Problem Sheet 1: Bayesian Theory

In the lectures, we showed how Bayesians use rules of conditional probability to learn about parameters in the model. In particular, the posterior is proportional to the prior times the likelihood function. Some questions in this problem sheet gives you some priors and likelihoods and asks you to work out posteriors. This allows you to get familiar with working with many common probability density functions. Appendix B of the textbook book contains definitions and properties of these and many more common distributions. The general solution strategy is to multiply the prior and the likelihood together. The result will be proportional to the posterior. One then examines (usually after re-arranging) this posterior to see if it belongs to any common class of distributions.

As another hint, note that when working with probability density functions we can usually ignore constant terms which do not involve the random variable. So, for instance, if x is a random variable with p.d.f. $p(x)$ then full knowledge of $p(x)$ is not always required. Writing $p(x) = cf(x)$, where c does not depend on x, for most derivations knowledge of $f(x)$ is all that is required. $f(x)$ is referred to as the kernel and c the integrating constant. So, for instance, you often see proofs where the researcher derives $f(x)$ and then says "this is the kernel of a Normal (or Gamma, etc.) p.d.f., hence it follows that x has a Normal distribution".

Exercise 1

Given the parameter θ where $0 < \theta < 1$, consider T i.i.d. Bernoulli random variables Y_t $(t = 1, 2, \cdots, T)$ each with:

$$p(y_t|\theta) = \begin{cases} \theta & \text{if } y_t = 1 \\ 1-\theta & \text{if } y_t = 0. \end{cases} \tag{1}$$

The likelihood function is, thus, $\prod_{t=1}^{T} p(y_t|\theta)$, which is

$$p(y|\theta) = \theta^m (1-\theta)^{T-m}, \tag{2}$$

where $m = T\overline{y}$ is the number of successes (i.e., $y_t = 1$) in T trials. Suppose prior beliefs concerning θ are represented by a Beta distribution (see definition in textbook Appendix B) with p.d.f.

$$p(\theta) \propto \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1}, \quad 0 < \theta < 1, \tag{3}$$

where $\underline{\alpha} > 0$ and $\underline{\delta} > 0$ are known. This class of priors can represent a wide range of prior opinions. Find the posterior density of θ.

Solution to Exercise 1

The posterior is obtained by multiplying (2) by (3):

$$\begin{aligned} p(\theta|y) &\propto \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1}\theta^m(1-\theta)^{T-m} \\ &= \theta^{\overline{\alpha}}(1-\theta)^{\overline{\delta}-1}, \quad 0 < \theta < 1. \end{aligned}$$

where

$$\begin{aligned}\overline{\alpha} &= \underline{\alpha} + m,\\ \overline{\delta} &= \underline{\delta} + T - m,\end{aligned}$$

Examing the form for the posterior, it can be seen to be a Beta p.d.f. with parameters $\overline{\alpha}$ and $\overline{\delta}$. Note, that, since prior and posterior both have Beta p.d.f.'s, that the conjugate family of prior distributions for a Bernoulli likelihood is the Beta family of p.d.f.'s

Exercise 2

A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for one percent of the healthy persons tested. If .1 percent of the population actually has the disease, what is the probability that a person has the disease given that her test result is positive?

Solution to Exercise 2

This problem can be solved by using Bayes' theorm. Let D denote the presence of the disease, D^c its absence and + denote a positive test result. Then $P(+|D) = .95$, $P(+|D^c) = .01$ and $P(D) = .001$. Then according to Bayes Theorem

$$P(D|+) = \frac{P(D)P(+|D)}{P(+)} = \frac{.001(.95)}{.001(.95) + .999(.01)} = .0868.$$

Exercise 3

Consider a random sample y_t $(t = 1, 2, \cdots T)$ from a $N(\theta_1, \theta_2^{-1})$ population. For reasons that will become clear as we proceed, it is convenient to work in terms of θ_2, the reciprocal of the variance (called the ***precision***). Assume θ_2 is known. Suppose prior beliefs for θ_1 are represented by the Normal distribution

$$\theta_1 \sim N(\underline{\mu}, \underline{h}^{-1}), \tag{4}$$

where $\underline{\mu}$ and $\underline{h} > 0$ are given. Find the posterior density of θ_1 as well as the marginal likelihood.

