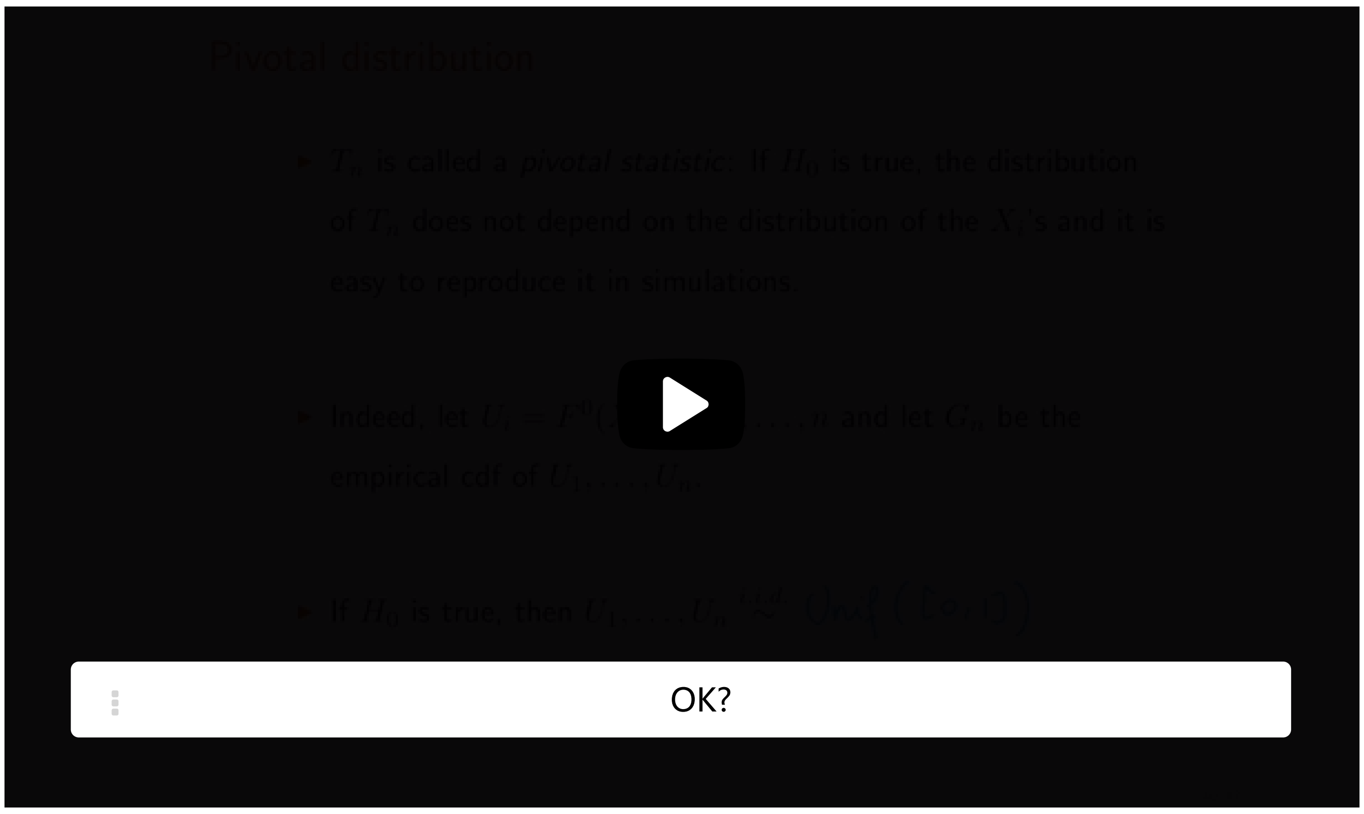
8. Kolmogorov-Smirnov Test Statistic Pivotal Under Null Non-asymptotic Distribution, Generating Data from a Given Distribution



13:16 / 13:16 1.0x

you don't have to go to the asymptotic,

no matter what n is, the distribution of tn

does not depend on any parameter of your specific test at hand.

