Statistics 2011-2012

1 i) Since the random variables follow the same distribution, the CLT gives an approximate distribution as:

$X \sim N(n\mu, n\sigma^2)$ <u>e</u>

ii)

Observed (O)

	GR	SK	
East	12	28	40
West	24	36	60
	36	64	100

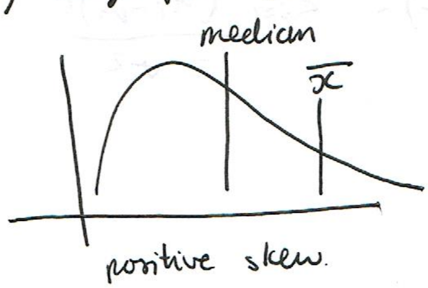
Expected (E)

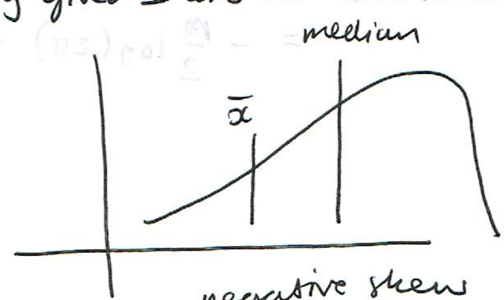
	GR	SK	
East	$\frac{36\times40}{100} = 14.4$	$\frac{64\times40}{100} = 25.6$	40
West	$\frac{36\times60}{100} = 21.6$	$\frac{64\times60}{100} = 38.4$	60
	36	64	100

		$O-E_i$	$(O-E_i)^2/E_i$
East	GR	-2.4	0.4
East	SK	2.4	0.225
West	GR	2.4	0.267
West	SK	-2.4	0.149
		Σ	1.04

$\therefore\ \chi^2 = 1.04$ <u>b</u>

iii) Says pair of statements but only gives 1 ans in mark scheme??





$\Rightarrow$ a & $\boxed{c}$? Log transform does reduce skewness.

iv) $1 = \int_1^b x \, dx = \left[\frac{x^2}{2} \right]_1^b$

$$1 = \frac{b^2}{2} - \frac{1}{2}$$

$$2 = b^2 - 1$$

$$3 = b^2$$

$$\pm\sqrt{3} = b$$

$\underline{c}$

v) RBG $= 3/4 \times \frac{3}{8} \times \frac{3}{7}$

RGB $= 3/4 \times 3/8 \times 3/7$

BRG = —— " ——

BGR = —— " ——

GRB = —— " ——

GBR = —— " ——

$\therefore \; 6 \times \frac{3}{4} \times \frac{3}{8} \times \frac{3}{7} = \frac{9}{28}$

$\underline{f}$

~~4~~

2 i) $l(\theta : x_1, x_2, \ldots, x_n) = f(x_1|\theta) f(x_2|\theta) \ldots f(x_n|\theta)$

$l(\mu, \sigma : x_1, x_2, \ldots, x_n) = f(x_1 | \mu, \sigma) f(x_2|\mu, \sigma) \ldots f(x_n|\mu, \sigma)$

$$l(\mu, \sigma) = \sum_{i=1}^{n} f(x_i | \mu, \sigma)$$

$$= \frac{1}{(\sqrt{2\pi})^n (\sigma)^n} e^{-\frac{1}{2}\left(\frac{\sum x_i - \mu}{\sigma}\right)^2}$$

n times.

from definition of normal distribution.

$$= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2}$$

ii) $\frac{\partial}{\partial \mu} l(\mu, \sigma) = 0$

$$\Rightarrow \quad 0 = \frac{\partial}{\partial \mu} l(\mu, \sigma) = \frac{-2 \times -1 \times \sum_{i=1}^{n} (x_i - \mu)^1}{2\sigma^2} .$$

$$0 = \frac{\sum_{i=1}^{n} (x_i - \mu)}{\sigma^2}$$

$$0 = \frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma^2}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$ the mean

$$\hat{\mu} = \bar{x}$$

checking Min/Max through second derivative

$$\frac{\partial^2}{\partial \mu^2} l(\mu, \sigma) = -\frac{n}{\sigma^2}$$

$\frac{-n}{\sigma^2} < 0$ for all values $\therefore$ $\hat{\mu}$ is a MLE (max ...)