Solution to Exercise 3

This can be interpreted as a simplied version of the simple regression model: it has only an intercept and the error variance is known. For later notational convenience, let

$$\begin{aligned} h &= [\theta_2^{-1}/T]^{-1} = T\theta_2 && (5)\\ \overline{h} &= \underline{h} + h && (6)\\ \overline{\mu} &= \overline{h}^{-1}(\underline{h}\underline{\mu} + h\overline{y}). && (7)\end{aligned}$$

It is useful to employ two identities so we can write things in terms of OLS quantities. The first identity is

$$\sum_{t=1}^{T}(y_t-\theta_1)^2=\sum_{t=1}^{T}(y_t-\overline{y})^2+T(\overline{y}-\theta_1)^2=\nu s^2+T(\overline{y}-\theta_1)^2 \tag{8}$$

where

$$\nu \;=\; T-1 \tag{9}$$

$$s^2 \;=\; \nu^{-1}\sum_{t=1}^{T}(y_t-\overline{y})^2. \tag{10}$$

The second identity is:

$$\underline{h}(\theta_1-\underline{\mu})^2+h(\overline{y}-\theta_1)^2=(\theta_1-\overline{\mu})^2+(\underline{h}^{-1}+h^{-1})^{-1}(\overline{y}-\underline{\mu})^2. \tag{11}$$

Now we apply Bayes Theorem to find the posterior density of θ_1. Using identity (8) and letting $\phi(y_t|\theta_1,\theta_2^{-1})$ denote the Normal p.d.f. with mean θ_1 and variance θ_2^{-1}, write the likelihood function as

$$\begin{aligned}
p\,(y|\theta_1) &= \prod_{t=1}^{T}\phi(y_t|\theta_1,\theta_2^{-1}) \qquad (12)\\
&= (2\pi\theta_2^{-1})^{-T/2}\exp\left(\sum_{t=1}^{T}(y_t-\theta_1)^2\right)\\
&= (2\pi\theta_2^{-1})^{-T/2}\exp\left(-\frac{h}{2T}\nu s^2+T(\overline{y}-\theta_1)^2\right)\\
&= c_1\phi(\overline{y}|\theta_1,h^{-1}),
\end{aligned}$$

where

$$c_1=(2\pi)^{\nu/2}T^{(-1/2)}\theta_2^{\nu/2}\exp\left(-\frac{1}{2}\theta_2\nu s^2\right) \tag{13}$$

does *not* depend on θ_1. Note that the factorization in (13) demonstrates that $\overline{y}$ is a sufficient statistic for θ_1 (i.e. the likelihood for θ_1 can be written so that data information enters only through $\overline{y}$). Also note that density $\phi(\overline{y}|\theta_1,h^{-1})$ corresponds to the sampling density of the sample mean, given θ_1.

Let us now multiply prior times likelihood

$$\begin{aligned}
p(\theta_1)p\,(y|\theta_1) &= \phi(\theta_1|\underline{\mu},\underline{h}^{-1})c_1\phi(\overline{y}|\theta_1,h^{-1}) \qquad (14)\\
&= c_1(2\pi\underline{h}^{-1})^{-1/2}\exp\left(-\frac{1}{2}\underline{h}(\theta_1-\underline{\mu})^2+h(\overline{y}-\theta_1)^2\right)\\
&= c_1(2\pi\underline{h}^{-1})^{-1/2}\exp\left(-\frac{1}{2}(\theta_1-\overline{y})^2+(\underline{h}^{-1}+h^{-1})^{-1}(\overline{y}-\underline{\mu})^2\right)\\
&= c_1\phi\left(\overline{y}|\underline{\mu},\underline{h}^{-1}+h^{-1}\right)\phi(\theta_1|\overline{\mu},\overline{h}^{-1}). \qquad (15)
\end{aligned}$$

