Homework 10

STAT 984

Emily Robinson

December 5, 2019

Exercise 8.1

Let $X_1, ..., X_n$ 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.

Consider

$$\begin{aligned}
& & L(\theta) &= \theta^n c^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)} \\
&\implies & \ell(\theta) &= n\log(\theta) + n\theta\log(c) - (\theta+1)\sum_{i=1}^n \log(x_i) \\
&\implies & \ell'(\theta) &= \frac{n}{\theta} + n\log(c) - \sum_{i=1}^n \log(x_i) \\
& & &= \frac{n}{\theta} - \sum_{i=1}^n \log\left(\frac{x_i}{c}\right) \\
&\implies & \ell''(\theta) &= -\frac{n}{\theta^2}.
\end{aligned}$$

Therefore, setting $\ell'(\theta) = 0$, implies

$$\begin{aligned}
& & \frac{n}{\theta} - \sum_{i=1}^n \log\left(\frac{x_i}{c}\right) &= 0 \\
&\implies & n - \theta\sum_{i=1}^n \log\left(\frac{x_i}{c}\right) &= 0 \\
&\implies & \hat{\theta}_n &= \frac{n}{\sum_{i=1}^n \log\left(\frac{x_i}{c}\right)}.
\end{aligned}$$

and from $\ell''(\theta)$, we obtain $I(\theta) = \frac{1}{\theta^2}$.

Thus, for a two -sided alternative,

$$W_n^2 = \frac{n}{\theta_0}(\hat{\theta}_n - \theta_0) = n\left(\frac{\hat{\theta}_n}{\theta_0} - 1\right)^2 = n\left(\frac{n}{\theta_0\sum_{i=1}^n \log(\frac{x_i}{c})} - 1\right)^2.$$

Reject when $W_n^2 > \chi^2_{1,(1-\alpha)}$.

$$R_n^2 = \frac{\theta_0^2}{n}\left[\frac{n}{\theta_0} - \sum_{i=1}^n \log\left(\frac{x_i}{\theta}\right)\right]^2 = n\left[1 - \frac{\theta_0 \sum_{i=1}^n \log\left(\frac{x_i}{c}\right)}{n}\right]^2.$$

Reject when $R_n^2 > \chi^2_{1,(1-\alpha)}$.

$$\begin{aligned}
2\Delta_n &= 2\left(\ell(\hat{\theta}_n) - \ell(\theta_0)\right) \\
&= 2\left[n\log\left(\frac{\hat{\theta}_n}{\theta_0}\right) + n\log(c)(\hat{\theta}_n - \theta_0) - (\hat{\theta}_n - \theta_0)\sum_{i=1}^n \log(x_i)\right] \\
&= 2\left[n\log\left(\frac{\hat{\theta}_n}{\theta_0}\right) - (\hat{\theta}_n - \theta_0)\sum_{i=1}^n \log\left(\frac{x_i}{c}\right)\right] \\
&= 2n\left[\log\left(\frac{\hat{\theta}_n}{\theta_0}\right) - (\hat{\theta}_n - \theta_0)\frac{1}{\hat{\theta}_n}\right] \\
&= 2n\left[\log\left(\frac{\hat{\theta}_n}{\theta_0}\right) - 1 = \frac{\theta_0}{\hat{\theta}_n}\right].
\end{aligned}$$

Reject when $2\Delta_n > \chi^2_{1,(1-\alpha)}$.

Exercise 8.2

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

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

Proof. In class, we showed

$$I(\theta) = \text{Diag}\left(\frac{1}{\boldsymbol{p}^*}\right) + \frac{\mathbf{1}\mathbf{1}^T}{p_k}$$

Let $Y_j = \frac{X_j}{n}$ for $1 \le j \le k-1$. Then, under the null, the Pearson Chi-square statistic is

$$\chi^2 = n\sum_{i=1}^k \frac{\left(Y_j - \theta_j^0\right)^2}{\theta_j^0}$$

where $\theta_k = 1 - \sum_{j=1}^k -1\theta_j$. Then we can show

$$n\sum_{i=1}^{k-1} \frac{\left(Y_j - \theta_j^0\right)^2}{\theta_j^0} = n\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)^T \text{Diag}\left(\frac{1}{\boldsymbol{\theta}^0}\right)\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)$$

and

$$n\frac{(Y_k - \theta_k^0)^2}{\theta_k^0} = n\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)^T \text{Diag}\left(\frac{1}{\boldsymbol{\theta}_k^0}\right)\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right).$$

Therefore,

$$\chi^2 = n\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)^T \text{Diag}\left(\frac{1}{\boldsymbol{\theta}^0} + \frac{1}{\boldsymbol{\theta}_k^0}\right)\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right) = n\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)^T I(\boldsymbol{\theta}^0)\left(\boldsymbol{Y} - \boldsymbol{\theta}^0\right)$$

Thus, the Wald test is the same as the usual Pearson chi-square test.

