MTH6136 STATISTICAL THEORY: Exercise Sheet 10

Please try to attempt these questions **before** the exercise class. You should post your solution **to Question 1 only** in the yellow box on the ground floor of the Mathematics Building by **11.00 on Thursday 1 April 2010**. This will be marked and returned to you the following week.

1. Suppose that $Y_1, \ldots, Y_{n_1}$ are independent Gamma(a, λ_1) random variables, and that $Y_{n_1+1}, \ldots, Y_{n_1+n_2}$ are independent Gamma(a, λ_2) random variables. We wish to test the hypothesis $H_0 : \lambda_1 = \lambda_2$ versus the general alternative $H_1 : \lambda_1 \neq \lambda_2$, and we assume that a is known. (Thus we want to compare two Gamma populations with a common known shape parameter).

   (a) Write down the likelihood $L = L_1 L_2$ where L_i is the likelihood from the i^{th} random sample $(i = 1, 2)$ and hence the loglikelihood $l(\underline{\theta}; \underline{y}) = l_1 + l_2$, where you should include the 'constants' (combined over the two samples) for part (e). Hence verify the maximum likelihood estimates are given by $\hat{\lambda}_i = a/\bar{y}_i$, $(i = 1, 2)$.

   (b) Find the restricted maximum likelihood estimate(s) $\underline{\hat{\theta}}_0$ under H_0 in terms of the overall sample mean $\bar{y}$, and hence show that the observed value of Wilks' statistic for testing H_0 is

   $$\begin{aligned} -2 \log \Lambda(\underline{y}) &= 2 \left[l(\underline{\hat{\theta}}; \underline{y}) - l(\underline{\hat{\theta}}_0; \underline{y}) \right] \\ &= 2a \left[n \log \bar{y} - n_1 \log \bar{y}_1 - n_2 \log \bar{y}_2 \right] \end{aligned}$$

   (c) What is the approximate sampling distribution of this statistic for large n? Write down (from a previous example) the test statistic based on an asymptotic pivot which has a standard Normal distribution under H_0, and show that the square of this statistic is *not* the Wilks' statistic (and hence the two tests are different).

   (d) Give a possible practical reason why the Wilks' statistic might be preferred.

   (e) Briefly outline how you would test H_0 when the common shape parameter a is unknown, writing down the two equations satisfied by $\hat{a}$ and $\hat{a}_0$ respectively.

2. The relationship between the volume of sales, in £1,000s, and the sales floor area, in square metres, for a chain of supermarkets is being investigated. The average weekly volume of sales for the ith supermarket is Y_i and the floor area is x_i for $i = 1, 2, \ldots, n$. Suppose that $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2 x_i)$ independently for $i = 1, 2, \ldots, n$, where σ^2 is known, and consider testing $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$.
   (a) Write down the likelihood, $L(\beta; \underline{y})$, and hence find the LR statistic, $\Lambda(\underline{y})$.
   (b) Show that the critical region of the generalised likelihood ratio test is of the form $R = \{\underline{y} : |\sum_{i=1}^n y_i - \sum_{i=1}^n x_i| \geq c\}$, where c is a constant chosen to give size α.
   (c) Explain why $\sum_{i=1}^n Y_i$ has a normal distribution, give its mean and variance, and hence show that, for a test at the 5% level of significance, $c = 1.96\sigma\sqrt{\sum_{i=1}^n x_i}$.
   (d) Now suppose that σ^2 is unknown. Briefly explain how Wilks' theorem may be used to find the critical region of a test with approximate size α for large samples.

MTH6136 STATISTICAL THEORY: Solutions to Exercise Sheet 10

1. (a)The likelihood is

$$L = \lambda_1^{n_1 a} e^{-\lambda_1 n_1 \bar{y}_1} \frac{\left(\prod_{i=1}^{n_1} y_i\right)^{a-1}}{[\Gamma(a)]^{n_1}} \lambda_2^{n_2 a} e^{-\lambda_2 n_2 \bar{y}_2} \frac{\left(\prod_{i=n_1+1}^{n_1+n_2} y_i\right)^{a-1}}{[\Gamma(a)]^{n_2}}, \quad \text{so}$$

$$l = l_1 + l_2 = n_1 a \log \lambda_1 - \lambda_1 n_1 \bar{y}_1 + n_2 a \log \lambda_2 - \lambda_2 n_2 \bar{y}_2 + (a-1)\sum_{i=1}^{n} \log y_i - n \log \Gamma(a).$$

