Homework 4

TV distance, KL-Divergence, Intro to MLE

1. Kullback-Leibler divergence

Homework due Jun 25, 2020 08:59 JST *Past Due*

Bookmark this page

Instructions:

For the following pairs of distributions $(\mathbf{P}, \mathbf{Q})$, compute the Kullback-Leibler divergence $\text{KL}(\mathbf{P}, \mathbf{Q})$.

If the KL divergence is $+\infty$ or $-\infty$, enter **+inf** or **-inf**.

(a)

1 point possible (graded)

$$\mathbf{P} = \mathcal{N}(a, \sigma^2), \quad \mathbf{Q} = \mathcal{N}(b, \sigma^2), \quad a, b \in \mathbb{R}, \ \sigma^2 > 0.$$

(If applicable, enter **ln(x)** for $\ln(x)$. Do NOT enter "log".)

$\text{KL}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** (a - b)^2/(2*sigma^2)

STANDARD NOTATION

Solution:

If we write $X \sim \mathbf{P}$, we can compute:

$$\begin{aligned}
\text{KL}(\mathbf{P}, \mathbf{Q}) &= \mathbb{E}_{\mathbf{P}}\left[\ln\left(\frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{-(X-a)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(X-b)^2}{2\sigma^2}\right)}\right)\right] \\
&= \mathbb{E}_{\mathbf{P}}\left[-\frac{(X-a)^2}{2\sigma^2} + \frac{(X-b)^2}{2\sigma^2}\right] \\
&= \frac{1}{2\sigma^2}\mathbb{E}_{\mathbf{P}}\left[2(a-b)(X-a) + (a-b)^2\right] \\
&= \frac{(a-b)^2}{2\sigma^2},
\end{aligned}$$

because $\mathbb{E}_{\mathbf{P}}[(X-a)] = 0$.

(b)

1 point possible (graded)

$$\mathbf{P} = \text{Ber}(a), \quad \mathbf{Q} = \text{Ber}(b), \quad a, b \in (0,1)$$

(If applicable, enter **ln(x)** for $\ln(x)$. Do NOT enter “log".)

$\text{KL}(\mathbf{P}, \mathbf{Q}) =$

Answer: a * ln(a/b) + (1-a) *ln((1-a)/(1-b))

STANDARD NOTATION

Solution:

If we write $X \sim \mathbf{P}$, $Y \sim \mathbf{Q}$, we have

$$\begin{aligned} \text{KL}(\mathbf{P}, \mathbf{Q}) = & \ \mathbf{P}(X=0) \ln \frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=0)} + \mathbf{P}(X=1) \frac{\mathbf{P}(X=1)}{\mathbf{P}(Y=1)} \\ = & \ a \ln \frac{a}{b} + (1-a) \ln \frac{1-a}{1-b}. \end{aligned}$$

(c)

2 points possible (graded)

$$P = \text{Unif}([0, \theta_1]), \quad Q = \text{Unif}([0, \theta_2]), \quad 0 < \theta_1 < \theta_2.$$

Hint: Note the support of each distribution when computing the expectation.

(If applicable, enter **ln(x)** for $\ln(x)$. Do NOT enter “log". If applicable, enter **theta_1** for θ_1 and **theta_2** for θ_2.)

$\text{KL}(\mathbf{P}, \mathbf{Q}) =$

Answer: ln(theta_2/theta_1)

$\text{KL}(\mathbf{Q}, \mathbf{P}) =$

Answer: inf

Solution:

We compute

$$\begin{aligned}\text{KL}(\mathbf{P},\ \mathbf{Q}) = & \ \mathbb{E}_{\mathbf{P}}\left[\ln\frac{\frac{1}{\theta_1}}{\frac{1}{\theta_2}}\right] \\ = & \ \ln\left(\frac{\theta_2}{\theta_1}\right).\end{aligned}$$

