x_i	Observed Frequency O_i	$x_i \cdot O_i$	Expected Frequency E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	12 } 22	00	2.6 } 12.2	7.87
1	10	10	9.6	
2	19	38	17.4	0.15
3	17	51	21.1	0.80
4	10	40	19.2	4.41
5	8	40	14.0	2.57
6	7	42	8.5	0.26
7	5	35	4.4	
8	5	40	2.0	
9	3 } 17	27	0.8 } 7.6	11.62
10	3	30	0.3	
11	1	11	0.1	
Total	100	364		27.68

$\therefore \bar{x} = \frac{364}{100} = 3.64$

$\Rightarrow \lambda = 3.64.$

Now, compute the probabilities associated with various values of x using the pmf with a computed λ. i.e.

$$p(0) = \frac{e^{-3.64} \cdot (3.64)^0}{0!} = 0.026$$

$$p(1) = \frac{e^{-3.64} \cdot (3.64)^1}{1!} = 0.096$$

$$p(2) = \frac{e^{-3.64} (3.64)^2}{2!} = 0.174 \quad \text{and so on.}$$

The expected frequecies are -

$E_i = n \cdot P(i)$

ie, $E_0 = 100 \cdot (P(0)) = 2.6$

$E_1 = 100 \cdot P(1) = 9.6$ and so on.

wherever expected frequencies are less than 5, combi with adjacent value.

Hence, 2.6 and 9.6 are combined.

Similarly, last five values (E_i) are combined.

The respective observed frequencies also have to be combined.

Now compute $\frac{(O_i - E_i)^2}{E_i}$ for each value of i.

Thus, $\chi_0^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$= 27.68$

The degrees of freedom $= k - s - 1$.

Due to the values being combined, the number of values counted for calculation happen to be 7.

Hence, $k = 7$.

And Poisson distribution has only one parameter λ

Hence, $s = 1$.

$\therefore$ degrees of freedom $= 7 - 1 - 1 = 5$.

Now from χ^2 table, it is found that $\chi^2_{0.05,5} = 11\cdot$

Since $\chi_0^2 > \chi^2_{0.05,5}$, the H_0 is rejected.

ie, given values are not following Poisson distribut

Example 2: The vehicles arriving at a petrol bunk in a 5 minute period between 6 am to 10 pm was monitored for 80 days given below.

Arrival per period	0	1	2	3	4	5	6	7	8	9	10
Frequency	15	12	8	10	10	7	4	3	5	4	2

Use χ^2 test to check whether the data follows Poisson distribution at 5% level of significance.

Solution: To check whether the given data follows Poisson distribution, the hypotheses are formed as –

H_0 : The random variable is Poisson distributed.

H_1 : – – – – not – – –

The pmf for poisson distribution is –

$$p(x) = \frac{e^{-\lambda}.\ \lambda^x}{x!}, \quad x = 0, 1, 2 \ldots$$

The mean, λ has to be calculated, as given in the previous example. (Refer the table in next page).

$$\lambda = \frac{274}{80} = 3.425$$

Now, compute the probabilities associated with various values of x using pmf and λ.

$$p(0) = \frac{e^{-3.425}.(3.425)^0}{0!} = 0.0325 \qquad E_0 = 80 \times 0.032 = 2.6$$

$$p(1) = \frac{e^{-3.425}.(3.425)^1}{1!} = 0.1113 \qquad E_1 = 80 \times 0.1113 = 8.904$$

$$p(2) = \frac{e^{-3.425}(3.425)^2}{2!} = 0.3812 \qquad E_2 = 80 \times 0.1906 = 15.248$$

Similarly, compute till $p(10)$ and E_{10}.

x_i	Observed Frequency O_i	$x_i O_i$	Expected frequency E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	15 } 27	00	2.6 } 11.504	20.8733
1	12	12	8.904	
2	8	16	15.248	3.4453
3	10	30	17.408	3.1525
4	10	40	14.904	1.6136
5	7	35	10.208	1.0082
6	4 } 18	24	5.832 } 10.528	5.3031
7	3	21	2.856	
8	5	40	1.224	
9	4	36	0.464	
10	2	20	0.152	
Total	80	274		35.396

Now, $\chi_0^2 = 35.396$ from our calculation.

The degrees of freedom $= 6 - 1 - 1$
$= 4$

From the chi-square table, it is found that,

$\chi^2_{0.05,4} = 9.49$

Since $\chi_0^2 > \chi^2_{0.05,4}$, the H_0 is rejected.

