Introduction to Bayesian Modeling

Homework Assignment #1 Solutions

1. An enzyme-linked immunosorbent assay (ELISA) test is performed to determine if the human immunodeficiency virus (HIV) is present in the blood of individuals. The ELISA test is not perfect. Suppose that the ELISA test correctly indicates HIV 99% of the time, and that the proportion of the time that it correctly indicates no HIV is 99.5%. Suppose that the prevalence among blood donors is known to be 1/10,000.

   (a) What proportion of blood that is donated will test positive using the ELISA test?

   Let E be the event that a blood sample tests positive and H be the event that a blood sample comes from a donor with HIV. Then $\Pr[H] = 0.0001$, $\Pr[E|H] = 0.99$, *and* $\Pr[E^c|H^c] = 0.995$. *Then using the law of total probability, we have:*

$$\begin{aligned}
\Pr[E] &= \Pr[E \mid H]\Pr[H] + \Pr[E \mid H^c]\Pr[H^c] \\
&= 0.99 \times 0.0001 + 0.005 \times 0.999 \\
&= 0.000099 + 0.004995 \\
&= 0.005094
\end{aligned}$$

   (b) What proportion of the blood that tests negative on the ELISA test is actually infected with HIV?

   From Part (a), we know $\Pr[E] = 0.005094$ *Therefore* $\Pr[E^c] = 0.994906$ *and by the definition of conditional probability:*

$$\begin{aligned}
\Pr[H \mid E^c] &= \frac{\Pr[E^c \mid H]\Pr[H]}{\Pr[E^c]} \\
&= \frac{0.01 \times 0.0001}{0.994906} \\
&\simeq 0.000001
\end{aligned}$$

   So approximately 1 in every 1,000,000 blood samples that test negative on the ELISA test is actually infected with HIV.

   (c) What is the probability that a positive ELISA outcome is truly positive, that is, what proportion of individuals with positive outcomes are actually infected with HIV?

Using a similar method to the one in Part (b), we have:

$$\begin{aligned}
\Pr[H \mid E] &= \frac{\Pr[E \mid H]\Pr[H]}{\Pr[E \mid H]\Pr[H] + \Pr[E \mid H^c]\Pr[H^c]} \\
&= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.005 \times 0.999} \\
&= \frac{0.000099}{0.000099 + 0.004995} \\
&= 0.0194
\end{aligned}$$

In other words, about 2% of blood that tests positive comes from individuals who truly have HIV. The other 98% are false positives.

2. An article in the July 22, 2009 edition of the International Herald Tribune indicated that 7% of British children attend special schools that cater to privileged parents and that 75% of all judges are known to have attended such schools. The implication was that underprivileged, but presumably intelligent and hardworking children were being excluded from high ranking professions like judgeships. We want to look at the relative probabilities of being a judge given that one did or did not attend an elite school.

   Specifically, let E denote attending an elite school with E^c the complement. Let J denote becoming a judge with J^c the complement. Let $p = \Pr(J)$ be the (unknown to us) proportion of judges among the populace in Great Britain. We are given that $\Pr(E|J) = 0.75$ and $\Pr(E) = 0.07$.

   (a) Use the definition of conditional probability to find $\Pr(J|E)$ and $\Pr(J|E^c)$ as functions of p. Find an actual number for the ratio $\Pr(J|E)/\Pr(J|E^c)$. What can you say about the effect of availability of elite schooling on the prospects of becoming a judge in Great Britain?

   By the definition of conditional probability, we have:

$$\begin{aligned}
\Pr[J \mid E] &= \frac{\Pr[E \mid J]\Pr[J]}{\Pr[E]} \\
&= \frac{0.75p}{0.07} \\
\Pr[J \mid E^c] &= \frac{\Pr[E^c \mid J]\Pr[J]}{\Pr[E^c]} \\
&= \frac{0.25p}{0.93}
\end{aligned}$$

   Then the ratio $\Pr(J|E)/\Pr(J|E^c)$ *is given by:*

$$\begin{aligned}
\frac{\Pr[J \mid E]}{\Pr[J \mid E^c]} &= \frac{{}^{0.75p}/_{0.07}}{{}^{0.25p}/_{0.93}} \\
&= \frac{0.75 \times 0.93}{0.07 \times 0.25} \\
&\simeq 39.857
\end{aligned}$$

(b) Let $q = \Pr(E)$. What value of q would correspond to no effect of E on the chances of becoming a judge later in life?

