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 <u>Privacy Policy</u>.





Lecture 9: Introduction to Maximum

Course > Unit 3 Methods of Estimation > Likelihood Estimation

> 5. Maximum Likelihood Estimator

Currently enrolled in Audit Track (expires December 25, 2019) Upgrade (\$300)

5. Maximum Likelihood Estimator

Review: Maximizing composite functions

0/1 point (graded)

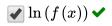
The **arguments of the minima** ($\mathit{resp.}$ **arguments of the maxima**) of a function f(x), denoted by $\mathop{\rm argmin} f(x)$ ($\mathit{resp.}$ $\mathop{\rm argmax} f(x)$), is the value(s) of x at which f(x) is minimum ($\mathit{resp.}$ maximum). We can also restrict to a subset S of the domain of f, and denote by $\mathop{\rm argmin}_{x \in S} f(x)$ ($\mathit{resp.}$ $\mathop{\rm argmax}_{x \in S} f(x)$) the value(s) of $x \in S$ at which f(x) is minimum ($\mathit{resp.}$ maximum) over S.

Let f(x) > 0 be continuous **positive** function with $\max_x f(x) = 1$. (Note that $\max_x f(x)$ is the maximum value of the function, which is different from $\underset{x}{\operatorname{argmax}} f(x)$, the value of the argument x at which the function is maximum.)

Which of the following functions of f(x) has the same argmax as f(x)? In other words, which of the following attain their maxima at the same x-value(s) as f(x)? (Choose all that apply.)







$$-\ln\left(\frac{1}{f(x)}\right)$$

$$\cos(f(x))$$

$$-\cos\left(2f\left(x
ight)
ight)$$



Solution:

We go through the choices in order.

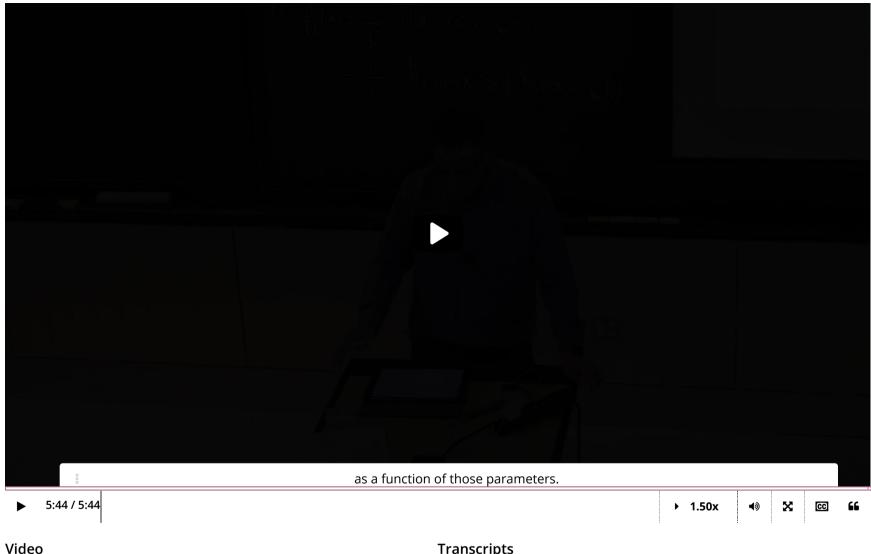
- Since y^2 , \sqrt{y} , $\ln{(y)} = -\ln{\left(\frac{1}{y}\right)}$ are all **strictly increasing** functions, their value increases as y increases. Hence, the functions $f(x)^2$, $\sqrt{f(x)}$, $\ln{(f(x))}$, $-\ln{\left(\frac{1}{f(x)}\right)}$ attain their maxima when f(x) attain its maximum, which is at $x = \operatorname{argmax} f(x)$.
- ullet The cosine function is strictly decreasing in $(0,\pi)$. Given $\max_x f(x) = 1 < \pi, \, \cos{(f(x))}$ is in fact minimum when f(x) is maximum.
- On the other hand, $-\cos{(2y)}$ is strictly increasing for $0 < 2y < \pi$. Since $\max_x 2f(x) = 2 < \pi$, we conclude that $-\cos{(2f(x))}$ is maximum again when f(x) is maximum, at $x = \operatorname{argmax} f(x)$.

