《統計學》

一、新北市捷運環狀線部分路段預定在2019年年底完工通車，目前已進入測試階段。已知新北市捷運電聯車的液壓式摩擦煞車系統的失效時間是服從具有自由度為2的卡方分配。今由工程師隨機抽樣測試該系統，得到一組樣本數為二之失效時間的隨機樣本，分別以 Y_1 與 Y_2 表示。（每小題10分，共40分）

(一)令變數 $D=Y_1-Y_2$，求出變數D之機率密度函數f(d)。

(二)令變數 $S=Y_1+Y_2$，求出變數S之變異數Var(S)。

(三)令變數R為全距，求出變數R之機率密度函數f(r)。

(四)令變數 $U=\min\{Y_1,Y_2\}$ 及 $V=\max\{Y_1,Y_2\}$，求出機率 $P(U<6,V>10)$。

試題評析	本題涉及隨機變數間之變數變換與順序統計量兩個主題，都是近年來常有命題的題型，老師也有在題庫本上提供學員很多練習機會，要拿到滿分不難。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第五章第四節。 《高點・高上統計學講義》第二回，趙治勳編撰，第九章第五節。

答：

樣本： $Y_1,Y_2\overset{iid}{\sim}\chi^2_{(2)}=Exp(\lambda=\frac{1}{2})$

(一) $f_{Y_1Y_2}(y_1,y_2)=f_{Y_1}(y_1)f_{Y_2}(y_2)=\frac{1}{4}e^{-\frac{y_1+y_2}{2}},0<y_1,y_2<\infty$

令 $\begin{matrix}D=Y_1-Y_2\\W=Y_2\end{matrix}\Rightarrow\begin{matrix}Y_1=D+W\\Y_2=W\end{matrix}$，$J=\begin{vmatrix}1&1\\0&1\end{vmatrix}=1$，$0<w<\infty,-w<d<\infty$

$$f_{DW}(d,w)=f_{Y_1Y_2}(y_1=d-w,y_2=w)\times|1|=\frac{1}{4}e^{-\frac{d+w+w}{2}}$$

$$=\frac{1}{4}e^{-\frac{d+2w}{2}},-w<d<\infty,0<w<\infty$$

$$f_D(d)=\begin{cases}\int_{-d}^{\infty}\frac{1}{4}e^{-\frac{d+2w}{2}}dw=\frac{1}{4}e^{\frac{d}{2}},-\infty<d<0\\\int_{0}^{\infty}\frac{1}{4}e^{-\frac{d+2w}{2}}dw=\frac{1}{4}e^{-\frac{d}{2}},0\le d<\infty\end{cases}$$

(二) $V(S)=V(Y_1+Y_2)=V(Y_1)+V(Y_2)=4+4=8$

(三) $U=\min(Y_1,Y_2),V=\max(Y_1,Y_2)$

$$g_{UV}(u,v)=2!\times\frac{1}{4}e^{-\frac{u+v}{2}}=\frac{1}{2}e^{-\frac{u+v}{2}},0<u\le v<\infty$$

令 $\begin{matrix}R=V-U\\W=U\end{matrix}\Rightarrow\begin{matrix}U=W\\V=R+W\end{matrix}$，$J=\begin{vmatrix}0&1\\1&1\end{vmatrix}=-1$，$\begin{matrix}0<w\le r+w<\infty\\\Rightarrow 0<w<\infty,0<r<\infty\end{matrix}$

$$f_{RW}(r,w)=g_{UV}(u=w,v=r+w)\times|-1|=\frac{1}{2}e^{-\frac{r+2w}{2}},\ 0<w<\infty,0<r<\infty$$

$$f_R(r)=\int_0^{\infty}\frac{1}{2}e^{-\frac{r+2w}{2}}dw=\frac{1}{2}e^{-\frac{r}{2}},0\le d<\infty$$

(四) $P(U<6\cap V>10)=\int_{10}^{\infty}\int_0^{6}g_{UV}(u,v)dudv=\int_{10}^{\infty}\int_0^{6}\frac{1}{2}e^{-\frac{u+v}{2}}dudv=0.0128$

二、某經濟學者長期研究國內農民的家戶所得，發現國內的農家所得是服從平均數為μ，變異數為σ^2之常態分配。針對評估農家所得分配是否不平均的狀況，此經濟學者提出了以所得分配的第90百分位數和第10百分位數的差（以θ表示），做為評估農家所得分配是否不平均的評估指標。今由全臺灣農家隨機抽取n筆家戶所得的資料，得到隨機樣本$X_1,X_2,...,X_n$。令F(x)為農家所得分配之累積分配函數（cumulative distribution function）。（每小題10分，共30分）

(一)假設μ和σ^2皆未知，求出上述評估指標θ之最小變異不偏估計量（uniformly minimum variance unbiased estimator），$\hat{\theta}$。

(二)假設σ^2已知但μ未知，求出上述累積分配函數F(x)之不偏估計量的變異數的Cram’er-Rao下限（Cram’er-Rao lower bound）。

(三)假設σ^2已知但μ未知，求出上述F(x)之信賴水準為100(1－α)%的信賴區間。

試題評析	本題難度較高，涉及數統範圍中UMVUE與CRLB，但考古題上都有出現過，（二）與（三）兩個小題是關鍵，其中有一個轉換為伯努力分配的過程，考生要很熟悉題意才有辦法得分。
考點命中	《迴歸分析熱門題庫》，高點文化出版，趙治勳編著，第一篇第四章。

