STATISTICS Part IB

Example Sheet 3 (of 3) RJS/Lent 2018

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. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Derive the distribution of X_1^2/σ^2. Using moment generating functions or otherwise, deduce that $\sum_{i=1}^n X_i^2 \sim \sigma^2 \chi_n^2$.

3. Let $X \sim N_n(\mu, \Sigma)$, and let X_1 denote the first n_1 components of X. Let μ_1 denote the first n_1 components of μ, and let Σ_{11} denote the upper left $n_1 \times n_1$ block of Σ. Show that $X_1 \sim N_{n_1}(\mu_1, \Sigma_{11})$.

4. Consider the simple linear regression model

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

where $\epsilon_1, \ldots, \epsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$ 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 $\hat{\sigma}^2 = n^{-1} \|Y - X\hat{\beta}\|^2$.

5. 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.

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

6. 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 $(\epsilon_{ij}) \overset{\text{iid}}{\sim} N(0, \sigma^2)$. 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 $\hat{\sigma}^2 = n^{-1}\|Y - X\hat{\beta}\|^2$.

7. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. By considering the distribution of the linearly transformed 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.

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

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

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

9. Consider again the simple linear regression model in **4.**. 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 $\hat{\sigma}^2$.

10. Consider the one-way ANOVA model of **6.**. Letting $\bar{Y}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$ and $\bar{Y} = n^{-1}\sum_{i=1}^I \sum_{j=1}^{n_i} Y_{ij}$ with $n = n_1 + \ldots + n_I$, show from first principles that the size α likelihood ratio test of equality of means rejects H_0 if

$$F \equiv \frac{\frac{1}{I-1}\sum_{i=1}^I n_i(\bar{Y}_{i\cdot} - \bar{Y})^2}{\frac{1}{n-I}\sum_{i=1}^I \sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2} > F_{I-1,n-I}(\alpha),$$

i.e. if 'the ratio of the between groups sum of squares to the within groups sum of squares is large'.

11. 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 $(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)$. (*Hint: Consider* $(X^T X)^{1/2}(\hat{\beta} - \beta)$, *where the square root is a matrix square root.*) 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?

$^+$**12.** Download **R** from `http://cran.r-project.org/`. Use it 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.*) 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?