Lecture 20 Math 341 5/11/17

NOT REQ'D...

Def: Systematic sweep Gibbs sampler for

$P(\theta_1, \ldots, \theta_p \mid X)$, the unknown posterior with p params.

where $P(\theta_j \mid \theta_{-j})$ is known $\forall j \in \{1, \ldots, p\}$

wh $\theta_{-j} := \{\theta_1, \theta_2, \ldots \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p\}$

i.e. all θ_j's except θ_j.

(See next page)

Step 1: Initialize $\vec{\theta}_0 = \langle \theta_{1,0}, \theta_{2,0}, \ldots, \theta_{p,0} \rangle$

Step 2: sample $\theta_{1,1}$ from $P(\theta_1 \mid \theta_2 = \theta_{2,0}, \ldots \theta_p = \theta_{p,0}, X)$

sample $\theta_{2,1}$ from $P(\theta_2 \mid \theta_1 = \theta_{1,1}, \theta_3 = \theta_{3,0}, \ldots \theta_p = \theta_{p,0}, X)$

$\vdots$

$\theta_{p,1}$ from $P(\theta_p \mid \theta_1 = \theta_{1,1}, \ldots \theta_{p-1} = \theta_{p-1,1}, X)$

and record $\langle \theta_{1,1}, \ldots \theta_{p,1} \rangle$ as a "sample"

Step 3: repeat step 2 ~~for~~ many times...

The claim is that the samples, given $t \to \infty$ come from $P(\theta_1, \ldots \theta_p \mid X)$. Proof...

Def. Consider $X_0, X_1, X_2, \ldots$ a sequence of r.v.'s (scalar or vector) with support $\mathcal{X}$.

If $P(X_t \in A \mid X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_{t-s}) = P(X_t \in A \mid X_{t-1})$ $\forall t, s$ $\forall A \subset \mathcal{X}$

trans. matrix if discr. / trans. kernel if cont.

then $X_1, X_2, \ldots$ is called a "Markov Chain".

Is the Gibbs sampler a Markov Chain?? YES!!! "MCMC"

Def. A Markov chain's "invariant distribution" is defined as:

$$P(X_{t+1}) = \int_{\mathcal{X}} P(X_{t+1}, X_t)\, dx_t = \int_{\mathcal{X}} P(X_{t+1} \mid X_t)\, P(X_t)\, dx_t$$

margin out the effect of the previous value

$$= \int_{\mathcal{X}} P(X_{t+1} \mid X_t)\, P(X_t \mid X_{t-1})\, P(X_{t-1})\, dx$$

same $= P_{inv}(X)$

known

$$= \int \left(\prod_{i=0}^{t+1} P(X_i \mid X_{i-1}) \right) P(X_0)\, dx$$

Cool equation huh?

Thm. for any starting distr. $P(X_0)$,

$$P_{lr}(X) = \lim_{t \to \infty} \int_{\mathcal{X}_t} \prod_{i=0}^{t+1} P(X_i | X_{i-1}) P(X_0) \, d\mathcal{X}$$

$\Rightarrow$ doesn't matter where you start, given enough "hops" or time, you wind up in the same steady state distr. AKA "long-term" / "equilibrium" / "limiting" / "stationary" distr.

Def: A jdf $P(X_1, \ldots, X_p)$ has the positivity cond.

if $\forall j$, $P(X_j) > 0$ $\forall x_j \in \text{Supp}(X_j)$

Thm: Consider $P(X_1, \ldots X_p)$ which has the positivity cond. then, $\forall \vec{a} \in \text{Supp}(\vec{X})$,

$$P(X_1, \ldots, X_p) \propto \prod_{j=1}^{p} \frac{P(X_j | X_1, \ldots, X_{j-1}, X_{j+1} = a_{j+1}, \ldots, X_p = a_p)}{P(X_j = a_j | X_1, \ldots, X_{j-1}, X_{j+1} = a_{j+1}, \ldots, X_p = a_p)}$$

Corr: If $P(X_1, \ldots, X_p)$ has the positivity cond.

$\Rightarrow$ $P(X_j | X_{-j}) > 0$ $\forall x_j \in \mathcal{X}$ i.e. all cond. densities are nonzero.

We need this for the proof...

What is the transition kernel for the Gibbs sampler?

$$P(\vec{\theta}_{t+1} | \vec{\theta}_t, X) = P(\theta_{t+1,1}, \ldots, \theta_{t+1,p} | \theta_{t,1}, \ldots, \theta_{t,p}, X)$$

I will not write X anymore... pretend its there

Systematic sweep steps ... NOT Bayes rule !!

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

not t+1!

