2. KS and KL Tests

The problems on this page concern the data set

$$S = \{0.28, 0.2, 0.01, 0.80, 0.1\}.$$

Let x_i denote the i'th element of the data set S.

The Empirical CDF

2/3 points (graded)

Let $F_5(t)$ denote the empirical cdf of the data set above.

What is $F(0.5)$?

4/5 ✔ **Answer:** 4/5

What is $F(0.1)$?

这里我没有sort，看错了

1/5 ✖ **Answer:** 2/5

What is $F(1)$?

1 ✔ **Answer:** 1

Solution:

Recall the definition of the empirical cdf:

$$F_5(t) := \frac{1}{5}\sum_{i=1}^{5} \mathbf{1}(x_i \le t).$$

Therefore

$$\begin{aligned} F_5(0.5) &= 4/5 \\ F_5(0.1) &= 2/5 \\ F_5(1) &= 1 \end{aligned}$$

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

QQ Plot

0/1 point (graded)

Consider the QQ-plot of the data set S against the distribution $\mathrm{Unif}(0,1)$. (You may graph the plot using computational tools.)

How many points in the QQ-plot lie above the line $y = x$?

4 ✖ **Answer:** 0

Solution:

Let F denote the cdf of $\mathrm{Unif}(0,1)$, and recall that $F_5(t) = t\mathbf{1}(t \in (0,1))$. The QQ-plot is given by the points

$$\boxed{\left(F^{-1}(1/i), x_i\right), \quad i = 1, \ldots, 5.}$$

Therefore, the plot consists of the points

$$\boxed{(1/5, 0.01), (2/5, 0.1), (3/5, 0.2), (4/5, 0.28), (1, 0.8).}$$

Since the y-coordinates of all these points are less that the corresponding x-coordinates, none of the point lies above the line $y = x$.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

KS Test Statistic

2/2 points (graded)
In this problem, you will test the null and alternative hypotheses

$$\begin{aligned} H_0 &= \text{the data set is distributed as } \mathrm{Unif}(0,1) \\ H_1 &= \text{the data set is not distributed as } \mathrm{Unif}(0,1). \end{aligned}$$

What is the value of the Kolmogorov-Smirnov test statistic on the data set S? Enter $T_5^{\mathrm{KS}}/\sqrt{5}$, the KS statistic without the factor of $\sqrt{n}$, below.

$T_5^{\mathrm{KS}}/\sqrt{5} =$ 0.52 ✔ Answer: 0.52

Does this test **reject** or **fail to reject** at level $\alpha = 0.1$?

Kolmogorov-Smirnov Tables

Show

- Reject ✔
- Fail to reject

Solution:

Recall that the KS test statistic is

$$T_n^{\mathrm{KS}} = \max_{i=1,\dots,n} \{\max\left(\left|\frac{i-1}{n} - F(x_i)\right|, \left|\frac{i}{n} - F(x_i)\right|\right)\}.$$

Therefore, T_5^{KS} is the largest of the following list of numbers: 每一个数据和他前一个分位点和后一个分位点比

$$\max\left(\left|0.01 - 0\right|, \left|0.01 - 1/5\right|\right) = 0.19$$

$$\max\left(\left|0.1 - 1/5\right|, \left|0.1 - 2/5\right|\right) = 0.1$$

$$\max\left(\left|0.2 - 2/5\right|, \left|0.2 - 3/5\right|\right) = 0.4$$

$$\max\left(\left|0.28 - 3/5\right|, \left|0.28 - 4/5\right|\right) = 0.52$$

$$\max\left(\left|0.8 - 4/5\right|, \left|0.8 - 1\right|\right) = 0.2$$

Hence,

$$T_5^{\mathrm{KS}} = \sqrt{5} * 0.52 \approx 1.163$$

is the correct response to the first question. For the second question, we consult a table for the KS test statistic to find that

$$P(T_5^{\mathrm{KS}} > 1.163) = P\left(\frac{T_5^{\mathrm{KS}}}{\sqrt{5}} > 0.52\right) \in (0.05, 0.10).$$

Therefore, we **reject** H_0 at level $\alpha = 0.1$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

KL Test Statistic

2.0/2 points (graded)
What is the sample mean $\hat{\mu}$ of the data set S?

$\hat{\mu} =$ 0.278 Answer: 0.278

What is the sample variance $\widehat{\sigma^2}$ of the data set S?
(You may use either the unbiased sample variance or the MLE of the variance.)

