Exam Statistics, Probability and Calculus

October 29, 2010, 10:00-13:00 h.

Use of Rice's book, class notes and pocket calculator is allowed.

1. A subject's genotype consists of two alleles and can be AA, Aa or aa. As you know, a genotype can be viewed as a pair of independent Bernoulli variables. In population I the A allele has population frequency 0.1. In population II the frequency of the A allele is 0.2.

   (a) Find the genotype frequencies in population I.

   (b) Find the genotype frequencies in population II.

   With probability 1/2 we select a subject from population I and otherwise we select a subject from population II. Let the random variable X be the number of A alleles of this person. Also, define the random variable Z as follows: $Z = 1$ if the subject we selected is from population I, and $Z = 0$ otherwise.

   (c) Find the expectation of X given $Z = 1$.

   (d) Find the variance of X given $Z = 1$.

   (e) Find the covariance between X and Z.

   Suppose that we selected a subject from population I. Define a random variable $Y = 3X + E$, where E has the standard normal distribution and X and E are independent.

   (f) Find the expectation of Y.

   (g) Find the variance of Y given $X = 1$.

   (h) Find the variance of Y.

2. Suppose X and Y have the joint density function

$$f(x, y) = k(x - y), \qquad 0 \leq y \leq x \leq 1.$$

   (a) Determine k such that $f(x, y)$ is a proper probability density function.

   (b) Determine the marginal density of X.

   (c) Determine the conditional density of Y given X.

Please turn over.

3. Suppose that X is a discrete random variable with

$$P(X=1)=\frac{2}{3}-\theta, \quad P(X=2)=\theta \text{ and } P(X=3)=\frac{1}{3}$$

Five independent observations of X are made: $X_1=3,\ X_2=3,\ X_3=1,\ X_4=2$ and $X_5=1$.

(a) Find the method of moments estimate of θ.

(b) What is the likelihood function of the data?

(c) Find the maximum likelihood estimator (MLE).

(d) Determine the approximate variance of the MLE.

4. A pharmaceutical company claims that 90% of all patients respond to their new drug. To test this claim, we give the drug to a group of 20 randomly selected patients. Let X denote the number of responders among these 20 patients. The null hypothesis is rejected if $X \leq 14$.

(a) Determine the level of significance (probability of a type I error) of this test.

(b) What is the power of the test against the alternative hypothesis that only 60% of all patients respond.

(c) If we reject the null hypothesis if $X \leq 15$, how would the level of significance change? How would the power change?

(d) Let p denote the (unknown) probability of being a responder. Suppose we want to test

$$H_0: p \geq 0.9 \text{ versus } H_A: p < 0.9$$

and we reject the null hypothesis if $X \leq 14$. Sketch the graph of the power as a function of the parameter p.

5. Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with parameter λ.

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}, \qquad x=0,1,2,\ldots$$

(a) Determine the likelihood ratio test for

$$H_0: \lambda=\lambda_0 \text{ versus } H_A: \lambda=\lambda_1,$$

where $\lambda_1 > \lambda_0$.

(b) Determine the generalized likelihood ratio test for

$$H_0: \lambda=\lambda_0 \text{ versus } H_A: \lambda>\lambda_0,$$

(c) Is the test you found uniformly most powerful (UMP)?