

### 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

$$F(x) = P(X \leq x) = \mathbb{E}[\mathbf{1}(X \leq x)],$$

where  $\mathbf{1}$  is the indicator function.

☒ True ✓

☐ False

#### Solution:

By definition,

$$\begin{aligned} F(x) = P(X \leq x) &= \int_{-\infty}^x f(t) dt \\ &= \int_{-\infty}^{\infty} f(t) \mathbf{1}(X \leq x) dt \\ &= \mathbb{E}[\mathbf{1}(X \leq x)]. \end{aligned}$$

Submit

You have used 1 of 1 attempt

**i** Answers are displayed within the problem

### Empirical Cumulative Distribution



And the point.

I have one.

So I'm actually here.

OK.

So that's correct, OK?

And then you just move by step like this.

Sometimes the steps are a little longer.

And then you end up here, at the last point.

You end up at 1, OK?

So that's what the empirical CDF looks

like, just this step function.

like, just this step function.

Video  
[Download video file](#)

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



Empirical Cumulative Distribution Function

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

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto \boldsymbol{P}(X \leq t). \end{aligned}$$

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

$$\begin{aligned} F_n : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq t). \end{aligned}$$

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

Example of Empirical CDF

5/5 points (graded)

Let  $X_1, \dots, X_5$  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  $F(t)$  be the empirical cdf of this sample. Find

Find  $F(-4)$ .

$F(-4) =$   ✔ Answer: 0

Find  $F(-3)$ .

$F(-3) =$   ✔ Answer: 1/5

Find  $F(10)$ .

$F(10) =$   ✔ Answer: 1

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

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

$A =$   ✔ Answer: 1.5

$B =$   ✔ 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 2 of 3 attempts

**i** Answers are displayed within the problem

## Discussion

Show 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