THE UNIVERSITY OF HONG KONG

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT1302/STAT2602 PROBABILITY AND STATISTICS II

May 13, 2016 Time: 9:30 a.m. - 11:30 a.m.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Answer **ALL** questions. Marks are shown in square brackets.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from a distribution with probability density function

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta > 0$ is an unknown parameter. Define $W = \prod_{i=1}^{n} X_i$.

(a) Show that $Y_i = -\ln X_i$ follows a Gamma(a, b) distribution with probability density function

$$g(y;a,b) = \begin{cases} \frac{1}{\Gamma(a)} \exp(-by) b^a y^{a-1}, & y, a, b > 0, \\ 0, & \text{otherwise}, \end{cases}$$

and determine a and b.

[5 marks]

(b) Construct the moment estimator of θ, namely $\tilde{\theta}$.

[3 marks]

(c) Construct the maximum likelihood estimator (MLE) of θ, namely $\hat{\theta}$.

[3 marks]

(d) Is the MLE $\hat{\theta}$ unbiased? If not, is it asymptotically unbiased?

[6 marks]

(e) Show that

$$\mathrm{Var}(\hat{\theta}) = \frac{n^2\theta^2}{(n-1)^2(n-2)}.$$

Hence show that $\hat{\theta}$ is a consistent estimator.

[7 marks]

(f) Determine the asymptotic distribution of $\hat{\theta}$.

[4 marks]

(g) Show that W is sufficient for θ.

[4 marks]

(h) Based on a random sample of size $n = 5$, derive the likelihood ratio test for testing

$$H_0: \theta = 1 \qquad \text{against} \qquad H_1: \theta = 2$$

and express your rejection region with significance level $\alpha = 0.01$ in term of a quantile from a Chi-square distribution. At the 0.01 level of significance, what is your conclusion if the sample drawn is $(0.6, 0.6, 0.8, 0.8, 0.9)$?

[12 marks]

(i) Find the uniformly most powerful test with significance level α for testing

$$H_0: \theta = 1 \qquad \text{against} \qquad H_1: \theta > 1.$$

[3 marks]

(j) Suppose that a random sample of size $n = 100$ is drawn from the population with $\overline{X} = 0.7$ and $\frac{1}{n}\sum_{i=1}^{n} \ln X_i = -0.4065$. Derive a large sample generalized likelihood ratio for testing

$$H_0: \theta = 2 \qquad \text{against} \qquad H_1: \theta \neq 2$$

at the 0.05 level of significance. State the rejection region, the test statistic and your conclusion clearly.

[10 marks]

[Total: 57 marks]

2. A bag contains a very large number of black and white marbles. A total of 2000 samples of 4 marbles are drawn from the bag with replacement. The frequencies of the number of black marbles, X, in these samples are tabulated below:

X	0	1	2	3	4	Total
Frequencies	20	100	430	830	620	2000

(a) Carry out an appropriate test for the hypothesis that X follows a Binomial distribution and that the ratio of the numbers of black to white marbles in the bag is 3:1 at the 5% level of significance. State the null and alternative hypotheses, the name of the test, the test statistic and your conclusion clearly.

[8 marks]

(b) Using the data above, provide an estimate for p, the true proportion of black marbles in the bag, based on the data above.

[2 marks]

(c) Carry out an appropriate test for the null hypothesis that the ratio of the numbers of black to white marbles in the bag is 3:1 at the 5% level of significance. State the null and alternative hypotheses, the name of the test, the test statistic and your conclusion clearly.

[5 marks]

(d) Are your conclusions in (a) and (c) consistent? Explain briefly.

[3 marks]

[Total: 18 marks]

3. Eddy claims that the median starting salary of female science graduates is higher than that of the male science graduates. To test for his claim, a paired difference experiment is conducted to compare the starting salary of male and female science graduates who find jobs. Pairs are formed by choosing a male and female with the same major, who took similar courses with similar grades. A random sample of $n = 15$ pairs is selected in this manner. Let X be the number of pairs with the starting salary of the male graduate higher than that of the female graduate among the 15 pairs.

   (a) Define θ, the parameter to be tested.

