STATISTICS   Part IB

Example Sheet 3 (of 3)   RJS/Lent 2010

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

1. Let $X \sim N_n(\mu, \Sigma)$, and let A be an arbitrary $m \times n$ matrix. Prove directly from the definition that AX has an m-variate normal distribution. Show further that $AX \sim N_m(A\mu, A\Sigma A^T)$. Give an alternative proof of this result using moment generating functions.

2. Consider the simple linear regression model

$$Y_i = a + bx_i + \epsilon_i, \quad i = 1, \ldots, n$$

where $\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$ and $\sum_{i=1}^n x_i = 0$. Derive from first principles explicit expressions for the MLEs $\hat{a}$, $\hat{b}$ and $\hat{\sigma}^2$. Show that we can obtain the same expressions if we regard the simple linear regression model as a special case of the general linear model $Y = X\beta + \epsilon$ and specialise the formulae $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{\sigma}^2 = n^{-1} \|Y - X\hat{\beta}\|^2$.

3. 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. Provide a 95% confidence interval for v based on a root which has a t-distribution.

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

4. Consider the one-way analysis of variance (ANOVA) model

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \ldots, I, j = 1, \ldots, n_i,$$

where $(\epsilon_{ij}) \overset{iid}{\sim} N(0, \sigma^2)$. Derive from first principles explicit expressions for the MLEs $\hat{\mu}_1, \ldots, \hat{\mu}_I$ and $\hat{\sigma}^2$. Show that we can obtain the same expressions if we regard the ANOVA model as a special case of the general linear model $Y = X\beta + \epsilon$ and specialise the formulae $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{\sigma}^2 = n^{-1} \|Y - X\hat{\beta}\|^2$.

5. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. 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.

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

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

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

7. 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.

8. Consider again the simple linear regression model in **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 $Z = AY$, where $Y = (Y_1, \ldots, Y_n)^T$, derive the joint distribution of $\hat{a}$, $\hat{b}$ and $\hat{\sigma}^2$.

9. Derive from first principles the form of the size α likelihood ratio test of equality of means in the one-way ANOVA model of **4.**.

10. In the standard linear model $Y = X\beta + \epsilon$ with $\epsilon \sim N_n(0, \sigma^2 I)$ and MLE $\hat{\beta}$, determine the distribution of the quadratic form $(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)$. Hence find a $(1 - \alpha)$-level confidence set for β based on a root which has an F-distribution. What shape is this confidence set?

11. Use **R** to compute a 95% confidence set for the vector of mean chick weights for the different food supplements in the `chickwts` data set (one of the in-built data sets in **R**). (*Hint: Type* `?model.matrix` *to find out how to obtain the design matrix.*)

12. (**Continuation**) Now use **R** to compute 95% confidence intervals for each of the individual mean chick weights. Which intervals exclude the estimate of the overall mean chick weight in the null model which assumes that the mean chick weight does not depend on the food supplement?