Preliminary Exam in Distribution Theory and Regression Analysis: January 2011

1. Continuous random variables (X, Y) have joint density,

$$f(x, y) = \begin{cases} e^{-y}, & \text{if } 0 < x < y < \infty \\ 0, & \text{elsewhere} \end{cases}$$

(i) Find the marginal densities of X and Y.

(ii) Find the best predictor of Y given X, under squared error *and* absolute error loss functions.

2. Suppose that the number of hurricanes during the hurricane season in a year can be modeled by a Poisson distribution; except that the average number (μ) of such hurricanes is influenced by environmental and global climatological factors (such as the El Ninõ effect) which in turn can be described via a Gamma (α, β) distribution with shape parameter α and scale parameter β. In other words; conditional on μ, the number of hurricanes $X \sim$ Poisson(μ), and our beliefs about μ is modeled by a Gamma (α, β) prior.

(i) Find the unconditional distribution of X.

(ii) What does this distrinution reduce to in the special case, when $\alpha = r$, a positive integer and $\beta = \frac{(1-p)}{p}$; $\quad 0 < p < 1$?

3. If the nonnegative function g, mapping the half-line $[0, \infty)$ into itself, satisfies $\int_0^\infty g(x)\, dx = 1$; show that the function

$$f(x, y) := \frac{2g(\sqrt{x^2 + y^2})}{\pi \sqrt{x^2 + y^2}}, \qquad 0 < x < \infty, \;\; 0 < y < \infty$$

and zero elsewhere, defines a joint density function of a random vector (X, Y).

4. Suppose X is a continuous random variable, with cumulative distribution function (*cdf*) $F(x)$ and a probabilty density function (*pdf*) $f(x)$ which is symmetric about $x = c$, for some real c.

(i) Show that if X has a finite mean $\mu := EX$, then we must have $c = \mu$.

(ii) Assuming a finite mean μ; prove that the distribution of X (equivalently its *pdf* $f(x)$) is symmetric if and only if the function

$$h(x) := x - G^{-1}F(x), \qquad x \in \{t \;:\; 0 < F(t) < 1\}$$

is constant; where G denotes the *cdf* of the r.v. $Y := -X$, and G^{-1} is its inverse function.

5. Consider the following multiple regression model,

$$Y_i = \beta_0 + \beta_i X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \qquad i = 1, 2,]cdots, n$$

which, using the matrix notation, can be expressed as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ (an $n \times 1$ response vector), $\boldsymbol{\beta}=(\beta_0, \ldots, \beta_{p-1})'$ (a $p \times 1$ parameter vector), $\boldsymbol{X}$ is the $n \times p$ design matrix whose ith row is $(1, X_{i1}, \ldots, X_{i,p-1})$, and $\boldsymbol{\epsilon}=(\epsilon_1, \ldots, \epsilon_n)'$, whose components are independent with ϵ_i distributed as $N(0, \sigma_i^2)$; $i = 1, \ldots, n$ (*Note.* If $\boldsymbol{X}$ is a matrix, then $\boldsymbol{X'}$ is the transpose of $\boldsymbol{X}$.)

With $w_i := \frac{1}{\sigma_i^2}$; *the weighted least squares estimator* of $\boldsymbol{\beta}$ is defined as the $p \times 1$ vector $\boldsymbol{b_w}$ which minimizes

$$Q_w = \sum_{i=1}^{n} w_i (Y_i - \beta_0 + \beta_i X_{i1} + \cdots + \beta_{p-1} X_{i,p-1})^2 .$$

Prove that,

(i) the weighted least squares estimator b_w of the parameter vector β is,

$$\boldsymbol{b_w} = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY},$$

where $\boldsymbol{W}$ is an $n \times n$ diagonal matrix whose ith element is w_i

(ii) the variance-covariance matrix of $\boldsymbol{b_w}$ is :

$$\boldsymbol{\Sigma_{b_w}} = (\boldsymbol{X'WX})^{-1}.$$

6. Consider the following simple linear regression model,

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, \ldots, n$$

where β_0 and β_1 are parameters, X_i are known constants, and ϵ_i are i.i.d. $N(0, \sigma^2)$. Answer the follwing questions, using the facts that the least squares estimators b_1 and b_0 for β_1 and β_0 *respectively*, are

$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \\ b_0 &= \bar{Y} - b_1 \bar{X}; \end{aligned}$$

where $\bar{X}$ and $\bar{Y}$ are the sample means of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively.

(i) What is the standard (plug-in) point estimator $(\hat{(Y)}_h)$ of the mean response $E(Y_h)$ when $X = X_h$?

(ii) Show that the variance of this point estimator, is given by

$$\sigma^2(\hat{Y}_h) = \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right).$$

[*You can use the following facts* without proof: [i] b_1 and $\bar{Y}$ are uncorrelated; and [ii] variance of the estimator of b_1 is var $(b_1) = \sigma^2 / \sum_{i=1}^n (X_i - \bar{X})^2$.]

(iii) Use the preceeding results to construct a 95% confidence interval for the mean response $E(Y_h)$ when $X = X_h$.