TOPIC:

DATE:

FILE UNDER:

PAGE:

Proof of 2-V Gibbs Sampler

Show that this chain is stationary at stationary dist.

It has $\pi(X,Y)$ as invariant distribution

Let $Z_n = (X_n, Y_n)$. Then the transition density is

$$K(Z_{n+1}|Z_n) = \pi_{x|y}(X_{n+1}|Y_n)\,\pi_{y|x}(Y_{n+1}|X_{n+1})$$

We need to show that given state Z_n and π

$$\int K(Z_{n+1}|Z_n)\,\pi(Z_n)\,dZ_n = \pi(Z_{n+1})$$

$$= \iint \pi(Y_{n+1}|Y_n)\,\pi(Y_{n+1}|X_{n+1})\,\pi(X_n, Y_n)\,dX_n\,dY_n$$

$$= \pi(Y_{n+1}|X_{n+1}) \iint \pi(X_{n+1}|Y_n)\,\pi(X_n, Y_n)\,dX_n\,dY_n$$

$$= \pi(Y_{n+1}|X_{n+1}) \int \pi(X_{n+1}|Y_n) \int \pi(X_n, Y_n)\,dX_n\,dY_n$$

$$= \pi(Y_{n+1}|X_{n+1}) \int \pi(X_{n+1}|Y_n)\,\pi(Y_n)\,dY_n$$

$$= \pi(Y_{n+1}|X_{n+1}) \int \pi(X_{n+1}, Y_n)\,dY_n$$

$$= \pi(Y_{n+1}|X_{n+1})\,\pi(X_{n+1})$$

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

$\square$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$$\cancel{p(\alpha,\sigma^2,\mu,\ldots y_n) = \prod_{i=1}^{n} p(y_i|\mu_i)\,p(\mu_i|\alpha,\sigma^2)}$$

$$\cancel{\mu^{y} e^{-\mu}}$$

Relationship b/w Gibbs sampling + Metropolis - Hastings

In M-H we have proposal $q(Y|X)$ and acceptance prob.

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

One can do

~~The Gibbs Sampler~~

Single-component M-H

Let $X^{(t)} = (X_1^{(t)}, X_2^{(t)}, \ldots X_K^{(t)})$ be a K-vector representing our quantities of interest (parameters) at iteration t.

Let $X_{-i}^{(t)} = (X_1^{(t)}, X_2^{(t)}, \ldots X_{i-1}^{(t)}, X_{i+1}^{(t)}, \ldots X_K^{(t)})$

SCMH, at iteration t updates the i^{th} component $X_i^{(t)}$ via:

1. Sample $Y_i \sim q_i(Y|X_i, X_{-i})$ for component i
2. let $\alpha(Y_i|X_i^{(t)}) = \min\left(\frac{\pi(Y_i|X_{-i})\,q(X_i|Y_i,X_{-i})}{\pi(X_i|X_{-i})\,q(Y_i|X_i,X_{-i})}, 1\right)$
3. Accept Y_i for component i w/prob $\alpha(Y_i|X_i^{(t)})$

Repeat K times.

$\pi(X_i|X_{-i})$ is a full conditional distribution

TOPIC:

DATE:

FILE UNDER:

PAGE:

In Gibbs sampling

$$q(y_i \mid X_i, X_{-i}) = \pi(y_i \mid X_{-i})$$

This gives us

$$\alpha(y_i \mid X_i, X_{-i}) = \min\left(\frac{\pi(y_i \mid X_{-i})\,\pi(X_i \mid X_{-i})}{\pi(X_i \mid X_{-i})\,\pi(y_i \mid X_{-i})}, 1\right)$$

$$= 1$$

So Gibbs sampling is like SCMH but it always accepts

General M-H is very exploratory, maybe too much.

Blocking

Hybrid Gibbs Sampler

Sometimes it is not possible to sample directly from a full conditional dist.

One can use a hybrid GS in that case. Suppose we have 2-variable problem w/ (X, Y) and we can easily sample $p(X \mid Y)$ but not $p(Y \mid X)$. Then for $Y_{n+1} \mid X_n$

1. Simulate $Y \sim q(Y \mid X_n)$, draw $u \sim \mathrm{unif}(0,1)$
2. Accept Y if $u \le \min\left(\frac{p(Y \mid X_n)\, q(Y_n \mid X_n)}{p(Y_n \mid X_n)\, q(Y \mid X_n)}, 1\right)$

We just need to simulate one value and move on for things to work

TOPIC:

DATE:

FILE UNDER:

PAGE:

236hkt

Hybrid GS

$Y_{ij} \sim \text{Bernoulli}(p_{ij})$

$\text{logit } p_{ij} = \cancel{\theta}\, \alpha_i \cancel{\theta_j}$

$\cancel{\theta_j \sim N(\mu, \sigma^2)}$ $\alpha_i \sim N(\mu, \sigma^2)$

$\cancel{\theta_j \sim N(}$ $\mu \sim N(0, D)$

$\sigma^2 \sim IG(a, b)$

$$p(Y, \alpha, \mu, \sigma^2) \propto \left[\prod_{i=1}^{n} \left[\prod_{j=1}^{G} p(y_{ij} | \alpha_i)\right] p(\alpha_i | \mu, \sigma^2)\right] \pi(\mu)\, \pi(\sigma^2)$$

