Course **Comp245**

Question		Marks & seen/unseen
1.		
Parts		
(i)	$x_{(\{n+1\}/2)}=3$, $\bar{x}=4$, IQR$=x_{(6)}-x_{(2)}=5$, $s_n=2\sqrt{2}$. Answer is (c).	seen sim.
(ii)	Using Bayes Theorem, $$\mathrm{P}(\overline{B}\mid A)=\frac{\mathrm{P}(A\cap\overline{B})}{\mathrm{P}(A)}=\frac{\mathrm{P}(A\mid\overline{B})\mathrm{P}(\overline{B})}{\mathrm{P}(A)}=\frac{(1-\mathrm{P}(\overline{A}\mid\overline{B}))(1-\mathrm{P}(B))}{\mathrm{P}(A)} = \frac{(1-0.8)(1-0.3)}{0.2}=0.7.$$ Answer is (e).	seen sim.
(iii)	The probability of not observing a 3 is $(5/6)^3\approx 0.58$, so the probability of at least one 3 is $1-0.58=0.42$. Answer is (c).	seen sim.
(iv)	Let X be the number of draws until the black ball is found. Without replacement, X is equally likely to be any number between 1 and 6, so $\mathrm{E}(X)=3.5$. With replacement, $X\sim$ Geometric(1/6), so $\mathrm{E}(X)=6$. Answer is (d).	unseen
(v)	By the Central Limit Theorem, an approximate distribution for S_n is $\mathrm{N}(n\mu,n\sigma^2)$. The distribution of $T_n=S_n/\sqrt{n}$ is then, approximately, $\mathrm{N}(\sqrt{n}\mu,\sigma^2)$. Answer is (a).	unseen
		4 marks each

Question 2.

Marks & seen/unseen

Parts

(i) $\mathrm{P}(X \geq x) = \sum_{k=x}^{\infty} p(k) = p \sum_{k=x}^{\infty} (1-p)^{k-1} = p \times \frac{(1-p)^{x-1}}{1-(1-p)} = (1-p)^{x-1}$.

seen — 4 marks

We then have $\mathrm{E}(X) = \sum_{x=1}^{\infty} \mathrm{P}(X \geq x) = \sum_{x=1}^{\infty} (1-p)^{x-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$.

unseen — 4 marks

(ii) If X is the number of rolls required, then $X \sim \mathrm{Geom}\left(\frac{1}{6}\right)$, so $\mathrm{E}(X) = \frac{1}{\frac{1}{6}} = 6$.

seen sim. — 3 marks

(iii) This is the same as the probability of failing to roll a 6 in five attempts; the first five unsuccessful rolls have no bearing on subsequent attempts (the Geometric is memoryless).

The probability is $\mathrm{P}(X \leq 10 | X > 5) = 1 - \mathrm{P}(X \geq 11 | X > 5) = 1 - \left(1 - \frac{1}{6}\right)^5 = 0.598$.

seen sim. — 4 marks

(iv) $\sum_{x=1}^{\infty} \mathrm{P}(X \geq x) = \sum_{x=1}^{\infty} \sum_{k=x}^{\infty} p(k) = \sum_{k=1}^{\infty} \sum_{x=1}^{k} p(k) = \sum_{k=1}^{\infty} k p(k) = \mathrm{E}(X)$.

unseen — 5 marks

Question

3.

Marks & seen/unseen

Parts

(i)

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \implies \ell(\lambda) = (\sum_{i=1}^{n} x_i) \log(\lambda) - n\lambda - \log(\prod_{i=1}^{n} x_i!)$$

$$\implies \frac{d}{d\lambda}\ell(\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \implies \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

This is a maximum since

$$\implies \frac{d^2}{d\lambda^2}\ell(\lambda) = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

is negative for all $\lambda > 0$.

seen

7 marks

(ii) $\hat{\lambda} = \bar{x}$ =1.615.

3 marks

(iii) Assuming a Poisson(1.61) distribution, we would have the following expected counts.

Number of emails (i)	0	1	2	3	≥4
Observed Frequency (O_i)	14	39	29	11	7
Expected Frequency (E_i)	19.89	32.12	25.94	13.96	8.09

These give a chi-square statistic of

$$X^2 = \sum_{i=0}^{4} \frac{(O_i - E_i)^2}{E_i} = 4.35$$

which is less than $\chi^2_{3,0.95}$=7.81. So insufficient evidence to reject the null hypothesis that the email counts follow a Poisson distribution.

seen sim.

5 marks

(iv) Now assuming a Poisson(2) distribution, we would have the following expected counts.

Number of emails (i)	0	1	2	3	≥4
Expected Frequency (E_i)	13.53	27.07	27.07	18.04	14.29

Here,

$$X^2 = \sum_{i=0}^{4} \frac{(O_i - E_i)^2}{E_i} = 11.88$$

which is greater than $\chi^2_{4,0.95}$=9.49, so we reject the null hypothesis that the email counts follow a Poisson(2) distribution. This suggests the mean number of emails per hour is not close to 2.

seen sim.

5 marks

Question 4.

Marks & seen/unseen

Parts

(i) $X_i \sim \mathrm{U}(0, b)$ has cumulative distribution function

seen 2 marks

$$F_{X_i}(x) = \begin{cases} 0, & x \le 0 \\ \frac{x}{b}, & 0 < x < b \\ 1, & x \ge b. \end{cases}$$

(ii) By independence of the $\{X_i\}$,

unseen 3 marks

$$F_{X_{(n)}}(x) = \mathrm{P}(X_{(n)} < x) = \prod_{i=1}^{n} \mathrm{P}(X_i < x) = \left(\frac{x}{b}\right)^n, \quad 0 \le x \le b.$$

The density is the first derivative of this distribution function, giving

seen 4 marks

$$f_{X_{(n)}}(x) = \frac{d}{dx} F_{X_{(n)}}(x) = \frac{d}{dx}\left(\frac{x}{b}\right)^n = \frac{n}{b}\left(\frac{x}{b}\right)^{n-1}, \quad 0 \le x \le b.$$

(iii)

seen 5 marks

$$\mathrm{E}(X_{(n)}|b) = \int_{-\infty}^{\infty} x\, f_{X_{(n)}}(x) dx = n \int_0^b \left(\frac{x}{b}\right)^n dx = \frac{n}{n+1} \frac{x^{n+1}}{b^n}\Big|_0^b = \left(\frac{n}{n+1}\right) b.$$

(iv) As an estimator of b,

unseen 3 marks

$$\mathrm{Bias}(X_{(n)}) = \mathrm{E}(X_{(n)}|b) - b = -\frac{b}{n+1}.$$

Since $\mathrm{E}(X_{(n)}|b) = \left(\frac{n}{n+1}\right) b$, consider a revised estimator $T = \left(\frac{n+1}{n}\right) X_{(n)}$.
Then clearly $\mathrm{E}(T|b) = b$, so T is now an unbiased estimator for b.

unseen 3 marks