

VE414 Lecture 13

Jing Liu

UM-SJTU Joint Institute

June 25, 2019

- Not to intentionally scrutinise a certain brand, but let us revisit our fish again



- Recall our analysis used the following made-up likelihoods

$$X_t \sim \text{Normal}(5, 1) \quad \text{and} \quad X_g \sim \text{Normal}(13, 9)$$

having observed $x = 7$, we pick the y value either maximises the likelihood

$$\mathcal{L}(y; 7) = (1 - y) \cdot f_{X_t}(7) + y \cdot f_{X_g}(7) \quad \text{over } y \in \{0, 1\}$$

or maximises the posterior $\mathcal{L}(y; 7) \cdot f_Y(y)$ over $y \in \{0, 1\}$.

- In practice, our understanding on the likelihood, e.g.

$$X \sim \text{Normal}(\mu, \sigma^2)$$

are obtained previously using some other data which are not available now.

- Suppose someone in the seafood market measures the length of each fish

$$\{x_1, \dots, x_n\}$$

which we assume to be independent. For simplicity, let us assume

$$\mu \sim \text{Normal}(\nu, \omega^2)$$

where ν and ω^2 are treated as knowns. If σ^2 is also known, then

$$\mu | \bar{x} \sim \text{Normal}\left(\frac{\omega^2 \bar{x} + \nu \sigma^2 / n}{\omega^2 + \sigma^2 / n}, \frac{\omega^2 \sigma^2 / n}{\omega^2 + \sigma^2 / n}\right), \quad \text{where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

can be obtained as we have done so in our discussion of conjugate priors.

- Of course, σ^2 is unavailable in practice, so let us model it as well as μ .
- Recall Jeffreys prior for unobserved Y and observed X is given by

$$f_Y(y) \propto \sqrt{I(y)}$$

where $I(y)$ is the variance of $\frac{\partial \ln \mathcal{L}(y; \textcolor{red}{X})}{\partial y}$ conditional on y

$$I(y) = \text{Var} \left[\frac{\partial \ln \mathcal{L}(y; \textcolor{red}{X})}{\partial y} \middle| y \right] = -\mathbb{E} \left[\frac{\partial^2 \ln \mathcal{L}(y; \textcolor{red}{X})}{\partial y^2} \middle| y \right]$$

- Given μ and $\{x_1, \dots, x_n\}$, the Jeffreys prior of $\sigma \in (0, \infty)$ is given by

$$\varphi_{\sigma^2} \propto \sigma^{-2}$$

which is an improper prior proposed by Jeffreys

$$\int_0^\infty \frac{1}{\sigma^2} d\sigma^2 \rightarrow \infty$$

- Therefore, we consider the following model

$$\begin{aligned} X \mid \{\mu, \sigma^2\} &\sim \text{Normal}(\mu, \sigma^2) \\ \mu &\sim \text{Normal}(\nu, \omega^2) \\ \sigma^2 &\sim \varphi_{\sigma^2} \end{aligned}$$

for which the joint posterior of μ and σ^2 is given by

$$\begin{aligned} f_{\{\mu, \sigma^2\} \mid \{x_1, \dots, x_n\}} &\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &\quad \cdot \frac{1}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{(\mu - \nu)^2}{2\omega^2}\right) \cdot \frac{1}{\sigma^2} \\ &\propto (\sigma^2)^{-(1+n/2)} \cdot \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2\right) - \frac{(\mu - \nu)^2}{2\omega^2}\right) \end{aligned}$$

Q: How can we sample from this posterior?

- Notice the sample mean \bar{x} and sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

are sufficient to capture everything in the data for this model since

$$\begin{aligned} f_{\{\mu, \sigma^2\} | \{x_1, \dots, x_n\}} &\propto (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) - \frac{(\mu - \nu)^2}{2\omega^2} \right) \\ &= (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{(\mu - \nu)^2}{2\omega^2} \right) \\ &= (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} \right) \cdot \exp \left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right) \\ &\quad \cdot \exp \left(-\frac{(\mu - \nu)^2}{2\omega^2} \right) \end{aligned}$$

- The prior μ and σ^2 are independent, but the posterior μ and σ are NOT.

- Since we are using an improper prior, we should double check whether

$$q_{\mu, \sigma^2} = (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{(\mu - \nu)^2}{2\omega^2} \right)$$

leads to any proper distribution before proceeding any further.

