Midterm - Redo.

Yifei Wang.

1. (a). $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

where A and B are events and $P(B) \neq 0$

(b). $z = \frac{\bar{\theta}_a - \bar{\theta}_b}{\sqrt{Var(\theta_a) + Var(\theta_b)}}$

where θ_a is the early interval of chain for parameter θ, and θ_b is the late interval. It describes the convergence of the MCMC chain.

(c). FALSE

(d) FALSE

(e) FALSE

2.

(a). Let $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$

We have $P(Y=Y) = \frac{1}{2\pi} \cdot |\Sigma_1|^{-1} \cdot exp\left(-\frac{1}{2}(y-\theta)^T \Sigma_1^{-1}(y-\theta)\right)$

For the posterior, we have,

$$P(\theta | Y, \Sigma_1) \propto P(Y|\theta, \Sigma_1) \cdot P(\theta)$$

$$\propto exp\left(-\frac{1}{2}(y-\theta)^T \Sigma_1^{-1}(y-\theta)\right)$$

Thus $\theta \sim N_2(Y, \Sigma_1)$

According to cholesky decomposition, we have

$$\Sigma_1 = LL^T \quad \text{where } L = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

We can sample $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$ and let. $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

$$\theta = y + LZ$$

(b).

$\theta_1 | \theta_2, y_1, y_2, \Sigma_1$ and $\theta_2 | \theta_1, y_1, y_2, \Sigma_1$ are conditional multivariate normal distribution, as we proved in part (a). They are normal distribution.

Let $\Lambda = \Sigma_1^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$

$$(y-\theta)^T \Sigma_1^{-1} (y-\theta)$$

$$= \frac{1}{1-\rho^2} \left[(y_1-\theta_1)^2 + (y_2-\theta_2)^2 - 2\rho (y_1-\theta_1)(y_2-\theta_2) \right]$$

$$= \frac{1}{1-\rho^2} \left[\theta_1^2 - 2(y_1 - y_2\rho \overset{+\rho\theta_2}{\vee})\theta_1 + (y_2-\theta_2)^2 + y_1^2 - 2\rho(y_1 y_2 - y_1\theta_2) \right]$$

We only cares about finding the mean and variance to determine a normal distribution, We have.

$$\sigma^2_{\theta_1|\theta_2} = 1-\rho^2$$

$$\mu_{\theta_1|\theta_2} = y_1 + \rho(\theta_2 - y_2)$$

Thus we have, $\theta_1 | \theta_2, y_1, y_2, \Sigma_1 \sim N(y_1 + \rho(\theta_2 - y_2), 1-\rho^2)$

Similarly, $\theta_2 | \theta_1, y_1, y_2, \Sigma_1 \sim N(y_2 + \rho(\theta_1 - y_1), 1-\rho^2)$

(c).

Part (a) has a bigger effective sample size and

(a) is more likely to provide a better approximation to the posterior.

because samples in (a) are independent.

(a) is MonteCarlo and

3.

(a). ① The data seems to have a bell shape and symmetric about some point, so it's possibly normal distributed.

② Some data are really far away from the high density region, which is very rare in a normal distribution, so normal might not be great here.

(b). This shows that the normal model could not capture well on everything. The data have a feature that the smallest observation is −44. However, none of the synthetic data have, or even close to, this feature. The normal model fails in capturing this feature, but it might be great at capturing the mean of the data.

