

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#).



8. Applying Huber's loss to the

[Course](#) > [Unit 3 Methods of Estimation](#) > [Lecture 12: M-Estimation](#) > Laplace distribution

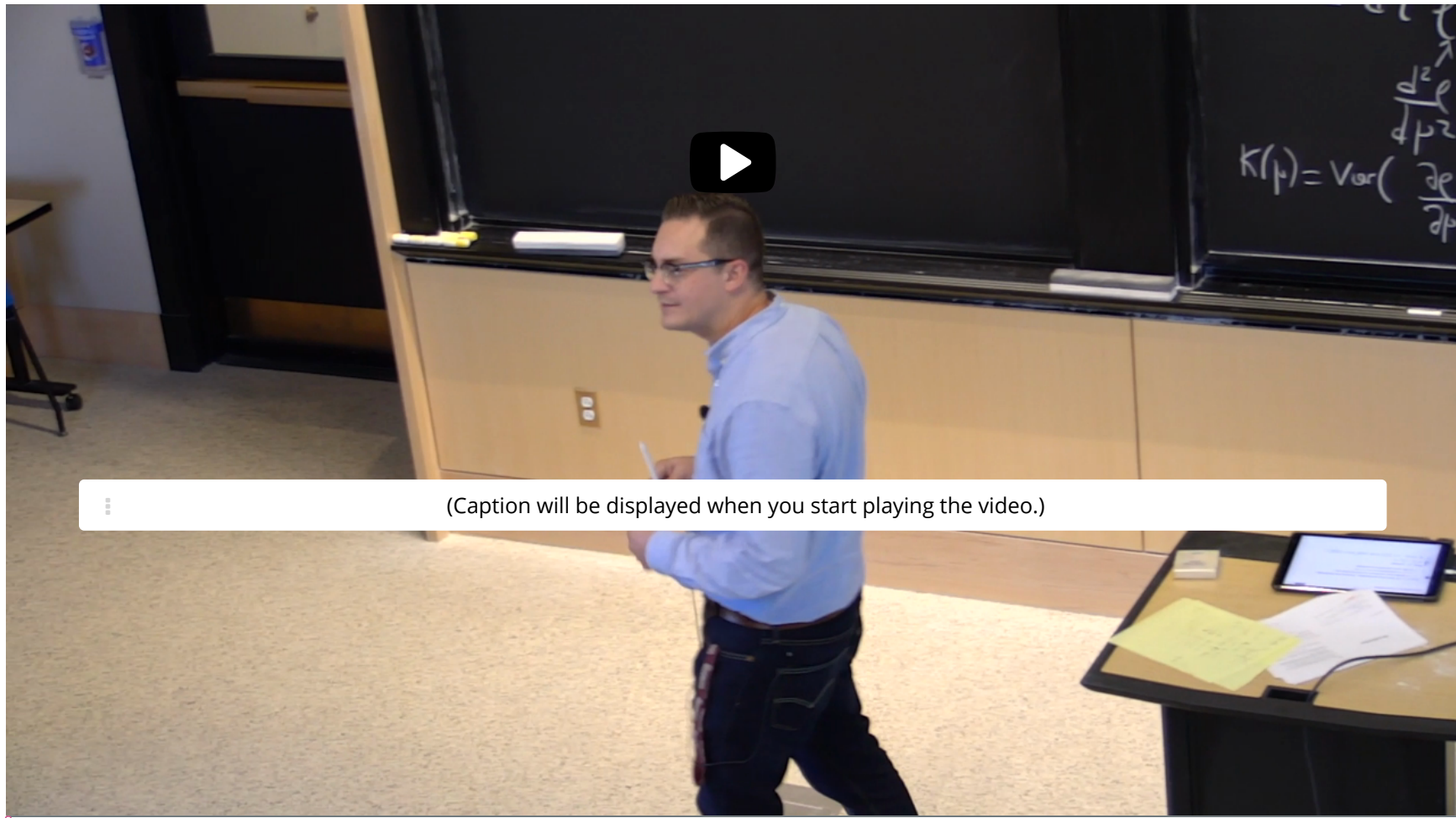
Audit Access Expires Dec 24, 2019

You lose all access to this course, including your progress, on Dec 24, 2019.