Equation (15) can be used to obtain both posterior and marginal likelihood. Since posterior is proportional to prior times likelihood and θ_1 only enters in the term $\phi(\theta_1|\overline{\mu}, \overline{h}^{-1})$ it follows that the posterior density of θ_1 is

$$p(\theta_1|y) = \phi\left(\theta_1|\overline{\mu}, \overline{h}^{-1}\right). \tag{16}$$

Remembering that the marginal likelihood, $p(y)$, is the integral of prior times likelihood we can use (15) as follows:

$$\begin{aligned} p(y) &= \int_{-\infty}^{\infty} p(\theta_1) p\left(y|\theta_1\right) d\theta_1 & (17) \\ &= c_1 \phi\left(\overline{y}|\underline{\mu}, \underline{h}^{-1} + h^{-1}\right) \int_{-\infty}^{\infty} \phi(\theta_1|\overline{\mu}, \overline{h}^{-1}) d\theta_1 & (18) \\ &= c_1 \phi\left(\overline{y}|\underline{\mu}, \underline{h}^{-1} + h^{-1}\right). & (19) \end{aligned}$$

Note that, to move from (18) to (19) we are using the fact that p.d.f.'s must integrate to 1.

The interpretation of quantities (6) and (7) is now clear: they are the posterior precision and posterior mean, respectively. Note that it is the additivity of precisions in these equations that motivates working with precisions rather than variances. Because posterior density (16) and prior density (4) are both members of the Normal family, it follows that the conjugate prior for the case of random sampling from a Normal population with *known* variance is itself a Normal density (Note: As we shall see in the next exercise, this no longer holds true when the variance in unknown).

Exercise 4

Consider Exercise 3, but with the population precision θ_2 also unknown. Suppose the joint prior distribution for $\theta = [\theta_1, \theta_2]'$ is the Normal-Gamma distribution, denoted $\theta \sim NG(\underline{\mu}, \underline{q}, \underline{s}^{-2}, \underline{\nu})$ with density

$$\begin{aligned} f_{NG}(\theta|\underline{\mu}, \underline{q}, \underline{s}^{-2}, \underline{\nu}) &= \phi(\theta_1|\underline{\mu}, \theta_2^{-1}\underline{q})\gamma(\theta_2|\underline{s}^{-2}, \underline{\nu}) & (20) \\ &= \left((2\pi\theta_2^{-1})^{-1/2} \exp\left[-\frac{1}{2}\theta_2 \underline{q}^{-1}(\theta_1 - \underline{\mu})^2\right]\right) \\ &\quad * \left[\left(\left[\frac{2}{\underline{\nu}\underline{s}^2}\right]^{\underline{\nu}/2} \Gamma(\underline{\nu}/2)\right)^{-1} \theta_2^{(\underline{\nu}-2)/2} \exp\left(-\frac{1}{2}\theta_2 \underline{\nu}\underline{s}^2\right)\right] \\ &\propto \theta_2^{(\underline{\nu}-1)/2} \exp\left(-\frac{1}{2}\theta_2[\underline{\nu}\underline{s}^2 + \underline{q}^{-1}(\theta_1 - \underline{\mu})^2]\right), \end{aligned}$$

where $\underline{\mu}$.,$\underline{q}$, $\underline{s}^2$ and $\underline{\nu}$ are known positive constants and $\gamma(\theta_2|\underline{s}^{-2}, \underline{\nu})$ is notation for the Gamma p.d.f.. Find the posterior density of θ.

