1 Q1

1.1

Let $K_i = \epsilon(t - X_i)$. $P(K_i = 1) = P(t - X_i \geq 0) = P(X_i \leq t) = F(t)$. And $P(K_i = 0) = P(t - X_i < 0) = P(X_i > t) = 1 - F(t)$. Since X_i is i.i.d, K_i is i.i.d also.

Then the pdf function for K_i is

$$P(K_i = k) = \begin{cases} F(t), & \text{if k = 1} \\ 1 - F(t), & \text{if k = 0} \\ 0, & \text{otherwise} \end{cases}$$

Then

$$E[\hat{F}_n(t)] = \frac{1}{n} nE[K_i] = E[K_i] = 1 * P(K_i = 1) + 0 * (1 - F(t)) = F(t) \tag{1}$$

$$Var[\hat{F}_n(t)] = \frac{1}{n^n} nVar[K_i] = \frac{1}{n} Var[K_i] = \frac{1}{n} \underbrace{(F(t)(1 - F(t)))}_{\text{By formula for Bernouli Distribution}} = \frac{1}{n}(F(t) - F(t)^2) \tag{2}$$

1.2

First, let put the multiplier into the absolute value sign, we obtain

$$P[\sqrt{\frac{n}{\alpha_n}} |\hat{F}_n(t) - F(t)| > \epsilon] = P[|\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t) - \sqrt{\frac{n}{\alpha_n}} F(t)| > \epsilon] \tag{3}$$

Now let's check if the random part $\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t)$. Follow the result from (a):

$$E[\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t)] = \sqrt{\frac{n}{\alpha_n}} E[\hat{F}_n(t)] = \sqrt{\frac{n}{\alpha_n}} F_n(t) \tag{4}$$

$$Var[\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t)] = \frac{n}{\alpha_n} Var[\hat{F}_n(t)] = \frac{n}{\alpha_n} \frac{1}{n} (F(t) - F(t)^2) = \frac{1}{\alpha_n} (F(t) - F(t)^2) \tag{5}$$

Because it is given $\alpha_n \to \infty$ and $(F(t) - F(t)^2) < \infty$, thus

$$\lim_{n \to \infty} Var[\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t)] = \lim_{n \to \infty} \frac{1}{\alpha_n} (F(t) - F(t)^2) \to 0$$

By Chebyshev's inequality, we have for arbitrary $\epsilon > 0$:

$$P[|\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t) - \sqrt{\frac{n}{\alpha_n}} F(t)| > \epsilon] \leq \frac{Var[\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t)]}{\epsilon^2} = \frac{\frac{1}{\alpha_n}(F(t) - F(t)^2)}{\epsilon^2} \xrightarrow{n \to \infty} 0 \tag{6}$$

Since probability is non-negative, by squeeze theorem:

$$P[|\sqrt{\frac{n}{\alpha_n}} \hat{F}_n(t) - \sqrt{\frac{n}{\alpha_n}} F(t)| > \epsilon] \xrightarrow{n \to \infty} 0 \tag{7}$$

1.3

For arbitrary $\epsilon > 0$ we have:

$$P[\sqrt{\frac{n}{\alpha_n}} \max_{1\le i\le k_n} |\hat{F}_n(t_i) - F(t_i)| > \epsilon] = P[\max_{1\le i\le k_n} |\sqrt{\frac{n}{\alpha_n}}\hat{F}_n(t_i) - \sqrt{\frac{n}{\alpha_n}}F(t_i)| > \epsilon] \tag{8}$$

$$\text{(max gives a subset of the total union set)} \Rightarrow \le P[\bigcup_{1\le i\le k_n} |\sqrt{\frac{n}{\alpha_n}}\hat{F}_n(t_i) - \sqrt{\frac{n}{\alpha_n}}F(t_i)| > \epsilon] \tag{9}$$

$$\text{(by the theorem about additivity from Math556)} \Rightarrow \le \sum_{1\le i\le k_n} P[|\sqrt{\frac{n}{\alpha_n}}\hat{F}_n(t_i) - \sqrt{\frac{n}{\alpha_n}}F(t_i)| > \epsilon] \tag{10}$$

$$\text{(follow from the what we did in (2))} \le \sum_{1\le i\le k_n} \frac{\frac{1}{\alpha_n}(F(t_i) - F(t_i)^2)}{\epsilon^2} \tag{11}$$

$$\text{(let } t^* \text{ denote the } t_i \text{ that maximize } |\hat{F}_n(t_i) - F(t_i)|) \Rightarrow \le \sum_{1\le i\le k_n} \frac{\frac{1}{\alpha_n}(F(t^*) - F(t^*)^2)}{\epsilon^2} \tag{12}$$

Here t^* denote the t_i that maximize $|\hat{F}_n(t_i) - F(t_i)|$ also maximizes $(\sum_{1\le i\le k_n} \frac{1}{\alpha_n}(F(t^*) - F(t^*)^2))$ since the calculation of variance depends on $(\hat{F}_n(t_i) - F(t_i))^2 = (|\hat{F}_n(t_i) - F(t_i)|)^2$. Therefore, we get the last inequality in the above.

$$P[\sqrt{\frac{n}{\alpha_n}} \max_{1\le i\le k_n} |\hat{F}_n(t_i) - F(t_i)| > \epsilon] \le \sum_{1\le i\le k_n} \frac{\frac{1}{\alpha_n}(F(t^*) - F(t^*)^2)}{\epsilon^2} \tag{13}$$

$$\text{it is given in the question that } \lim_{n\to\infty} \frac{k_n}{a_n} = 0 \Rightarrow = \frac{k_n}{\alpha_n} \frac{(F(t^*) - F(t^*)^2)}{\epsilon^2} \xrightarrow{n\to\infty} 0 \tag{14}$$

$$\tag{15}$$

Hence, for arbitrary $\epsilon > 0$ as $n \to \infty$, we have

$$P[\sqrt{\frac{n}{\alpha_n}} \max_{1\le i\le k_n} |\hat{F}_n(t_i) - F(t_i)| > \epsilon] \to 0 \tag{16}$$

2 Q2

2.1

By independence, the likelihood function is:

$$L(\lambda_1, \lambda_2|x, y) = \prod_{i=1}^{n} \frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!}\frac{e^{-\lambda_2}\lambda_1^{y_i}}{y_i!}$$

The log likelihood is:

$$l(\lambda_1, \lambda_2|x, y) = -(\lambda_1 + \lambda_2)n + \sum_{i=1}^{n}(x_i \log \lambda_1 + y_i \log \lambda_2) - \sum_{i=1}^{n}(\log x_i! + \log y_i!)$$

