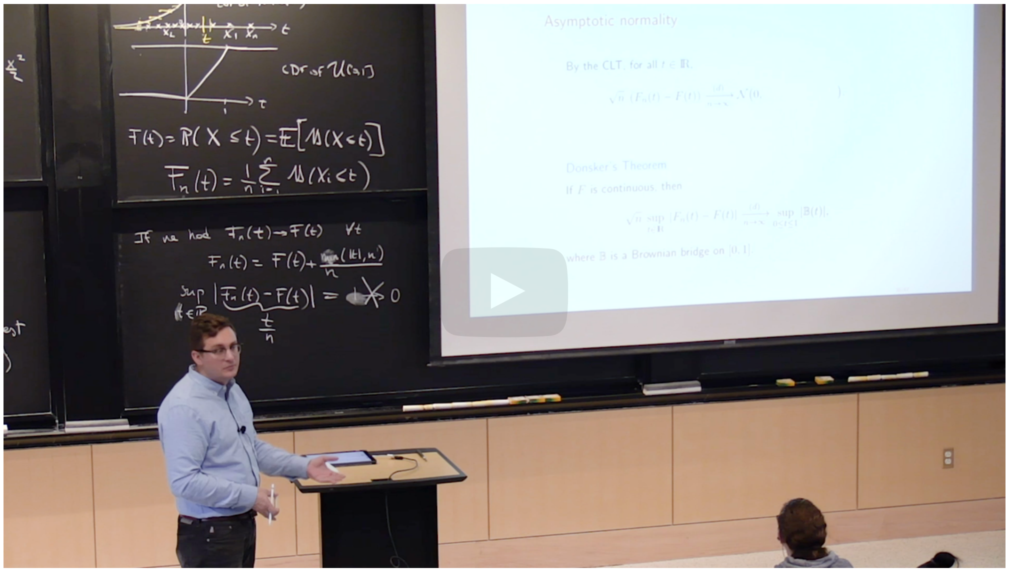
Exercises due Aug 3, 2021 19:59 EDT Completed

Asymptotic Normality of the Empirical CDF

Start of transcript. Skip to the end.



0:00 / 0:00 1.50x

OK, so now let's look at-- you know, we have convergence.

We have uniform convergence by Glovenko-Cantelli.

And now I would like to talk about asymptotic normality,

right?

Law of large numbers, central limit theorem,

that's the normal thing to do.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Pointwise Asymptotic Normality of the Empirical CDF

3.0/3 points (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathbf{P}$ for some distribution $\mathbf{P}$, and let F denote its cdf. Let F_n denote the empirical cdf. Then it holds for every $t \in \mathbb{R}$ that

$$\sqrt{n}\left(F_n(t) - F(t)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, \sigma^2\right)$$

for any fixed t and for some asymptotic variance σ^2.

What theorem implies the above convergence statement?

- ( ) Central limit theorem.
- (•) Law of large numbers.
- ( ) Glivenko-Cantelli theorem.

Which of the following is σ^2? Note that σ^2 is dependent on t.

- ( ) $F(t)$
- (•) $1 - F(t)$
- ( ) $\sqrt{F(t)(1 - F(t))}$
- ( ) $F(t)(1 - F(t))$

What is the asymptotic variance σ^2 of $F_n(0)$ in terms of the values of the cdf F? (Enter **F(x)** for $F(x)$ for any numerical value x.)

$\sigma^2 =$ [ ]

STANDARD NOTATION

Submit You have used 1 of 3 attempts

Donsker's Theorem

A stronger result than the one in the previous problem holds.

Let $X_1, \dots, X_n \stackrel{iid}{\sim} X$ for some distribution $\mathbf{P}$ with cdf F. Let F_n denote the empirical cdf of $X_1, \dots, X_n$.

Donsker's theorem states that if the true cdf F is continuous, then

$$\sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \to \infty]{(d)} \sup_{0 \le x \le 1} |\mathbb{B}(x)|,$$

where $\mathbb{B}$ is a random curve called a **Brownian bridge**.

The definition of $\mathbb{B}$ is outside the scope of this course. What we need to know about it is the fact that $\sup_{0 \le x \le 1} |\mathbb{B}(x)|$ is a **pivotal** distribution, i.e. it does not depend on the unknown distribution of the data, and hence we can look up its quantiles in tables or by using software. This will be important as we develop goodness of fit tests for continuous distributions.

Discussion

Hide Discussion

Topic: Unit 4 Hypothesis testing:Lecture 16: Goodness of Fit Tests Continued: Kolmogorov-Smirnov test, Kolmogorov-Lilliefors test, Quantile-Quantile Plots / 5. Asymptotic Normality of the Empirical CDF

Add a Post

Show all posts by recent activity

- Grab some extra here : Brownies — 4
  This speaks my heart's language for this brownie thing. https://www.youtube.com/watch?v=hqSnruUe3tA Hope it ...
- ? The difference between the 2 t's — 7
  I couldn't understand that part!
- ? Videos of Lecture 15 and 16 — 4
  Seems like the recent lectures (15 & 16) shows the professor talking to the in-person class and explaining what is o...
- Variance of empirical CDF — 2