2022 ■ Vol. 18 ■ no. 3

Computation of cumulative density of noncentral chi-square distribution in Excel

Prasanth Sambaraju [a]

[a] Independent Researcher

Abstract ■ Noncentral distributions are obtained by transformation of their respective central distributions, and are identified by a noncentrality parameter. The noncentrality parameter measures the degree to which mean of test statistics departs, when the null hypothesis is false. Central distributions are used to describe test statistics, when the null hypothesis is true. Noncentral distributions are used to calculate statistical power of a test in situations when the null hypothesis is false. The paper presents Visual Basic for Application code in Microsoft Excel to compute the cumulative distribution function for noncentral chi-square distributions. The results obtained were found to be comparable with the reported values.

Keywords ■ Power of chi-square test, cumulative distribution function, noncentral. **Tools** ■ VBA.

prasanth.sambaraju@gmail.com

10.20982/tqmp.18.3.p255

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Introduction

This article present a code in Excel Visual Basic for application (VBA) to compute the cumulative distribution function (cdf) for noncentral chi-square distributions. Noncentral chi-square distributions are used for calculating the power of test involving chi-square test statistics (Benton & Krishnamoorthy, 2003). The cdf of noncentral chi-square distribution is computed using the following equation (Narula & Desu, 1981), where v is the degrees of freedom, λ is the noncentrality parameter and $F(y \mid v+2i,0)$ is the cdf of a random variable that follows a central chi-squared distribution.

$$\mathrm{F}(\mathrm{y}|\mathrm{v},\lambda)=\sum_{i=0}^{\infty} e^{-\frac{\lambda}{2}} \frac{\left(\frac{\lambda}{2}\right)^{i}}{i!} \mathrm{F}\left(\mathrm{y} \mid \mathrm{v}+2\mathrm{i}, 0\right) \quad (1)$$

Method

When the cdf is computed in equation (1) using recursive computation starting with $i = 0$, would result in an "overflow" error in Excel, as the value of i exceeds 170. This is due to limitations of Excel double data type. So to overcome this limitation, logarithmic transformation was proposed by Nelson (2016). This transformation helps to overcome the "overflow" error and improve the accuracy of cdf estimation. The following terms in Equation 1 were transformed

(a) Exponential term $(e - \lambda/2) = (\lambda/2) \times -0.434294481903252$;
(b) Power term $(\lambda/2) = \log_{10}(\lambda/2)$;
(c) Factorial term $(i!) = \log_{10}(i)$.

After computing the summation term till convergence in Equation (1), the log value is converted by using the power function in Excel. The above logarithmic transformation works when $i > 0$, so when $i = 0$, the summation term is calculated without transformation.

Results

The accuracy of the code was assessed by comparing the results with the reported values and are shown in Table 1. As described by McCullough and Wilson (1999), log relative error (LRE) was used to measure the accuracy, which was calculated using equation (2). In the calculation, x is the estimated value whereas c is the reported value. The macro enabled excel workbooks contains the source code for the macro and user defined function (UDF). The macro can be executed by clicking on the shape, whereas the UDF can be called by writing the function name into a cell. The logarithmic transformation approach to calculate cdf can be extended to other noncentral distributions.

$$\mathrm{LRE}=-\log_{10}\left(\frac{|\mathrm{x}-\mathrm{c}|}{|\mathrm{c}|}\right) \quad (2)$$

Conclusions

When the above code was checked for large values of x, v and λ (de Oliveira & Ferreira, 2012), it gave either a value of 0 or `#VALUE!`. So a VBA equivalent of R code from the package `ncg` was written and tested. When this code was checked for these large values, the results were comparable with the reported values and are shown in Table 2.

References

Benton, D., & Krishnamoorthy, K. (2003). Computing discrete mixtures of continuous distributions: Noncentral chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient. *Computational Statistics & Data Analysis*, *43*(2), 249–267. doi:10.1016/s0167-9473(02)00283-9

de Oliveira, I. R. C., & Ferreira, D. F. (2012). Computing the noncentral gamma distribution, its inverse and the noncentrality parameter. *Computational Statistics*, *28*(4), 1663–1680. doi:10.1007/s00180-012-0371-8

McCullough, B. D., & Wilson, B. (1999). On the accuracy of statistical procedures in microsoft excel 97. *Computational Statistics & Data Analysis*, *31*(1), 27–37. doi:10.1016/s0167-9473(99)00004-3

Narula, S. C., & Desu, M. M. (1981). Algorithm as 170: Computation of probability and non-centrality parameter of a non-central chi-squared distribution. journal of the royal statistical society. *Series C (Applied Statistics)*, *30*(3), 349–352. doi:10.2307/2346368

Nelson, J. B. (2016). A robust function to return the cumulative density of non-central f distributions in microsoft office excel. *Psicológica*, *37*(1), 61–83.

Open practices

The *Open Material* badge was earned because supplementary material(s) are available on the journal's web site.

Citation

Sambaraju, P. (2022). Computation of cumulative density of noncentral Chi-square distribution in Excel. *The Quantitative Methods for Psychology*, *18*(3), 255–257. doi:10.20982/tqmp.18.3.p255

Received: 08/06/2022 ∼ Accepted: 13/08/2022

Tables 1 and 2 follow.

Table 1 ■ Comparison of reported and calculated cdf values

x	v	λ	reported values	calculated by the Excel VBA macro	LRE
0.00393	1	6	0.2498463724258039d-2	0.002498463724258	14.53
9.23636	5	1	0.8272918751175548d0	0.827291875117555	15.03
24.72497	11	21	0.2539481822183126d0	0.253948182218312	15.36
44.98534	31	6	0.8125198785064969d0	0.812519878506496	15.86
38.56038	51	1	0.8519497361859118d-1	0.085194973618591	14.89
82.35814	100	16	0.1184348822747824d-1	0.011843488227478	15.06
331.7885	300	16	0.7355956710306709d0	0.735595671030669	14.87
459.9261	500	21	0.2797023600800060d-1	0.027970236008001	14.58
0.00016	1	1	0.6121428929881423d-2	0.006121428929881	15.25
0.00393	1	1	0.303381422975380d-1	0.030338142297538	14.20
0.00393	1	6	0.2498463724258039d-2	0.002498463724258	14.53
9.23636	5	1	0.8272918751175548d0	0.827291875117555	15.03

Table 2 ■ Comparison of reported and calculated cdf values for large values of x, v and λ

x	v	λ	reported values	calculated by the Excel VBA macro	LRE
9999000	1000	10000000	0.375969658204496	0.375971820403499	5.24
20035000	30000	20000000	0.711879569275030	0.711879665653162	6.87
16050100000	1.50E+10	1050000000	0.705656960203322	0.705656917051403	7.21
5000002661439	5.00E+12	2.70E-20	0.799999926426758	0.799999925736534	9.06
0.00000001	0.05	0.00000002	0.628811844415446	0.628811844415446	15*
1.00E-40	0.005	1.80E-10	0.794093971798262	0.794093971798251	13.86

Note. * LRE equals to the number of significant digits in the reported value, when both values are same.