If we try to compute the KL divergence the other way round, we notice that $\mathbf{P}$ is not supported between for $\theta_1 < X < \theta_2$. We compute the expectation by integrating explicitly:

$$\begin{aligned}\text{KL}(\mathbf{Q},\ \mathbf{P}) = & \ \mathbb{E}_{\mathbf{Q}}\left[\ln\frac{q}{p}\right] \quad \text{where } p, q, \text{are the pdfs of } \mathbf{P}, \mathbf{Q} \text{ respectively} \\ = & \ \int_0^{\theta_1}\frac{1}{\theta_2}\ln\frac{1/\theta_2}{1/\theta_1}dx + \int_{\theta_1}^{\theta_2}\frac{1}{\theta_2}\ln\frac{1/\theta_2}{0}dx \\ = +\infty\end{aligned}$$

because the second term diverges to $+\infty$. **Remark:** In general, $\text{KL}(\mathbf{P},\mathbf{Q}) \neq \text{KL}(\mathbf{Q},\mathbf{P})$.

(d)

1 point possible (graded)

$$P = \text{Exp}(\lambda), \quad Q = \text{Exp}(\mu), \quad \lambda, \mu \in (0,\infty).$$

(If applicable, enter **ln(x)** for $\ln(x)$. Do NOT enter “log".)

$\text{KL}(\mathbf{P},\ \mathbf{Q}) =$ **Answer:** ln(lambda/mu) + mu/lambda - 1

STANDARD NOTATION

Solution:

If $X \sim P$, then

$$\begin{aligned}\text{KL}(\mathbf{P},\ \mathbf{Q}) = & \ \mathbb{E}_P\left[\ln\frac{\lambda e^{-\lambda x}}{\mu e^{-\mu x}}\right] \\ = & \ \mathbb{E}_P\left[\ln\frac{\lambda}{\mu} + (\mu-\lambda)X\right] \\ = & \ \ln\frac{\lambda}{\mu} + (\mu-\lambda)\frac{1}{\lambda} \\ = & \ \ln\frac{\lambda}{\mu} + \frac{\mu}{\lambda} - 1,\end{aligned}$$

because $\mathbb{E}_P[X] = \frac{1}{\lambda}$.

(a)

3 points possible (graded)

Compute the total variation distance between

$$\mathbf{P} = X \quad \text{and} \quad \mathbf{Q} = X + c, \qquad \text{where } X \sim \text{Ber}(p), p \in (0,1), \text{ and } c \in \mathbb{R}.$$

(If applicable, enter **abs(x)** for $|x|$. Simplify your answer to have the minimum number of absolute signs possible.)

For $c \notin \{-1, 0, 1\}$:

$\text{TV}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** 1

For $c = 0$:

$\text{TV}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** 0

For $c = 1$ or $c = -1$:

$\text{TV}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** 1/2*(1+abs(1-2*p))

Solution:

- For $c \notin \{-1, 0, 1\}$, the support of X and $X + c$ are disjoint, hence $\text{TV}(X, X + c) = 1$.
- For $c = 0$, by the definiteness property $\text{TV}(X, X) = 0$.
- For $c = 1$ (*resp.* $c = -1$), the support of X and $X + c$ intersect at $X = 1$ (*resp.* at $X = 0$). Hence

$$\begin{aligned} \text{TV}(X, X + c) &= \frac{1}{2}(|1 - p| + |p - (1 - p)| + |p|) \\ &= \frac{1}{2}(1 + |1 - 2p|) \qquad \text{where } c = 1, \text{ or } -1. \end{aligned}$$

(b)

2 points possible (graded)

Compute the total variation distance between

$$\mathbf{P} = \text{Ber}(p) \quad \text{and} \quad \mathbf{Q} = \text{Ber}(q), \quad \text{where } p, q \in [0,1].$$

(If applicable, enter **abs(x)** for $|x|$.)

$\text{TV}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** abs(p-q)

Let $X_1, \ldots, X_n$ be n i.i.d. Bernoulli random variables with some parameter $p \in [0,1]$, and $\bar{X}_n$ be their empirical average. Consider the total variation distance $\text{TV}(\text{Ber}(\bar{X}_n), \text{Ber}(p))$ between $\text{Ber}(\bar{X}_n)$ and $\text{Ber}(p)$ as a function of the random variable $\bar{X}_n$, and hence a random variable itself. Does $\text{TV}(\text{Ber}(\bar{X}_n), \text{Ber}(p))$ necessarily converge in probability to a constant? If yes, enter the constant below; if not; enter DNE.

