

**Kolmogorov–Smirnov distribution** (from <http://www.math.wm.edu/~leemis/chart/UDR/UDR.html>)

The symbol used for the Kolmogorov–Smirnov test statistic for a sample size  $n$  is typically  $D_n$ . Using a result from Birnbaum, Z.W. (1952), “Numerical Tabulation of the Distribution of Kolmogorov’s Statistic for Finite Sample Size,” *Journal of the American Statistical Association*, 47, 425–441, a Kolmogorov–Smirnov random variable  $D_n$  with parameter  $n$  has a cumulative distribution function of  $D_n - 1/(2n)$  of

$$P\left(D_n < \frac{1}{2n} + v\right) = n! \int_{\frac{1}{2n}-v}^{\frac{1}{2n}+v} \int_{\frac{3}{2n}-v}^{\frac{3}{2n}+v} \cdots \int_{\frac{2n-1}{2n}-v}^{\frac{2n-1}{2n}+v} g(u_1, u_2, \dots, u_n) du_n \dots du_2 du_1 \quad 0 \leq v \leq \frac{2n-1}{2n}$$

for all positive integer values of  $n$  with  $g(u_1, u_2, \dots, u_n) = 1$  over  $0 \leq u_1 \leq u_2 \leq \dots \leq u_n \leq 1$ , and 0 otherwise. Computing this integral is non-trivial, particularly for larger values of  $n$ .

The cumulative distribution function of  $D_1$  for  $n = 1$  is

$$F_{D_1}(t) = P(D_1 \leq t) = \begin{cases} 0 & t \leq \frac{1}{2} \\ 2t - 1 & \frac{1}{2} < t < 1 \\ 1 & t \geq 1. \end{cases}$$

The cumulative distribution function of  $D_2$  for  $n = 2$  is

$$F_{D_2}(t) = P(D_2 \leq t) = \begin{cases} 0 & t \leq \frac{1}{4} \\ 8\left(t - \frac{1}{4}\right)^2 & \frac{1}{4} < t < \frac{1}{2} \\ 1 - 2(1 - t)^2 & \frac{1}{2} < t < 1 \\ 1 & t \geq 1. \end{cases}$$

The general cumulative distribution function is mathematically intractable, but an algorithm to calculate it for specific values of  $n$  is given in Drew, Glen, and Leemis (2000), “Computing the Cumulative Distribution Function of the Kolmogorov–Smirnov Statistic,” *Computational Statistics and Data Analysis*, Volume 34, Number 1, July 2000, 1–15. Applying this algorithm when  $n = 6$  gives the cumulative distribution function of  $D_6$  as

$$F_{D_6}(t) = \begin{cases} 0 & t < \frac{1}{12} \\ 46080t^6 - 23040t^5 + 4800t^4 - \frac{1600}{3}t^3 + \frac{100}{3}t^2 - \frac{10}{9}t + \frac{5}{324} & \frac{1}{12} \leq t < \frac{1}{6} \\ 2880t^6 - 4800t^5 + 2360t^4 - \frac{1280}{3}t^3 + \frac{235}{9}t^2 + \frac{10}{27}t - \frac{5}{81} & \frac{1}{6} \leq t < \frac{1}{4} \\ 320t^6 + 320t^5 - \frac{2600}{3}t^4 + \frac{4240}{9}t^3 - \frac{785}{9}t^2 + \frac{145}{27}t - \frac{35}{1296} & \frac{1}{4} \leq t < \frac{1}{3} \\ -280t^6 + 560t^5 - \frac{1115}{3}t^4 + \frac{515}{9}t^3 + \frac{1525}{54}t^2 - \frac{565}{81}t + \frac{5}{16} & \frac{1}{3} \leq t < \frac{5}{12} \\ 104t^6 - 240t^5 + 295t^4 - \frac{1985}{9}t^3 + \frac{775}{9}t^2 - \frac{7645}{648}t + \frac{5}{16} & \frac{5}{12} \leq t < \frac{1}{2} \\ -20t^6 + 32t^5 - \frac{185}{9}t^3 + \frac{175}{36}t^2 + \frac{3371}{648}t - 1 & \frac{1}{2} \leq t < \frac{2}{3} \\ 10t^6 - 38t^5 + \frac{160}{3}t^4 - \frac{265}{9}t^3 - \frac{115}{108}t^2 + \frac{4651}{648}t - 1 & \frac{2}{3} \leq t < \frac{5}{6} \\ -2t^6 + 12t^5 - 30t^4 + 40t^3 - 30t^2 + 12t - 1 & \frac{5}{6} \leq t < 1 \\ 1 & t \geq 1. \end{cases}$$

The survivor, hazard, cumulative hazard, inverse distribution, moment generating, and characteristic functions on the support of  $X$  are also mathematically intractable.

The population mean, variance, skewness, and kurtosis of  $X$  are mathematically intractable. Using the APPL function  $\text{KSRV}(n)$ , however, one can calculate the moments of  $D_n$  for any value of  $n$ . For example, the population means of  $D_n$  for  $n = 1, 2, \dots, 6$  are given in the table below.

$n$	1	2	3	4	5	6
$E[D_n]$	3/4	13/24	293/648	813/2048	134377/375000	1290643/3919104