SOLUTION FOR HOMEWORK 6, STAT 6331

1. Exerc.7.22. It is given that $X_1, \ldots, X_n$ is a sample from $N(\theta, \sigma^2)$, and the Bayesian approach is used with $\Theta \sim N(\mu, \tau^2)$. The parameters σ^2, μ and τ^2 are given.

(a) Find the joinf pdf of $\bar{X}$ and Θ.

We know that $\bar{X} \sim N(\theta, \sigma^2/n)$, thus

$$f_{\bar{X},\Theta}(x, \theta) = (2\pi\sigma^2/n)^{-1/2} e^{-(x-\theta)^2/[2\sigma^2/n]} (2\pi\tau^2)^{-1/2} e^{-(\theta-\mu)^2/[2\tau^2]}$$

$$= [4\pi^2\sigma^2\tau^2/n]^{-1/2} \exp\left(-\frac{x^2 - 2x\theta + \theta^2}{2\sigma^2/n} - \frac{\theta^2 - 2\mu\theta + \mu^2}{2\tau^2}\right) \quad (1)$$

$$= [4\pi^2\sigma^2\tau^2/n]^{-1/2} \exp\left(-\frac{\tau^2 x^2 - 2\theta[\tau^2 x + \mu\sigma^2/n] + (\tau^2 + \sigma^2/n)\theta^2 + \mu^2\sigma^2/n}{2(\sigma^2/n)\tau^2}\right). \quad (2)$$

What can we say about this distribution? It is bivariate normal! Remember that $\mathbf{Y} = (Y_1, Y_2)^T$ is bivariate normal with mean $(\nu_1, \nu_2)^T$ and covariance matrix

$$\Sigma = \left\| \begin{matrix} d_1^2 & \rho d_1 d_2 \\ \rho d_1 d_2 & d_2^2 \end{matrix} \right\|, \quad d_1 > 0,\ d_2 > 0,\ -1 \le \rho \le 1,$$

iff

$$f_{Y_1,Y_2}(y_1, y_2) = (2\pi)^{-2/2} (\det(\Sigma))^{-1/2} \exp\{-(1/2)(\mathbf{y} - \nu)^T \Sigma^{-1} (\mathbf{y} - \nu)\} \quad (3)$$

$$= \frac{(1-\rho^2)^{-1/2}}{2\pi d_1 d_2} \exp\left\{\frac{1}{2(1-\rho^2)}\left[\frac{(y_1 - \nu_1)^2}{d_1^2} - 2\rho\frac{(y_1 - \nu_1)(y_2 - \nu_2)}{d_1 d_2} + \frac{(y_2 - \nu_2)^2}{d_2^2}\right]\right\}. \quad (4)$$

Let us now directly match (1) and (4). Because these two densities are the same for all x and θ, we can equate coefficients (first for x^2 and then for θ^2)

$$\begin{cases} -\frac{1}{2\sigma^2/n} = -\frac{1}{2(1-\rho^2)d_1^2} \\ -\frac{1}{2\tau^2} - \frac{1}{2\sigma^2/n} = -\frac{1}{2(1-\rho^2)d_2^2} \end{cases}$$

Now note that $d_2^2 = \text{Var}(\Theta) = \tau^2$, and using this we get the simplified system

$$\begin{cases} -\frac{1}{\sigma^2/n} = -\frac{1}{(1-\rho^2)d_1^2} \\ \frac{1}{\tau^2} + \frac{1}{\sigma^2/n} = \frac{1}{(1-\rho^2)\tau^2} \end{cases}$$

which yields

$$\begin{cases} d_1^2 = n^{-1}\sigma^2 + \tau^2 \\ \rho^2 = \frac{\tau^2}{n^{-1}\sigma^2 + \tau^2}. \end{cases}$$