$\text{TV}(\text{Ber}(\bar{X}_n), \text{Ber}(p)) \xrightarrow[n\to\infty]{(\mathbf{P})}$ **Answer:** 0

Solution:

To compute the total variation distance between two Bernoulli variables, again use the formula relating TV to the pmfs:

$$\begin{aligned}\text{TV}(\text{Ber}(p), \text{Ber}(q)) &= \frac{1}{2}[|p-q| + |(1-p)-(1-q)|] \\ &= |p-q|.\end{aligned}$$

Let $X_1, \ldots, X_n$ be n i.i.d. Bernoulli random variables with some parameter $p \in [0,1]$, and $\bar{X}_n$ be their empirical average. By the Law of Large Numbers, we know that $\bar{X}_n$ will converge to p. Now, imagine another Bernoulli distribution with parameter $\bar{X}_n$, say $\text{Ber}(\bar{X}_n)$. What can we infer about the two distributions? What is the total variation distance between $\text{Ber}(\bar{X}_n)$ and $\text{Ber}(p)$? Intuitively, the two distribution should behave similarly since

$$\bar{X}_n \xrightarrow[n\to\infty]{i.p.} p.$$

Recall that by definition, the convergence in probability means

$$P(|\bar{X}_n - p| > \epsilon) \xrightarrow[n\to\infty]{} 0.$$

Remember that the total variation distance between $\text{Ber}(q)$ and $\text{Ber}(p)$ is $|q-p|$. We want to calculate the total variation distance between $\text{Ber}(\bar{X}_n)$ and $\text{Ber}(p)$, that is, $|\bar{X}_n - p|$. This is the same as what we've seen above! By the Law of Large Numbers, we can say that the total variation distance will converge in probability to 0 as n goes to infinity.

$$\text{TV}(\text{Ber}(\bar{X}_n), \text{Ber}(p)) = |\bar{X}_n - p| \xrightarrow[n\to\infty]{p.} 0.$$

Compute the total variation distance between

$$P = \mathsf{Unif}([0, s]) \quad \text{and} \quad Q = \mathsf{Unif}([0, t]), \quad \text{where } 0 < s < t.$$

$\mathsf{TV}(\mathbf{P}, \mathbf{Q}) =$ **Answer:** 1 - s/t

STANDARD NOTATION

Solution:

To compute the total variation distance between two uniform distributions, denote the densities of the two distributions by

$$f_s(x) = \frac{1}{s}\mathbf{1}\{0 \le x \le s\}, \quad f_t(x) = \frac{1}{t}\mathbf{1}\{0 \le x \le t\}.$$

With this, we have

$$\begin{aligned}
\mathsf{TV}(\mathsf{Unif}([0, s]), \mathsf{Unif}([0, t])) &= \frac{1}{2}\int_{\mathbb{R}} |f_s(x) - f_t(x)|\, dx \\
&= \frac{1}{2}\left[\int_0^s \left|\frac{1}{s} - \frac{1}{t}\right| dx + \int_s^t \left|\frac{1}{t}\right| dx\right] \\
&= \frac{1}{2}\left[\left(1 - \frac{s}{t}\right) + \left(1 - \frac{s}{t}\right)\right] \\
&= 1 - \frac{s}{t}.
\end{aligned}$$

Hence, $\mathsf{TV}(\mathsf{Unif}([0, s]), \mathsf{Unif}([0, t]))$ is a continuous function in t that decreases to 0 as t approaches s.

Let $X \sim N(\mu, \sigma^2)$ and $Y \sim \mathsf{Ber}(p)$. Compute the total variation distance between the distributions of $\operatorname{sign}(X)$ and $Y - 1$. Note that $\operatorname{sign}(X)$ is a function of the random variable with

$$\operatorname{sign}(X) = \begin{cases} 1 & \text{if } X > 0 \\ 0 & \text{if } X = 0 \\ -1 & \text{if } X < 0. \end{cases}$$

(If applicable, enter **abs(x)** for $|x|$, **Phi(x)** for $\Phi(x) = \mathbf{P}(Z \le x)$ where $Z \sim \mathcal{N}(0, 1)$, and **q(alpha)** for q_α, the $1 - \alpha$ -quantile of a standard normal distribution, e.g. enter **q(0.01)** for $q_{0.01}$.)