Take the derivative and set the result to 0:

$$\frac{\partial}{\partial \lambda_1} l(\lambda_1, \lambda_2|x, y) = -n + \sum_{i=1}^{n} \frac{x_i}{\lambda_1} = 0 \tag{17}$$

$$\Rightarrow \hat{\lambda}_1 = \frac{\sum_{i=1}^{n} x_i}{n} \tag{18}$$

$$\frac{\partial}{\partial \lambda_2} l(\lambda_1, \lambda_2|x, y) = -n + \sum_{i=1}^{n} \frac{y_i}{\lambda_2} = 0 \tag{19}$$

$$\Rightarrow \hat{\lambda}_2 = \frac{\sum_{i=1}^{n} y_i}{n} \tag{20}$$

Then the MLE from our sample of size n is $\hat{\theta} = (\hat{\lambda}_1, \hat{\lambda}_2) = (\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n})$. By the invariance property of MLE, the MLE of τ is

$$\hat{\tau} = g(\hat{\theta}) = \frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2}$$

For consistency, let's first check the conditions for $\hat{\lambda}_1$ to be consistent (known i.i.d):

$$E[\hat{\lambda}_1] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}nE[X_i] = \frac{1}{n}n\lambda_1 = \lambda_1 \tag{21}$$

$$Var[\hat{\lambda}_1] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}nVar[X_i] = \frac{1}{n^2}n\lambda_1 = \frac{\lambda_1}{n} \tag{22}$$

$$\tag{23}$$

We see our estimator $\hat{\lambda}_1$ is unbiased and the variance goes to 0 as n goes to infinity and thus $\hat{\lambda}_1$ is consistent. The same argument applies to $\hat{\lambda}_2$ and thus $\hat{\lambda}_2$ is consistent also.

By the continuous mapping theorem where $\mathbf{g} : (\mathbf{R^2}, \mathbf{R^1})$, $g(\theta) = \frac{\lambda_1}{\lambda_1+\lambda_2}$ (assume $\lambda_1 + \lambda_2 \neq 0$), shown $\hat{\theta} = (\hat{\lambda}_1, \hat{\lambda}_2)$ is a consistent estimator (equivalent to convergent in probability), $(\frac{\hat{\lambda}_1}{\hat{\lambda}_1+\hat{\lambda}_2})$ is consistent (converges in probability) for $(\frac{\lambda_1}{\lambda_1+\lambda_2})$.

Or we could argue that respectively 1.$(\hat{\lambda}_1 + \hat{\lambda}_2)$ is a consistent estimator for $(\lambda_1 + \lambda_2)$ by showing

$$E[\hat{\lambda}_1 + \hat{\lambda}_2] = E[\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} Y_i] = \frac{1}{n}n(E[X_i] + E[Y_i]) = \lambda_1 + \lambda_2$$

$$Var[\hat{\lambda}_1 + \hat{\lambda}_2] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} Y_i] = \frac{1}{n^2}n(Var[X_i] + Var[Y_i]) = \frac{\lambda_1 + \lambda_2}{n},$$

they are unbiased and variance goes to 0 as n goes to infinity. 2. It is shown previously that $\hat{\lambda}_1$ is consistent. Thus, 1+2 (assume $\lambda_1+\lambda_2 \neq 0$) is consistent (converges in probability) since the numerator and denominator parts are consistent.

2.2

By the Asymptotic Normality of the MLE, as n goes to infinity, we have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}^{-1}(\theta))$$

where $\hat{\theta}$ denote the MLE from our sample of size n (here the sign n is omitted, $\hat{\theta} = (\hat{\lambda}_1, \hat{\lambda}_2) = \hat{\theta}_n = (\hat{\lambda}_{(n),1}, \hat{\lambda}_{(n)2})$.

Under some regularity conditions and known $g(\theta) = \frac{\lambda_1}{\lambda_1+\lambda_2}$ (assume $\lambda_1 + \lambda_2 \neq 0$) is continuous, we can apply Delta's Method and get:

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0, \nabla g(\theta)^T \mathcal{I}^{-1}(\theta) \nabla g(\theta)) \tag{24}$$

By continuous mapping theorem, (like dividing unnormalized random variable $Z_i \sim N(0, \nabla g(\theta)^T \mathcal{I}^{-1}(\theta) \nabla g(\theta))$ by its standard deviation) we obtain:

$$\sqrt{n} \frac{1}{\sqrt{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta) \nabla g(\theta)}} (g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0,1) \tag{25}$$

We've shown $\hat{\theta}$ is a consistent estimator of θ (converges in probability), then by continuous mapping theorem

$$k(\theta, \hat{\theta}) = \sqrt{\frac{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta)(\theta)}{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}} \xrightarrow{p} \sqrt{\frac{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta)(\theta)}{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta)(\theta)}} = 1 \tag{26}$$

Thus, by Slutsky's Theorem, we find the **pivot quantity** $Q(X, Y|\theta) \sim N(0,1)$:

$$\underbrace{\frac{\sqrt{n}}{\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}} (g(\hat{\theta}) - g(\theta))}_{=Q(X,Y|\theta)} = \frac{\sqrt{n}}{\sqrt{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta) \nabla g(\theta)}} \underbrace{\sqrt{\frac{\nabla g(\theta)^T \mathcal{I}^{-1}(\theta)(\theta)}{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}}_{k(\theta,\hat{\theta})} (g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0,1) \tag{27}$$

It is set that $\hat{\tau} = g(\hat{\theta})$. We get:

$$\underbrace{\frac{\sqrt{n}}{\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}} (\hat{\tau} - \tau)}_{=Q(X,Y|\theta)} \xrightarrow{d} N(0,1) \tag{28}$$

Let's $Z \sim N(0,1)$ and construct a double-sided $100(1-\alpha)\%$ confidence interval. By what we have shown, as n goes large, we can write

$$P(|Q(X,Y|\theta)| < z) = P(|\frac{\sqrt{n}}{\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}} (\hat{\tau} - \tau)| < z) = P(|Z| < z) = 1 - \alpha \tag{29}$$

We could control the significant level α by choosing the proper value of z. The confidence in the above at α level is given by

$$\textbf{C.I.} = (\hat{\tau} - \frac{z\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}{\sqrt{n}}, \hat{\tau} + \frac{z\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}{\sqrt{n}}) \tag{30}$$

