HONG KONG STATISTICAL SOCIETY

2015 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE – MODULE 5

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

The solutions are intended as learning aids and should not be seen as "model answers".

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

MCS 2015

1. Correlation of X and Y = $\frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$ (1) where $cov(X,Y) = E((X-E(X))(Y-E(Y))$ (1) 2

If $Corr(X,Y) = -1$ then there is an exact linear relationship between X and Y, $Y = a + bX$, with $b < 0$. (1) (1) 2

(i) Let X = father's height, Y = eldest son's height

$cov(X,Y) = Corr(X,Y)\sqrt{var(X)var(Y)}$ (1) $= 0.52 \times 7.0 \times 7.2 = 26.208$ (1) 2

(ii) Sample mean height of father and son = $\frac{X+Y}{2}$ (1). This is a linear function of X and Y and so has a (univariate) normal distribution.

$E\left(\frac{X+Y}{2}\right) = \frac{1}{2}(E(X) + E(Y)) = 177.0$ (1)

$var\left(\frac{X+Y}{2}\right) = \frac{1}{4}(var(X) + var(Y) + 2cov(X,Y))$ (1)

$= \frac{1}{4}(7.0^2 + 7.2^2 + (2 \times 26.208)) = \frac{153.256}{4} = 38.314$ (1)

$\therefore \frac{X+Y}{2} \sim N(177.0, 38.314)$ (1)

$P\left(\frac{X+Y}{2} > 180\right) = 1 - \Phi\left(\frac{180-177}{\sqrt{38.314}}\right) = 1 - \Phi(0.4847)$ (1)

$= 1 - 0.6861$ (using normal tables) $= 0.3139$. (1) 7

[any answer between 0.3130 and 0.3150 is acceptable]

(iii) Need to evaluate $P(Y > 1.05X) = P(Y - 1.05X > 0)$. (1)

$Y - 1.05X$ is univariate normal.

$E(Y - 1.05X) = 178.0 - (1.05 \times 176.0) = -6.8$ (1)

$var(Y - 1.05X) = var(Y) + (1.1025 \times var(X)) - 2.1cov(X,Y)$ (1)

$= 7.2^2 + (1.1025 \times 7.0^2) - (2.1 \times 26.208)$

$= 50.8257$ (1)

$\therefore Y - 1.05X \sim N(-6.8, 50.8257)$ (1)

$P(Y - 1.05X > 0) = 1 - \Phi\left(\frac{0-(-6.8)}{\sqrt{50.8257}}\right) = 1 - \Phi\left(\frac{6.8}{7.1292}\right)$

$= 1 - \Phi(0.9538) = 1 - 0.8299$

$= 0.1701$ (1) 7

[any answer between 0.1690 and 0.1710 is acceptable]

HC5 (2015)

2 (i) pgf $\pi_Y(t) = E(t^Y) = \sum_{k=0}^{\infty} t^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(t\lambda)^k}{k!}$ ① ①

$= e^{-\lambda}e^{t\lambda} = e^{-\lambda(1-t)}$ ① 3

(ii) $\frac{d\pi_Y(t)}{dt} = \lambda e^{-\lambda(1-t)}$ ①

$\frac{d^2\pi_Y(t)}{dt^2} = \lambda^2 e^{-\lambda(1-t)}$ ①

$E(Y) = \frac{d\pi_Y(t)}{dt}\Big|_{t=1} = \lambda$ ①

$E(Y(Y-1)) = E(Y^2) - E(Y) = \frac{d^2\pi_Y(t)}{dt^2}\Big|_{t=1} = \lambda^2$

$\therefore E(Y^2) = \lambda^2 + \lambda$ ①

$\text{var}(Y) = E(Y^2) - (E(Y))^2 = \lambda$ ① 5

(iii) Let $W = \sum_{i=1}^{n} Y_i$, so that

$\pi_W(t) = E(t^W) = E(t^{\sum Y_i}) = \prod_{i=1}^{n} E(t^{Y_i})$ ① $= \prod_{i=1}^{n} e^{-\lambda(1-t)} = e^{-n\lambda(1-t)}$ ①

which is the pgf of the Poisson distribution, parameter $n\lambda$.

