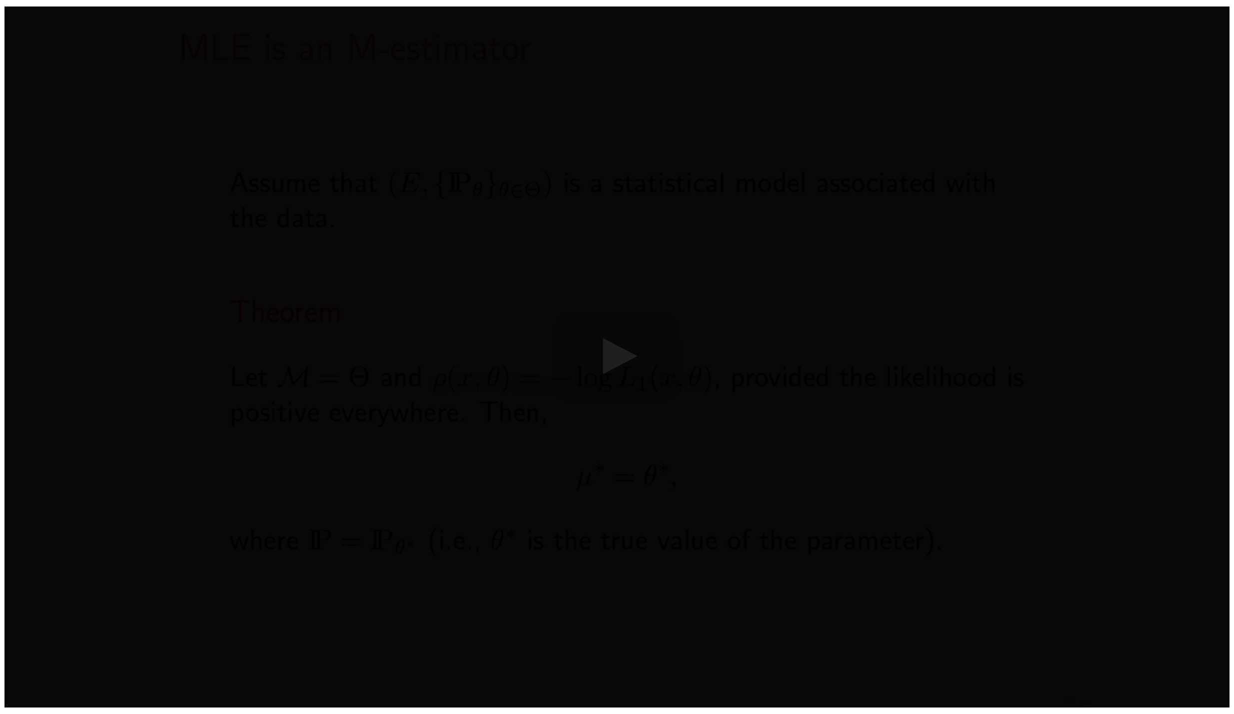
2. Introduction to M-estimation

Exercises due Jul 13, 2021 19:59 EDT

Video Note: The video below is the last 10 minutes of the previous lecture when Prof Rigollet began the discussion of m-estimation.

Introduction to M-estimation



10:47 / 10:47 1.50x

Now I have my--

if I minimize expectation of rho, I get--

basically, it's the same as minimizing my KL divergence,

and then I can replace my expectation by an average,

and that's how I got my maximum likelihood estimator.

So it's just an example of maximum likelihood.

It's just that it would be--

you would be hard-pressed to actually come up

with this very specific loss function

if you didn't come from the maximum likelihood.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

M-estimation

Let $X_1, \ldots, X_n$ be i.i.d. with some unknown distribution $\mathbf{P}$ and an associated parameter μ^* on a sample space E. We make no modeling assumption that $\mathbf{P}$ is from any particular family of distributions.

An **M-estimator** $\widehat{\mu}$ of the parameter μ^* is the **argmin of an estimator of a function** $\mathcal{Q}(\mu)$ **of the parameter** which satisfies the following:

- $\mathcal{Q}(\mu) = \mathbb{E}[\rho(X, \mu)]$ for some function $\rho : E \times \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the set of all possible values of the unknown true parameter $\mu*$;
- $\mathcal{Q}(\mu)$ attains a **unique** minimum at $\mu = \mu^*$, in $\mathcal{M}$. That is, $\mathrm{argmin}_{\mu \in \mathcal{M}} \mathcal{Q}(\mu) = \mu^*$.