- The corresponding conditional distribution of μ is fine

$$\begin{aligned} f_{\mu | \{\sigma^2, \bar{x}, s\}} &= \frac{f_{\{\mu, \sigma^2\} | \{\bar{x}, s\}}}{f_{\sigma^2 | \{\bar{x}, s\}}} \propto f_{\{\mu, \sigma^2\} | \{\bar{x}, s\}} \\ &\propto q_{\mu, \sigma^2} \\ &\propto \exp \left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{(\mu - \nu)^2}{2\omega^2} \right) \end{aligned}$$

since it is identical to what we had in our discussion of conjugate priors,

$$\mu | \{\sigma^2, \bar{x}, s^2\} \sim \text{Normal} \left(\frac{\omega^2 \bar{x} + \nu \sigma^2 / n}{\omega^2 + \sigma^2 / n}, \frac{\omega^2 \sigma^2 / n}{\omega^2 + \sigma^2 / n} \right)$$

- Thus, if the marginal

$$\begin{aligned} f_{\sigma^2 | \{\bar{x}, s^2\}}(\sigma^2 | \bar{x}, s^2) &= \int_{-\infty}^{\infty} f_{\{\mu, \sigma^2\} | \{\bar{x}, s^2\}}(\mu, \sigma^2 | \bar{x}, s^2) d\mu \\ &\propto \int_{-\infty}^{\infty} q_{\mu, \sigma^2}(\mu, \sigma^2) d\mu = J(\sigma^2) \end{aligned}$$

is a proper distribution, that is, $\int_0^{\infty} J(\sigma^2) d\sigma^2 < \infty$, then the joint is proper

$$f_{\{\mu, \sigma^2\} | \{\bar{x}, s^2\}} = f_{\mu | \{\sigma^2, \bar{x}, s^2\}} \cdot f_{\sigma^2 | \{\bar{x}, s^2\}}$$

- Integrating over μ boils down to the following

$$\begin{aligned} J(\sigma^2) &= \int_{-\infty}^{\infty} (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2} - \frac{(\mu-\nu)^2}{2\omega^2}\right) d\mu \\ &= (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp(\beta) d\mu \end{aligned}$$

- Completing the following square

$$\beta = -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{(\mu - \nu)^2}{2\omega^2} = a\mu^2 + b\mu + c$$

where

$$a = -\frac{n}{2\sigma^2} - \frac{1}{2\omega^2}; \quad b = \frac{n\bar{x}}{\sigma^2} + \frac{\nu}{\omega^2}; \quad c = -\frac{n\bar{x}^2}{2\sigma^2} - \frac{\nu^2}{2\omega^2}$$

we have

$$\beta = a \cdot (\mu - h)^2 + k \quad \text{where} \quad h = -\frac{b}{2a} \quad \text{and} \quad k = c - \frac{b^2}{4a}$$

which gives the following

$$\begin{aligned} J(\sigma^2) &= (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} + k\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(\mu - h)^2}{2 \cdot (-1/(2a))}\right) d\mu \\ &= (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} + k\right) \cdot \sqrt{-\frac{\pi}{a}} \end{aligned}$$

- Rearranging in the following way,

$$\begin{aligned} J(\sigma^2) &= (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} + k\right) \cdot \sqrt{-\frac{\pi}{a}} \\ &= (\sigma^2)^{-(1+n/2)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\nu - \bar{x})^2}{2(n\omega^2 + \sigma^2)}\right) \sqrt{\frac{2\pi\omega^2\sigma^2}{n\omega^2 + \sigma^2}} \end{aligned}$$

and applying the comparison test with the following

$$f_Z(z) \propto (z)^{-(1+n/2)} \exp\left(-\frac{ns^2}{2z}\right)$$

which is a proper distribution, namely, scaled inverse chi-squared distribution

$$Z \sim \text{Scaled Inverse } \chi^2(n, s^2)$$

- Therefore, $\int_0^\infty J(\sigma^2) d\sigma^2 < \infty$, and the marginal thus the joint must exist.

- Given the joint exists,

$$f_{\{\mu, \sigma^2\} | \{\bar{x}, s^2\}} \propto (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{(\mu - \nu)^2}{2\omega^2} \right)$$

we need to check the conditionals in order to use Gibbs sampling,

$$\begin{aligned}\mu | \{\sigma^2, \bar{x}, s\} &\sim \text{Normal} \left(\frac{\omega^2 \bar{x} + \nu \sigma^2 / n}{\omega^2 + \sigma^2 / n}, \frac{\omega^2 \sigma^2 / n}{\omega^2 + \sigma^2 / n} \right) \\ \sigma^2 | \{\mu, \bar{x}, s\} &\sim \text{Scaled Inverse } \chi^2 \left(n, \frac{(n-1)s^2}{n} + (\bar{x} - \mu)^2 \right)\end{aligned}$$