Submit

You have used 2 of 2 attempts

Answers are displayed within the problem

Definition of Maximum Likelihood Estimator and Log Likelihood



Download video file

Transcripts Download SubRip (.srt) file Download Text (.txt) file

Concept Check: Interpreting the Maximum Likelihood Estimator

1/1 point (graded)

Let $X_1,\ldots,X_n\stackrel{iid}{\sim} \mathbf{P}_{\theta^*}$ be discrete random variables. We construct a statistical model $(E,\{\mathbf{P}_{\theta}\}_{\theta\in\mathbb{R}})$ where \mathbf{P}_{θ} has pmf p_{θ} . We observe our sample to be $X_1=x_1,X_2=x_2,\ldots,X_n=x_n$. The **maximum likelihood estimator** for θ^* is defined to be

$$\hat{ heta_n}^{MLE} = \operatorname{argmax}_{ heta \in \mathbb{R}} \left(\prod_{i=1}^n p_{ heta}\left(X_i
ight)
ight).$$

Which of the following is a correct interpretation of the maximum likelihood estimator (MLE) when applied to the sample $X_1=x_1,X_2=x_2,\ldots,X_n=x_n$? (Choose all that apply.)

ightharpoonup The value of heta that maximizes the probability that $\mathbf{P}_{ heta}$ generates the data set (x_1,\ldots,x_n) .

 $m{\ell}$ The value of heta that minimizes an estimator of the KL divergence between ${f P}_ heta$ and the true distribution ${f P}_{ heta^*}.$

 \blacksquare It is the true parameter $heta^*$



Solution:

• "The value of θ that maximizes the probability that \mathbf{P}_{θ} generates the data set (x_1, \dots, x_n) ." is correct. Since the likelihood is the joint density of n iid samples from \mathbf{P}_{θ} ,

$$\mathbf{P}_{ heta}\left[X_1=x_1,\ldots,X_n=x_n
ight]=L_n(x_1,\ldots,x_n, heta).$$

Hence, the MLE finds $\hat{ heta}_n$ that maximizes the probability that x_1,\dots,x_n were sampled from $P_{\hat{ heta}_n}$.

- "The value of θ that minimizes the KL divergence between \mathbf{P}_{θ} and the true distribution \mathbf{P}_{θ^*} ." is correct. In fact, this is how the MLE was derived from KL divergence. See the third section "Parameter Estimation via KL Divergence" of this lecture to review this fact.
- "It is the true parameter θ^* " is incorrect. The MLE is an estimator– it is constructed from the finite amount of data x_1, \ldots, x_n that we are given– so we can't hope for it to exactly recover the true parameter.

Remark: Under some technical conditions the MLE is a **weakly consistent estimator** for θ^* , meaning that the MLE will converge to θ^* in probability under these conditions. However, there are examples of statistical models where the maximum likelihood estimator will **not** converge to the true parameter.

Submit

You have used 1 of 2 attempts

1 Answers are displayed within the problem

Discussion

Show all posts

Hide Discussion

by recent activity ▼

Topic: Unit 3 Methods of Estimation:Lecture 9: Introduction to Maximum Likelihood Estimation / 5. Maximum Likelihood Estimator

Add a Post

There are no posts in this topic yet.

×

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**

10/2/2019	5. Maximum Likelihood Es	5. Maximum Likelihood Estimator Lecture 9: Introduction to Maximum Likelihood Estimation 18.6501x Courseware edX			
		Learn About Verified Certificates			
		Learn About Vermed Certificates	© All Rights R	eserved	