if $\left[ P(X) = \lim_{t \to \infty} \int_X \left( \prod_{i=1}^{t+1} P(X_i | X_{i-1}) \right) P(X_0 = x)\, dx \right]$ then Sampler converges

Lec 21: Showing Gibbs sampling works continued

Note: $\theta_{t,p}$ — $t$: sample #, $p$: iteration

Let:

- $\overbrace{P\left(\vec{\theta}_{t+1} \,\middle|\, \vec{\theta}, X\right)}^{A} = \overbrace{P\left(\theta_{t+1,1}, \theta_{t+1,2}, \dots, \theta_{t+1,p} \,\middle|\, \theta_{t,1}, \theta_{t,2}, \dots, \theta_{t,p}, X\right)}^{\text{multiple iterations}}$

* ignoring X for convienence

$$A = \underbrace{\overbrace{P\left(\theta_{t+1,1} \,\middle|\, \theta_{t,\{2,\dots,p\}}\right)}^{\text{terms: } 1} \cdot \overbrace{P\left(\theta_{t+1,2} \,\middle|\, \theta_{t+1,1}, \theta_{t,\{3,\dots,p\}}\right)}^{2} \cdots \overbrace{P\left(\theta_{t+1,p} \,\middle|\, \theta_{t+1,\{1,\dots,p-1\}}\right)}^{p}}_{\text{p steps, systematic sweep steps}}$$

p steps, systematic sweep steps : we find $\theta_{t+1}$ one iteraction at a time using old data that is replaced by new data

So...

$$P(\vec{\theta}_{t+1} | X) = \int_{\Theta} P\left(\vec{\theta}_{t+1} \,\middle|\, \vec{\theta}_t, X\right) P(\vec{\theta}_t | X)\, d\vec{\theta}$$

... converges if $P(\vec{\theta}_{t+1}) = P(\vec{\theta}_t)$ since new data is the same as the old date

- $P\left(\theta_{t+1,\{1,2,\dots,p\}}\right) =$

- $= \int_{\theta_{t,1}} \int \cdots \int_{\theta_{t,p}} \underbrace{\overbrace{P\left(\vec{\theta}_{t+1} \,\middle|\, \vec{\theta}_t, X\right)}^{A}}_{\text{replace w/ the above expression}} P\left(\vec{\theta}_t \,\middle|\, \underbrace{X}_{\text{expandable via bayes rule}}\right) d\theta_{t,1}\, d\theta_{t,2} \cdots d\theta_{t,p}$

expandable via bayes rule: $P\left(\theta_{t,1} \,\middle|\, \theta_{t,\{2,\dots,p\}}\right) \cdot P\left(\underbrace{\theta_{t,\{2,\dots,p\}}}_{B}\right)$

* ignore X

$$= \int_{\theta_{t,2}} \int \cdots \int_{\theta_{t,p}} \underbrace{P\left(\vec{\theta}_{t+1} \,\middle|\, \vec{\theta}_t\right)}_{A} \cdot P\left(\underbrace{\theta_{t,\{2,\dots,p\}}}_{B}\right) \cdot \underbrace{\left[ \int_{\theta_{t,1}} P\left(\theta_{t,1} \,\middle|\, \theta_{t,\{2,\dots,p\}}\right) d\theta_{t,1} \right]}_{=1} d\theta_{t,2} \cdots d\theta_{t,p}$$

$$= \int_{\theta_{t,3}} \cdot \int \cdots \int_{\theta_{t,p}} \underbrace{\overset{A-1}{A}}_{\text{w/o the first term}} \cdot \int_{\theta_{t,2}} \overbrace{\underbrace{P\left(\theta_{t+1,1} \mid \theta_{t,2}, \ldots, \theta_{t,p}\right)}_{\text{1st term of } A} \, P\left(\theta_{t,2}, \ldots, \theta_{t,p}\right)}^{B} \, d\theta_{t,2} \cdots d\theta_{t,p}$$

$A_1$

$$= \int_{\theta_{t,3}} \cdot \int \cdots \int_{\theta_{t,p}} (A-1) \cdot \left[ \int_{\theta_{t,2}} P\left(\theta_{t+1,1}, \theta_{t,2}, \ldots, \theta_{t,p}\right) d\theta_{t,2} \right] \cdots d\theta_{t,p}$$

$$= \int_{\theta_{t,3}} \cdot \int \cdots \int_{\theta_{t,p}} (A-1) \cdot \underbrace{\left( P\ \theta_{t+1,1}, \theta_{t,3}, \ldots \theta_{t,p} \right)}_{\text{B w/o 1st term: } (B-1) \text{ Replaced by new } t+1 \text{ term}} \, d\theta_{t,3} \cdots d\theta_{t,p}$$

∴ repeat steps so that $A_{ith}$ term and $(B-i+1)$ th term combine to create the new $B$ term for next $A$

∴ at the end:

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

proves the Gibbs converges since each iteration for our old data changed to our new

