edX

3. The Empirical Cumulative Distribution Functions
True or False on CDF

1/1 point (graded)

True or False: Let X be a random variable with cdf $F(x)$. Then

$$F(x) \;=\; P(X \leq x) = \mathbb{E}\left[\mathbf{1}\left(X \leq x\right)\right],$$
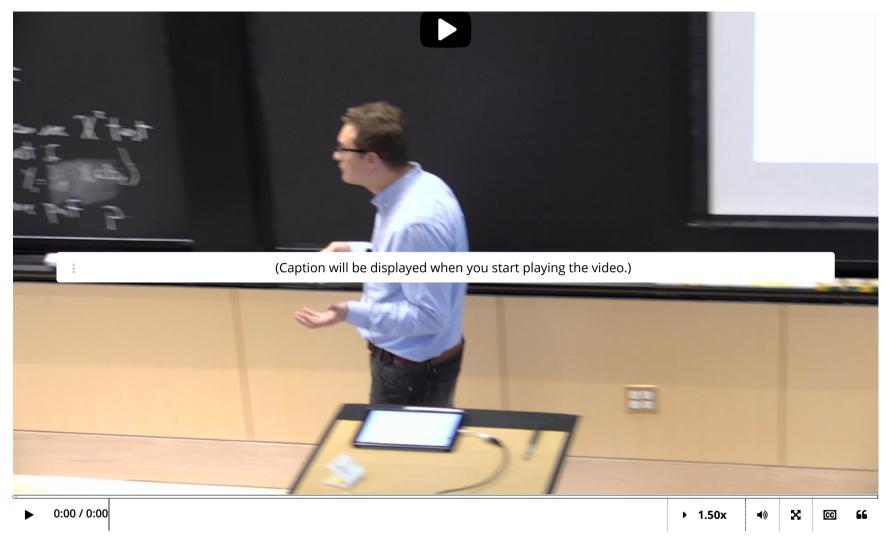
where $\mathbf{1}$ is the indicator function.

- ⦿ True

- ◯ False

✔

Solution:

By definition,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

$$= \int_{-\infty}^{\infty} f(t)\, \mathbf{1}(X \leq x)\, dt$$

$$= \mathbb{E}\left[\mathbf{1}(X \leq x)\right].$$

Submit   You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

Empirical Cumulative Distribution

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

▶ 0:00 / 0:00   ▸ 1.50x ◀)) ✕ CC "

Video
[Download video file](#)

Transcripts
[Download SubRip (.srt) file](#)
[Download Text (.txt) file](#)

Empirical Cumulative Distribution Function

Let X be a random variable with distribution $\mathbf{P}$. Recall the cdf of $\mathbf{P}$ is given by the function

$$F : \mathbb{R} \quad \to [0, 1]$$

$$t \quad \mapsto \mathbf{P}\left(X \le t\right).$$

Let $X_1, \ldots, X_n \overset{iid}{\sim} X$. The **empirical cumulative distribution function** , also called the **empirical cdf** , is the random function

$$F_n : \mathbb{R} \quad \to [0, 1]$$

$$t \quad \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(X_i \le t\right).$$

The empirical cdf depends on n and the observed data $X_i, i = 1, \ldots, n$.

Example of Empirical CDF

5/5 points (graded)

Let $X_1, \ldots, X_5$ be i.i.d. random variables. You obtain the sample $X_1 = 5$, $X_2 = 1.5$, $X_3 = -3$, $X_4 = 0.0$, $X_5 = 7$.

Let $F\left(t\right)$ be the empirical cdf of this sample. Find

Find $F\left(-4\right)$.

$F\left(-4\right) =$ | 0 | ✔ **Answer:** 0

Find $F\left(-3\right)$.

$F\left(-3\right) =$ | 1/5 | ✔ **Answer:** 1/5

Find $F\left(10\right)$.

$F\left(10\right) =$ | 1 | ✔ **Answer:** 1

Find the largest interval of t for which $F(t) = 3/5$. Answer by entering A and B in the equation below:

$F(t) = 3/5$ for $A \leq t < B$ where

$A =$ | 1.5 | ✔ **Answer:** 1.5

$B =$ | 5 | ✔ **Answer:** 5

Solution:

Given the sample $X_1 = 5$, $X_2 = 1.5$, $X_3 = -3$, $X_4 = 0.0$, $X_5 = 7$, the empirical cdf is

$$F(t) = \frac{1}{5}\sum_{i=1}^{5} \mathbf{1}(X_i \leq t)$$

$$= \begin{cases} 0 & \text{if } t < -3 \\ 1/5 & \text{if } -3 \leq t < 0.0 \\ 2/5 & \text{if } 0 \leq t < 1.5 \\ 3/5 & \text{if } 1.5 \leq t < 5 \\ 4/5 & \text{if } 5 \leq t < 7 \\ 1 & \text{if } 7 \leq t \end{cases}.$$

Hence $F(-4) = 0$, $F(-3) = 1/5$, $F(10) = 1$, and $F(t) = 3/5$ for $1.5 \leq t < 5$.
Remark: The empirical cdf is right-continuous, just like the cdf.

Submit | You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Hypothesis testing:Lecture 16: Goodness of Fit Tests Continued: Kolmogorov-Smirnov test, Kolmogorov-Lilliefors test, Quantile-Quantile Plots / 3. The Empirical Cumulative Distribution Functions

Add a Post

Show all posts ▼   by recent activity ▼

There are no posts in this topic yet.

✖