XVI-2. RESULTS OF NEW TESTS ON STATISTICAL DISTRIBUTIONS, CONTINUOUS FUNCTIONS

2. CONTINUOUS STATISTICAL DISTRIBUTIONS

2.1 COMPONENT FUNCTIONS

2.1.1 LOG GAMMA DISTRIBUTION: GAMMALN FUNCTION

The function is =GAMMALN(x)

This function provides values of the natural log of the gamma function. The function has never changed from Excel 97.

The gamma function is a restatement of factorials, and allows a definition of factorials for non-integer real numbers. The gamma function is defined for both positive and negative arguments (see Abramowitz (1963), page 255). The log of the gamma function using Stirling's Formula (Abramowitz (1963) equations 6.1.40 and 6.1.4) or the power series (eq. 6.1.33) cannot be used to obtain log gamma values for negative arguments. When the gamma value for negative arguments is positive (see Abramowitz (1963), figure 6.1), the log gamma function would exist. In this case it is best to use the gamma function recurrence equations (Abramowitz (1963) equations 6.1.5 or 6.1.6) to obtain existent values for negative arguments.

Smith (2002) did not provide a public log gamma function. His private functions and subs related to the log gamma function are not sufficiently clear so as to build a correct full range accurate log gamma function. According to Lewis, the problem was the fact that subtractions of the terms in the accepted equations (see Abramowitz) lead to errors that were higher than his standard for accuracy. The obvious solution is to build a xnumbers function that would retain 15 digit accuracy within the allowable IEEE-754 double range.

The reference function used to test GAMMALN was a basic spreadsheet multiplication of integers up and division of integers down from 100!. 100! is a reference point given in Abramowitz (1963) in table 6.6 to 20 decimal digits. The multiplications and divisions used existent Excel double precision arithmetic. The integer division down from 100 gave integer factorial values from 19 down to 1, indicating that division errors were not significant. It can then be assumed that the reference sequence was a correct and valid list of factorials. Using the Excel natural log function on the calculated factorials gave a series of log-gamma reference points. Comparing the GAMMALN function with these reference points gave LRE values from 9.76 (at 3!) to 13.03 (at 170!). The basic, untested assumption is that the natural log function (in Excel) is accurate.

One of the problems is that the 6.1.40 and 6.1.41 equations in Abramowitz (1963) are asymptotic formulas, and will give incorrect values for small arguments. Smith used equation 6.1.33 and others to get around this problem, but never could achieve his level of accuracy standard. An alternate in which the asymptotic formulas are used only for large arguments (i.e. >7), and the recurrence relationships used to go below 7, suffers from a loss in accuracy due to the accumulation of error from the natural log function. A NewLogGamma function in VBA was written, following this, and other improvements in the summation process. For arguments from 3 to 37, including fractional values in-between resulted in LRE values from 9.73 to 11.74, essentially the same as that from the above integer method.

A comparison of the Excel GAMMALN output with the tabulated values in table 6.1 of Abramowitz (1963) resulted in only differences of 1 in the last digit. The table gives values to 10 decimal values for arguments from 1 to 2 in steps of 0.005.

It can be assumed that Excel uses the same basic equations (6.1.33, 6.1.40 or 6.1.41), which are continuous in argument space, since there is no other known accurate alternates for non-integer arguments. Since the basic equations are smooth and continuous between integers and GAMMALN returns smooth and accurate values between 1 and 2, intermediate values over the entire range can be assumed to be as accurate as it is at the integers. Therefore the accuracy of GAMMALN is in excess of 9.76 over the entire range of n, with increasing accuracy as n increases.

A user should be aware that calculations involving differences of GAMMALN values would be much less accurate than the GAMMALN values themselves. This is a distinct limit on the accuracy of calculations involving combinations and permutations.