Solution to Exercise 4

Dropping irrelevant constants not depending on θ_1 and θ_2, the posterior is proportion to the likelihood (12) tiomes the prior in (20):

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p\left(y|\theta_1\right) & (21)\\ &\propto \left[\phi(\theta_1|\underline{\mu},\theta_2^{-1}\underline{q})\gamma(\theta_2|\underline{s}^{-2},\underline{\nu})\right]\left[c_1(\theta_2)\phi(\overline{y}|\theta_1,h^{-1})\right] \\ &\propto \theta_2^{(\underline{\nu}-1)/2}\exp\left(-\frac{\theta_2}{2}[\underline{\nu s}^2+\underline{q}^{-1}(\theta_1-\underline{\mu})^2]\right)\theta_2^{(\underline{\nu}+1)/2}\exp\left(-\frac{\theta_2}{2}[\nu s^2+T(\overline{y}-\theta_1)^2]\right) \\ &\propto \theta_2^{\overline{\nu}/2}\exp\left(-\frac{\theta_2}{2}[\underline{\nu s}^2+\nu s^2+\underline{q}^{-1}(\theta_1-\underline{\mu})^2+T(\overline{y}-\theta_1)^2]\right), & (22)\end{aligned}$$

where

$$\overline{\nu}=\underline{\nu}+T. \tag{23}$$

Using identity (11) with $\underline{h}=\underline{q}^{-1}\theta_2$ and $h=T\theta_2$, the last two terms in square brackets in (22) can be written as

$$\underline{q}^{-1}(\theta_1-\underline{\mu})^2+T(\overline{y}-\theta_1)^2=\overline{q}^{-1}(\theta_1-\overline{\mu})^2+(\underline{q}+T^{-1})^{-1}(\overline{y}-\underline{\mu})^2, \tag{24}$$

where

$$\overline{q}=(\underline{q}^{-1}+T)^{-1}, \tag{25}$$

and

$$\overline{\mu}=\overline{q}(\underline{q}^{-1}\underline{\mu}+T\overline{y}). \tag{26}$$

The letting

$$\begin{aligned} \overline{s}^2 &= \overline{\nu}^{-1}\left[\underline{\nu s}^2+\nu s^2+(\underline{q}+T^{-1})^{-1}(\overline{y}-\underline{\mu})^2\right] & (27)\\ &= \overline{\nu}^{-1}\left[\underline{\nu s}^2+\nu s^2+\underline{q}^{-1}\overline{q}T(\overline{y}-\underline{\mu})^2\right], & (28)\end{aligned}$$

it follows from (23) through (28) that posterior density (22) can be written

$$\begin{aligned} p(\theta|y) &\propto \theta_2^{\overline{\nu}/2}\exp\left(-\frac{\theta_2}{2}[\overline{q}^{-1}(\theta_1-\overline{\mu})^2+\overline{\nu s}^2]\right) & (29)\\ &\propto \theta_2^{1/2}\exp\left[-\frac{\theta_2}{2}\overline{q}^{-1}(\theta_1-\overline{\mu})^2\right]\theta_2^{(\overline{\nu}-2)/2}\exp\left[-\frac{\theta_2}{2}\overline{\nu s}^2\right] & (30)\end{aligned}$$

Comparing the formula for the Normal-Gamma p.d.f. (see equation 20 or Appendix B of the textbook) and (2.35), it can be seen that posterior density $p(\theta|y)$ corresponds to the kernel of a $NG(\overline{\mu},\overline{q},\overline{s}^{-2},\overline{\nu})$ distribution, with updating formulas (23) and (25) through (27). Because both prior and posterior are Normal-Gamma densities, the Normal Gamma is the conjugate prior for random sampling from a Normal population with unknown mean and variance.

Exercise 5

Natural conjugate priors have the desirable feature that prior information can be viewed as "fictitious sample information" in that it is combined with the sample in exactly the same way that additional sample information would be combined. To clarify this point, reconsider Exercises 1, 3 and 4.

(a) Using the setup and results from Exercise 1 show that a Beta prior distribution with parameters $\underline{\alpha}$ and $\underline{\delta}$ can be interpreted as the information contained in a sample of size $\underline{T} = \underline{\alpha} + \underline{\delta} - 2$ with $\underline{\alpha} - 1$ successes from the Bernoulli process of interest. (Of course if $\underline{\alpha}$ and $\underline{\delta}$ are not integers, then this interpretation must be loosely made).

(b) Using the setup and results from Exercise 3 (known variance) show that a Normal prior distribution (4) can be interpreted in terms of "equivalent sample information."