Upgrade by Nov 4, 2019 to get unlimited access to the course as long as it exists on the site. [Upgrade now](#)

8. Applying Huber's loss to the Laplace distribution

Applying Huber's loss to the Laplace distribution



▶ 0:00 / 0:00

▶ 1.50x 🔊 🗑️ 📄 🗣️

Video

[Download video file](#)

Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

The Laplace distribution

3/3 points (graded)

The **Laplace distribution** (also known as the **double-exponential distribution**) is a continuous distribution with **location parameter** $m \in \mathbb{R}$ and density given by

$$f_m(x) = \frac{1}{2} e^{-|x-m|}.$$

Let X denote a Laplace random variable with location parameter set to be $m = 0$.

What is $\mathbb{E}[X]$?

✓ Answer: 0

Does the variance $\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$ exist?

☒ Yes☐ No

Which of the following are true about X ? (Choose all that apply.)

Hint: The function $x^k e^{-|x|}$ is integrable, i.e. $\int_{-\infty}^{\infty} x^k e^{-|x|} dx$ is finite for all k .

☒ The distribution of X is symmetric in the sense that X and $-X$ have the same distribution.☐ The function $\ln f_m(x)$ has a continuous first derivative.☒ For any integer $k > 0$, the k -th moment $\mathbb{E}[X^k]$ exists.

**Solution:**

For the first question, observe that the function $xe^{-|x|}$ is odd and also integrable. Therefore,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{1}{2} xe^{-|x|} dx = 0.$$

For the second question, the function $x^2 e^{-|x|}$ is integrable. Hence,

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] = \int_{-\infty}^{\infty} \frac{1}{2} x^2 e^{-|x|} dx$$

For the third question, we examine the choices in order.

- "The distribution of X is symmetric in the sense that X and $-X$ have the same distribution." is correct. This is because the density $\frac{1}{2}e^{-|x|}$ is an even function.
- "The function $\ln f_m(x)$ has a continuous first derivative." is incorrect. This is because $\ln f_m(x) = -|x - m|$, which is not differentiable at $x = m$.
- "For any integer $k > 0$, the k -th moment $\mathbb{E}[X^k]$ exists." is correct. In general, the function $x^k e^{-|x|}$ is integrable on \mathbb{R} , so the k -th moment $\mathbb{E}[X^k]$ exists for all $k > 0$.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

The Sample Median

4/4 points (graded)

Let $S = x_1 < x_2 < \dots < x_n$ denote a sorted list of numbers. We define the **elementary median** $\text{med}_e(S)$ to be

$$\text{med}_e(S) := \begin{cases} x_{\lceil n/2 \rceil} & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \text{if } n \text{ is even} \end{cases}$$

In other words, when n is odd, the median is the middle number when the set S is sorted from smallest to largest. If n is even, we can just define the median to be the average of both middle numbers. This definition is likely familiar from prior math classes.

A more advanced definition, useful for statistical purposes, is to define the **sample median** $\text{med}_s(S)$ of a sample $S := X_1, X_2, \dots, X_n$ to be

$$\text{med}(S) := \underset{m}{\operatorname{argmin}} \sum_{i=1}^n |X_i - m|.$$

While the elementary median is unique, this is not always the case for the **sample median**, as you will see in the next few questions.

Consider the data set $S = \{1, 2, 3\}$.

What is the **elementary median** of S ?

✓ Answer: 2

What is the **sample median** of S ?

Hint: Use computational software to graph the objective function.

✓ Answer: 2

Now consider the data set $T = \{1, 2, 3, 4\}$.

What is the **elementary median** of T ?

✓ Answer: 2.5

Which of the following describes **all** of the sample medians of T ? (*Hint: Use computational software to graph the objective function*)

☐ 2.5☐ Any number in the open interval $(2, 3)$.☒ Any number in the closed interval $[2, 3]$.☐ None of the above.

Solution:

For the first question, the elementary median of S is 2. For the second question, we want to find m such that

$$F(m) = |1 - m| + |2 - m| + |3 - m|$$

is as small as possible. Note that $F(m)$ is a piecewise linear function with discontinuities in its first derivative precisely at the points $(1, 3)$, $(2, 2)$, and $(3, 3)$ (one corresponding to each summand of the form $|x - m|$). Drawing the line segments connecting these three points, we see that the minimizer of $F(m)$ is at $m = 2$. Therefore, for the second question, the sample median is 2.

