HONG KONG STATISTICAL SOCIETY

2016 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.

1(i) $P(X > k) = P(X = k+1) + P(X = k+2) + P(X = k+3) + \ldots$ 1 (method)

$= \theta(1-\theta)^k + \theta(1-\theta)^{k+1} + \theta(1-\theta)^{k+2} + \ldots$ 1 (terms substituted)

$= \theta(1-\theta)^k (1 + (1-\theta) + (1-\theta)^2 + \ldots)$ 1 (common factor)

Final term is sum to infinity of a GP with first term 1 and common ratio $1-\theta$ where $0 < 1-\theta < 1$.

So $P(X > k) = \dfrac{\theta(1-\theta)^k}{(1-(1-\theta))} = (1-\theta)^k$ as required. 1 (correct use of GP)

(Full marks also for correct solution by induction) TOTAL 4

(ii) $P(X = 1) = \theta(1-\theta)^0 = \theta$ and this is observed for 24 players. 1 (correct term for 24)

$P(X = 2) = \theta(1-\theta)^1$ and this is observed for 48 players. 1 (correct term for 48)

$P(X > 2) = (1-\theta)^2$ and this is observed for the other 128 players.

1 (correct term for 128)

So the likelihood for the data is given by

$$L(\theta) = \theta^{24}\left[\theta(1-\theta)\right]^{48}\left[(1-\theta)^2\right]^{128}$$ 1 (combining terms)

$$= \theta^{72}(1-\theta)^{304}$$ as required.

$l(\theta) = \log L(\theta) = 72\log\theta + 304\log(1-\theta).$ 1 (log likelihood)

$\dfrac{dl(\theta)}{d\theta} = \dfrac{72}{\theta} - \dfrac{304}{1-\theta}$ and setting this equal to zero we have 1 (correct derivative), 1 (set to 0)

$72(1-\hat{\theta}) = 304\hat{\theta} \Rightarrow \hat{\theta} = \dfrac{72}{376} = 0.1915.$ 1 (0.1915)

To check that this gives a maximum we differentiate again:

$\dfrac{d^2 l(\theta)}{d\theta^2} = -\dfrac{72}{\theta^2} - \dfrac{304}{(1-\theta)^2}$ which is negative. 1 (negative 2nd deriv) TOTAL 9

(iii) The standard error of $\hat{\theta}$ can be estimated by $\sqrt{\dfrac{1}{-d^2 l(\theta)/d\theta^2}}$ where the denominator is evaluated at $\hat{\theta}$. This gives $\sqrt{\dfrac{1}{\dfrac{72}{0.1915^2} + \dfrac{304}{0.8085^2}}} = \sqrt{0.0004118} = 0.0203$

1 (method), 1 (substitute), 1(0.0203 or equiv. variance)

[If candidates work out the expected information, this gives a variance of $\dfrac{\theta(1-\theta)}{n(2-\theta)}$ where $n = 200$. Substituting $\hat{\theta}$ into this gives a variance of 0.0004280. All 3 marks above, and appropriate follow-on marks below, should be awarded.]

Then the approximate 95% confidence interval is given by $\hat{\theta} \pm 1.96 s.e(\hat{\theta})$ i.e. 1 (1.96)

1 (method)

$0.1915 \pm 1.96 \times 0.0203$ which is (0.1517, 0.2313). 1(0.1517), 1(0.2313) TOTAL 7

2. (a) (i) $R(t)=\log M_X(t) \Rightarrow R'(t)=\frac{M_X'(t)}{M_X(t)}$. 1 ($1^{st}$ deriv correct)

Then using the quotient rule, $R''(t)=\frac{M_X(t)M_X''(t)-\left(M'(t)\right)^2}{\left(M_X(t)\right)^2}$. 1 ($2^{nd}$ deriv correct)

TOTAL 2

(ii) We have that $M_X'(0)=E(X)$ and $M_X''(0)=E(X^2)$. 1 (using zero to find expectations)

Also by definition, $M_X(0)=E(e^0)=E(1)=1$. 1 (MX(0)=1)

So $R'(0)=\frac{E(X)}{1}=E(X)$. Also $R''(0)=\frac{1\times E(X^2)-(E(X))^2}{1^2}=E(X^2)-(E(X))^2=\text{Var}(X)$.

