Stat260: Bayesian Modeling and Inference **Lecture Date: April 7, 2010**

Sampling Methods

Lecturer: Michael I. Jordan *Scribe: Taylor Berg-Kirkpatrick*

1 Gibbs Sampling

Let $x = (x_1, x_2, ..., x_p)$. In order to obtain samples $x^{(i)}$ from the joint distribution $P(x)$ do the following:

- Initialize $x^{(0)}$ and let $i = 0$.
- Repeatedly:

$$\begin{aligned}
&\text{Sample} \quad x_1^{(i+1)} \sim P(x_1 | x_2^{(i)}, x_3^{(i)}, x_4^{(i)}, ..., x_p^{(i)}) \\
&\text{Sample} \quad x_2^{(i+1)} \sim P(x_2 | x_1^{(i+1)}, x_3^{(i)}, x_4^{(i)}, ..., x_p^{(i)}) \\
&\quad\vdots \\
&\text{Sample} \quad x_p^{(i+1)} \sim P(x_p | x_1^{(i+1)}, x_2^{(i+1)}, x_3^{(i+1)}, ..., x_{p-1}^{(i+1)}) \\
&\text{Set} \quad i = i + 1
\end{aligned}$$

It is possible to do this block-wise, i.e. sample blocks of the x_i together. Various approaches exist (and can be justified) to ordering the variables in the sampling loop. One approach is random sweeps: variables are chosen at random to resample.

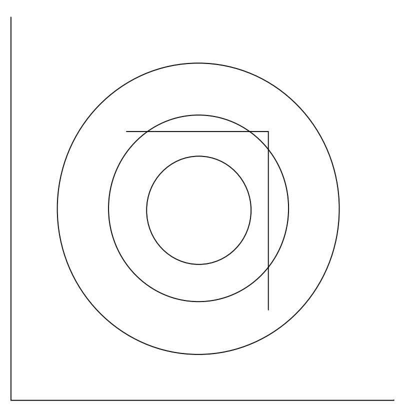

Figure 1: x_1, x_2 actually independent. Gibbs sampler makes big jumps. This is desirable.

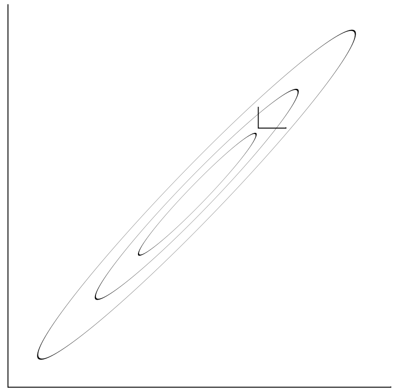

Figure 2: x_1, x_2 highly correlated. Gibbs sampler makes only small moves. This is called chattering and is undesirable.

Example 1 (Gibbs Sampling)**.**

$$\begin{aligned} y_{ij} \sim& \quad \mathcal{N}(\theta_j, \sigma^2) \\ \theta_j \sim& \quad \mathcal{N}(\mu, \tau^2) \\ (\mu, \sigma, \tau) \propto& \quad \tfrac{1}{\sigma} \end{aligned}$$

We want to sample all of $(\theta_1, ..., \theta_J, \mu, \sigma, \tau | y)$. Here's the Gibbs sampler:

$$\begin{aligned} \theta_j | \mu, \tau^2, \sigma^2, y \sim& \quad \mathcal{N}\left(\frac{\frac{1}{\tau^2}\mu + \frac{n_j}{\sigma^2}\bar{y}_{ij}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n_j}{\sigma^2}}\right) \\ \mu | \theta_1, ..., \theta_J, \tau^2 \sim& \quad \mathcal{N}\left(\frac{1}{J}\sum_j \theta_j, \frac{\tau^2}{J}\right) \\ \tau^2 | \theta, \mu \sim& \quad \text{IG}\left(\frac{J-1}{2}, \frac{\sum_j (\theta_j - \mu)^2}{2}\right) \\ \sigma^2 | \theta, y \sim& \quad \text{IG}\left(\frac{n}{2}, \frac{\sum_i \sum_j (y_{ij} - \theta_j)^2}{2}\right) \end{aligned}$$

2 Slice Sampling

Slice sampling is a special case of Gibbs sampling (in a product space). Consider the goal of obtaining samples from $P(x)$. Introduce a new random variable u, conditioned on x, in the following way:

$$\begin{aligned} x &\sim P(x) \\ u | x &\sim \text{Uniform}([0, P(x)]) \end{aligned}$$

This yields a joint distribution $P(x, u)$ such that the marginal distribution on x is the original $P(x)$. Hence, if we can obtain samples from $P(x, u)$, simply ignoring u will give us samples from $P(x)$. The Gibbs sampler for sampling from $P(x, u)$ has the following convenient updates:

Sample $\quad u^{(i+1)} \sim \text{Uniform}([0, p(x^{(i)})])$

Sample $\quad x^{(i+1)} \sim \text{Uniform}(\{x : p(x) > u^{(i+1)}\})$

See Neal (2003) for details on efficient book keeping in slice sampling, and in particular how to efficiently keep track of slices for x: $\{x : p(x) > u^{(i+1)}\}$.

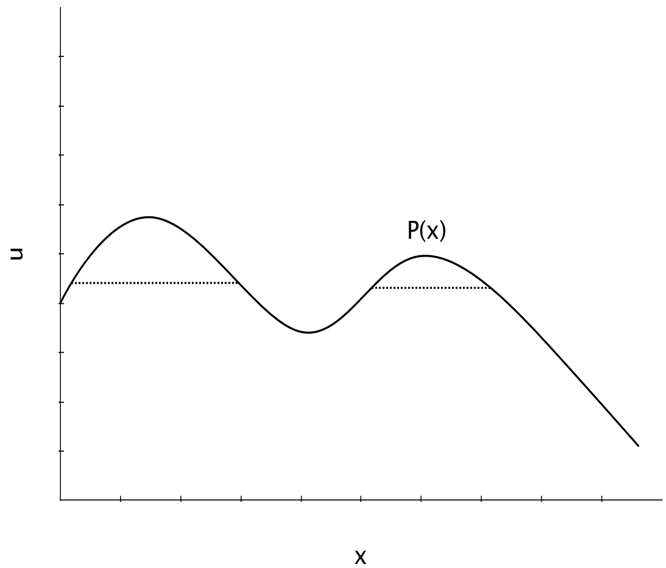

Figure 3: Multimodal distribution where slice for x has two distinct regions.

3 Simulated Annealing

Simulated annealing is mainly used for optimization, but can be used for sampling.

Define a temperature t_i at iteration i.

Sample $\quad x^{(i+1)} \sim (p(x))^{\frac{1}{t_i}}$ (usually via Metropolis-Hastings since it does not require the normalization constant.)

At each iteration t decreases: $t_{i+1} < t_i$. If t goes to 0, simulated annealing performs optimization. If t goes to 1, simulated annealing performs sampling.

Remark 2. Simulated tempering involves running multiple Metropolis-Hastings chains in parallel at different temperatures. Part of the proposal involves proposing to switch between different chains.

References

Neal, R. (2003). Slice sampling. In *Annals of Statistics*.