if 2^{nd} derivative > 0 then (minimum LE)

iii) $$0 = \frac{\partial}{\partial \sigma} l(\mu = \hat{\mu}, \sigma) = -\frac{n}{\sigma} - \frac{-2 \times \sum_{i}^{n} (x_i - \mu)^2}{2\sigma^3}$$

$$0 = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^3}$$

$$\frac{n}{\sigma} = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^3}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

iv) Since $E(s^2_{n-1}) = \sigma^2$, by linearity of expectations

$$E(\hat{\sigma}^2) = \frac{n-1}{n} E(s^2_{n-1}) = \frac{n-1}{n} \sigma^2$$

$$\text{bias}(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

$\hat{\sigma}^2$ is a biased estimator since a bias exists

8 i a) $X \sim B(25, 0.2)$

$E(x) = np = 25 \times 0.2 = 5$

$Var(x) = npq = 25 \times 0.2 \times 0.8 = 4$

b) $p(X=5) = C^{25}_{5}\, 0.2^{5}\, 0.8^{20} = 0.196$

c) $p(X \geq 6) = 1 - p(X \leq 5)$

$= 1 - [\, p(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$
$+ P(X=5))$

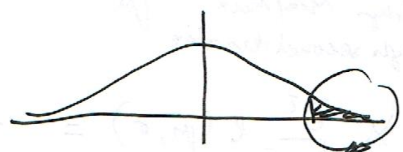
$= 1 - \left[C^{25}_{0}\, 0.2^{0}\, 0.8^{25} + C^{25}_{1}\, 0.2^{1}\, 0.8^{24} + C^{25}_{2}\, 0.2^{2}\, 0.8^{23} \right.$
$\left. + C^{25}_{3}\, 0.2^{3}\, 0.8^{22} + C^{25}_{4}\, 0.2^{4}\, 0.8^{21} + C^{25}_{5}\, 0.2^{5}\, 0.8^{20} \right]$

$= 1 - 0.6167$

$= 0.383$

ii a) $H_0: \mu = 0.2$
$H_1: \mu > 0.2$

one sided test →



want this part of the graph to be as close to 5% as possible. (α)

$P(X \geq 8) = 0.109$

$P(X \geq 9) = 0.047$ ← this probability is the closest in the rejection region.

∴ rejection region is $P(X \geq 9)$

$R = \{x \mid P(X \geq x) < \alpha\}$
$= \{9, 10, \ldots 25\}$

b) $6 \notin R$ so no significant evidence to reject null hypothesis

c) $p(X \geq 8) = 0.109$ ← closest number which is not in the rejection region.

↑

mark scheme is kinda wierd idk.

4 i) $1 = p_{XY}(x,y)$

$= p_{XY}(1,1) + p_{XY}(1,2) + p_{XY}(2,1) + p_{XY}(2,2)$

$= 2c + 3c + 3c + 4c$

$1 = 12c$

$\frac{1}{12} = c$

ii) $1 = \sum_x \sum_y p_{XY}(x,y)$

$= c \sum_{x=1}^{m_x} \sum_{y=1}^{m_y} (x+y)$

$= c \sum_{x=1}^{m_x} \left(m_y x + \frac{m_y(m_y+1)}{2} \right)$

$= c \left(m_y \frac{m_x(m_x+1)}{2} + \frac{m_x m_y(m_y+1)}{2} \right)$

$= \frac{c m_x m_y (m_x + m_y + 2)}{2}$

$c = \frac{2}{m_x m_y (m_x + m_y + 2)}$

iii) $p_X(x) = \sum_y p_{XY}(x,y)$

$= c \sum_{y=1}^{m_y} (x+y)$

$= c \left(m_y x + \frac{m_y(m_y+1)}{2} \right)$

$= \frac{2 m_y x + m_y(m_y+1)}{m_x m_y (m_x + m_y + 2)}$

$= \frac{2x + m_y + 1}{m_x(m_x + m_y + 2)}$

$P_x(y) = \sum_x p_{XY}(x,y)$

$= c \sum_{x=1}^{m_x} (x+y)$

$= c \left(m_x y + \frac{m_x(m_x+1)}{2} \right)$

$= \frac{2y + m_x + 1}{m_y(m_x + m_y + 2)}$

if $m_x = m_y = 2$

$p_X(x) = \frac{2x+3}{12}$

$P_Y(y) = \frac{(2y+3)}{12}.$

iv) $E(X) = E(Y) = \sum_x x\, p_X(x)$

		Y: 1	Y: 2	
X	1	2/12	3/12	5/12
	2	3/12	4/12	7/12
		5/12	7/12	1

$$E(X) = E(Y) = 1 \times 5/12 + 2 \times \frac{7}{12} = 1.582$$

v) $E(XY) = \sum_x \sum_y xy\, P_{XY}(x,y) = 1 \times 1 \times \frac{2}{12} + 1 \times 2 \times \frac{3}{12} + 2 \times 1 \times \frac{3}{12} + 2 \times 2 \times \frac{4}{12}$

$= 2.5$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

$= 2.5 - 1.582 \times 1.582$

$= -0.0069$ (negative value indicates that X & Y are not correlated).