Homework 3
of
STAT 632 Bayesian Statistics

Yanxun Xu

October 16, 2015

1 Problem 1

Solution:

As Σ is known, the prior distribution of $\boldsymbol{\mu}$ is

$$
\begin{aligned}
p(\boldsymbol{\mu}) &= (2\pi)^{-p/2}|\Lambda_0|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_0})^T\Lambda_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_0})\right)\\
&= (2\pi)^{-p/2}|\Lambda_0|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\Lambda_0^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\Lambda_0^{-1}\boldsymbol{\mu_0}-\frac{1}{2}\boldsymbol{\mu_0}^T\Lambda_0^{-1}\boldsymbol{\mu_0}\right)\\
&\propto \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\Lambda_0^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\Lambda_0^{-1}\boldsymbol{\mu_0}\right)\\
&= \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A_0}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{b_0}\right),
\end{aligned}
$$

where $\boldsymbol{A_0}=\Lambda_0^{-1}$ and $\boldsymbol{b_0}=\Lambda_0^{-1}\boldsymbol{\mu_0}$.

The data likelihood of Y is

$$
\begin{aligned}
p(\boldsymbol{Y}|\boldsymbol{\mu},\Sigma) &= \prod_{i=1}^{n}(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{y_i}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y_i}-\boldsymbol{\mu})\right)\\
&= (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y_i}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y_i}-\boldsymbol{\mu})\right)\\
&\propto \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A_1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{b_1}\right),
\end{aligned}
$$

where $\boldsymbol{A_1}=n\Sigma^{-1}$ and $\boldsymbol{b_1}=n\Sigma^{-1}\bar{\boldsymbol{y}}$.

According to Bayes' Rule, the posterior distribution can be written as

$$
\begin{aligned}
p(\boldsymbol{\mu}|\boldsymbol{Y},\Sigma) &\propto p(\boldsymbol{\mu})*p(\boldsymbol{Y}|\boldsymbol{\mu},\Sigma)\\
&= \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A_1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{b_1}\right)*\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A_0}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{b_0}\right)\\
&= \exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A_n}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{b_n}\right),
\end{aligned}
$$

where $\boldsymbol{A_n}=\boldsymbol{A_0}+\boldsymbol{A_1}=\Lambda^{-1}+n\Sigma^{-1}$ and $\boldsymbol{b_n}=\boldsymbol{b_0}+\boldsymbol{b_1}=\Lambda^{-1}\boldsymbol{\mu_0}+n\Sigma^{-1}\bar{\boldsymbol{y}}$.

2 Problem 2

Solution:

The importance sampling approach is based on the principle that

$$E[x] = \frac{E_g[x\frac{f(x)}{g(x)}]}{E_g[\frac{f(x)}{g(x)}]}.$$

Here, we chose $g \sim N(0,1)$ as our density function as it satisfies the corresponding requirements. Using the following code, we obtained $E[x] = 0.3587$.

```
set.seed(1);
sample.size = 10000;
x <- rep(NA, sample.size);
f <- rep(NA, sample.size);
g <- rep(NA, sample.size);
x <- rnorm(sample.size);
g <- dnorm(x);
for (i in 1:sample.size) {
  temp <- runif(1, 0, 1);
  if (temp <= 0.3) {
    f[i] = dbeta(x[i], 5, 2);
  } else {
    f[i] = dbeta(x[i], 2, 8);
  }
}
mu.is = sum(x*f/g)/sum(f/g);
```

Using the following code, we computed $P(0.45 \leq x \leq 0.55) = 0.0482$.

```
for (i in 1:sample.size) {
  temp <- runif(1, 0, 1);
  if (temp <= 0.3) {
    f[i] = rbeta(1, 5, 2);
  } else {
    f[i] = rbeta(1, 2, 8);
  }
}
sum(f >= 0.45 & f <= 0.55)/sample.size;
```

3 Problem 3

Solution:

Suppose $x \sim U_{[0;Mg(x)]}$, we have

$$\begin{aligned} P(U < f(x)) &= \int_0^{f(x)} \frac{1}{Mg(x)} du \\ &= \int_{-\infty}^{\infty} \frac{f(x)}{Mg(x)} g(x) dx \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{M}\int_{-\infty}^{\infty} f(x)dx \\
&= \frac{1}{M},
\end{aligned}
$$

and

$$
\begin{aligned}
P(U < f(x), X < x) &= \int_{-\infty}^{x}\int_{0}^{f(x)} \frac{1}{Mg(x)}dudx \\
&= \int_{-\infty}^{x} \frac{f(x)}{Mg(x)}g(x)dx \\
&= \frac{1}{M}\int_{-\infty}^{x} f(x)dx \\
&= \frac{F(x)}{M}.
\end{aligned}
$$