Note that "no effect" corresponds to having a ratio of conditional probabilities equal to 1. We can quickly see that:

$$\begin{aligned}
\Pr[J \mid E] &= \frac{\Pr[E \mid J]\Pr[J]}{\Pr[E]} \\
&= \frac{0.75p}{q} \\
\Pr[J \mid E^c] &= \frac{\Pr[E^c \mid J]\Pr[J]}{\Pr[E^c]} \\
&= \frac{0.25p}{1-q} \\
\frac{\Pr[J \mid E]}{\Pr[J \mid E^c]} &= \frac{{}^{0.75p}/_{q}}{{}^{0.25p}/_{1-q}} \\
&= 3\left(\frac{1-q}{q}\right)
\end{aligned}$$

Setting the left hand side equal to 1 gives us $q = 0.75$. *Alternatively (and much more simply), we may observe that the events "becoming a judge" and "attending an elite school" are independent when* $\Pr[E|J] = \Pr[E] = 0.75$.

3. Show that the mode of the Beta(a, b) distribution is $(a-1)/(a+b-2)$ when $a, b > 1$, zero when $a < 1 \leq b$ and one when $b < 1 \leq a$. Discuss the behavior of the mode when both a and b are less than one.

Begin by considering the pdf of a random variable, $X \sim \text{Beta}(a, b)$,

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}.$$

Further, observe that the natural log of $f(x)$ *is a monotonic transformation, and therefore the extrema of* $f(x)$ *and* $\ell n f(x)$ *will occur at the same points. Taking the first derivative of this function, we find that*

$$\begin{aligned}
\frac{d}{dx}\ell n f(x) &= \frac{d}{dx}\left[\ell n\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) + (a-1)\ell n x + (b-1)\ell n(1-x)\right] \\
&= \frac{a-1}{x} - \frac{b-1}{1-x}.
\end{aligned}$$

And setting the left hand side equal to zero, we obtain

$$\begin{aligned}\frac{b-1}{1-x} &= \frac{a-1}{x} \\ x(b-1) &= (a-1) - x(a-1) \\ x(a+b-2) &= a-1 \\ x &= \frac{a-1}{a+b-2}.\end{aligned}$$

Then the first derivative of $f(x)$ has one root, located at $x = (a-1)/(a+b-2)$. Moreover, since the natural log is a monotonic transformation, maxima of $f(x)$ will also be maxima of $\ell n f(x)$, and minima of $f(x)$ will also be minima of $\ell n f(x)$. Therefore, at the extrema the second derivative of $f(x)$ and $\ell n f(x)$ will have the same sign. Consider, then, the second derivative of $\ell n f(x)$ evaluated at $x = (a-1)/(a+b-2)$.

$$\begin{aligned}\frac{d^2}{dx^2}\ell n f(x) &= \frac{d}{dx}\frac{a-1}{x} - \frac{b-1}{1-x} \\ &= -\frac{a-1}{x^2} - \frac{b-1}{(1-x)^2}\end{aligned}$$

Note that when a and b are greater than 1, $\frac{d^2}{dx^2}\ell n f(x) < 0$ for all $x \in [0,1]$. This indicates that the function is concave down at the extremum. In other words, the extremum is a maximum—so when $a, b > 1$, the mode of the distribution occurs at $x = (a-1)/(a+b-2)$.

