Lecture 21 05/07/2020.

Problem with Gibbs sampling:

the samples $\left\langle \begin{bmatrix} \theta_{t,1} \\ \vdots \\ \theta_{t,p} \end{bmatrix}, \begin{bmatrix} \theta_{t+1,1} \\ \vdots \\ \theta_{t+1,p} \end{bmatrix}, \ldots, \begin{bmatrix} \theta_{S,1} \\ \vdots \\ \theta_{S,p} \end{bmatrix} \right\rangle$ AKA "Gibbs chain"

ordered set

are not independent. They're dependent because the $\vec{\theta}_{t+1}$ has sampled from a conditional distribution that contained information from $\vec{\theta}_t$.

Two r.v.'s X_1, X_2 then

Covariance $\sigma_{12} := \text{Cov}[X_1, X_2] := E[(X_1 - \mu_1)(X_2 - \mu_2)]$

Correlation $\rho_{12} := \text{Corr}[X_1, X_2] = \dfrac{\text{Cov}[X_1, X_2]}{SE[X_1]\, SE[X_2]}$

↳ is a measure of dependence

$= \dfrac{\sigma_{12}}{\sigma_1 \sigma_2} \in [-1, 1]$ (Proof in 621) ($= 0$ no dependence)

To estimate these parameters using n realization

$\left\langle \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}, \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}, \ldots, \begin{bmatrix} x_{n1} \\ x_{n2} \end{bmatrix} \right\rangle$

$$\sigma_{12} \approx S_{12} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$$

$$\rho_{12} \approx r_{12} = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^{n} (x_{i1} - \bar{x}_2)^2}}$$

Autocorrelation

$$\rho_{a,j} = \frac{Corr[\theta_{t,j}, \theta_{t-1,j}]}{\underbrace{SE[\theta_j]\, SE[\theta_j]}_{Var[\theta_j]}} \Rightarrow r_{a,j} = \frac{\sum_{t=B+1}^{S} (\theta_{tj} - \bar{\theta}_j)(\theta_{t-1,j} - \bar{\theta}_j)}{\sum_{t=B}^{S} (\theta_{tj} - \bar{\theta}_j)^2}$$

+1 avoids going back into the burn in samples

one iteration prior

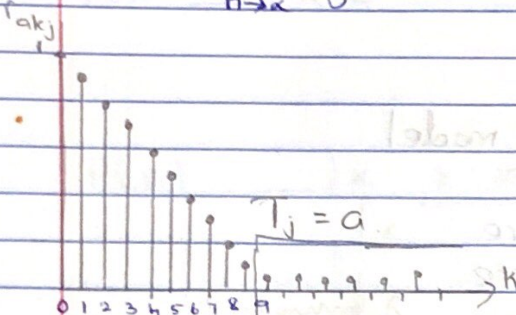
will be high all the time

$$Var[\theta_j] = \tfrac{1}{n-1} \sum (\theta_{tj} - \bar{\theta}_j)^2 \xrightarrow[n \to \infty]{d} \sigma^2$$

$$r_{a2j} = \frac{\sum_{t=B+2}^{S} (\theta_{tj} - \bar{\theta}_j)(\theta_{t-2,j} - \bar{\theta}_j)}{\text{same}}$$

$\vdots$

$$r_{akj} = \frac{\sum_{t=B+k}^{S} (\theta_{t,j} - \bar{\theta}_j)(\theta_{t-k,j} - \bar{\theta}_j)}{\text{same}}$$



$\exists k$ s.t. $r_{akj} \approx 0\ \forall j$ and then $k = T$. $T = \max[T_1, \ldots, T_p]$, the k where $\forall j\ r_{akj} \approx 0$. Now that you have T, ... you "thin" the chain and are left with an unordered set of iid samples from $P(\theta_1, \ldots, \theta_p | X)$:

$$\left\{ \begin{bmatrix} \theta_{B,1} \\ \vdots \\ \theta_{B,p} \end{bmatrix}, \begin{bmatrix} \theta_{B+T,1} \\ \vdots \\ \theta_{B+T,p} \end{bmatrix}, \begin{bmatrix} \theta_{B+2T,1} \\ \vdots \\ \theta_{B+2T,p} \end{bmatrix}, \ldots \right\}$$ which has $\frac{S-B}{T}$ total samples $= N$

average of

Now we can do Bayesian Inference:

$$\hat{\theta}_{j,MMSE} := E[\theta_j | X] \approx \tfrac{1}{N} \sum_{l=1}^{N} \theta_{l,j} = \bar{\theta}_j$$

$CR[\theta_j, 95\%] = [\text{samplequantile}[\theta_j\text{'s}, 2.5\%], \text{samplequantile}[\theta_j\text{'s}, 97.5\%]]$

$H_0: \theta_j \le \theta_0,\ H_a: \theta_j > \theta_0$

$$P_{val} = P(H_0 | X) = P(\theta_j \le \theta_0 | X) = \tfrac{1}{N} \sum_{l=1}^{N} \mathbb{1}_{\theta_{l,j} \le \theta_0}$$

the ⬭ of Gibbs samples $\le \theta_0$.

