《統計學》

試題評析	今年度試題分爲四大題，第一大題爲二元機率分配，考生只要掌握住二項分配是如何延伸出三項分配的觀念，就可以輕易地把一元超幾何分配延伸至二元超幾何分配；第二大題爲偏向數理統計範圍，但類似的題目在考古題中已經出現過，應可獲得較高分數；第三大題爲無母數統計中的兩相依母體Wilcoxon檢定，有準備到的考生應該可輕易得分；最後的第四大題有關時間序列，爲本份考卷中考最偏的一題，雖然計算不困難，可是應該有很多考生沒有準備到此主題。本卷基本分50分，程度較好的考生可以拿70分以上。

一、從一付撲克牌（共52張）以抽後不放回的方式隨機抽出2張，令X代表紅桃之張數，Y代表黑色牌之張數。
(一) 試求X與Y之聯合機率分配。（5分）
(二) 分別求出X與Y之邊際機率分配。（10分）
(三) 求出Y給定X＝1之下的條件機率分配。（5分）
(四) 求出Y給定X＝x之下的條件機率分配。（5分）

考點命中	《高點統計學講義》第二回，趙治勳編撰，第5.3節。

答：

(一) $(X,Y) \sim \text{Hyper}(N=52, m_1=13, m_2=26, n=2)$ 二元超幾何分配

$$f_{XY}(x,y)=\frac{\binom{13}{x}\binom{26}{y}\binom{13}{2-x-y}}{\binom{52}{2}},\ x=0,1,2,\ y=0,1,2,\ 0\le x+y\le 2$$

(二) $X \sim \text{Hyper}(N=52, m_1=13, n=2)$ 一元超幾何分配

$$f_X(x)=\sum_{y=0}^{2-x} f_{XY}(x,y)=\frac{\binom{13}{x}\binom{39}{2-x}}{\binom{52}{2}},\ x=0,1,2$$

$Y \sim \text{Hyper}(N=52, m_2=26, n=2)$ 一元超幾何分配

$$f_Y(y)=\sum_{x=0}^{2-y} f_{XY}(x,y)=\frac{\binom{13}{y}\binom{39}{2-y}}{\binom{52}{2}},\ y=0,1,2$$

(三) $Y|X=1 \sim \text{Hyper}(N-m_1=39, m_2=26, n-x=1)$ 一元超幾何=分配

$$f_{Y|X}(y|x=1)=\frac{f_{XY}(x=1,y)}{f_X(x=1)}=\frac{\binom{26}{y}\binom{13}{1-y}}{\binom{39}{1}},\ y=0,1$$

(四) $Y|X=x \sim Hyper(N-m_1=39, m_2=26, 2-x)$ 一元超幾何分配

$$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{\binom{26}{y}\binom{13}{2-x-y}}{\binom{39}{2-x}},\ x=0,1,2,\ y=0,1,2,\ 0\le x+y\le 2$$

二、令$X_1, X_2, \ldots, X_n$為一組隨機抽自均勻分配(Uniform distribution)之樣本,其機率密度函數為

$$f_x(x)=\frac{1}{\theta} \qquad \forall x\in[0,\theta]$$
$$=0 \qquad 其他$$

令$\bar{X}$代表樣本平均數。

(一) 試以動差法求得θ的點估計。(5分)
(二) 試求θ的充分統計量。(5分)
(三) 試求θ的最大概似估計。(5分)
(四) 試驗證θ的最大概似估計並非θ的不偏估計。(5分)
(五) 將θ的最大概似估計記作$\tilde{\theta}$。試求c使得$c\tilde{\theta}$為θ的不偏估計。(5分)
(六) 令(五)中之不偏估計為$\breve{\theta}$。並令$\hat{\theta}=2\bar{X}$。分別求$\breve{\theta}$與$\hat{\theta}$的變異數。(10分)
(七) 當$n>1$,說明是否$\breve{\theta}$相對於$\hat{\theta}$的相對效率(relative efficiency)大於1?(5分)