In general, the goal is to find the **loss function** ρ such $\mathcal{Q}(\mu) = \mathbb{E}[\rho(X, \mu)]$ has the properties stated above.

Note that the function $\rho(X, \mu)$ is in particular a function of the random variable X, and the expectation in $\mathbb{E}[\rho(X, \mu)]$ is to be taken against the **true distribution** $\mathbf{P}$ of X, with associated parameter value μ^*.

Because $\mathcal{Q}(\mu)$ is an expectation, we can construct a (consistent) estimator of $\mathcal{Q}(\mu)$ by replacing the expectation in its definition by the sample mean.

Example: multivariate mean as minimizer

Let $\mathbf{X} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ be a continuous random vector with density $f : \mathbb{R}^2 \to \mathbb{R}$. Recall the mean of $\mathbf{X}$ is

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X^{(1)}] \\ \mathbb{E}[X^{(2)}] \end{pmatrix}$$

$$\left(\mathbb{E}\left[X^{(2)}\right]\right)$$

Recall the square of the Euclidean norm function on $\mathbb{R}^2$:

$$\begin{aligned} \|\cdot\|^2 : \mathbb{R}^2 &\to \mathbb{R} \\ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &\mapsto (y_1)^2 + (y_2)^2. \end{aligned}$$

We now show that the (multivariate) mean of $\mathbf{X}$ satisfies:

$$\mathbb{E}[\mathbf{X}] \quad = \quad \mathbf{argmin}_{\vec{\mu}\in\mathbb{R}^2}\mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right].$$

(We will use subscripts to label the components of vectors below.)

First, expand $\mathcal{Q}(\vec{\mu}) = \mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right]$ as an integral expression, and write down both partial derivatives $\frac{\partial \mathcal{Q}}{\partial \mu_1}(\vec{\mu})$ and $\frac{\partial \mathcal{Q}}{\partial \mu_2}(\vec{\mu})$:

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left((x_1-\mu_1)^2+(x_2-\mu_2)^2\right) f(x_1,x_2)\,dx_1 dx_2 \\ \Longrightarrow \quad \frac{\partial}{\partial \mu_1}\mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right] &= -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mu_1) f(x_1,x_2)\,dx_1 dx_2 \\ \frac{\partial}{\partial \mu_2}\mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right] &= -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_2-\mu_2) f(x_1,x_2)\,dx_1 dx_2. \end{aligned}$$

To find the argmin of $\mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right]$, we set both partial derivatives to 0, and obtain:

$$\mathbf{argmin}_{\vec{\mu}\in\mathbb{R}^2}\mathbb{E}\left[\|\mathbf{X}-\vec{\mu}\|^2\right] = \begin{pmatrix} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1,x_2)\,dx_1 dx_2 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1,x_2)\,dx_1 dx_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left[X^{(1)}\right] \\ \mathbb{E}\left[X^{(2)}\right] \end{pmatrix}.$$

Concept check: M-estimators

1/1 point (graded)
Which of the following is true about M-estimation?
(Choose all that apply. Refer to the slides.)

- [ ] M-estimation involves estimating some parameter of interest related to the underlying, unknown distribution (e.g. its mean, variance, or quantiles)
- [ ] Maximum likelihood estimation is a special case of M-estimation.
- [ ] Unlike maximum likelihood estimation and the method of moments, no statistical model needs to be assumed to perform M-estimation.
- [x] M-estimation cannot be used for parametric statistical models.

Solution:

We examine the choices in order.

- "M-estimation involves estimating some parameter of interested related to the underlying, unknown distribution (e.g. its mean, variance, or quantiles)" is correct. This is precisely the goal of M-estimation, as stated in the slides. It is a flexible approach that applies even outside of parametric statistical models.
- "Maximum likelihood estimation is a special case of M-estimation." is correct. If we set the loss function to be the

negative log-likelihood, then the same optimization problem defining the MLE is the one considered for the M-estimator associated to this loss function.

- "Unlike maximum likelihood estimation and the method of moments, no statistical model needs to be assumed to perform M-estimation." is correct. As stated above, M-estimation is a flexible approach that can used to approximate relevant quantities of interest to a distribution, such as its moments.

