$$\frac{1}{S-1}\sum_{s=1}^{S}\underbrace{(\theta^{(s)}-\bar\theta)^2}_{\text{not iid}} = \frac{1}{S-1}\sum_{s=1}^{S}[\theta^{(s)}]^2 - \underbrace{\frac{S}{S-1}}_{\to 1}\bar\theta^2$$

By LLN,

$$\frac{1}{S-1}\sum_{s=1}^{S}[\theta^{(s)}]^2 = \underbrace{\frac{S}{S-1}}_{\to 1}\cdot\frac{1}{S}\sum_{s=1}^{S}[\theta^{(s)}]^2 \to E(\theta^2|y)$$

$$\bar\theta^2 \to [E(\theta|y)]^2$$

$$\frac{1}{S-1}\sum_{s=1}^{S}(\theta^{(s)}-\bar\theta)^2 \to E(\theta^2|y) - [E(\theta|y)]^2$$
$$= Var(\theta|y)$$

LLN $X_1,\cdots,X_n \overset{iid}{\sim} F$, mean of F is μ
var of F is $\sigma^2<\infty$

$$\frac{1}{n}\sum_{i=1}^{n}X_i \to \mu$$

CLT: $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right) \to N(0,\sigma^2)$

or roughly speaking, $\bar X = \frac{1}{n}\sum_{i=1}^{n}X_i \overset{approx.}{\sim} N(\mu,\frac{\sigma^2}{n})$

If $\theta^{(1)},\cdots,\theta^{(S)} \overset{iid}{\sim} p(\theta|y)$,

LLN $\frac{1}{S}\sum_{s=1}^{S}\theta^{(s)} \to E(\theta|y)$

CLT $\sqrt{S}\left(\underbrace{\frac{1}{S}\sum_{i=1}^{S}\theta^{(s)} - \underbrace{E(\theta|y)}_{(\mu)}}\right) \to N(0, \underbrace{Var(\theta|y)}_{(\sigma^2)<\infty})$

$$\frac{1}{S}\sum_{i=1}^{S}\theta^{(s)} \overset{approx.}{\sim} N\left(E(\theta|y), \frac{1}{S}\cdot Var(\theta|y)\right)$$

Monte Carlo method

$$\frac{1}{S}\sum_{s=1}^{S}\theta^{(s)} \to E(\theta|y)$$

posterior of $\theta|y$

$$\frac{1}{S}\sum_{s=1}^{S}g(\theta^{(s)}) \to E(g(\theta)|y) = \int g(\theta)\,p(\theta|y)\theta$$

predictive distribution

$$p(\tilde y|y) = \int \underbrace{p(\tilde y|\theta)}_{\to g(\theta)}\cdot p(\theta|y)\,d\theta \ \checkmark$$

$\theta^{(1)},\cdots,\theta^{(S)}$

$$= E_\theta[p(\tilde y|\theta)\,|\,y]$$
$$= E_{\theta|y}[p(\tilde y|\theta)] \approx \frac{1}{S}\sum_{s=1}^{S}p(\tilde y|\theta^{(s)})$$

$$p(\tilde y|y) = \int p(\tilde y|\theta)p(\theta|y)\,d\theta$$

$\tilde y, y$ independent given θ

$\Rightarrow p(\tilde y|\theta,y) = p(\tilde y|\theta)$

$$= \int p(\tilde y,\theta|\underbrace{y}_{\text{data}})\,d\theta$$

$\underbrace{y}_{\text{data}} \to \theta\sim p(\theta|y) \to \tilde y\sim p(\tilde y|\theta)$

This is equivalent to draw from $p(\tilde y,\theta|y)$

$(\theta^{(1)},\tilde y^{(1)}),\ \cdots,\ (\theta^{(S)},\tilde y^{(S)})$