TOPIC:

DATE:

FILE UNDER:

PAGE:

Gibbs Sampling (we want to sample from some posterior or other complicated density)

For K-dim distribution only need K full conditional dists, not K(K-1) conditionals 1, K ... w/ DA.

$p(x, y, z)$ → Target density

⇓

$p(x \mid y, z)$

$p(y \mid x, z)$

$p(z \mid x, y)$

Given the nth iteration (X_n, Y_n, Z_n)

① Sample $X_{n+1} \sim p(X \mid Y_n, Z_n)$

② Sample $Y_{n+1} \sim p(Y \mid X_{n+1}, Z_n)$

③ Sample $Z_{n+1} \sim p(Z \mid X_{n+1}, Y_{n+1})$

$$[X_n, Y_n, Z_n] \xrightarrow{D} [X, Y, Z]$$

$$[X_n] \xrightarrow{D} [X]$$

$$[Y_n] \xrightarrow{D} [Y]$$

$$[Z_n] \xrightarrow{D} [Z]$$

$$\frac{1}{N}\sum_{n=1}^{N} f(X_n, Y_n, Z_n) \longrightarrow \mathbb{E}[f(X, Y, Z)]$$

~~Gibbs~~ Geman + Geman

TOPIC: Intro to Gibbs

DATE:

FILE UNDER:

PAGE:

Data Augmentation

$p(x, y)$

$\Downarrow$

$\cancel{p(y|x) = \int p(y|x, z)}$

$$p(x) = \int p(x|y)\, p(y)\, dy$$

$$p(y) = \int p(y|x)\, p(x)\, dx$$

1. sample y_0
2. sample $x_0 \sim x | y_0$
3. sample $y_1 \sim y | x_0$

$p(x, y, z)$

$$p(x) = \int p(x, z | y)\, p(y)\, dy = \int p(x|z, y)\, p(z|y)\, p(y)\, dy$$

$$p(y) = \int p(y, x | z)\, p(z)\, dz = \int p(y|x, z)\, p(x|z)\, p(z)\, dz$$

$$p(z) = \int p(z, y | x)\, p(x)\, dx = \int p(z|y, x)\, p(y|x)\, p(x)\, dx$$

need 6 conditional distributions

TOPIC:

DATE:

FILE UNDER:

PAGE:

Indepence Metropolis Sampler seems to work well in same situations as rejection sampling.

If $C = \sup_x \frac{\pi(x)}{q(x)} < \infty$, then we have

$\|\pi_n - \pi\| < \cancel{K}\rho^n$ for $0 < \rho < 1$

Rate of convergence ρ depends on C being small (close to 1).

Gibbs Sampler (2 variables)

Let $\pi(x, y)$ be your target density.

Let $Z = (X, Y)$ so that we want to generate a MC $\{Z_n\}$ where π is the stationary dist of the chain. The full conditionals of π are

$$p(y|x) = \frac{\pi(x,y)}{p(x)} \propto \pi(x,y)$$

$$p(x|y) = \frac{\pi(x,y)}{p(y)} \propto \pi(x,y)$$

Let $Z_n = (X_n, Y_n)$. The Gibbs sampler obtains Z_{n+1} by

1. Simulate $X_{n+1} \sim p(x | Y_n)$
2. Simulate $Y_{n+1} \sim p(y | X_{n+1})$

$Z_{n+1} = (X_{n+1}, Y_{n+1})$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Ex. Generating Normal R.V.s

Let g be Unif$[-\delta, \delta]$ dist.

1. Simulate $\varepsilon \sim U(-\delta, \delta)$
2. ~~Let~~ Simulate $u \sim \text{unif}(0,1)$
3. Accept $y = x + \varepsilon$ if
$$u \leq \min\left\{ \frac{\varphi(y)}{\varphi(x)}, 1 \right\}$$

Independence Metropolis Algorithm (Tierney)

Define $q(y|x) = q(y)$.

So
$$\alpha(y|x) = \min\left(\frac{\pi(y)q(x)}{\pi(x)q(y)}, 1 \right)$$
$$= \min\left(\frac{w(y)}{w(x)}, 1 \right)$$

↳ importance weights.

1. Simulate $y \sim q(y)$
2. Simulate $u \sim \text{unif}(0,1)$
3. Accept y if $u \leq \min\left(\frac{w(y)}{w(x)}, 1 \right)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Variations on M–H

Random Walk

Let $q(y|x)$ be defined by

$$y = x + \varepsilon$$

where $\varepsilon \sim g$ and g is symmetric about 0.

