///

Ph.D. Prelim : Exam. A
Distribution Theory & Regression Analysis

May 31, 2011

1. Random variables X and Y are independent, both being exponentially distributed with means $EX = \lambda^{-1}$ and $EY = \mu^{-1}$. Instead of observing both X and Y, we can only observe the pair (Z, W), where

$$Z := \min(X, Y), \qquad \text{and} \qquad W := \begin{cases} 1, & \text{if } Z = X \\ 0, & \text{if } Z = Y \end{cases}$$

(a) Find the joint distribution of (Z, W), by computing $P(Z > z, W = i); i = 0, 1$.

(b) Prove that Z and W are statistically independent, by considering the conditional distributions of Z given W.

2. The *Cauchy family* $\{\mathcal{C}(\alpha, \beta) : -\infty < \alpha < \infty, \quad \beta > 0\}$ of absolutely continuous distributions with location and scale parameters α and β respectively, are defined by their density function

$$f(x; \alpha, \beta) := \frac{\beta}{\pi\{1 + \beta^2(x - \alpha)^2\}}; \quad -\infty < x < \infty, \quad -\infty < \alpha < \infty, \quad \beta > 0.$$

(a) Suppose R is *any positive random variable*, and Θ is a continuous random variable that is Uniformly distributed on $(0, \pi)$, and is independent of R. Then (R, Θ) can be interpreted as the polar coordinates of a point selected in the upper half plane. If (X, Y) denotes the corresponding Cartesian coordinates, defined via the usual transformation

$$X := R\cos\Theta, \qquad Y := R\sin\Theta;$$

show that the $\frac{X}{Y}$ has the standard Cauchy distribution $\mathcal{C}(0, 1)$.

Does the unspecified distribution of R need to be continuous with a density?

(b) Let X, Y be i.i.d. with the standard Normal distribution $N(0, 1)$. Using the transformation

$$U := X, \qquad W = \frac{X}{X + Y};$$

find the density of W, to show that it belongs to the *Cauchy family.*

3. $X_1, X_2, \cdots$ are i.i.d., distributed Uniformly on $(0, 1)$, and N is an integer valued random variable independent of the sequence $\{X_n\, ; n \geq 1\}$, such that

$$P(N = n) = \frac{c}{n!}, \quad n \geq 1$$

where c is a suitable constant.

(a) Find the p.d.f. of $Z := \min(X_1, \cdots, X_N)$.

(b) Compute EZ by first conditioning on N. Then compute EZ^2, to show that the coefficient of variation (η) of Z satisfies,

$$\eta^2 := \frac{\text{var } Z}{EZ^2} = \left(\frac{e}{e-2}\right).$$

4. This question has three parts.

(a) Prove the Bonferroni inequality for two events A and B, given by $P(A \cap B) \geq 1 - P(A) - P(B)$.

(b) Extend the Bonferroni inequality you proved in part **4(a)** to a finite number, say p, of events.

(c) For first-order normal linear models (usual multiple regression) with p parameters including the Y intercept, use part **4(b)** and mathematical arguments to obtain simultaneous confidence intervals for all the parameters with family confidence coefficient $1 - \alpha$.

5. This question has two independent parts.

(a) For the simple linear regression model show that the sample mean $\bar{Y}$ and the slope estimate are independent.

(b) For normal linear models, derive the weighted least squares estimator of $\boldsymbol{\beta}$.

6. This question has two independent parts.

(a) Explain how the odds ratio in logistic regression is connected with the odds ratio in the following 2×2 contingency table.

Response	$x_1 = 0$, Active Drug	$x_1 = 1$, Placebo
$y = 0$, not infected	n_{00}	n_{01}
$y = 1$, infected	n_{10}	n_{11}

(b) A multiple regression model is fitted with response y, and predictors x_1 and x_2. The ANOVA Table and some sums are given below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
Regression	651,996.1	2	325,983	267.2	4.74E-10
Residual	13,421,1	11	1220.1		
Total	665,387.2	13			

$$\sum_{i=1}^{n} x_{i1} = 3,125, \quad \sum_{i=1}^{n} x_{i2} = 60.72, \quad \sum_{i=1}^{n} x_{i1}^2 = 702,205, \quad \sum_{i=1}^{n} x_{i2} = 264.258,$$

$$\sum_{i=1}^{n} x_{i1}x_{i2} = 13,563.5, \quad \sum_{i=1}^{n} x_{i1}y_i = 3,952,040, \quad \sum_{i=1}^{n} x_{i2}y_i = 75,738.3.$$

Suppose it is of interest to investigate the contribution of x_2 to the original first-order model. Write down the hypotheses you want to test. Use the extra sum of squares approach to show clearly whether or not x_2 contributes significantly to the model.