بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

Simulation and Modeling-Lecture: 12

Kolmogorov- Smirnov Test

We do Kolmogorov-Smirnov Test or KS-test to test the hypothesis of numbers. Determining the numbers are uniform or not.

Algorithm:

Step-1: Define the hypothesis for uniformity

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Step-2: Arrange data in increasing order

$$R_i = i^m \ smallest\ integer$$

$$R_1 \leq R_2 \leq \cdots \ldots \ldots \ldots . \leq R_n$$

Step-3: Compute D^+ & D^- $(KS - test\ Parameter)$

$$D^+ = max\left\{\left(\frac{i}{N}\right) - R_i\right\}, 1 \leq i \leq N; i = 1,2,3 \ldots \ldots \ldots$$

$$D^- = max\left\{R_i - \left(\frac{i-1}{N}\right)\right\}, 1 \leq i \leq N; i = 1,2,3 \ldots \ldots \ldots$$

Step-4: Compute $D = max(D^+, D^-)$

Step-5: Determine the critical value D_α , for specified level of significant α. (This will be given)

Step-6: If $D > D_\alpha => H_0$ is rejected. That means numbers are not uniform

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Problem:

The sequence of numbers **0.63, 0.49, 0.24, 0.57, 0.71, 0.89** has been generated. Use KS-test with $\alpha = 0.05$ to determine if the hypothesis that the numbers are uniformly distributed on the interval [0,1] can be rejected. [$D_{0.05} = 0.521$]

Solution:

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Arranging numbers in increasing order,

$$0.24 \leq 0.49 \leq 0.57 \leq 0.63 \leq 0.71 \leq 0.89$$

Now, compute D^+ & D^-

i	1	2	3	4	5	6
R_i	0.24	0.49	0.57	0.63	0.71	0.89
i/N	0.17	0.33	0.50	0.67	0.83	1.00
$\frac{i-1}{N}$	0	0.17	0.33	0.50	0.67	0.83
$i/N - R_i$	-	-	-	0.04	0.12	0.11
$R_i - \frac{i-1}{N}$	0.24	0.32	0.24	0.13	0.04	0.06

$D^+ = \max(0.04, 0.12, 0.11) = 0.12$

$D^- = \max(0.24, 0.32, 0.24, 0.13, 0.04, 0.06) = 0.32$

$D = max(D^+, D^-) = \max(0.12, 0.32) = 0.32$

$D_\alpha = D_{0.05} = 0.521$

$D = 0.32 < D_{0.05} = 0.521\ ;\ H_0\ is\ not\ rejected\ /\ H_0\ is\ accepted$

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Problem:

The sequence of numbers **0.32, 0.51, 0.10, 0.87, 0.61, 0.29** has been generated. Use KS-test with $\alpha = 0.05$ to determine if the hypothesis that the numbers are uniformly distributed on the interval [0,1] can be rejected. [$D_{0.05} = 0.521$]

Solution:

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Arranging numbers in increasing order,

$$0.10 \leq 0.29 \leq 0.32 \leq 0.51 \leq 0.611 \leq 0.87$$

Now, ley's compute D^+ & D^-

i	1	2	3	4	5	6
R_i	0.10	0.29	0.32	0.51	0.61	0.87
i/N	0.17	0.33	0.50	0.67	0.83	1.00
$i-1/N$	0	0.17	0.33	0.50	0.67	0.83
$i/N - R_i$	0.07	0.04	0.18	0.16	0.22	0.13
$R_i - i-1/N$	0.10	0.12	-	0.01	-	0.04

$D^+ = \max(0.07, 0.04, 0.18, 0.16, .0.22, 0.13) = 0.22$

$D^- = \max(0.10, 0.12, 0.01, 0.01, 0.06, 0.04) = 0.12$

$D = max(D^+, D^-) = \max(0.12, 0.22) = 0.22$

$D_\alpha = D_{0.05} = 0.521$

$D = 0.22 < D_{0.05} = 0.521$; H_0 *is not rejected* /H_0 *is accepted*

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Chi-Square Test

If there is two different sample S_1 *and* S_2, then we apply Chi-Square test to determine that some points of S_1 is related to S_2 and some point of S_2 related to S_1. It is a comparison that there is any dependency among samples.