RELIABILITY ASSESSMENT OF GAMMALN IN EXCEL 2000, 2003 AND 2007

False Zeros	Non-Numeric Returns	Gross Errors	Logic Traps and Loops	System Halts and Crashes
None	None	None	None	None

RECOMMENDED EXCEL 2000, 2003 AND 2007 GAMMALN USAGE

Range of x values	ROUND level	Basis
2 to 8	10	Floating Point
8 to 27	11	Floating Point
27 to 98	12	Floating Point
98 up	13	Floating Point

2.2 DENSITY AND RELATED FUNCTIONS

2.2.1 EXPONENTIAL DISTRIBUTION, DENSITY: EXPONDIST

The function is =EXPONDIST(x, lambda, FALSE)

This is the direct output of the EXP function, a basic mathematical function. There was no alternate function to evaluate the accuracy of EXP. It can be assumed to be accurate to LRE values above 15.

2.2.2 GAMMA DISTRIBUTION, DENSITY: GAMMADIST

The function is =GAMMADIST(x, alpha, beta, FALSE)

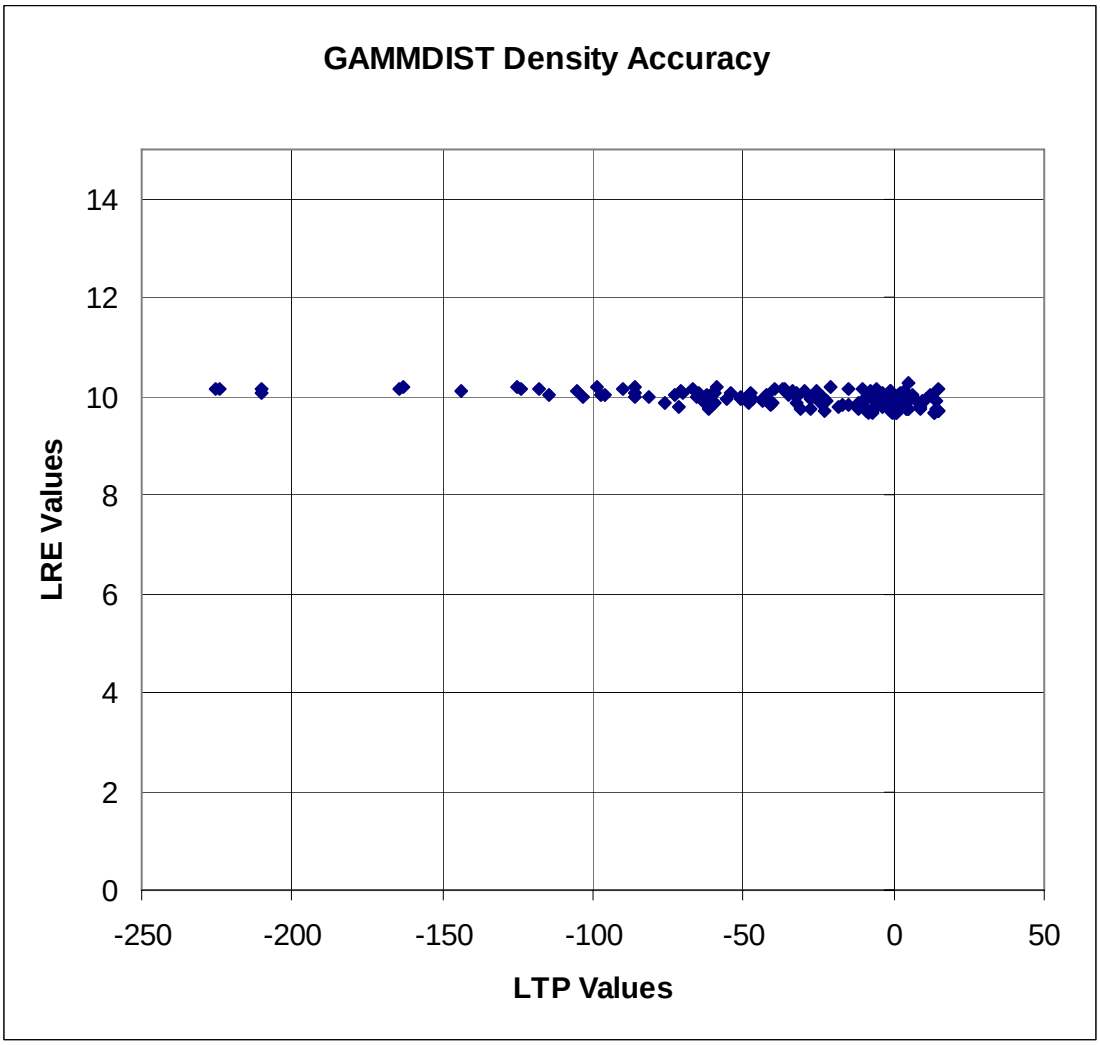
The Excel function as shown in Help, is a two- parameter (alpha and beta) gamma distribution, which is different from the standard one-parameter form given in chapter 6 of Abramowitz (1964). Abramowitz (1964) does not show the 2-parameter form anywhere. If the third parameter is set to one, then the standard form results, and cumulative values directly obtained from the chi distribution (Equation 26.4.19 in Abramowitz (1964).)

