Ma 494 — Theoretical Statistics

Test #2 — Solutions for April 14, 2010

Prof. Sawyer — Washington University

Take-home examination. Open book and notes. Due at end of period on 04/14/2010. Six (6) problems on 3 pages. Not all parts of problems will be equally weighted.

1. (See Section 8 of the Math494 notes for the background.)

(i) If P_i are the $n = 8$ listed P-values, then the values $X_i = 2\log(1/P_i)$ are 8.5374, 6.43775, 5.46674, 5.23459, 3.1213, 2.34237, 1.12424, and 0.278524, respectively, and $X = \sum_{i=1}^{8} X_i = 32.5429$. The P-value of Fisher's meta-analysis test for these 8 P-values is $P = P(\chi_{16} \geq 32.5429) = 0.00849139$, so that we reject H_0 at $\alpha = 0.01$ as well as at $\alpha = 0.05$.

(ii) Consider the power-law density $f(x, \alpha) = \alpha x^{\alpha-1}$ for $0 \leq x \leq 1$. If H_0 is true, then P_i are uniformly distributed in $(0, 1)$, so that $\alpha = 1$. Fisher's test is UMP (uniformly most powerful) against any alternative $\alpha < 1$, for which the P-values P_i are more concentrated near $P = 0$. (See Section 8 in the notes.)

2. By Section 7.1 in the Math494 notes, the GLRT statistic $\widehat{LR}_n(X)$ (with the hypothesis H_1 in the numerator) is $\widehat{LR}_n(X) = 1/(X_{\max})^n$, with rejection of H_0 for large values of $\widehat{LR}_n(X)$. Given H_0, $X_1, \ldots, X_n$ are uniformly distributed in $(0, 1)$. The critical regions of the GLRT test are $\mathcal{C}_\alpha = \{ X : X_{\max} \leq \lambda \}$ with $\lambda = \lambda_\alpha$ determined by the level of significance $P(\mathcal{C}_\alpha \mid H_0) = P(X_{\max} \leq \lambda \mid H_0) = P(\max_{1 \leq i \leq n} X_i \leq \lambda_\alpha) = \lambda^n = \alpha$, or $\lambda = \lambda_\alpha = \alpha^{1/n}$.

In particular $\lambda = 0.05^{1/n}$ if $\alpha = 0.05$. Thus $X_{\max} = 0.80$ is in the critical region (or $X \in \mathcal{C}_\alpha$) if $X_{\max} = 0.80 \leq \lambda = 0.05^{1/n}$, or if $\log(0.80) \leq (1/n)\log(0.05)$ or $n \geq \log(0.05)/\log(0.80) = 13.4251$, so that $X_{\max} = 0.80$ causes H_0 to be rejected in favor of H_1 if and only if $n \geq 14$.

3. By standard results, $S^2 = (1/(n-1))\sum_{j=1}^{n}(X_j - \overline{X})^2 \approx (\sigma^2/(n-1))Y$ where Y has a χ^2 distribution with $n - 1$ degrees of freedom (that is, $Y \approx \chi^2_{n-1}$). Thus $\mathrm{Var}(S^2) = (\sigma^4/(n-1)^2)\,\mathrm{Var}(Y)$. Since $Y \approx \chi^2_{n-1} \approx \mathrm{gamma}\big((n-1)/2, 1/2\big)$ and $\mathrm{Var}\big(\mathrm{gamma}(\alpha, \beta)\big) = \alpha/\beta^2$, we conclude $\mathrm{Var}(Y) = \big((n-1)/2\big)/(1/2)^2 = 2(n-1)$ and $\mathrm{Var}(S^2) = (\sigma^4/((n-1)^2)(2(n-1)) = 2\sigma^4/(n-1)$.

4. Abbreviating $X = (X_1, X_2, \ldots, X_n)$, the likelihood is

$$\begin{aligned} L(\theta, X) &= \prod_{i=1}^{n} f(X_i, \theta) = (1/2)^n \theta^n \left(\prod_{i=1}^{n} X_i^{-3/2} \right) \prod_{i=1}^{n} \exp\left(-\theta \frac{1}{\sqrt{X_i}} \right) \\ &= \theta^n \, C(X) \exp\left(-\theta \sum_{i=1}^{n} \frac{1}{\sqrt{X_i}} \right) \end{aligned}$$

where $C(X)$ depends only on $X_1, \ldots, X_n$. Thus

$$\log L(\theta, X) \;=\; n\log(\theta) + \log C(X) - \theta \sum_{i=1}^{n} \frac{1}{\sqrt{X_i}}$$

The MLE $\widehat{\theta}$ is found by solving

$$\frac{\partial}{\partial\theta} \log L(\theta, X) \;=\; \frac{n}{\theta} \;-\; \sum_{i=1}^{n} \frac{1}{\sqrt{X_i}} = 0$$

