Assignment 12

Exercise 7.12  (a) Derive the Jeffreys prior on $\sigma^2$ for a random sample from $N(0, \sigma^2)$. Is this prior proper or improper?

**Sketch of solution:** Letting $\theta = \sigma^2$, we differentiate the log-density twice to find that $I(\theta) = 1/(2\theta^2)$. Therefore, the Jeffreys prior on $(0, \infty)$ is the improper prior $1/\theta$.

(b) What is the Bayes estimator of $\sigma^2$ for the Jeffreys prior? Verify directly that this estimator is efficient.

**Sketch of solution:** The posterior density $p(\theta)$ is proportional to the prior times the likelihood, which means

$$p(\theta) \propto \theta^{-(n/2)-1} \exp\{-\beta/\theta\},$$

where $\beta = \sum_i^n X_i^2/2$. To find the posterior mean, we can evaluate

$$\frac{\int_0^\infty \theta^{-(n/2)} \exp\{-\beta/\theta\} d\theta}{\int_0^\infty \theta^{-(n/2)-1} \exp\{-\beta/\theta\} d\theta}$$

(which can be done using a change of variables $\eta = 1/\theta$ to change each integral into a gamma integral). Alternatively, we may identify the posterior density as an inverse gamma density with parameters $n/2$ and $\beta$. Either way, we find the posterior mean to be

$$\tilde{\theta}_n = \frac{\beta}{(n/2) - 1} = \frac{\sum_i^n X_i^2}{n - 2}.$$

To verify that this estimator is efficient, we note that there is a unique root of the likelihood equation, namely the maximum likelihood estimator, which equals $[(n - 2)/n] \tilde{\theta}_n$ and is efficient. Thus, it suffices to simply verify that

$$\sqrt{n} \tilde{\theta}_n \left(1 - \frac{n - 2}{n}\right) \overset{P}{\to} 0,$$

which follows because $\theta_n$ is bounded in probability.

Exercise 7.13 Let $X \sim \text{multinomial}(1, p)$, where $p$ is a $k$-vector for $k > 2$. Let $p^* = (p_1, \ldots, p_{k-1})$. Find $I(p^*)$.  

269
Sketch of solution: The log-likelihood is \( \sum_{i=1}^{k} X_i \log p_i \). Replacing \( p_k \) by \( 1 - p_1 - \cdots - p_{k-1} \), then differentiating, we obtain:

\[
\frac{\partial^2}{\partial p_i^2} \ell(p) = -\frac{X_i}{p_i^2} - \frac{X_k}{p_k^2},
\]

\[
\frac{\partial^2}{\partial p_i \partial p_j} \ell(p) = -\frac{X_k}{p_k^2}.
\]

Since \( E X_i = p_i \), we obtain:

\[
I_{ij}(p^*) = \begin{cases} 
    p_i^{-1} + p_k^{-1} & \text{if } i = j \\
    \frac{p_k^{-1}}{p_i} & \text{if } i \neq j 
\end{cases}
\]

This can be written as \( I(p^*) = \text{diag}(1/p^*) + 11^\top/p_k \), where \( 1/p^* \) is the vector of reciprocals of the elements of \( p^* \) and \( 1 \) is the \((k-1)\)-vector of all ones.

It was not assigned, but notice that you can easily invert this matrix by knowing that \( I^{-1}(p^*) \) is the asymptotic covariance matrix of the MLE \( X/n \) from an iid sample \( X_1, \ldots, X_n \). Thus, the CLT gives \( I^{-1}(p^*) \) directly as the upper \((k-1)\times(k-1)\) submatrix of \( \text{Cov}(X_1) \):

\[
I^{-1}(p^*) = \text{diag}(p^*) - p^*(p^*)^\top = 
\begin{pmatrix}
    p_1^0 (1 - p_1^0) & -p_1^0 p_2^0 & \cdots & -p_1^0 p_{k-1}^0 \\
    -p_1^0 p_2^0 & p_2^0 (1 - p_2^0) & \cdots & -p_2^0 p_{k-1}^0 \\
    \vdots & \ddots & \ddots & \vdots \\
    -p_{k-1}^0 p_1^0 & -p_{k-1}^0 p_2^0 & \cdots & p_{k-1}^0 (1 - p_{k-1}^0)
\end{pmatrix}.
\]

Exercise 8.1 Let \( X_1, \ldots, X_n \) be a simple random sample from a Pareto distribution with density

\[
f(x) = \theta c^\theta x^{-(\theta+1)} I\{x > c\}
\]

for a known constant \( c > 0 \) and parameter \( \theta > 0 \). Derive the Wald, Rao, and likelihood ratio tests of \( \theta = \theta_0 \) against a two-sided alternative.