When a and b are both less than 1, by contrast, $\frac{d^2}{dx^2}\ell n f(x) > 0$ for all $x \in [0,1]$. This means that the extremum represents a minimum, and the mode is to be found at either $x = 0$ or $x = 1$. Because the pdf is not well defined at these values, we consider the limits as x approaches either edge of its support. We find that when $a < b$,

$$\lim_{x\to 0^+} x^{a-1}(1-x)^{b-1} > \lim_{x\to 1^-} x^{a-1}(1-x)^{b-1},$$

and when $a > b$,

$$\lim_{x\to 0^+} x^{a-1}(1-x)^{b-1} < \lim_{x\to 1^-} x^{a-1}(1-x)^{b-1}.$$

That is, when $a < b < 1$, the mode of the distribution occurs at $x = 0$; and when $b < a < 1$, the mode occurs at $x = 1$.

When one of a or b is less than 1 and the other is 1 or greater, consideration of the first derivative reveals that $\frac{d}{dx}\ell n f(x) < 0$ for $x \in [0,1]$. Since the natural log is a monotonic transformation, this means that $f(x)$ itself is also monotone over this interval. It is then straightforward to show that, as above, when $a < b$ the mode occurs at $x = 0$ and when $a > b$ the mode occurs at $x = 1$.

4. Consider the expression

$$f(y|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta),$$

and define $f_i(y_i|\theta) \equiv f(y_i|\theta, x_i)$ where x_i denotes a known "covariate" variable. In particular, let the x_i's identify two groups, $x_i = 1$ for $i = 1, \ldots, k$ and $x_i = 0$ for $i = k+1, \ldots, n$. With $\theta = (\theta_1, \theta_2)'$, define $\lambda_i = \theta_1^{x_i}\theta_2^{1-x_i}$ so that λ_i equals θ_1 if $x_i = 1$ and equals θ_2 if $x_i = 0$.

Now consider $f(y_i|\theta, x_i) \equiv f(y_i|\lambda_i)$ for each of the choices:

$$\begin{aligned}
&\text{(i)}\ f(y_i|\lambda_i) = \lambda_i e^{-\lambda_i y_i},\\
&\text{(ii)}\ f(y_i|\lambda_i) = \lambda_i^{y_i} e^{-\lambda_i}/y_i!,\\
&\text{(iii)}\ f(y_i|\lambda_i) = 2\lambda_i y_i e^{-\lambda_i y_i^2}.
\end{aligned}$$

(a) Using Table 2.1, for each of the three choices identify the distribution with density $f(y_i|\lambda_i)$ when $x_i = 1$ and also when $x_i = 0$.

The three distributions are:

$$\begin{aligned}
&\text{(i)}\ y_i \sim \text{Exp}(\lambda_i),\\
&\text{(ii)}\ y_i \sim \text{Pois}(\lambda_i),\\
&\text{(iii)}\ y_i \sim \text{Weibull}(2, \lambda_i).
\end{aligned}$$

Then when $x_i = 1$, λ_i in these distributions is θ_1. When $x_i = 0$, λ_i is θ_2.

(b) For each choice, simplify the product $\prod_{i=1}^n f_i(y_i|\theta)$ so that as a function of θ it is proportional to a function that depends only on some combination of $\sum_{i=1}^k g(y_i)$, $\sum_{i=k+1}^n g(y_i)$, $\prod_{i=1}^k g(y_i)$, and $\prod_{i=k+1}^n g(y_i)$ where $g(x)$ can be one of x, x^2, or $x!$.

(i)

$$\begin{aligned}
\prod_{i=1}^n f_i(y_i|\theta) &= \left(\prod_{i=1}^k \theta_1 e^{-\theta_1 y_i}\right)\left(\prod_{i=k+1}^n \theta_2 e^{-\theta_2 y_i}\right)\\
&= \left(\theta_1^k e^{-\theta_1 \sum_{i=1}^k y_i}\right)\left(\theta_2^{n-k} e^{-\theta_2 \sum_{i=k+1}^n y_i}\right)\\
&= \theta_1^k \theta_2^{n-k} e^{-\theta_1 \sum_{i=1}^k y_i - \theta_2 \sum_{i=k+1}^n y_i}
\end{aligned}$$