Thus, we can obtain

$$
\begin{aligned}
P(X < x|U < f(x)) &= \frac{P(U < f(x), X < x)}{P(U < f(x))} \\
&= \frac{F(x)/M}{1/M} \\
&= F(x) \\
&= P(X < x) \\
&= P(Y < y),
\end{aligned}
$$

that can show the algorithm is equivalent to the Accept-Reject algorithm.

4 Problem 4

Solution:

The data likelihood of X is

$$p(X|\mu,\tau) \sim \prod_{i=1}^{n} normal(x_i;\mu,\tau).$$

The prior distribution of μ and τ is

$$p(\mu,\tau) \sim beta(\mu;2,2)lognormal(\tau;1,10).$$

The posterior distribution of μ and τ is

$$p(\mu,\tau|X) \sim beta(\mu;2,2)lognormal(\tau;1,10)\prod_{i=1}^{n} normal(x_i;\mu,\tau).$$

Assume the proposal density is symmetric, at iteration t, $\mu^{(t)}$ and $\tau^{(t)}$ are replaced with μ^* and τ^* with probability λ or remain the previous values $\mu^{(t-1)}$ and $\tau^{(t-1)}$ with probability $1-\lambda$, where

$$\lambda = \min{(1, \frac{p(\mu^*,\tau^*|X)}{p(\mu^{(t)},\tau^{(t)}|X)})}.$$

Using the following code, we computed the posterior probability $p(\mu \geq 0.5|X) = 0.8185$. The corresponding full period and after burn-in period density and trace plots in the parameter space and the autocorrelation plots of μ and τ are given in Figure 1.

```
set.seed(1);
x <- c(2.3656491, 2.4952035, 1.0837817, 0.7586751, 0.8780483, 1.2765341, 1.4598699,
 0.1801679, -1.0093589, 1.4870201, -0.1193149, 0.2578262);
sample.size = 100000;
mu <- rep(NA, sample.size);
tao <- rep(NA, sample.size);
p <- rep(NA, sample.size);
mu[1] <- rbeta(1, 2, 2);
tao[1] <- rlnorm(1, 1, 10);
p[1] <- 1/sqrt(tao[1]*2*pi)^length(x)*exp(-sum((x-mu[1])^2)/(2*tao[1]))*
 dbeta(mu[1], 2, 2)*dlnorm(tao[1], 1, 10);
count = 0;
for (i in 1:sample.size) {
  mu_new <- rbeta(1, 2, 2);
  tao_new <- rlnorm(1, 1, 10);
  p_new <- 1/sqrt(tao_new*2*pi)^length(x)*exp(-sum((x-mu_new)^2)/(2*tao_new))*
   dbeta(mu_new, 2, 2)*dlnorm(tao_new, 1, 10);
  r = min(p_new/p[i], 1);
  test <- runif(1);
  if(test < r) {
    mu[i+1] <- mu_new;
    tao[i+1] <- tao_new;
    p[i+1] <- p_new;
    count = count+1;
  }
  else {
    mu[i+1] <- mu[i];
    tao[i+1] <- tao[i];
    p[i+1] <- p[i];
  }
}
sum(mu >= 0.5)/sample.size;
par(mfrow = c(3, 2));
plot(mu, tao);
plot(mu, tao, type = "l")
plot(mu[50000:100001], tao[50000:100001]);
plot(mu[50000:100001], tao[50000:100001], type = "l");
acf(mu);
acf(tao);
```

5 Problem 6.2

Solution:

