Group Assignment 2

Introduction to Data Science 1MS041

Truls Karlsson, Nathalie Borglund, Noah Wassberg,
Andreas Larsson

November 2024

Problem 1

We have

$$f_{Y|X}(y, x) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad \lambda(x) = \exp(\alpha \cdot x + \beta) \tag{1}$$

Given this we want to follow the calculations from 4.2.1 to derive a loss that needs to be minimized with respect to α and β. From 4.2.1 we get the following

$$\begin{aligned}
\sum_{i=1}^{n} \ln(f_{X,Y}(X_i, Y_i)) &= -\sum_{i=1}^{n} \ln(f_{Y|X}(Y_i|X_i) f_X(X_i)) \\
&= -\sum_{i=1}^{n} \ln(f_{Y|X}(Y_i|X_i)) - \sum_{i=1}^{n} \ln(f_X(X_i)) \\
&= -\sum_{i=1}^{n} \ln\left(\frac{\lambda(x_i)^{y_i} e^{-\lambda(x_i)}}{y_i!}\right) - \sum_{i=1}^{n} \ln(f_X(X_i)) \\
&= -\sum_{i=1}^{n} y_i \ln(\lambda(x_i)) - \lambda(x_i) - \ln(y_i!) - \sum_{i=1}^{n} \ln(f_X(X_i))
\end{aligned} \tag{2}$$

From this point we can simplify by assuming that f_X is constant with respect to α and β since we cannot know what distribution f_X follows. $-\ln(y_i!)$ is also constant with respect to α and β. We can exclude the constant terms that do not depend on the parameters that are being optimized since they do not influence the optimization process. This leaves us with

$$\sum_{i=1}^{n} y_i \ln(\lambda(x_i)) - \lambda(x_i)$$
$$= -\sum_{i=1}^{n} y_i(\alpha \cdot x_i + \beta) - \exp(\alpha \cdot x_i + \beta) \tag{3}$$

Problem 2

The uniform CDF is:

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases} \tag{4}$$

For x in $\hat{\theta}$, $0 \leq x \leq \theta$ we get

$$F(x_i) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-0}{\theta-0} & \text{for } 0 \leq x \leq \theta \\ 1 & \text{for } x > \theta \end{cases} \tag{5}$$

That is for one x_i in $\hat{\theta}$. Since they are n samples in $\hat{\theta}$ and all are i.i.d the CDF for $\hat{\theta}$ is:

$$F_{\hat{\theta}}(x) = \begin{cases} 0 & \text{for } x < a \\ (\frac{x}{\theta})^n & \text{for } 0 \leq x \leq \theta \\ 1 & \text{for } x > \theta \end{cases} \tag{6}$$

The PDF can be derived by taking the derivative of the CDF:

$$f_{\hat{\theta}}(x) = \frac{d}{dx}\left[\left(\frac{x}{\hat{\theta}}\right)^n\right] = \frac{d}{dx}\left[\frac{x^n}{\hat{\theta}^n}\right] = \frac{nx^{n-1}}{\hat{\theta}^n} \tag{7}$$

To determine the bias we first must find the expectation using the general formula:

$$\mathbb{E}\big[x^k\big] = \int x^k f(x)dx, \text{where } f(x) \text{ is the PDF} \tag{8}$$

In our case we have:

$$\mathbb{E}\big[\hat{\theta}\big] = \int_0^\theta \frac{xnx^{n-1}}{\theta^n}dx = \int_0^\theta \frac{nx^n}{\theta^n}dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n\theta^{n+1}}{\theta^n(n+1)} = \frac{n\theta}{n+1} \tag{9}$$

Consequently the bias is:

$$bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta = \frac{n\theta}{n+1} - \theta = \frac{n\theta - \theta(n+1)}{n+1} = \frac{-\theta}{n+1} \tag{10}$$

The standard error (se) is defined as:

$$se(\hat{\theta}) = \sqrt{\mathbb{V}[\hat{\theta}]} \tag{11}$$

The variance can be determined by:

$$\mathbb{V}[\hat{\theta}] = \mathbb{E}[\hat{\theta^2}] - \left(\mathbb{E}[\hat{\theta}]\right)^2 \tag{12}$$

Now we calculate $\mathbb{E}[\hat{\theta^2}]$ using the same procedure as before.

$$\mathbb{E}[\hat{\theta^2}] = \int_0^\theta \frac{x^2 n x^{n-1}}{\theta^n} dx = \int_0^\theta \frac{n x^{n+1}}{\theta^n} dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^{n+2}}{\theta^2(n+2)} = \frac{n\theta^2}{n+2} \tag{13}$$

$$\mathbb{V}[\hat{\theta}] = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} \tag{14}$$

$$se(\hat{\theta}) = \sqrt{\mathbb{V}[\hat{\theta}]} = \sqrt{\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}} \tag{15}$$

