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 $l \in \{1, \ldots, L\}$

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

③ Repeat steps 1-2.

Gibbs is a Markov chain.

5/11/17

Algorithm: Systematic Sweep
Gibbs Sampler

$P(\theta_1, \ldots, \theta_p \mid X)$, the unknown posterior w/ p parameters.
Here, all conditionals, $\forall j$,

$P(\theta_j \mid \theta_{-j})$ where
$\theta_{-j} = \{\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p\}$ are known and can be "easily" sampled from.

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

Conditional on data $X, \ldots$

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

Step 2: Sample $\theta_{1,1}$ from
$P(\theta_{1,1} \mid \theta_2 = \theta_{0,2}, \ldots, \theta_p = \theta_{0,p})$

Sample $\theta_{1,2}$ from $P(\theta_{1,2} \mid \theta_1 = \theta_{1,1}, \theta_3 = \theta_{0,3}, \ldots,$
$\theta_p = \theta_{0,p})$
$\vdots$
Sample $\theta_{1,p}$ from $P(\theta_p \mid \theta_1 = \theta_{1,1}, \ldots, \theta_{p-1} = \theta_{1,p-1})$

Step 3: Repeat step 2 until "convergence."
Proof not req'd
Consider $X_0, X_1, X_2, \ldots$ a sequence of r.v.'s. (scalar or vector).

Each has support $\mathcal{X}$. If $P(X_t \in A | X_{t-1}, X_{t-2}, \ldots, X_0) = P(X_t \in A | X_{t-1})$ $\forall t, \forall A \subset \mathcal{X}$ then the sequence is called a "discrete-time" "Markov chain."

$\Rightarrow$ The Gibbs Sampler is a Markov chain. The Gibbs sampler is a form of "Markov Chain Monte Carlo" (MCMC).

$$P(X_{t+1}) = \int_{\mathcal{X}} P(X_{t+1}, X_t)\,dx = \int_{\mathcal{X}} \underbrace{P(X_{t+1}|X_t)}_{\text{transition kernel}} P(X_t)\,dx$$

If $P(X_{t+1}) = P(X_t)$, then this distribution is called the "invariant" "equilibrium" "stationary" "long term," "steady-state," "limiting" distr.

$$P(X_{t+1}) = P(X_t|X_{t-1})\,P(X_{t-1}|X_{t-2}) \cdots P(X_1|X_0)\,P(X_0)$$

$$= \underset{\text{inv.}}{P(X)} = \lim_{t\to\infty} \int_{\mathcal{X}} \quad dx_0$$

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

$= P(\theta_{t+1,1} | \theta_{t,2}, \ldots, \theta_{t,p}) \cdot (P(\theta_{t+1,2} | \theta_{t+1,1}, \theta_{t,3}, \ldots, \theta_{t,p}) \cdot P(\theta_{t+1,p-1} | \theta_{t+1,1}, \ldots, \theta_{t+1,p-1}, \theta_{t,p}) \cdot P(\theta_{t+1,p} | \theta_{t+1,1}, \ldots, \theta_{t+1,p-1})$

$$P(\vec{\theta}_{t+1}) = \int P(\vec{\theta}_{t+1} | \vec{\theta}_t)\,P(\vec{\theta}_t)\,d\vec{\theta}$$

$$P(\theta_{t+1,1}, \ldots, \theta_{t+1,p}) = \int_{\Theta_1}\int_{\Theta_2}\cdots\int_{\Theta_p} P(\theta_{t,1}, \ldots, \theta_{t,p})\,d\theta_1\,d\theta_2 \cdots d\theta_p$$

$\Rightarrow P(\theta_{t,1} | \theta_{t,2}, \ldots, \theta_{t,p}) \cdot P(\theta_{t,2}, \ldots, \theta_{t,p})$

$$= \int_{\Theta_2}\cdots\int_{\Theta_p} \text{kernel} \left[\int_{\Theta_1} P(\theta_{t,1} | \theta_{t,2}, \ldots, \theta_{t,p})\,d\theta_1\right] \cdot P(\theta_{t,2}, \ldots, \theta_{t,p})\,d\theta_2 \cdots d\theta_p$$

$$\int_{\Theta_3}\int_{\Theta_p} \text{next of the kernel} \left[\int_{\Theta_2} P(\theta_{t+1,1} | \theta_{t,2}, \ldots, \theta_{t,p})\,P(\theta_{t,2}, \ldots, \theta_{t,p})\,d\theta_2\right] d\theta_3 \cdots d\theta_p$$

