# Gibbs Sampler

Gibbs sampler is a generalization of slice sampler. It isn't limited to the uniform distribution of subgraph.

Gibbs sampler uses the true conditional distributions, *full conditional distribution*, associated with the target distribution to generate from that distribution.

## Slice Sampler v.s Gibbs Sampler

Similate an uniform distribution on the set:

$$\mathscr{S}(f)=\{(x,y,u):0\leq u\leq f(x,y)\}$$

### Algorithm for Slice Sampler

Move uniformly in one component at a time. Start at a point $(x,y,u)$ in $\mathcal{S}(f)$, we generate

1. $X$ along the $x$ -axis from the uniform distribution on $\{x:u\leq f(x,y)\}$
2. $Y$ along the $y$ -axis from the uniform distribution on $\{y:u\leq f\left(x^{\prime},y\right)\}$
3. $U$ along the $u$ -axis from the uniform distribution on $[0,f\left(x^{\prime},y^{\prime}\right)]$

### Algorithm for Two-Stage Gibbs Sampler

Take $X_0=x_0$. For $t=1,2,\ldots,$ generate

1. $Y_t\sim f_{Y|X}\left(\cdot\mid x_{t-1}\right)$
2. $X_t\sim f_{X|Y}\left(\cdot\mid y_t\right)$

### Disccusion

1. For slice sampler, if we treat $y$ as constant, then step 1. and step 3. is exactly step 1 in slice sampling, i.e. from $f(x|y)=f(x,y)/f(y)\propto f(x,y)$.
2. For slice sampler, the order $x-y-u$ can be changed.
3. For two-stage Gibbs sampler, if $f(x,y)$ satisfies the *positive condition* the stationary distribution for the chain $X$ is $f(x)$ and the transition density for it is

$$K\left(x,x^{*}\right)=\int f_{Y|X}(y\mid x)f_{X|Y}\left(x^{*}\mid y\right)dy$$

## EM Algorithm v.s. Gibbs Sampler

Consider a pair of random variables $(x, z)$, where $x$ is the observed part and $z$ is the missing part.

Their joint distribution is $f(x, z|\theta)$ and $g(x|\theta) = \int f(x, z|\theta)dz$. Then, the complete-data and imcomplete-data likelehood are:

$$L^c(\theta \mid \mathbf{x}, \mathbf{z}) = f(\mathbf{x}, \mathbf{z} \mid \theta) \text{ and } L(\theta \mid \mathbf{x}) = g(\mathbf{x} \mid \theta)$$

The conditional density of missing part is:

$$k(\mathbf{z} \mid \mathbf{x}, \theta) = \frac{L^c(\theta \mid \mathbf{x}, \mathbf{z})}{L(\theta \mid \mathbf{x})}$$

## Gibbs sampler

1. $\mathbf{z}|\theta \sim k(\mathbf{z} \mid \mathbf{x}, \theta)$
2. $\theta|\mathbf{z} \propto L^c(\theta \mid \mathbf{x}, \mathbf{z})$

## EM Algorithm

1. E-step

$$h(\theta) = \mathbb{E}_z \left[L^c(\theta \mid x, z) \mid x, \theta_t\right] = \int L^c(\theta \mid x, z) k\left(z \mid x, \theta_t\right) \mathrm{d}z$$

2. M-step

$$\theta_{t+1} = \arg\max h(\theta)$$

## Discussion

1. The step 1. of Gibbs sampler and EM algorithm is to deal with $z$.
    a. EM algorithm integrals $z$;
    b. Gibbs sampler samples $z$.

   Actually, in EM algorithm, $h(\theta)$ can be viewed as taking average likelihood according to the samples from $z|\theta$.
2. The step 2. of Gibbs sampler and EM algorithm is to update $\theta$.
    a. EM algorithm find $\theta$ maximize the Q-function;
    b. Gibbs sampler samples $\theta$ from the full conditional distribution.