Introduction to Data Science
Assignment 2 (Group 3)

Kartikey Sharma, Leire Unciti, Aviral Jain, Denice Wikström,
Alexander Ares Nilsson

November 08, 2024

1 Consider a supervised learning problem where we assume that $Y|X$ is Poisson distributed. That is, the conditional density of $Y|X$ is given by:- $f_{Y|X}(y,x) = \frac{\lambda^y e^{-\lambda}}{y!}, \lambda(x) = exp(\alpha x + \beta)$ Here α is a vector (slope) and β is a number (intercept). Follow the calculations from Section 4.2.1 to derive a loss that needs to be minimized with respect to α and β. Note: do we really need the factorial term?

Solution

The negative log-likelihood is given by:

$= -\sum_{i=1}^{n} ln(f_{X,Y}(X_i, Y_i))$

$= -\sum_{i=1}^{n} ln(f_{Y|X}(Y_i|X_i)) - \sum_{i=1}^{n} ln(f_X(X_i))$

$= -\sum_{i=1}^{n} ln(\frac{\lambda^{Y_i} e^{-\lambda}}{Y!}) - \sum_{i=1}^{n} ln(f_x(X_i))$

To minimize the loss with respect to α and β we can ignore both F_x and $Y!$ as they do not depend on the parameters.

$= -\sum_{i=1}^{n} ln(\lambda^{Y_i} e^{-\lambda})$

$= -\sum_{i=1}^{n} Y_i.ln(\lambda) - \lambda = \sum_{i=1}^{n} \lambda - Y_i.ln(\lambda) = \sum_{i=1}^{n} e^{\alpha X_i + \beta} - Y_i(\alpha X_i + \beta)$ +

The loss function to minimize is: $\sum_{i=1}^{n} e^{\alpha X_i + \beta} - Y_i(\alpha X_i + \beta)$

---

2 Let X_1,....,X_n be IID from Uniform(0,θ). Let $\hat{\theta}$ = max(X_1,....,X_n). First, find the distribution function of $\hat{\theta}$. Then compute the bias($\hat{\theta}$), se($\hat{\theta}$) and MSE($\hat{\theta}$).

Solution

CDF of $\hat{\theta}$

$= F_{\hat{\theta}}(x)$

$= P(\hat{\theta} \leq x)$

$= P(max(X_1, X_2, ..., X_n) \leq x)$

$=$ We know $X_i \sim Uniform(0, \theta) = F_{X_i}(x) = P(X_i \leq x) = \frac{x}{\theta}$ for $0 \leq x \leq \theta$

Since X_i are IID

$= P(max(X_1, ..., X_n) \leq x)$

$= P(X_1 \leq x).....P(X_n \leq x)$

Therefore $F_{\hat{\theta}}(x) = \begin{cases} 0, x < 0 \\ (\frac{x}{\theta})^n, 0 \leq x \leq n \\ 1, x > \theta \end{cases}$

PDF of $\hat{\theta}$

$= f_{\hat{\theta}}(x) = \frac{d}{dx}F_{\hat{\theta}}(x) = \frac{n}{\theta^n}x^{n-1}$, $0 \leq x \leq \theta$

E$[\hat{\theta}]$

$= \int_0^\theta x.f_{\hat{\theta}}(x)dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{\theta^n}[\frac{x^{n+1}}{n+1}]_0^\theta = \frac{n\theta}{n+1}$

E$[\hat{\theta}^2]$

$= \int_0^\theta x^2 f_{\hat{\theta}}(x)dx = \int_0^\theta x^2.\frac{n}{\theta^n}.x^{n-1}dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx = \frac{n\theta^2}{n+2}$

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

$SE(\hat{\theta}) = \sqrt{E[\hat{\theta}^2] - (E[\hat{\theta}])^2} = \sqrt{\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}} = \sqrt{\theta^2(\frac{n}{n+2} - \frac{n^2}{(n+1)^2})}$

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

---

3 Consider the continuous distribution with density.

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

(a) Find the distribution function F.

(b) Find the inverse distribution function F^{-1}.

(c) To sample using an Accept-Reject sampler, Algorithm 1, we need to find a density g such that $p(x) \leq Mg(x)$ for some $M > 0$. Find such a density g and find the value of M.

Solution

(a) CDF $F(x)$ can be calculated as

$= \int_{-\pi/2}^x p(t)dt$

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

$$F\_X(x) = \begin{cases} 0 \\ 1/2(\sin(x) + 1), -pi/2 <= x <= pi/2 \\ 1 \end{cases}$$

$= [\frac{1}{2}sin(t)]^{x}_{-\pi/2}$

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

(b) To find the inverse distribution $F^{-1}(y)$

Let $y = F(x)$

$\implies$ y = 1+sin(x)/2

$\implies$ 2y = 1 + sin(x)

$\implies$ (2y - 1) = sin(x)

$\implies$ x = arcsin(2y - 1)

so,

$\mathrm{F}^{-1}(y) = arcsin(2y - 1)$ domain?