   [2 marks]

   (b) State the null and alternative hypotheses for Eddy.

   [2 marks]

   (c) Eddy considers X as the test statistic and sets the rejection region as $RR = \{X \leq 5\}$. Calculate α, the probability of committing a type I error.

   [4 marks]

   (d) Suppose the true value of $\theta = 0.4$. Calculate β, the probability of committing a type II error.

   [4 marks]

   (e) Based on the α and β in (c) and (d), do you think this test is good enough in terms of power and the level of significance? Make a suggestion to Eddy so that both α and β can be reduced simultaneously.

   [3 marks]

   [Total: 15 marks]

4. Let $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ be two independent normal populations. Two samples of equal sizes $n_X = n_Y = n$ from the two populations produced sample means $\overline{x} = 250$, $\overline{y} = 280$ and sample standard deviations $s_X = 24.6577$ and $s_Y = 42.7083$, respectively.

   (a) At the 0.10 level of significance, one may conclude that the two population variances are insignificantly different that the null hypothesis of equal variance cannot be rejected. What is the largest possible value of n?

   [4 marks]

   (b) Using the value of n obtained in (a), carry out an appropriate test for

   $$H_0 : \mu_X \geq \mu_Y \qquad \text{vs} \qquad H_1 : \mu_X < \mu_Y$$

   at the 0.05 level of significance. State clearly the name of the test, the test statistic, the rejection region and your conclusion clearly.

   [6 marks]

   [Total: 10 marks]

********** END OF PAPER **********

A LIST OF STATISTICAL FORMULAE

1. $M_X(t) = \mathrm{E}(e^{tX})$. $\left.\left(\dfrac{\mathrm{d}^r}{\mathrm{d}t^r} M_X(t)\right)\right|_{t=0} = \mu'_r$.

2. Normal population $\Longrightarrow \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

3. $\mathrm{Bias}(\hat{\Theta}) = \mathrm{E}(\hat{\Theta}) - \theta$. $\mathrm{E}\left[(\hat{\Theta}-\theta)^2\right] = \mathrm{Var}(\hat{\Theta}) + \left[\mathrm{Bias}(\hat{\Theta})\right]^2$.

4. $I(\theta) = \mathrm{E}\left[\left(\dfrac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] = \mathrm{E}\left[-\dfrac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right]$. $\mathrm{Var}(\hat{\Theta}) \geq \dfrac{1}{nI(\theta)}$.

5. $f(x_1, x_2, \ldots, x_n; \theta) = g\left(u(x_1, x_2, \ldots, x_n), \theta\right) \cdot h(x_1, x_2, \ldots, x_n)$.

6. $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$. $\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$. $\dfrac{(n-1)S^2}{\sigma_0^2}$. $\dfrac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$.

7. $\dfrac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$. $\dfrac{\overline{X}-\overline{Y}-\delta}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, $S_p^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

8. $\dfrac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$, $df \approx \left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2 \Bigg/ \left[\dfrac{(s_1^2/n_1)^2}{n_1-1}+\dfrac{(s_2^2/n_2)^2}{n_2-1}\right]$. $\dfrac{S_1^2}{S_2^2}$, $f_{1-\alpha,m,n} = \dfrac{1}{f_{\alpha,n,m}}$.

9. $\dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, $\hat{p} = \dfrac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$. $-2\ln\Lambda \approx \displaystyle\sum_i \frac{(O_i-E_i)^2}{E_i} = \sum_i \frac{O_i^2}{E_i} - n$.

10. $\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$. $\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$. $\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$. $\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

11. $\overline{x}-\overline{y} \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$. $\overline{x}-\overline{y} \pm t_{\alpha/2,n_1+n_2-2}s_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$. $\overline{x}-\overline{y} \pm t_{\alpha/2,df}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$.

12. $\left(\dfrac{s_1^2}{s_2^2}\cdot\dfrac{1}{f_{\alpha/2,n_1-1,n_2-1}}, \dfrac{s_1^2}{s_2^2}\cdot f_{\alpha/2,n_2-1,n_1-1}\right)$. $\hat{p}_1-\hat{p}_2 \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.