By the 1-1 correspondence between distributions and pgf's, we can deduce that $W \sim$ Poisson, parameter $n\lambda$. ① 3

(iv) Likelihood, $L(\lambda) = \prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{Y_i}}{Y_i!} = \frac{e^{-n\lambda}\lambda^{\sum Y_i}}{\prod(Y_i!)}$ ①

Log likelihood, $l(\lambda) = -n\lambda + \sum Y_i \log(\lambda) + \text{constant}$ ①

$\frac{\partial l}{\partial \lambda} = -n + \frac{\sum Y_i}{\lambda}$ ①

$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{\sum Y_i}{\lambda^2} < 0$ ①

$\therefore$ Maximum occurs at $\frac{\partial l}{\partial \lambda} = 0$ i.e. $\hat{\lambda} = \frac{\sum Y_i}{n}$. ① 5

(v) Method of moments: set population mean, λ, equal to the sample mean, $\frac{\sum Y_i}{n}$, and solve for λ. ① Clearly m of an estimator of λ is $\frac{\sum Y_i}{n}$ i.e. $\hat{\lambda}$ ① 2

(vi) Central limit theorem: for n sufficiently large, $\bar{Y} = \frac{\sum Y_i}{n} \dot{\sim} N(E(Y), \frac{\text{var}(Y)}{n})$

i.e. $N(\lambda, \frac{\lambda}{n})$ ① $\therefore$ Approximate 95% confidence interval for λ is

$\hat{\lambda} \pm 1.96\sqrt{\frac{\hat{\lambda}}{n}}$ (since 1.96 is the 97.5%-pt of $N(0,1)$) ① 2

MC5 (2015)

3. (i) $\frac{dm}{dt} = 4\alpha^{\frac{1}{2}}(1-t\alpha^{\frac{1}{2}})^{-5}$ ①; $\frac{d^2m}{dt^2} = 20\alpha(1-t\alpha^{\frac{1}{2}})^{-6}$ ①

$\frac{d^3m}{dt^3} = 120\alpha^{\frac{3}{2}}(1-t\alpha^{\frac{1}{2}})^{-7}$ ①; $\frac{d^4m}{dt^4} = 840\alpha^2(1-t\alpha^{\frac{1}{2}})^{-8}$ ①

$E(X) = \frac{dm}{dt}|_{t=0} = 4\alpha^{\frac{1}{2}}$ ①

$E(X^2) = \frac{d^2m}{dt^2}|_{t=0} = 20\alpha$ ①

$E(X^3) = \frac{d^3m}{dt^3}|_{t=0} = 120\alpha^{3/2}$ ①

$E(X^4) = \frac{d^4m}{dt^4}|_{t=0} = 840\alpha^2$ ① 8

(ii) $E(\hat{\alpha}) = \frac{\sum E(X_i^2)}{20n} = \frac{20n\alpha}{20n} = \alpha \ \forall \alpha$ ①

$\therefore \hat{\alpha}$ is unbiased for α. ① 2

(iii) $var(X_i^2) = E(X_i^4) - (E(X_i^2))^2$ ① $= 840\alpha^2 - 400\alpha^2 = 440\alpha^2$ ①

$\therefore var(\hat{\alpha}) = \frac{1}{400n^2} \times 440n\alpha^2 = \frac{1.1\alpha^2}{n}$. ① 3

(iv) Likelihood, $L(\alpha) = \prod_{i=1}^{n} \frac{X_i^3 e^{-X_i/\sqrt{\alpha}}}{6\alpha^2} = \frac{\pi(X_i^3)\, e^{-\frac{\sum X_i}{\sqrt{\alpha}}}}{(6\alpha^2)^n}$ ①

