

Group Assignment 2

Daniel Cada, Albert Bargalló i Sales, Erik Schahine,
Michel Messo, and Moaaz Tameer Islam

November 9, 2024

1

Given that $Y|X$ follows a Poisson distribution with parameter $\lambda(x) = \exp(\alpha \cdot x + \beta)$, we need to minimize the negative log-likelihood.

The conditional density is:

$$f_{Y|X}(y, x) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where $\lambda = \lambda(x) = \exp(\alpha \cdot x + \beta)$

The negative log-likelihood for a single observation is:

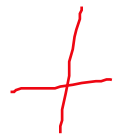
$$\begin{aligned} -\ln(f_{Y|X}(y, x)) &= -\ln\left(\frac{\lambda^y e^{-\lambda}}{y!}\right) \\ &= -[y \ln(\lambda) - \lambda - \ln(y!)] \\ &= -[y(\alpha \cdot x + \beta) - e^{\alpha \cdot x + \beta} - \ln(y!)] \\ &= [-y(\alpha \cdot x + \beta) + e^{\alpha \cdot x + \beta} + \ln(y!)] \end{aligned}$$

For n observations, our loss function becomes:

$$L(\alpha, \beta) = \sum_{i=1}^n [-y_i(\alpha \cdot x_i + \beta) + e^{\alpha \cdot x_i + \beta} + \ln(y_i!)]$$

The term $\ln(y_i!)$ doesn't depend on our parameters α and β . When we minimize this loss function with respect to these parameters, it will become as a constant. So the final equation turns out to:

$$L(\alpha, \beta) = \sum_{i=1}^n [-y_i(\alpha \cdot x_i + \beta) + e^{\alpha \cdot x_i + \beta}]$$



2

2.a Distribution Function

To find the distribution function of $\hat{\theta} = \max(X_1, \dots, X_n)$, we need to calculate $F_{\hat{\theta}}(t) = P(\hat{\theta} \leq t)$. For the maximum to be less than or equal to t , all individual observations must be less than or equal to t . Because they are all independent, we can separate them and then simplify the equation as follows:

$$\begin{aligned} P(\hat{\theta} \leq t) &= P(\max(X_1, \dots, X_n) \leq t) \\ &= P(X_1 \leq t \text{ and } X_2 \leq t \text{ and } \dots \text{ and } X_n \leq t) \\ &= P(X_1 \leq t) \times P(X_2 \leq t) \times \dots \times P(X_n \leq t) \\ &= [P(X_1 \leq t)]^n \end{aligned}$$

For a single observation $X \sim \text{Uniform}(0, \theta)$, we have:

$$P(X \leq t) = \begin{cases} \frac{t}{\theta} & \text{if } 0 \leq t \leq \theta \end{cases}$$

Therefore, the cumulative distribution function of $\hat{\theta}$ is:

$$F_{\hat{\theta}}(t) = \left(\frac{t}{\theta}\right)^n \quad \text{if } 0 \leq t \leq \theta$$



2.b Bias

The bias of $\hat{\theta}$ is defined as:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

We can calculate $E[\hat{\theta}]$ by using the PDF. Luckily we just calculated the CDF previously. We can take the derivative with respect to x to find the PDF:

$$f_{\hat{\theta}}(x) = \frac{d}{dx} F_{\hat{\theta}}(x) = \frac{d}{dx} \left(\frac{x}{\theta} \right)^n = \frac{n \cdot x^{n-1}}{\theta^n}, \quad \text{for } 0 \leq x \leq \theta.$$

Now we can find $E[\hat{\theta}]$ by integrating $x \cdot f_{\hat{\theta}}(x)$ over the range $[0, \theta]$:

$$E[\hat{\theta}] = \int_0^{\theta} x f_{\hat{\theta}}(x) dx = \int_0^{\theta} x \cdot \frac{n \cdot x^{n-1}}{\theta^n} dx.$$

$$E[\hat{\theta}] = \frac{n}{\theta^n} \int_0^{\theta} x^n dx.$$

$$E[\hat{\theta}] = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta.$$

Substituting $E[\hat{\theta}] = \frac{n}{n+1} \theta$, we get:

$$\text{Bias}(\hat{\theta}) = \frac{n}{n+1} \theta - \theta = -\frac{\theta}{n+1}.$$

$$\text{Bias}(\hat{\theta}) = -\frac{\theta}{n+1}.$$



2.c Standard Error

The standard error is the square root of the variance. Thus we can calculate the standard error by using the formula for calculating the variance:

$$\text{Var}(\hat{\theta}) = E[\hat{\theta}^2] - E[\hat{\theta}]^2. \quad (1)$$

$$E[\hat{\theta}]^2 = \left(\frac{n}{n+1} \theta \right)^2 = \frac{n^2 \theta^2}{(n+1)^2} \quad (2)$$

$$E[\hat{\theta}^2] = \int_0^{\theta} x^2 f_{\hat{\theta}}(x) dx = \frac{n}{\theta^n} \int_0^{\theta} x^{n+1} dx = \frac{n}{\theta^n} \left[\frac{x^{n+2}}{n+2} \right]_0^{\theta} = \frac{n}{\theta^n} \cdot \frac{\theta^{n+2}}{n+2} = \frac{n \theta^2}{n+2} \quad (3)$$