1 (1st deriv shown correct), 1 (2^{nd} deriv shown correct)

TOTAL 4

(b) (i) $M_Y(t)=E(e^{tY})=\sum_{y=0}^{\infty}e^{ty}\frac{\lambda^y e^{-\lambda}}{y!}=e^{-\lambda}\sum_{y=0}^{\infty}\frac{(\lambda e^t)^y}{y!}=e^{-\lambda}.e^{\lambda e^t} = e^{\lambda(e^t-1)}$.

1 (definition as expectation), 1 (integration step),
1 (recognition of sum)

TOTAL 3

(ii) $M_Y'(t)=\lambda e^t e^{\lambda(e^t-1)} \Rightarrow E(Y)=\lambda e^0 e^0=\lambda$. 1 ($1^{st}$ deriv correct), 1 (correct substitution)

$E\left[(Y-E(Y))^3\right]=E\left[(Y-\lambda)^3\right]=E(Y^3)-3\lambda E(Y^2)+3\lambda^2 E(Y)-\lambda^3$.

1 (substitute for E(Y)), 1 (expansion)

Using $E(Y^2), E(Y^3)$ given in the question, we have

$E\left[(Y-E(Y))^3\right]=\lambda+3\lambda^2+\lambda^3-3\lambda(\lambda+\lambda^2)+3\lambda^2.\lambda-\lambda^3$

$= \lambda+3\lambda^2+\lambda^3-3\lambda^2-3\lambda^3+3\lambda^3-\lambda^3=\lambda$ as required.

1 (substituted and simplified)

TOTAL 5

(iii) Mgf of sum of independent random variables is the product of their mgfs.

1 (mention independence),1 (state result)

Here they are identically distributed, so we raise to the *n*th power. 1 (*n*th power required)

$M_S(t)=\left[M_Y(t)\right]^n=e^{n\lambda(e^t-1)}$ 1 (power correctly applied)

This can be identified as the mgf of Poisson $(n\lambda)$. 1 (Poisson identified)

Since there is a unique one-to-one relationship between a distribution and its mgf, *S* follows a Poisson $(n\lambda)$ distribution. 1 (state uniqueness)

TOTAL 6

3. (i) Probabilities sum to 1, so 1 (sum to 1)

$$k(4+18+12+1+12+24+6+4+3)=84k=1\Rightarrow k=\frac{1}{84}.$$ 1 (1/84)

TOTAL 2

(ii) Sample size is only 3, so there cannot be 2 white dice and 2 blue dice. 1 (correct reason)

There are $\binom{9}{3}=84$ ways of drawing the sample altogether. If $X=2$ and $Y=0$ then there are 2 white dice and 1 red die drawn. The 2 white dice can be drawn in $\begin{bmatrix}3\\2\end{bmatrix}=3$ ways and the 1 red die in $\binom{4}{1}=4$ ways, so the probability is $\frac{3\times4}{84}=\frac{12}{84}$ as given.

1 (84 ways), 1 (3 ways), 1 (4 ways), 1 (probability calculation)

TOTAL 5

(iii) $P(Y=0)=\frac{4+18+12+1}{84}=\frac{35}{84}$. So the conditional distribution of X given that $Y=0$ is given by $P(X=x|Y=0)=\frac{P(X=x,Y=0)}{P(Y=0)}$ i.e. 1 (35/84), 1 (method)

x	0	1	2	3
P(X=x|Y=0)	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$

1 (pmf correct)

Then $E(X|Y=0)=\frac{18+24+3}{35}=\frac{45}{35}=\frac{9}{7}$. 1 (9/7)

$E(X^2|Y=0)=\frac{18+48+9}{35}=\frac{75}{35}=\frac{15}{7}$ so $\mathrm{Var}(X|Y=0)=\frac{15}{7}-\left(\frac{9}{7}\right)^2=\frac{24}{49}$. 1(15/7), 1(24/49)