- The conditional is also scaled inverse χ^2 , which is nonzero for $\sigma^2 \in (0, \infty)$

$$\begin{aligned}f_{\sigma^2 | \{\mu, \bar{x}, s\}} &= \frac{f_{\{\mu, \sigma^2\} | \{\bar{x}, s\}}}{f_{\mu | \{\bar{x}, s\}}} \propto f_{\{\mu, \sigma^2\} | \{\bar{x}, s\}} \\ &\propto (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right)\end{aligned}$$

- Therefore, we can reply on Gibbs sampling to sample the joint.

$$\mu^{(t)} \mid \{\sigma^{2(t-1)}, \bar{x}, s\} \sim \text{Normal} \left(\frac{\omega^2 \bar{x} + \nu \sigma^{2(t-1)}/n}{\omega^2 + \sigma^{2(t-1)}/n}, \frac{\omega^2 \sigma^{2(t-1)}/n}{\omega^2 + \sigma^{2(t-1)}/n} \right)$$

$$\sigma^{2(t)} \mid \{\mu^{(t)}, \bar{x}, s\} \sim \text{Scaled Inverse } \chi^2 \left(n, \frac{(n-1)s^2}{n} + (\bar{x} - \mu^{(t)})^2 \right)$$

Q: Given some initial value $\sigma^{2(0)}$, prior parameters ν , ω^2 , and data n , \bar{x} , s^2 , we can easily sample from the normal, how about the scaled inverse χ^2 ?

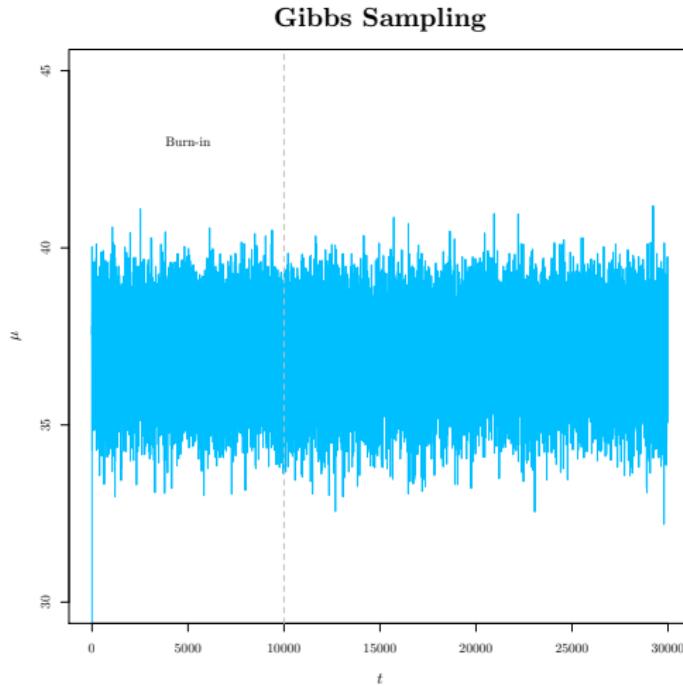
Q: You have seen what we need to do when uses Gibbs sampling, how about using Metropolis-Hastings algorithm? What do we need to do?

Q: Will the following proposal be a valid proposal?

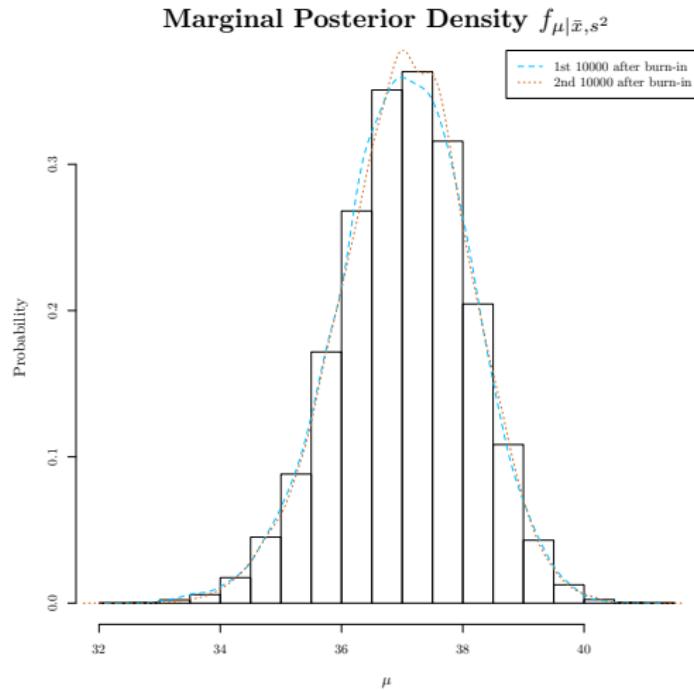
$$\begin{bmatrix} \mu^{(t)} & \sigma^{2(t)} \end{bmatrix}^T \sim \text{Normal} \left(\begin{bmatrix} \mu^{(t-1)} & \sigma^{2(t-1)} \end{bmatrix}^T, \begin{bmatrix} \gamma^2 & 0 \\ 0 & \gamma^2 \end{bmatrix} \right)$$