Sketch of solution: The log-likelihood and its first two derivatives are

\[
\ell(\theta) = n \log \theta + n\theta \log c - (\theta + 1) \sum_{i=1}^{n} \log X_i,
\]

\[
\ell'(\theta) = \frac{n}{\theta} + n \log c - \sum_{i=1}^{n} \log X_i,
\]

\[
\ell''(\theta) = -\frac{n}{\theta^2}.
\]
Setting $\ell'(\theta) = 0$ gives the MLE $\hat{\theta} = n/\sum_{i=1}^{n} \log(X_i/c)$. Also, from $\ell''(\theta)$ we obtain $I(\theta) = 1/\theta^2$. For two-sided alternatives, we may square the Wald and Rao statistics and then compare them to $\chi^2_1$.

Wald: $W_n^2 = \frac{n}{\hat{\theta}_0^2}(\hat{\theta} - \theta_0) = n \left( \frac{\hat{\theta}}{\theta_0} - 1 \right)^2$

Rao: $R_n^2 = \frac{\hat{\theta}_0^2}{n} \left[ \frac{n}{\theta_0} - \sum_{i=1}^{n} (\log X_i/c) \right] = n \left( 1 - \frac{\theta_0}{\hat{\theta}_n} \right)^2$

LR: $2\Delta_n = 2n \log \left( \frac{\hat{\theta}_n}{\theta_0} \right) - 2(\hat{\theta}_n - \theta_0) \sum_{i=1}^{n} \log(X_i/c) = 2n \left( \log \frac{\hat{\theta}_n}{\theta_0} - 1 + \frac{\theta_0}{\hat{\theta}_n} \right)$

Each test rejects when its corresponding statistic is greater than the $1 - \alpha$ quantile of $\chi^2_1$.

Exercise 8.2 Suppose that $X$ is multinomial$(n, p)$, where $p \in \mathbb{R}^k$. In order to satisfy the regularity condition that the parameter space be an open set, define $\theta = (p_1, \ldots, p_k - 1)$. Suppose that we wish to test $H_0 : \theta = \theta^0$ against $H_1 : \theta \neq \theta^0$.

(a) Prove that the Wald and score tests are the same as the usual Pearson chi-square test.

**Sketch of solution:** From Exercise 7.13, we obtain

$I(\theta) = \text{diag}(1/p^*) + 11^T/p_k = \left( \begin{array}{cccc} \frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2} + \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{array} \right)$.

If we define $Y$ to be the $(k-1)$-dimensional vector where $Y_j = X_j/n$, $1 \leq j \leq k-1$, then we may express the chi-square statistic under the null hypothesis $\theta = \theta^0$ as

$$\chi^2 = n \sum_{j=1}^{k} \frac{(Y_j - \theta^0_j)^2}{\theta^0_j},$$

where $\theta_k$ is defined to be $1 - \theta_1 - \cdots - \theta_{k-1}$. A bit of simple algebra shows that

$$n \sum_{j=1}^{k-1} \frac{(Y_j - \theta^0_j)^2}{\theta^0_j} = n(Y - \theta^0)^\top \text{diag}(1/\theta^0)(Y - \theta^0)$$
and
\[ n \frac{(Y_k - \theta_0^k)^2}{\theta_0^k} = n(Y - \theta^0)^\top \frac{11}{\theta_0^k} (Y - \theta^0). \]

Therefore, we conclude that
\[ \chi^2 = n(Y - \theta^0)^\top \left[ \text{diag}(1/\theta^0) + \frac{11}{\theta_0^k} \right] (Y - \theta^0) \]
\[ = n(Y - \theta^0)^\top I(\theta^0)(Y - \theta^0). \]

This can immediately be seen to be the Wald statistic because the MLE of \( \theta \) is \( Y \).