(c) Using the setup and results from Exercise 4 (unknown variance) show that a fictitious sample interpretation of prior (20) can also be given.

Solution to Exercise 5

(a) Such an interpretation follows directly from comparing (2) with (3) and remember that m is the number of successes.

(b) Define

$$\underline{T} \equiv \frac{\theta_2^{-1}}{\underline{h}^{-1}} = \frac{\underline{h}}{\theta_2}. \tag{31}$$

and write the prior variance as

$$\underline{h}^{-1} = \frac{\theta_2^{-1}}{\underline{T}}. \tag{32}$$

Similarly, using (6) define

$$\overline{T} \equiv \frac{\theta_2^{-1}}{\overline{h}^{-1}} = \frac{\overline{h}}{\theta_2} = \underline{T} + T, \tag{33}$$

and note that posterior mean (7) can be written

$$\overline{\mu} = \frac{\underline{T}\underline{\mu} + T\overline{y}}{\overline{T}^{-1}}. \tag{34}$$

Hence, prior distribution (4) can be interpreted as the information contained in a sample of size $\underline{T}$ from the underlying $N(\theta_1, \theta_2^{-1})$ population yielding a "sample" mean $\underline{\mu}$ and a variance in this prior mean equaling (32). Given this interpretation, (33) and (34) can be viewed as formulae for pooling the information from the actual sample and the fictitious prior sample.

(c) Let

$$\underline{T} = \frac{1}{\underline{q}}. \tag{35}$$

represent the prior sample size and let

$$\overline{T} = \frac{1}{\overline{q}} = \underline{T} + T \tag{36}$$

represent the total sample size. Then again posterior mean (26) can be written as the weighted average (34).

Exercise 6

Consider the Normal-Gamma prior density of (20). Find the marginal prior distribution of θ_1. Repeat the exercise for the Normal-Gamma posterior density of (30).

Solution to Exercise 6

The marginal p.d.f. of θ_1 is obtained by integrating out θ_2 from the joint p.d.f. (20). Note that the last line of (20) is the kernel of a $\gamma(a,b)$ density for θ_2 given all other quantities, where

$$a = \frac{\underline{\nu}+1}{\underline{\nu}\underline{s}^2}\left[1+\frac{1}{\underline{\nu}}\left(\frac{(\theta_1-\underline{\mu})^2}{\underline{s}^2\underline{q}}\right)\right]^{-1},$$

$$b = \underline{\nu}+1.$$

Thus, integrating (20) with respect to θ_2 is proportional to the integrating constant of a $\gamma(a,b)$ density. This can be found in the textbook Appendix B. Thus,

$$p(\theta_1|\underline{\mu},\underline{q},\underline{s}^2,\underline{\nu}) \propto \Gamma(b/2)\left(\frac{2a}{b}\right)^{b/2} \propto \left[1+\frac{1}{\underline{\nu}}\left(\frac{(\theta_1-\underline{\mu})^2}{\underline{s}^2\underline{q}}\right)\right]^{-1}, \tag{37}$$

which is the kernel of a $t(\theta_1|\underline{\mu},\underline{s}^2\underline{q},\underline{\nu}+1)$ distribution. The derivation for the

marginal posterior is exactly the same, except that all hyperparameters have upper bars. That is, we end up with$t(\theta_1|\overline{\mu},\overline{q},\overline{s}^{-2},\overline{\nu}+1)$.

Exercise 7

Consider a random sample Y_t $(t=1,2,\cdots,T)$ from a multivariate $N(\mu,\Sigma)$ distribution where μ is an $M\times 1$ vector and Σ is an $M\times M$ positive definite matrix. Define $Y=[Y_1,Y_2,\cdots,Y_T]$ and

$$\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t,$$

and

$$S = \frac{1}{T}\sum_{t=1}^{T}(Y_t-\overline{Y})(Y_t-\overline{Y})'.$$

Suppose $\underline{m}$ and $\underline{\mu}$ are both $M\times 1$, $\underline{T}>0$ and $\underline{\omega}>M$ are both scalars, and $\underline{S}$ is a $M\times M$ positive definite matrix. Consider the natural conjugate priors for the following three cases.