The 2-parameter form equation in Help after some algebra, agrees with the 2-parameter form given on page 165 of Lloyd and Lipow (1962). Lloyd and Lipow show that the variable (2*x / beta) is chi-square distributed with 2*alpha degrees of freedom. Note that the standard gamma distribution is not a fully normalized form of the general 2-parameter gamma distribution. The x / beta value in the general gamma distribution is not equal to the x value in the standard version. As a result values of GAMMADIST will not fully equal values of the one parameter standard form unless beta = 1.

EXCEL 2000

Values for the density were calculated for random values of x in the range of 0 to 100, random values of alpha in the range of 0 to 500, and random values of beta in the range of 0 to 1.

Figure 16-18: GAMMADIST Distribution Density Accuracy, Full Range

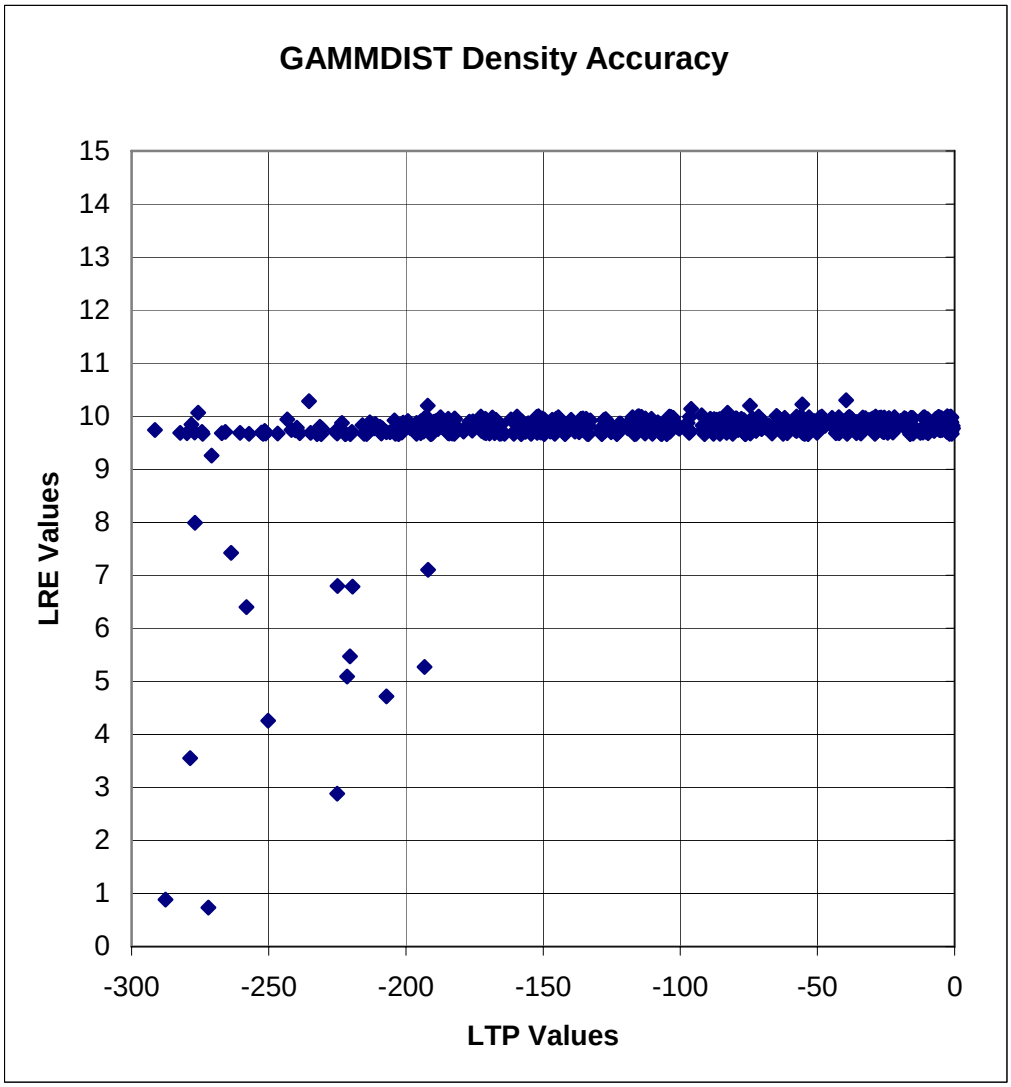


EXCEL 2003 and 2007

There was no change in the GAMMADIST density function for Excel 2003 and 2007.

Figure 16-19 shows the results of testing the Excel 2003 density function. The LTP values are just a straightforward log10 value of the density term. The data comes from random values of x, alpha and beta. X values vary from 0 to 1000, alpha values from 0 to 100 and beta values from 0 to 1.

Figure 16-19: Excel 2003 and 2007, GAMMADIST Density



The majority of the points lie above an LRE value of 9.6. For points from a smaller range of x and alpha (0 to 1 and 0 to 1) the values have LRE values all above 10. As the density value gets very small, the accuracy becomes less.

