2. Recap: Maximum Likelihood Estimators and Fisher information

Instructions:

For each of the following distributions, compute the maximum likelihood estimator based on n i.i.d. observations $X_1, \ldots, X_n$ and the Fisher information, if defined. If it is not, enter **DNE** in each applicable input box.

(a)

3/3 points (graded)

$$X_i \sim \text{Ber}(p), \quad p \in (0,1)$$

(Enter **barX_n** for the sample average $\overline{X}_n$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Maximum likelihood estimator $\hat{p} =$ barX_n

□ **Answer:** barX_n

Hint: Use the definition of Fisher information that leads to the shorter computation.

(If the Fisher information is not defined, enter **DNE**.)

Fisher information $I(p) =$ 1/(p*(1-p))

$\frac{1}{p\cdot(1-p)}$

□ **Answer:** 1/(p*(1-p))

Use Fisher Information to find the asymptotic variance $V(\hat{p})$ of the MLE $\hat{p}$.

$V(\hat{p}) =$ (p*(1-p))

$(p \cdot (1-p))$

□ **Answer:** p*(1-p)

STANDARD NOTATION

Solution:

The likelihood for one sample can be written as

$$L_1(X_1, p) = p^{X_1}(1-p)^{1-X_1}.$$

That means that the log likelihood for one sample is

$$\ell_1(X_1, p) = X_1 \ln(p) + (1 - X_1)\ln(1-p),$$

and for n samples this yields

$$\ell_n\left(X_1, \ldots, X_n, p\right) = \ln(p) \sum_{i=1}^{n} X_i + \ln(1-p)\left(n - \sum_{i=1}^{n} X_i\right).$$

Differentiating with respect to p yields

$$\frac{\partial}{\partial p} \ell_n(p) = \frac{1}{p} \sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right).$$

Setting this to zero to find the maximum likelihood estimator $\hat{p}$ then gives

$$\begin{aligned} 0 = & \quad (1-\hat{p}) \sum_{i=1}^{n} X_i - \hat{p}\left(n - \sum_{i=1}^{n} X_i\right) \\ \Longleftrightarrow \hat{p} = & \quad \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n. \end{aligned}$$

That this is indeed the global maximum can be verified by checking the concavity of the log likelihood, which in turn can be seen from the negativity of the second derivative, which we compute next.

The second derivative for one sample is

$$\frac{\partial^2}{\partial p^2} \ell_1(p) = -\frac{X_1}{p^2} - \frac{1-X_i}{(1-p)^2} < 0,$$

and by $\mathbb{E}\left[X_i\right] = p$, we obtain

$$I(p) = -\mathbb{E}\left[\frac{\partial^2}{\partial p^2} \ell_1(p)\right] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

Take the inverse of this to obtain the (asymptotic) variance: $V(\hat{p}) = p(1-p)$.

提交 你已经尝试了2次（总共可以尝试3次）

Answers are displayed within the problem

(b)

3/3 points (graded)

$$X_i \sim \text{Poiss}(\lambda), \quad \lambda > 0,$$

which means that each X_i has distribution

$$\mathbf{P}_\lambda(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

(Enter **barX_n** for the sample average $\overline{X}_n$.)

Maximum likelihood estimator $\hat{\lambda} =$ barX_n **Answer:** barX_n

(If the Fisher information is not defined, enter **DNE**.)

Fisher information $I(\lambda) =$ 1/lambda

$\frac{1}{\lambda}$

Answer: 1/lambda

Use Fisher Information to find the asymptotic variance $V(\hat{\lambda})$ of the MLE $\hat{\lambda}$.

$V(\hat{\lambda}) =$ lambda

λ

Answer: lambda

STANDARD NOTATION

Solution:

The likelihood for one sample can be written as

$$L_1(X_1, \lambda) = e^{-\lambda} \frac{\lambda^{X_1}}{X_1!}.$$

That means that the log likelihood for one sample is

$$\ell_1(X_1, \lambda) = -\lambda + X_1 \ln(\lambda) - \ln(X_1!),$$

and for n samples this yields

$$\ell_n(X_1, \ldots, X_n, p) = -n\lambda + n\overline{X}_n \ln(\lambda) - \sum_{i=1}^{n} \ln(X_i!).$$

Differentiating with respect to λ yields

$$\frac{\partial}{\partial \lambda} \ell_n(\lambda) = -n + n\overline{X}_n \frac{1}{\lambda}.$$

Setting this to zero to find the maximum likelihood estimator $\hat{\lambda}$ then gives

$$\hat{\lambda} = \overline{X}_n.$$

That this is indeed the global maximum can be verified by checking the concavity of the log likelihood, which in turn can be seen from the negativity of the second derivative, which we compute next.

The second derivative for one sample is

$$\frac{\partial^2}{\partial \lambda^2} \ell_1(\lambda) = -\frac{X_i}{\lambda^2} < 0,$$

and by $\mathbb{E}[X_i] = \lambda$, we obtain

$$I(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial \lambda^2} \ell_1(\lambda)\right] = \frac{\lambda}{\lambda^2} = \lambda.$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

(c)

3/3 points (graded)

$$X_i \sim \text{Exp}(\lambda), \quad \lambda > 0,$$

which means that each X_1 has density

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

(Enter **barX_n** for $\overline{X}_n$ the sample average.)

