香港統計學會考試

統計學高級證書，**2005**

試卷 **I**：統計理論
時間：**3** 小時

考生應解答五個題目
所有題目分數相等
每一題目的小題分數標明在括弧中

試場可提供畫圖紙和標準表格

考生可按照香港統計學會列載於考試指引的規例 (文件 Ex1)，使用計算器。

log 代表以 e 爲底的對數。
以其他數爲底的對數會明確標示，例如 $\log_{10}$.

注意：$\binom{n}{r}$ 等價於 ${}^{n}C_{r}$

本試卷共有 7 頁，**採用單面印刷**

本頁爲第一頁

題目 1 從第二頁開始

試卷共有八個題目

1. 英國的國家彩票參與者可從 1 至 49 的數字中隨機選取 6 個數字（不能重複）。 (2)

   (i) 證明從 49 個數字中選取 6 個數字共有 13,983,816 種方法。

   (ii) 假定有 10,000,000 個人參與這彩票抽獎。每名參與者以隨機及不重複方式，獨立地從 1 至 49 中選取 6 個數字。設 X 爲能夠選取 6 個得獎數字的總參與人數。請寫出 X 的正確分布，及有關分布的普阿松近似分布。從而計算以下的近似概率:

   (a) X = 0
   (b) X = 1 (9)

   (iii) 我們相信很多彩票參與者是從 1 至 31 中隨機猜測 6 個不重複的數字。請計算從 1 至 31 中選取 6 個不重複數字可選擇的總數目,並且推算得獎的 6 個數字均不會大過 31 的概率。 (4)

   (iv) 假定現在有 3,000,000 個彩票參與者從 1 至 31 中隨機選取 6 個不重複的數字，而同時有 7,000,000 個參與者從 1 至 49 中隨機選取 6 個不重複的數字。設 Y 爲選取了 6 個得獎數字的參與者總人數，寫出以下情況 Y 的（近似）分布。

   (a) 6 個得獎數字均是從 1 至 31 中選取。

   (b) 有一個或以上得獎數字不是從 1 至 31 中選取。 (5)

2. 我有三種方法上班。如果騎腳踏車，我的旅程時間是以 N(27,6.25) 分布，即平均數爲 27 分鐘及標準差爲 $\sqrt{(6.25)}$ = 2.5 分鐘的常態分布。若我乘坐巴士，我從家步行至巴士站的時間是以 N(7,4)分布，巴士旅程時間是以 N(13,20)分布,而從巴士站步行至工作地點則是以 N(5,1)分布。三段時間是獨立的。若我自己駕車，旅程時間是以 N(23,36)分布。

(i) 若我乘坐巴士，請找出我整個旅程時間的分布。 (3)

(ii) 那種交通方法能達到全程時間在 30 分鐘以內的機會最大，及其機會率是多少？ (5)

(iii) 那種交通方法能達到全程時間超過 35 分鐘的機會最小，而我最少需要 35 分鐘時間完成旅程的機會率是多少？ (5)

(iv) 經過多次上班的旅程後，我騎腳踏車、乘巴士或私家汽車的比例是 3:3:4。昨天，我的旅程時間在 30 分鐘以下，請計算三種交通方法分別的概率。 (7)

3. 一個通訊系統的失靈次數按一個比率參數爲 λ 的普阿松過程出現，故此在時間 t 內系统失靈次數的隨機變量 X 符合以下公式:

$$P(X = x) = \exp(-\lambda t)\frac{(\lambda t)^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

(i) 考慮在時間 t 內沒有任何失靈的概率或其他方法，證明在出現下次失靈前的時間 T 的概率分布爲

$$P(T > t) = \exp(-\lambda t), \quad t > 0,$$

及推演 T 的概率密度函數（pdf）。 (4)

(ii) 一組包括 n 個相同及獨立運作的上述通訊系統在時間 0 開始運作。T_i 表示第 i 個系統出現失靈的時間，i = 1, …, n。請寫出所有 n 個系統毫無失靈地連續運作而時間最少爲 t 的概率。請推演出 T_{min} = (T_1, …, T_n) 中最小值（即該組 n 個系統出現第一次失靈的時間）的概率密度函數爲指數格式，比率參數爲 n 及 λ 的函數，請註明 n 及 λ 的值。 (5)