(a) Suppose μ is unknown, Σ^{-1} is known and the prior distribution for μ is multivariate Normal with prior density $p(\mu)=\phi(\mu|\underline{\mu},\underline{\Sigma})$. Find the posterior distribution for μ

(b) Suppose both μ and Σ^{-1} are unknown with so-called Normal-Wishart prior distribution

$$\begin{aligned} p(\mu, \Sigma^{-1}) &= p(\mu|\Sigma^{-1})p(\Sigma^{-1}) \\ &= \phi(\mu|\underline{\mu}, \underline{T}^{-1}\Sigma)W(\Sigma^{-1}|\underline{S}^{-1}\underline{\omega}). \end{aligned}$$

Find the posterior distribution for μ and Σ^{-1}. Also find the marginal posterior distributions for μ and Σ^{-1}.

Solution to Exercise 7

This exercise repeats Exercises 3 and 4 for the matrix case and introduces you to the Wishart distribution, which is the matrix generalization of the Gamma. The likelihood function is common to parts a) and b) of the question. Similar to previous exercises, we are using the notation "$\phi(\mu|\underline{\mu}, \underline{\Sigma})$" to denote the multivariate Normal p.d.f.. Using the definition of the multivariate Normal distribution we have:

$$\begin{aligned} p(y|\mu, \Sigma^{-1}) &= \prod_{t=1}^{T}(2\pi)^{-M/2}|\Sigma^{-1}|^{(1/2)} \exp\left[-\frac{1}{2}(y_t-\mu)'\Sigma^{-1}(y_t-\mu)\right] \qquad (38) \\ &= (2\pi)^{-TM/2}|\Sigma^{-1}|^{(T/2)} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu)'\Sigma^{-1}(y_t-\mu)\right] \\ &= (2\pi)^{-TM/2}|\Sigma^{-1}|^{(T/2)} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}\text{tr}\left[\Sigma^{-1}(y_t-\mu)(y_t-\mu)'\right]\right] \\ &= (2\pi)^{-TM/2}|\Sigma^{-1}|^{(T/2)} \exp\left[-\frac{1}{2}\text{tr}\left(\Sigma^{-1}\sum_{t=1}^{T}(y_t-\mu)(y_t-\mu)'\right)\right] \end{aligned}$$

Noting the identity

$$\sum_{t=1}^{T}(y_t-\mu)(y_t-\mu)' = S + T(\overline{y}-\mu)(\overline{y}-\mu)',$$

likelihood (38) can be written as

$$L(\mu, \Sigma^{-1}; y) = (2\pi)^{-TM/2}|\Sigma^{-1}|^{(T/2)} \exp\left[-\frac{1}{2}\text{tr}\left(\Sigma^{-1}[S + T(\overline{y}-\mu)(\overline{y}-\mu)']\right)\right]. \qquad (39)$$

(a) Combining multivariate Normal prior density $p(\mu) = \phi(\mu|\underline{\mu}, \underline{\Sigma})$ with likelihood (39) yields

$$\begin{aligned} p(\mu|y) &\propto \exp\left[-\frac{1}{2}(\mu-\underline{\mu})'\Sigma^{-1}(\mu-\underline{\mu})\right] \times \qquad (40) \\ &\quad \exp\left[-\frac{1}{2}\text{tr}\left(\Sigma^{-1}[S + T(\overline{y}-\mu)(\overline{y}-\mu)']\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\left((\mu-\underline{\mu})'\Sigma^{-1}(\mu-\underline{\mu}) + (\mu-\overline{y})'(T\Sigma^{-1})(\mu-\overline{y})\right)\right]. \end{aligned}$$

