MATH 8000

Exam 2

Problem 1. Survival analysis: The Gamma-frailty PH model is commonly used to analyze correlated survival data. This problem will be aimed at investigating an interesting characteristic of this model. Let the random variables T_1 and T_2 denote two, correlated, failure times of interest; e.g., time until the onset of two different types of cancer. Under the Gamma-frailty PH model the conditional cumulative distribution function (CDF) for T_j is given by

$$F_j(t|\mathbf{X},\eta) = 1 - \exp\{-\Lambda_{0j}(t)\exp(\mathbf{X}'\boldsymbol{\beta}_j)\eta\}, \quad \text{for } j = 1, 2,$$

where $\Lambda_0(t)$ is an unknown differentiable function, $\mathbf{X} = (X_1, ..., X_p)'$ is a p-dimensional vector or predictor variables (e.g., age, ethnicity, gender, etc.), $\boldsymbol{\beta}_j$ is the corresponding vector of regression parameters, and η is a frailty term that is assumed to be distributed as Gamma$(\nu, 1/\nu)$. It is assumed that conditional on the frailty term that the failure terms are independent; i.e., $T_1|\eta \perp T_2|\eta$, where $\perp$ denotes statistical independence.

(a) State conditions on $\Lambda_{0j}(t)$ that ensure that $F_j(t|\mathbf{X},\eta)$ is a proper CDF.

(b) Derive the marginal survival functions for T_1 and T_2. Note, in survival analysis the survival function for T_j is defined to be

$$S_j(t|\mathbf{X}) = \text{pr}(T_j > t|\mathbf{X}).$$

(c) Derive the joint survival function of T_1 and T_2. Note, the joint survival function is defined as

$$S(t_1, t_2|\mathbf{X}) = \text{pr}(T_1 > t_1, T_2 > t_2|\mathbf{X}).$$

(d) Calculating the correlation between T_1 and T_2 in this context poses many issues, especially since the form of $\Lambda_{0j}(t)$ is unknown. A surrogate can be proposed in terms of Kendall's τ, which is given by

$$\tau = E[sign\{(T_{i1} - T_{j1})(T_{i2} - T_{j2})\}]$$

where (T_{i1}, T_{i2}) and (T_{j1}, T_{j2}) are independent and identically distributed copies of (T_1, T_2) and $sign(\cdot)$ is the usual sign function; i.e., it takes values -1, 0, and 1 when the argument is negative, zero, and positive, respectively. Show that for the Gamma-frailty PH model

$$\tau = (1 + 2\nu)^{-1}.$$

Note, if done properly you may assume an arbitrary joint distribution for the predictor variables.

Problem 2. Gamma distribution:

(a) Suppose that $X \sim \text{Gamma}(\alpha, \beta)$. Show that

$$Y = \frac{2X}{\beta} \sim \chi^2_{2\alpha}.$$

(b) Suppose $X_1 \sim \text{Gamma}(\alpha_1, \beta_1)$, $X_2 \sim \text{Gamma}(\alpha_2, \beta_2)$, and that X_1 and X_2 are independent. Find a function of X_1 and X_2 that has an F-distribution, be sure to identify the degrees of freedom.

(c) Suppose that $X_i \sim \text{Exponential}(\beta_1)$, for $i = 1, ..., n$, and $Y_j \sim \text{Exponential}(\beta_2)$, for $j = 1, ..., m$. Further, assume that $X_1, ..., X_n, Y_1, ..., Y_m$ are mutually independent. Under the assumption that $\beta_1 = \beta_2$, find a function of $\overline{X}$ and $\overline{Y}$ that has an F-distribution, be sure to identify the degrees of freedom. Note, $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\overline{Y} = m^{-1}\sum_{j=1}^{m} Y_j$.

