Mathematical Tripos: Part IB SMP/Lent 2013

Statistics: Example Sheet 3 (of 3)

Comments and corrections to s.pitts@statslab.cam.ac.uk

1. If $X \sim N(0,1)$ and $Y \sim \chi_n^2$ are independent, we say that $T = \frac{X}{\sqrt{Y/n}}$ has a t-distribution with n degrees of freedom and write $T \sim t_n$. Show that the probability density function of T is $f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(n\pi)^{1/2}} \frac{1}{(1+t^2/n)^{(n+1)/2}}, \quad t \in \mathbb{R}$.

2. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where σ^2 is unknown, and suppose we are interested in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Letting $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and $S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2$, show that the likelihood ratio can be expressed as
$$\Lambda_{\mathbf{X}}(H_0, H_1) = \left(1 + \frac{T^2}{n-1}\right)^{n/2},$$
where $T = \frac{n^{1/2}(\bar{X} - \mu_0)}{\{S_{XX}/(n-1)\}^{1/2}}$. Determine the distribution of T under H_0, and hence determine the size α likelihood ratio test.

3. Suppose that $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independent, with $X_i \sim N(\mu_X, \sigma^2)$, $i = 1, \ldots, m$, and $Y_i \sim N(\mu_Y, \sigma^2)$, $i = 1, \ldots, n$, where μ_X, μ_Y and σ^2 are unknown. Write down the distributions of $\bar{X} - \bar{Y}$ and $\frac{S_{XX} + S_{YY}}{\sigma^2}$. Find a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$.

4. Statisticians A and B obtain independent samples $X_1, \ldots, X_{10}$ and $Y_1, \ldots, Y_{17}$ respectively, both from a $N(\mu, \sigma^2)$ distribution with both μ and σ^2 unknown. They estimate (μ, σ^2) by $(\bar{X}, S_{XX}/9)$ and $(\bar{Y}, S_{YY}/16)$ respectively, where, for example, $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$ and $S_{XX} = \sum_{i=1}^{10} (X_i - \bar{X})^2$. Given that $\bar{X} = 5.5$ and $\bar{Y} = 5.8$, which statistician's estimate of σ^2 is more probable to have exceeded the true value by more than 50%? Find this probability (approximately) in each case. [Hint: This is something of a 'trick' question. Why? You may find χ^2 tables helpful.]

5. Suppose that $X_1, \ldots, X_m$ are iid $N(\mu_X, \sigma_X^2)$, and, independently, $Y_1, \ldots, Y_n$ are iid $N(\mu_Y, \sigma_Y^2)$, with μ_X, μ_Y, σ_X^2 and σ_Y^2 unknown. Write down the distributions of S_{XX}/σ_X^2 and S_{YY}/σ_Y^2. Derive a $100(1-\alpha)\%$ confidence interval for σ_X^2/σ_Y^2.

6. Consider the simple linear regression model $Y_i = a + bx_i + \varepsilon_i$, $i = 1, \ldots, n$, where $\varepsilon_1, \ldots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$ 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 $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and specialise the formulae $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$ and $\hat{\sigma}^2 = n^{-1} \|\mathbf{Y} - X\hat{\boldsymbol{\beta}}\|^2$.

7. Consider the model $Y_i = bx_i + \varepsilon_i$, $i = 1, \ldots, n$, where the ε_i are independent with mean zero and variance σ^2 (regression through the origin). Write this in the form $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and find the least squares estimator of b.

   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 $Y = \frac{v^2}{g}\sin(2\alpha) + \varepsilon$, where $g = 9.81$ and where ε has mean zero and variance σ^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

8. Consider the model $Y_i = \mu + \varepsilon_i$, $i = 1, \ldots, n$, where ε_i are iid $N(0, \sigma^2)$ random variables. Write this in matrix form $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and find the MLE $\hat{\boldsymbol{\beta}}$. Find the fitted values, the residuals and the residual sum of squares. Show how applying Theorem 3.7 (in lectures) to this case gives the independence of $\bar{Y}$ and S_{YY} for an iid sample from $N(\mu, \sigma^2)$. Write down an unbiased estimate $\tilde{\sigma}^2$ of σ^2.

9. Consider the one-way analysis of variance (ANOVA) model $Y_{ij} = \mu_i + \varepsilon_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$, where $(\varepsilon_{ij}) \overset{iid}{\sim} N(0, \sigma^2)$. 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 $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and specialise the formulae $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$ and $\hat{\sigma}^2 = n^{-1}\|\mathbf{Y} - X\hat{\boldsymbol{\beta}}\|^2$.

10. Consider the linear model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\mathbf{Y}$ is an $n \times 1$ vector of observations, X is a known $n \times p$ matrix of rank p, $\boldsymbol{\beta}$ is a $p \times 1$ unknown parameter vector and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of uncorrelated random variables with mean zero and variance σ^2 (i.e. we are *not* assuming that the ε_i are normally distributed). Let $\hat{\boldsymbol{\beta}}$ denote the least squares estimate of $\boldsymbol{\beta}$, $\hat{\mathbf{Y}}$ denote the vector of fitted values, and let $\mathbf{R}$ be the vector of residuals. Find $\mathbb{E}(\mathbf{R})$ and $\text{cov}(\mathbf{R})$. Show that $\text{cov}(\mathbf{R}, \hat{\boldsymbol{\beta}}) = 0$ and $\text{cov}(\mathbf{R}, \hat{\mathbf{Y}}) = 0$.

11. For the simple linear regresssion model $Y_i = a + bx_i + \varepsilon_i$, $i = 1, \ldots, n$, where $\sum_i x_i = 0$ and where the ε_i are iid $N(0, \sigma^2)$ random variables, the MLEs $\hat{a}$ and $\hat{b}$ were found in Question 6. Find the distribution of $\hat{\boldsymbol{\beta}} = (\hat{a}, \hat{b})^T$. Find a 95% confidence interval for b and for the mean value of Y when $x = 1$. [Hint: Look at "Applications of the distribution theory" in lectures.]

+12 Consider the one-way ANOVA model of Question 9. Letting $\bar{Y}_{i+} = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$ 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

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

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