The Mean Squared Error (MSE) or bias, variance trade-off is determined by squaring the adding the standard error and bias:

$$\begin{gathered} MSE(\hat{\theta}) = \left(\sqrt{\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}}\right)^2 + \left(\frac{-\theta}{n+1}\right)^2 = \\ \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} + \frac{\theta^2}{(n+1)^2} = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2 + \theta^2}{(n+1)^2} \end{gathered} \tag{16}$$

Problem 3

a)

The PDF is given as:

$$p(x) = \frac{1}{2}\cos x, \frac{-\pi}{2} < x < \frac{\pi}{2} \tag{17}$$

Since the interval on which $p(x)$ operates is open we integrate not over the full interval but on a range from $\left[\frac{-\pi}{2}, x\right]$.

$$P(x) = \frac{1}{2}\int_{\frac{-\pi}{2}}^{x} \cos x dx = \frac{1}{2}\bigg[\sin x\bigg]_{\frac{-\pi}{2}}^{x} = \frac{1}{2}(\sin x - \sin\frac{-\pi}{2}) = \frac{1}{2}(\sin x + 1) \quad (18)$$

The CDF can now be defined as:

$$P(x) = \begin{cases} 0 & \text{for } x \leq \frac{-\pi}{2} \\ \frac{1}{2}(\sin x + 1) & \text{for } \frac{-\pi}{2} < x < \frac{\pi}{2} \\ 1 & \text{for } x \geq \frac{\pi}{2} \end{cases} \quad (19)$$

b)

To find the inverse CDF, or quantile function, we set $P(x) = y$ and solve for x.

$$\begin{aligned} y &= \frac{1}{2}(\sin x + 1) \\ 2y &= \sin x + 1 \\ 2y - 1 &= \sin x \\ x &= \arcsin(2y - 1) \\ P^{-1}(y) &= \arcsin(2y - 1) \end{aligned} \quad (20)$$

Since the CDF $P(x)$ give outputs on the range $[0, 1]$, the inverse CDF, $P^{-1}(y)$ must take inputs on the same range. arcsin is continuos on the range $[-1, 1]$ which is the same range as its input for $y \in [0, 1], [2 \times 0 - 1, 2 \times 1 - 1]$.

c)

We look into $p(x) \leq Mg(x)$. Since $p(x)$ is a lower bound for $Mg(x)$ we find the maximum of $p(x)$. On the interval $\left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$ $\cos x$ takes the maximum value of 1 (when $x = 0$) which gives $p(x)$ a maximum value of $\frac{1}{2}$. Let us use a uniform distribution for g.

$$g(x) = \frac{1}{\frac{\pi}{2} - \frac{-\pi}{2}} = \frac{1}{\pi} \quad (21)$$

We use the max value of $p(x)$ and rearrange to solve for M:

$$\begin{aligned} p(x) &\leq Mg(x) \\ \frac{1}{2} &\leq M\frac{1}{\pi} \\ \frac{\pi}{2} &\leq M \end{aligned} \quad (22)$$

Problem 4

Let $Y_1, Y_2, ..., Y_n$ be a sequence of IID discrete random variables, where $P(Y_i = 0) = 0.1$, $P(Y_i = 0) = 0.3$, $P(Y_i = 2) = 0.2$ and $P(Y_i = 3) = 0.4$.

Let $X_n = max\{Y_1, ..., Y_n\}$ and let $X_0 = 0$.

We need to verify that $X_0, X_1, ..., X_n$ is a Markov Chain and its transition matrix P. Let's begin with showing that X_n is Markov Chain.

At each time n, X_n depends on X_{n-1} and the new observation Y_n: $X_n = max\{X_{n-1}, Y_n\}$. We know that Y_n is independent of the past due to it being an IID which means that the future state X_n only depends on the present state X_{n-1} and only that state. This satisfies the Markov property:

$$P(X_n = x_n | X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

+

Let us now find the transition matrix P and in order to find it, we need to first calculate the transition probabilities, that is

$$P_{ij} = P(X_n = j | X_{n-1} = i) \text{ for } i, j \in \{0, 1, 2, 3\}$$

which gives the probability of transitioning between state X_{n-1} to X_n. There are three different cases that we need to take into consideration when calculating the transition probabilities and they are as follows:

$$\begin{cases} P(X_n = j | X_{n-1} = i) = P(Y_n = j) & j > i \\ P(X_n = i | X_{n-1} = i) = P(Y_n \leq j) = \sum_{k=0}^{i} P(Y_n = k) & j = i \\ P(X_n = j | X_{n-1} = i) = 0 & j < i \end{cases}$$

