HONG KONG STATISTICAL SOCIETY

2015 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA – MODULE 1

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.

Question 1

(a)	(i) $P(A) = \sum_i P(A \mid E_i) P(E_i)$	**1**
	(ii) $P(A \mid E_j) = \dfrac{P(A \cap E_j)}{P(E_j)} \Rightarrow P(A \cap E_j) = P(A \mid E_j).P(E_j)$	**1**
	$P(E_j \mid A) = \dfrac{P(E_j \cap A)}{P(A)} \Rightarrow P(A \cap E_j) = P(E_j \mid A).P(A)$	**1**
	So $P(E_j \mid A).P(A) = P(A \mid E_j).P(E_j)$	**1**
	$\Rightarrow$ $P(E_j \mid A) = \dfrac{P(A \mid E_j).P(E_j)}{P(A)}$	**1**
	Using LTP, this means that	
	$P(E_j \mid A) = \dfrac{P(A \mid E_j).P(E_j)}{\sum_i P(A \mid E_i).P(E_i)}$	**1**
(b)	(i) Let E_1 = 'foetus has trisomy 21', E_2 = 'foetus does not have trisomy 21'.	
	Then E_1 and E_2 form a partition, with $P(E_1) = 0.001$, $P(E_2) = 0.999$.	**1**
	Let A = 'foetus tests positive'. Then	
	$P(A) = P(A \mid E_1).P(E_1) + P(A \mid E_2).P(E_2)$	**1**
	$= (0.90 \text{ x } 0.001) + (0.05 \text{ x } 0.999)$	**1**
	$= 0.05085$	**1**
	(ii) $P(E_2 \mid A) = \dfrac{P(A \mid E_2).P(E_2)}{P(A)} = \dfrac{(0.05\text{x}0.999)}{0.05085} = 0.9823$	**1,1,1**
	(iii) X = no.of foetuses that do not have trisomy 21 ~ Bi(1,000, 0.9823)	**1**
	Approximately, X ~ N(982.3, 17.3867) *(1 mark for mean, 1 for variance)*	**1, 1**
	$P(X \geq 990) \approx P\left(Z \geq \dfrac{989.5 - 982.3}{\sqrt{17.3867}} \right)$	**1**
	$= P(Z \geq 1.73)$	**1**
	$= 1 - \Phi(1.73)$	**1**
	$= 0.0418$	**1**

Question 2

(i)	There are $\binom{25}{5}$ different ways to choose the committee; these are all equally-likely outcomes when the committee is selected at random. For x = 0, 1, …, 5 and y = 0, 1, …, 5–x, with $x+y \geq 1$, there are: $\binom{7}{x}$ different ways of choosing x Support staff $\binom{14}{y}$ different ways of choosing y Teaching staff and $\binom{4}{5-x-y}$ different ways of choosing $5 - x - y$ Management staff so $\binom{7}{x}\binom{14}{y}\binom{4}{5-x-y}$ different ways of choosing a social committee	**1** **1** **1**
(ii)	$$P(X=1, Y=3) = \frac{\binom{7}{1}\binom{14}{3}\binom{4}{1}}{\binom{25}{5}} = 0.1918$$ (*1 mark for identifying correct probability, 1 mark for correct binomial coefficients, 1 mark for correct answer*)	**1, 1, 1**
(iii)	Selection of the committee is by random sampling without replacement from the finite population of 25 members of staff. For the marginal distribution of X, we need only consider the staff as 7 Support and 18 Others. Hence X has the hypergeometric distribution $$P(X=x) = \frac{\binom{7}{x}\binom{18}{5-x}}{\binom{25}{5}}$$	**1** **1** **1**
(iv)	Let Z be the number of Management staff on the committee. Z is also a hypergeometric random variable, and $$P(Z \geq 1) = 1 - P(Z=0) = 1 - \frac{\binom{4}{0}\binom{21}{5}}{\binom{25}{5}} = 1 - 0.3830 = 0.6170$$ (*1 mark each for first two steps and 1 mark for correct answer*) **… continued**	**1** **1, 1, 1**