$\mathsf{TV}(\operatorname{sign}(X), Y - 1) =$

Answer: 0.5*(Phi(mu/sigma)+p+abs(1-p-Phi(-mu/sigma)))

STANDARD NOTATION

Solution:

Observe that $\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Hence

$$\operatorname{sign}(X) = \begin{cases} -1 & \text{with probability } \Phi\left(-\frac{\mu}{\sigma}\right) \\ 1 & \text{with probability } 1 - \Phi\left(-\frac{\mu}{\sigma}\right) = \Phi\left(\frac{\mu}{\sigma}\right) \end{cases}$$

Hence,

$$\begin{aligned}
2\mathsf{TV}(\operatorname{sign}(X), Y - 1) &= |\Phi\left(\frac{\mu}{\sigma}\right)| + |p| + |(1 - p) - \Phi\left(-\frac{\mu}{\sigma}\right)| \\
&= \Phi\left(\frac{\mu}{\sigma}\right) + p + |1 - p - \Phi\left(-\frac{\mu}{\sigma}\right)|.
\end{aligned}$$

Compute the total variation distance between

$$\mathbf{P} = \text{Ber}(p) \quad \text{and} \quad \mathbf{Q} = \text{Poiss}(p), \quad \text{where } p \in (0,1).$$

$\text{TV}(\mathbf{P}, \mathbf{Q}) =$ [ ] **Answer:** p*(1-e^(-p))

STANDARD NOTATION

Solution:

Recall the pmf $f_X(x)$ for $X \sim \text{Poiss}(p)$ is

$$f_X(x) = e^{-p}\frac{p^x}{x!} \qquad \text{for } x = 0, 1, 2 \ldots.$$

Hence,

$$\begin{aligned} 2\text{TV}(\text{Ber}(p), \text{Poiss}(p)) &= |e^{-p} - (1-p)| + |pe^{-p} - p| + e^{-p}\left(\frac{p^2}{2!} + \frac{p^3}{3!} + \cdots\right) \\ &= \left(e^{-p} - (1-p)\right) + \left(p\left(1 - e^{-p}\right)\right) + e^{-p}\left(e^p - (1+p)\right) \quad \text{since } e^{-p} > (1-p) \text{ for } p > 0 \\ &= 2\left(p\left(1 - e^{-p}\right)\right). \\ \iff \text{TV}(\text{Ber}(p), \text{Poiss}(p)) &= p\left(1 - e^{-p}\right). \end{aligned}$$

(a)

3 points possible (graded)

Are the following functions concave, convex, or neither?

$$f_1(x) = \ln x, \quad x > 0.$$

- (●) Concave ✔
- ( ) Convex
- ( ) Not concave and not convex

$$f_2(x) = -x^4 + x^2 - 40x, \quad x \in \mathbb{R}$$

- ( ) Concave
- ( ) Convex
- (●) Not concave and not convex ✔

$$f_3(x) = \frac{1}{\exp(x) - 1}, \quad x > 0$$

- ○ Concave
- ◉ Convex ✔ ✔
- ○ Not concave and not convex

Solution:

Recall that for a twice continuously differentiable function f, we can check concavity by testing whether $f''(x) \leq 0$ for all x in the (convex) domain in question.

To begin, compute

$$\begin{aligned} f_1'(x) &= \frac{1}{x} \\ f_1''(x) &= -\frac{1}{x^2} < 0, \quad \text{for } x > 0, \end{aligned}$$

so f_1 is concave.

$$\begin{aligned} f_2'(x) &= -4x^3 + 2x - 40 \\ f_2''(x) &= -12x^2 + 2, \end{aligned}$$

which means $f_2''(x) > 0$ for $x \in \left(-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$, but $f_2''(x) < 0$ for $x \notin \left[-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right]$, hence f_2 is neither concave or convex.

$$\begin{aligned} f_3'(x) &= -\frac{e^x}{(e^x - 1)^2} \\ f_3''(x) &= -\frac{e^x(e^x - 1) - 2e^{2x}}{(e^x - 1)^3} \\ &= \frac{e^{2x} + e^x}{(e^x - 1)^3} > 0, \quad \text{for } x > 0. \end{aligned}$$

That means that f is convex for $x > 0$.