- The proposal will work, however, the parameter γ^2 needs to tuned at least.

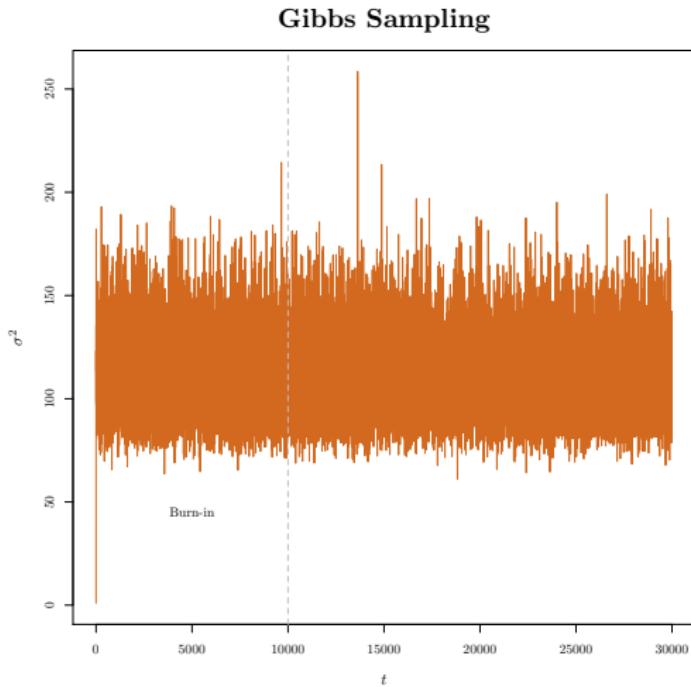
- For this simple model, the Markov Chain converges fairly quickly,