(ii)

$$\begin{aligned}
\prod_{i=1}^n f_i(y_i|\theta) &= \left(\prod_{i=1}^k \theta_1^{y_i} e^{-\theta_1}/y_i!\right)\left(\prod_{i=k+1}^n \theta_2^{y_i} e^{-\theta_2}/y_i!\right)\\
&= \left(\frac{\theta_1^{\sum_{i=1}^k y_i} e^{-k\theta_1}}{\prod_{i=1}^k y_i!}\right)\left(\frac{\theta_2^{\sum_{i=k+1}^n y_i} e^{-(n-k)\theta_2}}{\prod_{i=k+1}^n y_i!}\right)\\
&= \frac{\theta_1^{\sum_{i=1}^k y_i}\theta_2^{\sum_{i=k+1}^n y_i} e^{-k\theta_1 - (n-k)\theta_2}}{\prod_{i=1}^n y_i!}
\end{aligned}$$

(iii)

$$
\begin{aligned}
\prod_{i=1}^{n} f_i(y_i|\theta) &= \left(\prod_{i=1}^{k} 2\theta_1 y_i e^{-\theta_1 y_i^2}\right)\left(\prod_{i=k+1}^{n} 2\theta_2 y_i e^{-\theta_2 y_i^2}\right) \\
&= \left(2^k \theta_1^k \left(\prod_{i=1}^{k} y_i\right) e^{-\theta_1 \sum_{i=1}^{k} y_i^2}\right)\left(2^{n-k}\theta_2^{n-k}\left(\prod_{i=k+1}^{n} y_i\right) e^{-\theta_2 \sum_{i=k+1}^{n} y_i^2}\right) \\
&= 2^n \theta_1^k \theta_2^{n-k}\left(\prod_{i=1}^{n} y_i\right) e^{-\theta_1 \sum_{i=1}^{k} y_i^2 - \theta_2 \sum_{i=k+1}^{n} y_i^2}
\end{aligned}
$$

5. Derive the following posterior densities and show that these densities are the densities shown in Table 2.3.

   (a) An exponential sample with a Gamma (α, β) prior.

   Let $X_i \overset{\text{iid}}{\sim} \text{Exp}\,(\theta)$ *for* $i \in \{1, \ldots, n\}$ *and let* $\theta \sim \text{Gamma}\,(\alpha, \beta)$. *Then*

$$
\begin{aligned}
p(\theta \mid x_1, \ldots, x_n) &\propto p(\theta) \prod_{i=1}^{n} f(x_i \mid \theta) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \prod_{i=1}^{n} \theta e^{-\theta x_i} I_{(0,\infty)}(\theta) \\
&\propto \theta^{n+\alpha-1} e^{-\theta\left(\beta + \sum_{i=1}^{n} x_i\right)} I_{(0,\infty)}(\theta)
\end{aligned}
$$

   We recognize this to be the kernel of a Gamma $(\alpha + n, \beta + \sum_{i=1}^{n} x_i)$ *distribution.*

   (b) A normal sample with known mean, μ, and a Gamma (α, β) prior on the precision, τ.

   Let $X_i \overset{\text{iid}}{\sim} \text{N}\left(\mu, {}^1\!/\!_\tau\right)$ *for* $i \in \{1, \ldots, n\}$ *and let* $\tau \sim \text{Gamma}\,(\alpha, \beta)$. *Then*

$$
\begin{aligned}
p(\tau \mid x_1, \ldots, x_n) &\propto p(\tau) \prod_{i=1}^{n} f(x_i \mid \tau) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau} \prod_{i=1}^{n} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x_i-\mu)^2} I_{(0,\infty)}(\tau) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau} \left(\frac{\tau}{2\pi}\right)^{n/2} e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2} I_{(0,\infty)}(\tau) \\
&\propto \tau^{\frac{n}{2}+\alpha-1} e^{-\tau\left(\beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)} I_{(0,\infty)}(\tau)
\end{aligned}
$$

   We recognize this to be the kernel of a Gamma $\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$ *distribution.*

6. Let $y \sim \mathrm{N}\left(\mu, {}^{1}/_{\tau}\right)$. Derive the posterior distribution for μ and τ under the conjugate prior. Use the *complete the square formula*

$$r(\mu - v)^2 + s(\mu - w)^2 = (r+s)(\mu - \hat{\mu})^2 + \frac{rs}{r+s}(v-w)^2, \quad \hat{\mu} = \frac{r}{r+s}v + \frac{s}{r+s}w,$$

and adapt the arguments illustrated in the previous subsection for the SIR prior.

