Longrightarrow$ slow decorrelation.
- Too large $\Longrightarrow$ high rejection rate.

Slice sampling provides an adaptive step size that is automatically adjusted to match the characteristics of the distribution.

Slice sampling involves augmenting z with an additional variable u, and draw samples from the joint (z, u) space.

- Goal: Sample uniformaly from

$$\hat{p}(z, u) = \begin{cases} 1/Z_p & \text{if} 0 \leq \hat{p}(z) \\ 0 & \text{otherwise} \end{cases}$$

- The marginal distribution over z is given by

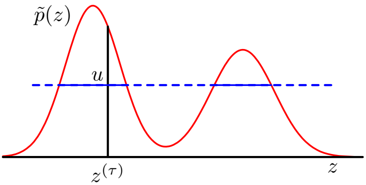
$$\int \hat{p}(z, u) du = \int_0^{\hat{p}(z)} \frac{1}{Z_p} du = \frac{\hat{p}(z)}{Z_p} = p(z)$$

- We can sample from $p(z)$ by sampling from $\hat{p}(z, u)$ and ignore u.

Procedure

- Alternatively sample z and u.
- Given z, sample u uniformly from the range $0 \leq u \leq \hat{p}(z)$.
- Given u, sample z uniformly from the *slice* $\{z : \hat{p}(z) > u$ as

$$p(z|u) = \frac{p(z,u)}{p(u)} = \frac{1/Z_p}{\int_{z:\hat{p}(z)>u} 1/Z_p du} = U\{z : \hat{p}(z) > u\}$$



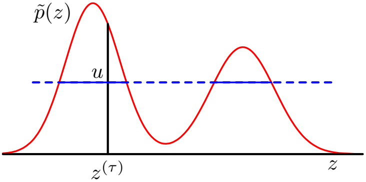
(a)

- Comment: idea is very similar to rejection sampling, but here a Gibbs sampling procedure is employed.
- Key: How to draw z from the *slice*?

Procedure

- Alternatively sample z and u.
- Given z, sample u uniformly from the range $0 \leq u \leq \hat{p}(z)$.
- Given u, sample z uniformly from the *slice* $\{z : \hat{p}(z) > u$ as

$$p(z|u) = \frac{p(z,u)}{p(u)} = \frac{1/Z_p}{\int_{z:\hat{p}(z)>u} 1/Z_p du} = U\{z : \hat{p}(z) > u\}$$



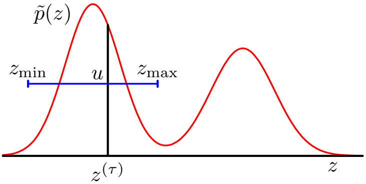
(a)

- Comment: idea is very similar to rejection sampling, but here a Gibbs sampling procedure is employed.
- Key: How to draw z from the *slice*?

- Suppose the current value of z is $z^{(\tau)}$ and we have obtained a corresponding sample u. The next value of z is obtained by considering a region

$$z_{min} \leq z^{(\tau)} \leq z_{max}$$

- The sampling scheme satisfy the detailed balance.
- The region encompass as much of the slice as possible to allow for large moves in z space, while has as little space as possible outside the slice.



(b)

- We can adapt the choice of the region.

Region Adaptation

- Starting with a region containing $z^{(\tau)}$ having some length w and testing each of the end points to see if they lie within the slice.
- If either one does not, extend the region by an increment of w until the end point lies outside the slice.
- A candidate value z' is chose uniformly from the region.
- If it lies within the slice, then forms $z^{(\tau+1)}$. Otherwise, shrunk the region such that the region still contains $z^{(\tau)}$, and another candidate is drawn uniformly from the reduced region and so on.
- How about multi-mode?

Comment

- Might not be as efficient as a well designed Metropolis scheme and Gibbs Sampling. But slice sampling methods will often require less effort to implement and tune.
- Can be efficient form some applications as random walk behavior is supressed.
- Can use Gibbs scheme to handle multiple variables.

Importance Sampling for Model Comparison

$$p_E(\mathbf{z}) = \frac{1}{Z_E} exp(-E(\mathbf{z}))$$

Z_E is the partition function.

- Usually, Z_E is difficult to evaluate directly.
- For model comparison, it is actuallly the ratio of the partition functions for two models.

$$\begin{aligned} \frac{Z_E}{Z_G} &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}) + G(\mathbf{z})) \exp(-G(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \mathbb{E}_{G(\mathbf{z})}[\exp(-E + G)] \\ &\simeq \sum_{l} \exp(-E(\mathbf{z}^{(l)}) + G(\mathbf{z}^{(l)})) \end{aligned}$$

- Yield accurate results if the importance sampling distribution p_G is closed matched to p_E.

Markov chain sampling

- Sampling from p_G can be complicated.
- Alternate approach: Use samples obtained from a Markov chain to define the importance.
- If the transition probability for the Markov chain is given by $T(\mathbf{z}, \mathbf{z}')$, and the sample set is given by $\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(L)}$, then

$$\frac{1}{Z_G} \exp\left(-G(\mathbf{z})\right) = \sum_{l=1}^{L} T(\mathbf{z}^{(l)}, \mathbf{z})$$

- How to estimate the transition probability?

Chaining

- Introducing a succession of intermediate distributions $p_2, \cdots, p_{M-1}$. Then
$$\frac{Z_M}{Z_1} = \frac{Z_2}{Z_1}\frac{Z_3}{Z_2}\cdots\frac{Z_M}{Z_{M-1}}$$
in which the intermmediate ratios can be determined using Monte Carlo methods.
- One way to construct a sequence of intermediate systems is to use an energy function :
$$E_\alpha(\mathbf{z}) = (1-\alpha)E_1(\mathbf{z}) + \alpha E_M(\mathbf{z})$$
- If Monte Carlo methods, then one single Markov chain run is enough. Run initially for the system p_1 and then after some sutiable number of steps, switch to the next distribution. ???
- Must remain close to the equilibrium distribution at each stage.