$\hat{\theta}_2 = \frac{\sum X_i(1-I_i)}{\sum 1-I_i}$, $\bar{X}_{mix\,2}$

Get $\hat{\sigma}_1^2$

$$\frac{\partial}{\partial \sigma_1^2}[\ell] = -\frac{1+\frac{1}{2}\sum I_i}{\cancel{\sigma_1^2}} + \frac{\sum I_i (X_i-\theta)^2}{2(\sigma_1^2)^2} = 0$$

$$\Rightarrow 1+\tfrac{1}{2}\sum I_i = \frac{1}{2\sigma_1^2}\sum I_i (X_i-\theta_1)^2$$

$$\Rightarrow 2+\sum I_i = \frac{1}{\sigma_1^2}\sum I_i (X_i-\theta_1)^2$$

$$\Rightarrow \hat{\sigma}_1^2 = \frac{\sum I_i (X_i-\theta)^2}{2+\sum I_i}$$

Likewise, $\hat{\sigma}_2^2 = \frac{\sum (1-I_i)(X_i-\theta)^2}{2+\sum(1-I_i)}$

$\hat{p}$ get $\frac{\partial}{\partial p}[\ell] \stackrel{set}{=} 0$

$$= \frac{\sum I_i}{p} - \frac{\sum 1-I_i}{1-p} = 0$$

Use data until converges

Gibbs

$$\frac{\sum I_i}{p} = \frac{\sum 1-I_i}{1-p} = n-\sum I_i$$

$$\sum I_i - p\sum I_i = pn - p\sum I_i \Rightarrow \boxed{\hat{p} = \frac{\sum I_i}{n}}$$

5/9/17

Recall $X_1,\dots,X_n|\theta,\sigma^2 \stackrel{iid}{\sim} N(\theta,\sigma^2)$

$\theta \sim N(\mu_0,\tau^2)$

$\sigma^2 \sim InvG\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

Do not sample from InvGamma

$\Rightarrow p(\theta,\sigma^2|X) \propto N(\theta_p,\sigma_p^2)\,K(\sigma^2|X)$

Recall

$p(\theta|\sigma^2,X) = N(\theta_p,\sigma_p^2)$,

$p(\sigma^2|\theta,X) = InvG\left(\frac{n_0+n}{2}, \frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2}\right)$

Can we use these two cond. distributions to tell us about $P(\theta, \sigma^2 | X)$?

$$P(AB) = P(A|B) \cdot P(B) = P(B|A) P(A)$$

$$P(\theta, \sigma^2 | X) = \underbrace{P(\theta | \sigma^2, X)} P(\sigma^2 | X) = \underbrace{P(\sigma^2 | \theta, X)} P(\theta | X)$$

What if we used an iterative algorithm?

① Assume an arbitrary value of θ_0.

② Draw σ_0^2 from $P(\sigma^2 | \theta = \theta_0, X) = \text{InvG}$

③ Draw θ_1 from $P(\theta | \sigma^2 = \sigma_0^2, X)$

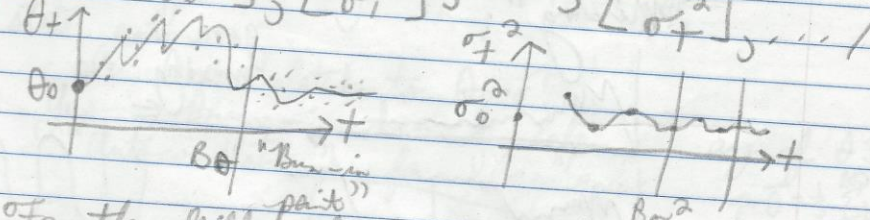
④ Draw σ_1^2 from $P(\sigma^2 | \theta = \theta_1, X)$

④ Repeat 2, 3 until "convergence."

This algorithm is called "Gibbs sampling" or "the Gibbs sampler." (1980's)

Output of a Gibbs sampler looks like

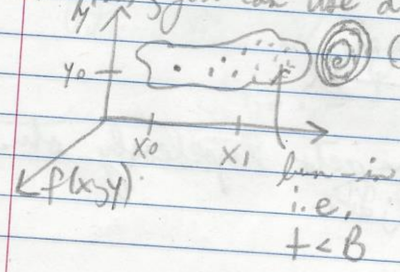
2-D RV $\left\langle \begin{bmatrix} \theta_0 \\ \sigma_0^2 \end{bmatrix}, \begin{bmatrix} \theta_1 \\ \sigma_1^2 \end{bmatrix}, \ldots, \begin{bmatrix} \theta_t \\ \sigma_t^2 \end{bmatrix}, \ldots \right\rangle$



For the Gibbs chart, the burn-in is the max B_j for all θ_j's.