(a) Derive the conditional density $p(\mu|\tau, y)$. Show that it is the normal density given in this subsection.

Recall that the conjugate prior for $\mathrm{N}\left(\mu, {}^{1}/_{\tau}\right)$ data is

$$\tau \sim \mathrm{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right), \quad \mu \mid \tau \sim \mathrm{N}\left(\mu_0, \frac{1}{\omega_0 \tau}\right).$$

Then we have

$$\begin{aligned} p(\mu \mid \tau, y) &\propto L(\mu \mid \tau, y) p(\mu \mid \tau) \\ &= \sqrt{\frac{\tau}{2\pi}} \exp\left[-\frac{\tau}{2}(y-\mu)^2\right] \times \sqrt{\frac{\omega_0 \tau}{2\pi}} \exp\left[-\frac{\omega_0 \tau}{2}(\mu - \mu_0)^2\right] \\ &= \frac{\tau\sqrt{\omega_0}}{2\pi} \exp\left[-\frac{\tau}{2}\left((\mu - y)^2 + \omega_0(\mu-\mu_0)^2\right)\right] \end{aligned}$$

And using the complete the square formula, we obtain

$$\begin{aligned} p(\mu \mid \tau, y) &\propto \exp\left[-\frac{\tau}{2}\left((\omega_0+1)\left[\mu - \left(\frac{1}{1+\omega_0}y + \frac{\omega_0}{1+\omega_0}\mu_0\right)\right]^2\right)\right] \\ &\quad \times \exp\left[-\frac{\tau}{2}\left(\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2\right)\right] \\ &\propto \left[-\frac{\tau(1+\omega_0)}{2}\left(\mu - \left[\frac{1}{1+\omega_0}y + \frac{\omega_0}{1+\omega_0}\mu_0\right]\right)^2\right] \end{aligned}$$

We recognize that this is the kernel of a $\mathrm{N}\left(\hat{\mu}, {}^{1}/_{\tau(1+\omega_0)}\right)$ distribution, where

$$\hat{\mu} = \frac{1}{1+\omega_0}y + \frac{\omega_0}{1+\omega_0}\mu_0.$$

Then $\mu \mid \tau, y \sim \mathrm{N}\left(\hat{\mu}, {}^{1}/_{\tau(1+\omega_0)}\right)$.

(b) Using the result in (a), obtain the marginal densities $p(\mu|y)$ and $p(\tau|y)$.

Observe that $p(\mu, \tau \mid y) = p(\mu \mid \tau, y) p(\tau \mid y)$. Then

$$p(\tau \mid y) = \frac{p(\mu, \tau \mid y)}{p(\mu \mid \tau, y)},$$

where $p(\mu \mid \tau, y)$ is as we have already established in (a). We thus consider $p(\mu, \tau \mid y)$.

