Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. It only takes a minute to sign up.
Sign up to join this community
Anybody can ask a question
Anybody can answer
The best answers are voted up and rise to the top





# Empirical CDF vs CDF

Asked 4 years, 7 months ago Active 2 years, 5 months ago Viewed 22k times

Ask Question

23

I'm learning about the Empirical Cumulative Distribution Function. But I still don't understand

1. Why is it called 'Empirical'?
2. Is there any difference between Empirical CDF and CDF?

distributions terminology cumulative-distribution-function empirical-cumulative-distr-fn

Share Cite Improve this question Follow

edited Aug 7 '17 at 11:59 Firebug 13.8k 3 49 116

asked Oct 13 '16 at 5:40 Gammaries 255 1 2 6

2 Check here stats.stackexchange.com/questions/222120/... – Tim Oct 13 '16 at 6:11

There is a simple, straightforward, elegant explanation in terms of tickets in a box models: the CDF describes what is in the original box. The ECDF is what you get when you put your sample (which is a set of tickets drawn from the original box: so-called "empirical" data) into an empty box. – whuber Oct 13 '16 at 18:57

One thing to be aware of is that your empirical distribution is usually bounded by the way it's constructed, while the CDF may not be. For instance, if you build empirical CDF from observations of Poisson variable, the obtained ECDF is going to be bounded by the highest observed frequency, while the true CDF is unbounded. – Aksakal Oct 17 '16 at 17:37

Add a comment

## 5 Answers

Active Oldest Votes

30

Let $X$ be a random variable.

- The cumulative distribution function $F(x)$ gives the $P(X \leq x)$.
- An empirical cumulative distribution function function $G(x)$ gives $P(X \leq x)$ based on the observations in your sample.

The distinction is which probability measure is used. For the empirical CDF, you use the probability measure defined by the frequency counts in an empirical sample.

### Simple example (coin flip):

Let $X$ be a random variable denoting the result of a single coin flip where $X = 1$ denotes heads and $X = 0$ denotes tails.

The CDF for a fair coin is given by:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2} & \text{for } 0 \leq x < 1 \\ 1 & \text{for } 1 \leq x \end{cases}$$

If you flipped 2 heads and 1 tail, the empirical CDF would be:

$$G(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{2}{3} & \text{for } 0 \leq x < 1 \\ 1 & \text{for } 1 \leq x \end{cases}$$

The empirical CDF would reflect that in your sample, 2/3 of your flips were heads.

### Another example ($F$ is CDF for normal distribution):

Let $X$ be a normally distributed random variable with mean 0 and standard deviation 1.

The CDF is given by:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Let's say you had 3 IID draws and obtained the values $x_1 < x_2 < x_3$. The empirical CDF would be:

$$G(y) = \begin{cases} 0 & \text{for } y < x_1 \\ \frac{1}{3} & \text{for } x_1 \leq y < x_2 \\ \frac{2}{3} & \text{for } x_2 \leq y < x_3 \\ 1 & \text{for } x_3 \leq y \end{cases}$$

With enough IID draws (and certain regularity conditions are satisfied), the empirical CDF would converge on the underlying CDF of the population.

Share Cite Improve this answer Follow Add a comment

edited Dec 10 '18 at 13:45

answered Oct 13 '16 at 6:36 Matthew Gunn 19.8k 1 8 43 80

14

> Is there any difference between Empirical CDF and CDF?

Yes, they're different. An empirical cdf is a proper cdf, but empirical cdfs will always be discrete even when not drawn from a discrete distribution, while the cdf of a distribution can be other things besides discrete.

If you treat a sample as if it were a population of values, each one equally probable (i.e. place probability 1/n on each observation) then the cdf of that distribution would be the ECDF of the data.

> Why does it called 'Empirical'?

It's an estimate of the population cdf based on the sample; specifically if you treat the proportions of the sample at each distinct data value and treat it like it was a probability in the population, you get the ECDF.

Empirical has a meaning something like "by observation rather than theory", and that's exactly what it means in this case ... using the observations to determine the distribution function.

Share Cite Improve this answer Follow Add a comment

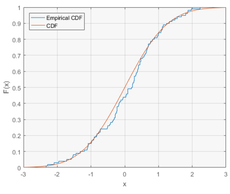
edited Dec 10 '18 at 15:17

answered Oct 13 '16 at 10:11 Glen_b 247k 28 519 891

12

The empirical CDF is built from an actual data set (in the plot below, I used 100 samples from a standard normal distribution). The CDF is a theoretical construct - it is what you would see if you could take infinitely many samples.

