Ph.D. Prelim: Exam A

Distribution Theory & Regression Analysis

Jan 2017

1. Given the cumulative distribution function (c.d.f)

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2 + 0.2, & 0 \le x < 0.5, \\ x, & 0.5 \le x < 1, \\ 1, & 1 \le x. \end{cases}$$

   (a) Is $F_X(x)$ a discrete or continuous distribution?

   (b) Write $F_X(x)$ in the form of $cF_1(x) + dF_2(x)$, where F_1 and F_2 are c.d.f's and c and d are known constants. Find c, d, F_1 and F_2.

2. (a) Let $X \ge 0$ be a random variable with finite $E(X^2)$. Prove that

$$E(X^2) = \int_0^\infty 2xP(X > x)dx$$

   (b) Let X and Y be independent random variables having cumulative distribution functions $F(t)$ and $G(t)$ respectively. Suppose that $(1 - G(t)) = (1 - F(t))^\alpha$, for all $t > 0$ and where $\alpha > 0$. Prove that $Z = \min(X, Y)$ and $\delta = I(X \le Y)$ are independent random variables.

3. Let f be a density with support on $(0, \infty)$. Let $g(x, y) = f(x + y)/(x + y)$ for $x > 0, y > 0$ and 0 otherwise. Prove that g is a probability density function on R^2 and find its covariance matrix.

4. (a) Prove the following Theorem: If $y \sim N(0, \sigma^2 I)$ and M is a symmetric idempotent matrix of rank m, then

$$\frac{y'My}{\sigma^2} \sim \chi^2(tr(M)).$$

(b) Use above Theorem to show that in a normal error simple linear regression $Y_i = \beta_0 + X_i\beta_1 + \epsilon_i$ for $i \in \{1, 2, \ldots, n\}$ with $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)' \sim N(0, \sigma^2 I)$, the residual sum of squares

$$RSS/\sigma^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2/\sigma^2 \sim \chi^2_{n-2}.$$

5. Consider the general linear model $Y = X\beta + \epsilon$ with $E(\epsilon) = 0$ and $\text{cov}(\epsilon) = \sigma^2 I$, where X is $n \times p$ of rank p. Let $r_i = Y_i - \hat{Y}_i = Y_i - x_i'\hat{\beta}$ be the ith residual, where $\hat{\beta}$ is the least squares estimator of β.

(a) Show that $\sum_{i=1}^{n} \hat{Y}_i(Y_i - \hat{Y}_i) = 0$.

(b) Now consider the linear model with an intercept term included explicitly. That is, the first column of X is all $1'$s so $Y_i = \beta_0 + \sum_{j=1}^{p-1} x^{\star}_{ij}\beta_j + \epsilon_i$, $i = 1, 2, \ldots, n$ with the same assumptions on the ϵ's as above. Show that in this case $\sum_{i=1}^{n} r_i = 0$ and

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

6. Consider the general linear model $Y = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2 I)$, where X is $n \times p$ of rank p. Let $\theta = H\beta$ where H is $q \times p$ of rank q and

$$Q = (H\hat{\beta} - \theta)'(H(X'X)^{-1}H')^{-1}(H\hat{\beta} - \theta)/(q \times \hat{\sigma}^2)$$

where $\hat{\beta}$ is the least squares estimator of β. What is the distribution of Q? State what general result you are applying and how it applies here. (You can use the fact that $(n-p)\hat{\sigma}^2/\sigma^2$ is distributed chi-square with $n - p$ degrees of freedom and $\hat{\sigma}^2$ is independent of $\hat{\beta}$).