(iii) 證明所有 n 個系统已經在時間 t 前出現失靈的概率爲 $(1-e^{-\lambda t})^n$，並且推演 T_{max} = (T_1, …, T_n) 中最大值（即該組 n 個系統中最後失靈出現的時間）的概率密度函數。

已知 n = 10 及 λ = 0.002，找出會出現以下情形的時間 t_1 及 t_2

$$P(T_{\min} < t_1) = P(T_{\max} > t_2) = 0.05 .$$ (11)

4. (i) 連續隨機變量 X 的概率密度函數分布 f(x)爲

$$f(x) = \alpha\,(1-x)^{\alpha-1}\,,\quad 0 < x < 1,\quad \alpha > 0\,.$$

找出 X 的累積分布函數，F(x)，從而得出 X 的中位數。並且以 $\alpha = 3$，繪製 f(x)及 F(x)的曲線圖。 (8)

(ii) 從這分布中選取一隨機樣本 $x_1, \ldots, x_n$，以估計未知參數 α。請寫出這些數據的似然函數，記作 $L(x_1, x_2, \ldots, x_n \mid \alpha)$，及證明 α 的極大概似估計量（MLE）爲

$$\hat{\alpha} = \frac{-n}{\sum_{i=1}^{n} \log(1-x_i)}\,. \qquad (4)$$

此外，以 α 及 n 的函數得出 $\dfrac{d^2 \log L(x_1, x_2, \ldots, x_n \mid \alpha)}{d\,\alpha^2}$。假定 $\hat{\alpha}$ 是近似常態分布的，平均數爲 α 及方差爲

$$-1 \Big/ \left[\frac{d^2 \log L(x_1, x_2, \ldots, x_n \mid \alpha)}{d\,\alpha^2}\right]$$

推演 α 的近似９０%置信區間。根據樣本爲 0.12, 0.43, 0.07, 0.87, 0.29，計算這區間。 (8)

5. (i) 一社會科學家正進行一個學生濫用藥物的調查。她的調查是根據一個從 Wackford Squeers University 隨機抽選一個有 n 位學生的樣本，然後進行面談訪問，當中 y 個學生對她說有濫用藥物。假設在被抽選的人口中，濫用藥物者的真正百分比爲 p，他們的回應是誠實的，樣本中濫用藥物的人數服從二項分佈 $B(n, p)$。寫出這些數據的似然函數；求 p 的極大似然估計，記作 $\hat{p}$。並寫出 $\mathrm{Var}(\hat{p})$和進而求 $\hat{p}$ 的標準誤差的估計，記作 $\mathrm{SE}(\hat{p})$。已知 n=100 和 y = 20，計算 $\hat{p}$ 的值和 $\mathrm{SE}(\hat{p})$。 (8)

(ii) 一統計師現忠告她，因爲藥物濫用問題的敏感性，一些學生可能沒有誠實回答，因此構成偏差。他建議採用一個匿名的策略去鼓勵誠實的答案；他提供一個有偏差的錢幣，一面寫上“濫用藥物”，另一面則寫上“沒有濫用藥物”。當擲這錢幣時，它給出每一面的概率分別爲 0.75 和 0.25。

獨立於第一個樣本，抽選另一個也是 n 容量的隨機樣本。n 個學生每人被要求在沒有被訪問員看見的情況下擲這錢幣，如果他／她是在錢幣顯示的組別內，要回答“是”，其他則回答“否”。

假設回應是誠實的，證明：“是”的概率，記作 θ，爲

$\theta = 0.25 + 0.5p$.

此外假設學生回答“是”的人數，記作 Z，服從 $B(n, \theta)$ 的分布，有觀察值 z。寫出 θ的極大似然估計，記作 $\hat{\theta}$，和它的標準誤差的估計。用 θ 與 p 之間的關係去推演 p 的極大似然估計，記作 $\tilde{p}$，和 SE($\tilde{p}$)。已知 $n = 100$ 和 $z = 45$，計算 $\tilde{p}$ 的值和 SE($\tilde{p}$)。 (8)

(iii) 一新聞記者比較了上述兩個調查的結果，說第一個調查較爲可靠。你是否同意？你爲什麼同意或不同意？ (4)

