WPI Mathematical Sciences Ph.D. General Comprehensive Exam

MA 541 Probability and Mathematical Statistics-II

August, 2017

Note: Please make sure to write down your thinking process in bullet points even if you cannot solve the problems perfectly.

1. Let $X = (X_1, \ldots, X_n)^T$ be a random vector, and $\gamma = (\gamma_1, \ldots, \gamma_n)^T$ be a vector of coefficients. Consider the following model
$$X = Z\gamma + \eta,$$
where Z is a $n \times n$ identity matrix, and $\eta = (\eta_1, \ldots, \eta_n)^T$ is a vector of standard normal random variables (i.e., i.i.d. N(0, 1)).

   (a) Consider the following ridge regression objective function
$$\sum_{i=1}^{n} \left(x_i - \sum_{i=1}^{n} z_{ij}\gamma_j \right)^2 + \sum_{i=1}^{n} \gamma_i^2.$$
   Calculate the ridge regression estimator of the parameter vector γ.

   (b) Calculate the bias and variance of the ridge regression estimator.

   (c) Compare the bias and variance that you obtained from the ridge regression with the standard least square estimator and provide an explanation.

2. Let $X_1, \ldots, X_n$ be a random sample from the probability density function $f(x|\theta) = \theta/x^2, \theta < x < \infty, 0 < \theta < \infty$.

   (a) Find a minimal sufficient statistic for θ.

   (b) Prove that the minimal sufficient statistic for θ is complete.

   (c) Obtain the UMVUE of θ.

3. Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$, where σ^2 is unknown. Derive (with sufficient details) a 90% confidence interval for μ based on the likelihood ratio test (LRT).

4. In a study we aim to estimate the probability that no patients at all would arrive in an emergency room between 10pm and 11pm. To do so, we collect the data regarding the number of patients arriving during this period of time: $X_1, ..., X_n$. We also know it is reasonable to assume that X_i can be considered as following a Poisson(λ) distribution. There could be many different ways to do estimation.

   (a) Derive the UMVUE (uniform minimum variance unbiased estimator) for the probability that no patients arrive in this period of time.

   (b) Derive the MLE (maximum likelihood estimator) for the probability that no patients arrive in this period of time.

(c) Ideally, estimator with a smaller variation should be better. Derive the asymptotic variances for the above two estimators.

5. Let $X_1, \ldots, X_n, X_{n+1} \mid \theta \overset{iid}{\sim} f(x \mid \theta)$, a density function. Suppose $X_1, \ldots, X_n$ are observed but X_{n+1} is missing. Let $X_{(r)}$ and $X_{(r+k)}$ denote order statistics of $X_1, \ldots, X_n$. Find $P\{X_{(r)} < X_{n+1} < X_{(r+k)}\}$. How can you use this result to obtain an approximate 95% prediction ('confidence') interval for X_{n+1}?

6. In an experiment, 20 people tossed a coin (probability of heads .25), each once. If the coin comes heads, A is answered ('yes' or 'no') and if the coin comes tails, B (a different question) is answered ('yes' or 'no'); 2 people answered 'yes'. Independently, in another experiment, 30 people are asked the same questions but a different coin (probability of heads is .75) is tossed; 15 people said 'yes'. Let π_1 and π_2 denote the probabilities of 'yes' on A and B respectively. Find the maximum likelihood estimates (MLEs) of π_1 and π_2. Are you surprised? explain.