$\widehat{\sigma^2} =$ 0.09552 Answer: 0.076

Solution:

The sample mean is

$$\hat{\mu} = \frac{0.28 + 0.2 + 0.01 + 0.8 + 0.1}{5} \approx 0.278.$$

We can use two different estimators of the variance, which are both sometimes called the sample variance, the MLE:

$$\widehat{\sigma^2}^{\mathrm{MLE}} = \frac{0.28^2 + 0.2^2 + 0.01^2 + 0.8^2 + 0.1^2}{5} - (0.278)^2 \approx 0.076,$$

or the unbiased sample variance:

$$\widehat{\sigma^2}^{\mathrm{unbiased}} = \frac{5}{4}\widehat{\sigma^2}^{\mathrm{MLE}} = 0.09552.$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

KL Test Statistic

2/2 points (graded)
In this problem, you will consider the following null and alternative hypotheses.

$$\begin{aligned} H_0 &= \text{the data set is distributed as a Gaussian distribution (for some choice of mean and variance)} \\ H_1 &= \text{the data set is not distributed as a Gaussian (for any choice of mean and variance).} \end{aligned}$$

What is the Kolmogorov-Lilliefors test statistic evaluated on the data set S? Enter $T_5^{\mathrm{KL}}/\sqrt{5}$, the KL statistic without the factor of $\sqrt{n}$, below.
(You are encouraged to use computational tools.)

Kolmogorov-Lilliefors Tables

Show

$T_5^{\mathrm{KL}}/\sqrt{5} =$ 0.29741839452455765 Answer: 0.297

Do you **reject** or **fail to reject** H_0 at level $\alpha = 0.1$ on the data set S?

- Reject
- Fail to reject ✔

Solution:

Recall that the KL test statistic is given by

$$T_n^{\mathrm{KL}} = \max_{i=1,\ldots,n} \{\max \left(\left| \frac{i-1}{n} - \Phi_{\hat{\mu},\widehat{\sigma^2}}(x_i) \right|, \left| \frac{i}{n} - \Phi_{\hat{\mu},\widehat{\sigma^2}}(x_i) \right| \right) \}.$$

To find $\Phi_{\hat{\mu},\widehat{\sigma^2}}(x_i)$, we make change of variables:

$$\boxed{\Phi_{\hat{\mu},\widehat{\sigma^2}}(x_i) = \frac{x_i - \hat{\mu}}{\sqrt{\widehat{\sigma^2}}}.}$$

Then we use the following formula to find $T_5^{\text{KL}}/\sqrt{5}$:

$$\max_{i=1,...5}\left\{\max\left(\left|\frac{i-1}{5}-\Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(x_i)\right|,\left|\frac{i}{5}-\Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(x_i)\right|\right)\right\}$$

We now proceed to get the numerical answer for the two choices of $\widehat{\sigma^2}$.

If we use $\boxed{\widehat{\sigma^2}=\widehat{\sigma^2}^{\text{MLE}}}$, then

$$\boxed{\left(\Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{MLE}}}(0.01),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{MLE}}}(0.1),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{MLE}}}(0.2),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{MLE}}}(0.28),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{MLE}}}(0.8),\ \right)\ =\ (0.16615080.25981560.38890870.50288630.9705093)\,.}$$

Therefore, $T_5^{\text{KL}}/\sqrt{5}$ is given approximately by the largest of the following list of numbers

$$\max\left(\left|0-0.17\right|,\left|1/5-0.17\right|\right)=0.17$$

$$\max\left(\left|1/5-0.26\right|,\left|2/5-0.26\right|\right)=0.14$$

$$\max\left(\left|2/5-0.39\right|,\left|3/5-0.39\right|\right)=0.21$$

$$\max\left(\left|3/5-0.5\right|,\left|4/5-0.5\right|\right)=0.3$$

$$\max\left(\left|4/5-0.97\right|,\left|1-0.97\right|\right)=1.7.$$

We conclude that $\mathbf{T_n^{KL}}/\sqrt{\mathbf{5}}\ \mathbf{=\approx\ 0.2971137}$.

On the other hand, if we use $\boxed{\widehat{\sigma^2}=\widehat{\sigma^2}^{\text{unbiased}}}$, then

$$\left(\Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(0.01),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(0.1),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(0.2),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(0.28),\ \Phi_{\hat{\mu},\widehat{\sigma^2}^{\text{unbiased}}}(0.8),\ \right)\ =\ (0.19293350.28232980.40037540.50258160.9543$$

And similar as above, $T_5^{\text{KL}}/\sqrt{5}$ is given by 0.2974184.

Hence, the two choices of estimators of the variance give the same KL statistic up to 3 decimal places in this example.

Finally, from the KL statistic table, we see that the 0.9-quantile of the KL statistic is $0.315,$ which is greater than $T_5^{\text{KL}}/\sqrt{5}=0.297$ from our data. Hence, we fail to reject the null hypothesis.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 4 Hypothesis testing:Homework 8 / 2. KS and KL Tests