<u>THE UNIVERSITY OF HONG KONG</u>

<u>DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE</u>

STAT2602 PROBABILITY AND STATISTICS II

May 11, 2017 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 SIX questions. Marks are shown in square brackets.

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

   (i) Calculate the moment generating function of X_1. **[3 marks]**

   (ii) Calculate the moment generating function of $Y = \sum_{i=1}^{n} X_i$. **[4 marks]**

   (iii) Calculate $\mathrm{E}(Y^s)$ for $s = 1, 2$. **[8 marks]**

   [Total: 15 marks]

2. Let $X_1, X_2, \cdots, X_n$ be an independent random sample from $N(2\theta, 5\theta^2)$ with $\theta \neq 0$. Define $\hat{\theta} = \frac{A}{n}\sum_{i=1}^{n} X_i$, and let $\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[(\hat{\theta} - \theta)^2\right]$ be its mean squared error.

   (i) Choose the value of A such that $\hat{\theta}$ is an unbiased estimator of θ. **[3 marks]**

   (ii) Choose the value of A such that $\mathrm{MSE}(\hat{\theta})$ is minimized. **[5 marks]**

   (iii) Show that $T_1 = \left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right)^2$ is a consistent estimator of θ^2. **[3 marks]**

   (iv) Show that T_1 is an asymptotically unbiased estimator of θ^2. **[3 marks]**

   (v) Calculate $\mathrm{E}X_1^2$, and then construct T_2, a method of moments estimator of θ^2. **[3 marks]**

   (vi) Show that T_2 is an unbiased estimator of θ^2. **[3 marks]**

   [Total: 20 marks]

3. Let $X_1, X_2, \cdots, X_5$ be an independent random sample of SAT mathematics scores, assumed to be $N(\mu_X, \sigma^2)$, and let $Y_1, Y_2, \cdots, Y_4$ be an independent random sample of SAT verbal scores, assumed to be $N(\mu_Y, \sigma^2)$. Suppose the following data are observed:

   $$x_1 = 644 \quad x_2 = 493 \quad x_3 = 532 \quad x_4 = 462 \quad x_5 = 565$$

   $$y_1 = 623 \quad y_2 = 472 \quad y_3 = 492 \quad y_4 = 661$$

   (i) Find a 90% confidence interval for μ_X. **[5 marks]**

   (ii) Find a 90% confidence interval for $\mu_X - \mu_Y$. **[5 marks]**

   [Total: 10 marks]

4. Families were selected at random from a certain population in which the parents had been married for at least 10 years but less than 15 years. The data were tabled by two attributes in the following manner.

		Number of Children			sum
		0	1	>1	
Family Income in Dollars	<6000	11	24	93	128
	$6000-12,000$	9	28	70	107
	$>12,000$	6	15	31	52
sum		26	67	194	287

Test, at the 1% significance level, whether the attributes — family income and number of children — are independent.

[Total: 10 marks]

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

   (i) Find the MLEs of μ and σ^2 on the space
   $$\Omega_1 = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}.$$
   [8 marks]

   (ii) Find the MLEs of μ and σ^2 on the space
   $$\Omega_2 = \{(\mu, \sigma^2) : \mu_0 \leq \mu \leq \mu_1, \sigma^2 > 0\},$$
   where μ_0 and μ_1 are two given finite constants with $\mu_0 < \mu_1$. **[10 marks]**

   (iii) Will the MLE of μ in part (ii) be the UMVUE of μ? Explain it. **[2 marks]**

   (iv) Suppose that $\mu = 0$ and $n = 20$. Find a rejection region of size 0.05 for testing hypotheses
   $$H_0 : \sigma^2 = 1 \quad \text{versus} \quad H_1 : \sigma^2 = 2.$$
   [10 marks]

[Total: 30 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]**

   (ii) Use the central limit theorem to find n and c so that this test approximately has the size 0.1 under H_0 and the power 0.8 under H_1. **[10 marks]**

[Total: 15 marks]

This page is intentionally left blank

A LIST OF STATISTICAL FORMULAE

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

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

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

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