Introduction to Data Science group assignment 2

Max Callenmark Kasper Lindkvist Eskil Worm Forss
Rami Abou Zahra Oskar Ådahl

Division Of Work

Task 1: Oskar Ådahl
Task 2: Max Callenmark
Task 3: Kasper Lindkvist
Task 4: Rami Abou Zahra
Task 5: Eskil Worm Forss

1)

The negative log-likelihood loss is given by

$$\begin{aligned}
-\sum_{i=1}^{n} \log(f_{Y|X}(Y_i, X_i)) &= -\sum_{i=1}^{n} \log\left(\frac{\lambda(X_i)^{Y_i} e^{-\lambda(X_i)}}{Y_i!}\right) \\
&= -\sum_{i=1}^{n} (Y_i \log(\lambda(X_i)) - \lambda(X_i) - \log(Y_i!)) \\
&= -\sum_{i=1}^{n} \left(Y_i(\alpha X_i + \beta) - e^{\alpha X_i + \beta} - \log(Y_i!)\right) \\
&= \sum_{i=1}^{n} \left(e^{\alpha X_i + \beta} - Y_i(\alpha X_i + \beta)\right) - \sum_{i=1}^{n} \log(Y_i!).
\end{aligned}$$

We see that the right sum does not depend on α or β, so we only need to minimize the left sum.

2)

distribution function of $\hat{\theta}$:

Since $X_1, X_2, \ldots, X_n$ are IID we have that :

$$F_{\hat{\theta}}(\hat{\theta}) = \left(F_X(\hat{\theta})\right)^n$$

since $X_i \sim U(0, \theta)$ we have that the distribution function of F_X is given by:

$$F_X(x) = \begin{cases} 0 \quad \text{for} \quad x < 0 \\ \dfrac{x}{\theta} \quad \text{for} \quad 0 \leq x \leq \theta \\ 1 \quad \text{for} \quad x > \theta \end{cases}$$

So the distribution function for $\hat{\theta}$ is then given by:

$$F_{\hat{\theta}}(\hat{\theta}) = \begin{cases} 0 \quad \text{for} \quad \hat{\theta} < 0 \\ (\frac{\hat{\theta}}{\theta})^n \quad \text{for} \quad 0 \leq \hat{\theta} \leq \theta \\ 1 \quad \text{for} \quad \hat{\theta} > \theta \end{cases}$$

Calculate bias($\hat{\theta}$):

The bias($\hat{\theta}$) is given by $E[\hat{\theta}] - \theta$, so we start by finding the density function for $\hat{\theta}$:

$$f_{\hat{\theta}}(\hat{\theta}) = \frac{d}{d\hat{\theta}} F_{\hat{\theta}}(\hat{\theta}) = \frac{d}{d\hat{\theta}} \frac{\hat{\theta}^n}{\theta^n} = n\frac{\hat{\theta}^{n-1}}{\theta^n}$$

Then we find $E[\hat{\theta}]$:

$$E[\hat{\theta}] = \int \hat{\theta} f_{\hat{\theta}} \; d\hat{\theta} = \int_0^{\theta} \hat{\theta} \cdot n\frac{\hat{\theta}^{n-1}}{\theta^n} \; d\hat{\theta} = \frac{n}{\theta^n} \int_0^{\theta} \hat{\theta} \; d\hat{\theta} = \frac{n}{n+1} \cdot \theta$$

Then we can calculate **bias**($\hat{\theta}$):

$$\textbf{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta = \frac{n}{n+1} \cdot \theta - \theta = \left(\frac{n}{n+1} - 1\right) \cdot \theta = \left(\frac{-1}{n+1}\right) \cdot \theta = -\frac{\theta}{n+1}$$

Calculate $\mathbf{se}(\hat{\theta})$:

The standard error of $\hat{\theta}$ is given by $\sqrt{\text{Var}(\hat{\theta})}$, so we start by finding $\text{Var}(\hat{\theta})$:

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

$$E[\hat{\theta}^2] = \int \hat{\theta}^2 f_{\hat{\theta}} \, d\hat{\theta} = \int_0^\theta \hat{\theta}^2 n \frac{\hat{\theta}^{n-1}}{\theta^n} \, d\hat{\theta} = \frac{n}{\theta^n} \int_0^\theta \hat{\theta}^{n+1} \, d\hat{\theta} = \frac{n}{n+2} \cdot \theta^2 \implies$$

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

So the $\mathbf{se}(\hat{\theta})$ is then given by:

$$\mathbf{se}(\hat{\theta}) = \sqrt{\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)\theta^2}$$