$\frac{\partial l}{\partial \lambda_i} = 0 \;\rightarrow\; \frac{n_i a}{\lambda_i} = n_i \bar{y}_i$, $(i = 1, 2)$, giving the required MLEs.
(b) Under H_o, the data becomes a single random sample size n from Gamma(a, λ) where λ is the common (scale) parameter so the loglikelihood is

$$na \log \lambda - \lambda n \bar{y} + (a-1)\sum_{i=1}^{n} \log y_i - n \log \Gamma(a).$$

Differentiating this with respect to λ and setting to zero gives the restricted MLEs $\hat{\lambda}_{10} = \hat{\lambda}_{20} = \hat{\lambda} = a/\bar{y}$. Thus $-2\log\Lambda(\underline{y})$

$$\begin{aligned} &= 2\left[n_1 a \log \hat{\lambda}_1 - a n_1 + n_2 a \log \hat{\lambda}_2 - a n_2 - na \log \hat{\lambda} + na\right] \\ &= 2\left[n_1 a \log a + n_2 a \log a - na \log a + na \log \bar{y} - n_1 a \log \bar{y}_1 - n_2 a \log \bar{y}_2\right] \end{aligned}$$

which gives the required value of Wilks' statistic.
(c) The large sample distribution is in general $\chi^2_{p-p_0}$ and here $p = 2$, $p_0 = \dim \omega = 1$ (corresponding to 1 restriction under H_0, so we use χ^2_1 (and we reject H_0 if Wilks' statistic is greater than $\chi^2_{1;\alpha}$ for a test of size α). From the previous example in chapter 4 (replacing $\hat{a}$ by a when it is known)

$$\frac{\sqrt{a}(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\frac{\bar{Y}_1^2}{n_1} + \frac{\bar{Y}_1^2}{n_1}}} \sim N(0,1) \quad \text{approx. for large } n_1,\, n_2 \text{ under } H-0$$

or alternatively, since $\hat{\lambda}_i \sim N[\lambda_i; \overbrace{\text{CRLB}(\lambda_i)}^{\lambda_i^2/(n_i a)}]$ are independent $(i = 1, 2)$,

$$\frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\sqrt{\widehat{\text{CRLB}}(\lambda_1) + \widehat{\text{CRLB}}(\lambda_2)}} = \frac{\sqrt{a}\left(\frac{1}{\bar{Y}_1} - \frac{1}{\bar{Y}_2}\right)}{\sqrt{\frac{1}{n_1 \bar{Y}_1^2} + \frac{1}{n_2 \bar{Y}_2^2}}},$$

which gives the previous expression after a little more algebra. Clearly the square of this is not the Wilks' statistic even though they are both zero when $\bar{y}_1 = \bar{y}_2$.
(d) The approximation for Wilks' statistic requires that only n is large whereas the Normal approximation requires both n_1 and n_2 large. Thus when one of the samples is large, the other can be small and we can still use Wilks' statistic with $\chi^2_{p-p_0}$.

(e) If a is unknown then (a) gives $\hat{\lambda}_i = \hat{a}/\bar{y}_i$, $(i = 1, 2)$, where $\hat{a}$ solves

$$\frac{\partial l}{\partial a} = 0 \quad \rightarrow \quad n_1 \log \hat{\lambda}_1 + n_2 \log \hat{\lambda}_2 + \sum_{i=1}^{n} \log y_i = \frac{n\Gamma'(a)}{\Gamma(a)}, \quad \text{or}$$

$$n \log a + \sum_{i=1}^{n} \log y_i - n_1 \log \bar{y}_1 - n_2 \log \bar{y}_2 = \frac{n\Gamma'(a)}{\Gamma(a)}.$$

Applying this argument to the likelihood under H_0, $\hat{a}_0$ solves

$$n \log \lambda + \sum_{i=1}^{n} \log y_i = \frac{n\Gamma'(a)}{\Gamma(a)}, \quad \text{or} \quad n \log a + \sum_{i=1}^{n} \log y_i - n \log \bar{y} = \frac{n\Gamma'(a)}{\Gamma(a)}.$$

The Wilks' statistic is now $2[l(\hat{\lambda}_1, \hat{\lambda}_2, \hat{a}; \underline{y}) - l(\hat{\lambda}, \hat{\lambda}, \hat{a}_0; \underline{y})]$ which has a chisquare distribution under H_0 with $3 - 2 = 1$ degrees of freedom.