Let's simplify it. First, for the $(\hat{\theta})$

$$g(\hat{\theta}) = (\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2}) \tag{31}$$

$$\nabla g(\hat{\theta})^T = [\frac{\partial}{\partial \hat{\lambda}_1}(\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2}), \frac{\partial}{\partial \hat{\lambda}_2}(\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2})] = [\frac{\hat{\lambda}_2}{(\hat{\lambda}_1 + \hat{\lambda}_2)^2}, \frac{-\hat{\lambda}_1}{(\hat{\lambda}_1 + \hat{\lambda}_2)^2}] \tag{32}$$

Next, for the $\mathcal{I}(\theta)$

$$i, j \in \{1,2\}, i \neq j, \text{ known } X_i \text{ and } Y_i \text{ are i.i.d)} \tag{33}$$

$$\Rightarrow \mathcal{I}(\theta)_{ij} = Cov(\frac{\partial}{\partial \hat{\lambda}_i} \log f_\theta(X,Y), \frac{\partial}{\partial \hat{\lambda}_j} \log f_\theta(X,Y)) = 0 \tag{34}$$

$$i \in \{1,2\} \text{ we know poisson pdf is twice differentiable, case i=1 is shown below, similar for i=2} \tag{35}$$

$$\Rightarrow \mathcal{I}(\theta)_{11} = -E[\frac{\partial^2}{\partial \hat{\lambda}_1^2} \log f_\theta(X,Y)] \tag{36}$$

$$= -E[\frac{\partial^2}{\partial \hat{\lambda}_1^2}(-(\lambda_1+\lambda_2)n + \sum_{i=1}^{n}(X_i \log \lambda_1 + Y_i \log \lambda_2) - \sum_{i=1}^{n}(\log X_i! + \log Y_i!))] \tag{37}$$

$$= -E[\sum_{i=1}^{n} -\frac{X_i}{\lambda_1^2}] \tag{38}$$

$$X_i \text{ is i.i.d} = \frac{1}{\lambda_1^2} * n * E[X_i] \tag{39}$$

$$= \frac{n}{\lambda_1} \tag{40}$$

$$\Rightarrow \mathcal{I}(\theta) = \begin{pmatrix} n/\lambda_1 & 0 \\ 0 & n/\lambda_2 \end{pmatrix} \tag{41}$$

$$\text{by the invariance property of MLE} \tag{42}$$

$$\Rightarrow \mathcal{I}(\hat{\theta}) = \begin{pmatrix} n/\hat{\lambda}_1 & 0 \\ 0 & n/\hat{\lambda}_2 \end{pmatrix} \quad \text{and} \quad \mathcal{I}(\hat{\theta})^{-1} = \begin{pmatrix} \hat{\lambda}_1/n & 0 \\ 0 & \hat{\lambda}_2/n \end{pmatrix} \tag{43}$$

Plug these in, we get

$$g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta}) = \frac{\hat{\lambda}_1^2 \hat{\lambda}_2 + \hat{\lambda}_1 \hat{\lambda}_2^2}{n(\hat{\lambda}_1+\hat{\lambda}_2)^4} \tag{44}$$

$$\frac{z\sqrt{g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}{\sqrt{n}} = \frac{z\sqrt{\hat{\lambda}_1^2 \hat{\lambda}_2 + \hat{\lambda}_1 \hat{\lambda}_2^2}}{n(\hat{\lambda}_1+\hat{\lambda}_2)^2} \tag{45}$$

The the confidence interval is

$$\textbf{C.I.} = (\hat{\tau} - \frac{z\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}{\sqrt{n}}, \hat{\tau} + \frac{z\sqrt{\nabla g(\hat{\theta})^T \mathcal{I}^{-1}(\hat{\theta}) \nabla g(\hat{\theta})}}{\sqrt{n}}) \tag{46}$$

$$= (\frac{\hat{\lambda}_1}{\hat{\lambda}_1+\hat{\lambda}_2} - \frac{z\sqrt{\hat{\lambda}_1^2 \hat{\lambda}_2 + \hat{\lambda}_1 \hat{\lambda}_2^2}}{n(\hat{\lambda}_1+\hat{\lambda}_2)^2}, \frac{\hat{\lambda}_1}{\hat{\lambda}_1+\hat{\lambda}_2} + \frac{z\sqrt{\hat{\lambda}_1^2 \hat{\lambda}_2 + \hat{\lambda}_1 \hat{\lambda}_2^2}}{n(\hat{\lambda}_1+\hat{\lambda}_2)^2}) \tag{47}$$

z is the number chosen by us that gives $P(|Z| < z) = 1 - \alpha$ where $Z \sim N(0,1)$

3 Q3

$$P_\theta(X=x) = \frac{\alpha(x)\theta^x}{f(\theta)} = \alpha(x) * \frac{1}{f(\theta)} * \exp\{x\log\theta\} \tag{48}$$

Let $h(x) = \alpha(x)$, $c(\theta) = \frac{1}{f(\theta)}$, $w_1(\theta) = \log\theta$ and $t_1(x) = x$. $P_\theta(X=x)$ is a member of exponential family. Then let

$$k(x|\theta) = \prod_{i=1}^{n} \frac{\alpha(x_i)\theta^{x_i}}{f(\theta)} = (\prod_{i=1}^{n} \alpha(x_i)) * \frac{1}{f(\theta)^n} * \exp\{(\log\theta) * (\sum_{i=1}^{n} x_i)\}$$

As we see, for $T_n(X) = \sum_{i=1}^{n} X_i$ 1.it is sufficient follows from Fisher-Neyman Factorization Theorem, 2.it is a member of exponential family (joint pdf still takes the form of exponential family) with full-rank pmf d=k=1 (parameter space contains an open set in $\mathbf{R}$) and thus it is complete (Theorem in lecture note). $T_n(X) = \sum_{i=1}^{n} X_i$ is complete sufficient statistic.
Let's calculate the pmf of T_n (using the notation given in the question):

$$\begin{aligned}
P(T=t) &= \sum_{\vec{x}\in R(t)} P(X_1 = x_1, \ldots, X_n = x_n) && (49)\\
\text{(by i.i.d again)} &= \sum_{\vec{x}\in R(t)} \prod_{i=1}^{n} P_\theta(X=x) && (50)\\
&= \sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \alpha(x_i)) * \frac{1}{f(\theta)^n} * \exp\{(\log\theta) * (\sum_{i=1}^{n} x_i)\} && (51)\\
&= \frac{1}{f(\theta)^n} \sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \alpha(x_i)) * \exp\{(\log\theta) * (t)\} && (52)\\
&= \frac{\theta^t}{f(\theta)^n} \sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \alpha(x_i)) && (53)\\
&= \frac{\theta^t}{f(\theta)^n} c(t,n) && (54)
\end{aligned}$$