(b)

2 points possible (graded)

A symmetric 2×2 matrix $\mathbf{A}$ (i.e. $\mathbf{A}^T = \mathbf{A}$) is negative semi- definite, i.e. $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^2$, if and only if both of the following is true:

- $\operatorname{tr}(\mathbf{A}) \leq 0$
- $\det(\mathbf{A}) \geq 0$

(This fact can be explained in terms the eigenvalues of $\mathbf{A}$. Let λ_1 and λ_2 be the eigenvalues of $\mathbf{A}$, then $\operatorname{tr}(\mathbf{A}) = \lambda_1 + \lambda_2$ while $\det(\mathbf{A}) = \lambda_1\lambda_2$. The two conditions above ensure that $\lambda_1, \lambda_2 \leq 0$.)

Use the fact given above to determine whether the following functions concave, convex, or neither.

$$f_4(\theta_1, \theta_2) = -\theta_1^2 + \frac{1}{2}(\theta_1 - \theta_2)^2 - 3\theta_2^2, \quad (\theta_1, \theta_2) \in \mathbb{R}^2$$

- (•) Concave ✔
- ( ) Convex
- ( ) Not concave and not convex

$$f_5(\theta_1, \theta_2) = -\theta_1^4 - \theta_2^4 - (\theta_2 - \theta_1)^3, \quad (\theta_1, \theta_2) \in \mathbb{R}^2, \text{ with } \theta_1 < \theta_2$$

- (•) Concave ✔
- ( ) Convex
- ( ) Not concave and not convex

Solution:

If f is function from $\Omega \subseteq \mathbb{R}^d \to \mathbb{R}$, then it is concave if the Hessian of f is negative semi-definite. In the special case of two dimensions, this can be checked by testing whether both $\operatorname{tr} \nabla^2 f \leq 0$ and $\det \nabla^2 f \geq 0$ are true.

$$\nabla f_4(\theta_1, \theta_2) = \begin{pmatrix} -\theta_1 - \theta_2 \\ -\theta_1 - 5\theta_2 \end{pmatrix}$$
$$H f_4(\theta_1, \theta_2) = \begin{pmatrix} -1 & -1 \\ -1 & -5 \end{pmatrix}.$$

Since $\operatorname{tr} \nabla^2 f_4 = -6 < 0$ and $\det \nabla^2 f_4 = 4 > 0$, we have $\nabla^2 f$ is negative semi-definite for all θ, and in turn, f_4 is concave.

$$\nabla f_5(\theta_1, \theta_2) = \begin{pmatrix} -4\theta_1^3 + 3(\theta_2 - \theta_1)^2 \\ -4\theta_2^3 - 3(\theta_2 - \theta_1)^2 \end{pmatrix}$$
$$H f_5(\theta_1, \theta_2) = \begin{pmatrix} -12\theta_1^2 - 6(\theta_2 - \theta_1) & 6(\theta_2 - \theta_1) \\ 6(\theta_2 - \theta_1) & -12\theta_2^2 - 6(\theta_2 - \theta_1) \end{pmatrix}.$$

We again check

$$\operatorname{tr} \nabla^2 f_5(\theta_1, \theta_2) = -12\theta_1^2 - 12(\theta_2 - \theta_1) - 12\theta_2^2 < 0, \quad \text{if } \theta_1 < \theta_2,$$

$$\begin{aligned} \det \nabla^2 f_5(\theta_1, \theta_2) = &\ (12\theta_1^2 + 6(\theta_2 - \theta_1))(12\theta_2^2 + 6(\theta_2 - \theta_1)) - 36(\theta_2 - \theta_1)^2 \\ = &\ 144\theta_1^2\theta_2^2 + 72(\theta_1^2 + \theta_2^2)(\theta_2 - \theta_1) > 0, \quad \text{if } \theta_1 < \theta_2. \end{aligned}$$

Combined, f_5 is concave on $\{\theta_1 < \theta_2\}$.

