Solution to Chapter 7 Analytical Exercises

1. (a) Since $a(\mathbf{w}) \neq 1 \Leftrightarrow f(y|\mathbf{x};\boldsymbol{\theta}) \neq f(y|\mathbf{x};\boldsymbol{\theta}_0)$, we have $\text{Prob}[a(\mathbf{w}) \neq 1] = \text{Prob}[f(y|\mathbf{x};\boldsymbol{\theta}) \neq f(y|\mathbf{x};\boldsymbol{\theta}_0)]$. But $\text{Prob}[f(y|\mathbf{x};\boldsymbol{\theta}) \neq f(y|\mathbf{x};\boldsymbol{\theta}_0)] > 0$ by hypothesis.

   (b) Set $c(x) = \log(x)$ in Jensen's inequality. $a(\mathbf{w})$ is non-constant by (a).

   (c) By the hint, $\text{E}[a(\mathbf{w})|\mathbf{x}] = 1$. By the Law of Total Expectation, $\text{E}[a(\mathbf{w})] = 1$.

   (d) By combining (b) and (c), $\text{E}[\log(a(\mathbf{w}))] < \log(1) = 0$. But $\log(a(\mathbf{w})) = \log f(y|\mathbf{x};\boldsymbol{\theta}) - \log f(y|\mathbf{x};\boldsymbol{\theta}_0)$.

2. (a) (The answer on p. 505 is reproduced here.) Since $f(y \mid \mathbf{x};\boldsymbol{\theta})$ is a hypothetical density, its integral is unity:

$$\int f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y = 1. \tag{1}$$

   This is an identity, valid for any $\boldsymbol{\theta} \in \Theta$. Differentiating both sides of this identity with respect to $\boldsymbol{\theta}$, we obtain

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times 1)}{\mathbf{0}}. \tag{2}$$

   If the order of differentiation and integration can be interchanged, then

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y = \int \frac{\partial}{\partial\boldsymbol{\theta}} f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y. \tag{3}$$

   But by the definition of the score, $\mathbf{s}(\mathbf{w};\boldsymbol{\theta}) f(y \mid \mathbf{x};\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} f(y \mid \mathbf{x};\boldsymbol{\theta})$. Substituting this into (3), we obtain

$$\int \mathbf{s}(\mathbf{w};\boldsymbol{\theta}) f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times 1)}{\mathbf{0}}. \tag{4}$$

   This holds for any $\boldsymbol{\theta} \in \Theta$, in particular, for $\boldsymbol{\theta}_0$. Setting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we obtain

$$\int \mathbf{s}(\mathbf{w};\boldsymbol{\theta}_0) f(y \mid \mathbf{x};\boldsymbol{\theta}_0)\mathrm{d}y = \text{E}[\mathbf{s}(\mathbf{w};\boldsymbol{\theta}_0) \mid \mathbf{x}] = \underset{(p\times 1)}{\mathbf{0}}. \tag{5}$$

   Then, by the Law of Total Expectations, we obtain the desired result.

   (b) By the hint,

$$\int \mathbf{H}(\mathbf{w};\boldsymbol{\theta}) f(y|\mathbf{x};\boldsymbol{\theta})\mathrm{d}y + \int \mathbf{s}(\mathbf{w};\boldsymbol{\theta})\,\mathbf{s}(\mathbf{w};\boldsymbol{\theta})' f(y \mid \mathbf{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times p)}{\mathbf{0}}.$$

   The desired result follows by setting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

3. (a) For the linear regression model with $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \sigma^2)'$, the objective function is the average log likelihood:

$$Q_n(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2.$$

To obtain the concentrated average log likelihood, take the partial derivative with respect to σ^2 and set it equal to 0, which yields

$$\sigma^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 \equiv \frac{1}{n} SSR(\boldsymbol{\beta}).$$