Known $T_n(x)$ is complete sufficient, by Theorem 2.22 (Lehmann-Scheffe) in the lecture's note, if there exists an unbiased estimator $h(T)$ for $d(\theta)$, then UMVUE is given by $E(h(t)|T=t)$. Then to find a UMVUE, we need to first obtain an unbiased estimator $h(T)$ for $d(\theta) = \theta^r$. Then we can show:

$$\begin{aligned}
E(h(T)) &= \theta^r && (55)\\
\sum_{t=0}^{\infty} h(t) \frac{\theta^t}{f(\theta)^n} c(t,n) &= \theta^r && (56)\\
\sum_{t=0}^{\infty} h(t)c(t,n)\theta^t &= f(\theta)^n \theta^r && (57)\\
&= (\sum_{x=0}^{\infty} \alpha(x)\theta^x)^n \theta^r && (58)\\
\text{(polynomial multiplication, group by terms using formula)} &= (\alpha(0)\theta^0 + \alpha(2)\theta^2 \ldots)^n \theta^r && (59)\\
&= \sum_{t=0}^{\infty} \{ \sum_{\vec{x}\in R(t-r)} \prod_{i=1}^{n} \alpha(x_i) \} \theta^{r+t} && (60)\\
&= \sum_{t=0}^{\infty} c(t-r,n)\theta^{r+t} && (61)\\
& && (62)
\end{aligned}$$

$$\textbf{Finally we arrives at} \Rightarrow \quad \sum_{t=0}^{\infty} h(t)c(t,n)\theta^t = \sum_{t=0}^{\infty} c(t-r,n)\theta^{r+t}$$

Again, by some algebra, the h(t) satisfies the above condition is (let $Y_r(t)$ denote $h(t)$ here):

$$Y_r(t) = \begin{cases} 0, & \text{if t < r} \\ \frac{c(t-r,n)}{c(t,n)} & \text{if t} \geq \text{r} \end{cases}$$

Such $Y_r(t)$ makes every term in the summation equal. Since $Y_r(t)$ is based on the complete sufficient statistic T_n and it is shown to be unbiased. It is the UMVUE by Theorem 2.22 (Lehmann-Scheffe).

For poisson distribution, we have $P_\theta(X = x) = \frac{e^{-\theta}\theta^x}{x!}$. It is a member of exponential family and here $\alpha(x) = \frac{1}{x!}$. For $x = 0, 1, 2, \ldots$, $\alpha(x) > 0$. Use the result we've obtained, **write backward**, here we have

$$c(t,n) = \sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \alpha(x_i)) = \sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \frac{1}{x_i!}) \tag{63}$$

$$c(t,n)\theta^t = (\sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \frac{1}{x_i!}))\theta^t \tag{64}$$

$$\sum_{t=0}^{\infty} c(t,n)\theta^t = \sum_{t=0}^{\infty}\{\sum_{\vec{x}\in R(t)} (\prod_{i=1}^{n} \frac{1}{x_i!})\}\theta^t \tag{65}$$

$$= \underbrace{(\frac{1}{0!}\theta^0 + \frac{1}{1!}\theta^1 + \frac{1}{2!}\theta^2 \ldots)^n}_{\text{Taylor Expansion for } e^{n\theta}} = e^{n\theta} = 1 + \frac{(n\theta)^1}{1!} + \frac{(n\theta)^2}{2!} + \ldots = \sum_{t=0}^{\infty} \frac{\theta^t n^t}{t!} \tag{66}$$

Then, the expression for $c(t,n)$ that makes this equation holds could be

$$c(t,n) = \frac{n^t}{t!}$$

Plug in our result for $c(t,n)$ into the general form of UMVUE which is $Y_r(t)$,

$$\frac{c(t-r,n)}{c(t,n)} = \frac{n^{t-r}}{(t-r)!} * \frac{t!}{n^t} = \frac{t!}{(t-r)!n^r}$$

$$Y_r(t) = \begin{cases} 0, & \text{if t < r} \\ \frac{t!}{(t-r)!n^r} & \text{if t} \geq \text{r} \end{cases}$$

4 Q4

4.1

Let $f(x|\theta) = \prod_{i=1}^{n} f_\theta(x_i)$.

It is given that T_n is a sufficient statistic and **by the definition of sufficient statistic** we can write:

$$h(x) = f(x|T(x)) = \frac{f(x, T(x)\theta)}{g(T(x)|\theta)} = \frac{f(x|\theta)}{g(T(x)|\theta)} \tag{67}$$

where 1.$h(x)$ denotes the conditional distribution of x and indicates it doesn't rely on θ, 2. the event for $\{X = x\}$ is a subset of the event for $\{T(X) = T(x)\}$ and thus $f(x, T(x)|\theta)) = f(x|\theta))$, 3.**we let** $g(T(x)|\theta)$ **denote the pmf of** $T(x)$**.**

Hence we have

$$f(x|\theta) = \underbrace{g(T(x)|\theta)}_{\text{pmf for T(x)}} * \underbrace{h(x)}_{\text{Doesn't rely on }\theta} \tag{68}$$

Now we take the logarithm and the derivative. We obtain:

$$\frac{\partial}{\partial\theta}\log[f(x|\theta)] = \frac{\partial}{\partial\theta}\log[g(T(x)|\theta)h(x)] \tag{69}$$

$$= \frac{\partial}{\partial\theta}\log[g(T(x)|\theta)] + \frac{\partial}{\partial\theta}\log[h(x)] \tag{70}$$

$$= \frac{\partial}{\partial\theta}\log[g(T(x)|\theta)]. \tag{71}$$

By the definition of Fisher Information, for our sample $\vec{X} = (X_1, X_2, \ldots, X_n)$ (known i.i.d and pmf is twice differentiable), we have:

$$\mathcal{I}_x(\theta) = E[(\frac{\partial}{\partial\theta}\log[f(x|\theta)])^2|\theta] = E[(\frac{\partial}{\partial\theta}\log[g(T(x)|\theta)])^2|\theta] = \mathcal{I}_{T_n(x)}(\theta) \tag{72}$$

As we see, since $E[(\frac{\partial}{\partial\theta}\log[f(x|\theta)])^2|\theta] = E[(\frac{\partial}{\partial\theta}\log[g(T(x)|\theta)])^2|\theta]$ the Fisher information of T_n is equal to the Fisher information contained in the whole sample $\vec{X}$.