Now I note that $\nu_2 = E(\Theta) = \mu$. Keeping this in mind, I compare (1) and (4) in terms of coefficients for x (note that there is no term x in (1)),

$$0 = \frac{\nu_1}{(1-\rho^2)d_1^2} - \frac{\rho\nu_2}{(1-\rho^2)d_1 d_2}.$$

This yields

$$\nu_1 = \rho \frac{\mu d_1}{d_2} = \mu.$$

We conclude that (1) is the bivariate normal with

$$\nu_1 = \mu,\ \ \nu_2 = \mu,\ \ d_1^2 = n^{-1}\sigma^2 + \tau^2,\ \ d_2^2 = \tau^2,\ \ \rho^2 = \frac{\tau^2}{n^{-1}\sigma^2 + \tau^2}. \tag{5}$$

(ii) Can we figure out (5) faster/simpler? Of course! We know that $\nu_2 = \mu$ and $d_2^2 = \tau^2$. Then using $\bar{X} \sim N(\theta, \sigma^2)$, we can write,

$$\nu_1 = E(\bar{X}) = E\{E(\bar{X}|\Theta)\} = E(\Theta) = \mu,$$

and

$$d_1 = \mathrm{Var}(\bar{X}) = E(\bar{X})^2 - (E\bar{X})^2 = E(\bar{X}^2) - \mu^2.$$

Now, our next step is to calculate

$$E(\bar{X})^2 = E\{E\{(\bar{X})^2|\Theta\}\} = E\{[\mathrm{Var}(\bar{X}|\Theta) + (E\{\bar{X}|\Theta\})^2\}$$

$$= E\{[n^{-1}\sigma^2 + \Theta^2]\} = n^{-1}\sigma^2 + E(\Theta^2) = n^{-1}\sigma^2 + [\mathrm{Var}(\Theta) + (E\Theta)^2] = n^{-1}\sigma^2 + \tau^2 + \mu^2. \tag{6}$$

This allows us to conclude that

$$d_1 = E(\bar{X}^2) - \mu^2 = n^{-1}\sigma^2 + \tau^2.$$

Finally, because $E\{(\bar{X} - \mu)(\Theta - \mu)\} = E(\bar{X}\Theta) - \mu^2$, we get

$$\rho = \frac{E(\bar{X}\Theta) - \mu^2}{d_1 d_2} = \frac{E\{E(\bar{X}\Theta|\Theta)\} - \mu^2}{d_1 d_2} = \frac{E(\Theta^2) - \mu^2}{d_1 d_2} =$$

$$\text{I use (6) to continue } \frac{\tau^2 + \mu^2 - \mu^2}{\tau(n^{-1}\sigma^2 + \tau^2)^{1/2}} = \frac{\tau}{(n^{-1}\sigma^2 + \tau^2)^{1/2}}.$$

We gor the same parameters!

(b) We have shown this in Section (a), but let us do this directly to improve our calculus skills. Write

$$f_{\bar{X}}(x) = \int_{-\infty}^{\infty} f_{\bar{X},\Theta}(x, \theta) d\theta \text{ [I use (2)]}$$

$$\frac{1}{2\pi(\tau^2\sigma^2/n)^{1/2}} \int_{-\infty}^{\infty} \Big[\exp\{-\frac{\tau^2 x^2}{2(\sigma^2/n)\tau^2}\} \tag{7}$$

$$\times \exp\{-\frac{(\tau^2 + \sigma^2/n)[\theta^2 - 2\theta(\tau^2 x + \mu\sigma^2/n)(\tau^2 + \sigma^2/n)^{-1} + (\tau^2 x + \mu\sigma^2/n)^2/(\tau^2 + \sigma^2/n)^2]}{2(\sigma^2/n)\tau^2}\} \tag{8}$$

$$\times \exp\{\frac{(\tau^2 x + \mu\sigma^2/n)^2(\tau^2 + \sigma^2/n)^{-1} - \mu^2\sigma^2/n}{2(\sigma^2/n)\tau^2}\}\Big] d\theta. \tag{9}$$