It does not depend on the unknown f,

nor does it depend on f not.

In particular, it means that you can go into lab

and compute its p values, or its critical values.

OK?

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

CDF as a Random Function

3/3 points (graded)

Let X be a random varible with invertible cdf F_X. Define another random variable $Y = F_X(X)$. Find the cdf F_Y of Y.

For $t < 0$:

$F_Y(t) =$ 0 ✔ Answer: 0

0

For $t \geq 1$:

$F_Y(t) =$ 1 ✔ Answer: 1

1

For $0 \leq t < 1$:

$F_Y(t) =$ t ✔ Answer: t

t

(What is the distribution of Y?)

STANDARD NOTATION

Solution:

Given $Y = F_X(X)$ where F_X is a cdf, Y only takes values between 0 and 1. This means that $F_Y(t) = 0$ for all $t \leq 0$ and $F_Y(t) = 1$ for all $t > 1$.

In the region $0 \leq t < 1$

$$F_Y(t) \;=\; P(F_X(X) \leq t) \;=\; P(X \leq F^{-1}(t)) \;=\; F(F^{-1}(t)) \;=\; t.$$

We see that the cdf of Y is that of a uniform distribution with support in $[0,1]$, i.e. $Y \sim \mathrm{Unif}(0,1)$.

Remark 1: Note that $Y = F_X(X) \sim \mathrm{Unif}(0,1)$ regardless of the distribution of X as long as F_X is invertible. In the case when F_X is not invertible, modifications can be made to obtain similar result.
Remark 2: Inverting the result gives $X \sim F_X^{-1}(Y)$ where $Y \sim \mathrm{Unif}(0,1)$. This is useful for simulating data from a given distribution with cdf F_X. Start by sampling from $\mathrm{Unif}(0,1)$, and apply F_X^{-1} to the sample. The resulting sample will be from a distribution with cdf F_X.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Kolmogorov-Smirnov Test Statistic as a Pivotal Distribution Under Null Hypothesis

Let $X_1, \ldots, X_n$ be iid samples with unknown cdf F_X. For simplicity, restrict to the cases when F_X is invertible.

Recall the goal of the Kolmogorov-Smirnov Test goodness of fit test is to decide between the hypotheses

$$\begin{aligned} H_0 &: F_X = F^0 \\ H_1 &: F_X \neq F^0. \end{aligned}$$

Recall also the Kolmogorov-Smirnov test statistic:

$$T_n \;=\; \sqrt{n} \sup_{t \in \mathbb{R}} \left| F_n(t) - F^0(t) \right|$$

Assuming H_0 is true, then T_n becomes

$$T_n \;=\; \sqrt{n} \sup_{t \in \mathbb{R}} \left| F_n(t) - F_X(t) \right|$$

We will see that under the null hypothesis, the distribution of T_n does not depend on the distribution of the data X_i, i.e. T_n is pivotal, and this is true for any n, not only for large n.

The trick is to make a change of variables. Let $\tilde{t} = F_X(t)$, then $t = F_X^{-1}(\tilde{t})$. We have

$$\begin{aligned} T_n &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| F_n(t) - F_X(t) \right| \\ &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq t) \right) - F_X(t) \right| \qquad \text{(definition of empirical cdf)} \\ &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(F_X(X_i) \leq F_X(t)) \right) - F_X(t) \right| \qquad \text{(apply } F_X \text{ to both sides of inequality)} \end{aligned}$$

$$= \quad \sqrt{n} \sup_{\tilde{t} \in (0,1)} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(Y_i \leq \tilde{t}\right) \right) - \tilde{t} \right| \qquad \text{where } Y_i \sim \mathsf{Unif}\,(0,1)\,.$$

Discussion

Show Discussion

Topic: Unit 4 Hypothesis testing:Lecture 16: Goodness of Fit Tests Continued: Kolmogorov-Smirnov test, Kolmogorov-Lilliefors test, Quantile-Quantile Plots / 8. Kolmogorov-Smirnov Test Statistic Pivotal Under Null