4.2

By the Chain Rule and the derivative of an inverse function, we have ($\theta = h^{-1}(\eta)$):

$$\frac{\partial}{\partial\eta}\ln f_\theta(X) = \frac{\partial}{\partial\eta}\ln f(x, h^{-1}(\eta)) = [\frac{1}{f(x, h^{-1}(\eta))}][\frac{\partial}{\partial(h^{-1}(\eta))}f(x, h^{-1}(\eta))][\frac{\partial}{\partial\eta}h^{-1}(\eta)] \tag{73}$$

Where

$$[\frac{1}{f(x, h^{-1}(\eta))}][\frac{\partial}{\partial(h^{-1}(\eta))}f(x, h^{-1}(\eta))] = [\frac{1}{f(x,\theta)}][\frac{\partial}{\partial\theta}f(x,\theta)] = [\frac{\partial}{\partial\theta}\ln f_\theta(x)]$$

and

$$[\frac{\partial}{\partial\eta}h^{-1}(\eta)] = \frac{1}{h'(h^{-1}(\eta))} = \frac{1}{h'(\theta)}.$$

Hence

$$\frac{\partial}{\partial\eta}\ln f_\theta(X) = [\frac{\partial}{\partial\theta}\ln f_\theta(x)][\frac{1}{h'(\theta)}] \tag{74}$$

Since $h'(\theta)$ doesn't has any random part which is x, when taking the expectation, we can take them out:

$$\mathcal{I}(\eta) = E[(\frac{\partial}{\partial\eta}\ln f_\theta(X))^2] = E[[\frac{\partial}{\partial\theta}\ln f_\theta(x)]^2[\frac{1}{h'(\theta)}]^2] = [\frac{1}{h'(\theta)}]^2 E[(\frac{\partial}{\partial\theta}\ln f_\theta(x))^2] = \frac{\mathcal{I}(\theta)}{[h'(\theta)]^2}$$

5 Q5

5.1

The log likelihood is:

$$l(\theta|x) = \log L(\theta|x) = -\frac{n}{2}(\ln(2\pi) + \ln(\sigma^2)) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2.$$

We want to maximize this likelihood, take the derivative w.r.t σ^2 and set it to 0:

$$\frac{\partial l(\theta|x)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_i^n x_i^2}{2\sigma^4} = 0 \quad (75)$$

$$\Rightarrow \hat{\sigma}^2_{ML} = \frac{\sum_i^n x_i^2}{n} \quad (76)$$

By the invariance property of MLE, the MLE of σ is

$$\hat{\sigma}_{ML} = \sqrt{\hat{\sigma}^2_{ML}} = \sqrt{\frac{\sum_i^n x_i^2}{n}}.$$

For the consistency, let's check the conditions:

It is known $\mu = 0$, and X_i are i.i.d from $N(0, \sigma^2)$, therefore we have

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

where $Z_i \sim N(0,1)$ and thus $\sum_{i=1}^n Z_i^2 \sim \chi^2_n$.

1. For the expectation, we need $\lim_{n\to\infty} Bias(\hat{\sigma}_{ML}) = 0$

$$E(\hat{\sigma}_{ML}) = E(\sqrt{\frac{\sum_i^n X_i^2}{n}}) \quad (77)$$

$$= \sqrt{\frac{\sigma^2}{n}} E(\sqrt{\sum_{i=1}^{n} Z_i^2}) \quad (78)$$

$$\text{Plug in the pdf } \chi^2_n \Rightarrow = \frac{\sigma}{\sqrt{n}} \int_0^{+\infty} \sqrt{x}\frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} dx \quad (79)$$

$$= \frac{\sigma}{\sqrt{n}} \frac{(1/2)^{n/2}}{\Gamma(n/2)} \frac{\Gamma((n+1)/2)}{(1/2)^{(n+1)/2}} \int_0^{\infty} \overbrace{\frac{(1/2)^{(n+1)/2}}{\Gamma((n+1)/2)} x^{(n+1)/2-1} e^{-x/2} dx}^{\text{The pdf of } \chi^2_{n+1}} \quad (80)$$

$$\text{The integral of pdf is 1} \Rightarrow = \frac{\sigma}{\sqrt{n}} \frac{(1/2)^{n/2}}{\Gamma(n/2)} \frac{\Gamma((n+1)/2)}{(1/2)^{(n+1)/2}} \quad (81)$$

$$= \sigma\sqrt{\frac{2}{n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \quad (82)$$

The bias is:

$$Bias(\hat{\sigma}_{ML}) = \sigma - E(\hat{\sigma}_{ML}) = \sigma(1 - \sqrt{\frac{2}{n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}) \quad (83)$$

$$\text{Let } k = \frac{n}{2} \Rightarrow = \sigma(1 - \sqrt{\frac{1}{k}} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k)}) \quad (84)$$

$$\text{Apply the formula for } \alpha \in \mathbf{R},\ lim_{k\to\infty} \frac{\Gamma(k+\alpha)}{\Gamma(k)k^\alpha} = 1 \Rightarrow = \sigma(1 - \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k)\sqrt{k}}) \xrightarrow{k\to\infty} \sigma(1-1) = 0 \quad (85)$$

Thus as n goes to infinity, k goes to infinity also and thus bias goes to 0.

2. For the variance, we need $\lim_{n\to\infty} Var(\hat{\sigma}_{ML}) = 0$

$$
\begin{aligned}
E(\hat{\sigma}^2_{ML}) &= E(\sum_{i=1}^{n} \frac{X_i^2}{n}) && (86)\\
\text{by i.i.d and } \mu = \mu_i = 0 \Rightarrow &= \frac{1}{n}\ n\ E((X_i - \mu_i)^2) && (87)\\
&= \sigma^2 && (88)\\
Var(\hat{\sigma}_{ML}) &= E(\hat{\sigma}^2_{ML}) - (E(\hat{\sigma}_{ML}))^2 && (89)\\
&= \sigma^2 - (\sigma\sqrt{\frac{2}{n}}\ \frac{\Gamma((n+1)/2)}{\Gamma(n/2)})^2 && (90)
\end{aligned}
$$

In 1., we've shown bias goes to 0 and $g(x) = x^2$ is a continuous function on R so $lim_{n\to\infty}(E(\hat{\sigma}_{ML}))^2 = \sigma^2$. Therefore

$$\lim_{n\to\infty} Var(\hat{\sigma}_{ML}) = \sigma^2 - \sigma^2 = 0.$$

Since as n goes to infinity, bias and variance of $\hat{\sigma}_{ML}$ converges to 0, we conclude that $\hat{\sigma}_{ML}$ is a consistent estimator for σ.