For the third question, the elementary median is 2.5. For the fourth question, we want to find m such that

$$G(m) = |1 - m| + |2 - m| + |3 - m| + |4 - m|$$

is as small as possible. As before, $G(m)$ is a piecewise linear function. Its discontinuities are at the points $(1, 6)$, $(2, 4)$, $(3, 4)$, and $(4, 6)$. Hence, $G(m)$ is a horizontal line segment on the interval $[2, 3]$, and $G(m)$ has positive slope otherwise. Hence, the correct response to the fourth question is "Any number in the closed interval $[2, 3]$ ".

Remark: In general, one can show that for an ordered sample $S = x_1 < \dots < x_n$ that

$$\text{med}(S) = \begin{cases} x_{\lceil n/2 \rceil} & \text{if } n \text{ is odd} \\ \text{Any number in the interval } [x_{n/2}, x_{n/2+1}] & \text{if } n \text{ is even} \end{cases}$$

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Maximum Likelihood Estimator for the Laplace Distribution

2/2 points (graded)

Consider a Laplace statistical model $(\mathbb{R}, \{P_m\}_{m \in \mathbb{R}})$ where P_m denotes the Laplace distribution with location parameter m . Let $X_1, \dots, X_n \stackrel{iid}{\sim} P_{m^*}$ denote a sample from a Laplace distribution with unknown parameter m^* . Recall that the density of P_m is given by

$$f_m(x) = \frac{1}{2} e^{-|x-m|}.$$

What is the log-likelihood $\ell_n(X_1, \dots, X_n; m)$ for this statistical model?

☒ $-n \ln(2) - \sum_{i=1}^n |X_i - m|$

☐ $-\sum_{i=1}^n |X_i - m|$

☐ $-n \ln(2) - \sum_{i=1}^n (X_i - m)^2$

☐ $-n \ln(2) + \sum_{i=1}^n |X_i - m|$



Recall that the maximum likelihood estimator $\widehat{m}_n^{\text{MLE}}$ is given by

$$\widehat{m}_n^{\text{MLE}} = \operatorname{argmin}_{m \in \mathbb{R}} -\ell_n(X_1, \dots, X_n; m).$$

Suppose you observe the sample

$$S = 0.5, 1.2, 0.6, -0.7, -0.2.$$

What is the value of the MLE for m^* for this data set? *Hint:* Use the previous question, in particular the remark at the end of the solution.

✓ Answer: 0.5

Solution:

For the first question, the likelihood for n observations is given by

$$\prod_{i=1}^n f_m(X_i) = \frac{1}{2^n} \prod_{i=1}^n e^{-|x-m|}.$$

Therefore,

$$\ell_n(X_1, \dots, X_n; m) = -n \ln(2) - \sum_{i=1}^n |X_i - m|.$$

For the second question, we need to minimize the quantity

$$5 \ln(2) + |m - 0.5| + |m - 1.2| + |m - 0.6| + |m + 0.2| + |m + 0.7|.$$

with respect to m . As stated in the previous quantity, any m that minimizes the above is a sample median of the data set S . Since S is odd, we have that $\widehat{m}_n^{\text{MLE}} = 0.5$.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Concept Question: Maximum Likelihood Estimator for the Laplace distribution

1/1 point (graded)

As in the previous problem, let $\widehat{m}_n^{\text{MLE}}$ denote the MLE for an unknown parameter m^* of a Laplace distribution.

Can we apply the theorem for the asymptotic normality of the MLE to $\widehat{m}_n^{\text{MLE}}$? (You must choose the correct answer that also has the correct explanation.)

☐ No, because the log-likelihood is not concave.

☒ No, because the log-likelihood is not twice-differentiable, so the Fisher information does not exist.

☐ Yes, because the log-likelihood is concave.

☐ Yes, because the other technical conditions required to apply the theorem are satisfied.

**Solution:**

We examine the choices in order.

- "No, because the log-likelihood is not concave." is incorrect. A sum of concave functions is concave, and for any constant c , the function $x \rightarrow -|x - c|$ is concave. Therefore, the log-likelihood

$$\ell_n(X_1, \dots, X_n; m) = -n \ln(2) - \sum_{i=1}^n |X_i - m|$$

is also concave. Hence, the reasoning for this response is incorrect.

