

kstest2

Two-sample Kolmogorov–Smirnov test

Syntax

```
H = kstest2(X1,X2)
H = kstest2(X1,X2,alpha,tail)
[H,P,KSSTAT] = kstest2(X1,X2,cdf,alpha,tail)
```

Description

`H = kstest2(X1,X2)` performs a two-sample Kolmogorov–Smirnov test to compare the distributions of values in the two data vectors `x1` and `x2` of length `n1` and `n2`, respectively, representing random samples from some underlying distribution(s). The null hypothesis for this test is that `x1` and `x2` are drawn from the same continuous distribution. The alternative hypothesis is that they are drawn from different continuous distributions. The result `H` is 1 if you can reject the hypothesis that the distributions are the same, or 0 if you cannot reject that hypothesis. You reject the hypothesis if the test is significant at the 5% level.

For each potential value `x`, the Kolmogorov–Smirnov test compares the proportion of `x1` values less than `x` with proportion of `x2` values less than `x`. The `kstest2` function uses the maximum difference over all `x` values as its test statistic. Mathematically, this can be written as

$$\max(|F1(x) - F2(x)|)$$

where $F1(x)$ is the proportion of `x1` values less than or equal to `x` and $F2(x)$ is the proportion of `x2` values less than or equal to `x`. Missing observations, indicated by NaNs are ignored.

`H = kstest2(X1,X2,alpha)` performs the test at the $(100*\alpha)\%$ significance level.

The decision to reject the null hypothesis occurs when the significance level, `alpha`, equals or exceeds the P-value.

`H = kstest2(X1,X2,alpha,tail)` accepts a string `tail` that specifies the type of test. `tail` can have one of the following values:

- 'unequal'
- 'larger'
- 'smaller'

The tests specified by these values are described in [Tests Specified by tail](#)

`[H,P] = kstest2(...)` also returns the asymptotic p-value P . The asymptotic p-value becomes very accurate for large sample sizes, and is believed to be reasonably accurate for sample sizes n_1 and n_2 such that $(n_1*n_2)/(n_1 + n_2) \geq 4$.

`[H,P,ksstat] = KSTEST2(...)` also returns the Kolmogorov–Smirnov test statistic $KSSTAT$ defined above for the test type indicated by *tail*.

Tests Specified by tail

Let $S_1(x)$ and $S_2(x)$ be the empirical distribution functions from the sample vectors x_1 and x_2 , respectively, and $F_1(x)$ and $F_2(x)$ be the corresponding true (but unknown) population CDFs. The two-sample Kolmogorov–Smirnov test tests the null hypothesis that $F_1(x) = F_2(x)$, for all x , against the alternative hypothesis specified by *tail*, as described in the following table.

tail	Alternative Hypothesis	Test Statistic
'unequal'	$F_1(x)$ does not equal $F_2(x)$ (two-sided test)	$\max S_1(x) - S_2(x) $
'larger'	$F_1(x) > F_2(x)$ (one-sided test)	$\max[S_1(x) - S_2(x)]$
'smaller'	$F(x) < F_2(x)$ (one-sided test)	$\max[S_2(x) - S_1(x)]$

Examples

The following commands compare the distributions of a small evenly-spaced sample and a larger normal sample:

```
x = -1:1:5
y = randn(20,1);
[h,p,k] = kstest2(x,y)

h =
    1
```

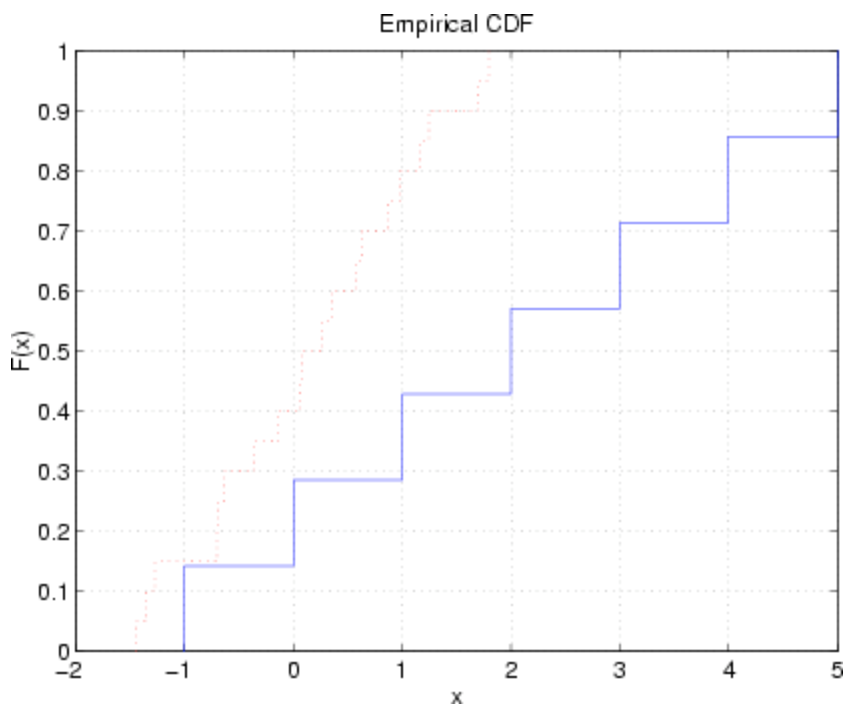
```
p =  
    0.0403  
k =  
    0.5714
```

The difference between their distributions is significant at the 5% level ($p = 4\%$). To visualize the difference, you can overlay plots of the two empirical cumulative distribution functions. The Kolmogorov–Smirnov statistic is the maximum difference between these functions. After changing the color and line style of one of the two curves, you can see that the maximum difference appears to be near $x = 1.9$. You can also verify that the difference equals the k value that `kstest2` reports:

```
cdfplot(x)  
hold on  
cdfplot(y)  
h = findobj(gca,'type','line');  
set(h(1),'linestyle',':','color','r')
```

```
1 - 3/7
```

```
ans =  
    0.5714
```



See Also

[kstest](#), [lillietest](#)

◀ kstest

kurtosis ▶

© 1984–2006 The MathWorks, Inc. · [Terms of Use](#) · [Patents](#) · [Trademarks](#) · [Acknowledgments](#)