Maximum likelihood estimator $\hat{\lambda} =$ 1/barX_n **Answer:** 1/barX_n (If the Fisher information is not defined, enter **DNE**.)

Fisher information $I(\lambda) =$ 1/lambda^2 **Answer:** 1/(lambda^2)

$\frac{1}{\lambda^2}$

Use Fisher Information to find the asymptotic variance $V(\hat{\lambda})$ of the MLE $\hat{\lambda}$.

$V(\hat{\lambda}) =$ lambda^2 **Answer:** lambda^2

λ^2

STANDARD NOTATION

Solution:

The likelihood for one sample can be written as

$$L_1(X_1, \lambda) = \lambda e^{-\lambda X_1}.$$

That means that the log likelihood for one sample is

$$\ell_1(X_1, \lambda) = \ln(\lambda) - \lambda X_1,$$

and for n samples this yields

$$\ell_n(X_1, \ldots, X_n, p) = n \ln(\lambda) - n\lambda \overline{X}_n.$$

Differentiating with respect to λ yields

$$\frac{\partial}{\partial \lambda} \ell_n(\lambda) = \frac{n}{\lambda} - n\overline{X}_n.$$

Setting this to zero to find the maximum likelihood estimator $\hat{\lambda}$ then gives

$$\hat{\lambda} = \frac{1}{\overline{X}_n}.$$

That this is indeed the global maximum can be verified by checking the concavity of the log likelihood, which in turn can be seen from the negativity of the second derivative, which we compute next.

The second derivative for one sample is

$$\frac{\partial^2}{\partial \lambda^2} \ell_1 (\lambda) = -\frac{1}{\lambda^2}.$$

and hence

$$I(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial \lambda^2} \ell_1 (\lambda)\right] = \frac{1}{\lambda^2}.$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

(d)

7.0/7 points (graded)

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \sigma^2 > 0,$$

which means that each X_1 has density

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Hint: Keep in mind that we consider σ^2 as the parameter, not σ. You may want to write $\tau = \sigma^2$ in your computation.

(Enter **barX_n** for the sample average $\overline{X}_n$ and **bar(X_n^2)** for the sample average of second moments $\overline{X_n^2}$.)

Maximum likelihood estimator $\hat{\mu} =$ barX_n   **Answer:** barX_n

(Enter **barX_n** for the sample average $\overline{X}_n$ and **bar(X_n^2)** for the sample average of second moments $\overline{X_n^2}$.)

Maximum likelihood estimator $\widehat{\sigma^2} =$ bar(X_n^2) - barX_n^2   **Answer:** bar(X_n^2)-(barX_n)^2

Hint: One of the formulas for Fisher information will lead to a much shorter computation.

(If the Fisher information is not defined, enter **DNE** for all boxes below.)

$[I(\mu, \sigma^2)]_{1,1} =$ 1/(sigma^2)   **Answer:** 1/(sigma^2)
$\frac{1}{\sigma^2}$

, $[I(\mu, \sigma^2)]_{1,2} =$ 0   **Answer:** 0
0

$[I(\mu, \sigma^2)]_{2,1} =$ 0   **Answer:** 0
0

, $[I(\mu, \sigma^2)]_{2,2} =$ 1/(2*sigma^4)   **Answer:** 1/(2*sigma^4)
$\frac{1}{2 \cdot \sigma^4}$

Using the Fisher Information you obtain above, what is the asymptotic variance $V(\widehat{\sigma^2})$ of the MLE $\widehat{\sigma^2}$?Compare this with your result from Homework 5 Problem 3.

$V\left(\widehat{\sigma^2}\right) =$ 2*sigma^4

$2 \cdot \sigma^4$

Answer: 2*sigma^4

STANDARD NOTATION

Solution:

The likelihood for one sample can be written as

$$L_1\left(X_1, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(X_1 - \mu\right)^2}{2\sigma^2}\right).$$

That means that the log likelihood for one sample is

$$\ell_1\left(X_1, \mu, \sigma^2\right) = -\frac{1}{2}\left(\ln\left(2\right) + \ln\left(\pi\right) + \ln\left(\sigma^2\right)\right) - \frac{\left(X_1 - \mu\right)^2}{2\sigma^2},$$

and for n samples this yields

$$\ell_n\left(X_1, \ldots, X_n, \mu, \sigma^2\right) = -\frac{n}{2}\left(\ln\left(2\pi\right) + \ln\left(\sigma^2\right)\right) - \sum_{i=1}^{n} \frac{\left(X_i - \mu\right)^2}{2\sigma^2}.$$

Differentiating with respect to μ yields

$$\frac{\partial}{\partial \mu} \ell_n\left(\mu, \sigma^2\right) = \sum_{i=1}^{n} \frac{X_i - \mu}{2\sigma^2}.$$