Question 2

	Continued …	
(v)	$\mathbb{E}(X) = 1.4 \quad \text{Var}(X) = 0.84$	**1**
	$\mathbb{E}(Y) = 2.8 \quad \text{Var}(Y) = 1.027$	**1**
	$\mathbb{E}(Z) = 0.8 \quad \text{Var}(Z) = 0.56$	**1**
(vi)	The random variable $T = X + Y + Z$ takes the value 5 with probability 1. Therefore:	**1**
	$\mathbb{E}(T) = 5 \quad \text{Var}(T) = 0$	**1, 1**
	$\text{Var}(X+Y+Z) \neq \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)$ because the three random variables are not independent	**1**

Question 3

(i)	X = no. of dice on which this player's number comes up ~ Bi(3, $^1/_6$) W = player's winnings has the following probability distribution: $P(W=-1) = P(X=0) = (^5/_6)^3 = {}^{125}/_{216}$ $P(W=1) = P(X=1) = 3(^5/_6)^2(^1/_6) = {}^{75}/_{216}$ $P(W=2) = P(X=2) = 3(^5/_6)(^1/_6)^2 = {}^{15}/_{216}$ $P(W=3) = P(X=3) = (^1/_6)^3 = {}^{1}/_{216}$ So $\mathbb{E}(W) = \frac{-125+75+30+3}{216} = -\frac{17}{216}$	 **1** **1** **1** **1** **1, 1**
(ii)	$\mathbb{E}(W^2) = \frac{125+75+60+9}{216} = \frac{269}{216}$ $\text{Var}(W) = \mathbb{E}(W^2) - \{\mathbb{E}(W)\}^2 = \frac{269}{216} - \frac{17^2}{216^2} = 1.2392$	**1** **1**
(iii)	Let T denote the three friends' total winnings. If the random variables W_A, W_B and W_C are the winnings of A, B and C respectively, they are <u>independent</u> random variables so: $\mathbb{E}(T) = \mathbb{E}(W_A) + \mathbb{E}(W_B) + \mathbb{E}(W_C) = -\frac{51}{216} = -0.2361$ $\text{Var}(T) = \text{Var}(W_A) + \text{Var}(W_B) + \text{Var}(W_C) = 3.7176$	 **1** **1** **1**
(iv)	P(none of them wins) = $(^1/_2)^3 = {}^9/_{72}$ — $T = -3$ P(one wins on 1 die)= P(Aoo etc.) =$9(^1/_6)(^1/_2)^2 = {}^{27}/_{72}$ — $T = -1$ P(one wins on 2 dice) = P(AAo etc.) = $9(^1/_6)^2(^1/_2) = {}^9/_{72}$ — $T = 0$ P(one wins on 3 dice) = P(AAA etc.) = $3(^1/_6)^3 = {}^1/_{72}$ — $T = 1$ P(two win on 1 die each) = P(ABo etc.) = $18(^1/_6)^2(^1/_2) = {}^{18}/_{72}$ — $T = 1$ P(one wins on 2 dice, other on 1) = P(AAB etc.) = $18(^1/_6)^3 = {}^6/_{72}$ — $T = 2$ P(three win on 1 die each) = P(ABC) = $6(^1/_6)^3 = {}^2/_{72}$ — $T = 3$ (*Deduct ½ mark for each wrong probability or value of T or missing value of T.*) So $\mathbb{E}(T) = (-27-27+0+1+18+12+6)/72 = -17/72$ $\mathbb{E}(T^2) = (81+27+0+1+18+24+18)/72 = 169/72$ $\text{Var}(T) = \mathbb{E}(T^2) - \{\mathbb{E}(T)\}^2 = 2.291$ The expected value of T is not changed by the friends' strategy but the variance of T is reduced.	**1** **4** (lines 2–7) **1** **1** **1** **1**