Now see how I am finishing. The integral with respect to θ, according to (8), is a constant. Now, I expect to have

$$f_{\bar{X}}(x) = (2\pi d^2)^{-1/2} \exp\{-(x-\nu)^2/(2d^2)\} = (2\pi d^2)^{-1/2} \exp\{-\frac{x^2}{2d^2} + \frac{2x\nu}{2d^2} - \frac{\nu^2}{2d^2}\}.$$

Thus, let us look at terms with e^{-cx^2} and e^{-cx}. From (7) and (9) we get that

$$\exp\{-\frac{x^2}{2d^2}\} = \exp\{-\frac{\tau^2 - \tau^4(\tau^2+\sigma^2/n)^{-1}}{2(\sigma^2/n)\tau^2} x^2\} = \exp\{-\frac{x^2}{2(\tau^2+\sigma^2/n)}\}.$$

This yields that $d^2 = \tau^2 + \sigma^2/n$ (what was expected to see). Further,

$$\exp\{\frac{2\nu}{2d^2}\} = \exp\{\frac{2\tau^2\mu(\sigma^2/n)(\tau^2+\sigma^2/n)^{-1}}{2(\sigma^2/n)\tau^2}\} = \exp\{\frac{\mu}{\tau^2+\sigma^2/n}\}.$$

This implies the wished $\mu = \nu$.

(c) The posterior is

$$\pi(\theta|\bar{x}, \sigma^2, \mu, \tau^2) = \frac{f_{\Theta,\bar{X}}(\theta, \bar{x})}{f_{\bar{X}}(\bar{x})}.$$

We did this in class so I just remind you the idea. You wish to see

$$\pi(\theta|\bar{x}, \sigma^2, \mu, \tau^2) = (2\pi d^2)^{-1/2} \exp\{\frac{-(\theta-\nu)^2}{2d^2}\}. \tag{10}$$

The only θ what we see in (2) are

$$\pi(\theta|\bar{x}, \sigma^2, \mu, \tau^2) \propto \exp\{\frac{-(\tau^2+\sigma^2/n)\theta^2 - 2\theta[\tau^2\bar{x} + \mu\sigma^2/n]}{2\sigma^2/n)\tau^2}\}. \tag{11}$$

Thus we conclude vie a comparison of (10) and (11) that

$$\begin{cases} \exp\{-\frac{\theta^2}{2d^2}\} = \exp\{-\frac{\tau^2+\sigma^2/n)\theta^2}{2(\sigma^2/n)\tau^2}\} \\ \exp\{\frac{2\theta\nu}{2d^2}\} = \exp\{\frac{2\theta[\tau^2\bar{x}+\mu\sigma^2/n]}{2(\sigma^2/n)\tau^2}\}. \end{cases}$$

Solving the system yields

$$\begin{cases} d^2 = \frac{(\sigma^2.n)\tau^2}{\tau^2+\sigma^2/n} \\ \nu = \bar{x}\frac{\tau^2}{\tau^2+\sigma^2/n} + \mu\frac{\sigma^2/n}{\tau^2+\sigma^2/n}. \end{cases}$$

Finally, note that $\nu = E(\Theta|\bar{X}) = \hat{\theta}_{Bayes}$.

2. Exerc. 7.23. First of all, let is discuss the following. If

$$S_n^2 := (n-1)^{-1}\sum_{l=1}^{n}(X_l - \bar{X})^2,$$

then why is $(n-1)S_n^2/\sigma^2$ is Chi-squared with $(n-1)$ degrees of freedom?