- "No, because the log-likelihood is not twice-differentiable, so the Fisher information does not exist." is correct. This is because $\ell_n(X_1, \dots, X_n; m)$ has discontinuities in its first derivative with respect to m at $m = X_i$ for $i = 1, \dots, n$.
- "Yes, because the log-likelihood is concave." is incorrect. Although the log-likelihood is concave, the Fisher information does not exist, as discussed in the analysis of the previous two responses.
- "Yes, because the other technical conditions required to apply the theorem are satisfied." is incorrect. The remaining technical conditions are not enough to guarantee asymptotic normality since the Fisher information does not exist, as discussed above.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Applying Huber's loss to a Laplace distribution I

2/2 points (graded)

As above, let m^* denote an unknown parameter for a Laplace distribution. In this problem, we will use the principles of M-estimation and the smoothness of Huber's loss function to construct an asymptotically normal estimator for m^* . Let P_m denote the Laplace distribution with location parameter m .

Recall Huber's loss is defined as

$$h_\delta(x) = \begin{cases} \frac{x^2}{2} & \text{if } |x| < \delta \\ \delta(|x| - \delta/2) & \text{if } |x| > \delta \end{cases}.$$

As computed in lecture, the derivative of Huber's loss is the **clip function** :

$$\text{clip}_\delta(x) := \frac{d}{dx} h_\delta(x) = \begin{cases} \delta & \text{if } x > \delta \\ x & \text{if } -\delta \leq x \leq \delta \\ -\delta & \text{if } x < -\delta \end{cases}$$

Find the value of

$$\left. \frac{\partial}{\partial m} \mathbb{E}_{X \sim P_{m^*}} [h_\delta(X - m)] \right|_{m=m^*}.$$

Hint: You are allowed to switch the derivative and expectation.

✓ Answer: 0

In the framework of M-estimation, our loss function is not Huber's loss itself, but rather

$$\rho(x, m) := h_\delta(x - m)$$

Recall the functions

$$J(m) = \mathbb{E} \left[\frac{\partial^2 \rho}{\partial m^2} (X_1, m) \right]$$

$$K(m) = \text{Var} \left[\frac{\partial \rho}{\partial m} (X_1, m) \right]$$

Do the functions K and J exist for a Laplace statistical model?

- ☐ No, because the log-likelihood is not twice-differentiable.
- ☐ No, because $J(m)$ exists but $K(m)$ does not.
- ☐ Yes, because the Fisher information is well-defined for a Laplace statistical model.
- ☒ Yes, because the function $\rho(x, m)$ as defined above is twice-differentiable.



Solution:

The answer to the first question is 0. To see this, observe that

$$\begin{aligned} \frac{\partial}{\partial m} \mathbb{E}_{X \sim P_m^*} [h_\delta(X)] &= \mathbb{E}_{X \sim P_m^*} \left[\frac{\partial}{\partial m} h_\delta(X) \right] \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \text{clip}_\delta(x - m) e^{-|x - m^*|} dx \\ &= \frac{1}{2} \left(\delta \int_{m+\delta}^{\infty} e^{-|x - m^*|} dx - \delta \int_{-\infty}^{-\delta+m} e^{-|x - m^*|} dx + \int_{-\delta+m}^{\delta+m} x e^{-|x - m^*|} dx \right). \end{aligned}$$

Applying the change of variables $y = x - m$, we have

$$= \frac{1}{2} \left(\delta \int_{\delta}^{\infty} e^{-|y+m-m^*|} dy - \delta \int_{-\infty}^{-\delta} e^{-|y+m-m^*|} dy + \int_{-\delta}^{\delta} y e^{-|y+m-m^*|} dy \right).$$

Setting $m = m^*$, we have

$$\frac{\partial}{\partial m} \mathbb{E}_{X \sim P_m^*} [h_{\delta}(X)] \Big|_{m=m^*} = \frac{1}{2} \left(\delta \int_{\delta}^{\infty} e^{-|y|} dy - \delta \int_{-\infty}^{-\delta} e^{-|y|} dy + \int_{-\delta}^{\delta} y e^{-|y|} dy \right) = 0.$$

Remark: The function $m \mapsto \mathbb{E}_{X \sim P_m^*} [h_{\delta}(X)]$ is strictly convex, so this means the loss function has a unique critical point, and this is where the minimum is attained. The above calculation guarantees that the minimum is at $m = m^*$, the value of the true parameter.

"No, because the log-likelihood is not twice-differentiable.", "No, because $J(m)$ exists but $K(m)$ does not.", "Yes, because the Fisher information is well-defined for a Laplace statistical model.", "Yes, because the function $\rho(x, m)$ as defined above is twice-differentiable."

For the second question, we consider the responses in order.