Setting this to zero to find the maximum likelihood estimator $\hat{\mu}$ then gives

$$\hat{\mu} = \overline{X}_n.$$

Similarly, differentiating the log likelihood with respect to σ^2, we find

$$\frac{\partial}{\partial \sigma^2} \ell_n\left(\mu, \sigma^2\right) = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{\left(X_i - \mu\right)^2}{2\sigma^4}.$$

Setting this to zero gives

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \hat{\mu}\right)^2 = \frac{1}{n}\left(X_i - \overline{X}_n\right)^2 = \overline{X_n^2} - \overline{X}_n^2.$$

That this is indeed the global maximum can be verified by checking the concavity of the log likelihood, which in turn can be seen from the negative definiteness of the Hessian, which we compute next.

We compute the second derivatives:

$$\begin{aligned} \frac{\partial^2}{\partial \mu^2} \ell_1\left(\mu, \sigma^2\right) &= -\frac{1}{\sigma^2}, \\ \frac{\partial^2}{\partial \mu \partial\left(\sigma^2\right)} &= -\frac{1}{\left(\sigma^2\right)^2}\left(X_1 - \mu\right), \end{aligned}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2} = \frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(X_1-\mu)^2,$$

from where negative definiteness can be checked by testing $\mathbf{tr}\, H\ell_1 < 0$ and $\mathbf{det} H\ell_1 > 0$. Moreover,

$$\begin{aligned} I(\mu, \sigma^2) &= -\begin{pmatrix} \mathbb{E}\left[\frac{1}{\sigma^2}\right] & \frac{\mathbb{E}[X_1-\mu]}{(\sigma^2)^2} \\ \frac{\mathbb{E}[X_1-\mu]}{(\sigma^2)^2} & \mathbb{E}\left[\frac{1}{2(\sigma^2)^2} - \frac{(X_1-\mu)^2}{(\sigma^2)^3}\right] \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}. \end{aligned}$$

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

(e)

6.0/6 points (graded)

X_i follows a shifted exponential distribution with parameters $a \in \mathbb{R}$ and $\lambda > 0$. That means each X_i has density

$$f_{a,\lambda}(x) = \lambda e^{-\lambda(x-a)} \mathbf{1}\{x \geq a\}, \quad x \in \mathbb{R}.$$

(Enter **barX_n** for the sample average $\overline{X}_n$, and if applicable, use **min_i(X_i)** for $\min_{1\leq i\leq n} X_i$).

Maximum likelihood estimator $\hat{a} =$ min_i(X_i) **Answer:** min_i(X_i)

Maximum likelihood estimator $\hat{\lambda} =$ 1/(barX_n - min_i(X_i)) **Answer:** 1/(barX_n-min_i(X_i))

Hint: Think of the effect of the indicator function on the derivatives. (If the Fisher information is not defined, enter **DNE** in all boxes below.
)

$[I(a,\lambda)]_{1,1} =$ DNE **Answer:** DNE , $[I(a,\lambda)]_{1,2} =$ DNE **Answer:** DNE

DNE DNE

$[I(a,\lambda)]_{2,1} =$ DNE **Answer:** DNE , $[I(a,\lambda)]_{2,2} =$ DNE **Answer:** DNE

DNE DNE

STANDARD NOTATION

Solution:

The likelihood for one sample can be written as

$$L_1(X_1, a, \lambda) = \lambda e^{-\lambda(X_i - a}\mathbf{1}\{X_i \geq a\}.$$

That means the likelihood for n samples is

$$L_n(X_1, \ldots, X_n, a, \lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n}(X_i - a)\right) \mathbf{1}\{\min X_i \geq a\}.$$

Considering two cases, we note that if $a > \min X_i$, then $\mathbf{1}\{\min X_i \geq a\} = 0$ and hence $L_n(a, \lambda) = 0$. On the other hand, if $a \leq \min X_i$, then $\mathbf{1}\{\min X_i \geq a\} = 1$, and

$$L_n(a, \lambda) = e^{n\lambda a} \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} X_i\right),$$

which is an increasing function of a. Hence, for any fixed value of λ, the maximum likelihood estimator for a is

$$\hat{a} = \min_{1 \leq i \leq n} X_i.$$

For this value of a, we can now optimize the log likelihood

$$\ln L_n(X_1, \ldots, X_n, \hat{a}, \lambda) = n\lambda(\lambda) - \lambda(n\overline{X}_n - n\hat{a}),$$

$$\frac{\partial}{\partial \lambda} \ln L_n(\hat{a}, \lambda) = \frac{n}{\lambda} - n(\overline{X}_n - \hat{a}).$$

Setting the first derivative to zero then yields

$$\hat{\lambda} = \frac{1}{\overline{X}_n - \min X_i}.$$

For this example, the Fisher information is not well-defined since the support of the distribution depends on the parameter a.

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

讨论

显示讨论

主题: Unit 3 Methods of Estimation:Homework 6 Maximum Likelihood Estimation and Method of Moments / 2. Recap: Maximum Likelihood Estimators and Fisher information