The following shows the calculations for the transition probabilities that we will insert later into the transition matrix P:

i = 0:

$$P_{00} = P(Y_i = 0) = 0.1$$

$$P_{01} = P(Y_i = 1) = 0.3$$

$$P_{02} = P(Y_i = 2) = 0.2$$

$$P_{03} = P(Y_i = 3) = 0.4$$

i = 1:

$$P_{10} = 0$$

$$P_{11} = P(Y_i \leq 1) = 0.1 + 0.3 = 0.4$$

$$P_{12} = 0.2$$

$$P_{13} = 0.4$$

i = 2:

$$P_{20} = 0$$

$$P_{21} = 0$$

$$P_{22} = 0.1 + 0.3 + 0.2 = 0.6$$

$$P_{23} = 0.4$$

i = 3:

$$P_{30} = 0$$

$$P_{31} = 0$$

$$P_{32} = 0$$

$$P_{33} = 0.1 + 0.3 + 0.2 + 0.4 = 1$$

We are now done with the transition probabilities. Our transition matrix P is defined as follows:

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{pmatrix}$$

Which when inserting our values that we calculated gives us the following matrix:

$$P = \begin{pmatrix} 0.1 & 0.3 & 0.2 & 0.4 \\ 0 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1.0 \end{pmatrix}$$

Answer: The transition matrix P is as follows:

$$P = \begin{pmatrix} 0.1 & 0.3 & 0.2 & 0.4 \\ 0 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1.0 \end{pmatrix}$$

Problem 5

Given a sequence $X_1, X_2, ..., X_n$ that are IID, from an unknown distribution F, we want to estimate the quantile p of F using the empirical distribution function $\hat{F}_n(x)$. We then want to find a confidence interval for p using Dvoretzky-Kiefer-Wolfowitz (DKW) inequality.

Since we don't know the true distribution of F, we approximate the p-quantile using the empirical distribution function $\hat{F}_n(x)$, which is calculated as following,

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} IX_i \leq x \tag{23}$$

Next, we use the Dvoretzky-Kiefer-Wolfowitz (DKW) Inequality (Theorem 5.28),

$$P\left(\sup_x \left|\hat{F}(x) - F(x)\right| > \epsilon\right) \leq 2e^{-2n\epsilon^2} \tag{24}$$

This inequality gives an upper bound on the probability that the empirical distribution function $\hat{F}_n(x)$ deviates from the true distribution function $F(x)$ by more than ϵ.

Using the DKW inequality, we can create a confidence set for the distribution function F. We define the lower and upper limits as follows.

$$L(x) = max\{\hat{F}(x) - \epsilon_n, 0\} \tag{25}$$

$$U(x) = min\{\hat{F}(x) + \epsilon_n, 1\} \tag{26}$$

From this, it follows that.

$$P(L(x) \leq F(x) \leq U(x) \text{ for all } x) \geq 1 - \alpha \tag{27}$$

Next, we calculate the margin error, considering the confidence level $1-\alpha$. To achieve this, we set the probability in the DKW inequality to α.

$$P\left(\sup_x \left|\hat{F}(x) - F(x)\right| > \epsilon\right) \leq \alpha \tag{28}$$

From this we get that $2e^{-2n\epsilon^2} = \alpha$. Now, by setting the probability in the DKW inequality to α, we can solve for ϵ to find the margin error that ensures the confidence level.

$$\begin{aligned}
2e^{-2n\epsilon^2} &= \alpha \\
\ln(2e^{-2n\epsilon^2}) &= \ln(\alpha) \\
\ln(2) - 2n\epsilon^2 &= \ln(\alpha) \\
-2n\epsilon^2 &= \ln(\alpha) - \ln(2) \\
\epsilon^2 &= \frac{\ln(2/\alpha)}{2n} \\
\epsilon &= \sqrt{\frac{\ln(2/\alpha)}{2n}}
\end{aligned} \tag{29}$$

Now that we have found the margin of error ϵ, we can define the confidence interval for the estimated quantile p. This interval is given by.

$$[L(p), U(p)] \tag{30}$$

By substituting the expressions for the lower L and upper U bounds, we can express the confidence interval as.

$$[X_{(p-\epsilon)}, X_{(p+\epsilon)}] \tag{31}$$

Then finally, We substitute $\epsilon = \sqrt{\frac{\ln(2/\alpha)}{2n}}$ into the interval.

$$[X_{(p-\sqrt{\frac{\ln(2/\alpha)}{2n}})}, X_{(p+\sqrt{\frac{\ln(2/\alpha)}{2n}})}] \tag{32}$$

The following diagram illustrates the confidence interval, showing that with confidence $1 - \alpha$, the true p-quantile lies within the range.