考點命中	《高點統計學講義》第三回,趙治勳編撰,第8.3節及第8.4節。

答:

母體: $X\sim U(0,\theta)$

樣本: $X_1,X_2,\cdots,X_n \overset{iid}{\sim} U(0,\theta)$

(一) 令 $E(X)=\frac{\theta}{2}=\bar{X} \Rightarrow \hat{\theta}_{MME}=2\bar{X}$

(二) $f(x_1,x_2,\cdots,x_n;\theta)=\frac{1}{\theta^n}I_{(x_{(n)},\infty)}(\theta)=h(x_1,x_2,\cdots,x_n)g(x_{(n)};\theta)$

其中$h(x_1,x_2,\cdots,x_n)=1$, $g(x_{(n)};\theta)=\frac{1}{\theta^n}I_{(x_{(n)},\infty)}(\theta)$

由Neyman-Fisher分解定理, $X_{(n)}=\max(X_1,X_2,\cdots,X_n)$爲$\theta$之充分統計量

(三) $L(\theta)=f(x_1,x_2,\cdots,x_n;\theta)=\frac{1}{\theta^n}I_{(x_{(n)},\infty)}(\theta)$

當$\theta=X_{(n)}$時, $L(\theta)$具有最大值

$\therefore X_{(n)}=\max(X_1,X_2,\cdots,X_n)$爲$\theta$之最大概似估計量

(四) $X_1,X_2,\cdots,X_n \overset{iid}{\sim} U(0,\theta)$

令$Y_i=\frac{X_i}{\theta}\sim U(0,1)$

$Y_1,Y_2,\cdots,Y_n \overset{iid}{\sim} U(0,1)$

$Y_{(n)}=\frac{X_{(n)}}{\theta}\sim Beta(n,1)$

故$E(Y_{(n)})=E(\frac{X_{(n)}}{\theta})=\frac{n}{n+1}\quad\Rightarrow E(X_{(n)})=\frac{n}{n+1}\theta(\neq\theta)$

$X_{(n)}$不為θ之不偏估計量

也可得，$V(Y_{(n)})=V(\frac{X_{(n)}}{\theta})=\frac{n}{(n+1)^2(n+2)}\Rightarrow V(X_{(n)})=\frac{n}{(n+1)^2(n+2)}\theta^2$

(五) $\tilde{\theta}=X_{(n)}$

$E(\frac{n+1}{n}X_{(n)})=\theta$　故$\frac{n+1}{n}X_{(n)}$為θ之不偏估計量

(六) $\breve{\theta}=\frac{n+1}{n}X_{(n)}$

$$V(\breve{\theta})=V(\frac{n+1}{n}X_{(n)})=(\frac{n+1}{n})^2V(X_{(n)})=(\frac{n+1}{n})^2\frac{n}{(n+1)^2(n+2)}\theta^2=\frac{\theta^2}{n(n+2)}$$

$\hat{\theta}=2\overline{X}$

$$V(\hat{\theta})=4V(\overline{X})=4(\frac{\frac{\theta^2}{12}}{n})=\frac{\theta^2}{3n}$$

也可得，$E(\hat{\theta})=2E(\overline{X})=2(\frac{\theta}{2})=\theta$

(七) $MSE(\breve{\theta})=V(\breve{\theta})+[E(\breve{\theta})-\theta]^2=V(\breve{\theta})=\frac{\theta^2}{n(n+2)}$

$$MSE(\hat{\theta})=V(\hat{\theta})+[E(\hat{\theta})-\theta]^2=V(\hat{\theta})=\frac{\theta^2}{3n}$$

$$\frac{MSE(\breve{\theta})}{MSE(\hat{\theta})}=\frac{\frac{\theta^2}{n(n+2)}}{\frac{\theta^2}{3n}}=\frac{3}{n+2}<1\quad(\because n>1)$$

三、王老師想瞭解學生在經過他所設計之聽力訓練後，英文聽力是否有進步，全班學生在訓練前與訓練後分別接受難易度相似的英聽測驗。隨機從班上抽出20名學生，下表是這些學生訓練前後的成績差異。其中，成績差異＝訓練後的成績－訓練前的成績。