- "M-estimation cannot be used for parametric statistical models." is incorrect. M-estimation can be used in both a parametric and non-parametric context, though in this lecture, we will only see it applied in parametric examples.

Submit You have used 1 of 2 attempts

Relating M-estimation and Maximum Likelihood Estimation

1/1 point (graded)

Let $(E, \{\mathbf{P}_\theta\}_{\theta\in\Theta})$ denote a discrete statistical model and let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbf{P}_{\theta^*}$ denote the associated statistical experiment, where θ^* is the true, unknown parameter. Suppose that $\mathbf{P}_\theta$ has a probability mass function given by p_θ. Let $\hat{\theta}_n^{\text{MLE}}$ denote the maximum likelihood estimator for θ^*.

The maximum likelihood estimator can be expressed as an M-estimator– that is,

$$\hat{\theta}_n^{\text{MLE}} = \text{argmin}_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \rho\left(X_i, \theta\right)$$

for some function ρ.

Which of the following represents the correct choice of the function ρ so that the equation above is satisfied?

- ○ $-\ln p_\theta\left(X_i\right)$
- ○ $\ln p_\theta\left(X_i\right)$
- ○ $p_\theta\left(X_i\right)$
- ◉ None of the above.

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Median as a Minimizer

3/3 points (graded)

Assume that X is a continuous random variable with density $f : \mathbb{R} \to \mathbb{R}$. Then a **median** of X is defined to be any point $\text{med}\left(X\right) \in \mathbb{R}$ such that

$$P\left(X > \text{med}\left(X\right)\right) = P\left(X < \text{med}\left(X\right)\right) = \frac{1}{2}.$$

(Recall that for a continuous distribution, $P\left(X > \text{med}\left(X\right)\right) = P\left(X \geq \text{med}\left(X\right)\right)$.) Note: A median of a distribution is *not necessarily* unique.)

In this problem, you will show that any median satisfies

$$\mathrm{med}\left(X\right) = \mathrm{argmin}_{\mu \in \mathbb{R}} \mathbb{E}\left[\left|X - \mu\right|\right].$$

Which of the following correctly expresses $\mathbb{E}\left[\left|X - \mu\right|\right]$ in terms of the density $f\left(x\right)$?

- ○ $\int_{-\infty}^{\infty} x f\left(x\right)\, dx - \mu$
- ○ $\int_{-\infty}^{\infty} x f\left(x\right)\, dx - \mu\left(-\int_{\mu}^{\infty} f\left(x\right)\, dx + \int_{-\infty}^{\mu} f\left(x\right)\, dx\right)$
- ◉ $\int_{\mu}^{\infty} x f\left(x\right)\, dx - \int_{-\infty}^{\mu} x f\left(x\right)\, dx - \mu$
- ○ $\int_{\mu}^{\infty} x f\left(x\right)\, dx - \int_{-\infty}^{\mu} x f\left(x\right)\, dx - \mu\left(\int_{\mu}^{\infty} f\left(x\right)\, dx - \int_{-\infty}^{\mu} f\left(x\right)\, dx\right)$

Let $\mathcal{Q}\left(\mu\right) = \mathbb{E}\left[\left|X - \mu\right|\right]$ denote the expression obtained in the previous question. Then $\mathcal{Q}\left(\mu\right)$ consists of a sum of terms, each of which can be differentiated with respect to μ.

What is $\mathcal{Q}'\left(\mu\right) = \frac{d}{d\mu}\mathcal{Q}\left(\mu\right)$?

Hint: Use the product rule and the fundamental theorem of calculus.

- ○ 1
- ○ $\int_{-\infty}^{\mu} f\left(x\right)\, dx - \int_{\mu}^{\infty} f\left(x\right)\, dx$
- ◉ $4\mu f\left(\mu\right) + \int_{-\infty}^{\mu} f\left(x\right)\, dx - \int_{\mu}^{\infty} f\left(x\right)\, dx$
- ○ $4\mu f\left(\mu\right) + 1$

Using your response from the previous question and the definition of median, what is $\mathcal{Q}'\left(\mathrm{med}\left(X\right)\right)$?

- ○ 0
- ◉ 1
- ○ $4\mathrm{med}\left(X\right) f\left(\mathrm{med}\left(X\right)\right) + 1$
- ○ Cannot be determined.