(c) For s in 1...20:

Start with $\theta^{(0)}$

① sample one $\sigma^{2(s)}$ from $\sigma^2 | Y, \theta^{(s-1)} \sim$ InverseGamma

② Sample one $\mu^{(s)}$ from $\theta | Y, \sigma^{2(s)} \sim$ Normal

③ Sample 66 $\tilde{Y}_i^{(s)}$ from $\tilde{Y} | \theta^{(s)}, \sigma^{2(s)} \sim$ Normal$(\theta^{(s)}, \sigma^{2(s)})$

④ Compute and save the smallest value in $Y_i^{(s)}$ for i in 1...66 as $t_{min}^{(s)}$

Plot a histogram among all $t_{min}^{(s)}$ for S in 1...20

4. We know that
$$\theta_A = \mu + \delta \sim N(\mu_0 + \delta_0,\ \sigma_0^2 + \tau_0^2)$$
$$\theta_B = \mu - \delta \sim N(\mu_0 - \delta_0,\ \sigma_0^2 + \tau_0^2)$$

and
$$\mathrm{Cov}(\theta_A, \theta_B) = \mathrm{Cov}(\mu+\delta, \mu-\delta) = \mathrm{Cov}(\mu,\mu) + \mathrm{Cov}(\delta,\mu) - \mathrm{Cov}(\mu,\delta) - \mathrm{Cov}(\delta,\delta) = \sigma_0^2 - \tau_0^2$$

let $\sigma_A = \sigma_B = \sqrt{\sigma_0^2 + \tau_0^2}$, $\mu_A = \mu_0 + \delta_0$, $\mu_B = \mu_0 - \delta_0$
$$\rho = \frac{\mathrm{Cov}(A,B)}{\sigma_A \sigma_B} = \frac{\sigma_0^2 - \tau_0^2}{\sigma_0^2 + \tau_0^2}$$

and reparametrize θ_A, θ_B to be.
$$\begin{cases} \theta_A = \sigma_A Z_A + \mu_A \\ \theta_B = \sigma_B\left[\rho Z_A + \sqrt{1-\rho^2}\, Z_B\right] + \mu_B \end{cases}$$
where $Z_A, Z_B \overset{iid}{\sim} N(0,1)$

We can easily varify this reparametrization by identifying
$$E[\theta_A] = \mu_A = \mu_0 + \delta_0, \quad \mathrm{Var}[\theta_A] = \sigma_A^2 = \sigma_0^2 + \delta_0^2$$
$$E[\theta_B] = \mu_B = \mu_0 + \delta_0, \quad \mathrm{Var}[\theta_B] = \sigma_B^2 \rho^2 + \sigma_B^2(1-\rho^2) = \sigma_0^2 + \delta_0^2$$

We know the joint distribution for Z_A, Z_B is $f(Z_A, Z_B) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(Z_A^2 + Z_B^2)\right]$

Now we perform a ~~ch~~ multivariate change of variables on this.

We have
$$\begin{cases} Z_A = \dfrac{\theta_A - \mu_A}{\sigma_A} \\ Z_B = \dfrac{1}{\sqrt{1-\rho^2}}\left[\dfrac{\theta_B - \mu_B}{\sigma_B} - \rho\dfrac{\theta_A - \mu_A}{\sigma_A}\right] \end{cases}$$

and let
$$|J| = \det\begin{bmatrix} \dfrac{\partial Z_A}{\partial \theta_A} & \dfrac{\partial Z_A}{\partial \theta_B} \\ \dfrac{\partial Z_B}{\partial \theta_A} & \dfrac{\partial Z_B}{\partial \theta_B} \end{bmatrix} = \det\begin{bmatrix} \dfrac{1}{\sigma_A} & 0 \\ \dfrac{-\rho}{\sigma_A\sqrt{1-\rho^2}} & \dfrac{1}{\sigma_B\sqrt{1-\rho^2}} \end{bmatrix} = \frac{1}{\sigma_A \sigma_B \sqrt{1-\rho^2}}$$