Question 4

(a)	(i) For $y = 0, 1, 2, \ldots$ and $z = 0, 1, 2, \ldots$, $P(Y = y, Z = z) = P(Y = y, X = y+z)$ $= P(Y = y \mid X = y+z).P(X = y+z)$ Given that $X = y + z$, $Y \sim \text{Bi}(y + z, \theta)$. So $P(Y = y, Z = z) = \binom{y+z}{y}\theta^y(1-\theta)^z.\frac{\exp(-\lambda)\lambda^{y+z}}{(y+z)!}$ $= \frac{\exp(-\lambda)}{y!z!}\lambda^{y+z}\theta^y(1-\theta)^z$ (ii) For y = 0, 1, 2, …, the marginal distribution of Y is $P(Y = y) = \sum_{z=0}^{\infty} \text{P}(Y = y, Z = z)$ $= \frac{\exp(-\lambda)}{y!}(\lambda\theta)^y \sum_{z=0}^{\infty}\frac{\{\lambda(1-\theta)\}^z}{z!}$ $= \frac{\exp(-\lambda)}{y!}(\lambda\theta)^y \exp(\lambda(1-\theta))$ $= \frac{\exp(-\lambda\theta)}{y!}(\lambda\theta)^y$ i.e. Y has the Poisson distribution with parameter $\lambda\theta$. (iii) By the symmetry of the situation, Z has the Poisson distribution with parameter $\lambda(1 - \theta)$. [*This might be derived in a similar way as in (ii).*] So $P(Z = z) = \frac{\exp(-\lambda(1-\theta))}{z!}(\lambda(1-\theta))^z$ So $P(Y = y, Z = z) = P(Y = y) \text{ x } P(Z = z)$, for all (y, z) Therefore Y and Z are independent random variables. **Continued …**	 **1** **1** **1** **1, 1** **1** **1** **1** **1** **1** **1** **1** **1** **1**

Question 4

	… Continued	
(b)	For $t > 0$,	
	$P(T \le t) = 1 - P(T > t)$	**1**
	$= 1 - P(0 \text{ arrivals in next } t \text{ minutes})$	**1**
	$= 1 - P(X_t = 0)$	**1**
	where X_t has the Poisson distribution with parameter λt	
	Therefore, T has cumulative distribution function	
	$F_T(t) = P(T \le t) = 1 - \exp(-\lambda t),\ t > 0$	**1**
	and probability density function	
	$f_T(t) = \frac{d}{dt} F_T(t) = \lambda \exp(-\lambda t), \quad t > 0$	**1**
	i.e. T has the exponential distribution with parameter λ	**1**

Question 5

(i)	$f(x) = 1$ and $F(x) = x$ (0 < x < 1) The p.d.f. of $X_{(j)}$ is $g(x) = \frac{n!}{(j-1)!(n-j)!}[F(x)]^{j-1}[1-F(x)]^{n-j} f(x)$ $= \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j}, \quad 0 < x < 1$ $\mathbb{E}(X_{(j)}) = \frac{n!}{(j-1)!(n-j)!}\int_0^1 x^j (1-x)^{n-j}\, dx$ $= \frac{n!}{(j-1)!(n-j)!} \frac{j!(n-j)!}{(n+1)!} = \frac{j}{n+1}$ $\mathbb{E}(X_j^2) = \frac{n!}{(j-1)!(n-j)!}\int_0^1 x^{j+1}(1-x)^{n-j}\, dx$ $= \frac{n!}{(j-1)!(n-j)!}\frac{(j+1)!(n-j)!}{(n+2)!} = \frac{j(j+1)}{(n+1)(n+2)}$ $\text{var}(X_{(j)}) = \frac{j(j+1)}{(n+1)(n+2)} - \left(\frac{j}{(n+1)}\right)^2 = \frac{j(n+1)-j^2}{(n+1)^2(n+2)}$ When n is odd, the sample median is $X_{((n+1)/2)}$. Using the results above, $\mathbb{E}(X_{((n+1)/2)}) = \frac{(n+1)/2}{n+1} = \frac{1}{2}$ $\text{var}(X_{((n+1)/2)}) = \frac{(n+1)^2/2-(n+1)^2/4}{(n+1)^2(n+2)} = \frac{1}{4(n+2)}$	**1** **1** **1** **1** **1** **1** **1** **1** **1**
(ii)	The joint p.d.f. of $X_{(j)}$ and $X_{(k)}$ is $g(x_j, x_k) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}[F(x_j)]^{j-1}[F(x_k)-F(x_j)]^{k-j-1}$ X $[1-F(x_k)]^{n-k} f(x_j) f(x_k)$ $= \frac{n!}{(j-1)!(k-j-1)!(n-k)!} x_j^{\,j-1}(x_k - x_j)^{k-j-1}(1-x_k)^{n-k},$ $0 < x_j < x_k < 1$ **Continued …**	**1** **1**