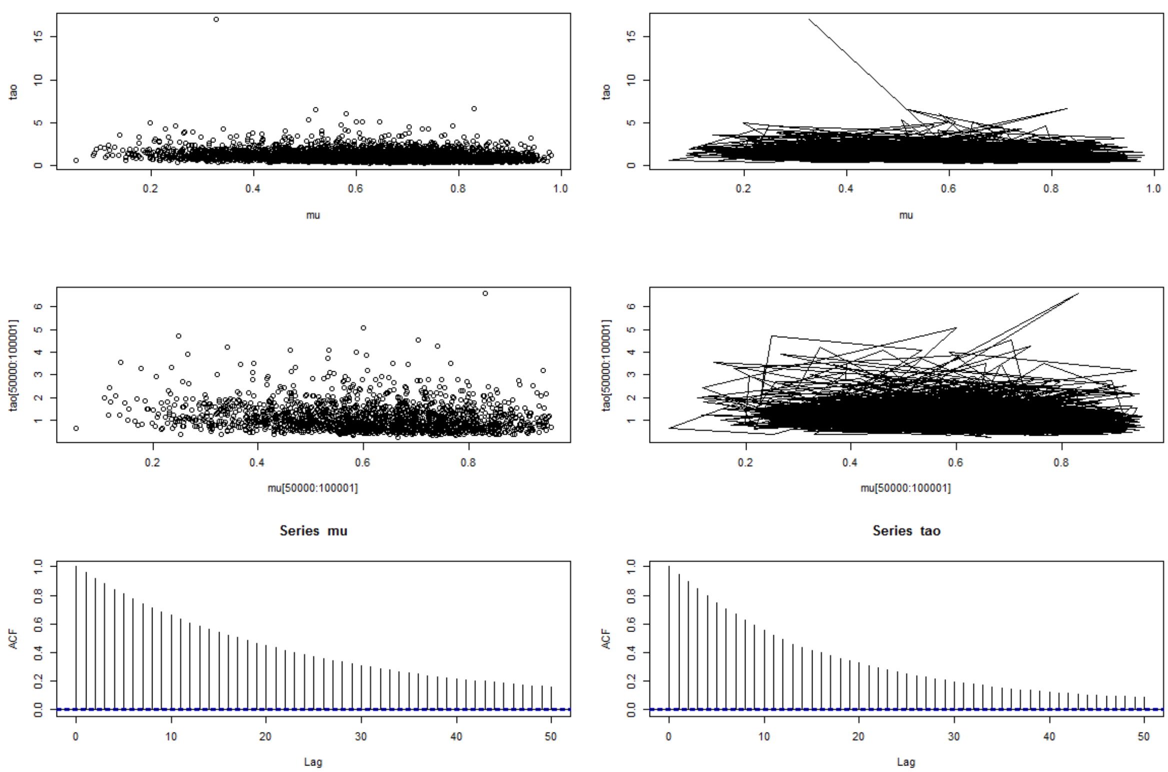

Figure 1: (Top) The full period density and trace plots in the parameter space; (Middle) The after burn-in period density and trace plots in the parameter space; (Bottom) The autocorrelation plots of μ and τ

(a) After uploading the `glucose` data set, we made a histogram of the data as shown in Figure 2, which is highly right-skewed. The code is given as

```
glucose = read.table("glu.txt", header = FALSE);
y <- glucose[,1];
hist(y, freq=FALSE);
lines(density(y), col = 2, lwd = 2);
```

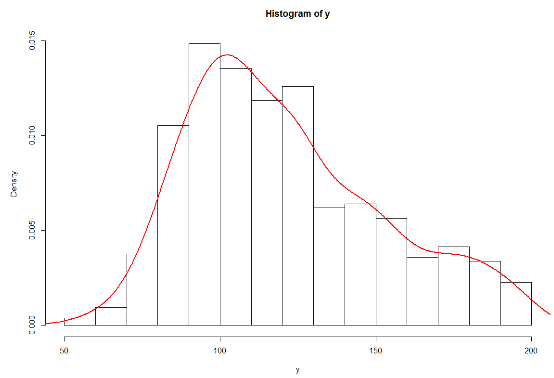

Figure 2: The histogram of the data

(b) Let $n_1 = \sum_{\{i:X_i=1\}} 1$, $n_2 = \sum_{\{i:X_i=2\}} 1 = n - n_1$, $\bar{y}_1 = \sum_{\{i:X_i=1\}} y_i$, and $\bar{y}_2 = \sum_{\{i:X_i=2\}} y_i$. Obviously, according to the model description, we could obtain the full conditional distributions of each parameter as

$$p(X_i = 1|p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, Y) = \frac{p * normal(y_i; \theta_1, \sigma_1^2)}{p * normal(y_i; \theta_1, \sigma_1^2) + (1-p) * normal(y_i; \theta_2, \sigma_2^2)}, i = 1, \cdots, n,$$