5.2

Let's first check it is unbiased:

$$
\begin{aligned}
E(T_n) &= E(\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n}\frac{|X_i|}{n}) && (91)\\
\text{by i.i.d} \Rightarrow &= \sqrt{\frac{\pi}{2}}\ n\ \frac{1}{n}\ E(|X_i|) && (92)\\
&= \sqrt{\frac{\pi}{2}}\int_{-\infty}^{\infty}|x|\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2}}dx && (93)\\
\text{by symmetry} \Rightarrow &= \sqrt{\frac{\pi}{2}}\frac{2}{\sqrt{2\pi\sigma^2}}\int_{0}^{\infty} xe^{-\frac{x^2}{2\sigma^2}}dx && (94)\\
\text{let } u = -\frac{x^2}{2\sigma^2} \Rightarrow &= \sqrt{\frac{\pi}{2}}\frac{2\sigma^2}{\sqrt{2\pi\sigma^2}}\int_{0}^{-\infty} -e^u du && (95)\\
&= \sqrt{\frac{\pi}{2}}\frac{2\sigma^2}{\sqrt{2\pi\sigma^2}}[-e^u]_0^{-\infty} && (96)\\
&= \sigma(0-(-1)) && (97)\\
&= \sigma && (98)
\end{aligned}
$$

For T_n to be consistent, we also need $\lim_{n\to\infty} Var(T_n) = 0$. Use decomposition of variance $Var(X) = E(X^2) - (E(X))^2$ and check the convergence of $Var(T_n)$

$$\begin{aligned}
Var(T_n) &= Var(\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n}\frac{|X_i|}{n}) && (99)\\
&= \frac{\pi}{2n^2}Var(\sum_{i=1}^{n}|X_i|) && (100)\\
\text{by i.i.d} \Rightarrow &= \frac{\pi}{2n^2}\ n\ Var(|X_i|) && (101)\\
&= \frac{\pi}{2n}\{E(|X_i|^2) - E(|X_i|)^2\} && (102)\\
\text{remove the absolute value sign for } (|X_i|)^2 &= \frac{\pi}{2n}\{E(X_i^2) - E(|X_i|)^2\} && (103)\\
\text{we've got } E(|X_i|) \text{ when checking unbiasedness} &= \frac{\pi}{2n}\{(Var(X_i) + (E(X_i))^2 - (\frac{\sigma\sqrt{2}}{\sqrt{\pi}})^2\} && (104)\\
&= \frac{\pi}{2n}(\sigma^2 + 0^2 - \frac{2\sigma^2}{\pi}) && (105)\\
&= \frac{\pi\sigma^2}{2n}(1 - \frac{2}{\pi}) && (106)
\end{aligned}$$

As we see, for finite variance, $\lim_{n\to\infty} Var(T_n) = 0$. Hence, we conclude that T_n is a consistent estimator for σ.

5.3

By the high order moment of normal distribution, we have

$$Var(X_i^4) = E(X_i^4) - (E(X_i^2))^2 = E(X_i^4) - (Var(X_i) + (E(X_i))^2)^2 = 3\sigma^4 - (\sigma^2 - 0)^2 = 2\sigma^4.$$

Having shown $E(\hat{\sigma}^2_{ML}) = E(\sum_{i=1}^{n}\frac{X_i^2}{n}) = \sigma^2$ in (a), then by the CLT we have

$$\sqrt{n}(\sum_{i=1}^{n}\frac{X_i^2}{n} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4). \quad (107)$$

Consider the continuous transformation on $y \in [0,\infty)$, $g(y) = \sqrt{y}$ and then get $\hat{\sigma}_{ML} = g(\hat{\sigma}^2_{ML})$. Apply Delta's Method and continuous mapping theorem (dividing both side by $\sqrt{n}$), for large n, we obtain the limiting distribution:

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}_{ML} - \sigma) = \sqrt{n}(g(\sum_{i=1}^{n}\frac{X_i^2}{n}) - g(\sigma^2)) &\xrightarrow{d} N(0, (g'(\sigma))^2\sigma^4) && (108)\\
\sqrt{n}(\hat{\sigma}_{ML} - \sigma) &\xrightarrow{d} N(0, \frac{\sigma^2}{2}) && (109)\\
(\hat{\sigma}_{ML} - \sigma) &\xrightarrow{d} N(0, \frac{\sigma^2}{2n}) && (110)
\end{aligned}$$

Then, as n goes to infinity:

$$\lim_{n\to\infty}\frac{V(\hat{\sigma}_{ML})}{V(T_n)} = \frac{\frac{\sigma^2}{2n}}{\frac{\pi\sigma^2}{2n}(1-\frac{2}{\pi})} = \frac{1}{\pi - 2} = \frac{1}{1.14} < 1$$

The denominator part $(\pi - 2) \simeq 2.14$. Thus according to asymptotic relative eciency, we'd like to choose σ_{ML} which has smaller variance and gives a narrower confidence interval since

$$\lim_{n\to\infty}\frac{V(\hat{\sigma}_{ML})}{V(T_n)} = \frac{\frac{1}{2}}{(\frac{\pi}{2} - 1)} = \frac{1}{\pi - 2} = \frac{1}{1.14} < 1$$

Alternatively, we can also show this without using CLT and Delta's Method (if CLT and Delta's Method are not eligible here.....). From asymptotic expansion of the ratio of gamma functions, for large n, we have

$$(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)})^2 = \frac{n}{2} - \frac{1}{4} + \mathcal{O}(n^{-1})$$

Then go back to our ratio:

$$\frac{V(\hat{\sigma}_{ML})}{V(T_n)} = \frac{\sigma^2 - (\sigma\sqrt{\frac{2}{n}}\,\frac{\Gamma((n+1)/2)}{\Gamma(n/2)})^2}{\frac{\pi\sigma^2}{2n}(1-\frac{2}{\pi})} = \frac{1 - (\sqrt{\frac{2}{n}}\,\frac{\Gamma((n+1)/2)}{\Gamma(n/2)})^2}{\frac{\pi}{2n}(1-\frac{2}{\pi})} = \frac{n - 2(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)})^2}{(\frac{\pi}{2}-1)} \tag{111}$$

$$\xrightarrow{as\ n\to\infty} = \frac{n - 2(\frac{n}{2}-\frac{1}{4})}{(\frac{\pi}{2}-1)} = \frac{\frac{1}{2}}{(\frac{\pi}{2}-1)} = \frac{1}{\pi-2} \tag{112}$$