Then the joint density of θ_A and θ_B is given by :

$$f(\theta_A, \theta_B) = f(Z_A, Z_B)\,|J|,$$

$$= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(Z_A^2 + Z_B^2)\right) \cdot \frac{1}{\sigma_A \sigma_B \sqrt{1-\rho^2}}$$

$$= \frac{1}{2\pi \sigma_A \sigma_B \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left[\left(\frac{\theta_A - \mu_A}{\sigma_A}\right)^2 + \frac{1}{1-\rho^2}\left(\frac{\theta_B - \mu_B}{\sigma_B} - \rho \frac{\theta_A - \mu_A}{\sigma_A}\right)^2\right]\right)$$

$$= \frac{1}{2\pi \sigma_A \sigma_B \sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(\theta_A - \mu_A)^2}{\sigma_A^2} + \frac{(\theta_B - \mu_B)^2}{\sigma_B^2} - 2\rho \frac{\theta_A - \mu_A}{\sigma_A} \frac{\theta_B - \mu_B}{\sigma_B}\right]\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \cdot |\Sigma_1|^{-1} \cdot \exp\left(-\frac{1}{2}(\theta - \bar{\mu})^T \Sigma_1^{-1} (\theta - \bar{\mu})\right)$$

where $\bar{\mu} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$, $\Sigma_1 = \begin{pmatrix} \sigma_A^2 & \rho\sigma_A\sigma_B \\ \rho\sigma_A\sigma_B & \sigma_B^2 \end{pmatrix}$

~~Thus we have~~ Thus the induced joint pri- on θ_A and θ_B

is bivariate normal :

$$\begin{bmatrix} \theta_A \\ \theta_B \end{bmatrix} \sim N_2\left(\begin{bmatrix} \mu_0 + \delta_0 \\ \mu_0 - \delta_0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 + \tau_0^2 & \sigma_0^2 - \tau_0^2 \\ \sigma_0^2 - \tau_0^2 & \sigma_0^2 + \tau_0^2 \end{bmatrix}\right)$$

5. (a). We could have different Poison models with different parameters (per-day email rate)

(b). $P(S_1 \cdots S_{10} | \theta) = \prod_{i=1}^{16} \frac{\exp(-\theta)\theta^{S_i}}{S_i!} \propto \theta^{\Sigma S_i} \exp(-10\theta)$, $S_i|\theta \sim \text{Poison}(\theta)$.

$P(P_1 \cdots P_{20} | \theta) = \prod_{i=1}^{20} \frac{\exp(-10\theta)(10\theta)^{P_i}}{P_i!} \propto \theta^{\Sigma P_i} \exp(-200\theta)$, $P_i|\theta \sim \text{Poison}(10\theta)$

(c). $P(\theta|S,P) \propto P(S,P|\theta) \cdot P(\theta)$

$= P(S|\theta)\, P(P|\theta)\, P(\theta)$

$\propto \theta^{C_1} \exp(-C_2\theta) \cdot P(\theta)$.

Thus $P(\theta)$ should at least have θ^{C_1} and $\exp(-C_2\theta)$, and a gamma prior with parameters α, β could satisfy this: $\theta \sim \text{Gamma}(\alpha, \beta)$.

(d). $\text{mean}(5,7,9) = 7$ $\quad \text{Var}(5,7,9) = \frac{8}{3}$

If in the literature we find that these three studies have similar sample sizes, and we want to have a relatively weak prior, we can set prior mean $\frac{\alpha}{\beta} = 7$ and prior variance $\frac{\alpha}{\beta^2} = \frac{8}{3}$. This gives us $\begin{cases} \alpha = \frac{147}{8} \\ \beta = \frac{21}{8} \end{cases}$

(e). $P(\theta|S,P) \propto P(S|\theta)\, P(P|\theta)\, P(\theta)$

$\propto \theta^{20}\exp(-10\theta) \cdot \theta^{1200}\exp(-200\theta) \cdot \theta^{\frac{147}{8}-1}\exp(-\frac{21}{8}\theta)$

$= \theta^{1280+\frac{147}{8}-1} \exp(-(210+\frac{21}{8})\theta)$

Thus, $\theta|S,P \sim \text{Gamma}(1280+\frac{147}{8}, 210+\frac{21}{8})$

$E[\theta|S,P] = (1280+\frac{147}{8}) / (210+\frac{21}{8})$