Recall that if $\xi_1, \dots, \xi_m$ are iid standard normal then

$$\chi_m^2 \stackrel{D}{=} \sum_{j=1}^m \xi_j^2$$

is the (centered) chi-squared RV with m degrees of freedom. A direct calculation shows that

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + [(n-1)/n](X_n - \bar{X}_{n-1})^2, \tag{12}$$

where $\bar{X}_{n-1} = (n-1)^{-1}\sum_{l=1}^{n-1} X_l$. Now we use the induction. For $n=2$ we have (I see this directly since $(1/2)(X_2 - X_1) \sim N(0,1)$)

$$S_2^2 = (1/2)(X_2 - X_1)^2 \sim \chi_1^2.$$

Next, if $(m-1)S_m^2 \stackrel{D}{=} \chi_{m-1}^2$, we should prove that $mS_{m+1}^2 \stackrel{D}{=} \chi_m^2$. To verify this we use (11) and write,

$$mS_{m-1}^2 = (m-1)S_m^2 + (m/(m+1))(X_{m+1} - \bar{X}_m)^2.$$

On the right side, the first term is χ_{m-1}^2 by the induction's assumption, and the second is independent of S_m because X_{m+1} is obviously independent and $\bar{X}_m$ is independent by Basu's Theorem. Plus we can check directly that $(m/(m+1))(X_{m+1} - \bar{X})^2 \stackrel{D}{=} \chi_1^2$. We conclude that mS_{m+1} has chi-squared distribution with m degress of freedom.

Remark: If you note that $\sum_{l=1}^n X_l^2 = n\bar{X} + \sum_{l=1}^n (X_l - \bar{X})^2$ where $n^{1/2}\bar{X} \sim N(0, \sigma^2)$ and it is independent of $\sum_{l=1}^n (X_l - \bar{X})^2$, then this is another way to look (establish) the property.

Now let us look at the problem in the text. Let $t = s^2$. Then I use the chi-squared density

$$F_{S_n|\sigma^2}(t|\sigma^2) = \frac{1}{\Gamma((n-1)/2)2^{(n-1)/2}}[(n-1)\sigma^{-2}t]^{(n-1)/2-1}e^{-(n-1)t/2\sigma^2}[(n-1)/\sigma].$$

Here I used the fact that if A RV Z has the pdf $f_Z(z)$ then (do you remember how to verify it?)

$$f_{bZ}(y) = b^{-1}f_Z(y/b).$$

Now I use the given prior pfd $\pi(\sigma^2)$ and get (as usual in Bayesian analysis, we are interested in factors depending only on σ^2)

$$\pi(\sigma^2|t) \propto [(1/\sigma^2)^{(n-1)/2-1}e^{-(n-1)t/(2\sigma^2)}\sigma^{-2}][\sigma^{-2\alpha-1}e^{-1/(\beta\sigma^2}]$$

$$\propto (1/\sigma^2)^{(n-1)/2+\alpha+1}\exp\{-\sigma^{-2}[(n-1)t/2 + (1/\beta)]\}.$$

As you see, this is again the inverted gamma pdf, that is, $\pi(\sigma^2|t)$ is $IG(a,b)$ with

$$a = (n-1)/2 + \alpha, \quad b = [(n-1)t/2 + (1/\beta)]^{-1}.$$

Because for the IG distribution the mean is $1/[(a-1)b]$, we get

$$\hat{\sigma}_{Bayes}^2 = E(\sigma^2|t) = \frac{\frac{n-1}{2}t + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1} = \frac{\frac{n-1}{2}S_n^2 + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1}.$$

3. Exers. 7.24. Let $X_1, \ldots, X_n$ be a sample from $Poisson(\lambda)$ and $\lambda \sim Gamma(\alpha, \beta)$.
(a). What is the posterior distribution of Λ? Write

$$\pi(\lambda|\mathbf{x}) = \frac{\pi(\lambda) f_{\mathbf{X}|\lambda}(\mathbf{x}|\lambda)}{f_{\mathbf{X}}(\mathbf{x})} \propto \lambda^{\beta-1} e^{-\lambda/\beta} e^{-n\lambda} \lambda^{\sum_{l=1}^n x_l}$$

$$= \lambda^{(\beta+\sum_{l=1}^n x_l)-1} \exp\{-\frac{\lambda}{\beta/(n\beta+1)}\} \sim Gamma(\alpha + \sum_{l=1}^n x_l, \frac{\beta}{n\beta+1}).$$

(b) Write (using the previous result and a formula for the mean of a Gamma RV,

$$\hat{\lambda}_{Bayes} = E(\Lambda|\mathbf{X}) = (\alpha + \sum_{l=1}^n X_l)\frac{\beta}{n\beta+1} + \frac{1}{n\beta+1}(\alpha\beta).$$

Note that we again see the familiar structure in the Bayes estimate as an average because the prior mean and the MLE.