Solution:

For the first question, we have

$$\begin{aligned}
\mathbb{E}\left[\left|X - \mu\right|\right] &= \int_{-\infty}^{\infty} \left|x - \mu\right| f\left(x\right)\, dx \\
&= \int_{\mu}^{\infty} \left(x - \mu\right) f\left(x\right)\, dx + \int_{-\infty}^{\mu} \left(-x + \mu\right) f\left(x\right)\, dx \\
&= \int_{\mu}^{\infty} x f\left(x\right)\, dx - \int_{-\infty}^{\mu} x f\left(x\right)\, dx - \mu\left(\int_{\mu}^{\infty} f\left(x\right)\, dx - \int_{-\infty}^{\mu} f\left(x\right)\, dx\right)
\end{aligned}$$

$$= \int_{\mu}^{\infty} xf(x)\,dx - \int_{-\infty}^{\mu} xf(x)\,dx - \mu\left(\int_{\mu}^{\infty} f(x)\,dx - \int_{-\infty}^{\mu} f(x)\,dx\right)$$

Therefore, "$\int_{\mu}^{\infty} xf(x)\,dx - \int_{-\infty}^{\mu} xf(x)\,dx - \mu\left(\int_{\mu}^{\infty} f(x)\,dx - \int_{-\infty}^{\mu} f(x)\,dx\right)$" is the correct answer to the first question.

For the second question, we differentiate the previous answer term by term. We have, by the fundamental theorem of calculus and the product rule that

$$\begin{aligned} \frac{d}{d\mu}\left(\int_{\mu}^{\infty} xf(x)\,dx\right) &= -\mu f(\mu) \\ \frac{d}{d\mu}\left(-\int_{-\infty}^{\mu} xf(x)\,dx\right) &= -\mu f(\mu) \\ \frac{d}{d\mu}\left(-\mu\left(\int_{\mu}^{\infty} f(x)\,dx - \int_{-\infty}^{\mu} f(x)\,dx\right)\right) &= -\int_{\mu}^{\infty} f(x)\,dx + \int_{-\infty}^{\mu} f(x)\,dx + 2\mu f(\mu)\,. \end{aligned}$$

Adding these terms, we have cancellations, yielding

$$\frac{d}{d\mu}\mathcal{Q}(\mu) = -\int_{\mu}^{\infty} f(x)\,dx + \int_{-\infty}^{\mu} f(x)\,dx.$$

Therefore, the correct response to the second question is "$\int_{-\infty}^{\mu} f(x)\,dx - \int_{\mu}^{\infty} f(x)\,dx$".

For the third question, by definition, the median $\text{med}(X)$ of X is a real number that satisfies $P(X > \text{med}(X)) = P(X < \text{med}(X))$. Therefore,

$$\mathcal{Q}'(\text{med}(X)) = \int_{-\infty}^{\text{med}(X)} f(x)\,dx - \int_{\text{med}(X)}^{\infty} f(x)\,dx = P(X < \text{med}(X)) - P(X > \text{med}(X)) = 0.$$

The correct response is "0".

Submit

You have used 2 of 2 attempts

Quantile as a Minimizer

7/7 points (graded)

Recall from the lecture that the **check function** is defined as

$$C_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x < 0 \\ \alpha x & \text{if } x \geq 0. \end{cases}$$

Assume that X is a continuous random variable with density $f : \mathbb{R} \to \mathbb{R}$. Define the **$\alpha$-quantile** of X to be $Q_X(\alpha) \in \mathbb{R}$ such that

$$\mathbf{P}(X \leq Q_X(\alpha)) = \alpha.$$

(Here we have used a different convention of the definition of the quantile function from before, where for a standard normal distribution, q_α is such that $P(X > q_\alpha) = \alpha$.)

Just like for the median, whether Q_α is unique depends on the distribution.