$$
\begin{aligned}
p(\mu,\tau \mid y) &\propto L(\mu,\tau \mid y)p(\mu \mid \tau)p(\tau) \\
&= \sqrt{\frac{\tau}{2\pi}} \exp\left[-\frac{\tau}{2}(y-\mu)^2\right] \times \sqrt{\frac{\omega_0\tau}{2\pi}} \exp\left[-\frac{\omega_0\tau}{2}(\mu-\mu_0)^2\right] \\
&\qquad \times \frac{(b/2)^{a/2}}{\Gamma(a)}\tau^{\frac{a}{2}-1} \exp\left(-\frac{\tau b}{2}\right) I_{(0,\infty)}(\tau) \\
&= \frac{\tau^{a/2}\omega_0^{1/2}(b/2)^{a/2}}{2\pi\Gamma(a)} \exp\left(-\frac{\tau}{2}\left[b+(\mu-y)^2+\omega_0(\mu-\mu_0)^2\right]\right) \\
&\propto \tau^{a/2} \exp\left(-\frac{\tau}{2}\left[(1+\omega_0)(\mu-\hat{\mu})^2+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2+b\right]\right) \\
&= \tau^{1/2} \exp\left(-\frac{\tau(1+\omega_0)}{2}(\mu-\hat{\mu})^2\right) \\
&\qquad \times \tau^{\frac{a}{2}-\frac{1}{2}} \exp\left(-\frac{\tau}{2}\left[b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2\right]\right) \\
&\propto [\tau(1+\omega_0)]^{1/2} \exp\left(-\frac{\tau(1+\omega_0)}{2}(\mu-\hat{\mu})^2\right) \\
&\qquad \times \tau^{\frac{a+1}{2}-1} \exp\left(-\tau\,\frac{b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2)}{2}\right)
\end{aligned}
$$

Here we can recognize the product of $p(\mu \mid \tau, y)$ and additional terms which form the kernel for a Gamma $\left(\frac{a+1}{2}, \frac{b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2}{2}\right)$ distribution. Then following our line of argument,

$$
\tau \mid y \sim \text{Gamma}\left(\frac{a+1}{2}, \frac{b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2}{2}\right).
$$

Now, to find the marginal distribution $p(\mu \mid y)$, we want to take the joint distribution which we've now found and marginalize over τ. We begin by writing the full posterior distribution.

$$
\begin{aligned}
p(\mu,\tau \mid y) &= \sqrt{\frac{\tau(1+\omega_0)}{2\pi}} \exp\left(-\frac{\tau(1+\omega_0)}{2}(\mu-\hat{\mu})^2\right) \\
&\qquad \times \frac{\left[\frac{b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2}{2}\right]^{\frac{a+1}{2}}}{\Gamma\left(\frac{a+1}{2}\right)}\tau^{\frac{a+1}{2}-1} \exp\left(-\tau\,\frac{b+\frac{\omega_0}{1+\omega_0}(y-\mu_0)^2)}{2}\right) \\
&= \sqrt{\frac{1+\omega_0}{2\pi}}\left(\frac{\left[\frac{b}{2}+\frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a+1}{2}}}{\Gamma\left(\frac{a+1}{2}\right)}\right)\tau^{\left(\frac{a}{2}+1\right)-1} \\
&\qquad \times \exp\left(-\tau\left[\frac{b}{2}+\frac{1+\omega_0}{2}(\mu-\hat{\mu})^2+\frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]\right)
\end{aligned}
$$

In this expression, we can recognize a Gamma kernel for τ conditional on μ, giving us

$$
\tau \mid \mu \sim \text{Gamma}\left(\frac{a}{2}+1, \frac{b}{2}+\frac{1+\omega_0}{2}(\mu-\hat{\mu})^2+\frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right).
$$

Knowing this, we can multiply and divide by the constant of integration for this Gamma distribution, which yields

$$
\begin{aligned}
p(\mu \mid y) &= \int_{\text{Supp}(\tau)} p(\mu, \tau \mid y) d\tau \\
&= \sqrt{\frac{1+\omega_0}{2\pi}} \left(\frac{\left[\frac{b}{2} + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a+1}{2}}}{\Gamma\left(\frac{a+1}{2}\right)} \right) \\
&\quad \times \left(\frac{\Gamma\left(\frac{a}{2}+1\right)}{\left[\frac{b}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2 + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a}{2}+1}} \right) \\
&\quad \times \int_0^\infty \frac{\left[\frac{b}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2 + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a}{2}+1}}{\Gamma\left(\frac{a}{2}+1\right)} \tau^{\left(\frac{a}{2}+1\right)-1} \\
&\qquad \times \exp\left(-\tau\left[\frac{b}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2 + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]\right) d\tau \\
&= \sqrt{\frac{1+\omega_0}{2\pi}} \left(\frac{\Gamma\left(\frac{a}{2}+1\right)}{\Gamma\left(\frac{a}{2}+\frac{1}{2}\right)} \right) \left(\frac{\left[\frac{b}{2} + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a}{2}+\frac{1}{2}}}{\left[\frac{b}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2 + \frac{\omega_0}{2(1+\omega_0)}(y-\mu_0)^2\right]^{\frac{a}{2}+1}} \right)
\end{aligned}
$$