For the variance we similarly get

$$\mathrm{Var}(\hat{\lambda}_{Bayes}|\mathbf{X}) = (\alpha + \sum_{l=1}^n X_l)\frac{\beta^2}{(n\beta+1)^2}.$$

4. Exerc. 7.25. A sequence of RVs $X_1, \ldots, X_n$ is observed. It is known that $X_i|\theta_i \sim N(\theta_i, \sigma^2)$ and $\theta_i \sim N(\mu, \tau^2)$.

We begin with part (b) because we will use it in part (a).

(b) Assertion: If $X_i|\theta_i$ are independent with the pdf $f(x|\theta_i)$, and Θ_i are iid according to the pdf $\pi(\theta|\tau)$ then $X_1, \ldots, X_n$ are iid. Let us prove this assertion.

The joint pdf of $(\mathbf{X}, \Theta)$ is

$$f(\mathbf{x}, \theta|\tau) = \prod_{l=1}^n f(x_l, \theta_l|\tau) = \prod_{l=1}^n \pi(\theta_l|\tau) f(x_l|\theta).$$

Then

$$f(\mathbf{x}|\tau) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{l=1}^n \pi(\theta_l|\tau) f(x_l|\theta_l) d\theta_l$$

$$= \int_{-\infty}^{\infty} \cdots \{\int_{-\infty}^{\infty} \pi(\theta_1|\tau) f(x_1|\theta_1, \tau) dx_1\} \prod_{l=2}^n \pi(\theta_l|\tau) f(x_l|\theta_l) d\theta_l.$$

Note that the integral in the curly brackets is $f(x_1|\tau)$, and then repeating taking integrals (note that in the ofirst line there are n integrals) and taking into account that all obtaoined pdfs are the same we obtain that

$$f(\mathbf{x}|\tau) = \prod_{l=1}^n f(x_l|\tau).$$

This shows that marginally all X's are iid.

(a) Using the result of part (b) we can consider

$$f_{X|\mu,\sigma^2,\tau^2}(x) = \int_{-\infty}^{\infty} f(x|\theta,\sigma^2)\pi(\theta|\mu,\tau^2)d\theta$$

$$= \int_{-\infty}^{\infty} [2\pi\sigma\tau]^{-1} e^{-(x-\theta)^2/(2\sigma^2)} e^{-(\theta-\mu)^2/(2\tau^2)} d\theta.$$

Now we do some calculations for the product of the two exponential functions in the last integral. Write

$$e^{-(x-\theta)^2/(2\sigma^2)} e^{-(\theta-\mu)^2/(2\tau^2)} = \exp\{-\frac{(\theta - [\frac{x\tau^2}{\sigma^2+\tau^2} + \frac{\mu\sigma^2}{\sigma^2+\tau^2}])^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} - \frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\}.$$

Only the first term in the exponent involves θ. Let us denote it as $-T(\theta)$. Then it is the kernel of a normal pdf so

$$\int_{-\infty}^{\infty} e^{-T(\theta)} d\theta = (2\pi)^{1/2} \frac{\sigma\tau}{(\sigma^2+\tau^2)^{1/2}}.$$

Combining the results we conclude that

$$f_{X|\mu,\sigma^2,\tau^2}(x) = (2\pi\sigma\tau)^{-1}(2\pi)^{1/2} \frac{\sigma\tau}{(\sigma^2+\tau^2)^{1/2}} \exp\{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\}.$$

Simplify it a bit and you can clearly see that this is the pdf of $N(\mu, \sigma^2+\tau^2)$. (As we knew before, sure.)