That is, the given data is not following Poisson distribution.

Chi-Square Test for Exponential Distribution

In previous section, chi-square test had been applied on Poisson distribution. It is a discrete distribution. When the chi-square test has to be applied on continuous distribution, we need to follow a different procedure. It is explained below with the help of exponential distribution —

In case of continuous distributions, class intervals to be formed in such a way that the probability of all the class intervals must be equal. (Note that, in case of discrete distributions, class interval were of equal width). This can be done by taking $p_i = \frac{1}{k}$, where k is number of class intervals and p_i is the probability of each class interval.

It is known that, the expected frequency,

$$E_i = n.p_i \geq 5$$

$$\Rightarrow n.\frac{1}{k} \geq 5$$

$$\Rightarrow k \leq \frac{n}{5}$$, i.e. width of the class interval should be $\leq \frac{n}{5}$.

Now, we know that pdf of exponential distribution is —

$$f(x) = \lambda . e^{-\lambda x}$$

and the cdf is —

$$F(x) = 1 - e^{-\lambda x}$$

The end-points of each class intervals should be computed using cdf as —

$F(a_i) = 1 - e^{-\lambda a_i}$, where a_i is the end-point of i^{th} interval.

Since, $F(a_i)$ is the cumulative area from 0 to a_i,

$F(a_i) = i.p.$

Thus, $ip = 1 - e^{-\lambda a_i}$, $\quad i = 0, 1, \dots k$

$\Rightarrow \quad e^{-\lambda a_i} = 1 - ip$

$\Rightarrow \quad -\lambda a_i = \log_e (1 - ip)$

$\Rightarrow \quad a_i = \frac{-1}{\lambda} \ln(1 - ip)$ (Remember, $\ln$ is representation for $\log_e$)

Thus, with the help of computed k and a_i, we will formulate class intervals. Then the given data are distributed into these intervals for applying chi-square test.

Example1:

Life tests were performed on a random sample of electronic components at 1.5 times the nominal volatage. Their lifetime (or time to failure), in days, are given below. Apply chi-square test to verify whether the given data follows exponential distribution.

79.919	3.081	0.062	1.961	5.845
3.027	6.505	0.021	0.013	0.123
6.769	59.899	1.192	34.760	5.009
18.387	0.141	43.565	24.420	0.433
144.695	2.663	17.967	0.091	9.003
0.941	0.878	3.371	2.157	7.579
0.624	5.380	3.148	7.078	23.960
0.590	1.928	0.300	0.002	0.543
7.004	31.764	1.005	1.147	0.219
3.217	14.382	1.008	2.336	4.562

Solution: Let X be a random variable representing life time of electronic components in days. Now the hypotheses are set as-

H_0 : X is exponentially distributed.

H_1 : X is not " " .

Now, to compute class intervals, we should know the value of parameter λ, of exponential distribution. We know that, for exponential distribution, the mean,

$$\bar{X} = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{\bar{X}}$$

But, we know that $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

Hence, compute or average, $\bar{X}$. for a given numbers.

$$\bar{X} = \frac{594.674}{50}$$

$$= 11.8935$$

$$\therefore \lambda = \frac{1}{11.8935} = 0.0841$$

Now, the number of class intervals, k, should be -

$$k \le \frac{n}{5}$$

$$\Rightarrow k \le \frac{50}{5}$$

$$\Rightarrow k \le 10.$$

Let us take $k = 8$. Then $p = \frac{1}{k} = 0.125$.

Now, the end-points of each class interval a_i should be computed. Note that, always $a_0 = 0$ and $a_k = \infty$

so, $a_1, a_2, \dots a_7$ have to be computed using the formula -

$$a_i = -\frac{1}{\lambda} \ln(1 - ip), \quad i = 0, 1 \dots k$$

$$a_1 = \frac{-1}{0.0841} \cdot \ln(1 - 1(0.125)) = 1.5878$$

$$a_2 = \frac{-1}{0.0841} \cdot \ln(1 - 2.(0.125)) = 3.4207$$

$$a_3 = \frac{-1}{0.0841} \cdot \ln(1 - 3.(0.125)) = 5.5886$$

similarly,

$a_4 = 8.2419$, $\quad a_5 = 11.6627$

$a_6 = 16.4839$ $\quad a_7 = 24.7258$

Now taking these end-points of class intervals, create the class-intervals and the actual (observed) number of elements belonging to each of these class intervals as shown in the following table -

