Maximum Likelihood Estimation with Indicator Function

Asked 2 years, 10 months ago Active 2 years, 10 months ago Viewed 2k times

4

2

I need to solve this exercise from the book below.

Mathematical Statistics, Knight (2000)

Problem 6.17

Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with frequency function

$$f(x;\theta) = \begin{cases} \theta & \text{for } x = -1. \\ (1-\theta)^2\theta^x & \text{for } x = 0, 1, 2, \ldots \end{cases}$$

(a) Find the Cramer-Rao lower bound for unbiased estimators based on $X_1, \ldots, X_n$.

(b) Show that the maximum likelihood estimator of θ based on $X_1, \ldots, X_n$ is

$$\hat{\theta}_n = \frac{2\sum_{i=1}^n I_{(X_i=-1)} + \sum_{i=1}^n X_i}{2n + \sum_{i=1}^n X_i}$$

and show that $\{\hat{\theta}_n\}$ is consistent for θ.

(c) Show that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2(\theta))$ and find the value of $\sigma^2(\theta)$. Compare $\sigma^2(\theta)$ to the Cramer-Rao lower bound in part (a).

No clue on how to solve (a) or (c).

I started to solve (b) but I can't seem to arrive at the desired solution. I'm getting this:

$$\mathcal{L} = \prod_{i=1}^n (1-\theta)^2 \theta^{x_i I_{(X_i\geq 0)} + I_{(X_i=-1)}}$$
$$\mathcal{L} = (1-\theta)^{2\sum_{i=1}^n I_{(X_i\geq 0)}} \theta^{\sum_{i=1}^n x_i I_{(X_i\geq 0)} + \sum_{i=1}^n I_{(X_i=-1)}}$$
$$\log\mathcal{L} = 2\sum_{i=1}^n I_{(X_i\geq 0)} \log(1-\theta) + \sum_{i=1}^n x_i I_{(X_i\geq 0)} \log\theta + \sum_{i=1}^n I_{(X_i=-1)} \log\theta$$

FOC

$$0 = -\frac{2\sum_{i=1}^{n} I_{(X_i \geq 0)}}{1-\theta} + \frac{\sum_{i=1}^{n} x_i I_{(X_i \geq 0)}}{\theta} + \frac{\sum_{i=1}^{n} I_{(X_i = -1)}}{\theta}$$

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n} I_{(X_i=-1)} + \sum_{i=1}^{n} x_i I_{(X_i \geq 0)}}{\sum_{i=1}^{n} I_{(X_i=-1)} + 2\sum_{i=1}^{n} I_{(X_i \geq 0)} + \sum_{i=1}^{n} x_i I_{(X_i \geq 0)}}$$

which differs from the result I'm given...

Any help would be greatly appreciated.

probability maximum-likelihood

edited Dec 2 '16 at 7:06
heropup
70k 9 67 112

asked Dec 2 '16 at 5:35
Slakeon
43 4

2 Answers

4

The solution that is provided suggests that for a sample $\boldsymbol{x} = (x_1, \ldots, x_n)$, we should define a random variable $K = \sum_{i=1}^{n} \mathbb{1}_{x_i=-1}$; that is, K counts the number of observations that equal to -1. Then

$$\begin{aligned}\mathcal{L}(\theta \mid \boldsymbol{x}) &= \theta^K \prod_{j=1}^{n-K} (1-\theta)^2 \theta^{x_{[j]}} \\ &= \theta^K (1-\theta)^{2(n-K)} \theta^{\sum_{j=1}^{n-K} x_{[j]}}\end{aligned}$$

where $x_{[j]}$ represents the j^{th} observation of $\boldsymbol{x}$ that is nonnegative. But note that

$$\sum_{j=1}^{n-K} x_{[j]} = K + \sum_{i=1}^{n} x_i = K + n\bar{x}.$$

So we may write the log-likelihood as

$$\ell(\theta \mid \boldsymbol{x}) = 2(n-K)\log(1-\theta) + (2K + n\bar{x})\log\theta.$$

Thus the log-likelihood is maximized at a critical point satisfying

$$0 = \frac{\partial \ell}{\partial \theta} = -\frac{2(n-K)}{1-\theta} + \frac{2K + n\bar{x}}{\theta},$$

or

$$\hat{\theta} = \frac{2K + n\bar{x}}{n(2+\bar{x})}.$$

This is equivalent to the stated solution (just written more compactly).