5. Exerc. 7.27 (a) First, let us write down the log-likelihood,

$$l(\beta,\tau|(\mathbf{x},\mathbf{y})) = \sum_{l=1}^{n} [-\beta\tau_l + y_l \ln(\beta) + y_l \ln(\tau_l) - \ln(y_l!) - \tau_l + x_l \ln(\tau_l) - \ln(x_l!)].$$

Take partial derivatives with respect to the parameters and then equate them to zero. This allows us to find extremes. The equations are

$$-\sum_{l=1}^{n} \tau_l + \beta^{-1} \sum_{l=1}^{n} y_l = 0,$$

and

$$-\beta + \tau_l^{-1} y_l - 1 + \tau_l^{-1} x_l = 0, \quad l = 1, 2, \ldots, n.$$

We got a system of $n+1$ equations with $n+1$ parameters which we can solve. Write,

$$\beta^* = \frac{\sum_{l=1}^n y_l}{\sum_{l=1}^n \tau_l}, \qquad \tau_l^* = \frac{x_l + y_l}{\beta^* + 1}, \;\; l = 1, 2, \ldots, n.$$

Taking the sum of τ_l^* yields $\sum_{l=1}^n \tau_l = (\beta+1)^{-1} \sum_{l=1}^n (x_l + y_l)$, and we can simplify the relation for β^* in the following way,

$$\beta^* = \frac{(\beta^*+1)\sum_{l=1}^n y_l}{\sum_{l=1}^n (x_l + y_l)}$$

which in its turn yields

$$\beta^* = \frac{\sum_{l=1}^n y_l}{\sum_{l=1}^n x_l}.$$

Now you should check that the extreme point is the maximum. Here either graphic, or the following analysis can be used. Note that for any vector τ an extrema in β is the maximum (the second derivative is $-\beta^{-2}\sum_{l=1}^n y_l$, it is negative with the exception of the event with the exponentially vanishing probability when $\sum_{l=1}^n y_l = 0$ which is a special case to consider! What do you think the MLE should be in this case?). The same is true for any τ_l. We conclude that $(\beta^*, \tau_l^*, l = 1, 2, \ldots, n)$ are the MLE.

(b) Let $\hat\beta_n$ and $\hat\tau_l$ are the limit points. Then they satisfy

$$\hat\beta = \frac{\sum_{l=1}^n y_l}{\hat\tau_1 + \sum_{l=2}^n x_l}, \quad \hat\tau_1 = \frac{\hat\tau_1 + y_1}{\hat\beta + 1}, \quad \hat\tau_i = \frac{x_i + y_i}{\hat\beta + 1}, \; i = 2, \ldots, n.$$

Solve this system of equations and you get (7.2.16). Well, what is the simplest way to do that? You may begin with the second equation and get $\hat\tau_1 = y_1/\hat\beta$, substitute it into the first equation and get $\hat\beta = \frac{\sum_{l=2}^n y_l}{\sum_{l=2}^n x_l}$. Then you sum over i in the third equation, substitute $\hat\beta$ and get $\sum_{i=2}^n \hat\tau_i = \sum_{i=2}^n x_i$, and then plug into the first equation; this yields the first equation in (7.2.16). The two others have being already obtained.

(c) We already did this in part (b).

6. Exerc. 7.30 Note that if $U \sim f(u)$, $W \sim g(w)$ and $Z \sim Bernoulli(p)$, then

$$X \stackrel{D}{=} ZU + (1 - Z)W.$$

This is how you generate a RV with the mixture distribution!

This remark can help you to understand the considered setting of "missed data" where observations of Z are missed.