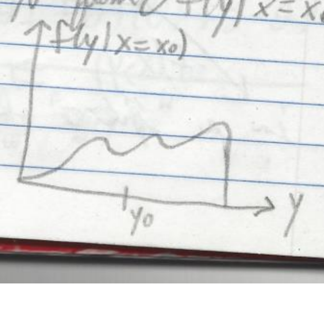
For $t \geq B$, the claim $f(x, y)$ is "burned-in."

If you seek $f(x,y)$ but only know $f(x|y)$ & $f(y|x)$, you can use a Gibbs sampler.



① Begin at arbitrary x_0.

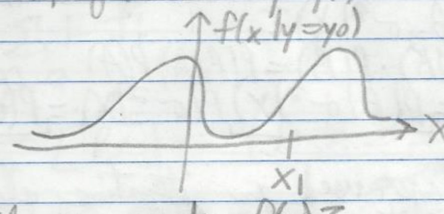
② Draw y_0 from $f(y | x = x_0)$



9:30-9:40

Hamiltonian Monte Carlo

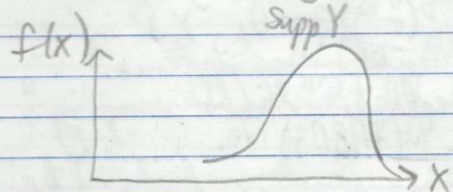
③ Draw x_1 from $f(x|y=y_0)$



$\cancel{\left\langle \binom{x_0}{y_0}, \binom{x_1}{y_1} \right.}$, $(X_B)(X_{B+1}),\dots(X_T)\rangle$

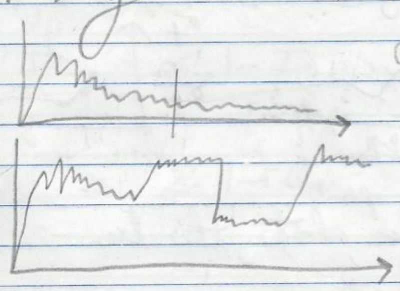
What if you want $f(x)$?

$$f(x) = \int_{\text{supp } Y} f(x,y)\,dy$$

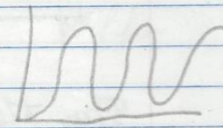


Main Problem w/ Gibbs Sampler

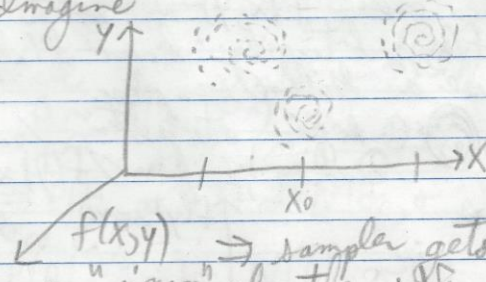
"Mixing"



good mixing

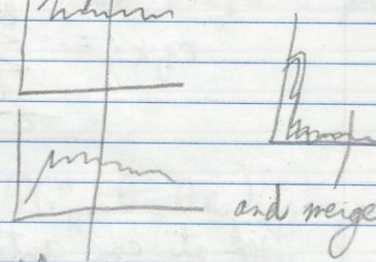


Imagine



$\Rightarrow$ sampler gets hopelessly stuck in a "piece" of the pdf.

a possible solution: use multiple Gibbs chains from different x_0's.



a smaller problem (which is more fixable) is as follows. Is θ_1 related to θ_0? Yes.

and merge chains

$$\text{Cor}[X, Y] := \frac{\text{Cov}(X,Y)}{SE(X)SE(Y)} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$\text{Cov}[\theta_1, \theta_0] \neq 0$$

continuous cor.

$$r : \frac{S_{xy}}{S_x S_y} = \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})/n-1}{\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n-1}}}$$

estimators for stats. err.

Is θ_{1000} related to θ_{999}?

Yes $\Rightarrow$ burn-in doesn't help.