Completing the square on μ, i.e., writing

$$(\mu - \underline{\mu})'\Sigma^{-1}(\mu - \underline{\mu}) + (\mu - \overline{y})'(T\Sigma^{-1})(\mu - \overline{y}) = \qquad (41)$$
$$(\mu - \overline{\mu})'(\underline{\Sigma}^{-1} + T\Sigma^{-1})(\mu - \overline{\mu}) + (\underline{\mu} - \overline{y})'[\underline{\Sigma}^{-1}(\underline{\Sigma}^{-1} + T\Sigma^{-1})^{-1}T\Sigma^{-1}](\underline{\mu} - \overline{y}),$$

where

$$\overline{\mu} = (\underline{\Sigma}^{-1} + T\Sigma^{-1})^{-1}(\underline{\Sigma}^{-1}\underline{\mu} + T\Sigma^{-1}\overline{y}), \qquad (42)$$

posterior kernel (40) simplifies to

$$p(\mu|y) \propto \exp\left[-\frac{1}{2}(\mu - \overline{\mu})'(\underline{\Sigma}^{-1} + T\Sigma^{-1})(\mu - \overline{\mu})\right], \qquad (43)$$

which is immediately recognized as a multivariate Normal kernel. Therefore,

$$\mu|y \sim N\left(\overline{\mu}, [\underline{\Sigma}^{-1} + T\Sigma^{-1}]^{-1}\right). \qquad (44)$$

(b) The posterior for μ and Σ^{-1} is

$$\begin{aligned} p(\mu, \Sigma^{-1}|y) \quad \propto \quad & |\underline{T}\Sigma^{-1}|^{(1/2)} \exp\left[-\frac{1}{2}(\mu - \underline{\mu})'\underline{T}\Sigma^{-1}(\mu - \underline{\mu})\right] |\Sigma^{-1}|^{(\underline{\omega}-M-1)/2} * \qquad (45) \\ & \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\underline{S}^{-1}\right] \left(|\Sigma^{-1}|^{(T/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}[S + T(\overline{y} - \mu)(\overline{y} - \mu)'])\right]\right) \\ \propto \quad & |\Sigma^{-1}|^{(1/2)} \exp\left[-\frac{1}{2}(\mu - \underline{\mu})'\underline{T}\Sigma^{-1}(\mu - \underline{\mu})\right] \exp\left[-\frac{1}{2}(\overline{y} - \mu)'T\Sigma^{-1}(\overline{y} - \mu)\right] * \\ & \left(|\Sigma^{-1}|^{(T+\underline{\omega}-M-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}[\underline{S}^{-1} + S])\right]\right). \end{aligned}$$

analogous to (41)-(42) we have

$$(\mu - \underline{\mu})'\underline{T}\Sigma^{-1}(\mu - \underline{\mu}) + (\mu - \overline{y})'(T\Sigma^{-1})(\mu - \overline{y}) = \qquad (46)$$
$$(\mu - \overline{\mu})'[(\underline{T} + T)\Sigma^{-1}](\mu - \overline{\mu}) + (\mu - \overline{y})[\underline{T}T(\underline{T} + T^{-1}\Sigma^{-1})](\mu - \overline{y}),$$

where

$$\overline{\mu} = \frac{\underline{T}^{-1}\mu + T\overline{y}}{\underline{T} + T}. \qquad (47)$$

Using (46)-(47), posterior (45) simplifies to

$$\begin{aligned} p(\mu, \Sigma^{-1}|y) \quad \propto \quad & |\overline{\omega}\Sigma^{-1}|^{(1/2)} \exp\left[-\frac{1}{2}(\mu - \overline{\mu})'\overline{\omega}\Sigma^{-1}(\mu - \overline{\mu})\right] * \qquad (48) \\ & \left(|\Sigma^{-1}|^{(T+\underline{\omega}-M-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\tilde{S}^{-1})\right]\right), \end{aligned}$$

where

$$\tilde{S} = \underline{S}^{-1} + S + \frac{\underline{T}T}{\underline{T} + T}(\overline{y} - \mu)(\overline{y} - \mu)'. \qquad (49)$$

This can be seen to be of the same form as the Normal-Wishart prior distribution (confirming the conjugacy of this prior).