Lecture 20

④ After $r_{ak} \approx 0$ we Thin out, we use every Kth iteration after B to have iid samples

⑤ repeat process for σ^2

let k for θ and k for σ^2 $= T$

- Result:
$$\left\{ \begin{array}{ccccc} l_1 & l_2 & l_3 & \cdots & L \\ \begin{bmatrix} \theta_B \\ \sigma^2_B \end{bmatrix}, & \begin{bmatrix} \theta_{B+T} \\ \sigma^2_{B+T} \end{bmatrix}, & \begin{bmatrix} \theta_{B+2T} \\ \sigma^2_{B+2T} \end{bmatrix}, & \ldots, & \begin{bmatrix} \theta_S \\ \sigma^2_S \end{bmatrix} \end{array} \right\} L \text{ elements}$$

* these values are as good as $P(\theta|X)$ and $P(\sigma^2|X)$

- Point estimation of Thinned samples:

$$\hat{\theta}_{MMSE} = E[\theta|X] \approx \bar{\theta} = \left[\frac{1}{L} \sum_{i=1}^{L} \theta_i \right]$$

$$\hat{\theta}_{MAE} = Med[\theta|X] \approx \text{quantile}[0.5, \theta_l]$$

Note: θ are ordered

$$\hat{\theta}_{MAD} = Mod[\theta|X] = \text{too hard computationally}$$

- Credible Region: $CR_{\theta;1-\alpha} = \left[\text{quantile}\left(\frac{\alpha}{2}, \theta_l\right), \text{quantile}\left(1-\frac{\alpha}{2}, \theta_l\right) \right]$

- $P_{val} := P(H_0|X) = P(\theta \geq \theta_0|X) = \frac{1}{L} \sum_{i=1}^{L} \mathbb{1}_{\theta_i \geq \theta_0}$

$H_0: \theta \geq \theta_0$

$H_\alpha: \theta < \theta_0$

Note: $\mathbb{1}_{\theta_i \geq \theta_0} = \begin{cases} 1 & \text{if } \theta_i \geq \theta_0 \\ 0 & \text{if } \theta_i < \theta_0 \end{cases}$

- Posterior Predictive distr.

$P(X^*|X)$

$= \int_{\Theta} P(X^*|\theta) P(\theta|X) \, d\theta$

by sampling

① Pick $l \in \{1, 2, \ldots, L\}$ to get θ_l

θ from Burned-in thin chain

② Draw X^* from $P(X^* | \theta = \theta_l)$

③ Repeat steps 1, 2 many times.

To Show That Gibb's sampling works; we need:

- Systemic Sweep Gibbs Sampler

Assume: Posterior $P\{\theta_1 \dots \theta_p \mid X\}$ is unknown

$P\{\theta_j \mid \theta_{-j}, X\}$ Known $\forall_j$

Note: θ_{-j} is all values of θ except θ_j

① Initialize: $\vec{\theta}_0 = < \theta_{1,0}, \theta_{2,0}, \dots, \theta_{p,0} >$, thinned chain

② sample: New $\theta_{\{1,2,\dots,p\},1}$

$\theta_{1,1}$ from $P(\theta_1 \mid \theta_{-1,0} \mid X)$

$\theta_{2,1}$ from $P(\theta_2 \mid \theta_{-2,0} \mid X)$

$\vdots$

$\theta_{p,1}$ from $P(\theta_p \mid \theta_{-p,0} \mid X)$

Note: $\theta_{a,b}$ — a: iteration, b: sample #

③ Return: $\vec{\theta}_1 = < \theta_{1,1}, \theta_{2,1}, \dots, \theta_{p,1} >$

④ Repeat step 2, 3 for

$\theta_{\{1,2,3\dots p\}, n+1}$ now Knowing $\vec{\theta}_{n=1}$

*Claim: $P(\theta_{\{1,2,\dots,p\},t} \mid X)$ & $\lim_{t \to \infty}$ $\to P(\theta_1, \theta_2, \dots, \theta_p \mid X)$

- Proof./ consider $X_0, X_1, X_2, \dots$ r.v.'s w/ common support $\mathcal{X}$

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

then $X_1, X_2, \dots$ is called a Markov Chain

its distribution is defined as:

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

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

Basically the Same

Note: $P(X_{t+1} \mid X_t)$ is Known

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

$\vdots$

$$P(X_{t+1}) = \int_{\mathcal{X}} \left(\prod_{i=1}^{t+1} P(X_i \mid X_{i-1}) \right) \underbrace{P(X_0)}_{\text{(Starting distribution)}}\, dx_i$$

- Thm for any starting distribution $P(X_0)$:

$$P(X) = \lim_{t\to\infty} \int_{\mathcal{X}} \left(\prod_{i=1}^{t+1} P(X_i \mid X_{i-1}) \right) P(X_0 = x)\, dx$$

* So it doesnt matter where you start eg: $P\left(\theta_{\{1,2,\dots,p\},0} \mid X\right)$
* we will always end up in the same distr. eg: $P(\theta_1, \dots \theta_p \mid X)$

* Steady state distribution:
long term distribution
equilibrium
etc.

- Joint density function : jdf of $P(X_1, \dots, X_p)$ has a positivity condition

$$\text{if } \left[\forall_j \; P(X_j) > 0\right] \;\&\; \left[\forall x_j \in \text{supp}[X_j]\right]$$

- Thm: $P(X_1, \dots, X_p)$ w/ positivity condition (Pos. cond.) ; $\forall \bar{a} \in \text{Supp}[X_j]$

$$P(X_1, \dots, X_p) \propto \prod_{j=1}^{p} \frac{P\left(X_j \mid X_1, X_2, \dots, X_{j-1}, X_{j+1} = a_{j+1}, \dots, X_p = a_p\right)}{P\left(X_j = a_j \mid X_1, X_2, \dots, X_{j-1}, X_{j+1} = a_{j+1}, \dots, X_p = a_p\right)}$$

- Corrallory: if $P(X_1 \dots X_p)$ has Pos. Cond.

$$\text{then } P\left(X_j \mid X_{-j}\right) > 0 \quad \forall x_j \in \mathcal{X}$$

* all conditional densities are non-zero