TOTAL 6

(iv) X marginal distribution:

0	1	2	3
$\frac{20}{84}$	$\frac{45}{84}$	$\frac{18}{84}$	$\frac{1}{84}$

So $E(X)=\frac{45+36+3}{84}=1.$

1 (X marginal correct), 1(E(X)=1)

Y marginal distribution:

0	1	2
$\frac{35}{84}$	$\frac{42}{84}$	$\frac{7}{84}$

So $E(Y)=\frac{42+14}{84}=\frac{56}{84}$.

1 (Y marginal correct), 1(E(Y)=56/84)

$$E(XY)=\frac{24}{84}\times1\times1+\frac{6}{84}\times1\times2+\frac{3}{84}\times2\times1=\frac{42}{84}.$$ 1(E(XY)=42/84)

So $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{42}{84}-\frac{56}{84}=-\frac{14}{84}=-\frac{1}{6}$. 1 (Cov = -1/6)

Negative covariance makes sense: the more white dice in the sample, the fewer spaces there are for blue dice. 1 (valid comment)

TOTAL 7

4. (i) Since we know that $\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$, then since $E(\sum_{i=1}^{n}X_i^2)=\sum_{i=1}^{n}E(X_i^2)$,

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2).$$

1 (alternative form), 1(written as expectations)

Now $\sigma^2=E(X_i^2)-\mu^2\Rightarrow E(X_i^2)=\sigma^2+\mu^2$ and so $\sum_{i=1}^{n}E(X_i^2)=n(\sigma^2+\mu^2)$.

1 (E(X^2)), 1 (multiply by *n*)

$\text{Var}(\bar{X})=\frac{\sigma^2}{n}=E(\bar{X}^2)-\mu^2\Rightarrow E(\bar{X}^2)=\frac{\sigma^2}{n}+\mu^2.$ 1 (var of mean), 1 (final expression)

We require $\sum_{i=1}^{n}X_i^2-nE(\bar{X}^2)=n(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right)=(n-1)\sigma^2.$ So

$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{1}{n}(n-1)\sigma^2$ as required. 1 (correctly combined)

For an unbiased estimator, we require the expectation to equal σ^2. From the previous line, $\frac{n}{n-1}E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{n}{(n-1)}.\frac{(n-1)}{n}\sigma^2=\sigma^2$, so $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ is an unbiased estimator for σ^2.

1 (require expectation = σ^2), 1 (final estimator correct)
TOTAL 9

(ii) $E(S_a^2)=E\left[a\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=aE\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=a(n-1)\sigma^2.$

1 (constant out of sum), 1 (correct final expression)

So $Bias(S_a^2)=a(n-1)\sigma^2-\sigma^2=(an-a-1)\sigma^2.$ 1 (method), 1 (bias correct)

TOTAL 4

(iii) $\text{Var}(S_a^2)=\text{Var}\left[a\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=a^2\text{Var}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=2a^2(n-1)\sigma^4.$

1 (constant squared)

So $MSE(S_a^2)=2a^2(n-1)\sigma^4+(an-a-1)^2\sigma^4.$ 1 (MSE correct)

For max/min, set $\frac{dM}{da}$ equal to zero: 1 (method)

$$\frac{dM}{da}=4a(n-1)\sigma^4+2(an-a-1)(n-1)\sigma^4=0.$$ 1 (deriv correct)

Dividing through by $2(n-1)\sigma^4$: $2a+an-a-1=0\Rightarrow a(n+1)=1\Rightarrow a=\frac{1}{n+1}.$ 1(1/(n+1))

To show that this gives a minimum value, look at the second derivative:

$\frac{d^2M}{da^2}=4(n-1)\sigma^4+2(n-1)^2\sigma^4$, which is clearly positive, so the required value of a is $\frac{1}{n+1}$.

1 (2nd deriv correct), 1 (positive)

TOTAL 7