8. Kolmogorov-Smirnov Test Statistic Pivotal Under Null

Exercises due Aug 3, 2021 19:59 EDT

Non-asymptotic Distribution, Generating Data from a Given Distribution

Start of transcript. Skip to the end.

0:00 / 0:00 1.50x

So It turns out that this result is also true.

So what did we want?

Why did we use the Donsker's theorem?

Because Donsker's theorem tells us

that no matter what the true f is, asymptotically,

or even what f0 is, asymptotically,

square root of n, fn minus f not, if f not equal to f,

will converge to something which is

Video
Download video file

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

CDF as a Random Function

3 points possible (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 $t < 0$:

$F_Y(t) =$

For $t \geq 1$:

$F_Y(t) =$

For $0 \leq t < 1$:

$F_Y(t) =$

(What is the distribution of Y?)

STANDARD NOTATION

Submit You have used 0 of 3 attempts

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

$$\begin{aligned} T_n &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| F_n(t) - F_X(t) \right| \\ &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq t) \right) - F_X(t) \right| && \text{(definition of empirical cdf)} \\ &= \sqrt{n} \sup_{t \in \mathbb{R}} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(F_X(X_i) \leq F_X(t)) \right) - F_X(t) \right| && \text{(apply } F_X \text{ to both sides of inequality)} \\ &= \sqrt{n} \sup_{\tilde{t} \in (0,1)} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(Y_i \leq \tilde{t}) \right) - \tilde{t} \right| && \text{where } Y_i \sim \mathsf{Unif}(0,1). \end{aligned}$$

Discussion

Hide 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

Add a Post

Show all posts by recent activity

Any hint for the excersice? 4

< Previous Next >