(a)

1 point possible (graded)

Compute the likelihood function and the maximum likelihood estimator for θ for

$$f_\theta(x) = \tau\theta^\tau x^{-(\tau+1)} \mathbf{1}(x \geq \theta), \quad \theta > 0,$$

where $\tau > 0$ is a known constant.

- ( ) $\max X_i$
- (x) $\min X_i$ ✔
- ( ) $\frac{1}{n}\sum X_i$
- ( ) $n\tau$
- ( ) ∞

Solution:

The likelihood function is

$$L = \tau^n \theta^{n\tau} \prod_i X_i^{-(\tau+1)} \mathbf{1}\{\min_i X_i \geq \theta\}$$

For $\theta \leq \min_i X_i$, the log-likelihood function is

$$l = n \ln \tau + n\tau \ln \theta - (\tau + 1) \sum_{i=1} \ln X_i$$

Take the derivative with respect to θ :

$$\frac{\partial l}{\partial \theta} = \frac{n\tau}{\theta} > 0.$$

Thus, L is an increasing function on $(0, \min_i X_i]$, and is 0 for $\theta > \min_i X_i$. Therefore,

$$\hat{\theta} = \min_i X_i$$

(b)

3 points possible (graded)

Compute the likelihood function and the maximum likelihood estimator for θ for

$$f_\theta(x) = \sqrt{\theta} x^{\sqrt{\theta}-1} \mathbf{1}(0 < x < 1), \quad \theta > 0.$$

You will find that the maximum likelihood estimator for θ is of the form

$$\hat{\theta}^{\text{MLE}} = c_1 n^{c_2} \left(\sum_{i=1}^{n} \ln X_i \right)^{c_3}.$$

Enter the numbers $c_1,\ c_2,\ c_3$ below.

$c_1 =$ **Answer:** 1

$c_2 =$ **Answer:** 2

$c_3 =$ **Answer:** -2

Solution:

The likelihood function is

$$L = \theta^{n/2} \prod_i X_i^{\sqrt{\theta}-1} \mathbf{1}\{0 \leq X_i \leq 1\}.$$

The log-likelihood function is

$$l = \frac{n}{2} \ln \theta + (\sqrt{\theta} - 1) \sum_i \ln X_i.$$

Take the derivative with respect to θ and set it to 0 :

$$\frac{\partial l}{\partial \theta} = \frac{n}{2\theta} + \frac{1}{2\theta^{1/2}} \sum_i \ln X_i = 0.$$

Then we get

$$\hat{\theta} = \frac{n^2}{(\sum \ln X_i)^2}.$$

Compute the likelihood function and the maximum likelihood estimator for θ

$$f_\theta(x) = \theta \tau x^{\tau-1} \exp\{-\theta x^\tau\} \mathbf{1}(x \geq 0), \quad \theta > 0,$$

where $\tau > 0$ is a known constant.

You will find that the maximum likelihood estimator for θ is of the form

$$\hat{\theta}^{\text{MLE}} = c_1 n^{c_2} \left(\sum_{i=1}^{n} X_i^{c_3} \right)^{c_4}.$$

Enter the c_1, c_2, c_3, c_4 in terms of τ if applicable.

(Enter **tau** for τ.)

$c_1 =$ **Answer:** 1

$c_2 =$ **Answer:** 1

$c_3 =$ **Answer:** tau

$c_4 =$ **Answer:** -1

Solution:

The likelihood function is

$$L = \theta^n \tau^n \prod_i X_i^{\tau-1} \exp\{-\theta \sum_i X_i^\tau\} \mathbf{1}\{X_i \geq 0\}.$$

The log-likelihood function is

$$l = n \ln \theta + n \ln \tau + (\tau - 1) \sum_i \ln X_i - \theta \sum_i X_i^\tau.$$

Take the derivative with respect to θ and set it to 0

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} - \sum_i X_i^\tau = 0,$$

we get

$$\hat{\theta} = \frac{n}{\sum_i X_i^\tau}.$$

5. Constrained maximum likelihood estimator

Homework due Jun 25, 2020 08:59 JST *Past Due*

Bookmark this page

Instruction:

What can we do when we have prior knowledge about the parameter? Imagine that an expert told you that the parameter θ lies between a and b. Would that additional knowledge change the MLE calculation? We will start by calculating just the standard MLE. We will then think about what we can do with this additional knowledge in part (c).