Semi-conjugate inverse gamma

$P(\theta_{t+1}, \theta_{t,2}, \ldots \theta_{t,p})$ (after kernel)

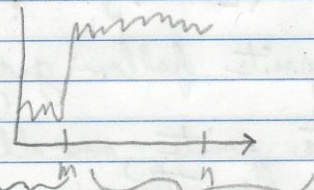
$P(\theta_{t+1}, \theta_{t,2}, \ldots, \theta_{t,p})$

$\underbrace{\int_{\Theta_4} \int_{\Theta_p}}_{\text{rest of kernel}} \left[\int_{\Theta_3} p(\theta_{t+1,2} \mid \theta_{t+1,1}, \theta_{t,3}, \ldots, \theta_{t,p}) \cdot P(\theta_{t+1,1}, \theta_{t,3}, \ldots, \theta_{t,p}) d\theta_3 \right] d\theta_4 \cdots d\theta_p$

$P(\theta_{t+1,1}, \theta_{t+1,2}, \theta_{t,4}, \ldots, \theta_{t,p})$

$= \ldots P(\theta_{t+1,1}, \ldots, \theta_{t+1,p})$

Change Point Model



$\overset{iid}{\sim} \text{Poisson}(\lambda_1)$ $\overset{iid}{\sim} \text{Poisson}(\lambda_2)$

Parameters:

λ_1 - mean of "first process"

λ_2 - mean of second process

m - the "change point"

Priors: (good review)

Uninformative priors are good

$P(\lambda_1) = \text{Gamma}(\alpha, \beta)$

$P(\lambda_2) = \text{Gamma}(\alpha, \beta)$

$P(m) = \text{Uniform}(\{0, \ldots, n\}) = \frac{1}{n} \; \forall m$

$P(\lambda_1)P(\lambda_2)P(m)$

$\|$

$P(\lambda_1, \lambda_2, m \mid X_1, \ldots, X_n) \propto P(X_1, \ldots, X_n \mid \lambda_1, \lambda_2, m) P(\lambda_1, \lambda_2, m)$

$\propto \left(\prod_{i=1}^{m} \frac{e^{-\lambda_1} \lambda_1^{x_i}}{x_i!} \right) \left(\prod_{i=m+1}^{n} \frac{e^{-\lambda_2} \lambda_2^{x_i}}{x_i!} \right) \left(\lambda_1^{\alpha-1} e^{-\beta \lambda_1} \right) \cdot$

$\left(\lambda_2^{\alpha-1} e^{-\beta \lambda_2} \right) \propto e^{-m\lambda_1} \lambda_1^{\sum_{i=1}^{m} x_i} \; \sum_{i=1}^{m} x_i = a \; e^{-(n-m)\lambda_2} \lambda_2^{\sum_{i=m+1}^{n} x_i} \; \sum_{i=m+1}^{n} x_i = b$

$\lambda_1^{\alpha-1} e^{-\beta\lambda_1} \lambda_2^{\alpha-1} e^{-\beta\lambda_2}$

$$= e^{-(m+\beta)\lambda_1} \lambda_1^{a+\alpha-1} e^{-(n-m+\beta)\lambda_2} \lambda_2^{b+\alpha-1}$$ unknown distribution

$$P(\lambda_1 | X_1, \dots, X_n, \lambda_2, m) \propto e^{-(m+\beta)\lambda_1} \lambda_1^{a+\alpha-1} \propto$$

$\text{Gamma}(a+\alpha, m+\beta)$

$$P(\lambda_2 | X_1, \dots, X_n, \lambda_1, m) \propto e^{-(n-m+\beta)\lambda_2} \lambda_2^{b+\alpha-1}$$

$\propto \text{Gamma}(b+\alpha, n-m+\beta)$

dependent on m

$$P(m | X_1, \dots, X_n, \lambda_1, \lambda_2) \propto \underbrace{e^{-m(\lambda_1-\lambda_2)} \lambda_1^{a} \lambda_2^{b}}_{h(m)}$$

$$\propto \frac{h(m)}{\sum_{k=0}^{n} h(k)} \quad \left\langle \begin{bmatrix} \lambda_{0,1} \\ \lambda_{0,2} \\ m_0 \end{bmatrix}, \begin{bmatrix} \lambda_{1,1} \\ \lambda_{1,2} \\ m_1 \end{bmatrix}, \dots \right]$$

Burn-in line: points follow gray line when it burned-in

Delete other data points; now all correlated b/c one vector comes from other

Take posterior and graph credible region

Middle bar is mean ($\hat{\theta}_{MMSE}$)

$CR_{\theta, 95\%}$: capture θ value 95% of the time