(c) We know that max(p(x)) = $\frac{1}{2}$

g(x) can be a uniform distribution over the interval $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$

where $g(x) = \frac{1}{\pi}$ (since this integrates to 1 over the interval)

We want $p(x) \leq Mg(x)$ which implies:-

$\implies$ 1/2cos(x) $\leq M.1/\pi$

$\implies$ M $\geq \frac{\pi}{2}$

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

---

4 Let Y_1,Y_2,....,Y_n be a sequence of IID discrete random variables, where $\mathbb{P}(Y_i = 0) = 0.1$, $\mathbb{P}(Y_i = 1) = 0.3$, $\mathbb{P}(Y_i = 2) = 0.2$, and $\mathbb{P}(Y_i = 3) = 0.4$. Let $X_n = \max\{Y_1,....,Y_n\}$. Let $X_0 = 0$ and verify that X_0,X_1,....,X_n is a Markov chain. Find the transition matrix P.

Solution

We are given a sequence of discrete random variables Y_1, Y_2, Y_n with probabilities

$P(Y_i = 0) = 0.1$

$P(Y_i = 1) = 0.3$

$P(Y_i = 2) = 0.2$

$P(Y_i = 3) = 0.4$

A sequence $X_n = max\{Y_1, ..., Y_n\}$ where $X_0 = 0$. We want to show that $X_0, X_1, ..., X_n$ is a Markov chain and find the transition matrix P for this chain. Possible states of X_n are 0,1,2,3 matching the range of values that Y_i can take. Since $X_{n+1} = max(X_n, Y_{n+1})$, the next state X_{n+1} depends only on the current state X_n and the value of Y_{n+1}. This satisfies the Markov property. Let's define the transition matrix P where $P_j = P(X_{n+1} = j|X_n = i)$ for $i, j \in \{0, 1, 2, 3\}$

Case i = 0

$X_n = 0$. Since $X_{n+1} = max\{0, Y_{n+1}\} = Y_{n+1}$. The distribution of X_{n+1} is the same as the distribution of Y.

$P_{0j} = P(X_{n+1} = j | X_n = 0) = P(Y = j)$

$P_{00} = 0.1$

$P_{01} = 0.3$

$P_{02} = 0.2$

$P_{03} = 0.4$

Case i = 1

$X_n = 1$ then $X_{n+1} = max\{1, Y_{n+1}\}$

if $Y_{n+1} = 0$ then $X_{n+1} = 1$

if $Y_{n+1} = 1$ then $X_{n+1} = 1$

if $Y_{n+1} = 2$ then $X_{n+1} = 2$

if $Y_{n+1} = 3$ then $X_{n+1} = 3$

This means:

$P_{10} = 0$

$P_{11} = 0.1 + 0.3 = 0.4$

$P_{12} = 0.2$

$P_{13} = 0.4$

Case i = 2

$X_n = 2$ then $X_{n+1} = max\{2, Y_{n+1}\}$

if $Y_{n+1} = 0 or Y_{n+1} = 1$ then $X_{n+1} = 2$

if $Y_{n+1} = 2$ then $X_{n+1} = 2$

if $Y_{n+1} = 3$ then $X_{n+1} = 3$

So we get:-

$P_{20} = 0$

$P_{21} = 0$

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

$P_{23} = 0.4$

Case i = 3

Since 3 is the maximum value X_{n+1} will always be 3 once it reaches 3. So, $P_{30} = 0$, $P_{31} = 0$, $P_{32} = 0$, $P_{33} = 1$.
Based on these cases, for the possible states of X_{ij} the transition matrix will therefore be:-

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

$+$

5 Let X_1,....,X_n be IID from some distribution F that is unknown. Let $\hat{F_n}$ be the empirical distribution function, use this to find an estimate of the quantile p of F. Use theorem 5.28 to find a confidence interval for p.

Solution

Theorem 5.28

$X_1, ..., X_n(IID)$

For any $\epsilon \geq 0 : P(sup(|\hat{F_n}(x) - F(x)|) \leq 2e^{-2n\epsilon^2}$

To estimate the p-quantile of F using $\hat{F_n}$, we define the empirical quantile value $\hat{x}p$ such that $\hat{F_n}(\hat{x}p) = p$. The empirical quantile is obtained by ordering the sample $X_1, ..., X_n$ and selecting the value at position floor(np).

Confidence interval:

Apply theorem 5.28

$2e^{-2n\epsilon^2} = \alpha \Leftrightarrow -2n\epsilon^2 = ln(\frac{\alpha}{2}) \Leftrightarrow \epsilon = \sqrt{\frac{ln(2/\alpha)}{2n}}$

With probability at least $1 - \alpha$, $|\hat{F_n}(xp) - F(xp)| \leq \epsilon$

So, the confidence interval for the quantile xp can be approximated by identifying the interval of value x around $\hat{x}p$ such that $\hat{F_n}(x)$ lies within $\{p - \epsilon, p + \epsilon\}$