Question 5

(ii)	**… Continued**	
	$\mathbb{E}\left(X_{(j)}X_{(k)}\right) \propto \int_0^1 \int_0^{x_k} x_j^{\,j}\left(x_k - x_j\right)^{k-j-1} x_k \left(1-x_k\right)^{n-k}\, dx_j\, dx_k$	**1**
	$= \int_0^1 x_k \left(1-x_k\right)^{n-k} \int_0^{x_k} x_j^{\,j}\left(x_k - x_j\right)^{k-j-1} dx_j\, dx_k$	
	$= \dfrac{j!(k-j-1)!}{k!} \int_0^1 x_k^{\,k+1}\left(1-x_k\right)^{n-k} dx_k$	**1**
	$= \dfrac{j!(k-j-1)!}{k!} \dfrac{(k+1)!(n-k)!}{(n+2)!}$	**1**
	$\mathbb{E}\left(X_{(j)}X_{(k)}\right) = \dfrac{n!}{(j-1)!(k-j-1)!(n-k)!} \dfrac{j!(k-j-1)!}{k!} \dfrac{(k+1)!(n-k)!}{(n+2)!}$	**1**
	$= \dfrac{j(k+1)}{(n+1)(n+2)}$	**1**
(iii)	When n is even, the sample median is $M = \frac{1}{2}\left(X_{(n/2)} + X_{((n/2)+1)}\right)$. Using the results above,	
	$\mathbb{E}(M) = \frac{1}{2}\left\{\mathbb{E}\left(X_{(n/2)}\right) + \mathbb{E}\left(X_{((n/2)+1)}\right)\right\} = \frac{1}{2}\left\{\dfrac{n/2}{n+1} + \dfrac{(n/2)+1}{n+1}\right\} = \frac{1}{2}$	**1**
	$\mathrm{cov}\left(X_{(n/2)}, X_{((n/2)+1)}\right) = \mathbb{E}\left(X_{(n/2)}X_{((n/2)+1)}\right) - \mathbb{E}\left(X_{(n/2)}\right)\mathbb{E}\left(X_{((n/2)+1)}\right)$	
	$= \dfrac{n^2}{4(n+1)^2(n+2)}$	**1**
	$\mathrm{var}(M) = \frac{1}{4}\left\{\mathrm{var}\left(X_{(n/2)}\right) + \mathrm{var}\left(X_{((n/2)+1)}\right) + 2\,\mathrm{cov}\left(X_{(n/2)}, X_{((n/2)+1)}\right)\right\}$	**1**
	$= \dfrac{n}{4(n+1)(n+2)}$	**1**

