<u>THE UNIVERSITY OF HONG KONG</u>
<u>DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE</u>

STAT2602 PROBABILITY AND STATISTICS II

December 11, 2021 Time: 6:30 p.m. - 8:30 p.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 SIX questions. Marks are shown in square brackets.

1. Let X be a random variable which has a probability density function given by

$$f(x) = \begin{cases} 2e^{-2x}, & \text{for } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

   (i) Calculate the moment generating function of X. [6 marks]

   (ii) Calculate $\mathrm{E}(X^s)$ for $s = 1, 2, 3$. [9 marks]

[Total: 15 marks]

2. Let $X_1, X_2, \cdots, X_n$ be an independent random sample from a Bernoulli distribution with parameter p.

   (i) Find the maximum likelihood estimator (MLE) of p. [5 marks]

   (ii) Show that this MLE is a UMVUE of p. [10 marks]

   (iii) Construct a method of moments estimator (MME) of p, which is different from the MLE. [5 marks]

   (iv) Which estimator is better when n is large? Explain it. [5 marks]

[Total: 25 marks]

3. Let $X_1, X_2, \cdots, X_n$ be an independent random sample from $N(\mu, \sigma^2)$.

   (i) Find the MLEs of μ and σ^2 on the space

$$\Omega = \{(\mu, \sigma^2) : \mu_0 \le \mu \le \mu_1, \sigma^2 > 0\},$$

   where μ_0 and μ_1 are two given finite constants with $\mu_0 < \mu_1$. [10 marks]

   (ii) Suppose that $\mu = 0$ and $n = 10$. Find a rejection region of size 0.1 for testing hypotheses

$$H_0 : \sigma^2 = 1 \quad \text{versus} \quad H_1 : \sigma^2 = 2.$$

[10 marks]

[Total: 20 marks]

4. Let X_1, X_2, X_3, X_4 be an independent random sample of SAT mathematics scores, assumed to be $N(\mu_X, \sigma^2)$, and let $Y_1, Y_2, \cdots, Y_5$ be an independent random sample of SAT verbal scores, assumed to be $N(\mu_Y, \sigma^2)$. Suppose the following data are observed:

$$x_1 = 661 \quad x_2 = 492 \quad x_3 = 472 \quad x_4 = 623$$

$$y_1 = 565 \quad y_2 = 493 \quad y_3 = 462 \quad y_4 = 532 \quad y_5 = 644$$

   (i) Does the 95% confidence interval for μ_X contain 600? [5 marks]

(ii) Does the 95% confidence interval for $\mu_X - \mu_Y$ contain 0? [5 marks]

[Total: 10 marks]

5. A journal reported that, in a bag of m&m's chocolate peanut candies, there are 25% brown, 35% yellow, 20% blue, 15% red and 5% green candies. Suppose you purchase a bag of m&m's chocolate peanut candies at a store and find 17 brown, 20 yellow, 10 blue, 10 red, and 3 green candies, for a total of 60 candies. At the 0.05 level of significance, does the bag purchased agree with the distribution suggested by the journal?

[Total: 10 marks]

6. Let $X_1, X_2, \cdots, X_n$ be an independent random sample from a Bernoulli distribution with parameter p.

(i) Show that the most powerful test for hypotheses:

$$H_0 : p = \frac{1}{2} \quad \text{versus} \quad H_1 : p = \frac{1}{3}$$

has the rejection region $\left\{ \sum_{i=1}^{n} X_i \leq c \right\}$ for some constant c. [5 marks]

Furthermore, suppose that $n = 40$.

(ii) Use the central limit theorem to find c so that this test approximately has the size 0.05. [10 marks]

(iii) Calculate the power of this test approximately. [5 marks]

[Total: 20 marks]

A LIST OF STATISTICAL FORMULAE

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

2. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$. $\quad S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2$.

3. CLT: $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$ for large n.

4. Normal population $\Longrightarrow \dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$, $\dfrac{\overline{X} - \mu}{S/\sqrt{n-1}} \sim t_{n-1}$.

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

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

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

8. $\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}$. $\quad \overline{x} - \overline{y} \pm t_{\alpha/2,n_x+n_y-2}\sqrt{\dfrac{n_x s_x^2 + n_y s_y^2}{n_x + n_y - 2}\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}$.

9. $-2\ln\Lambda \approx \displaystyle\sum_i \dfrac{(O_i - E_i)^2}{E_i} = \sum_i \dfrac{O_i^2}{E_i} - n$. $\quad -2\ln\Lambda \approx n\left(\displaystyle\sum_i\sum_j \dfrac{O_{i,j}^2}{n_{i\cdot}n_{\cdot j}} - 1\right)$.

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