Finally, to join up with the result in the book, consider the transformed random variable $\theta \mid y = \frac{\mu-\hat{\mu}}{\sqrt{\hat{\sigma}^2/(1+\omega_0)}} \mid y$, where $\hat{\mu}$ is as previously specified and

$$
\hat{\sigma}^2 = \frac{b + \frac{\omega_0}{1+\omega_0}(y-\mu_0)^2}{a+1}.
$$

If $\theta \mid y = f(\mu \mid y)$, then $f^{-1}(\theta \mid y) = \sqrt{\frac{\hat{\sigma}^2}{1+\omega_0}}\theta + \hat{\mu}$ and the Jacobian will be $\left|\sqrt{\frac{\hat{\sigma}^2}{1+\omega_0}}\right|$. This gives us

$$
\begin{aligned}
p(\theta \mid y) &= \sqrt{\frac{1+\omega_0}{2\pi}} \left(\frac{\Gamma\left(\frac{a}{2}+1\right)}{\Gamma\left(\frac{a}{2}+\frac{1}{2}\right)} \right) \left(\frac{\left[(a+1)\frac{\hat{\sigma}^2}{2}\right]^{\frac{a}{2}+\frac{1}{2}}}{\left[(a+1)\frac{\hat{\sigma}^2}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2\right]^{\frac{a}{2}+1}} \right) \left| \sqrt{\frac{\hat{\sigma}^2}{1+\omega_0}} \right| \\
&= \sqrt{\frac{1}{2\pi}} \left(\frac{\Gamma\left(\frac{a}{2}+1\right)}{\Gamma\left(\frac{a}{2}+\frac{1}{2}\right)} \right) \left(\frac{\left[(a+1)\frac{\hat{\sigma}^2}{2}\right]^{\frac{a}{2}+1}}{\left[(a+1)\frac{\hat{\sigma}^2}{2} + \frac{1+\omega_0}{2}(\mu-\hat{\mu})^2\right]^{\frac{a}{2}+1}} \right) \left(\sqrt{\frac{1}{\frac{a}{2}+\frac{1}{2}}} \right) \\
&= \frac{1}{\sqrt{2\pi\left(\frac{a}{2}+\frac{1}{2}\right)}} \left(\frac{\Gamma\left(\frac{a}{2}+1\right)}{\Gamma\left(\frac{a}{2}+\frac{1}{2}\right)} \right) \left(\frac{(a+1)\hat{\sigma}^2}{(a+1)\hat{\sigma}^2 + (\omega_0+1)(\mu-\hat{\mu})^2} \right)^{\frac{a}{2}+1} \\
&= \frac{1}{\sqrt{(a+1)\pi}} \left(\frac{\Gamma\left(\frac{a+2}{2}\right)}{\Gamma\left(\frac{a+1}{2}\right)} \right) \left(1 + \frac{(\omega_0+1)(\mu-\hat{\mu})^2}{(a+1)\hat{\sigma}^2} \right)^{-\frac{a+2}{2}}
\end{aligned}
$$

And since $\theta^2 = \frac{(\omega_0+1)(\mu-\hat{\mu})^2}{\hat{\sigma}^2}$, we can recognize this as a pdf for a t distribution. Specifically,

$$\theta \mid y = \frac{\mu - \hat{\mu}}{\sqrt{\hat{\sigma}^2/(1+\omega_0)}} \mid y \sim \mathrm{t}\,(a+1).$$

(c) Derive the predictive density for a future observation based on this model.

A similar argument to the one above should get the predictive density, but this question was much harder than I anticipated and I'm tired of writing solutions, so if you made it this far, excellent job!