and

$$p(X|p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, Y) = \prod_{i=1}^{n} p(X_i|p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, Y);$$

$$p(p|X, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, Y) \sim beta(p; a + n_1, b + n_2);$$

$$p(\theta_1|X, p, \theta_2, \sigma_1^2, \sigma_2^2, Y) \sim normal(\mu_n, \tau_n^2),$$

where $\mu_n = \frac{\mu_0/\tau_0^2 + n_1\bar{y}_1/\sigma_1^2}{1/\tau_0^2 + n_1/\sigma_1^2}$ and $\tau_n^2 = \frac{1}{1/\tau_0^2 + n_1/\sigma_1^2}$;

$$p(\theta_2|X, p, \theta_1, \sigma_1^2, \sigma_2^2, Y) \sim normal(\mu_n, \tau_n^2),$$

where $\mu_n = \frac{\mu_0/\tau_0^2 + n_2\bar{y}_2/\sigma_2^2}{1/\tau_0^2 + n_2/\sigma_2^2}$ and $\tau_n^2 = \frac{1}{1/\tau_0^2 + n_2/\sigma_2^2}$;

$$p(\sigma_1^2|X, p, \theta_1, \theta_2, \sigma_2^2, Y) \sim inverse - gamma(\nu_n/2, \nu_n\sigma_n^2/2),$$

where $\nu_n = \nu_0 + n_1$ and $\sigma_n^2 = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + \sum_{\{i:X_i=1\}}(y_i - \theta_1)^2)$;

$$p(\sigma_2^2|X, p, \theta_1, \theta_2, \sigma_1^2, Y) \sim inverse - gamma(\nu_n/2, \nu_n\sigma_n^2/2),$$

where $\nu_n = \nu_0 + n_2$ and $\sigma_n^2 = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + \sum_{\{i:X_i=2\}}(y_i - \theta_2)^2)$.

(c) Using the following code, we computed and plotted the autocorrelation of $\theta_{(1)}^{(s)}$ and $\theta_{(2)}^{(s)}$, as shown in Figure 3. Their effective sample sizes are 420.4008 and 216.5206, respectively.

```
set.seed(1);
n <- dim(glucose)[1];
sample.size <- 10000;
x <- matrix(NA, sample.size, n);
p <- rep(NA, sample.size);
theta_1 <- rep(NA, sample.size);
theta_2 <- rep(NA, sample.size);
sigma2_1 <- rep(NA, sample.size);
sigma2_2 <- rep(NA, sample.size);
a <- 1;
b <- 1;
mu_0 <- 120;
tao2_0 <- 200;
sigma2_0 <- 1000;
nu_0 <- 10;
p[1] <- rbeta(1, 1, 1);
x[1, ] <- rbinom(n, 1, p[1]);
theta_1[1] <- rnorm(1, mu_0,sqrt(tao2_0));
theta_2[1] <- rnorm(1, mu_0,sqrt(tao2_0));
sigma2_1[1] <- 1/rgamma(1, nu_0/2, nu_0*sigma2_0/2);
sigma2_2[1] <- 1/rgamma(1, nu_0/2, nu_0*sigma2_0/2);
for (i in 2:sample.size) {
  # Update x
  for (j in 1:n) {
    temp_1 <- p[i-1]*dnorm(y[j], theta_1[i-1], sqrt(sigma2_1[i-1]));
    temp_2 <- (1-p[i-1])*dnorm(y[j], theta_2[i-1], sqrt(sigma2_2[i-1]));
    x[i,j] <- rbinom(1, 1, temp_1/(temp_1+temp_2));
  }
  # Update p
  c <- sum(x[i,]);
  p[i] <- rbeta(1, a+c, b+n-c);
  # Update theta_1
  y_1.bar <- mean(y[x[i,] == 1]);
  mu_n <- (mu_0/tao2_0+c*y_1.bar/sigma2_1[i-1])/(1/tao2_0+c/sigma2_1[i-1]);
  tao2_n <- 1/(1/tao2_0+c/sigma2_1[i-1]);
  theta_1[i] <- rnorm(1, mu_n, sqrt(tao2_n));
  # Update sigma2_1
  nu_n <- nu_0+c;
  s2_n <- sum((y[x[i,] == 1]-theta_1[i])^2)/c;
  sigma2_n <- (nu_0*sigma2_0+c*s2_n)/nu_n;
```

```
  sigma2_1[i] <- 1/rgamma(1,nu_n/2,nu_n*sigma2_n/2);
  # Update theta_2
  y_2.bar <- mean(y[x[i,] == 0]);
  mu_n <- (mu_0/tao2_0+(n-c)*y_2.bar/sigma2_2[i-1])/(1/tao2_0+(n-c)/sigma2_2[i-1]);
  tao2_n <- 1/(1/tao2_0+(n-c)/sigma2_2[i-1]);
  theta_2[i] <- rnorm(1, mu_n, sqrt(tao2_n));
  # Update sigma2_2
  nu_n <- nu_0+(n-c);
  s2_n <- sum((y[x[i,] == 0]-theta_2[i])^2)/(n-c);
  sigma2_n <- (nu_0*sigma2_0+(n-c)*s2_n)/nu_n;
  sigma2_2[i] <- 1/rgamma(1,nu_n/2,nu_n*sigma2_n/2);
}
theta_11 <- rep(NA, sample.size);
theta_22 <- rep(NA, sample.size);
for (i in 1:sample.size) {
  theta_11[i] <- min(theta_1[i], theta_2[i]);
  theta_22[i] <- max(theta_1[i], theta_2[i]);
}
acf(theta_11);
acf(theta_22);
effectiveSize(theta_11);
effectiveSize(theta_22);
```

