4. Consistency of the Empirical CDF

Concept Checks: Empirical CDF

3/3 points (graded)

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} X$, with (true) cdf $F(t)$, and let $F_n(t)$ be the empirical cdf of $X_1, \ldots, X_n$.

What is the domain of F_n? That is, what are all the values of t for which F_n is defined.

- ○ $0 \le t \le 1$
- ◉ $-\infty \le t \le \infty$ ✔

For any t (in the domain of F_n), the empirical cdf $F_n(t)$ is

- ◉ random ✔
- ○ deterministic

For any t (in the domain of F_n), the true cdf $F(t)$ is

- ○ random
- ◉ deterministic ✔

Solution:

- Since $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le t)$ and $\mathbf{1}(X_i \le t)$ is defined for all $t \in \mathbb{R}$, the domain of F_n is also all $t \in \mathbb{R}$.
- For any t (in the domain of F_n), $F_n(t)$ is a function of the random variables $X_1, \ldots, X_n$ and hence is random.
- For any t, $F(t) = P(X \le t)$ is deterministic.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Pointwise and Uniform Convergence of Functions

A sequence of functions $g_n(x)$ **converges pointwise** to a function $g(x)$ if for each x, $\lim_{n\to\infty} g_n(x) = g(x)$.

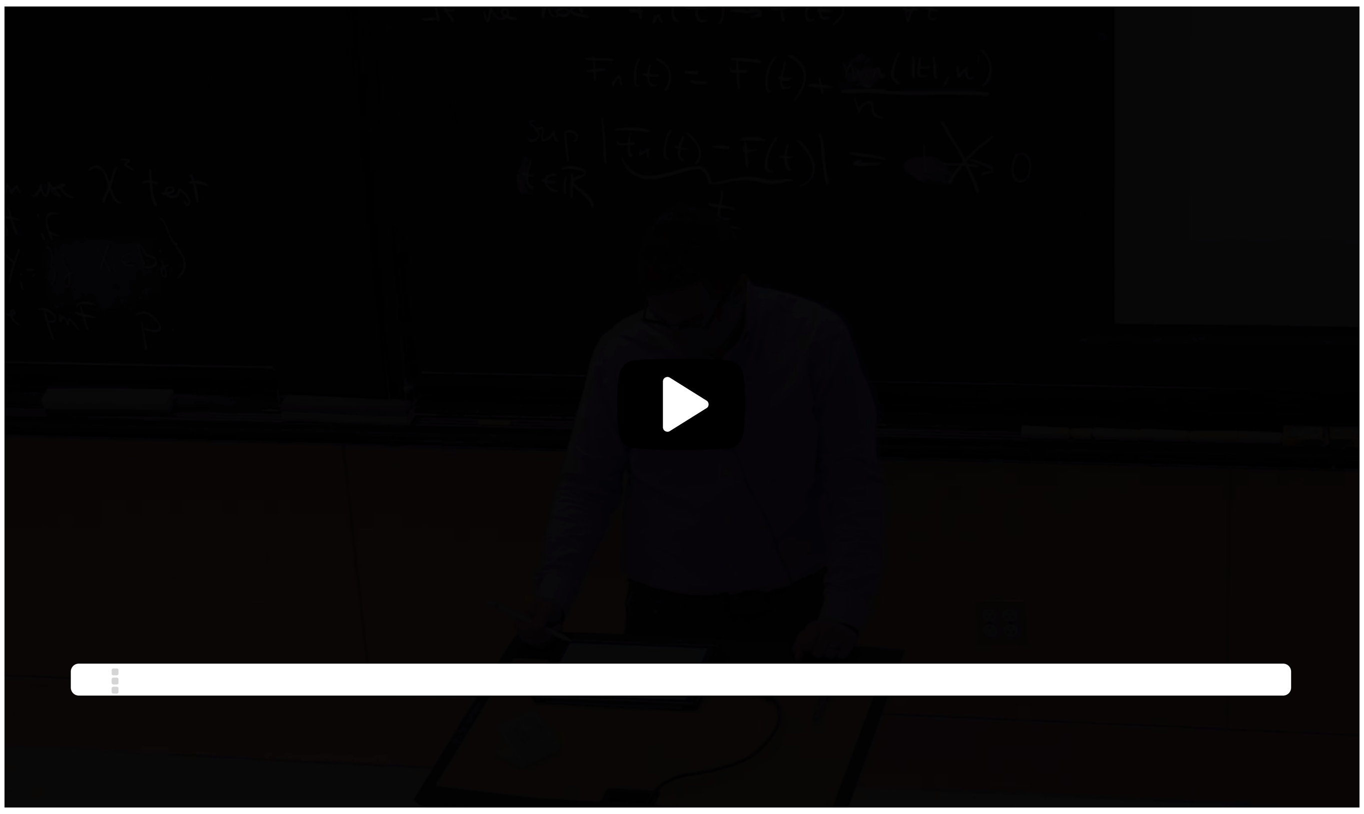
Example: In the region $x > 1$, $g_n(x) = \frac{1}{x^n}$ converges **pointwise** to $g(x) = 0$. For any fixed $x > 1$, $\frac{1}{x^n} \xrightarrow[n\to\infty]{} 0$.