(1) $p(\alpha_i | \sim) \propto \left[\prod_{j=1}^{G} p(Y_{ij} | \alpha_i)\right] p(\alpha_i | \mu, \sigma^2)$ $i = 1, \to n$

(2) $p(\mu | \sim) \propto \left(\prod_{i=1}^{n} p(\alpha_i | \mu, \sigma^2)\right] \pi(\mu)$

$$= N\left(\frac{D}{D+\sigma^2} \bar{\alpha}, \frac{\sigma^2}{\sigma^2 + D} D\right)$$

(3) $p(\sigma^2 | \sim) \propto \left[\prod_{i=1}^{n} p(\alpha_i | \mu, \sigma^2)\right] \pi(\sigma^2)$

$$IG\left(a + \frac{n}{2}, b + \frac{1}{2} \sum (\alpha_i - \mu)^2\right)$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

~~If we use~~

If we use a Metropolis step at both iterations, then we have SCMH.

Other possibilities are rejection sampling

K-Variable Gibbs sampler

Let $(X_1^{(n)}, X_2^{(n)}, \ldots, X_K^{(n)})$ be the state at step n. The Gibbs sampler obtains the next state $(n+1)$ by

$$X_1^{(n+1)} \sim X_1 \mid X_2^{(n)}, X_3^{(n)}, \ldots, X_K^{(n)}$$

$$X_2^{(n+1)} \sim X_2 \mid X_1^{(n+1)}, X_3^{(n)}, \ldots, X_K^{(n)}$$

$$\vdots$$

$$X_K^{(n+1)} \sim X_K \mid X_1^{(n+1)}, \ldots, X_{K-1}^{(n+1)}$$

Notes on Gibbs Sampling

1. It is sometimes useful to update groups of variables at a time (i.e. vector update rather than univariate). This is called "block Gibbs".
2. Starting values are a guess. Can try frequentist values (means, medians, modes) or sample from priors. Maybe multiple starts
3. Good to use a burn in period before taking averages / storing values

TOPIC:

DATE:

FILE UNDER:

PAGE:

Start J chains

Monitoring Convergence (Gelman + Rubin)

let $X_j^{(0)}, X_j^{(1)}, \ldots$ be the j^{th} Markov Chain

(1) Discard $X_j^{(0)}, X_j^{(1)}, \rightarrow X_j^{(D-1)}$ values (burn in)

Use values $\underbrace{X_j^{(D)}, X_j^{(D+1)}, \rightarrow X_j^{(D+L-1)}}_{L \text{ values}}$

$$X_j^{(1)}, \rightarrow X_j^{(L)}$$

(2) Calculate

$$\bar{X}_j = \frac{1}{L}\sum_{t=1}^{L} X_j^{(t)} \quad \text{(chain mean)}$$

$$\bar{X}_\bullet = \frac{1}{J}\sum_{j=1}^{J} \bar{X}_j \quad \text{(grand mean)}$$

$$B = \frac{L}{J-1}\sum_{j=1}^{J}(\bar{X}_j - \bar{X}_\bullet)^2 \quad \text{(between chain variance)}$$

$$s_j^2 = \frac{1}{L-1}\sum_{t=1}^{L}\left(X_j^{(t)} - \bar{X}_j\right)^2 \quad \text{(within chain variance)}$$

$$W = \frac{1}{J}\sum_{j=1}^{J} s_j^2$$

(3) let $R = \dfrac{\frac{L-1}{L}W + \frac{1}{L}B}{W}$

$\sqrt{R} < 1.2$ acceptable?

TOPIC:

DATE:

FILE UNDER:

PAGE:

(7) Gibbs sampler w/ improper priors is bad because it will run even if posterior is improper (no warning)

ex $X \sim N(\theta, \sigma^2)$

$$p(\theta, \sigma) = \frac{1}{\sigma^2}$$

but

$$\iint p(x \mid \sigma, \theta)\, p(\theta, \sigma)\, d\theta\, d\sigma = \iint \frac{1}{\sqrt{2\pi}\, \sigma^3} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}\, d\theta\, d\sigma$$

$$= \int \frac{1}{\sigma^2}\, d\sigma = \infty$$

$\Rightarrow p(\theta, \sigma)$ is not a valid prior

But

$$p(\theta \mid \sigma, x) \propto e^{-\frac{1}{2\sigma^2}(x-\theta)^2} = N(x, \sigma^2)$$

$$p(\sigma \mid \theta, x) \propto \frac{1}{\sigma^3} e^{-(\theta-x)^2/2\sigma^2}$$

$$= \frac{1}{\sqrt{e}} \quad \text{where} \quad e \sim \mathrm{Exp}\left(\frac{(\theta-x)^2}{2}\right)$$

Gibbs Sampler is easy to run!