Class Interval	Observed Frequency O_i	Expected Frequency E_i	$\frac{(O_i - E_i)^2}{E_i}$
[0, 1.5878)	19	6.25	26.01
[1.5878, 3.4207)	10	6.25	2.25
[3.4207, 5.5886)	3	6.25	1.69
[5.5886, 8.2419)	6	6.25	0.01
[8.2419, 11.6627)	1	6.25	4.41
[11.6627, 16.4839)	1	6.25	4.41
[16.4839, 24.7258)	4	6.25	0.81
[24.7258, ∞)	6	6.25	0.01
Total	50	50.00	39.60

The expected frequencies are computed as –

$E_i = n \cdot p_i$

$= 50 \cdot (0.125) = 6.25 \quad \forall i$

From the table,

$\chi_0^2 = 39.60$

The exponential distribution has only one parameter, λ. Hence, the degrees of freedom $= k - s - 1$

$= 8 - 1 - 1 = 6.$

Now, for level of significance $\alpha = 0.05$, the χ^2-table value is –

$\chi^2_{0.05, 6} = 12.6.$

As, $\chi_0^2 > \chi^2_{0.05, 6}$, the H_0 is rejected. That is the given data are not exponentially distributed

Example 2: The time (in minutes) required for 30 different employees to compute and record the number of hours worked during a week are given:

1.88	2.62	1.49	0.35	0.82	2.03
0.54	0.21	0.39	0.90	2.03	2.16
1.90	0.63	0.17	0.03	0.45	0.36
0.15	2.03	4.29	0.04	1.73	0.92
2.81	0.05	5.5	2.16	0.48	0.18

Use χ^2-test to test the hypothesis that these service times are exponentially distributed at 5% level of significance. Let the number of intervals be $k = 6$ and critical value $= 9.49$.

Solution: Let X be a random variable representing service time of employes.

H_0: X follows exponential distribution.

H_1: X do not follow ''

Now, $\bar{x} = \frac{\sum x_i}{n}$

$= \frac{39.3}{30} = 1.31$

$\therefore \lambda = \frac{1}{\bar{x}} = \frac{1}{1.31} = 0.7634$

In the question, it is given that,

$k = 6$

$\therefore p = \frac{1}{6} = 0.1667$

The expected frequencies, $E_i = n \cdot p_i$

$= 30\left(\frac{1}{6}\right)$

$= 5$ for all i

Now, compute end-points of class intervals –

$a_i = -\frac{1}{\lambda} \ln(1 - ip)$

$a_1 = \frac{-1}{0.7634} \ln\left(1 - 1 \cdot \frac{1}{6}\right) = 0.2388$

$a_2 = \frac{-1}{0.7634} \ln\left(1 - 2 \cdot \frac{1}{6}\right) = 0.5311$

$a_3 = 0.9080$ $\quad a_4 = 1.4391$

$a_5 = 2.3471$

Using these endpoints, create class intervals and the respective frequencies as in the table below –

Class Intervals	Observed Frequency O_i	Expected Frequency E_i	$\frac{(O_i - E_i)^2}{E_i}$
[0, 0.2388)	7	5	0.8
[0.2388, 0.5311)	5	5	0.0
[0.5311, 0.9080)	4	5	0.2
[0.9080, 1.4391)	1	5	3.2
[1.4391, 2.3471)	9	5	3.2
[2.3471, ∞)	4	5	0.2
	30	30	7.6

$\chi_0^2 = 7.6$

Degrees of freedom $= k - s - 1 = 6 - 1 - 1 = 4$.

$\chi^2_{0.05,4} = 9.49$

As, $\chi_0^2 < \chi^2_{0.05,4}$, the H_0 is accepted. That is, the data is exponentially distributed.

Kolmogorov – Smirov Goodness-of-fit Test

The chi-square test discussed in the previ
section makes use of class intervals. If the
of the class-interval is changed, the frequen
of that class-interval changes. This will c
the value of χ_0^2. Hence, based of the wi
of class intervals, acceptance/rejection of H0 m
happen. To avoid this problem, K-S test
used, for checking exponential distribution.

If any data is exponentially distributed
the range [0, T] with values {T1, T2, ---
then they have to be cumulatively added as
T1, T1+T2, T1+T2+T3, T1+T2+T3+T4, ----
Then these values are normalized for the ran
[0, 1] by dividing each value by T. No
the new set of values

$$\frac{T1}{T}, \frac{T1+T2}{T}, \frac{T1+T2+T3}{T}, \ldots$$

follows uniform distribution. On these num
K-S test has to be performed as earli
unit 3).