You may find that working with K and avoiding the unnecessary use of additional indicator functions (you really only need one, namely $1_{x_i=-1}$) will reduce your chances of making errors. Please feel free to attempt the other parts of the question.

If you find it difficult to follow the above solution, it is helpful to consider a numeric example. Suppose you are given the sample

$$\boldsymbol{x} = (-1, 0, 1, 3, -1, 5, -1).$$

Then $n = 7$, $K = 3$, and the sample total is $n\bar{x} = 6$. We observe that $K + n\bar{x} = 3 + 6 = 9$, which is equal to the sum of nonnegative observations $\sum_{j=1}^{n-K} x_{[j]} = 0 + 1 + 3 + 5 = 9$.

The resulting likelihood function is

$$\mathcal{L}(\theta \mid \boldsymbol{x}) = \theta^3 (1-\theta)^{2(7-3)} \theta^{0+1+3+5} = \theta^{12}(1-\theta)^8.$$

This is maximized when $\hat{\theta} = 12/(8+12) = 3/5$.

edited Dec 2 '16 at 6:24

answered Dec 2 '16 at 6:15

heropup
70k 9 67 112

Thanks! Your approach was great! Once I have that, how should I work with K to find the Fisher Information Matrix? – Slakeon Dec 2 '16 at 6:27

@Slakeon You've got the first derivative of the log-likelihood already. Now calculate the second derivative, and compute its expectation. The tricky part is to notice that

$$K \mid \theta \sim \operatorname{Binomial}(n, p = \theta),$$

because by definition $\Pr[X = -1 \mid \theta] = \theta$; thus $K \mid \theta$ which counts the number of such outcomes, is a binomial random variable and its expectation is $\operatorname{E}[K \mid \theta] = n\theta$. What is $\operatorname{E}[n\bar{X}]$, the expected sample total? – heropup Dec 2 '16 at 6:41

Thanks again! Nice observation, never would have thought that! – Slakeon Dec 2 '16 at 6:46

3

$$L(\theta) = \prod_{i=1}^{n} \theta^{I_{x_i=-1}} (1-\theta)^{2 I_{x_i \geq 0}} \theta^{x_i I_{x_i \geq 0}} .$$

$$\ell(\theta) = \log L(\theta) = (\log \theta) \sum_{i=1}^{n} (I_{x_i=-1} + x_i I_{x_i \geq 0}) + 2(\log(1-\theta)) \sum_{i=1}^{n} I_{x_i \geq 0}$$

$$\ell'(\theta) = \frac{1}{\theta} \sum_{i=1}^{n} (I_{x_i=-1} + x_i I_{x_i \geq 0}) - \frac{2}{1-\theta} \sum_{i=1}^{n} I_{x_i \geq 0} = \frac{A}{\theta} - 2\frac{B}{1-\theta}$$

$$= 0 \text{ if and only if } A(1-\theta) - 2B\theta = 0,$$

$$\text{and that holds precisely if } A = 2B\theta + A\theta = (A+2B)\theta, \text{ so}$$

$$\theta = \frac{A}{A+2B} = \frac{\sum_{i=1}^{n} (I_{x_i=-1} + x_i I_{x_i \geq 0})}{\sum_{i=1}^{n} (I_{x_i=-1} + x_i I_{x_i \geq 0}) + 2\sum_{i=1}^{n} I_{x_i \geq 0}}.$$

answered Dec 2 '16 at 6:38

Michael Hardy
215k 23 210 491

How would you get to the desired result proceeding like this? – Slakeon Dec 2 '16 at 6:56

@Slakeon Recalling that we defined $K = \sum_{i=1}^{n} 1_{x_i=-1}$, then you can see that the numerator is is simply $K + \sum_{j=1}^{n-K} x_{[j]} = K + K + n\bar{x} = 2K + n\bar{x}$, as shown in my answer. The denominator simplifies in a similar fashion, noting that $\sum_{i=1}^{n} 1_{x_i \geq 0} = n - K$. This illustrates the power of a careful choice of convenient and concise notation. – heropup Dec 2 '16 at 6:58

@Slakeon : More tomorrow . . . – Michael Hardy Dec 2 '16 at 7:23