Note
$$\left.\begin{aligned} q(y|x) &= g(\varepsilon) \\ &\,\parallel \\ q(x|y) &= g(-\varepsilon) \end{aligned}\right\} \cancel{=g(\varepsilon)}$$

So the Metropolis acceptance prob is

$$\alpha(y|x) = \min\left\{\frac{\pi(y)\,q(x|y)}{\pi(x)\,q(y|x)}, 1\right\}$$

$$= \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}$$

Start at $X_n = x$.

1. Simulate $\varepsilon \sim g$. Let $Y = X_n + \varepsilon$
2. Simulate $U \sim \text{unif}(0,1)$.
3. If $U \leq \alpha(Y|X_n) = \min\left(\frac{\pi(Y)}{\pi(X_n)}, 1\right)$

   Then $X_{n+1} = Y$, else $X_{n+1} = X_n$.

TOPIC: Metropolis Hastings Time-Reversibility

DATE:

FILE UNDER:

PAGE:

(1)

$$K(y|x) = \alpha(y|x)\, q(y|x)$$

$$+ \underbrace{\mathbb{1}\{y=x\}}_{\delta_x(y)} \left[1 - \int \alpha(s|x)\, q(s|x)\, ds\right]$$

(2)

(1) Show $K(y|x)\,\pi(x) = K(x|y)\,\pi(y)$

$$\Leftrightarrow \quad \alpha(y|x)\,q(y|x)\,\pi(x) = \alpha(x|y)\,q(x|y)\,\pi(y)$$

$$\Leftrightarrow \quad \min\left(\cancel{\frac{\pi(y)}{\pi(x)}\frac{q(x|y)}{q(y|x)}}, 1\right) q(y|x)\,\pi(x)$$

$$= \min\left(\frac{\pi(x)}{\pi(y)}\frac{q(y|x)}{q(x|y)}, 1\right) q(x|y)\,\pi(y)$$

$\int a\, \cancel{q\, dy} + \int$

$\int a \cdot q\, dy + [1 - \int a q\, ds] \int \mathbb{1}\{y=x\}\, dy$

$$\Leftrightarrow \quad \min\big(\pi(y)\,q(x|y),\ q(y|x)\,\pi(x)\big) = \min\big(\pi(x)\,q(y|x),\ q(x|y)\,\pi(y)\big)$$

(2) $\underbrace{\cancel{\mathbb{1}\{y=x\}}}_{\delta_x(y)}\, r(x) = \mathbb{1}\{y=x\}\left[1 - \int \alpha(s|x)\, q(s|x)\, ds\right]$

$$\cancel{\pi} \quad \delta_x(y)\, r(x)\, \pi(x) = \delta_y(x)\, r(y)\, \pi(y)$$

over set of points where $y = X$. Obvious!

TOPIC:

DATE:

FILE UNDER:

PAGE:

Metropolis-Hastings

Let $q(y|x)$ be a transition density from which we can easily simulate.

Let $X_n = x$

$\frac{p(x,y)}{p(x)}$

$p(y)$

① Simulate a candidate $Y \sim q(y|x)$

② Let $\alpha(y|x) = \min\left\{\frac{\pi(y)\, q(x|y)}{\pi(x)\, q(y|x)}, 1\right\}$

not necessary

③ Simulate $U \sim \text{unif}(0,1)$ ~~at accept~~

If $U \leq \alpha(y|x)$ then set $X_{n+1} = Y$.
Otherwise set $X_{n+1} = X$.

$K(dy|x) \approx \mathbb{P}(X_{n+1} \in dy \mid X_n = x)$

Why does this work? ~~We need to show that~~ Let $K(y|x)$ be the transition density of the Metropolized chain.