The empirical CDF usually approximates the CDF quite well, especially for large samples (in fact, there are theorems about how quickly it converges to the CDF as the sample size increases).




Share Cite Improve this answer Follow

answered Oct 14 '16 at 7:30 Chris Taylor 3,314 1 8 23 28

Add a comment

10

Empirical is something you build from data and observations. For instance, suppose you want to know about the distribution of the height of people in a country. You start by measuring people and come up with a histogram that can be approximated to a distribution. Then you calculate the empirical CDF.

If you are using a statistical distribution (a deterministic formula that gives the exact same output with the same parameters) you can calculate its CDF also.

You can say "The height of the people in this country is distributed similar to normal distribution with the mean 1.75 m and the standard deviation 0.1 m. Then you can use CDF of $\sim N(\mu = 1.75\ \mathrm{m}, \sigma = 0.1\ \mathrm{m})$ instead of the constructed CDF of the empirical distribution.

Share Cite Improve this answer Follow

edited Oct 17 '16 at 15:44 Nick Cox 46.1k 8 106 151

answered Oct 13 '16 at 5:49 berkorbay 338 2 10

> Is there a confidence measurement employed that expresses the likelihood that CDF and Empirical CDF describe the same population in the limit of all the experimental sampling in the world? This would seem to have application to Electoral polling, for instance. (though maybe not, since the output is not strictly describable as a function...) – BenPen Oct 13 '16 at 15:00

Add a comment

4

According to Dictionary.com, the definitions of "empirical" include:

> derived from or guided by experience or experiment.

Hence, the Empirical CDF is the CDF you obtain from your data. This contrasts with the theoretical CDF (often just called "CDF"), which is obtained from a statistical or probabilistic model such as the Normal distribution.

Share Cite Improve this answer Follow

answered Oct 13 '16 at 6:34 Waldir Leoncio 2,687 8 35 40

Add a comment

## Your Answer

Sign up or log in
Sign up using Google
Sign up using Facebook
Sign up using Email and Password
Post as a guest
Name
Email
Required, but never shown
Post Your Answer
By clicking "Post Your Answer", you agree to our terms of service, privacy policy and cookie policy

Not the answer you're looking for? Browse other questions tagged distributions terminology cumulative-distribution-function empirical-cumulative-distr-fn or ask your own question.

Featured on Meta
Testing three-vote close and reopen on 13 network sites
We are switching to system fonts on May 10, 2021
Should we reduce the vote threshold for closing questions?

Linked
26 Can anyone clarify the concept of a "sum of random variables"
6 Why is the empirical cumulative distribution of 1:1000 a straight line?
Related
7 Use Empirical CDF vs Distribution CDF?
0 Calculate expectation from empirical cdf
7 What inferential method produces the empirical CDF?
8 Discrete analog of CDF: "cumulative mass function"?
3 Goodness of fit (cdf: empirical vs theoretical)?
1 When to use empirical cdf?
3 In trouble with CDF graph
0 obtaining empirical CDF of a given data
Hot Network Questions
What language is the GPT-3 engine written in?
Grandpa's Crazy Math. Yes? No?
How to replace edges by gaps?
What is Narcissis?
Grandmother keeps calling my daughter "Good girl"
What is the intuition behind Chebychev's Inequality in Measure Theory
Are Labyrinth and Pan's Labyrinth based on some myth/legend?
An Anvil made of never melting ice
EGARCH(1,1) mean
Structure of "affirming the consequent fallacy"
What is this icon from shell32.dll
Government's standing in criminal trials
Is it true that for Cobb-Douglas preferences, the demand function is always iso-elastic?
Tikz Graph How to drop shadow like in figure
Why did Dumbledore pretend to not understand post-time-travel Harry and Hermione?
Multiple cubes as just one object?
32-bit PCI riser cards: different types?
How to explain to my supervisor that her style of input is hindering my progress?
Why are there more research papers on Android malware than iOS malware?
Making a 9 digit number divisible by 11
Wiring Ethernet through Telephone Cable
How to respond to "unfair to write English" comments?
How to prevent guns from working?
In which Chinese fairy tale does a painter disappear into his own painting?
Question feed



site design / logo © 2021 Stack Exchange Inc; user contributions licensed under cc by-sa. rev 2021.5.12.39277