

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](#).



[Course](#) > [Unit 3 Methods of Estimation](#) > [Lecture 12: M-Estimation](#) > 3. Recall the Ideas for M-estimation

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](#)

3. Recall the Ideas for M-estimation

Recall: Ideas for M-estimation

Definition

replace E with $\frac{1}{n} \sum_{i=1}^n$

▶ Define $\hat{\mu}_n$ as a minimizer of:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{y_i - \mu}{\sigma}\right)$$

▶ Examples: Empirical mean, empirical median, empirical quantiles, MLE, etc.

▶ 6:27 / 6:27

▶ 1.50x

🔊 🔍 📄 🗣️

Video

[Download video file](#)

Transcripts

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

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

Concept check: Defining M-estimators

1/1 point (graded)

Suppose we have access to a distribution \mathbf{P} which has an unknown parameter μ^* that we would like to estimate from samples $X_1, \dots, X_n \stackrel{iid}{\sim} \mathbf{P}$. Suppose we have a **loss function** $\rho(x, \mu)$ with the property that

$$\mu^* = \operatorname{argmin}_{\mu \in \mathbb{R}} \mathbb{E}_{X \sim \mathbf{P}} [\rho(X, \mu)].$$

What commonly used statistical trick is used to define an M-estimator? (Refer to the slides.)

☐ Using the KL divergence instead of TV distance.

☐ The method of moments.

☒ Replacing expectations with averages.



Solution:

The correct response is "Replacing expectations with averages." Indeed, we have that the equation

$$\mu^* = \operatorname{argmin}_{\mu \in \mathbb{R}} \mathbb{E}_{X \sim \mathbf{P}} [\rho(X, \mu)]$$

becomes

$$\hat{\mu} = \operatorname{argmin}_{\mu \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n [\rho(X_i, \mu)]$$

upon replacing the expectation by an average over the sample. Here, $\hat{\mu}$ is precisely the M-estimator associated with $\rho(x, \mu)$.

The response "Using the KL divergence instead of TV distance." is incorrect. Rather, the KL divergence was used specifically in the context of maximum likelihood estimation. It does not play a role in the context of M-estimation.

The response "The method of moments." is also incorrect. The method of moments is a tool for parameter estimation which is distinct from M-estimation. The method of moments is not what is used to define an M-estimator.

[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 / 3. Recall the Ideas for M-estimation

[Add a Post](#)

Show all posts ▼

by recent activity ▼

There are no posts in this topic yet.

x[Learn About Verified Certificates](#)

© All Rights Reserved