6. 隨機變量 X 具有幾何分佈 $f(x)$，其中

$$f(x) = (1-p)^x p, \quad x = 0, 1, 2, \ldots, \quad 0 < p < 1 \ .$$

(i) 爲 $0 \le x \le 5$，劃出 $p = 1/3$ 時 $f(x)$的圖。 (4)

(ii) 證明：X 的概率母函數是 $G(s)$，其中

$$G(s) = \frac{p}{1-(1-p)s}, \quad |s| < \frac{1}{1-p} \ ,$$

並根據這函數或其他函數，求 X 的均值和方差。 (6)

(iii) 對任何非負整數 x，證明：$P(X \ge x) = (1-p)^x$，並推演對任何非負整數 l 和 m

$$P(X \ge l+m \mid X \ge l) = P(X \ge m).$$

解釋這結果。 (4)

(iv) 隨機變量 Y 具有幾何分佈 $g(y)$，其中

$$g(y) = (1-\theta)^y \theta, \quad y = 0, 1, 2, \ldots, \quad 0 < \theta < 1 \quad .$$

X 和 Y 是獨立的；隨機變量 Z 定義爲 X 和 Y 的最小值，即 $Z = \min(X, Y)$。由於留意到 $P(Z \geq z) = P(X \geq z$ 和 $Y \geq z)$，求 $P(Z \geq z)$的表達式，其中 z 是任何非負整數。考察 $P(Z \geq z) - P(Z \geq z + 1)$，由此或其他，證明

$$P(Z = z) = \left[(1-p)(1-\theta)\right]^z (p + \theta - p\theta), \quad z = 0, 1, 2, \ldots \quad .$$

分辨這分佈的形式；從而寫出 $E(Z)$和 $\mathrm{Var}(Z)$。 (6)

7. 以標準松樹樣本每平方米牛頓計，不同木紋角度 $x°$量度的爆破壓力如下表。

爆破壓力 $y\,(\mathrm{N/m^2}) \times 10^{-10}$	0.987	1.064	1.337	1.912	2.740	5.771	11.494
角度 $x°$	0	15	30	45	60	75	90

(i) 把這些數據劃圖。用以下提供的摘要資料，計算 x 與 y 之間相關 $\mathrm{corr}(x, y)$的積矩係數；並簡要評論。 (7)

(ii) 對 y 以 x 的方式表達，Hankinson's 公式的形式是

$$y = \left(\frac{1 - \sin^2 x}{a} + \frac{\sin^2 x}{b} \right)^{-1} ,$$

其中 a 和 b 分別爲平行和垂直於木紋的標準爆破壓力。證明：這關係怎樣可以線性方式表達。

$Y = A + BX$,

其中 $Y = 1/y$ 和 $X = \sin^2 x$，A 和 B 是你所求未知參數 a 和 b 的函數。 (3)

(iii) 以 Y 相對於 X 劃圖。評論這關係對回歸分析的適合性。用最小二乘去估計 A 和 B，推導 a 和 b 相應的估計；並計算 corr(X, Y)和比較這與(i)計算的相關。 (10)

註釋: (A) 清楚給出任何沒有證明的假設公式。

(B)

$$\sum x = 315,\ \sum x^2 = 20475,\ \sum y = 25.305,\ \sum y^2 = 180.474,\ \sum xy = 1773.795,$$
$$\sum \sin^2 x = 3.5,\quad \sum \sin^4 x = 2.75,\quad \sum\left(\frac{1}{y}\right) = 3.849,\quad \sum\left(\frac{1}{y^2}\right) = 2.9136,\quad \sum\frac{\sin^2 x}{y} = 1.03385.$$

8. X 和 Y 的聯合概率質量函數如下表

		Y的值		
		0	1	2
X的值	−1	1/6	1/12	1/12
	0	1/12	1/6	1/12
	1	1/12	1/12	1/6

(i) 求 X 和 Y 的邊際分佈，從而計算 $E(X)$, $E(Y)$, Var(X)和 Var(Y)。 (5)

(ii) 爲每個 X 的可能值，求 Y 的條件分佈，從而證明：$E(Y | X = x)$是一個 x 的線性函數。 (5)

(iii) 求 $E(XY)$，並推演 Cov(X, Y)和 corr(X, Y)。X 和 Y 是否獨立？ (5)

(iv) 求 $Z = X^3 + (Y - 1)^3$ 的概率分佈。 (5)