MATHEMATICAL TRIPOS: PART IB **(1995)**

Statistics – Example Sheet 2 **AY**

[*Second sheet of 3*]

1. A positive random variable X has density function

$$f(x|\theta) = \frac{\theta}{(\theta + x)^2}, \quad x > 0$$

where $\theta > 0$ is an unknown parameter. Find the best test of size 0.05 of $H_0 : \theta = 1$ vs $H_1 : \theta = 2$, and calculate the Type II error probability for this test.

2. Let $X_1, X_2, \ldots, X_n$ be IID random variables, each with the Poisson distribution of parameter θ (and therefore of mean θ and variance θ). Find the best size α test of $H_0 : \theta = 1$ against $H_1 : \theta = 1.21$. Use the Central Limit Theorem to show that the smallest value of n required to make $\alpha = 0.05$ and $\beta \leq 0.1$ (where α and β are the Type I and Type II error probabilities) is somewhere near 213.

3. Find the best size α test of $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ ($> \theta_0$) and write down an expression for the power function when we have IID random variables $X_1, \ldots, X_n$ whose densities are exponential, with parameter θ.

Use χ^2 tables to find the least sample size which will allow us to test $H_0 : \theta = 1$ against $H_1 : \theta = 3$ with $\alpha = 5\%$ and $\beta \leq 10\%$. Describe the appropriate critical region numerically. [*Hint*: Use Question 3, Example Sheet 1.]

4.* Let $X_1, X_2, \ldots, X_n$ be a sample of size n from the distribution with probability density function

$$\tfrac{1}{2}\lambda e^{-\lambda|x|}, \quad -\infty < x < \infty, \quad \lambda > 0,$$

where λ is unknown. It is desired to test the hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$ where $\lambda_1 > \lambda_0$. Find the test that minimizes the sum of the probabilities of the two types of error. Denote the size of this test by α. Is this the most powerful test of size α for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_2$, for each $\lambda_2 > \lambda_0$?

5. Let $X_1, \ldots, X_n$ be a random sample from a distribution with the density function

$$f(x|\theta_1) = \theta_1 e^{-\theta_1 x}, \quad x > 0;$$

and let $Y_1, \ldots, Y_n$ be a random sample from the density

$$f(y|\theta_2) = \theta_2 e^{-\theta_2 y}, \quad y > 0.$$

Assume that the samples are independent.

Find the form of the likelihood ratio test for testing $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$. Show that the test can be expressed in terms of the statistic

$$T = \Sigma X_i / (\Sigma X_i + \Sigma Y_j).$$

By showing that the distribution of T does not depend on $\theta = \theta_1 = \theta_2$ when H_0 is true, construct a test of exact size α of H_0 against H_1 based on T.

6. (i) It is known that an observation made on a certain system will yield a result falling into one of k categories, $C_1, C_2, \ldots, C_k$. To each value of the real parameter θ in a given interval corresponds a probability distribution $p_i(\theta)$ ($i = 1, 2, \ldots, k$). The null hypothesis is made that, for some unknown value of θ, the probability of a result falling into category C_i is $p_i(\theta)$. A total of n observations is made, of which n_i fall into category C_i ($i = 1, 2, \ldots, k$). Show carefully how to use a χ^2 distribution in testing the above hypothesis, and describe how you would carry out the test, using a statistic of the form

$$\sum_{i=1}^{k} (n_i - e_i)^2 / e_i,$$

where e_i is the expected number of observations in category C_i under (an appropriate case of) the null hypothesis.

(ii) A scientist gets observations in three categories, across which he suspects a linear trend of probability, in which case

$$p_1(\theta) = \ - \theta, \quad p_2(\theta) = \ , \quad p_3(\theta) = \ + \theta$$

for some value of θ such that $- < \theta <$. The numbers observed are $n_1 = 12$, $n_2 = 24$, $n_3 = 24$. Test the hypothesis of linearity.

7. A machine in a factory produces plastic articles in bunches of three articles at a time. The process is rather unreliable, and quite a few defective articles are observed. In an experimental run of the machine, 512 bunches were produced. Of these, the numbers of bunches with $i = 0, 1, 2, 3$ defective articles were 213 ($i = 0$), 228 ($i = 1$), 57 ($i = 2$), and 14 ($i = 3$). Test the hypothesis that each article has a constant (but unknown) probability θ of being defective, independently of all other articles.

8. Show that in a 2×2 contingency table, with observed frequencies Y_{ij} ($i, j = 1, 2$), the usual expression for the statistic X^2 may be put in the form

$$Y_{++}(Y_{11}Y_{22} - Y_{12}Y_{21})^2 / (Y_{1+}Y_{2+}Y_{+1}Y_{+2})$$

where $+$ denotes summation over both values of the corresponding subscript.

9. From each of six batches of seed a random sample of 100 seeds was selected for sowing. The numbers of seeds that failed to germinate in the six samples of 100 seeds were

$$12, 20, 9, 17, 24, 16.$$

Test the hypothesis that the proportion of non-germinating seeds was the same for all batches.