

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](#).



6. Robust Statistics and Cauchy's

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

Audit Access Expires Dec 24, 2019

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

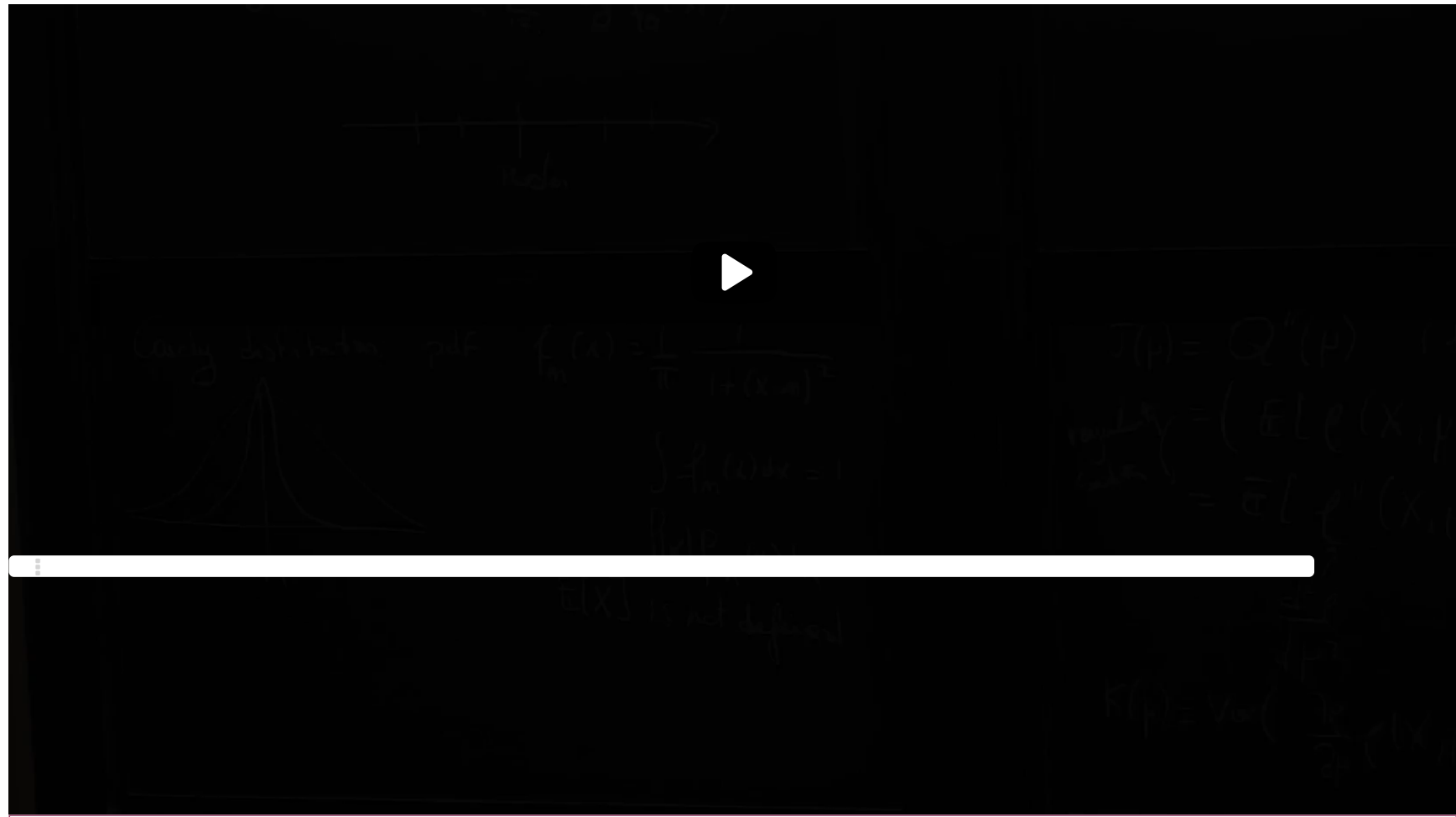
Upgrade by Nov 4, 2019 to get unlimited access to the course as long as it exists on the site. [Upgrade now](#)

6. Robust Statistics and Cauchy's Distribution

Video Note:

At 1:13, Professor Rigllet misspoke and said “variance”, but what he should have said was “median”. Using variance instead of median will make the point not hold as the variance is sensitive to outliers.

Robust Statistics, Cauchy Distribution



Video

[Download video file](#)

Transcripts

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

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

Robust Statistics and the Median

4/4 points (graded)

In this problem, you will see how some estimators are more resilient to corruptions or mistakes in the data than others. Such estimators are referred to as **robust**.

Researchers in a lab observe the following data set

$$\{0.5, 1.8, -2.3, 0.86, 0.32\}.$$

In reality, these numbers are generated from the standard normal distribution $\mathcal{N}(0, 1)$, but as in most statistical problems, this distribution is unknown to the researchers. Their goal is to estimate the mean of this unknown distribution, and they will try two different statistics for doing so: the sample mean and the median.