答：

(一)母體：$X\sim N(\mu,\sigma^2)$

令母體第90個百分位數為$x_{0.9}$與母體第10個百分位數為$x_{0.1}$

$$P(X\le x_{0.9})=P(Z\le\frac{x_{0.9}-\mu}{\sigma})=0.9\Rightarrow\frac{x_{0.9}-\mu}{\sigma}=z_{0.1}=1.282\Rightarrow x_{0.9}=\mu+1.282\sigma$$

$$P(X\le x_{0.1})=P(Z\le\frac{x_{0.1}-\mu}{\sigma})=0.1\Rightarrow\frac{x_{0.1}-\mu}{\sigma}=z_{0.9}=-1.282\Rightarrow x_{0.1}=\mu-1.282\sigma$$

$$\theta=x_{0.9}-x_{0.1}=2.564\sigma$$

樣本：$X_1,X_2,...,X_n\overset{iid}{\sim}N(\mu,\sigma^2)$　　$S^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

已知$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}$，　$S^2=\frac{\sigma^2}{n-1}\frac{(n-1)S^2}{\sigma^2}\sim Gamma(\alpha=\frac{n-1}{2},\lambda=\frac{n-1}{2\sigma^2})$

S^2為σ^2之c.s.s.

$$E(2.564S)=2.564E[(S^2)^{\frac{1}{2}}]=2.564\frac{\Gamma(\frac{n-1}{2}+\frac{1}{2})}{\Gamma(\frac{n-1}{2})(\frac{n-1}{2\sigma^2})^{\frac{1}{2}}}=2.564\frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n-1}\Gamma(\frac{n-1}{2})}\sigma$$

$$\Rightarrow E(2.564\frac{\sqrt{n-1}\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}S)=2.564\sigma=\theta$$

由Lehmann-Scheffe, $2.564\frac{\sqrt{n-1}\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}S$ 為 θ 之UMVUE

(二)令 $Y_i=\begin{cases}1, X_i\le x\\ 0, o.w.\end{cases}\sim Ber(p=F_X(x))$

樣本： $Y_1,Y_2,...,Y_n\overset{iid}{\sim}Ber(p=F_X(x))$

$$CRLB(p=F_X(x))=\frac{[\frac{d}{dp}p]^2}{-nE(\frac{d^2}{dp^2}\ln f(Y\mid p))}=\frac{1}{-nE(-\frac{Y}{p^2}+\frac{1-Y}{(1-p)^2})}$$

$$=\frac{1}{-n(-\frac{1}{p}+\frac{1}{1-p})}=\frac{p(1-p)}{n}$$

其中 $f(y\mid p)=p^y(1-p)^{1-y}$

$\ln f(y\mid p)=y\ln p+(1-y)\ln(1-p)$

$\frac{d}{dp}\ln f(y\mid p)=\frac{y}{p}-\frac{1-y}{1-p}$

$\frac{d^2}{dp^2}\ln f(y\mid p)=-\frac{y}{p^2}+\frac{1-y}{(1-p)^2}$

(三)母體： $Y\sim Ber(p=F_X(x))$

樣本： $Y_1,Y_2,...,Y_n\overset{iid}{\sim}Ber(p=F_X(x))$

點估計： $\bar{Y}\underset{byC.L.T.}{\sim}N(p,\frac{p(1-p)}{n})$

樞紐量： $\frac{\bar{Y}-p}{\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}}\underset{byC.L.T.}{\sim}N(0,1)$

機率區間: $P(-z_{\frac{\alpha}{2}}\le\frac{\bar{Y}-p}{\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}}\le z_{\frac{\alpha}{2}})=1-\alpha$

信賴區間: $P(\bar{Y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}\le p\le\bar{Y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}})=1-\alpha$

結論: $p=F_X(x)$ 之 $(1-\alpha)100\%$ 信賴區間為

$$(\bar{Y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}},\bar{Y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}})$$

三、同時丟擲三個均勻的骰子（均勻的骰子，指骰子出現每一點的機率均等）5次，令變數T代表5次丟擲中三個骰子出現的點數皆不同的次數，且令變數W代表5次丟擲中三個骰子出現的點數皆相同的次數。（每小題10分，共30分）

(一)求出變數T與W之聯合機率密度函數f(t,w)。

(二)求出給定T＝t之下，W之條件機率密度函數 $f(w|t)$。

(三)求出條件變異數的期望值 $E[V(W|T)]$。

試題評析	本題涉及三項分配，都是近年來常有命題的章節，老師也有在題庫本上提供學員很多練習機會，要拿到滿分不難。
考點命中	《高點・高上統計學講義》第二回，趙治勳編撰，第八章第二節。

答：

(一)由題意，$(T,W)\sim$ 三項$(n=5,p_T=\frac{120}{216},p_W=\frac{6}{216})$

$$f_{TW}(t,w)=\frac{5!}{t!w!(5-t-w)!}(\frac{120}{216})^t(\frac{6}{216})^w(\frac{90}{216})^{5-t-w},\ t,w=0,1,2,3,4,5,\ 0\le t+w\le 5$$

(二) $W|T=t\sim \text{Bin}(5-t,\frac{p_W}{1-p_T}=\frac{1}{16})$

$$f_{W|T}(w|t)=\frac{(5-t)!}{w!(5-t-w)!}(\frac{1}{16})^w(\frac{15}{16})^{5-t-w},\ t,w=0,1,2,3,4,5,\ 0\le t+w\le 5$$

(三) $E[V(W|T)]=E[(5-T)\frac{1}{16}\frac{15}{16}]=\frac{15}{256}[5-E(T)]$

$$=\frac{15}{256}[5-\frac{25}{9}]=\frac{25}{192}$$

其中 $T\sim \text{Bin}(5,p_T=\frac{120}{216})$ ，$E(T)=5\times\frac{120}{216}=\frac{25}{9}$