Calculate $\mathbf{MSE}(\hat{\theta})$:

We have that $\mathbf{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = (\mathbf{se}(\hat{\theta}))^2 + (\mathbf{bias}(\hat{\theta}))^2 = \text{Var}(\hat{\theta}) + (\mathbf{bias}(\hat{\theta}))^2$. We have already calculated $\text{Var}(\hat{\theta})$ and $\mathbf{bias}(\hat{\theta})$ so the $\mathbf{MSE}(\hat{\theta})$ is given by:

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

3)

a)

The cumulative distribution function F is:

$$F(x) = \int_{-\infty}^x p(t)dt = \int_{-\frac{\pi}{2}}^x \frac{1}{2}\cos(t)dt = \left[\frac{1}{2}\sin(t)\right]_{-\frac{\pi}{2}}^x = \frac{1}{2}\left(\sin(x) - \sin(-\frac{\pi}{2})\right) = \frac{\sin(x)+1}{2}$$

So, $F(x) = \dfrac{\sin(x)+1}{2}$, for $-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}$

b)

The inverse distribution function F^{-1} is:

$$F(x) = \frac{\sin(x)+1}{2} = y \Rightarrow 2y = \sin(x)+1 \Rightarrow \sin(x) = 2y+1 \Rightarrow x = \arcsin(2y-1) = F^{-1}(y)$$

So, $F^{-1}(y) = \arcsin(2y-1)$, for $0 < y < 1$, since arcsin takes arguments $[-1, 1]$.

c)

We want density function $g(x)$ such that $p(x) \leq Mg(x)$ for some constant $M > 0$, $\forall x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. A natural choice of distribution for g is the uniform distribution over the support.

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

Now $p(x) = \frac{1}{2}cos(x) \leq \frac{1}{2} * 1 = \frac{1}{2}$, so we choose M such that $Mg(x) = \frac{1}{2}$.

$$Mg(x) = M * \frac{1}{\pi} = \frac{M}{\pi} = \frac{1}{2} \Rightarrow M = \frac{\pi}{2} > 0$$

So, we choose $g(x) = \frac{1}{\pi}$ and $M = \frac{\pi}{2}$.

4)

In order to show that this is a Markov process, we need to show that the Markov property holds, i.e

$$P(X_n \mid X_{n-1}) = P(X_n \mid X_{n-1}, \dots, X_0)$$

Since X_n is the max value, we really only need to know the previous value of X in order to determine if X_n is larger than X_{n-1}. The filtration is increasing, and thus the Markov property holds.

In order to determine the transition matrix, we start off by listing the state space, which in this case is $\{0, 1, 2, 3\}$. The transitions p_{ij} are the probabilities of moving from state i to state j in one step.

Since it is impossible to move from a higher state to a lower state, for all $j < i$ the probability is 0. Likewise, if $j > i$, then the probabilities are just the ones that are given in the problem. However, if $j = i$ then we may have observed all the values lower than i and hence not updated the max-value, or we just observed i. This yields the following transition matrix:

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

5)

What we want to find a confidence interval for is:

$$q = \inf_x x : F(x) \geq p$$

by definition we have that $F(q) \geq p$ and $x < q \implies F(x) < p$ now we define $\hat{F}_n^{-1}(x)$ by

$$\hat{F}_n^{-1}(x) = \inf x : \hat{F}_n(x) \geq p$$

If we assume the event $A = \sup_x |\hat{F}_n(x) - F(x)| \leq \varepsilon$ to be true (i.e for $\omega \in \Omega$ such that the statement holds) we get that.

$$x < q \implies (F(x) < p \implies \hat{F}_n(x) < p : +\varepsilon)$$

$$x \geq q \implies (F(x) \geq p \implies \hat{F}_n(x) \geq p : -\varepsilon)$$

then we get that because $\hat{F}_n^{-1}$ is increasing

$$\hat{F}_n^{-1}(p - \varepsilon) \leq q \leq \hat{F}_n^{-1}(p + \varepsilon)$$

this is an interval for which q must be in (under our assumption) hence it is a confidence interval with confidence at least $\mathbb{P}(A)$. Now if we choose some confidence-level $1 - \alpha$ we can set $\varepsilon = \sqrt{\frac{1}{2n} \cdot \ln \frac{2}{\alpha}}$ to get that, by using DKW

$$\mathbb{P}(\sup_x |\hat{F}_n(x) - F(x)| \leq \varepsilon) > 1 - \alpha$$

then finally

$$[\hat{F}_n^{-1}(p - \varepsilon), ; \hat{F}_n^{-1}(p + \varepsilon)]$$

is our confidence interval for q.