Exam Statistics, Probability and Calculus

October 30, 2009, 10:00-13:00 h.

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

If you perform any computations with `R`, please provide the commands you used.

1. Country A has ten million citizens and 10 per 1500 of these are illiterate. Country B has twenty million citizens and 75 per 1500 of these are illiterate.

   (a) For country A, compute the probability that of 500 randomly selected citizens at most 4 are illiterate. Suppose the selection is done with replacement.

   (b) Use both the Poisson and the normal distribution to approximate the probability of part (a).

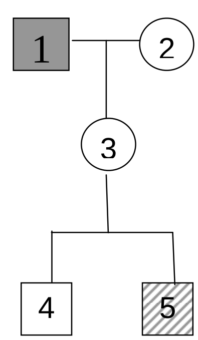
   (c) Suppose you meet a random person from country A or B who is illiterate. What is the probability that he or she is from country A?

2. Consider genotypes AA, AB and BB with frequencies 0, 0.001 and 0.999 respectively. Persons carrying genotype AB have probability 0.8 to have a disease D. Persons with genotype BB will not have disease D. Define the random variable X to be 1 if a person is AB and 0 otherwise and the random variable Y to be 1 if a person has disease D and 0 otherwise.

   (a) Find the prevalence of disease D.

   (b) Find the variance of Y conditional on $X = 1$.

   (c) Find the unconditional variance of Y.



   (d) In a particular family (see picture), person 1 has disease D. Person 3 is healthy and wants to know the probability for her newborn son (5) to have the disease. This son's older brother (4) is healthy. Find the probability that person 5 has the disease.

(e) Denote by X_i the number of A alleles of person i. Let X_f denote the number of A alleles of the father of persons 4 and 5. Suppose that their mother (3) has genotype AB ($X_3 = 1$) and the father has genotype BB ($X_f = 0$). Conditional on this information, the joint distribution of X_4 and X_5 is given by the following table.

	$X_4 = 0$	$X_4 = 1$
$X_5 = 0$	1/4	1/4
$X_5 = 1$	1/4	1/4

From the table, find the correlation of X_4 and X_5 conditional on $X_3 = 1$ and $X_f = 0$.

(f) Find the unconditional correlation of X_4 and X_5.

3. Consider a sample $X_1, X_2, \ldots, X_n$ from a probability distribution with density function

$$f(x;\theta) = \theta^2 x e^{-\theta x}, \quad x \geq 0, \quad \theta > 0.$$

(a) Find the method of moments estimator of θ. You can use integration by parts to compute the moment(s) of this distribution.

(b) Find the maximum likelihood estimator of θ.

(c) Find the asymptotic variance of the maximum likelihood estimator.

4. Consider a random sample of size n from an exponential distribution with density

$$f(x;\lambda) = \lambda e^{-\lambda x}, \qquad x \geq 0, \quad \lambda > 0.$$

Denote the mean by $\theta = 1/\lambda$. We want to test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1,$$

where $\theta_1 < \theta_0$.

(a) Show that the most powerful test rejects for small values of $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

(b) Now suppose we want to test

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0,$$

Is the test you found at part (a) uniformly most powerful (UMP)?

5. Suppose we have a single observation X from the uniform distribution on the interval $[0, \theta]$. We want to test

$$H_0 : \theta = 2 \quad \text{versus} \quad H_1 : \theta > 2.$$

(a) We decide to reject the null hypothesis if $X > 1.9$. What is the level of significance of this test?

(b) What is the power of the test against the alternative $\theta = 2.3$?

(c) Suppose we observe $X = 1.77$. What is the p value of this observation?

(d) Construct a 95% confidence interval for θ on the basis of the observation $X = 1.77$.