The differences between figures 16-18 and 16-19 have to do with the ranges of x, alpha and beta. Higher values of x in relation to smaller values of alpha result in the loss of accuracy for small density values.

For density computations of values in the above range, the frequency of #NUM!-returns is about 0.1%. Each #NUM! is a true zero. The false zero rate is about 6%. Given the loss in accuracy at very small density values and the increasing false zero rate, density values less than 1E-150 should be taken as zero.

RELIABILITY ASSESSMENT OF GAMMADIST (DENSITY) IN EXCEL 2000, 2003 AND 2007

False Zeros	Non-Numeric Returns	Gross Errors	Logic Traps and Loops	System Halts and Crashes
Varies, 6%?	0.1%	None	None	None

RECOMMENDED EXCEL 2000, 2003 AND 2007 GAMMADIST (DENSITY) USAGE

Range of x Values	Range of alpha Values	Range of beta Values	Restrictions	ROUND level	Basis
0 to 1000	0 to 100	0 to 1	Density value greater than 1E-150	9	Floating Point

2.2.3 NORMAL DISTRIBUTION, DENSITY: NORMDIST

The function is =NORMDIST(x, mean, standard_dev, FALSE)

This function takes the input values and directly calculates a density value as shown in Help. The accuracy of the returned value does not depend on any statistical distribution, only on Excel's inherent accuracy of performing arithmetic operations and on the accuracy of the EXP function.

2.2.4 WEIBUL DISTRIBUTION, DENSITY: WEIBULL

The function is =WEIBULL(x, alpha, beta, FALSE)

The probability values, both density and cumulative come directly by equation. Help shows the two equations. There were no tests made, since values come directly from an equation.