A sequence of functions $g_n(x)$ **converges uniformly** to a function $g(x)$ if $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |g_n(x) - g(x)| = 0$. That is, for every $M > 0$, there exists an n_M such that $\sup_x |g_n(x) - g(x)| < M$ for all $n \ge n_M$.

Example: In the region $x > 2$, $g_n(x) = \frac{1}{x^n}$ converges **uniformly** to $g(x) = 0$, since $\sup_{x>2} g_n(x) = \sup_{x>2} \frac{1}{x^n} = \frac{1}{2^n} \xrightarrow[n\to\infty]{} 0$.

Example of pointwise but not uniform convergence:
The sequence of functions $g_n(x) = \frac{1}{x^n}$ does **not** converge uniformly to $g(x) = 0$ in the region $x > 1$, since $\sup_{x>1} g_n(x) = \sup_{x>1} \frac{1}{x^n} = 1$, which does not converge to 0 as $n \to \infty$.

Consistency of Empirical CDF, Uniform versus Pointwise Convergence, Fundamental Theorem of Statistics



6:30 / 6:30 1.0x

compact set,

things would actually work much nicer for me.

And here, the fact that I let t become large as n becomes large

is a problem.

And so essentially what you're using is the fact that Fn of t

only takes a very small number of values,

and therefore, you can actually control that.

So that's basically the idea.

End of transcript. Skip to the start.

Video
Download video file

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

xuetangX.com 学堂在线

Consistency of the Empirical cdf

2/2 points (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} X$ be i.i.d. random variables with cdf $F(t)$.

Recall the empirical cdf is the random function

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

Then following convergence holds almost surely:

$$F_n(0) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le 0) \xrightarrow[n\to\infty]{a.s.} L$$

for some value L. What is L?
(Choose all that apply.)

- [ ] 0

- [ ] 1
- [x] $F(0)$ ✔
- [ ] $F(1)$
- [x] $\mathbb{E}[\mathbf{1}(X \le 0)]$ ✔

✔

What result is invoked to obtain the value of L?

- ○ central limit theorem
- ◉ (strong) law of large numbers ✔
- ○ Slutsky's theroem

Solution:

Observe that for all i, $\mathbf{1}(X_i \le 0)$ is a Bernoulli random variable with mean $P(X_i \le 0) = F(0)$. By the law of large numbers,

$$F_n(0) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le 0) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[\mathbf{1}(X_1 \le 0)] = F(0)\,.$$

Therefore, $L = F(0)$.

Remark: It holds in general that for any $t \in \mathbb{R}$,

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le t) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[\mathbf{1}(X_1 \le t)] \;=\; F(t)$$

Submit You have used 3 of 3 attempts

ℹ Answers are displayed within the problem

Let $X_1, \ldots, X_n \overset{iid}{\sim} X$ be i.i.d. random variables with cdf $F(t)$ and empirical cdf $F_n(t)$.

The **Glivenko-Cantelli theorem** , also known as the **Fundamental Theorem of Statistics** , states that

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n\to\infty]{a.s.} 0.$$

This is a stronger result than the one in the problem above in that the convergence happens **uniformly** over t. This means for all large enough n and for any $\delta > 0$, the difference $|F_n(t) - F(t)|$ is bounded above by δ for all t. Almost sure convergence means that for all $\delta > 0$ and $\epsilon > 0$, there exists $N = N(\delta, \epsilon)$ such that the event $\sup_t |F_n(t) - F(t)| < \delta$ occurs with probability at least $1 - \epsilon$ for all $n > N$. In other words, with probability approaching 1, the function F_n is a close L_∞ (the sup-norm) approximation of F.