For the score test (warning: tedious algebra to follow!), we first evaluate the score vector at \( \theta^0 \), which equals
\[ \nabla \ell(\theta^0) = \left( \frac{X_1}{\theta_0^1} - \frac{X_k}{1 - \theta_0^1 - \cdots - \theta_0^{k-1}}, \cdots, \frac{X_{k-1}}{\theta_0^{k-1}} - \frac{X_k}{1 - \theta_0^1 - \cdots - \theta_0^{k-1}} \right)^\top \]
\[ = \text{diag}(1/\theta^0)(nY) - \frac{X_k}{\theta_k} 1. \]

From the solution sketch for Exercise 7.13, we know that
\[ I^{-1}(\theta^0) = \text{diag}(\theta^0) - \theta^0(\theta^0)^\top = \begin{pmatrix} \theta_0^1(1 - \theta_0^1) & -\theta_0^1\theta_0^2 & \cdots & -\theta_0^1\theta_0^{k-1} \\ -\theta_0^1\theta_0^2 & \theta_0^2(1 - \theta_0^2) & \cdots & -\theta_0^2\theta_0^{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ -\theta_0^1\theta_0^{k-1} & -\theta_0^2\theta_0^{k-1} & \cdots & \theta_0^{k-1}(1 - \theta_0^{k-1}) \end{pmatrix}. \]

Thus, if we write \( \Gamma_0 = \text{diag}(\theta^0) \), the Rao score statistic is
\[ R_n = \frac{1}{n} \nabla \ell(\theta^0)^\top I^{-1}(\theta^0) \nabla \ell(\theta^0) \]
\[ = \frac{1}{n} \left[ \Gamma_0^{-1}(nY) - \frac{X_k}{\theta_0^k} 1 \right] \left[ \Gamma_0^{-1}(nY) - \frac{X_k}{\theta_0^k} 1 \right]. \]

Direct evaluation of this expression gives
\[ R_n = \frac{1}{n} \left[ nY^\top - \frac{X_k}{\theta_0^k} (\theta^0)^\top - nY^\top 1(\theta^0)^\top + \frac{X_k}{\theta_0^k} 1^\top \theta^0(\theta^0)^\top \right] \left[ \Gamma_0^{-1}(nY) - \frac{X_k}{\theta_0^k} 1 \right] \]
\[ = \frac{1}{n} \left[ nY^\top - \frac{X_k}{\theta_0^k} (\theta^0)^\top - nY^\top 1(\theta^0)^\top \right] \Gamma_0^{-1} \left[ \Gamma_0^{-1}(nY) - \frac{X_k}{\theta_0^k} \Gamma_0 1 \right] \]
\[ = \left[ Y^\top \Gamma_0^{-1} - 1^\top \right] \left[ nY - \frac{X_k}{\theta_0^k} \Gamma_0 1 \right]. \]
because \( 1 - 1^\top \theta^0 = \theta_k^0 \) and \( nY^\top 1 = n - X_k \). Continuing, we obtain

\[
R_n = nY^\top \Gamma^{-1}Y - \frac{X_k}{\theta_k^0} Y^\top 1 - n1^\top Y + \frac{X_k Y_k}{\theta_k^0} \Gamma_0 1
\]

which may be rewritten as

\[
R_n = n \sum_{j=1}^k \frac{Y_j^2}{\theta_j^0} - n = n \sum_{j=1}^k \frac{(Y_j - \theta_j^0)^2}{\theta_j^0}
\]

because \( 1 = \sum_j Y_j = \sum_j \theta_j^0 \). This completes the proof.

(b) Derive the likelihood ratio statistic \( 2\Delta_n \).

**Sketch of solution:** The log-likelihood function, ignoring a constant that does not depend on \( \theta \), is

\[
\ell(\theta) = \theta_1^{nY_1} \theta_2^{nY_2} \cdots \theta_{k-1}^{nY_{k-1}} (1 - \theta_1 - \cdots - \theta_{k-1})^{n - nY_1 - \cdots - nY_{k-1}}.
\]

Therefore, we obtain (since \( \hat{\theta}_n = Y \))

\[
2\Delta_n = 2n \sum_{j=1}^{k-1} Y_j \log \frac{\theta_j^0}{Y_j} + 2n(1 - Y_1 - \cdots - Y_{k-1}) \log \frac{1 - \theta_1 - \cdots - \theta_{k-1}}{1 - Y_1 - \cdots - Y_{k-1}}.
\]

Comparing \( 2\Delta_n \) to a \( \chi^2_{k-1} \) distribution is an interesting alternative to the usual Pearson chi-square test.