Substituting this into the average log likelihood, we obtain the concentrated average log likelihood (concentrated with respect to σ^2):

$$Q_n(\boldsymbol{\beta}, \frac{1}{n}SSR(\boldsymbol{\beta})) = -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log(\frac{1}{n}SSR(\boldsymbol{\beta})).$$

The unconstrained ML estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$ of $\boldsymbol{\theta}_0$ is obtained by maximizing this concentrated average log likelihood with respect to $\boldsymbol{\beta}$, which yields $\widehat{\boldsymbol{\beta}}$, and then setting $\widehat{\sigma}^2 = \frac{1}{n}SSR(\widehat{\boldsymbol{\beta}})$. The constrained ML estimator, $(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2)$, is obtained from doing the same subject to the constraint $\mathbf{R}\boldsymbol{\beta} = \mathbf{c}$. But, as clear from the expression for the concentrated average log likelihood shown above, maximizing the concentrated average log likelihood is equivalent to minimizing the sum of squared residuals $SSR(\boldsymbol{\beta})$.

(b) Just substitute $\widehat{\sigma}^2 = \frac{1}{n}SSR(\widehat{\boldsymbol{\beta}})$ and $\widetilde{\sigma}^2 = \frac{1}{n}SSR(\widetilde{\boldsymbol{\beta}})$ into the concentrated average log likelihood above.

(c) As explained in the hint, both $\widehat{\sigma}^2$ and $\widetilde{\sigma}^2$ are consistent for σ_0^2. Reproducing (part of) (7.3.18) of Example 7.10,

$$-\,\mathrm{E}\big[\mathbf{H}(\mathbf{w}_t;\boldsymbol{\theta}_0)\big] = \begin{bmatrix} \frac{1}{\sigma_0^2}\,\mathrm{E}(\mathbf{x}_t\mathbf{x}_t') & \mathbf{0} \\ \mathbf{0}' & \frac{1}{2\sigma_0^4} \end{bmatrix}. \qquad (7.3.18)$$

Clearly, both $\widehat{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ are consistent for $-\,\mathrm{E}\big[\mathbf{H}(\mathbf{w}_t;\boldsymbol{\theta}_0)\big]$ because both $\widehat{\sigma}^2$ and $\widetilde{\sigma}^2$ are consistent for σ_0^2 and $\frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'$ is consistent for $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$.

(d) The $\mathbf{a}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$ in Table 7.2 for the present case are

$$\mathbf{a}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\beta} - \mathbf{c}, \quad \mathbf{A}(\boldsymbol{\theta}) = \Big(\underset{(r\times K)}{\mathbf{R}} \;\vdots\; \underset{(r\times 1)}{\mathbf{0}}\Big).$$

Also, observe that

$$\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t(y_t - \mathbf{x}_t'\widetilde{\boldsymbol{\beta}}) \\ -\frac{1}{2\widetilde{\sigma}^2} + \frac{1}{2\widetilde{\sigma}^4}\frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\widetilde{\boldsymbol{\beta}})^2 \end{bmatrix} = \frac{1}{SSR_R}\begin{bmatrix} \mathbf{X}'(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \\ 0 \end{bmatrix}$$

and

$$Q_n(\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log(\frac{1}{n}SSR_U),\;\; Q_n(\widetilde{\boldsymbol{\theta}}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log(\frac{1}{n}SSR_R).$$

Substitute these expressions and the expression for $\widehat{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ given in the question into the Table 7.2 formulas, and just do the matrix algebra.

(e) The hint is the answer.

(f) Let $x \equiv \frac{SSR_R}{SSR_U}$. Then $x \geq 1$ and $W/n = x - 1$, $LR/n = \log(x)$, and $LM/n = 1 - \frac{1}{x}$. Draw the graph of these three functions of x with x in the horizontal axis. Observe that their values at $x = 1$ are all 0 and the slopes at $x = 1$ are all one. Also observe that for $x > 1$, $x - 1 > \log(x) > 1 - \frac{1}{x}$.