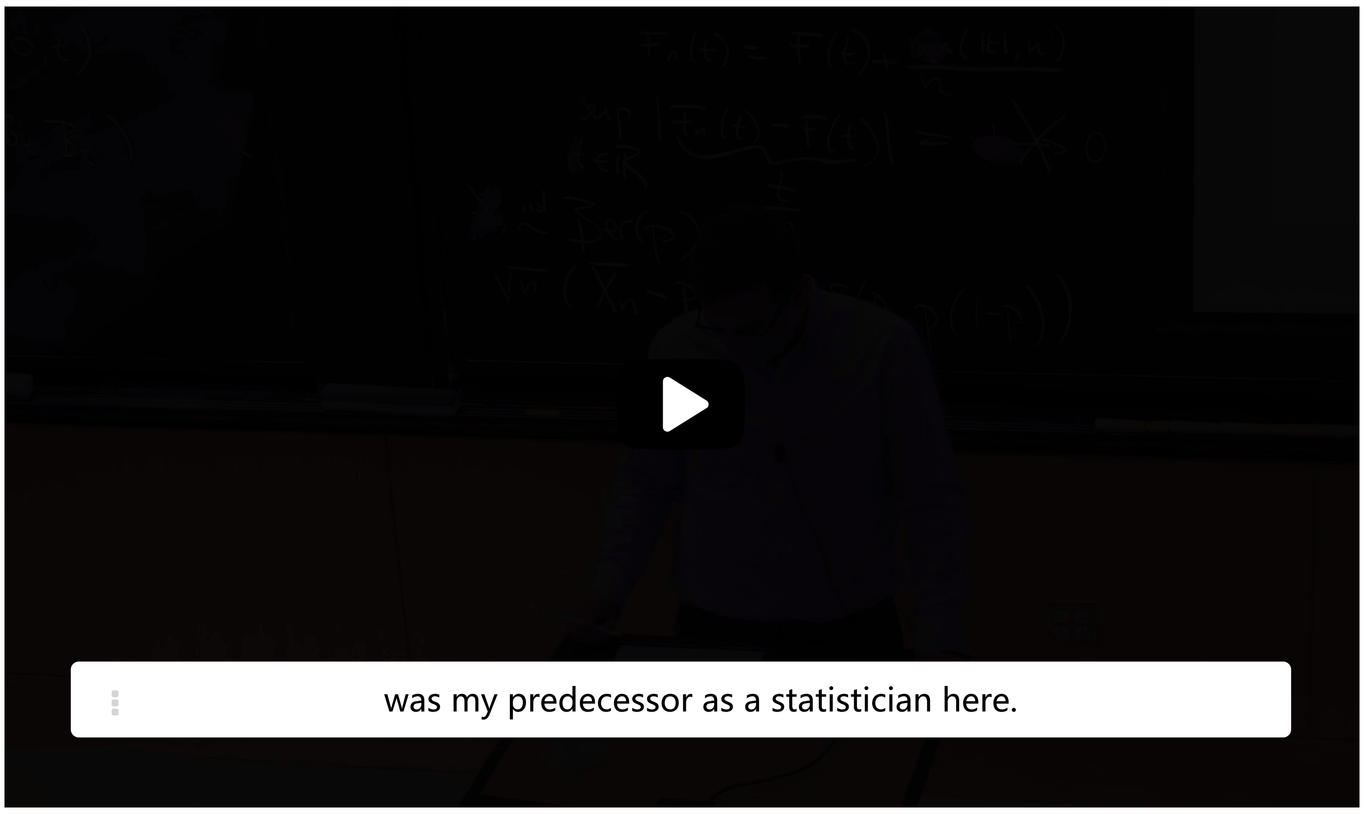
5. Asymptotic Normality of the Empirical CDF

Asymptotic Normality of the Empirical CDF



9:19 / 9:19 1.0x

the soup over,

you end up with a Brownian bridge.

That's Donsker's theorem.

So you should think of this as being a uniform central limit

theorem.

That's what it is.

There was actually a book on uniform central limit theorems

that was written by Richard Dudley, who

was my predecessor as a statistician here.

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

Pointwise Asymptotic Normality of the Empirical CDF

3/3 points (graded)

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbf{P}$ 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

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

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

What theorem implies the above convergence statement?

- (x) 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.)

$\sigma^2 =$ F(0)*(1-F(0)) ✔ **Answer:** F(0)*(1 - F(0))

STANDARD NOTATION

Solution:

Recall that the empirical CDF is given by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le t).$$

Moreover, $\mathbb{E}[\mathbf{1}(X_i \le t)] = P(X_1 \le t) = F(t)$, by definition. Therefore,

$$\sqrt{n}(F_n(t) - F(t)) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{1}(X_i \le t) - \mathbb{E}[\mathbf{1}(X_i \le t)])\right).$$

The random variables $\mathbf{1}(X_i \le t)$ are iid Bernoulli with mean $P(X_i \le t) = F(t)$, so the **central limit theorem** applies:

$$\sqrt{n}(F_n(t) - F(t)) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, F(t)(1 - F(t))).$$

In summary, the correct answers are

- **Central limit theorem**, for the first question,
- $F(t)(1 - F(t))$ for the second question, and
- $F(0)(1 - F(0))$ for the third question when $t = 0$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Donsker's Theorem

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

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

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

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

Show Discussion