Lecture 21. NO. 1 05/07/2020.

Problem with Gibbs sampling : samples $\left[\begin{bmatrix}\theta_{t,1}\\ \vdots \\ \theta_{t,p}\end{bmatrix}, \cdots \begin{bmatrix}\theta_{t+1,1}\\ \vdots \\ \theta_{t+1,p}\end{bmatrix} \cdots \begin{bmatrix}\theta_{S,1}\\ \vdots \\ \theta_{S,p}\end{bmatrix}\right]$ are not independent.

They're dependent because the θ_{t+1} has sampled from a conditional distribution that contained information from $\overline{\theta_t}$.

Two random variables X_1, X_2, then covariance $\sigma_{12} := Cov[X_1, X_2]$

$:= E[(X_1-\mu_1)(X_2-\mu_2)]$

correlation $\rho_{12} := Corr[X_1, X_2]$

↓ is a measure of dependence.

$$= \frac{Cov[X_1, X_2]}{SE[X_1]SE[X_2]}$$

$$= \frac{\sigma_{12}}{\sigma_1 \sigma_2} \in [-1, 1] \quad (=0, \text{ no dependence.})$$

To estimate these parameters using n realization $\left[\begin{bmatrix}x_{11}\\ x_{12}\end{bmatrix}, \begin{bmatrix}x_{21}\\ x_{22}\end{bmatrix}, \cdots, \begin{bmatrix}x_{n1}\\ x_{n2}\end{bmatrix}\right]$

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

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

- Autocorrelation.

$$\rho_{a1j} = \frac{Corr[\theta_{t,j}, \theta_{t-1,j}]}{SE[\theta_j]SE[\theta_j]} \Rightarrow ra_{1j} = \frac{\sum_{t=B+1}^{S}(\theta_{tj}-\bar{\theta}_j)(\theta_{t-1,j}-\bar{\theta}_j)}{\sum_{t=B}^{S}(\theta_{tj}-\bar{\theta}_j)^2}$$

+1 avoids going back into the burn in samples.

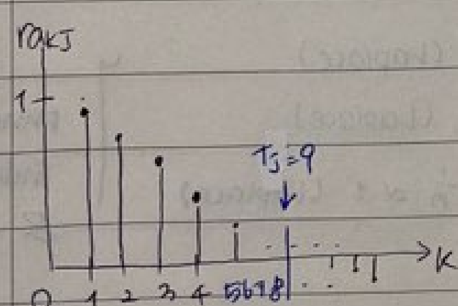
1 iteration prior

$\uparrow var[\theta_j]$

will be high all the time.

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

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



$$ra_{kj} := \frac{\sum_{t=B+k}^{S}(\theta_{tj}-\bar{\theta}_j)(\theta_{t-k,j}-\bar{\theta}_j)}{\text{same.}}$$

$\exists k$ s.t. $ra_{kj} \approx 0 \; \forall j$, and then $k = T$

$T = \max\{T_1, \ldots, T_p\}$, the k where $\forall j \; ra_{kj} \approx 0$

$\{[\theta_{*,1} \dots \theta_{*,p}], [\theta_{*+1,1} \dots \theta_{*+1,p}] \cdots [\theta_{S,1} \dots \theta_{S,p}]\}$ NO

Now that you have T..., AKA "Gibbs chain". ↖ ordered set.

You "thin" the chain and are left with an unordered set of iid samples

from $P(\theta_1,\dots,\theta_p|X)$: $\{ \begin{bmatrix}\theta_{B,1}\\ \vdots \\ \theta_{B,p}\end{bmatrix}, \begin{bmatrix}\theta_{B+T,1}\\ \vdots \\ \theta_{B+T,p}\end{bmatrix}, \begin{bmatrix}\theta_{B+2T,1}\\ \vdots \\ \theta_{B+2T,p}\end{bmatrix}, \dots \}$ $\frac{S-B}{T}$ total samples $= N$

average of

which has.

Now we can do Bayesian Inference.

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

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

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

pvalue $= P(H_0|X) = P(\theta_j \le \theta_0|X) = \frac{1}{N}\sum_{l=1}^{N} \mathbb{1}_{\theta_{l,j} \le \theta_0}$ the p- of Gibbs sampler $\le \theta_0$.

sample from $P(X_*|X)$? Recall $P(X_*|X) = \int_{\Theta_1} \cdots \int_{\Theta_p} p(X_*|\theta_1,\dots,\theta_p) P(\theta_1,\dots,\theta_p|X) d\theta_1 \dots d\theta_p$.

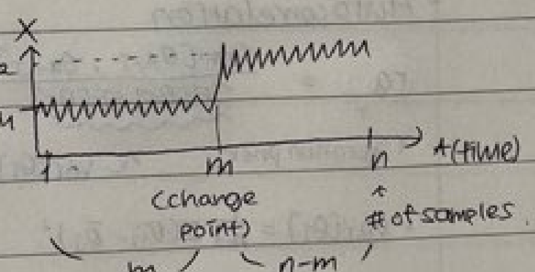
① Sample $\vec{\theta}_{samp}$ from the Gibbs set

② Sample X_* from likelihood model $P(X_*|\vec{\theta} = \vec{\theta}_{samp})$

③ Repeat many times.

Change point detection model

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



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

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

Priors $\lambda_1 \sim Gamma(1,0) \propto 1$ (Laplace)

$\lambda_2 \sim Gamma(1,0) \propto 1$ (Laplace)

$m \sim Unif(\{1,2,\dots,n\}) = \frac{1}{n} \propto 1$ (Laplace)

principle of indifference & independent

NO.

$$P(\lambda_1, \lambda_2, m \mid x_1, \dots, x_n) \propto P(x_1, \dots, x_n \mid \lambda_1, \lambda_2, m) \cdot P(\lambda_1, \lambda_2, m)$$

$$\propto 1 \quad \propto 1 \quad \propto 1$$

$$= P(x_1, \dots, x_m \mid \lambda_1) P(x_{m+1}, \dots, x_n \mid \lambda_2) P(\lambda_1) P(\lambda_2) P(m)$$

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

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

$$\propto e^{-m\lambda_1} e^{-n\lambda_2} e^{m\lambda_2} \lambda_1^{\sum_{t=1}^{m} x_t} \lambda_2^{\sum_{t=m+1}^{n} x_t}$$

Note: this is not a kernel of a known distri.

Let's use Gibbs sampling ...

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

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

$$P(m \mid \text{——}) \propto e^{-(\lambda_1 - \lambda_2)m} \lambda_1^{\sum_{t=1}^{m} x_t} \lambda_2^{\sum_{t=m+1}^{n} x_t} = k(m \mid \text{—})$$

$\uparrow$ use grid sampling

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