國立臺灣師範大學 109 學年度碩士班招生考試試題

科目：**機率與統計**　　　　　　　　　　　適用系所：**數學系**

注意：1.本試題共 2 頁，請依序在答案卷上作答，並標明題號，不必抄題。2.答案必須寫在指定作答區內，否則依規定扣分。

1. (30分) Suppose that X is a geometric random variable with distribution given by

$$P(X=k)=\theta^{k-1}(1-\theta),\ 0<\theta<1,\ k=1,2,3,\ldots,\infty\ .$$

Now consider a new random variable Y, where r is a positive integer and

$$Y=\begin{cases} X, & \text{if } X\le r; \\ r+1, & \text{if } X\ge r+1. \end{cases}$$

(a) Show that the distribution of Y is given by $P(Y=k)=\theta^{k-1}(1-\theta)$, $k=1,2,\ldots,r$; $P(Y=r+1)=\theta^r$.

(b) Find the expectation of Y, i.e., $E[Y]$.

(c) Suppose that $Y_1,Y_2,\ldots,Y_n$ are a random sample of size n from the distribution of Y given in (a). Show that the maximum likelihood estimator (MLE) of θ is given by $\widehat{\theta}=\dfrac{\sum_{i=1}^n Y_i-n}{\sum_{i=1}^n Y_i-M}$, where M is the number of indices i such that $Y_i=r+1$.

2. (20分) An epidemiologist is interested in determining whether or not the rate λ_1 of development of new disease cases in Community 1 is different from the rate λ_2 in Community 2. Suppose that Y_i ($i=1,2$) is a random variable denoting the number of disease cases developing in Community i during the year 2019, and that P_i represents the corresponding mid-year population as determined from census information. (You may assume that P_1 and P_2 are known constants.) The epidemiology is willing to assume that $Y_1\sim\text{Poisson}(P_1\lambda_1)$ and $Y_2\sim\text{Poisson}(P_2\theta\lambda_1)$, where the rate ratio $\theta=\lambda_2/\lambda_1$. Also, Y_1 and Y_2 are assumed to be independent random variables.

(a) Develop an appropriate likelihood ratio test statistic for testing $H_0:\theta=1$ versus $H_1:\theta\neq 1$.

(b) Suppose that this epidemiologist obtains the following data: $Y_1=100$, $Y_2=200$, $P_1=2\times10^5$, and $P_2=10^5$. Using the likelihood ratio statistic derived in (a), show that these data strongly favor rejection of $H_0:\theta=1$.

3. (30分) In a certain genetic experiment, the observed frequencies in the four distinct genotype classes AB, Ab, aB and ab were found to be n_1,n_2,n_3 and n_4, respectively, where $N=(n_1+n_2+n_3+n_4)$. Based on a certain genetic theory, the corresponding expected proportions are $(2+\theta)/4,(1-\theta)/4,(1-\theta)/4$ and $\theta/4$, respectively, where θ is an unknown parameter.

(a) Derive an explicit expression for the large-sample variance $\text{Var}(\widehat{\theta})$, where $\widehat{\theta}$ is the maximum likelihood estimator (MLE) of θ.

(b) Consider an alternative estimator $\widehat{\theta}^*$, where $\widehat{\theta}^*=(n_1-n_2-n_3+5n_4)/2N$. Prove that $\widehat{\theta}^*$ is an unbiased estimator for θ.

(c) Derive an explicit expression for the exact variance, $\text{Var}(\widehat{\theta}^*)$ and for the efficiency, $\text{Var}(\widehat{\theta})/\text{Var}(\widehat{\theta}^*)$.

4. (20分) Let $X_1, X_2, \ldots, X_n$ be n independent and identically distributed continuous random variables with the probability density function (pdf) $f(x)$ and distribution function $F(x)$. Let $Z_{n:1} \leq \ldots \leq Z_{n:n}$ be the order statistics corresponding to $X_1, \ldots, X_n$.

   (a) For an arbitrary k $(1 \leq k \leq n-1)$, write down the joint pdf of $Z_{n:1}, \ldots, Z_{n:k+1}$.

   (b) Hence, or otherwise, obtain the conditional pdf of $Z_{n:k+1}$, given $Z_{n:1}, \ldots, Z_{n:k}$.

   (c) Show that the answer to (b) depends on $Z_{n:1}, \ldots, Z_{n:k}$ only through $Z_{n:k}$.