Then

$$\Delta\ell(\boldsymbol{\theta}^0) = \text{Diag}\left(\frac{1}{\boldsymbol{\theta}^0}\right)(n\boldsymbol{Y}) = \frac{X_k}{\theta_k}\mathbf{1}.$$

From Exercise 7.13, $I^{-1}(\boldsymbol{\theta}^0) - \boldsymbol{\theta}^0(\boldsymbol{\theta}^0)^T$. Therefore,

$$\begin{aligned}
R_n &= \frac{1}{n}\Delta\ell(\boldsymbol{\theta}^0)I^{-1}(\boldsymbol{\theta}^0)\Delta\ell(\boldsymbol{\theta}^0) = \left[\boldsymbol{Y}^T\text{Diag}(\boldsymbol{\theta}^0) - \mathbf{1}^T\right]\left[(n\boldsymbol{Y}) - \frac{X_k}{\theta_k^0}\text{Diag}\left(\boldsymbol{\theta}^0\right)\mathbf{1}\right] \\
&= n\boldsymbol{Y}^T\boldsymbol{Y} + \frac{X_kY_k}{\theta_k^0} - n \\
&= n\sum_{j=1}^{k}\frac{Y_j^2}{\theta_j^0} - n \\
&= n\sum j = 1^k\frac{(Y_j - \theta_j^0)^2}{\theta_j^0}
\end{aligned}$$

Thus, the Score test is the same as the usual Pearson chi-square test. □

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

Consider

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

Therefore, since $\hat{\boldsymbol{\theta}}_n = \boldsymbol{Y}$,

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

. Reject when $2\Delta_n > \chi^2_{k-1,(1-\alpha)}$.

Exercise 8.8

Let $X_1, ..., X_n$ be an independent sample from an exponential distribution with mean λ, and $Y_1, ..., Y_n$ be an independent sample from an exponential distribution with mean μ. Assume that X_i and Y_i are independent. We are interested in testing the hypothesis $H_0 : \lambda = \mu$ verses $H_1 : \lambda > \mu$. Consider the statistic

$$T_n = 2\sum_{i=1}^{n}(I_i - 1/2)/\sqrt{n},$$

where I_i is the indicator variable $I_i = I(X_i > Y_i)$.

(a) Derive the asymptotic distribution of T_n under the null hypothesis.

Under the null, $\lambda = \mu$ implies $X_i \stackrel{d}{=} Y_i$. Therefore, $I_i \sim$Bern$(1/2)$ and $\sum_{i=1}^n I_i \sim$Bin$(n, 1/2)$. Therefore,

$$\begin{aligned}
T_n &= \frac{2(I_1 - 1/2 + I_2 - 1/2 + ... + I_n - 1/2)}{\sqrt{n}} \\
&= \frac{2[\sum_{i=1}^n (I_i) - n/2]}{\sqrt{n}} \\
&= 2 \cdot \frac{1}{\sqrt{n}} \left[I_n - \frac{1}{2n} \right] \cdot \frac{n}{n} \\
&= 2\sqrt{n} \left[\frac{I_n}{n} - \frac{1}{2} \right] \\
&\xrightarrow{d} 2 \cdot N(0, 1/4) \\
&\stackrel{d}{=} N(0, 1).
\end{aligned}$$

(b) Use the Lindeberg Theorem to show that, under the local alternative hypothesis $(\lambda_n, \mu_n) =)\lambda + n^{-1/2}\delta, \lambda)$, where $\delta > 0$,

$$\frac{\sum_{i=1}^n (I_i - \rho_n)}{\sqrt{n\rho_n(1-\rho_n)}} \xrightarrow{\mathbb{L}} N(0, 1), \text{ where } \rho_n = \frac{\lambda_n}{\lambda_n + \mu_n} = \frac{\lambda + n^{-1/2}\delta}{2\lambda + n^{-1/2}\lambda}.$$

Under the local alternative, $I_i \sim$Bern(ρ_n). Then by (4.15) in Example 4.17, we know

$$\frac{\sum_{i=1}^n (I_i - \rho_n)}{\sqrt{n\rho_n(1-\rho_n)}} = \frac{I_n - n\rho_n}{\sqrt{n\rho_n(1-\rho_n)}} \xrightarrow{d} N(0, 1)$$

whenever $n\rho_n(1-\rho_n) \to \infty$ as $n \to \infty$.

Exercise 8.9

Suppose $X_1, ...X_m$ is a simple random sample and $Y_1, ..., Y_n$ is another simple random sample independent of the X_i, with $P(X_i \le t) = t^2$ for $t \in [0, 1]$ and $P(Y_i \le t) = (t - \theta)^2$ for $t \in [\theta, \theta + 1]$. Assume $m/(m + n) \to \rho$ as $m, n \to \infty$ and $0 < \theta < 1$.

Find the asymptotic distribution of $\sqrt{m+n}[g(\bar{Y} - \bar{X}) - g(\theta)]$.

Consider $E[X] = 2/3$, $E[Y] = \theta + 2/3$, and $Var(X) = Var(Y) = 1/18$. Then from the CLT, we know $(\bar{X} - 2/3) \xrightarrow{d} N\left(0, \frac{1}{18m}\right)$ and $\left(\bar{Y} - (\theta + 2/3)\right) \xrightarrow{d} N\left(0, \frac{1}{18m}\right)$. Therefore, since X_i and Y_i are independent,

$$(\bar{Y} - \bar{X}) - (\theta + 2/3 - 2/3) \xrightarrow{d} N\left(0, \frac{1}{18m} + \frac{1}{18n}\right).$$

Therefore,

$$\sqrt{m+n}\left[(\bar{Y} - \bar{X}) - \theta\right] \xrightarrow{d} N\left(0, \frac{\left(\frac{n+m}{m}\right)\left(\frac{n+m}{n}\right)}{18}\right) \stackrel{d}{=} N\left(0, \frac{1}{18\rho(1-\rho)}\right).$$

Thus, using the delta rule,

$$\sqrt{m+n}\left[g(\bar{Y}-\bar{X})-g(\theta)\right] \xrightarrow{d} N\left(0, \frac{g'(\theta)^2}{18\rho(1-\rho)}\right).$$