"Autocorrelation" for "lag" one:

avg. of θ's

$$\bar{\theta} = \frac{1}{S}\sum_{t=B}^{B+S} \theta_t$$

$$r_{\theta,1} := \frac{\sum_{t=B}^{B+S-1} (\theta_t - \bar{\theta})(\theta_{t+1} - \bar{\theta})}{\sum_{t=B}^{B+S} (\theta_t - \bar{\theta})^2}$$

B+S is total # of iterations

Autocorrelation w/ lag 2:

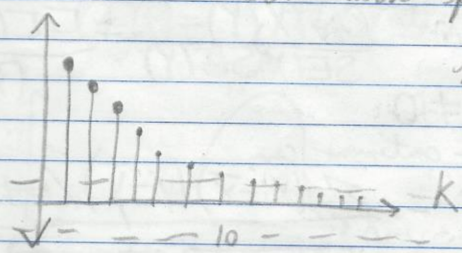
$$r_{\theta,2} := \frac{\sum_{t=B}^{B+S-2} (\theta_t - \bar{\theta})(\theta_{t+2} - \bar{\theta})}{\sum_{t=B}^{B+S} (\theta_t - \bar{\theta})^2}$$

Autocorrelation for lag k:

$$r_{\theta,k} := \frac{\sum_{t=B}^{B+S-K} (\theta_t - \bar{\theta})(\theta_{t+k} - \bar{\theta})}{\sum_{t=B}^{B+S} (\theta_t - \bar{\theta})^2}$$

Look at an "autocorrelation plot"

Time Series



Retain $T \bmod T = 0$
Thin every 10 here

Gibbs is smaller than Metropolis

"T" $\rightarrow$ "this #"

independent draws $\Rightarrow$ $\underbrace{\left\langle \begin{bmatrix} \theta_B \\ \sigma^2_B \end{bmatrix}, \begin{bmatrix} \theta_{B+T} \\ \sigma^2_{B+T} \end{bmatrix}, \begin{bmatrix} \theta_{B+2T} \\ \sigma^2_{B+2T} \end{bmatrix}, \ldots \right\rangle}_{\text{burned + thinned chain}}$

good to have ≥ 1000 length L

$$\hat{\theta}_{MMSE} = E[\theta|X] \approx \bar{\theta} = \frac{1}{L} \sum_{l=1}^{L} \theta_l$$

$\hat{\theta}_{MAE} = Med[\theta|X] \approx$ order θ's from lowest to highest

$\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(L)}$ and use $\theta_{(\frac{L}{2})}$.

$$P(X^*|X) = \int_{\Theta} p(X^*|\theta) P(\theta|X) \, d\theta$$

FINAL

Metropolitan - Hastings
model verification

How do we sample from this distribution?
Sample θ from posterior then sample X from θ.
Pick an arbitrary θ

① Pick $l \in \{1, \dots, L\}$

② Sample X^* from $P(X^* | \theta = \theta_l)$

③ Repeat steps 1-2.

Gibbs is a markov chain.

5/11/17

Algorithm: Systematic Sweep
Gibbs Sampler
$P(\theta_1, \dots, \theta_p | X)$, the unknown posterior w/ p parameters.
Here, all conditionals, $\forall j$,
$P(\theta_j | \theta_{-j})$ where
$\theta_{-j} = \{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p\}$ are known and can be "easily" sampled from.

$P(\vec{\theta}_{t+1} | \vec{\theta}_t)$

Conditional on data $X, \dots$

Step 1: Initialize $\vec{\theta} = \langle \theta_{0,1}, \theta_{0,2}, \dots, \theta_{0,p} \rangle$

Step 2: Sample $\theta_{1,1}$ from
$P(\theta_{1,1} | \theta_2 = \theta_{0,2}, \dots, \theta_p = \theta_{0,p})$
Sample $\theta_{1,2}$ from $P(\theta_{1,2} | \theta_1 = \theta_{1,1}, \theta_3 = \theta_{0,3}, \dots, \theta_p = \theta_{0,p})$
⋮
Sample $\theta_{1,p}$ from $P(\theta_p | \theta_1 = \theta_{1,1}, \dots, \theta_{p-1} = \theta_{1,p-1})$

Step 3: Repeat step 2 until "convergence."

Proof not req'd

Consider $X_0, X_1, X_2, \dots$ a sequence of r.v.'s. (scalar or vector).