Sample from $P(x_*|x)$?

Recall $P(x_*|x) = \int_{\Theta_1} \cdots \int_{\Theta_p} P(x_*|\theta_1,\ldots\theta_p) P(\theta_1,\ldots\theta_p)\, d\theta_1,\ldots,d\theta_p$

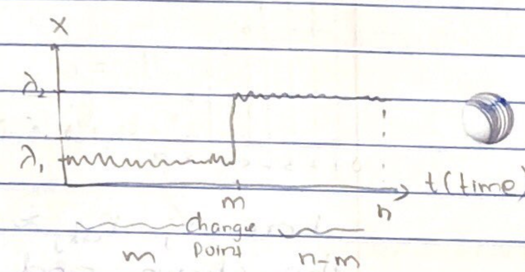
(I) Sample $\vec{\theta}_{samp}$ from the Gibbs set

(II) Sample x_* from likelihood model $P(x_*|\vec{\theta} = \vec{\theta}_{samp})$

(III) Repeat many times

Change point detection model

There's some process where parameter changes somewhere in time.



Let $X_1,\ldots,X_m \overset{iid}{\sim} \text{Poisson}(\lambda_1)$

$X_{m+1},\ldots,X_n \overset{iid}{\sim} \text{Poisson}(\lambda_2)$

Priors
$\lambda_1 \sim \text{Gamma}(1,0) \propto 1$ (laplace)
$\lambda_2 \sim \text{Gamma}(1,0) \propto 1$ (laplace)
$m \sim \text{Unif}(\{1,2,\ldots,n\}) = 1/n \propto 1$ (Laplace)

Principle of Indifference & independent

$$P(\lambda_1,\lambda_2,m|x_1,\ldots,x_n) \propto P(x_1,\ldots,x_n|\lambda_1,\lambda_2,m)\, P(\lambda_1,\lambda_2,m)$$

$$= P(x_1,\ldots,x_m|\lambda_1)\, P(x_{m+1},\ldots,x_n|\lambda_2)\, \underset{\propto 1}{P(\lambda_1)} \underset{\propto 1}{P(\lambda_2)} \underset{\propto 1}{P(m)}$$

$$= \prod_{t=1}^{m} \frac{e^{-\lambda_1}\lambda_1^{x_t}}{x_t!} \prod_{t=m+1}^{n} \frac{e^{-\lambda_2}\lambda_2^{x_t}}{x_t!}$$

$$= \frac{e^{-m\lambda_1}\lambda_1^{\sum_{t=1}^{m} x_t}}{\prod_{t=1}^{m} x_t!} \cdot \frac{e^{-(n-m)\lambda_2}\lambda_2^{\sum_{t=m+1}^{n} x_t}}{\prod_{t=m+1}^{n} x_t!}$$

$$\propto e^{-m\lambda_1} \underbrace{e^{-n\lambda_2} e^{m\lambda_2}}_{e^{-(n-m)\lambda_2}} \lambda_1^{\sum_{t=1}^{m} x_t} \lambda_2^{\sum_{t=m+1}^{n} x_t}.$$

Note: This is not a kernal of a known distribution.

Let's use Gibb's Sampling . . .

$$P(\lambda_1 \mid \_\_\_) \propto e^{-m\lambda_1} \lambda_1^{\sum_{t=1}^{m} x_t (+1-1)} \propto \text{Gamma}\left(\sum_{t=1}^{m} x_t + 1, m\right)$$

$$P(\lambda_2 \mid \_\_) \propto e^{-(n-m)\lambda_2} \lambda_2^{\sum_{t=m+1}^{n} x_t.} \propto \text{Gamma}\left(\sum_{t=m+1}^{n} x_t + 1, n-m\right)$$

$$P(m \mid \_\_\_) \propto e^{m(\lambda_2 - \lambda_1)} \lambda_1^{\sum_{t=1}^{m} x_t.} \lambda_2^{\sum_{t=m+1}^{n} x_t} = K(m \mid \_\_\_)$$

Use grid sampling.

$$P(n \mid \_\_\_) = \frac{k(m \mid \_\_\_)}{\sum_{m=1}^{n} k(m \mid \_\_)}$$