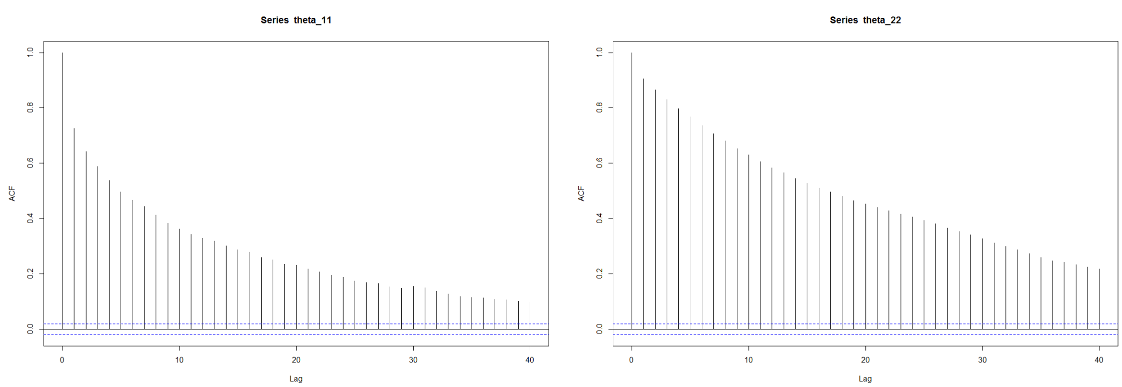

Figure 3: The autocorrelation plots of (Left) $\theta_{(1)}^{(s)}$ and (Right) $\theta_{(2)}^{(s)}$

(d) We made a histogram of the new data as shown in Figure 4, which is as right-skewed as the one in part (a). However, in order to discuss the adequacy of this two data, we should draw the Q-Q plot, as shown in Figure 5, which reveals that the two data has almost the same distribution. The code is given as

```
x_new <- rbinom(length(p), 1, p);
y_new <- rep(NA, sample.size);
for (i in 1:sample.size) {
  if (x_new[i] == 1) {
    y_new[i] <- rnorm(1, theta_1[i], sqrt(sigma2_1[i]));
  } else {
```

```
    y_new[i] <- rnorm(1, theta_2[i], sqrt(sigma2_2[i]));
  }
}
hist(y_new, freq=FALSE);
lines(density(y), col = 2, lwd = 2);
qqplot(y, y_new);
```