(a) Write for $z \in \{0, 1\}$,

$$f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}|p) = f_{\mathbf{Z}}(\mathbf{z}|p) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}, p) = \prod_{l=1}^n [pf(x_l)]^{z_l}[(1-p)g(x_l)]^{1-z_l}.$$

To see why this expression is correct, note that $f_{X,Z}(x, 1|p) = pf(x)$ and $f_{X,Z}(x, 0|p) = (1 - p)g(x)$. (Note that there is a misprint in the text on page 361, first line).

(b) Consider

$$f_{Z|X}(z|x, p) = \frac{f_{Z,X}(z, x|p)}{f_X(x|p)} = \frac{[pf(x)]^z[(1-p)g(x)]^{1-z}}{\sum_{z=0}^1 [pf(x)]^z[(1-p)g(x)]^{1-z}} = \frac{[pf(x)]^z[(1-p)g(x)]^{1-z}}{(1-p)g(x) + pf(x)},$$

which is the pmf of $Bernoulli(pf(x)/[pf(x) + (1 - p)g(x)])$.

(c) Using results of part (b) we can write,

$$E\{\ln(L(p|\mathbf{X}, \mathbf{Z}))|p = \hat p, \mathbf{X} = \mathbf{x}\} = E\{\sum_{l=1}^n [Z_l \ln(pf(x_l)) + (1 - Z_l)\ln((1-p)g(x+l))]|p = \hat p, \mathbf{X} = \mathbf{x}\}$$

$$= \sum_{l=1}^{n} \Big[\frac{\hat{p} f(x_l)}{\hat{p} f(x_l) + (1-\hat{p}) g(x_l)} \ln(p f(x_l)) + \frac{(1-\hat{p}) g(x_l)}{\hat{p} f(x_l) + (1-\hat{p}) g(x_l)} \ln((1-p) g(x_l)) \Big].$$

The last expression, as a function in p, has a maximum (check this!) and the $p*$ is the solution of the equation

$$p^{-1} \sum_{l=1}^{n} \frac{\hat{p} f(x_l)}{\hat{p} f(x_l) + (1-\hat{p}) g(x_l)} - (1-p)^{-1} \sum_{l=1}^{n} \frac{(1-\hat{p}) g(x_l)}{\hat{p} f(x_l) + (1-\hat{p}) g(x_l)} = 0.$$

Solve it and get

$$\hat{p}^{(r+1)} = n^{-1} \hat{p}^{(r)} \sum_{l=1}^{n} \frac{f(x_l)}{\hat{p}^{(r)} f(x_l) + (1-\hat{p}^{(r)}) g(x_l)}.$$

7. Exerc. 7.31 The part (b) is well explained — just apply Jensen's inequality. For part (a), the first assertion is plain — just plug $\hat{\theta}^{(r)}$ in (7.2.19). Then, using (7.2.19) we can write,

$$\log L(\theta|\mathbf{y}) = E\{\log(L(\theta|\mathbf{y}, \mathbf{X}))|\theta^{(r)}, \mathbf{Y} = \mathbf{y}\} - E\{\log k(\mathbf{X}|\theta, \mathbf{y})|\theta^{(r)}, \mathbf{Y} = \mathbf{y}\}.$$

Now, we choose $\theta^{(r+1)}$ which maximizes the expected complete-data log-likelihood, that is

$$\theta^{(r+1)} := \mathrm{argmax}_{\theta \in \Omega} E\{\log(L(\theta|\mathbf{Y}, \mathbf{X}))|\theta^{(r)}, \mathbf{Y}\}.$$

At the same time, thanks to the part (b),

$$E\{\log k(\mathbf{X}|\theta^{(r)}, \mathbf{Y})|(\theta^{(r)}, \mathbf{Y}\} \geq E\{\log k(\mathbf{X}|\theta^{(r+1)}, \mathbf{Y})|(\theta^{(r)}, \mathbf{Y}\}.$$

This yields $\log(L(\theta^{(r+1)}|\mathbf{y})) \geq \log(L(\theta^{(r)}|\mathbf{y}))$.

Remark: Please pay attention to the Jensen inequality — it is very useful in solving many problems.