Problem 3. A Bayesian approach: Consider the situation in which $X_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$, for $i = 1, ..., n$, and further that the X_i are mutually independent. Note, in this model we are treating the parameters μ and σ^2 as random variables. Moreover we assume due to "conjugacy" (more on this later) that $\mu|\sigma^2 \sim N(\mu_0, \sigma^2/n_0)$ and $\sigma^2 \sim InverseGamma(\nu_0/2, \nu_0\sigma_0^2/2)$, in the Bayesian paradigm we refer to these distributions as "priors." Note, if $Z \sim InverseGamma(a, b)$ then the probability density function of Z is given by

$$f_Z(z) = \frac{b^a}{\Gamma(a)} z^{-a-1} e^{-\frac{b}{z}}.$$

(a) Calculate the $E(X_1)$ and $V(X_1)$.

(b) Find the distribution of $\mu|\mathbf{X}, \sigma^2$, where $\mathbf{X} = (X_1, ..., X_n)'$; i.e., if we view $X_1, ..., X_n$ as observed data we are finding the updated prior distribution of μ, which is referred to as the "posterior" distribution of μ. If done correctly, one will notice that the posterior distribution of μ is in the same family as its prior, this is why we say that the proposed prior distribution is conjugate.

(c) Find the distribution of $\sigma^2|\mathbf{X}$. Using this distribution provide your "best" guess at what value of σ^2 was used to generate $X_1, ..., X_n$, you should justify how you obtained your guess.

(d) Combining the above results, find the distribution of $\mu|\mathbf{X}$. Using this distribution provide your "best" guess at what value of μ was used to generate $X_1, ..., X_n$, you should justify how you obtained your guess.

Problem 4. Consider the situation in which $X \sim Bernoulli(p_1)$ and $Y \sim Bernoulli(p_2)$. In this problem we will investigate the bounds on the correlation that can exist between these two random variables.

(a) Note the correlation (ρ) between these two random variables does not necessarily exist between [-1,1]. The restriction on ρ is required to ensure that the joint probability mass function of X and Y are non-negative for all outcomes. Derive the upper and lower bounds for ρ.

(b) Under what special case can ρ be exactly 1 and -1.

Problem 5. Let $X_1, ..., X_n$ be mutually independent random variables all arising from a common distribution. If $\text{pr}\left(\sum_{i=1}^n X_i = 0\right) = 0$, find the following expectation

$$E\left(\frac{X_1}{\sum_{i=1}^n X_i}\right).$$

Problem 6. Let X be any non-negative integer-valued random variable with positive expectation. Prove the following inequality,

$$\frac{E(X)^2}{E(X^2)} \leq \text{pr}(X \neq 0) \leq E(X).$$

Problem 7. Consider the random vector $\mathbf{X} = (X_1, ..., X_n)'$ which is distributed $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; i.e., $\mathbf{X} = (X_1, ..., X_n)'$ follows a Multivariate Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Note, $\boldsymbol{\mu} = (E(X_1), ..., E(X_n))'$ and the ijth entry in $\boldsymbol{\Sigma}$ is given by $\boldsymbol{\Sigma}_{ij} = \text{Cov}(X_i, X_j)$. The joint PDF of $\mathbf{X}$ can be expressed as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}.$$

The MGF of $\mathbf{X}$ can be expressed as

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathbf{t}'\boldsymbol{\mu} + \frac{\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}{2}\right\}$$

where $\mathbf{t} = (t_1, ..., t_n)' \in \mathbb{R}^n$.

(a) Derive the MGF of $\mathbf{X}$. Hint: You will need to also argue that $\boldsymbol{\Sigma}$ is symmetric.

(b) Find the distribution of $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is a $q \times p$ matrix and $\mathbf{b}$ is a $q \times 1$ vector.

(c) Consider $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1 = (X_1, ..., X_p)$ and $\mathbf{X}_2 = (X_{p+1}, ..., X_n)$. Find the marginal distribution of $\mathbf{X}_1$ and the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1$.

(d) Let $Z_1, ..., Z_n$ be mutually independent random variables each following a standard normal distribution; i.e., $Z_i \sim N(0, 1)$, for $i = 1, ..., n$. Using the above results find the joint distribution of the following random variables

$$Y_1 = \sum_{i=1}^{n} a_i Z_i \qquad Y_2 = \sum_{i=1}^{n} b_i Z_i$$

where $\sum_{i=1}^{n} a_i b_i = 0$. State, and justify your assertion, whether Y_1 and Y_2 are independent.