Question 6

(i)	Since X and Y are independent, $f(x,y) = f_X(x).f_Y(y) = 1, \quad 0 \le x \le 1,\ 0 \le y \le 1$	**1**
	Now $\quad U = (-2\ln X)^{1/2}.\sin(2\pi Y) \qquad V = (-2\ln X)^{1/2}.\cos(2\pi Y)$	
	so $\quad U^2 + V^2 = (-2\ln X), \qquad \dfrac{U}{V} = \tan(2\pi Y)$	**1, 1**
	and $\quad X = \exp[-½(U^2 + V^2)], \qquad Y = \dfrac{1}{2\pi}\tan^{-1}\left(\dfrac{U}{V}\right)$	**1, 1**
	$J = \begin{vmatrix} \dfrac{\partial X}{\partial U} & \dfrac{\partial X}{\partial V} \\ \dfrac{\partial Y}{\partial U} & \dfrac{\partial Y}{\partial V} \end{vmatrix}$ $= \begin{vmatrix} -U.\exp\left[-\dfrac{1}{2}(U^2+V^2)\right] & -V.\exp\left[-\dfrac{1}{2}(U^2+V^2)\right] \\ \dfrac{1}{2\pi}.\dfrac{1}{1+\left(\dfrac{U}{V}\right)^2}.\dfrac{1}{V} & \dfrac{1}{2\pi}.\dfrac{1}{1+\left(\dfrac{U}{V}\right)^2}.\dfrac{-U}{V^2} \end{vmatrix}$	**4@1**
	$= \dfrac{1}{2\pi}\exp\left[-\dfrac{1}{2}(U^2+V^2)\right]$	**1**
	Hence $f(u,v) = \dfrac{1}{2\pi}\exp\left[-\dfrac{1}{2}(u^2+v^2)\right], \qquad -\infty < u,v < \infty$	**1**
(ii)	The joint p.d.f. of U and V factorises. Using the Factorisation Theorem, therefore, U and V are independent. Also	**1**
	$f(u) \propto \exp\left[-\dfrac{1}{2}u^2\right], \qquad -\infty < u < \infty$	**1**
	so $\quad f(u) = \dfrac{1}{\sqrt{2\pi}}\exp\left[-\dfrac{1}{2}u^2\right], \qquad -\infty < u < \infty$	**1**
	i.e. $\quad U \sim \mathrm{N}(0, 1)$	**1**
	Similarly, $V \sim \mathrm{N}(0, 1)$.	**1**

Question 6

	… Continued	
(iii)	Using any method of generating Uniform random variates, generate a pair of (pseudo-independent) random numbers between 0 and 1. These are x and y in the above theory. Now form u and v, as described in part (i). Then these are (pseudo-independent) random variates from the standard normal distribution.	**1** **1**
	Also, U^2 and V^2 are independent $\chi^2(1)$ random variables, so $U^2 + V^2$ is a $\chi^2(2)$ random variable. So, in order to generate a random variate from this distribution, proceed as above and then form $u^2 + v^2$.	**1** **1**

Question 7

(i)	$M_X(t) = \mathbb{E}(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx$	**1, 1**
	Now $xt - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 = -\frac{1}{2\sigma^2}\left\{x^2 - 2x\mu + \mu^2 - 2\sigma^2 xt\right\}$	**1**
	$= -\frac{1}{2\sigma^2}\left\{\left[x - \left(\mu + \sigma^2 t\right)\right]^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right\}$	
	$= -\frac{1}{2\sigma^2}\left\{\left[x - \left(\mu + \sigma^2 t\right)\right]^2\right\} + \mu t + \frac{1}{2}\sigma^2 t^2$	**1**
	So $M_X(t) = \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right)\underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\left[\mu+\sigma^2 t\right]}{\sigma}\right)^2\right\}dx}_{=1}$	**1** **1**
	$= \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right)$	**1**
(ii)	$M_{aX+b}(t) = \mathbb{E}(e^{t(aX+b)}) = e^{tb}\,\mathbb{E}(e^{taX}) = e^{tb}M_X(at)$	**1, 1**
	For Z, apply this result when $a = 1/\sigma$ and $b = -\mu/\sigma$, to obtain	**1**
	$M_Z(t) = e^{-\mu t/\sigma}M_X(t/\sigma) = \exp\left(-\frac{\mu t}{\sigma} + \frac{\mu t}{\sigma} + \frac{\sigma^2 t^2}{2\sigma^2}\right) = \exp\left(t^2/2\right)$	**1**
	By part (i), this is the m.g.f. of the N(0, 1) distribution. Hence, by the uniqueness property of m.g.f.'s, $Z \sim$ N(0, 1).	**1**
(iii)	$K(X) = \frac{E\left(\left[X-\mu\right]^4\right)}{\sigma^4} - 3 = E\left(\left[\frac{X-\mu}{\sigma}\right]^4\right) - 3 = E(Z^4) - 3$	**1, 1**
	$M'_Z(t) = t.\exp\left(t^2/2\right)$	**1**
	$M''_Z(t) = \exp\left(t^2/2\right) + t^2\exp\left(t^2/2\right) = (1+t^2)\exp\left(t^2/2\right)$	**1**
	$M'''_Z(t) = 2t.\exp\left(t^2/2\right) + (1+t^2).t.\exp\left(t^2/2\right) = (3t+t^3)\exp\left(t^2/2\right)$	**1**
	$M''''_Z(t) = (3+3t^2)\exp\left(t^2/2\right) + (3t+t^3).t.\exp\left(t^2/2\right)$	**1**
	$\mathbb{E}\left(Z^4\right) = M''''_Z(0) = 3 \Rightarrow K(X) = 0$	**1, 1**