Discussion

Hide Discussion

Add a Post

All Posts

Almost sure convergence

discussion posted 4 days ago by mrBB (Community TA)

The very last sentence on the page is: "Almost sure convergence means that the above happens for every point in a set with probability 1." I don't understand the part "for every point in a set". I thought we just concluded that uniform convergence is not a property of single points but of an entire function. It seems meaningless to me to say we have UC of $F_n(t)$ to $F(t)$ at point $t = a$ but not at $t = b$.

Can't/shouldn't we just say "Almost sure convergence means that the above happens with probability 1."

This post is visible to everyone.

Add a Response

2 responses

Mark_B2
4 days ago

A function is a map from one set of points to another one.

Add a comment

younhun (Staff)
4 days ago

There's two degrees of freedom here. One is in t, the point at which you are evaluating the CDF.

For every choice of t, there is an associated sequence of random variables, $F_n(t)$, and this is the other degree of freedom. The claim is that the sequence $F_1(t), F_2(t), \ldots$ converges almost surely to $F(t)$, in a uniform way over all t. E.g.

$\mathbf{Pr}(|F_n(t) - F(t)| < \delta) \to 1$ **for all t** (Statement #1)

in such a way that

$\mathbf{Pr}(|F_n(t) - F(t)| < \delta \; \forall \; t) \to 1$. (Statement #2 -- the event inside the parenthesis is what the words "above happens for all points in a set" is referring to.)

Do you see why Statement #2 is stronger than #1? In #1, you can have the pathological behavior where the "large enough n" changes depending on t, and so for any finite n, the difference $|F_n(t) - F(t)|$ could be large somewhere other than the t you care about.

In other words, #2 exhibits **uniformity** of the random function F_n whereas #1 does not, but both imply almost sure convergence of the random variable $F_n(t)$. Hence we obtain the much stronger statement $\sup_t |F_n(t) - F(t)| \to 0$, which says "for large enough n, the function F_n is a pretty good guess for F in terms of infinity-norm error $\|F_n - F\|_\infty$."

It seems you misunderstood my remark. If I understand your post correctly you're explaining the difference between pointwise and uniform convergence to me. But I don't think I had a problem understanding these two concepts. I was just commenting on the formulation "**Almost sure convergence** means that the above happens for every point in a set with probability 1."

I repeat myself, but "the above" refers to $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n\to\infty]{a.s.} 0$ (note the $\sup$ part of the expression) and the sentence implies that this can hold for some points but not for all, which seems nonsensical (as this is not a statement about a single point but a statement about a function). I.e. either "the above" happens, or it doesn't happen, but it can't happen only for some points: either we have uniform convergence or we don't.

I get the impression, but I'm speculating here, (the writer of) the sentence starts off giving a definition of a.s. convergence but then halfway the sentence changes mind and starts defining uniform convergence.

And an additional comment. The preceding sentence "This means for all large enough n, the difference $|F_n(t) - F(t)|$ is bounded above by the same number δ for all t." also seems inaccurate or at least not strong enough, as it doesn't express that we can take for δ an arbitrarily small positive number. (Similar issue as discussed here.)

posted 2 days ago by mrBB (Community TA)

All good points, but the misunderstanding here from both sides (the person who wrote this originally and you) is that in the pre-build source code of the notes, "the above" really does refer to what I said, because that remark about "$|F_n(t) - F(t)|$ being bounded by δ" does appear right above that sentence. So it's a wording issue that didn't translate well to the edX page rendering.

And yes, it should be "for all $\delta > 0$ and $\epsilon > 0$, there exists $N = N(\delta, \epsilon)$ such that $\mathbf{Pr}(|F_n(t) - F(t)| < \delta) > 1 - \epsilon$ is true for all $n > N$", if we wanted to be super rigorous (there's kind of an implicit consideration here that we really only care about small δ's.)

I'll fix both, but just wanted to reassure you that you have it correct.

posted 2 days ago by **younhun** (Staff)

That's a very nice fix! 😃

posted a day ago by **mrBB** (Community TA)

Add a comment

Showing all responses

Add a response:

Preview

Submit