Then we see as n goes large, the numerator part:

$$\frac{V(\hat{\sigma}_{ML})}{V(T_n)} \to \frac{1}{\pi-2} \tag{113}$$

6 Q6

6.1

It is given observed pairs (X_i, Y_i) are mutually independent and vector X and Y_i themselves are independent also. Then, we can write pdf

$$f(x,y) = \frac{m!}{x_1! \dots x_n!} \tau_1^{x_1} \dots \tau_n^{x_n} \prod_{i=1}^{n} \frac{e^{-m\beta\tau_i}(m\beta\tau_i)^{y_i}}{y_i!} \tag{114}$$

6.2

By taking the logarithm, we have:

$$\begin{aligned}
l(\theta, \tau | x, y) &= \log(m! \prod_{i=1}^{n} \frac{\tau_i^{x_i}}{x_i!}) + \log(\prod_{i=1}^{n} \frac{e^{-m\beta\tau_i}(m\beta\tau_i)^{y_i}}{y_i!}) && (115)\\
&= \log m! + \sum_{i=1}^{n} x_i \log \tau_i - \sum_{i=1}^{n} \log x_i! + \sum_{i=1}^{n} (-m\beta\tau_i) + \sum_{i=1}^{n} y_i \log(m\beta\tau_i) - \sum_{i=1}^{n} \log y_i! && (116)\\
\sum_i \tau_i = 1 \Rightarrow &= \log m! + \sum_{i=1}^{n} x_i \log \tau_i - \sum_{i=1}^{n} \log x_i! - m\beta + \sum_{i=1}^{n} y_i \log(m\beta\tau_i) - \sum_{i=1}^{n} \log y_i! && (117)
\end{aligned}$$

Add the Lagrange Multiplier:

$$l^*(\theta, \tau | x, y) = \lambda(1 - \sum_{i=1}^{n} \tau_i) + l(\theta, \tau | x, y)$$

Take the derivative w.r.t the parameters and set it to 0:

$$\begin{aligned}
\frac{\partial\, l^*(\theta, \tau | x, y)}{\partial \beta} &= -m \frac{\partial}{\partial \beta}(\beta) + \sum_{i=1}^{n} y_i \frac{\partial}{\partial \beta}(\log(m\beta\tau_i)) = 0 && (118)\\
-m + \sum_{i=1}^{n} \frac{y_i}{\beta} &= 0 && (119)\\
\Rightarrow \hat{\beta} &= \frac{\sum_{i=1}^{n} y_i}{m} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} && (120)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\, l^*(\theta, \tau | x, y)}{\partial \tau_i} &= -\lambda \sum_{j=1}^{n} \frac{\partial}{\partial \tau_i}(\tau_j) + \sum_{j=1}^{n} x_i \frac{\partial}{\partial \tau_i}(\log \tau_j) + \sum_{j=1}^{n} y_i \frac{\partial}{\partial \tau_i}(m\beta\tau_j) = 0 && (121)\\
-\lambda + \frac{x_i}{\tau_i} + \frac{y_i}{\tau_i} &= 0 && (122)\\
\frac{x_i + y_i}{\tau_i} &= \lambda && (123)\\
\Rightarrow \hat{\tau}_i &= \frac{x_i + y_i}{\lambda} && (124)
\end{aligned}$$

Go back and consider our constraint again:

$$\text{The derived MLE of } \tau_i \text{ sum up to 1: } \sum_{i=1}^{n} \hat{\tau}_i = \sum_{i=1}^{n} \frac{x_i + y_i}{\lambda} = 1 \tag{125}$$

$$\Rightarrow \lambda = \sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{126}$$

$$\text{It is given } m = \sum_{i=1}^{n} x_i \text{ and from the MLE of } \beta \text{, we have } \sum_{i=1}^{n} y_i = m\hat{\beta} \tag{127}$$

$$\Rightarrow \lambda = m + \sum_{i=1}^{n} y_i = m + m\hat{\beta} \tag{128}$$

Hence, rewrite MLE of τ_i as:

$$\hat{\tau}_i = \frac{x_i + y_i}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i} = \frac{x_i + y_i}{m + \sum_{i=1}^{n} y_i} = \frac{x_i + y_i}{m + m\hat{\beta}}$$

6.3

Let $m^* = \sum_{i=2}^{n} x_i$ and thus $m = m^* + x_1$:

$$E[l(\theta, \tau | x, y) | \tau_i, \beta, (x_{(-1)}, y)] \tag{129}$$

$$= \sum_{x_1=0}^{\infty} \{\log m! + \sum_{i=1}^{n} x_i \log \tau_i - \sum_{i=1}^{n} \log x_i! - m\beta + \sum_{i=1}^{n} y_i \log(m\beta\tau_i) - \sum_{i=1}^{n} \log y_i!\} \tag{130}$$

$$* \binom{m}{x_1} (\tau_1^{(r)})^{x_1} (1 - \tau_1^{(r)})^{m - x_1} \quad (\text{Plug in } m = m^* + x_1) \tag{131}$$

$$= \sum_{x_1=0}^{\infty} \{\log(m^* + x_1)! + \sum_{i=1}^{n} x_i \log \tau_i! - (m^* + x_1)\beta + \sum_{i=1}^{n} y_i \log(m^* + x_1) + \sum_{i=1}^{n} y_i \log(\beta\tau_i) \tag{132}$$

$$- (\sum_{i=1}^{n} \log x_i! + \sum_{i=1}^{n} \log y_i!\}) \; * \frac{1}{(1 - \tau_1^{(r)})} * \underbrace{\binom{m^* + x_1}{x_1} (\tau_1^{(r)})^{x_1} (1 - \tau_1^{(r)})^{m^*+1}}_{Negative\ Binomial\ (r,p) = (m^*+1, \tau_1^{(r)})} \tag{133}$$

Let k be the term in the first parenthesis, the summation $\sum_{x_1=0}^{\infty}$ gives the result $(k * \frac{1}{(1-\tau_1^{(r)})})$ if it doesn't contain random part X_1 as the pmf of Negative Binomial distribution sum to 1. We want to maximize the expectation. Take such simplification of summation when considering the derivative. Also we can ignore the parts that are not relavant to the parameter β and τ_i:

Set the derivative to 0:

$$\frac{\partial}{\partial \beta} E[l(\theta,\tau|x,y)|\tau_i,\beta,(x_{(-1)},y)] = \frac{1}{(1-\tau_1^{(r)})} * [-(m^* + E[x_1]) + \frac{\sum_{i=1}^n y_i}{\beta}] = 0 \tag{134}$$

$$\Rightarrow \hat{\beta}^{(r+1)} = \frac{\sum_{i=1}^n y_i}{m^* + E[x_1]} = \frac{\sum_{i=1}^n y_i}{\sum_{i=2}^n x_2 + m\hat{\tau}_1^{(r)}} \tag{135}$$