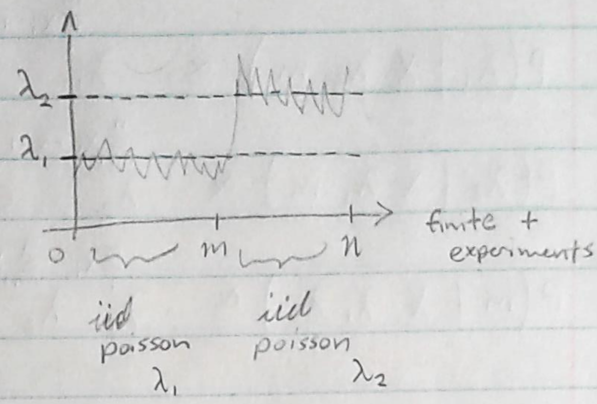
* so it doesnt matter where $\vec{\theta}_t \mid X$ begins, it will eventually converge at some point such that the next sample will be identical to the previous

- ## Change point Model

ex: two iid Poissons st.

$$\vec{\theta} = \begin{cases} \lambda_1 \in (0,\infty) \\ \lambda_2 \in (0,\infty) \\ m \in \{1,2,\dots,n\} \end{cases}$$

must change at some distinct pt.




all unknown

- $$P(\lambda_1 \lambda_2 m \mid \underbrace{X_1 X_2 \dots X_n}_{X}) \propto \underbrace{P(X \mid \lambda_1 \lambda_2 m)}_{\text{Break into parts}} \left[ P(\lambda_1 \lambda_2 m) \right]^{*}$$

origin

Prior based on distribution of value

$\lambda_1, \lambda_2, m$ are all independent so $= P(\lambda_1) \cdot P(\lambda_2)\, P(m)$

$$\begin{aligned} P(\lambda_1) &\sim \text{Gamma}(\alpha,\beta) \propto \lambda_1^{\alpha-1} e^{-\beta\lambda_1} \\ P(\lambda_2) &\sim \text{Gamma}(\alpha,\beta) \propto \lambda_2^{\alpha-1} e^{-\beta\lambda_2} \end{aligned}$$

uninformative if $\alpha = 1$, $\beta = 0$

$$P(m) \sim U\{1,\dots,n\} = 1/n$$

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

posterior

$$P(X_1 \dots X_m \mid \lambda_1) \cdot P(X_{m+1} \dots X_n \mid \lambda_2) \cdot P(\lambda_1)\, P(\lambda_2) \cdot P(m)$$

$$\propto \left( e^{-m\lambda_1} \lambda_1^{\sum_1^n x_i} \right) \left( e^{-(n-m)\lambda_2} \lambda_2^{\sum_{m+1}^n x_i} \right) \left( \lambda_1^{\alpha-1} e^{-\beta\lambda_1} \right) \left( \lambda_2^{\alpha-1} e^{-\beta\lambda_2} \right) \cancel{\left( \frac{1}{n} \right)}$$

constant

- $$K(\lambda_1 \lambda_2 m \mid X) \propto \left( e^{-(m+\beta)\lambda_1} \lambda_1^{\sum_1^m x_i + \alpha - 1} \right) \left( e^{-(n-m+\beta)\lambda_2} \lambda_2^{\sum_{m+1}^n x_i + \alpha - 1} \right)$$
    - Kernel of posterior
    - Unknown form : non conjuate
    - Gibbs sample to find values
        - * create conditionals

* Create conditionals from <u>posterior</u> :

$$P(\lambda_1 \mid X\ \lambda_2\ m) \propto \lambda_1^{\sum_1^m X_i + \alpha - 1} e^{-(m+\beta)\lambda_1} \propto \text{Gamma}\left(\sum_1^m X_i + \alpha,\ m+\beta\right)$$

$$P(\lambda_2 \mid X\ \lambda_1\ m) \propto \lambda_2^{\sum_{m+1}^n X_i + \alpha - 1} e^{-(n-m+\beta)\lambda_2} \propto \text{Gamma}\left(\sum_{m+1}^n X_i + \alpha,\ n-m+\beta\right)$$

$$P(m \mid X\ \lambda_1\ \lambda_2) \propto \lambda_1^{\sum_1^m X_i} \lambda_2^{\sum_{m+1}^n X_i} e^{m(\lambda_2 - \lambda_1)} \propto k(m \mid \lambda_1 \lambda_2 X)$$

unknown

sample from these starting at,

$\lambda_{1_0} = 1$ let
$\lambda_{2_0} = 1$ $\alpha = 1$
$m_0 = \frac{n}{2}$ $\beta = 0$

<u>Note</u> Sampling from kernel

- Grid over support : $< 1, 2, \dots, n >$ (or close)
- find $k(\text{Grid} \mid X)$ for each pt.

so

- $P(m \mid X) = \dfrac{k(m \mid X)}{\sum_1^n k(i \mid X)}$ } discrete sample