STATISTICS Part IB

Example Sheet 3 (of 3) RJS/Lent 2009

Comments and corrections to r.samworth@statslab.cam.ac.uk

1. Consider the simple linear regression model

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

where $\epsilon_1, \ldots, \epsilon_n \stackrel{iid}{\sim} N(0, \sigma^2)$ and $\sum_{i=1}^n x_i = 0$. Derive from first principles explicit expressions for the MLEs $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$. Let A be an $n \times n$ orthogonal matrix where the entries in the first row are all equal to $1/\sqrt{n}$, and where the jth entry in the second row is $x_j/\sqrt{S_{xx}}$. By considering the distribution of $Z = AY$, derive the joint distribution of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$.

2. The relationship between the range in metres, Y, of a howitzer with muzzle velocity v metres per second fired at angle of elevation α degrees is assumed to be

$$Y = \frac{v^2}{g} \sin(2\alpha) + \epsilon,$$

where $g = 9.81$ and where $\epsilon \sim N(0, \sigma^2)$. Estimate v from the following independent observations made on 9 shells, and provide a 95% confidence interval for v.

α (deg)	5	10	15	20	25	30	35	40	45
$\sin 2\alpha$	0.1736	0.3420	0.5	0.6428	0.7660	0.8660	0.9397	0.9848	1
range (m)	4860	9580	14080	18100	21550	24350	26400	27700	28300

3. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. By considering the distribution of the random vector

$$(\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}),$$

where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, show that $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are independent. Hence give an alternative proof to the one from lectures of the fact that $\bar{X}$ and $S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2$ are independent.

4. Consider the linear model $Y = X\beta + \epsilon$, where $\mathbb{E}(\epsilon) = 0$ and $\mathrm{Cov}(\epsilon) = \sigma^2\Sigma$, for some unknown parameter $\sigma > 0$ and known positive definite matrix Σ. Derive the form of the Generalised Least Squares estimator $\tilde{\beta}^{GLS}$, defined by

$$\tilde{\beta}^{GLS} = \mathrm{argmin}_\beta (Y - X\beta)^T \Sigma^{-1} (Y - X\beta).$$

State and prove a version of the Gauss–Markov theorem for $\tilde{\beta}^{GLS}$.

5. Suppose $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are independent, with $X_1, \ldots, X_m \sim N(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_n \sim N(\mu_Y, \sigma^2)$, where σ is unknown. Derive the likelihood ratio test of size α of $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X \neq \mu_Y$, showing in particular that it can be expressed in terms of

$$T = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{S_{XX} + S_{YY}}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}}.$$

6. Download the 'Mobile phone data' at `http://www.statslab.cam.ac.uk/~rjs57/Teaching.html`, and read the description that accompanies the data. Explain why, if each student carries out the task both with and without a mobile phone, we should consider a single sample t-test based on the matched paired differences, whereas if students are assigned only to one group or the other, we should consider a two sample t-test. Carry out both tests at the 5% level.

7. Suppose $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are independent, with $X_1, \ldots, X_m \sim N(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_n \sim N(\mu_Y, \sigma_Y^2)$, and we wish to test $H_0 : \sigma_X^2 = \sigma_Y^2$ against $H_1 : \sigma_X^2 > \sigma_Y^2$ with μ_X, μ_Y unknown. Show that for $S_{XX}/S_{YY} > m/n$, the likelihood ratio is an increasing function of S_{XX}/S_{YY}, and derive a size α test.

8. Download the 'Cars data demo' at `http://www.statslab.cam.ac.uk/~rjs57/Teaching.html` and work through the commands in **R**. Carry out the exercise at the end.

9. Derive from first principles the form of the size α likelihood ratio test of equality of means in a one-way Analysis of Variance model.

10. If $X_1, \ldots, X_n$ are independent with $X_i \sim N(\mu_i, 1)$, we say $X = \sum_{i=1}^n X_i^2$ has a *non-central chi-squared distribution* with n degrees of freedom and non-centrality parameter $\lambda = \sum_{i=1}^n \mu_i^2$ and write $X \sim \chi_n^2(\lambda)$. In this case X has probability density function

$$f(x; n, \lambda) = \sum_{r=0}^{\infty} \frac{e^{-\lambda/2}\lambda^r}{2^r r!} \frac{x^{\frac{n}{2}+r-1} e^{-x/2}}{2^{\frac{n}{2}+r}\Gamma(\frac{n}{2}+r)}, \quad x \in (0, \infty).$$

Show that if $R \sim \mathrm{Poi}(\lambda/2)$ and the conditional distribution of X given R is χ_{n+2R}^2, then $X \sim \chi_n^2(\lambda)$. Hence or otherwise compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

11.* (Extension of Cochran's theorem) Let $Y \sim N_n(\mu, \sigma^2 I)$, and let $A_1, \ldots, A_k$ be $n \times n$ symmetric matrices with $\mathrm{rank}(A_i) = r_i$ and $A_1 + \ldots + A_k = I$. Suppose that $r_1 + \ldots r_k = n$. Write down expressions for A_i^2 and $A_j A_i$ for $j \neq i$. Hence show that we have the orthogonal direct sum decomposition $\mathbb{R}^n = \mathrm{Im}(A_1) \oplus \ldots \oplus \mathrm{Im}(A_k)$, where $\mathrm{Im}(A_i) = \{A_i x : x \in \mathbb{R}^n\}$ denotes the image of A_i. Deduce that there exists an $n \times n$ orthogonal matrix Q such that

$$Q^T A_i Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{r_i} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

for all $i = 1, \ldots, k$, where the ones occur in diagonal entries $\sum_{l=1}^{i-1} r_l + 1, \ldots, \sum_{l=1}^{i} r_l$. Finally, show that $Y^T A_1 Y, \ldots, Y^T A_k Y$ are independent, with $Y^T A_i Y \sim \sigma^2 \chi_{r_i}^2(\frac{1}{\sigma^2}\mu^T A_i \mu)$.

12. Let $Y = X\beta + \epsilon$, where X and β are partitioned as $X = (X_0 \; X_1)$ and $\beta^T = (\beta_0^T, \beta_1^T)$ (where β_0 has p_0 components and β_1 has $p - p_0$ components). Recall that the likelihood ratio statistic for testing $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ is $\|PY - P_0 Y\|^2 / \|Y - PY\|^2$, where $P = X(X^T X)^{-1} X^T$ and $P_0 = X_0 (X_0^T X_0)^{-1} X_0^T$. Determine the joint distribution of $\|PY - P_0 Y\|^2$ and $\|Y - PY\|^2$ under H_1.