In this problem, you will convince yourself that any α-quantile of X satisfies

$$Q_X(\alpha) = \text{argmin}\ \mathbb{E}[C_\alpha(X - \mu)]$$

$$Q_X(\alpha) = \text{argmin}_{\mu\in\mathbb{R}} \mathbb{E}\left[C_\alpha(X-\mu)\right].$$

First, compute $\mathbb{E}\left[C_\alpha(X-\mu)\right]$. Answer by entering the coefficients $A,\ B,\ C,\ D$ in terms of α and μ in the expression below:

$$\begin{aligned}\mathbb{E}\left[C_\alpha(X-\mu)\right] &= A\int_{-\infty}^{\mu} xf(x)\ dx + B\int_{\mu}^{\infty} xf(x)\ dx\\ &\quad +C\int_{-\infty}^{\mu} f(x)\ dx + D\int_{\mu}^{\infty} f(x)\ dx.\end{aligned}$$

$A =$

$B =$

$C =$

$D =$

Second, let $F(\mu) = \mathbb{E}\left[C_\alpha(X-\mu)\right]$ denote the expression obtained in the question above. Find $F'(\mu)$. Answer by entering the coefficients $E,\ G,\ H,$ in terms of α and μ below:

$$F'(\mu) = \left(\mathbb{E}\left[C_\alpha(X-\mu)\right]\right)' = E + G\left(\mu f(\mu)\right) + H\int_{-\infty}^{\mu} f(x)\ dx.$$

$E =$

$G =$

$H =$

Finally, set $F'(\mu) = 0$ to find $\textbf{argmin}_{\mu\in\mathbb{R}} F(\mu) = \textbf{argmin}_{\mu\in\mathbb{R}} \mathbb{E}\left[C_\alpha(X-\mu)\right]$. (There is no answer box for this question.)

STANDARD NOTATION

Submit You have used 1 of 3 attempts

Correct (7/7 points)

(Optional) Convexity of the Expectation of the Loss Function

Strict convexity of $\mathcal{Q}(\mu) = \mathbb{E}\left[\rho(X,\mu)\right]$ ensures that it has a unique minimum, and this is guaranteed by strict convexity of $\rho(X,\mu)$ in μ. We will explain the univariate case below.

Expectation of a convex function is convex

Let X be a random variable with some unknown distribution $\mathbf{P}$ with some associated parameter μ^* on some sample space E. Let $\rho : E\times\mathcal{M}\to\mathbb{R}$, where $\mathcal{M}$ is the set of all possible values of the unknown true parameter μ^* and let $\mathcal{Q}(\mu) = \mathbb{E}\left[\rho(X,\mu)\right]$.

$$\rho(X,\mu)\ \ \textbf{strictly convex in}\ \mu \quad\Longrightarrow\quad \mathbb{E}\left[\rho(X,\mu)\right]\ \ \textbf{strictly convex in}\ \mu.$$

Proof:

Recall that ρ being strictly convex in μ means that

$$t\rho(x,\mu_1)+(1-t)\rho(x,\mu_2)-\rho(x,t\mu_1+(1-t)\mu_2)>0 \quad \text{for all } x.$$

Taking the expectation of the above inequality gives:

$$\mathbb{E}\left[t\rho(X,\mu_1)+(1-t)\rho(X,\mu_2)-\rho(X,t\mu_1+(1-t)\mu_2)\right] \quad = \quad t\mathbb{E}\left[\rho(X,\mu_1)\right]+(1-t)\mathbb{E}\left[\rho(X,\mu_2)\right]-\mathbb{E}\left[\rho(X,t\mu_1+(1-t$$

for any $\mu_1 \neq \mu_2 \in \mathcal{M}$, and $t \in (0,1)$. This is because $\mathbb{E}[f(X)]>0$ if $f(X)>0$. (Think about why: in the discrete case, this can roughly be read as "the weighted average of a collection of positive numbers is positive".) The above inequality exactly implies strict convexity of $\mathbb{E}[\rho(X,\mu)]$.

Hide

Discussion

Hide Discussion

Topic: Unit 3 Methods of Estimation:Lecture 12: M-Estimation / 2. Introduction to M-estimation

Add a Post

Show all posts by recent activity

- ? Loss function 2
- Calculus references 2
  Can someone point me to some useful materials for the calculus used here please?
- missing steps to evaluate the intergrals 7
  hi, can someone please post the full details how to evaluate the intergrals x*f(x) please. It is not immediately obvious to me how to get the -mu*f...
- Beautiful exercise 5
  I had to go back to Kahn Academy to recall the Fundamental Theorem of Calculus, but the final result was worth the effort. This check function i...
- ? Notation for "as a function of"? 1
- L12Q4 - Quantile as a Minimizer 8
  Community TA
- Note: mu is not the mean here 2
  For the last two problems (and associated parts of lecture) mu is just being used as an unknown parameter/variable, not as the mean of X. That ...

Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store