- "No, because the log-likelihood is not twice-differentiable." is incorrect. In the problem "Huber's loss" on the page "Robust Statistics and Huber's Loss", we showed that $\rho(x, m) = h_{\delta}(x - m)$ is twice-differentiable with respect to m .
- "No, because $J(m)$ exists but $K(m)$ does not." is also incorrect. Both $K(m)$ and $J(m)$ exist because ρ is twice-differentiable, and its derivatives are integrable.
- "Yes, because the Fisher information is well-defined for a Laplace statistical model." is incorrect. The Fisher information does not exist for a Laplace statistical model, as was shown in the problem "Concept Question: Maximum Likelihood Estimator for the Laplace Distribution" on the page "Applying Huber's loss to the Laplace distribution."
- "Yes, because the function $\rho(x, m)$ as defined above is twice-differentiable." is the correct answer. In a previous problem, we showed that $\rho(x, m) = h_{\delta}(x - m)$ is twice-differentiable with respect to m .

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Applying Huber's loss to a Laplace distribution II

1/1 point (graded)

We use the same set-up as in the previous problem. Recall that m^* is an unknown location parameter for a Laplace distribution.

The M-estimator \widehat{m} for m^* associated to the loss function $\rho(x, m) = h_\delta(x - m)$ is given by

$$\widehat{m} = \operatorname{argmin}_{m \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n h_\delta(X_i - m).$$

Consider the slide *Asymptotic normality* in Unit 3. Suppose that the technical conditions in 3. of slide are satisfied. Also, note that for any fixed x , the function $m \mapsto h_\delta(x - m)$ is strictly convex. (You can see this by observing $h_\delta(x - m) = h_\delta(m - x)$, so the graph of $h_\delta(x - m)$ as a function of m for a fixed x is the same as the graph of $h_\delta(x - m)$ as a function of x for a fixed m .) Finally, consider the calculation of $J(m)$ from lecture.

Can we apply the theorem on the slide *Asymptotic normality* to conclude that \widehat{m} is asymptotically normal?

(Choose the correct answer, 'Yes' or 'No', that also has a correct explanation.)

- ☐ No, because m^* is not the unique minimizer of the function $m \mapsto \mathbb{E}_{X \sim P_{m^*}} [\rho(X, m)]$.
- ☐ No, because $J(m)$ is not invertible.
- ☐ Yes, because m^* is the unique minimizer of the function $m \mapsto \frac{1}{n} \sum_{i=1}^n h_\delta(X_i - m)$ and $J(m)$ is invertible.
- ☒ Yes, because m^* is the unique minimizer of the function $m \mapsto \mathbb{E}_{X \sim P_{m^*}} [\rho(X, m)]$ and $J(m)$ is invertible.



Solution:

"Yes, because m^* is the unique minimizer of the function $m \mapsto \mathbb{E}_{X \sim P_{m^*}} [\rho(X, m)]$ and $J(m)$ is invertible." is the correct answer. As shown in the Remark in the solution to the previous problem, m^* is the unique minimizer of the loss function $m \mapsto \mathbb{E}_{X \sim P_{m^*}} [\rho(X, m)]$. Moreover,

$$J(m) = 1 - e^{-\delta}$$

as was shown in the lecture. Hence $J(m)$ is invertible. Assuming that the required technical conditions are satisfied, we conclude that the estimator \widehat{m} is asymptotically normal.

You have used 2 of 2 attempts

i Answers are displayed within the problem

Applying Huber's loss to the Laplace distribution: Computation

Statistical analysis



► Let $J(\mu) = + \frac{\partial^2 Q}{\partial \mu \partial \mu^\top}(\mu)$ ($= + \mathbb{E} \left[\frac{\partial^2 \rho}{\partial \mu \partial \mu^\top}(X_1, \mu) \right]$ under

some regularity conditions).

► Let $K(\mu) = \text{Cov} \left[\frac{\partial \rho}{\partial \mu}(X_1, \mu) \right]$.

(Caption will be displayed when you start playing the video.)

► **Remark:** In the log-likelihood case (write $\mu = \theta$),

$$J(\theta) = K(\theta) = \mathcal{I}(\theta) \quad (\text{Fisher information})$$

► 0:00 / 0:00

► 1.50x



Video

[Download video file](#)

Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

Discussion

Hide Discussion

Topic: Unit 3 Methods of Estimation:Lecture 12: M-Estimation / 8. Applying Huber's loss to the Laplace distribution

Add a Post

Show all posts ▼

by recent activity ▼

There are no posts in this topic yet.

✕

[Learn About Verified Certificates](#)

© All Rights Reserved