GroupAssignment2_Team19

Ismat Halabi, Anudeep Battu, Muhammad Hasnain Saeed, Lavy Selvaraj, Muhammad Sajjad

November 11, 2024

1.

$f_{Y|X}(y|x) = \dfrac{\lambda^y e^{-\lambda}}{y!}, \qquad \lambda(x) = exp(\alpha \cdot x + \beta)$

The negative log likelihood is:

$$\begin{aligned}
-\sum_{i=1}^{n} \ln(f_{X,Y}(X_i, Y_i)) &= -\sum_{i=1}^{n} \ln(f_{Y|X}(Y_i|X_i)) - \sum_{i=1}^{n} \ln(f_X(X_i)) \\
&= -\sum_{i=1}^{n} \ln\left(\frac{\lambda(x_i)^{y_i} e^{-\lambda(x_i)}}{y_i!}\right) - C_1 n, \text{ we assume } f_X \text{ does not depend on any parameters} \\
&= -\sum_{i=1}^{n} \ln\left(e^{y_i \ln(\lambda(x_i))}\right) - \sum_{i=1}^{n} \ln\left(e^{-\lambda(x_i)}\right) + \sum_{i=1}^{n} \ln(y_i!) + C_1 n \\
&= -\sum_{i=1}^{n} y_i \ln(\lambda(x_i)) + \sum_{i=1}^{n} \lambda(x_i) + C_2 n + C_1 n \\
&= -\sum_{i=1}^{n} y_i \ln(\exp(\alpha \cdot x_i + \beta)) + \sum_{i=1}^{n} \exp(\alpha \cdot x_i + \beta) + n(C_1 + C_2) \\
&= -\sum_{i=1}^{n} y_i(\alpha \cdot c_i + \beta) + \sum_{i=1}^{n} exp(\alpha \cdot c_i + \beta) + nC_3 \\
&= \sum_{i=1}^{n} exp(\alpha \cdot c_i + \beta) - \sum_{i=1}^{n} y_i(\alpha \cdot c_i + \beta), \text{ the constant term does not depend on } \alpha \text{ or } \beta
\end{aligned}$$

+

2.

Definitions:

- Let $X_1, X_2, ..., X_n \sim$ I.I.D Uniform$(0, \theta)$.
- Estimator $\hat{\theta} = max(X_1, X_2, ..., X_n)$

Since $\hat{\theta} = max(X_1, X_2, ..., X_n)$, we can determine its distribution by noting that $P(\hat{\theta} \leq x)$ is equivalent to the probability that all $X_i's$ are less than or equal to x. Given that the random variables $X_i's$ are I.I.D and follow a uniform distribution, we have:

$$F_{\hat{\theta}}(x) = P(\hat{\theta} \leq x) = P[max(X_1, ...X_n)] \leq x = P(X_1, ..., X_n) \leq x = \prod_{i=1}^{n} P(X_i \leq x) = \frac{x^n}{\theta^n} \tag{1}$$

$$F_{\hat{\theta}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^n}{\theta^n} & \text{if } 0 \leq x \leq \theta \\ 1 & \text{if } x > \theta \end{cases} \tag{2}$$

The probability density function (PDF) is the derivative of its cumulative distribution function (CDF). Hence:

$$f_{\hat{\theta}}(x) = \begin{cases} \frac{n}{\theta^n}(x)^{n-1} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The bias of the estimator $\hat{\theta}$ is given by:

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

To calculate $\mathbb{E}[\hat{\theta}]$, we use the PDF of $\hat{\theta}$:

$$\mathbb{E}[\hat{\theta}] = \int_0^\theta x\, f_{\hat{\theta}}(x)\, dx$$

Substituting $f_{\hat{\theta}}(x)$ into the expectation:

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

Therefore,

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

Finally, substituting back to find the bias:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}$$

Thus, the bias of $\hat{\theta}$ is:

$$\textbf{Bias}(\boldsymbol{\hat{\theta}}) = -\frac{\boldsymbol{\theta}}{\boldsymbol{n+1}}$$

The variance of the estimator $\hat{\theta}$ is defined as:

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

$$\mathbb{E}[\hat{\theta}^2] = \int_0^\theta x^2 \cdot \frac{n}{\theta^n}(x)^{n-1}\,dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,dx = \frac{n}{\theta^n}\cdot\frac{\theta^{n+2}}{n+2}$$

Therefore,

$$\mathbb{E}[\hat{\theta}^2] = \frac{n\theta^2}{n+2}$$

Now, we can substitute into the variance formula:

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

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

Simplifying:

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

The standard error (SE) is the square root of the variance:

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

Thus, the standard error of $\hat{\theta}$ is:

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

The mean squared error (MSE) of the estimator $\hat{\theta}$ is defined as:

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

We can substitute the variance and bias squared into the above formula:

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

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

Thus, the mean squared error of $\hat{\theta}$ is:

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

3.

$$p(x) = \frac{1}{2}cos(x), -\frac{\pi}{2} < x < \frac{\pi}{2} \tag{4}$$

(a)

$$F(x) = \int_{-\pi/2}^{x} \frac{1}{2}\cos(t)dt = \frac{1}{2}sin(t)\Big|_{-\pi/2}^{x} = \frac{1}{2}\left[\sin(x) - (-1)\right] = \underline{\frac{1}{2}\left[\sin(x) + 1\right]} \tag{5}$$

domain?