Question 8

(a)	Since U has the uniform distribution on the interval 0 to 1, then its cumulative distribution function is $P(U \le u) = u$ for $0 \le u \le 1$.	**1**
	Therefore, the cumulative distribution function of Y is	
	$G(y) = P(Y \le y)$	
	$= P(F^{-1}(U) \le y)$	**1**
	$= P(U \le F(y))$	**1**
	$= F(y)$	**1**
	Since Y and X have the same cumulative distribution function, they must have the same distribution.	**1**
(b)	Denote the cumulative distribution function of the target probability distribution by F.Suppose we have a method for obtaining pseudo-random variates from the uniform distribution on the interval 0 to 1; denote these variates by $u_1, u_2, u_3, \ldots$. Let $x_i = F^{-1}(u_i)$. The theory above shows that $x_1, x_2, x_3, \ldots$ are pseudo-random variates from the target distribution.	**1** **1** **1**
(c)	(i) For this distribution, $F(x) = 4x^3 - 3x^4$ $(0 \le x \le 1)$.	**1**
	Using the inverse c.d.f. method would mean solving for x the equation	
	$4x^3 - 3x^4 = u$	
	This cannot be done in closed form, so would have to be done numerically for every different value of u which is computationally inefficient.	**1**
	(ii) $f(x) = 12x^2(1-x) \Rightarrow f'(x) = 24x - 36x^2 \Rightarrow f'(x) = 0$ when $x = {}^2/_3$	**1**
	$f''(x) = 24 - 72x \Rightarrow f''({}^2/_3) = -24 < 0$	**1**
	$\therefore f(x)$ has a local maximum at $x = 2/3$. The maximum value is ${}^{16}/_9$.	**1**
	STEP 1 – Generate two pseudo-random variates from the uniform distribution on the interval 0 to 1; denote these by u_1 and u_2.	**1**
	STEP 2 – Consider the point $(u_1, {}^{48}/_{27}u_2)$. If this point lies below the curve $y = f(x)$, that is if ${}^{48}/_{27}u_2 < 12u_1^2(1-u_1)$, then accept u_1 as a pseudo-random variate from $f(x)$. Otherwise, reject this value and begin at STEP 1 again.	**1** **1, 1**
	(iii) The point $(U_1, {}^{48}/_{27}U_2)$ is uniformly distributed on the rectangle (0, 1) x (0, ${}^{48}/_{27}$), which has area ${}^{48}/_{27}$. The area under the curve $y = f(x)$ is 1. So there is probability ${}^{27}/_{48}$ of accepting a simulation, independently of all other simulations. Therefore, on average, need 2 x ${}^{48}/_{27} = 3.56$ uniform random variates (*1 mark for the value of ${}^{48}/_{27}$, 1 mark for the factor of 2*) to generate one variate from the target distribution.	**1** **1, 1**