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with probability density function

$$f_\theta(x) = \theta x^{-\theta-1}, \quad \theta > 0, \quad x \geq 1.$$

To encourage you to do the computations carefully rather than eliminate choices, you will be given only **1-2 attempts per question** .

(a)

1 point possible (graded)

What is the likelihood function for θ ?

- $\theta^n \prod_{i=1}^n x_i^{-\theta-1}$ ✔
- $\theta^n \prod_{i=1}^n x_i^{-\theta-1} \mathbf{1}\{\min_i X_i \geq 1\}$ ✔
- $\theta^n \prod_{i=1}^n x_i^{-\theta-1} \mathbf{1}\{\min_i X_i < 1\}$
- $\theta^n \prod_{i=1}^n x_i^{-\theta-1} \mathbf{1}\{\max_i X_i \geq 1\}$
- $\theta^n \prod_{i=1}^n x_i^{-\theta-1} \mathbf{1}\{\max_i X_i < 1\}$
- $n \ln \theta - (\theta + 1) \sum_{i=1}^n \ln X_i$

Solution:

$$
\begin{aligned}
L_n &= \prod_{i=1}^{n} \theta x_i^{-\theta-1} \mathbf{1}\{X_i \geq 1\} \\
&= \theta^n \prod_{i=1}^{n} x_i^{-\theta-1} \mathbf{1}\{\min_i X_i \geq 1\}
\end{aligned}
$$

But since we assume our statistical model to be well-specified, $\min_i X_i \geq 1$ will always be satisfied, and so we can drop the corresponding indicator function. Hence, $L_n = \theta^n \prod_{i=1}^n x_i^{-\theta-1}$ is correct under the well-specified assumption.

(b)

1 point possible (graded)

What is the maximum likelihood estimator for θ ?

- $\frac{n}{\sum_{i=1}^n \ln X_i}$ ✔
- $-\frac{n}{\sum_{i=1}^n \ln X_i}$
- $\frac{\sum_{i=1}^n \ln X_i}{n}$
- $-\frac{\sum_{i=1}^n \ln X_i}{n}$
- $\frac{\sum_{i=1}^n X_i}{n}$
- $\frac{n}{\sum_{i=1}^n X_i}$

Solution:

Take the derivative of the likelihood function with respect to θ .

$$\frac{\partial L_n}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \ln X_i = 0$$

Solving the equation for θ , we get

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln X_i}$$

(c)

1 point possible (graded)

Suppose we have two numbers $0 < a < b$. We are interested in the value of θ that maximizes the likelihood in the set $[a, b]$.

Let $\hat{\theta}$ denote the maximum likelihood estimator you found in part (b) above, and let $\hat{\theta}_{\text{const}}$ denote the maximum likelihood estimator within the interval $[a, b]$, where $0 < a < b$. Choose all correct answers.

- [ ] If $b \leq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = a$
- [ ] If $b \leq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = b$ ✔
- [ ] If $b \leq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = \hat{\theta}$
- [ ] If $a < \hat{\theta} < b$, then $\hat{\theta}_{\text{const}} = a$
- [ ] If $a < \hat{\theta} < b$, then $\hat{\theta}_{\text{const}} = b$
- [ ] If $a < \hat{\theta} < b$, then $\hat{\theta}_{\text{const}} = \hat{\theta}$ ✔
- [ ] If $a \geq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = a$ ✔
- [ ] If $a \geq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = b$
- [ ] If $a \geq \hat{\theta}$, then $\hat{\theta}_{\text{const}} = \hat{\theta}$

Solution:

Take the second derivative of the likelihood function with respect to θ .

$$\frac{\partial^2 L_n}{\partial \theta^2} = -\frac{n}{\theta^2} < 0$$

Since the second derivative is strictly less then 0 , the function is strictly concave with respect to θ . Therefore, depending on the value of $\frac{n}{\sum_{i=1}^{n} \ln X_i}$, which is the maximum, the largest value that likelihood function can take in the set $[a, b]$ changes.