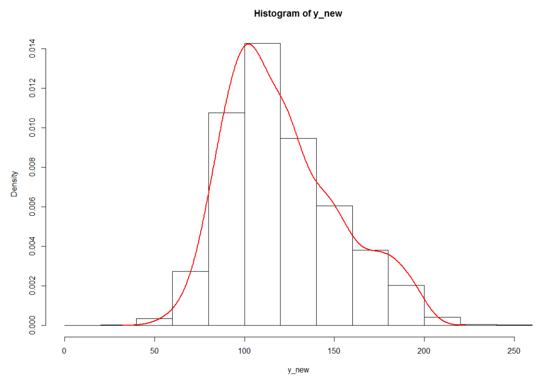

Figure 4: The histogram of the new data

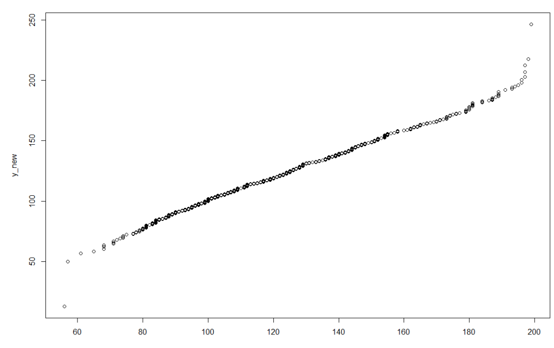

Figure 5: Q-Q plot of the original and new data

6 Problem 7.1

Solution:

(a) As

$$\int\int p_J(\theta,\Sigma)d\theta d\Sigma \propto \int\int |\Sigma|^{-(p+2)/2}d\theta d\Sigma = \infty,$$

which contradicts with the fact that the integration of probability density should be 1. Thus, p_J cannot be a probability density for (θ, Σ).

(b)

$$\begin{aligned}
p(\boldsymbol{\theta}, \Sigma|\boldsymbol{Y}) &\propto p(\boldsymbol{\theta}, \Sigma) \times p(\boldsymbol{Y}|\boldsymbol{\theta}, \Sigma) \\
&\propto |\Sigma|^{-\frac{p+2}{2}} \prod_{i=1}^{n} (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y_i} - \theta)^T \Sigma^{-1} (\boldsymbol{y_i} - \theta)\right) \\
&= |\Sigma|^{-\frac{p+2}{2}} (2\pi)^{-\frac{nk}{n}} |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y_i} - \theta)^T \Sigma^{-1} (\boldsymbol{y_i} - \theta)\right) \\
&\propto |\Sigma|^{-(\frac{p+n}{2}+1)} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y_i} - \theta)^T \Sigma^{-1} (\boldsymbol{y_i} - \theta)\right),
\end{aligned}$$

where

$$\begin{aligned}
\sum_{i=1}^{n}(\boldsymbol{y_i} - \theta)^T \Sigma^{-1} (\boldsymbol{y_i} - \theta) &= tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{y_i} - \theta)(\boldsymbol{y_i} - \theta)^T\right) \\
&= tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{y_i} - \bar{\boldsymbol{y}} + \bar{\boldsymbol{y}} - \theta)(\boldsymbol{y_i} - \bar{\boldsymbol{y}} + \bar{\boldsymbol{y}} - \theta)^T\right) \\
&= tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{y_i} - \bar{\boldsymbol{y}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^T + 2(\boldsymbol{y_i} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}} - \theta) + (\bar{\boldsymbol{y}} - \theta)(\bar{\boldsymbol{y}} - \theta)^T\right) \\
&= tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{y_i} - \bar{\boldsymbol{y}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^T + n(\bar{\boldsymbol{y}} - \theta)\Sigma^{-1}(\bar{\boldsymbol{y}} - \theta)^T\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
p(\boldsymbol{\theta}, \Sigma|\boldsymbol{Y}) &\propto |\Sigma|^{-(\frac{p+n}{2}+1)} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}S) - \frac{1}{2}n(\bar{\boldsymbol{y}} - \theta)\Sigma^{-1}(\bar{\boldsymbol{y}} - \theta)^T\right) \\
&= |\Sigma|^{-(\frac{p+n}{2}+1)} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}S)\right) \exp\left(-\frac{1}{2}(\theta - \bar{\boldsymbol{y}})(\Sigma/n)^{-1}(\theta - \bar{\boldsymbol{y}})^T\right) \\
&\propto Inverse - Wishart(n-1, S) \times N(\bar{\boldsymbol{y}}, \Sigma/n),
\end{aligned}$$

where $S = \sum_{i=1}^{n}(\boldsymbol{y_i} - \bar{\boldsymbol{y}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^T$.

Then,

$$\begin{aligned}
p(\Sigma|\boldsymbol{Y}) &\propto Inverse - Wishart(n-1, S) \\
p(\boldsymbol{\theta}|\Sigma, \boldsymbol{Y}) &\propto N(\bar{\boldsymbol{y}}, \Sigma/n).
\end{aligned}$$