- Again, the following shows the Markov Chain has converged,

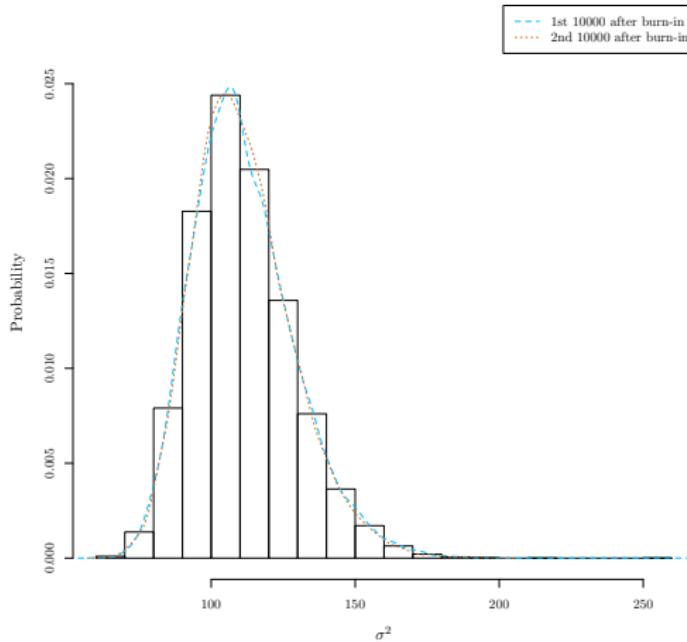


- For this simple model, the Markov Chain converges fairly quickly,



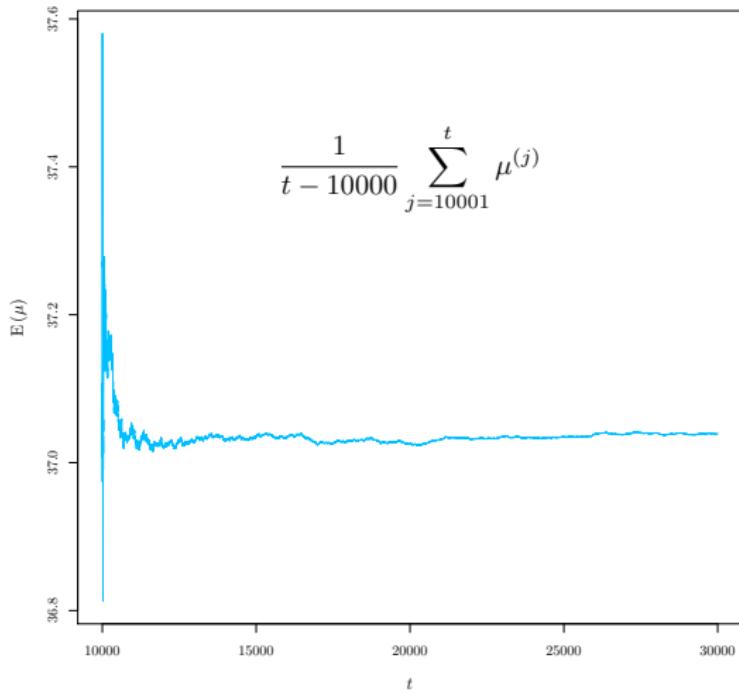
- Again, the following shows the Markov Chain has converged,

Marginal Posterior Density $f_{\sigma^2|\bar{x}, s^2}$



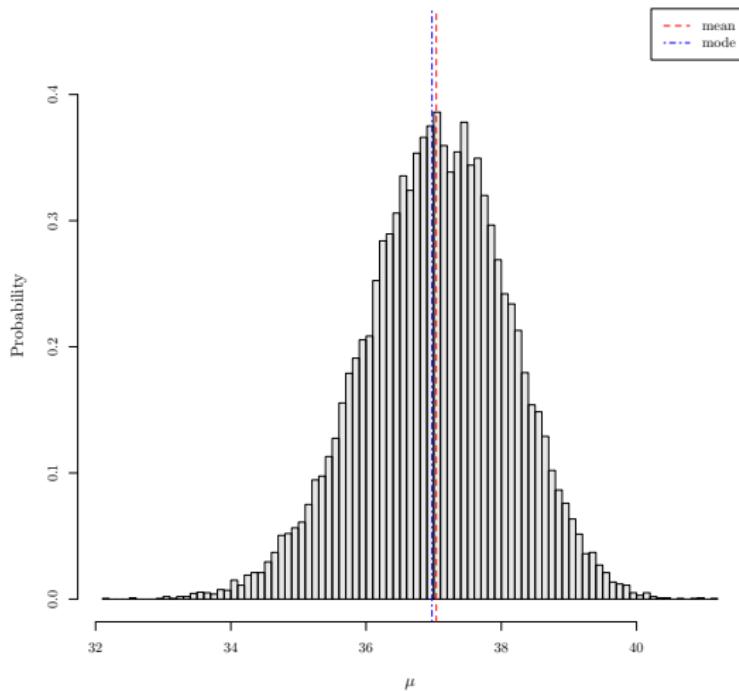
- Suppose we are primarily interested in μ , specifically, $f_{\mu|\{\bar{x}, s^2\}}$

Gibbs Sampling



Q: Is there any way other than MCMC we can use to obtain a point estimate

Marginal Posterior Density $f_{\mu|\bar{x}, s^2}$



- Note we could use numerical optimisations to find the mode in this case,

$$f_{\mu|\{\bar{x}, s^2\}} = \frac{f_{\{\mu, \sigma^2\}|\{\bar{x}, s^2\}}}{f_{\sigma^2|\{\mu, \bar{x}, s^2\}}}$$

since we have the joint up to a constant A depends only on \bar{x} and s^2 ,

$$\begin{aligned} Af_{\{\mu, \sigma^2\}|\{\bar{x}, s^2\}} &= q_{\mu, \sigma^2} \\ &= (\sigma^2)^{-(1+n/2)} \cdot \exp \left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2} - \frac{(\mu-\nu)^2}{2\omega^2} \right) \end{aligned}$$

and we have worked the full conditional

$$\sigma^2 | \{\mu, \bar{x}, s\} \sim \text{Scaled Inverse } \chi^2 \left(n, \frac{(n-1)s^2}{n} + (\bar{x}-\mu)^2 \right)$$

- In general, the marginal is intractable, or difficult to obtain for the very least

$$f_{\phi|X} = \int_{-\infty}^{\infty} f_{\{\phi, \gamma\}|X} d\gamma$$