11.5	15.0	8.5	3.5	–4.5	–2.5	–8.0	7.5	10.0	20.0
–4.0	–1.5	14.0	–3.0	–6.5	18.0	–7.0	13.0	12.0	19.0

假設訓練前後的成績差異不是常態分配。將成績差異絕對值由小而大做排序，給予對應之排名；舉例說明，第12位學生成績差異的絕對值最小，所以排名為1，第6位學生的排名為2，以此類推。令$W=\sum_{i=1}^{20}(iU_i)$，其中

$U_i=1$若排名i對應的成績差異是正數，

$=0$若排名i對應的成績差異是負數。

(一)若該訓練對學生英聽能力並無影響，試問U_i服從什麼分配？（5分）

(二)若該訓練對學生英聽能力並無影響，試求W的期望值與變異數。（10分）

(三)利用統計量W以及(二)中的結果，在顯著水準0.05下，檢定是否學生的英聽能力有進步。（5分）

考點命中	《高點統計學講義》第四回，趙治勳編撰，頁84。

答：

(一)

若該訓練對學生英聽能力並無影響的話，成績差異可以期望會有一半正數一半負數，故

$$U_i \overset{iid}{\sim} Ber(p=\frac{1}{2}), i=1,2,\cdots,20$$

(二)

$$E(W)=E(\sum_{i=1}^{20} iU_i)=(1+2+\cdots+20)\frac{1}{2}=\frac{20(20+1)}{2}\frac{1}{2}=105$$

$$V(W)=V(\sum_{i=1}^{20} iU_i)=(1^2+2^2+\cdots+20^2)\frac{1}{2}\frac{1}{2}=\frac{20(20+1)(2(20)+1)}{6}\frac{1}{2}\frac{1}{2}=717.5$$

(三) 設$W=\min\{+\text{等級和},-\text{等級和}\}=\min\{168,42\}=42$

H_0:該訓練對學生英聽能力並無影響 vs H_1:該訓練對學生英聽能力有影響

T.S.： $Z=\dfrac{W-E(W)}{\sqrt{V(W)}} \overset{d}{\sim} N(0,1)$

R.R.：Reject H_0 at $\alpha=0.05$ if $Z^*<-z_{0.05}$=-1.645

$\because Z^*=\dfrac{42-105}{\sqrt{717.5}}=-2.352$ $\therefore$ reject H_0

我們有足夠證據去推論該訓練對學生英聽能力具有顯著影響

四、蘭花餐廳王老闆分析該餐廳自98年第一季至102年第四季（共20季），每季之營業額（萬元）。首先，他算出四季之季節指數如下：

第一季	第二季	第三季	第四季
1.25	1.45	0.70	0.60

其次，王老闆將這些營業額分別除以其對應之季節指數，得到一組去除季節因子之時間數列y_1, $y_2, \ldots, y_{20}$；接下來，利用簡單線性模型，他求出去除季節因子之時間數列的長期趨勢估計式$\hat{y}_t=21.5+0.8t$

(一)各季之營業額與四季平均營業額之比較如何？（5分）

(二)預測103年第一季至第四季的營業額。（10分）

答：

(一) 由題目給定四季之季節指數可得，

第一季及第二季之營業額較四季平均營業額爲高

而第三季及第四季之營業額較四季平均營業額爲低

(二) 注意：資料已經除以季節指數

103年第一季的營業額預測(t=21), $\hat{y}_{21}=21.5+0.8(21)=38.3$(萬元)

103年第二季的營業額預測(t=22), $\hat{y}_{22}=21.5+0.8(22)=39.1$(萬元)

103年第三季的營業額預測(t=23), $\hat{y}_{23}=21.5+0.8(23)=39.9$ (萬元)

103年第四季的營業額預測(t=24), $\hat{y}_{24}=21.5+0.8(24)=40.7$ (萬元)