Algorithm:

Step-1: Define the hypothesis for uniformity

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Step-2: Divide total no of observation, N into mutually exclusive equal numbered classes n,

$$E_i \geq 5 \;;\; E_i = Expected$$

Step-3: Test statistics,

$$\chi_0^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i}\right] \;;\; O_i = Observation\ from\ Statistics$$

Step-4: Determine critical value for given level of significant with $(n-1)$

Step-5:

if, $\chi_0^2 > \chi_{\alpha,\ n-1}^2 => H_0\ is\ rejected$

else,
$no\ difference\ between\ detected\ sample\ distribution\ and\ uniform\ distribution$

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Problem:

For the following samples, apply Chi-Square Test,

0.43, 0.09, 0.52, 0.98, 0.78, 0.44, 0.21, 0.12, 0.38, 0.67, 0.97, 0.46, 0.07, 0.18, 0.49, 0.47, 0.69, 0.99, 0.77 ,0.76, 0.65, 0.14, 0.25, 0.37, 0.74, 0.03, 0.71, 0.28, 0.39, 0.56, 0.73, 0.99, 0.71, 0.99, 0.64, 0.5, 0.66, 0.01, 0.24, 0.73, 0.15, 0.45, 0.10, 0.18, 0.82, 0.96, 0.43, 0.27, 0.34, 0.65, 0.79, 0.03, 0.49 , 0.69 ,0.85 ,0.6, 0.93, 0.48, 0.42, 0.04, 0.46, 0.04, 0.91, 0.81 ,0.62, 0.79, 0.88, 0.46, 0.74, 0.06, 0.11, 0.64, 0.76, 0.22, 0.47, 0.94, 0.37,0.5, 0.97, 0.26, 0.92, 0.87, 0.88, 0.27, 0.12, 0.10, 0.29, 0.65, 0.13, 0.4, 0.8, 0.82, 0.25, 0.78, 0.99, 0.36, 0.24, 0.18, 0.2, 0.1

$N = 100, n = 10 \; and \; \alpha = 0.05, \; \chi^2_{0.95, \;\; 9} = \; 16.9$

Solution:

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim \; U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Now let n, such that $E_i \geq 5$

$$=> \frac{N}{n} \geq 5 \;\; [E_i = \frac{N}{n}]$$

$$=> \frac{100}{n} \geq 5$$

$$=> \frac{100}{5} \geq n$$

$$\therefore \; n \leq 20$$

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Test Statistics

n	$Interval$	O_i	$E_i = \frac{N}{n}$	$\frac{(O_i - E_i)^2}{E_i}$
1	$(0.0 - 0.1)$	8	10	0.4
2	$(0.1 - 0.2)$	12		0.4
3	$(0.2 - 0.3)$	12		0.4
4	$(0.3 - 0.4)$	6		1.6
5	$(0.4 - 0.5)$	14		1.6
6	$(0.5 - 0.6)$	4		3.6
7	$(0.6 - 0.7)$	11		0.1
8	$(0.7 - 0.8)$	13		0.9
9	$(0.8 - 0.9)$	8		0.4
10	$(0.9 - 1.0)$	12		0.4
$\chi_0^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$				9.8

$\alpha = 0.05\ and\ n = 10$ So, $\chi^2_{\alpha,9} = 16.9$

$$9.8 < 16.9$$

$$\therefore \chi_0^2 < \chi^2_{\alpha.9}$$

$$So, H_0\ \ is\ not\ rejected\ Or, \qquad H_0\ is\ accepted$$

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Problem:

For the following samples, apply Chi-Square Test

0.77 ,0.76, 0.65, 0.14, 0.25, 0.37, 0.74, 0.03, 0.71, 0.28, 0.39, 0.56, 0.73, 0.99, 0.71, 0.99, 0.64, 0.5, 0.66, 0.01

$N = 20, n = 4 \ and \ \alpha = 0.95$, $X^2_{\alpha,3} = 7.82$

Solution:

$$H_0 : R_i \sim U[0,1]$$

$$H_1 : R_i \nsim U[0,1]$$

Hypothesis H_0 indicates that R_i is uniformly distributed between [0,1] and hypothesis H_1 indicates that R_i isn't uniformly distributed between [0,1]

Now let n, such that $E_i \geq 5$

$$=> \frac{N}{n} \geq 5$$

$$=> \frac{20}{n} \geq 5$$

$$=> \frac{20}{5} \geq n$$

$$\therefore \ n \leq 4$$

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University

Test Statistics

	$Interval$	O_i	$E_i = \frac{N}{n}$	$\frac{(O_i - E_i)^2}{E_i}$
1	$(0.0 - 0.25)$	3	5	0.8
2	$(0.25 - 0.50)$	4		0.2
3	$(0.50 - 0.75)$	9		3.2
4	$(0.75 - 1.00)$	4		0.2
$\chi_0^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right]$				4.4

$\alpha = 0.05 \; and \; n = 4$ So, $\chi^2_{\alpha,3} = 7.82$

$$4.4 < 7.82$$

$$\therefore \chi_0^2 < \chi^2_{\alpha,3}$$

So, H_0 is Accepted

Created by
Shahadat Hossain
Lecturer, Dept of CSE
City University