Thus

$$\widehat{\theta} \;=\; \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{\sqrt{X_i}}}$$

5. (i) By definition, χ_n^2 has the same distribution as $Z_1^2 + Z_2^2 + \ldots + Z_n^2$, where $Z_1, Z_2, \ldots, Z_n$ are independent normal $N(0,1)$ random variables. In particular

$$P\left(\frac{\chi_n^2 - n}{\sqrt{2n}} \;\le\; y\right) \;=\; P\left(\frac{Z_1^2 + \ldots + Z_n^2 - n}{\sqrt{2n}} \;\le\; y\right)$$

Since $Z_i^2 \approx \chi_1^2 \approx \text{gamma}(1/2, 1/2)$ and, if $Y \approx \text{gamma}(\alpha, \beta)$, $E(Y) = \alpha/\beta$ and $\text{Var}(Y) = \alpha/\beta^2$, it follows that $\mu = E(Z_i^2) = 1$ and $\sigma^2 = \text{Var}(Z_i^2) = 2$. Thus by the central limit theorem applied to the Z_i^2

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{\chi_n^2 - n}{\sqrt{2n}} \;\le\; y\right) \;&=\; \lim_{n\to\infty} P\left(\frac{Z_1^2 + \ldots + Z_n^2 - n}{\sqrt{2n}} \;\le\; y\right) \\
&=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(1/2)x^2}\, dx
\end{aligned}$$

for all y. This means that $Z = (\chi_n^2 - n)/\sqrt{2n}$ is approximately standard normal. Solving for χ_n^2 leads to $\chi_n^2 = n + Z\sqrt{2n}$.

(ii) If $A = (\chi_{48}^2)_{0.90}$, then, using the approximation in part (ii), $P(\chi_{48}^2 \le A) = P(48 + Z\sqrt{96} \le A) = P\big(Z \le (A-48)/\sqrt{96}\big) = 0.90$ and $(A-48)/\sqrt{96}$ is approximately 1.28155. Thus A is approximately $48 + 1.28155\sqrt{96} = 60.557$. From Table A.3 pp856–857 in the back of the text, $(\chi_{48}^2)_{0.90} = 60.907$, so that the approximate value is approximately 0.350 or 0.57% too small.

(iii) By the same argument, $(\chi_{150}^2)_{0.90} = 150 + 1.28155\sqrt{300} = 172.197$.

6. By equation (3) in the problem,

$$2\log\widehat{LR}_n(X) \;=\; 2\left(\frac{n}{2}\log\left(1\;+\;\frac{T(X)^2}{n-1}\right)\right) \;=\; n\log\left(1\;+\;\frac{T(X)^2}{n-1}\right)$$

Thus

$$\begin{aligned} P\big(2\log\widehat{LR}_n(X)\le y\big) \;&=\; P\left(n\log\left(1\;+\;\frac{T(X)^2}{n-1}\right)\;\le\;y\right) \qquad (1)\\ =\;P\left(\log\left(1\;+\;\frac{T(X)^2}{n-1}\right)\;\le\;\frac{y}{n}\right)\;&=\;P\left(1\;+\;\frac{T(X)^2}{n-1}\;\le\;e^{y/n}\right)\\ =\;P\left(T(X)^2\;\le\;(n-1)\big(e^{y/n}-1\big)\right)& \qquad (2)\end{aligned}$$

Note

$$\lim_{n\to\infty}(n-1)\big(e^{y/n}-1\big)\;=\;\lim_{n\to\infty}n\big(e^{y/n}-1\big)\;=\;y \qquad (3)$$

by L'Hôpital's rule. Given H_0, $T(X)$ has a Student-t distribution with $n-1$ degrees of freedom, so that $T(X)^2$ has the same distribution as

$$\frac{Z_0^2}{\frac{1}{n-1}\displaystyle\sum_{i=1}^{n-1}Z_i^2}$$

where $Z_0, Z_1, Z_2, \ldots$ are independent standard normal variables. By the law of large numbers, $\lim_{n\to\infty}\big(1/(n-1)\big)\sum_{i=1}^{n-1}Z_i^2 \;=\; 1$. Thus, in distribution, by the hints in Problem 6,

$$\begin{aligned}\lim_{n\to\infty}P\big(T(X)^2\le y\big)\;&=\;\lim_{n\to\infty}P\left(\frac{Z_0^2}{\frac{1}{n-1}\sum_{i=1}^{n-1}Z_i^2}\;\le\;y\right) \qquad (4)\\ =\;P\big(Z_0^2\le y\big)\;&=\;P\big(\chi_1^2\le y\big)\end{aligned}$$

Putting together (1), (2), (3), (4), and part (ii) of the hint implies

$$\lim_{n\to\infty}P\big(2\log\widehat{LR}_n(X)\le y\big)\;=\;P(\chi_1^2\le y)$$

for all y, which was to be proven.