What is the sample mean \bar{X}_5 of the data set?

✓ Answer: 0.236

What is the median M_5 of the data set? (For a discrete data set, we define the median to be the middle number when the data set is sorted from smallest to largest.)

✓ Answer: 0.5

Suppose now that one of the lab assistants makes a typo, and instead records the data set

$$\{0.5, 1.8, -23, 0.86, 0.32\}.$$

(Note that the number -2.3 from the sample has been erroneously changed to -23 .)

What is the sample mean of the new, corrupted data set?

✓ Answer: -3.9

What is the median of the new, corrupted data set?

✓ Answer: 0.5

Solution:

The mean of the first data set is

$$\frac{-2.3 + 0.5 + 1.8 + 0.86 + 0.32}{5} \approx 2.36$$

From smallest to largest, the data set reads

$$-2.3, 0.32, 0.5, 0.86, 1.8$$

, so 0.5 is the median.

The mean of the second data set is

$$\frac{-23 + 0.5 + 1.8 + 0.86 + 0.32}{5} \approx -3.90.$$

From smallest to largest, the data set reads,

$$-23, 0.32, 0.5, 0.86, 1.8$$

so 0.5 is still the median.

Remark: This simple example illustrates how the median is in general a more *robust* estimator than the mean. That is, errors or corruptions in the data set have a limited effect on how much the median changes. However, the same is not true for the mean.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Cauchy distribution I

2/2 points (graded)

The **Cauchy distribution** is a continuous distribution with a parameter m , known as the **location parameter**, and with density given by

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

Suppose X is a random variable distributed as the Cauchy distribution.

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

☐ 0☐ $\frac{1}{2}$ ☐ 1☒ Does not exist.

Recall that the **median** of a continuous distribution is any number M such that $P(X > M) = P(X < M) = 1/2$. For the Cauchy distribution, it turns out that the median is unique.

If the location parameter is set to be $m = 1/2$, what is $\text{med}(X)$?

1/2

✓ Answer: 0.5

STANDARD NOTATION

Solution:

The correct answer to the first question is "Does not exist". To show this, let us temporarily set the location parameter to be $m = 0$. If we were to try to compute the mean, we would write down the integral

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{x}{1+x^2} dx.$$

However, this improper integral does not converge. The antiderivative is

$$\frac{1}{2\pi} \ln(1+x^2),$$

which is unbounded as $|x| \rightarrow \infty$.

The answer to the second question is "1/2". This is because

$$P(X > 1/2) = \int_{1/2}^{\infty} \frac{1}{\pi} \cdot \frac{1}{1+(x-1/2)^2} dx = - \int_{1/2}^{-\infty} \frac{1}{\pi} \cdot \frac{1}{1+(-y+1/2)^2} dy = P(X < 1/2).$$

The third equation follows from making the substitution $x = -y + 1$.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Cauchy distribution II

1/1 point (graded)

As in the previous problem, let X denote a random variable distributed as the Cauchy distribution with location parameter m .

Which of the following are true about the random variable $X - m$? (Choose all that apply.)

- ☒ The expectation (first moment) of $X - m$ is not defined.
- ☒ $X - m$ is distributed as a Cauchy random variable with location parameter set to be 0.
- ☒ $X - m$ is **symmetric** in the sense that $X - m$ and $m - X$ both have the same distribution.
- ☐ The method of moments can be used to estimate the location parameter m .



Solution:

Let us examine the choices in order.

- "The expectation (first moment) of $X - m$ is not defined." is correct. As we showed in a previous problem, the improper integral

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{x}{1 + (x - m)^2} dx$$

does not converge, so the expectation of X is not defined.

- " $X - m$ is distributed as a Cauchy random variable with location parameter set to be 0." is correct. We show that $X - m$ has the same cdf as a Cauchy random variable Y with location parameter set to be 0. Indeed

$$P(X - m < t) = \int_{-\infty}^{t+m} \frac{1}{\pi} \cdot \frac{1}{1 + (x - m)^2} dx = \int_{-\infty}^t \frac{1}{\pi} \cdot \frac{1}{1 + y^2} dy = P(Y < t).$$

Here we made the substitution $y = x - m$.

- " $X - m$ is **symmetric** in the sense that $X - m$ and $m - X$ both have the same distribution." is correct. By the previous question, $X - m$ has a density given by $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. This is an even function, so it follows that $X - m$ and $m - X$ have the same distribution.
- "The method of moments can be used to estimate the location parameter m ." is incorrect. Since the moments of a Cauchy random variable do not exist, the method of moments cannot be used for parameter estimation for this family of distributions.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Methods of Estimation:Lecture 12: M-Estimation / 6. Robust Statistics and Cauchy's 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