2. (a) The likelihood is

$$\begin{aligned} L(\beta; \underline{y}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 x_i}} \exp\left\{ -\frac{(y_i - \beta x_i)^2}{2\sigma^2 x_i} \right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} x_i \right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{x_i} \right\}, \end{aligned}$$

and so the log-likelihood is

$$\ell(\beta; \underline{y}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \log(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{x_i}.$$

Thus, solving the equation

$$\frac{d\ell}{d\beta} = \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta x_i) = 0,$$

the maximum likelihood estimate of β is $\hat{\beta} = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} x_i$. The restricted maximum likelihood estimate of β under H_0 is $\hat{\beta}_0 = 1$. Hence, the generalised likelihood ratio is

$$\begin{aligned} \Lambda(\underline{y}) = \frac{L(\hat{\beta}_0; \underline{y})}{L(\hat{\beta}; \underline{y})} &= \frac{\exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(y_i - \hat{\beta}_0 x_i)^2}{x_i} \right\}}{\exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(y_i - \hat{\beta} x_i)^2}{x_i} \right\}} \\ &= \exp\left[-\frac{(\hat{\beta} - \hat{\beta}_0)}{2\sigma^2} \sum_{i=1}^{n} \{2y_i - (\hat{\beta} + \hat{\beta}_0) x_i\} \right] \\ &= \exp\left\{ -\frac{\sum_{i=1}^{n} x_i}{2\sigma^2} \left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} - 1 \right)^2 \right\}. \end{aligned}$$

(b) Since the critical region is $R = \{\underline{y} : \Lambda(\underline{y}) \le a\}$, we reject H_0 if and only if

$$-\frac{\sum_{i=1}^n x_i}{2\sigma^2}\left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} - 1\right)^2 \le \log a \Rightarrow \left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} - 1\right)^2 \ge b,$$

where a and b are constants chosen to give significance level α. Thus, we reject H_0 if and only if $|\sum_{i=1}^n y_i - \sum_{i=1}^n x_i| \ge c$, where c is a constant chosen to give size α.
(c) Since we have $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2 x_i)$ independently for $i = 1, 2, \ldots, n$, it follows that $\sum_{i=1}^n Y_i \sim \mathrm{N}(\beta \sum_{i=1}^n x_i, \sigma^2 \sum_{i=1}^n x_i)$. So, under H_0, $\sum_{i=1}^n Y_i \sim \mathrm{N}(\sum_{i=1}^n x_i, \sigma^2 \sum_{i=1}^n x_i)$,

$$\frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n x_i}{\sqrt{\sigma^2 \sum_{i=1}^n x_i}} \sim \mathrm{N}(0, 1).$$

Thus, we reject H_0 at the 5% level of significance if and only if

$$\left|\sum_{i=1}^n y_i - \sum_{i=1}^n x_i\right| \ge z_{0.025}\sigma\sqrt{\sum_{i=1}^n x_i} = 1.96\sigma\sqrt{\sum_{i=1}^n x_i}.$$

(d) If σ^2 were unknown, the generalised likelihood ratio would become

$$\Lambda(\underline{y}) = \frac{L(\hat{\beta}_0, \hat{\sigma}_0^2; \underline{y})}{L(\hat{\beta}, \hat{\sigma}^2; \underline{y})},$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2 and $\hat{\sigma}_0^2$ is the restricted maximum likelihood estimate of σ^2 under H_0. Since there are $p = 2$ unknown parameters in total and $p_0 = 1$ unknown parameters under H_0, we have $s = p - p_0 = 1$. Therefore, by Wilks' theorem, for large samples, we reject H_0 at the $100\alpha\%$ level of significance if and only if $-2\log\{\Lambda(\underline{y})\} \ge \chi^2_{1,\alpha}$.