↳ This chain is "null recurrent". There is a stationary/invariant function, but it is not a density.

(8) Your Gibbs Sampler may converge but that isn't always a good thing

TOPIC:

DATE:

FILE UNDER:

PAGE:

For MCs, need

- Irreducibility — positive probability of visiting any set of measure > 0
- Aperiodic — no regularly scheduled visits
- Positive Harris Recurrence
  - ↳ Any set with $\mu(A) > 0$ is visited infinitely often

true if sample space is product

} Imply $X_n \mid X_0 = x \longrightarrow \pi$

stationary dist

Slice Sampling

Given state X_n, Generate a

$$Y_{n+1} \sim \text{Unif}[0, \pi(X_n)]$$

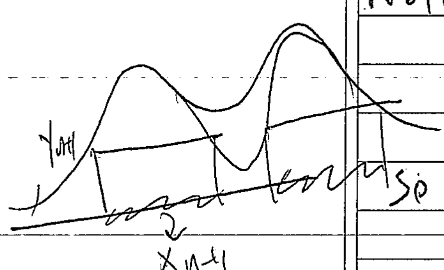
auxiliary variable

Given Y_{n+1}, generate

$$X_{n+1} \sim \text{Unif}[\{x : \pi(x) \geq Y_n\}]$$

Note that

$$\pi(x) = \int \mathbb{1}\{0 \leq y \leq \pi(x)\}\, dy$$



so that

$$f(x,y) = \mathbb{1}\{0 \leq y \leq \pi(x)\}$$ is

a joint density

TOPIC:

DATE:

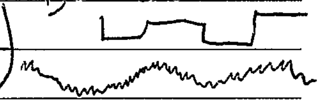
FILE UNDER:

PAGE:

Hit and run

(4) Monitoring convergence is central
- look at marginal trace plots
- try multiple chains with widely dispersed starting values
- Monitor the acceptance rate (~30% is good for > 2 dimensions, ~50% for 1,2-D.

too low acceptance → (slow convergence)
too high acceptance → (slow mixing)



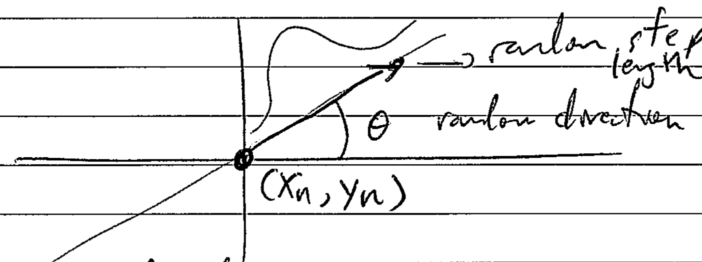
(5) Order of updating need not be sequential. One
- can do 1,2,3,4,3,2,1,2,3 ...
- or randomly choose a component ("random scan")
- randomly permute the order of updating
(1,2,3,4), (1 3 4 2), (4 3 2 1) ......

(6) Suppose X is multivariate and X_n is the current state. Choose a random direction e_n. Sample a scalar r from the density

$$p(r) \propto \pi(X_n + r e_n)$$

Set $X_{n+1} = X_n + r e_n$

In 2-D



- searches in more directions (than just along axes)
- how to sample from $p(r)$?
- using a Metropolis step results in random walk Metropolis

TOPIC:

DATE:

FILE UNDER:

PAGE:

MCMC Diagnostics

(9) You need to babysit your Gibbs sampler. Look at output!

(10) Try exhausting all iid sampling options

(8) For estimation, one can use batch means; Suppose we have chain $X_1, X_2, X_3, \ldots$ and we want $\mathbb{E}h(X)$. Then we have

$$\underbrace{h(X_1) \quad h(X_2) \quad h(X_3) \ldots h(X_K)}_{\text{Batch 1}}, \underbrace{h(X_{K+1}), \ldots}_{\text{Batch 2}}$$

Let $b_1 = \frac{1}{K}\sum_{i=1}^{K} h(X_i)$ $\qquad$ $b_2 = \frac{1}{K}\sum_{i=K+1}^{2K} h(X_i)$

Things far apart are roughly independent

~~Ergodic Theorem says~~

$$\bar{b} = \frac{1}{M}\sum_{i=1}^{M} b_i = \frac{1}{KM}\sum_{i=1}^{KM} h(X_i) \longrightarrow \mathbb{E}[h(X)]$$

# of batches

The batches should be roughly independent (by ergodic theorem) so for large M

$$\sqrt{M}\left(\frac{\bar{b} - \mathbb{E}h(x)}{s}\right) \longrightarrow N(0,1)$$

where $s^2 = \frac{1}{M}\sum(b_i - \bar{b})^2$, the sample variance of b_i's.