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

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

$$\vdots$$

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

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

represents the p steps

$$P(\vec{\theta}_{t+1}) = \int_{\Theta} P(\vec{\theta}_{t+1} \mid \vec{\theta}_t)\, P(\vec{\theta}_t)\, d\vec{\theta}$$

if $P(\vec{\theta}_{t+1}) = P(\vec{\theta}_t)$ then this distr is the invariant distr

$$P(\theta_{t+1,1}, \theta_{t+1,2}, \ldots, \theta_{t+1,p}) = \int_{\theta_{t,1}}\int_{\theta_{t,2}} \cdots \int_{\theta_{t,p}} P(\theta_{t,1}, \theta_{t,2}, \ldots, \theta_{t,p})\, d\theta_{t,1}\, d\theta_{t,2} \cdots d\theta_{t,p}$$

$$\| $$

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

$$= \int_{\theta_{t,2}} \cdots \int_{\theta_{t,p}} P(\theta_{t,2}, \ldots \theta_{t,p}) \underbrace{\int_{\theta_{t,1}} P(\theta_{t,1} \mid \theta_{t,2}, \ldots \theta_{t,p})\, d\theta_{t,1}}_{1} \cdot d\theta_{t,2} \cdots d\theta_{t,p}$$

$$= \int_{\theta_{t,3}} \cdots \int_{\theta_{t,p}} \underset{\text{rest of Kernel}}{} \underbrace{\int_{\theta_{t,2}} \underbrace{P(\theta_{t+1,1} \mid \theta_{t,2}, \ldots, \theta_{t,p})\, P(\theta_{t,2}, \ldots \theta_{t,p})\, d\theta_{t,2}}_{P(\theta_{t+1,1}, \theta_{t,2}, \ldots, \theta_{t,p})}}_{P(\theta_{t+1,1}, \theta_{t,3}, \ldots \theta_{t,p})} d\theta_{t,3} \cdots d\theta_{t,p}$$

$$\int_{\theta_{t,4}} \cdots \int_{\theta_{t,p}} \overset{\text{last of }\text{these}}{\cdots\cdots} \underbrace{\int_{\theta_{t,3}} P(\theta_{t+1,2} \mid \theta_{t+1,1}, \theta_{t,3}, \ldots, \theta_{t,p})\, P(\theta_{t+1,1}, \theta_{t,3}, \ldots, \theta_{t,p})\, d\theta_{t,3}}_{P(\theta_{t+1,2},\, \theta_{t+1,1},\, \theta_{t,4}, \ldots, \theta_{t,p})}\, d\theta_{t,4} \cdots d\theta_{t,p}$$

$\vdots$

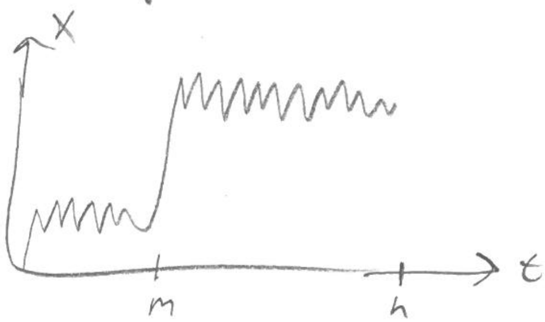
$P(\theta_{t+1,1}, \ldots \theta_{t+p,p})$

proves that the Gibbs Sampler converges.

Some examples....

Suppose you want to do some advanced models...

Change pt model



$\overset{iid}{\sim}$ Poisson(λ_1) $\overset{iid}{\sim}$ Poisson(λ_2)

Both λ_1, λ_2 unknown and m (the change pt) unknown.

Priors

$\lambda_1 \sim \text{Gamma}(\alpha, \beta)$, $\lambda_2 \sim \text{Gamma}(\alpha, \beta)$, $P(m) \propto \frac{1}{n}$ i.e. Uniform discrete

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

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

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

$$= e^{-(m+\beta)\lambda_1} \lambda_1^{\sum_{i=1}^{m} x_i + \alpha - 1} e^{-(n-m+\beta)\lambda_2} \lambda_2^{\sum x_i + \alpha - 1}$$

non-conj!

$$P(\lambda_1 \mid X_1, \ldots X_n, \lambda_2, m) \propto \text{Gamma}\left(\sum_{i=1}^{m} x_i + \alpha,\ m + \beta\right)$$

$$P(\lambda_2 \mid X_1, \ldots X_n, \lambda_1, m) \propto \text{Gamma}\left(\sum_{i=m+1}^{n} x_i + \alpha,\ n - m + \beta\right)$$

$$P(m \mid X_1, \ldots, X_n, \lambda_1, \lambda_2) \propto e^{-m(\lambda_1 - \lambda_2)} \lambda_1^{\sum x_i} \lambda_2^{\sum x_i} = f(m)$$

$$= \frac{f(m)}{\sum_{m=1}^{n} f(m)}$$ since uniform discrete... easy to sample from