STAT2602 Mid-term Class Test

NOT TO BE TAKEN AWAY

[Total: 50 marks]

Fall, 2024 **Time: 10:30am - 12:10pm**

1. Suppose that a random variable X has a p.d.f given by

$$f(x) = \begin{cases} 3e^{-3x}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

(i) Calculate the moment generating function of X. **[5 marks]**

(ii) Calculate $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$. **[2 marks]**

(iii) Calculate $\mathrm{E}(e^{X/2})$ $\mathrm{E}(e^{X})$, and $\mathrm{E}(e^{4X})$, if exists. **[3 marks]**

[Total: 10 marks]

2. The number of male and female customers who visited a company on n successive days were recorded as $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, such that there were X_i male and Y_i female visitors on the i-th day. Assume that for each i, X_i and Y_i follow Poisson distributions with means λ and $\beta\lambda$ respectively, for $\lambda, \beta > 0$, and that the $2n$ observations are independent.

It is known that the Poisson distribution with mean $\theta > 0$ has the p.d.f (or p.m.f.) function

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(i) Show that the likelihood function of (λ, β) based on the observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is

$$L(\lambda, \beta) = \frac{e^{-n(1+\beta)\lambda}\beta^{\sum_{i=1}^n Y_i}\lambda^{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i}}{\prod_{i=1}^n (X_i! Y_i!)}.$$

[5 marks]

(ii) If it is known that $\beta = 1$, show that the statistic.

$$T = \sum_{i=1}^n (X_i + Y_i)$$

is sufficient for λ. **[5 marks]**

(iii) If both λ and β are unknown, show that the statistic

$$T = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i \right)$$

is sufficient for (λ, β). **[5 marks]**

(iv) From now on suppose that $\beta = \lambda$, but the value of λ is unknown.

(a) Show that the statistic

$$T = \sum_{i=1}^{n} (X_i + 2Y_i)$$

is sufficient for λ. **[5 marks]**

(b) Show that the Fisher information based on the $2n$ observations is

$$I(\lambda) = n \left(4 + \frac{1}{\lambda} \right).$$

[5 marks]

(c) Show that the maximum likelihood estimator of λ is

$$\hat{\lambda} = \sqrt{\frac{1}{16} + \frac{T}{2n}} - \frac{1}{4}, \qquad \text{where} \qquad T = \sum_{i=1}^{n} (X_i + 2Y_i)$$

[5 marks]

(d) If n is large, the maximum likelihood estimator $\hat{\lambda}$ has an approximately normal distribution. State the mean and variance of this normal distribution. **[5 marks]**

(e) It is observed that the total numbers of male and female visitors on $n = 20$ successive days are

$$\sum_{i=1}^{20} X_i = 40, \qquad \text{and} \qquad \sum_{i=1}^{20} Y_i = 10$$

respectively. Calculate the maximum likelihood estimate of λ and its standard error. **[5 marks]**

[Total: 40 marks]

A LIST OF STATISTICAL FORMULAE

1. $M_X(t) = \mathrm{E}(e^{tX})$. $\quad \mathrm{E}(X^r) = \left(\dfrac{\mathrm{d}^r}{\mathrm{d}t^r} M_X(t)\right)\Big|_{t=0}$.
2. For $X \sim \text{Uniform}(a, b)$,
$$f(x) = \frac{x}{b-a}$$
3. For $X \sim \mathrm{N}(\mu, \sigma^2)$,
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$
4. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$. $\quad S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$.
5. $\text{Bias}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$. $\quad \mathrm{E}\left[(\hat{\theta} - \theta)^2\right] = \text{Var}(\hat{\theta}) + \left[\text{Bias}(\hat{\theta})\right]^2$.
6. $I(\theta) = \mathrm{E}\left[\left(\dfrac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right] = \mathrm{E}\left[-\dfrac{\partial^2 \ln f(X;\theta)}{\partial \theta^2}\right]$. $\quad \text{Var}(\hat{\theta}) \geq \dfrac{1}{nI(\theta)}$.
7. $I_n(\theta) = \mathrm{E}\left[\left(\dfrac{\partial \ell(\theta)}{\partial \theta}\right)^2\right] = \mathrm{E}\left[-\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right] = nI(\theta)$, where $\ell(\theta) = \ln L(\theta)$, $L(\theta) = \mathbf{f}(X_1, X_2, \ldots, X_n; \theta)$ for $\theta \in \Omega$, and $\mathbf{f}(x_1, x_2, \ldots, x_n; \theta)$ is the joint p.d.f. for the random sample $X_1, X_2, \ldots, X_n$.
8. Factorization: $\mathbf{f}(x_1, \cdots, x_n; \theta) = g\big(T(x_1, \cdots, x_n); \theta\big) h(x_1, \cdots, x_n)$.
9. Expontial family: $f(x;\theta) = h(x)c(\theta)\exp\left(\displaystyle\sum_{i=1}^{s} p_i(\theta) t_i(x)\right)$.
10. CLT: $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$ for large n.
11. MLE: $\dfrac{\hat{\theta} - \theta}{\sqrt{1/I_n(\theta)}} \approx N(0,1)$ for large n.

********** **END OF PAPER** **********