Calculating the variance:

$$\text{Var}(\hat{\theta}) = \frac{n \theta^2}{n+2} - \frac{n^2 \theta^2}{(n+1)^2} = \frac{n \theta^2}{(n+1)^2 (n+2)}$$

Thus, the standard error is:

$$\text{SE} = \sqrt{\frac{n \theta^2}{(n+1)^2 (n+2)}}$$

2.d Mean Squared Error

The mean squared error (MSE) is given by:

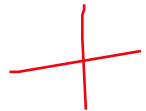
$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2.$$

$$\text{MSE}(\hat{\theta}) = \frac{n \theta^2}{(n+1)^2 (n+2)} + \left(-\frac{\theta}{n+1} \right)^2.$$

$$\text{MSE}(\hat{\theta}) = \frac{n \theta^2}{(n+1)^2 (n+2)} + \frac{\theta^2}{(n+1)^2} = \frac{2 \theta^2}{(n+2)(n+1)}.$$

Finally we are left with:

$$\text{MSE}(\hat{\theta}) = \frac{2 \theta^2}{(n+2)(n+1)}.$$



3

3.a

The cumulative distribution function (CDF) $F(x)$ is defined as:

$$F(x) = \int_{-\frac{\pi}{2}}^x p(t) dt = \int_{-\frac{\pi}{2}}^x \frac{1}{2} \cos(t) dt.$$

$$F(x) = \frac{1}{2} [\sin(t)]_{-\frac{\pi}{2}}^x = \frac{1}{2} \left(\sin(x) - \sin\left(-\frac{\pi}{2}\right) \right).$$

Since $\sin\left(-\frac{\pi}{2}\right) = -1$, the cumulative distribution function is:

$$F(x) = \frac{1}{2} (\sin(x) + 1).$$

domain?

3.b

To find the inverse distribution function F^{-1} , we solve for x in terms of u , where $u = F(x)$. :

$$u = \frac{1}{2} (\sin(x) + 1).$$

$$\sin(x) = 2u - 1.$$

$$x = \arcsin(2u - 1).$$

Thus, the inverse distribution function $F^{-1}(u)$ is:

$$F^{-1}(u) = \arcsin(2u - 1).$$

for $0 < u < 1$.

3.c

Since $p(x) = \frac{1}{2} \cos(x)$, which is bounded within $-\frac{\pi}{2} < x < \frac{\pi}{2}$, we can choose a uniform distribution over this interval as $g(x)$:

$$g(x) = \frac{1}{\pi}, \quad -\frac{\pi}{2} < x < \frac{\pi}{2}.$$

We need M such that $p(x) \leq M g(x)$. This means:

$$\frac{1}{2} \cos(x) \leq M \cdot \frac{1}{\pi}.$$

$$M \geq \frac{\pi}{2} \cos(x).$$

Since $\cos(x)$ reaches its maximum value of 1 at $x = 0$, we can take $M = \frac{\pi}{2}$ to satisfy the inequality for all x in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Therefore, we have:

$$g(x) = \frac{1}{\pi}, \quad M = \frac{\pi}{2}.$$

4

First, X_0, X_1, \dots, X_n is a Markov chain, since the current value of X_n only depends on the value of X_{n-1} at the previous state and the current value of Y_n .

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.2 & 0.4 \\ 0 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

5

To estimate the quantile p of an unknown distribution F based on independent and identically distributed (IID) samples X_1, X_2, \dots, X_n , we use the empirical distribution function \hat{F}_n . The empirical distribution function is defined as:

$$\hat{F}_n(x) = \frac{1}{n} \sum i = 1^n \mathbf{1}_{\{X_i \leq x\}},$$

where $\mathbf{1}_{\{X_i \leq x\}}$ is the indicator function as in the notes.

Since $F(x_p) = p$, where x_p is the p -th quantile of F , we can estimate p by evaluating \hat{F}_n at x_p :

$$\hat{p} = \hat{F}_n(x_p) = \frac{1}{n} \sum i = 1^n \mathbf{1}_{\{X_i \leq x_p\}}.$$

The DKW inequality provides a uniform bound on the difference between the empirical distribution function \hat{F}_n and the true distribution function F :

$$P\left(\sup_x |\hat{F}_n(x) - F(x)| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

We then choose a confidence level $1 - \alpha$, where $\alpha \in (0, 1)$.

$$2e^{-2n\varepsilon^2} = \alpha.$$

$$\varepsilon = \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\alpha}\right)}.$$

With probability at least $1 - \alpha$, the true p lies within ε of \hat{p} :

$$P\left(|\hat{F}_n(x_p) - p| \leq \varepsilon\right) \geq 1 - \alpha.$$

Therefore, the confidence interval for p for our earlier choice of epsilon is:

$$[\hat{p} - \varepsilon, \hat{p} + \varepsilon].$$

Estimate of p :

$$\hat{p} = \hat{F}_n(x_p) = \frac{1}{n} \sum i = 1^n \mathbf{1}_{\{X_i \leq x_p\}}.$$

Confidence Interval for p :

$$\left[\hat{F}_n(x_p) - \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\alpha}\right)}, \quad \hat{F}_n(x_p) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\alpha}\right)} \right],$$

where α is the significance level, and n is the sample size.

technically the CI should be for q , as a function of p , but it's the same i