Log likelihood, $\mathcal{L}(\alpha) = -2n\log(\alpha) - \alpha^{-\frac{1}{2}}\sum X_i$ + constant ①

$\frac{\partial \mathcal{L}}{\partial \alpha} = -\frac{2n}{\alpha} + \frac{1}{2}\alpha^{-\frac{3}{2}}\sum X_i$ ①

$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \frac{2n}{\alpha^2} - \frac{3}{4}\alpha^{-\frac{5}{2}}\sum X_i$ ①

$E\left(-\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right) = -\frac{2n}{\alpha^2} + \frac{3}{4}\alpha^{-\frac{5}{2}} \times 4n\alpha^{\frac{1}{2}}$

$= \alpha^{-2} n(-2+3) = \alpha^{-2} n$ ①

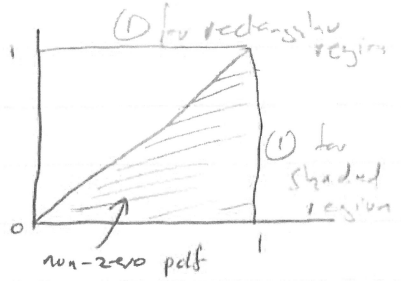
$\therefore$ Lower bound on variance $= \alpha^2/n$ ①

$\therefore$ Efficiency of $\hat{\alpha} = \frac{\alpha^2/n}{\frac{1.1\alpha^2}{n}} = \frac{1}{1.1} = \frac{10}{11} \simeq 0.9091$. ① 7

HC 5 (2015)

4.

(i) $f(x,y) = f(y|x)f(x) = 2$ (1) (1) $0 < y < x < 1$



4

(ii) $f(y) = \int_{x=y}^{1} f(x,y)\,dx$ (1) $= \int_{x=y}^{1} 2\,dx$

$= 2(1-y)$ (1) $\quad 0 < y < 1$

$E(Y) = \int_0^1 y f(y)\,dy$ (1) $= \int_0^1 (2y - 2y^2)\,dy$ (1)

$= \left[y^2 - \frac{2}{3}y^3\right]_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$ (1)

5

(iii) $f(x|y) = \frac{f(x,y)}{f(y)}$ (1) $= \frac{2}{2(1-y)} = \frac{1}{1-y}$ (1) $\quad 0 < y < x < 1$

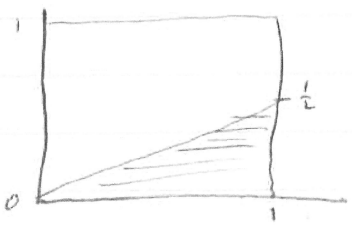
$E(X|Y=\tfrac{1}{2}) = \int_{1/2}^{1} x\, f(x)_{y=\frac{1}{2}}\,dx$ (1) $= \int_{\frac{1}{2}}^{1} \frac{x}{1-\frac{1}{2}}\,dx$ (1)

$= \left[x^2\right]_{\frac{1}{2}}^{1}$ (1) $= 1 - \frac{1}{4} = \frac{3}{4}$ (1)

6

[Alter: note that $X|y \sim U(y,1)$ and deduce expectation from this.]

(iv) $Y < \frac{1}{2}X$ corresponds to the shaded region



$\therefore P(Y < \frac{1}{2}X) = \int_{x=0}^{1} \int_{y=0}^{x/2} f(x,y)\,dy\,dx$ (1)

$= \int_0^1 \int_0^{\frac{x}{2}} 2\,dy\,dx$ (1)

$= \int_0^1 \left[2y\right]_0^{\frac{x}{2}} dx$ (1) $= \int_0^1 x\,dx$ (1) $= \left[\frac{1}{2}x^2\right]_0^1$

5

$= \frac{1}{2}$. (1)

[Alter: note that the required probability = 2x shaded area and use a geometric argument.]