Exam Statistics, Tinbergen Institute M.Phil.-course
21 December 2005

---

1. Let $X = (X_1, X_2)^\top$ be a vector of independent random variables that both have a normal $N(0, \sigma^2)$ distribution ($\sigma^2 > 0$). Let $Y = (Y_1, Y_2)^\top$ with $Y = AX$, where A is the matrix

$$A = \begin{pmatrix} a & -1 \\ b & ab \end{pmatrix},$$

for real numbers a and b ($b \neq 0$).

   (a) Compute the covariance matrix of Y.

   (b) What is the distribution of Y?

   (c) Show that Y_1 and Y_2 are independent random variables.

   (d) Show that Y_1^2 and Y_2^2 are independent random variables.

   (e) For certain real constants λ_1 and λ_2 put $U = \lambda_1 Y_1^2 + \lambda_2 Y_2^2$. How do we have to choose λ_1 and λ_2 such that U has a χ_2^2-distribution?

   (f) How to choose the constants λ_1 and λ_2 in the previous part such that U has an exponential distribution with parameter 1?

2. We observe $X_1, \ldots, X_n$, independent random variables with a common exponential distribution depending on a parameter $\theta > 0$ with density

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta},\ x > 0.$$

   (a) What is the probability density function of the random vector $(X_1, \ldots, X_n)$?

   (b) Suppose that we already know that $U = \sum_{k=1}^{n-1} X_k$ has a Gamma distribution with density

$$f_{n-1}(u|\theta) = \frac{u^{n-2}}{\theta^{n-1}(n-2)!} e^{-u/\theta},\ u > 0.$$

Show by computing the convolution integral that $S = \sum_{k=1}^{n} X_k = U + X_n$ has density

$$f_n(s|\theta) = \frac{s^{n-1}}{\theta^n (n-1)!} e^{-s/\theta},\ s > 0.$$

N.B.: Also S thus has a gamma distribution.

(c) Show that S/θ has density

$$f_n(s|1) = \frac{s^{n-1}}{(n-1)!} e^{-s}.$$

(d) Consider the hypotheses $H_0 : \theta = \theta_0$ and $H_A : \theta = \theta_1$, where $\theta_1 > \theta_0$. Show that the Neyman-Pearson test rejects the null hypothesis for "large values" of S, $S > c$ say.

(e) If α is the significance level of the test, show that $c = \theta_0 \gamma_\alpha$, where γ_α satisfies $\int_{\gamma_\alpha}^{\infty} f_n(s|1)\, ds = \alpha$.

(f) The power of this test in θ_A is $\pi(\theta_A) = \mathbb{P}(S > \theta_0 \gamma_\alpha | \theta_A)$. Compute $\lim_{n\to\infty} \pi(\theta_A)$.

(g) Is the Neyman-Pearson test uniformly most powerful for testing H_0 against the alternative $\theta > \theta_0$?

3. Consider the multivariate regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where the design matrix X is of size $n \times p$, and where the elements e_i of the vector $\mathbf{e}$ are independent random variables with $\mathbb{E}\, e_i = 0$ and $\operatorname{Var} e_i = \sigma^2$. The least squares estimator of β is given by $\hat{\beta}_n = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$.

(a) Suppose that one has an additional (row) vector of design variables x_{n+1}. The corresponding response variable Y_{n+1} is then predicted by $\hat{Y}_{n+1|n} = x_{n+1}\hat{\beta}_n$. Let $\varepsilon_{n+1|n} = \hat{Y}_{n+1|n} - Y_{n+1}$ be the prediction error. Why are $\hat{Y}_{n+1|n}$ and Y_{n+1} independent?

(b) Compute the expectation $\mathbb{E}\, \varepsilon_{n+1|n}$ and the variance $\operatorname{Var}\, \varepsilon_{n+1|n}$.

(c) If we also observe Y_{n+1} we can compute a new least squares estimator $\hat{\beta}_{n+1}$ following the usual least squares procedure, but now based on $n+1$ observations. It turns out that the following recursive relationship holds

$$\hat{\beta}_{n+1} = \hat{\beta}_n + \frac{1}{1+d} (\mathbf{X}^\top \mathbf{X})^{-1} x_{n+1}^\top (Y_{n+1} - x_{n+1}\hat{\beta}_n),$$

where $d = x_{n+1}(\mathbf{X}^\top\mathbf{X})^{-1}x_{n+1}^\top$. Using the estimator $\hat{\beta}_{n+1}$, we predict Y_{n+1} by $\hat{Y}_{n+1} = x_{n+1}\hat{\beta}_{n+1}$. Show that

$$\hat{Y}_{n+1} = \frac{1}{1+d}\hat{Y}_{n+1|n} + \frac{d}{1+d}Y_{n+1}.$$

(d) Let ε_{n+1} be the associated prediction error, $\varepsilon_{n+1} = \hat{Y}_{n+1} - Y_{n+1}$. Compute $\mathbb{E}\,\varepsilon_{n+1}$ and $\operatorname{Var}\varepsilon_{n+1}$.

(e) Which of the two predictors $\hat{Y}_{n+1|n}$ and $\hat{Y}_{n+1}$ would you prefer?

(f) Suppose that we also know that the e_i are $N(0,\sigma^2)$ distributed random variables with *unknown* σ^2. Show that $x_{n+1}\hat{\beta}_{n+1}$ has a $N(x_{n+1}\beta, \frac{d\sigma^2}{1+d})$ distribution.

(g) Let $R = \sum_{i=1}^{n+1}(Y_i - x_i\hat{\beta}_{n+1})^2$. It is known that $\frac{R}{\sigma^2}$ has a χ^2_{n+1-p} distribution. Show that

$$T := \frac{x_{n+1}(\hat{\beta}_{n+1} - \beta)}{\sqrt{\frac{R}{n+1-p}\frac{d}{1+d}}}$$

has a t_{n+1-p}-distribution.

(h) Construct a $(1-\alpha)$-confidence interval for $x_{n+1}\beta$ based on $\hat{\beta}_{n+1}$.