Add Lagrange Multiplier, for i = 1 we have: (136)

$$\frac{\partial}{\partial \tau_1}(E[l(\theta,\tau|x,y)|\tau_i,\beta,(x_{(-1)},y)] + \lambda(1-\sum_{i=1}^n \tau_i)) = \frac{1}{(1-\tau_1^{(r)})} * [-\lambda + \frac{E[X_1]}{\tau_1} + \frac{y_i}{\tau_1}] = 0 \tag{137}$$

For i = 2,3....n we have: (138)

$$\frac{\partial}{\partial \tau_i}(E[l(\theta,\tau|x,y)|\tau_i,\beta,(x_{(-1)},y)] + \lambda(1-\sum_{i=1}^n \tau_i)) = \frac{1}{(1-\tau_1^{(r)})} * [-\lambda + \frac{x_i}{\tau_i} + \frac{y_i}{\tau_i}] = 0 \tag{139}$$

Consider the constraint (τ_1 sum to 1) in the similar way as we did in (b), we can write (140)

$$\Rightarrow \hat{\tau_i}^{(r+1)} = \frac{x_i + y_i}{E[X] + \sum_{i=2}^n x_i + \sum_{i=1}^n y_i} = \frac{x_i + y_i}{m\hat{\tau}_1^{(r)} + \sum_{i=2}^n x_i + \sum_{i=1}^n y_i} \tag{141}$$

We substitute (x_1) the value of expectation $E[X_1]$ in the above by the one about $m\hat{\tau}_1^{(r)}$ assigned at the r-th round.

We get the EM sequence (j=1,2,3,.....n):

$$\hat{\beta}^{(r+1)} = \frac{\sum_{i=1}^n y_i}{m\hat{\tau}_1^{(r)} + \sum_{i=2}^n x_i} \tag{142}$$

$$\text{(for j =1) } \hat{\tau}_1^{(r+1)} = \frac{m\hat{\tau}_1^{(r)} + y_i}{m\hat{\tau}_1^{(r)} + \sum_{i=2}^n x_i + \sum_{i=1}^n y_i} \tag{143}$$

$$\text{(for j } \neq 1) \hat{\tau}_j^{(r+1)} = \frac{x_i + y_i}{m\hat{\tau}_1^{(r)} + \sum_{i=2}^n x_i + \sum_{i=1}^n y_i} \tag{144}$$

We see the result is quite the same except the missing x_1 is replaced by its expectation.

7 Q7

Following the same fashion of the proof of Neyman-Pearson lemma in the lecture's note, we construct the equation in the similar way. Let

$$\mathcal{I} = \int (\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx \tag{145}$$

We check this integrand is greater than or equal to 0. Rewrite this as:

$$\mathcal{I} = (\int_{:f_0(x) > \sum_{j=1}^k a_j f_j(x)} + \int_{:f_0(x) = \sum_{j=1}^k a_j f_j(x)} + \int_{:f_0(x) < \sum_{j=1}^k a_j f_j(x)})(\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx \tag{146}$$

By our definition of ϕ_0 and known $0 \leq \phi \leq 1$, we have

1. For any $x \in \{x : f_0(x) > \sum_{j=1}^k a_j f_j(x)\}$, $\phi_0(x) - \phi(x) = 1 - \phi(x) \geq 0$. Here $(f_0(x) - \sum_{j=1}^k a_j f_j(x)) > 0$. Both two parts are non-negative. Thus

$$\int_{:f_0(x) > \sum_{j=1}^k a_j f_j(x)} (\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx \geq 0.$$

2. For any $x \in \{x : f_0(x) < \sum_{j=1}^k a_j f_j(x)\}$, $\phi_0(x) - \phi(x) = -\phi(x) \leq 0$. Here $(f_0(x) - \sum_{j=1}^k a_j f_j(x)) < 0$. One part is non-positive and one part is negaive. Thus

$$\int_{:f_0(x) < \sum_{j=1}^k a_j f_j(x)} (\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx \geq 0.$$

3. For any $x \in \{x : f_0(x) = \sum_{j=1}^k a_j f_j(x)\}$, here $(f_0(x) - \sum_{j=1}^k a_j f_j(x)) = 0$. One part is 0 . Thus

$$\int_{:f_0(x) = \sum_{j=1}^k a_j f_j(x)} (\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx = 0.$$

Therefore, we get

$$\mathcal{I} = \int (\phi_0(x) - \phi(x))(f_0(x) - \sum_{j=1}^{k} a_j f_j(x))dx \geq 0 \tag{147}$$

$$\Rightarrow \sum_{j=1}^{k} a_j \int (\phi_0(x) - \phi(x)) f_j(x) dx = \int (\phi_0(x) - \phi(x)) \sum_{j=1}^{k} a_j f_j(x) dx \leq \int (\phi_0(x) - \phi(x)) f_0(x) dx \tag{148}$$

$$\Rightarrow \sum_{j=1}^{k} a_j \int (\phi_0(x) - \phi(x)) f_j(x) dx \leq \int (\phi_0(x) - \phi(x)) f_0(x) dx \tag{149}$$

$$\tag{150}$$

Here the left part of the inequality is greater than or equal to 0,

$$0 \leq \sum_{j=1}^{k} a_j \int (\phi_0(x) - \phi(x)) f_j(x) dx$$

because it is given in the question that $0 \leq \phi \leq 1$ and

$$\int \phi(x) f_j(x) dx \leq \int \phi_0(x) f_j(x) dx \tag{151}$$

$$0 \leq \int \phi_0(x) f_j(x) dx - \int \phi(x) f_j(x) dx \tag{152}$$

$$0 \leq \int (\phi_0(x) - \phi(x)) f_j(x) dx \tag{153}$$

$$(\text{known } a_j \geq 0) \qquad 0 \leq \sum_{j=1}^{k} a_j \int (\phi_0(x) - \phi(x)) f_j(x) dx \tag{154}$$

.

Hence it follows that

$$0 \leq \sum_{j=1}^{k} a_j \int (\phi_0(x) - \phi(x)) f_j(x) dx \leq \int (\phi_0(x) - \phi(x)) f_0(x) dx \tag{155}$$

$$0 \leq \int (\phi_0(x) - \phi(x)) f_0(x) dx \tag{156}$$

$$\int \phi(x) f_0(x) dx \leq \int \phi_0(x) f_0(x) dx \tag{157}$$

As we see ϕ_0 maximizes it among all ϕ.