We need to show

$$\pi(y)\, K(x|y) = \pi(x)\, K(y|x)$$

Constants of prop. cancel out

$K(y|x) = \alpha(y|x)\, q(y|x)$, so we need

$$\alpha(y|x)\, q(y|x)\, \pi(x) = \alpha(x|y)\, q(x|y)\, \pi(y)$$

$\Leftrightarrow$

$$\min\left(\frac{\pi(y) q(x|y)}{\pi(x) q(y|x)}, 1\right) q(y|x)\, \pi(x) = \min\left(\frac{\pi(x) q(y|x)}{\pi(y) q(x|y)}, 1\right) q(x|y)\, \pi(y)$$

$\Leftrightarrow$

$$\min\left(\pi(y) q(x|y),\ q(y|x)\pi(x)\right) = \min\left(\pi(x) q(y|x),\ q(x|y)\pi(y)\right)$$

© bothway

TOPIC:

DATE:

FILE UNDER:

PAGE:

Note That

$$\mathbb{P}(X_n = j \mid X_{n-1} = i) = \frac{\mathbb{P}(X_n = j, X_{n-1} = i)}{\pi_i}$$

$$TR \quad = \frac{\mathbb{P}(X_n = i, X_{n-1} = j)}{\pi_i}$$

$$= \mathbb{P}(X_{n-1} = j \mid X_n = i)$$

$\Rightarrow P(i \to j) = P(j \to i)$ "flux" of $i \to j = j \to i$

Main ~~I~~deas:

(1) We want to sample some complicated density π.

(2) We Know that certain Markov chains will converge to a stationary distribution

(3) How to construct a MC s.t. the sequence of values $\{X_n\}$ converges to a target distribution π?

MCMC $\Rightarrow$

We Know that if a MC with transition matrix P or Kernel $K(X,Y)$ is time reversible ~~then~~ wrt π, then π must be the stationary dist of the MC.

Given the chain we start at some point and run it until convergence

TOPIC:

DATE:

FILE UNDER:

PAGE:

$$K(Z_{n+1}|Z_n) = p(y_{n+1}|x_{n+1})\, p(x_{n+1}|y_n)$$

(no x_n)

Ex: Simulate $N(\mu, \Sigma) = N\left((\mu_1, \mu_2), \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right)$

Set $Z_n = (X_n, Y_n)$

(1) Simulate $X_{n+1} \sim N\left(\mu_1 + \square (y_n - \mu_2),\ \sigma_1^2(1-\rho)\right)$ $\quad \rho = \frac{c}{\sigma_1\sigma_2}$

(2) Simulate $Y_{n+1} \sim N\left(\mu_2 + \square (\mu_1 - \mu_2),\ \sigma_2^2(1-\rho)\right)$

Ex: Let $y_1, \ldots, y_n \sim N(\mu, \tau^{-1})$

↳ precision

Suppose we put priors (indep.).

$\mu \sim N(0, w^{-1})$

$\tau \sim \text{Gamma}(\alpha, \beta)$

No conjugacy, so we need Gibbs sampling to explore posterior $\mu, \tau \mid y_1, \ldots, y_n$.

$p(\mu, \tau, y) = L(\mu, \tau|y)\, p(\mu)\, p(\tau)$

$p(\mu|\tau, y) \propto L(\mu, \tau|y)\, p(\mu)$

$p(\tau|\mu, y) \propto L(\mu, \tau|y)\, p(\tau)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$$p(\mu,\tau,y) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left(-\frac{\tau}{2}\sum(y_i-\mu)^2\right)$$

$$\times \left(\frac{w}{2\pi}\right)^{1/2} \exp\left(-\frac{w}{2}\mu^2\right)$$

$$\times \frac{\tau^{\alpha-1}}{\Gamma(\alpha)}\beta^{\alpha}\exp\left(-\tau\beta\right)$$

$$p(\tau|\mu,y) \propto \tau^{n/2}\exp\left(-\frac{\tau}{2}\sum(y_i-\mu)^2\right)$$

$$\times \tau^{\alpha-1}\exp(-\tau\beta)$$

$$= \tau^{\alpha-1+n/2}\exp\left(-\tau\left[\beta+\frac{1}{2}\sum(y_i-\mu)^2\right]\right)$$

$$= \text{Gamma}\left(\alpha+\frac{n}{2},\ \beta+\frac{1}{2}\sum(y_i-\mu)^2\right)$$

$$p(\mu|\tau,y) \propto \exp\left(-\frac{\tau}{2}\sum(y_i-\mu)^2\right)$$

$$\times \exp\left(-\frac{w}{2}\mu^2\right)$$

$$= \exp\left(-\frac{(n\tau+w)}{2}\mu^2-\left(\sum y_i\right)\tau\mu\right)$$

$$= N\left(\left(\frac{\tau}{n\tau+w}\right)\sum y_i\ ,\ \frac{1}{n\tau+w}\right)$$

Gibbs sampler iterates between those densities