Integrating $F(x)$ over $[-\pi/2, +\pi/2]$, we get $\int_{-\pi/2}^{+\pi/2} \frac{1}{2}\cos(t)dt = \frac{1}{2}sin(t)\Big|_{-\pi/2}^{+\pi/2} = \frac{1}{2}\left[+1 - (-1)\right] = \frac{1}{2} * 2 = 1$. Hence $p(x)$ is a density function.

(b)

We have :

$$F(x) = y = \frac{1}{2}\left[\sin(x) + 1\right] \implies 2y - 1 = \sin(x) \implies x = F^{-1}(y) = \arcsin(2y - 1), 0 < y < 1 \tag{6}$$

(c)

We need to find $M > 0$ and $g(x)$ such that $p(x) \leq Mg(x)$ for the Accept-Rejection sampler.
$x \in (-\pi/2, +\pi/2)$, $p(x) \in [0, 1/2]$, $p(0) = 1/2$, which is the maximum value, $p(x)$ is continuous.
We choose $g(x)$ as a uniform density distribution in this algorithm for simplicity. It is defined in this case over $(-\pi/2, +\pi/2)$. It must be that $g(x) = 1/\pi \ in \ (-\pi/2, +\pi/2)$ to have an area of 1.
Finally, $M = p(0)/g(x) = \frac{1/2}{1/\pi} = \pi/2$.

4.

Y_i can only take values in the set $\{0, 1, 2, 3\}$. $X_n = max\{Y_1, ..., Y_n\}$. The state space of X_i is also $\{0, 1, 2, 3\}$. Let m be an arbitrary number between 1 and n. Now consider

$$\begin{aligned}\mathbb{P}(X_m = x|X_{m-1}, X_{m-2}, ..., X_0) &= \mathbb{P}(\max\{X_{m-1}, Y_m\} = x|X_{m-1}, X_{m-2}, ..., X_0) \\ &= \mathbb{P}(\max\{X_{m-1}, Y_m\} = x|X_{m-1}) \; Y_m \; is \; independent \; of \; Y_{m-1}, ..., Y_1 \\ &= \mathbb{P}(X_m = x|X_{m-1})\end{aligned}$$

Hence, the Markov property is satisfied and $X_0, .., X_n$ is a Markov chain.
To compute the transition matrix let us start with staying at the state of 0.

$$\mathbb{P}(X_n = 0|X_{n-1} = 0) = \mathbb{P}(\max\{X_{n-1}, Y_n\} = 0|X_{n-1} = 0) = \mathbb{P}(Y_n = 0) = 0.1$$

Similarly let us compute the transition probability from state 0 to state 1:

$$\mathbb{P}(X_n = 1|X_{n-1} = 0) = \mathbb{P}(\max\{X_{n-1}, Y_n\} = 1|X_{n-1} = 0) = \mathbb{P}(Y_n = 1) = 0.3$$

By computing the same way for 0 to 2 state and 0 to 3 state we get If $X_n = 1$ then we need to consider only the state values 1,2 and 3.

$$\mathbb{P}(X_n = 1|X_{n-1} = 1) = \mathbb{P}(\max\{X_{n-1}, Y_n\} = 1|X_{n-1} = 1 = \mathbb{P}(Y_n = 1) = 0.3$$

In the same vein, we see that

$$\mathbb{P}(X_n = 2|X_{n-1} = 2) = \mathbb{P}(Y_n \leq 2) = 0.6 \;\; max \; depends \; on \; Y_n \; being \; less \; than \; or \; equal \; to \; 2.$$

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

(a) Estimating the Quantile $q(p)$

To estimate the p-th quantile $q(p)$ of the distribution F, we define $q(p)$ as the value such that:

$$F(q(p)) = p$$

Using the empirical distribution function $\hat{F}_n(x)$, the estimate $\hat{q}_p$ for the quantile $q(p)$ is:

$$\hat{q}_p = \inf\{x : \hat{F}_n(x) \geq p\}$$

(b) Constructing the Confidence Interval for $q(p)$ Using the DKW Inequality

Using Theorem 5.28, the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality provides an upper bound on the deviation between the empirical distribution function $\hat{F}_n(x)$ and the true distribution function $F(x)$. The inequality states:

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

To achieve a confidence level of $1-\alpha$, we set:

$$2e^{-2n\epsilon^2} = \alpha$$

Solving for ϵ, we get:

$$\epsilon = \sqrt{\frac{\ln(2/\alpha)}{2n}}$$

Thus, with probability at least $1-\alpha$, we have:

$$\sup_x \left|\hat{F}_n(x) - F(x)\right| \leq \epsilon$$

This implies that for any x, with probability at least $1-\alpha$:

$$F(x) \in \left[\hat{F}_n(x) - \epsilon, \hat{F}_n(x) + \epsilon\right]$$

Since $F(q(p)) = p$, we conclude that $q(p)$ must lie in the set:

$$\{x : p - \epsilon \leq \hat{F}_n(x) \leq p + \epsilon\}$$

Therefore, the confidence interval for the quantile $q(p)$ at confidence level $1-\alpha$ is:

$$\left[\hat{F}_n^{-1}(p-\epsilon), \hat{F}_n^{-1}(p+\epsilon)\right]$$

where $\hat{F}_n^{-1}$ represents the inverse of the empirical distribution function, which provides the value x such that $\hat{F}_n(x)$ corresponds to the specified probability.