

The Mean Difference for Lognormal Distribution

Giovanni Girone, Fabio Manca

University of Bari "Aldo Moro", Bari, Italy

Email: giovanni.girone@uniba.it, fabio.manca@uniba.it

Received 30 March 2016; accepted 23 May 2016; published 26 May 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The calculation of the mean difference for the lognormal distribution involves several hard integrals featuring the error function. In this paper, considering two particular cases of an integral of the exponential function for the complement to one of the error functions, and using various symmetries, we have achieved the result of an extremely simple and useful formula of the mean difference for the lognormal distribution.

Keywords

Mean Difference, Function Lognormal

1. Introduction

In several previous studies [1]-[4], we have obtained close formula of the Gini mean difference for many continuous and discrete distributive models. In those papers, due to the difficulty in finding a solution for the various integrals featuring the error function, we were unable to deduce the formula of the mean difference of the lognormal model. The aim of this study is to bridge that gap.

2. Lognormal Distribution

The density function of the lognormal distribution is:

$$f(x) = \frac{e^{-\frac{[\log(x-\theta)]^2}{2\gamma^2}}}{\lambda x \sqrt{2\pi}}, \quad x > \theta, \gamma > 0.$$

Without going into generalizations, given that the mean difference is independent of the position parameter

and homogenous of the scale parameter, we consider the standardized model with $\theta = 0$

$$f(x) = \frac{e^{-\frac{[\log x^2]}{2\gamma^2}}}{\lambda x \sqrt{2\pi}}, \quad x > \theta, \gamma > 0.$$

The cumulative distribution function of the lognormal distribution is:

$$F(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\log x}{\sqrt{2}\lambda} \right) \right].$$

3. The Mean Difference

Given a continuous variable with density function:

$$f(x) = 0, \quad -\infty < x < a,$$

$$f(x) > 0, \quad a < x < b,$$

$$f(x) = 0, \quad b < x < \infty,$$

cumulative distribution function

$$F(x) = 0, \quad -\infty < x < a,$$

$$F(x) = \int_a^x f(u) du, \quad a < x < b,$$

$$F(x) = 1, \quad b < x < \infty,$$

and first incomplete moment

$$F_1(x) = 0, \quad -\infty < x < a,$$

$$F_1(x) = \int_a^x uf(u) du, \quad a < x < b,$$

$$F_1(x) = \mu, \quad b < x < \infty,$$

the mean difference of the distribution can be expressed by various formulas. The original formula is

$$\Delta = 2 \int_a^b \int_a^x (x-y) f(x) f(y) dy dx$$

the formula based on the density function, and the cumulative distribution function is

$$\Delta = 2 \int_a^b x [2F(x) - 1] f(x) dz,$$

another option being the formula based on the density function, the cumulative distribution function and the first incomplete moment:

$$\Delta = 2 \int_a^b [xF(x) - F_1(x)] f(x) dx,$$

and, lastly, the formula based solely on the cumulative distribution function is

$$\Delta = 2 \int_a^b F(x) [1 - F(x)] dx.$$

For our purposes of calculating the mean difference of the lognormal distribution, we utilize the last formula.

4. Selected Integrals for the Error Function

The error function is a particular integral pertaining to a variant of the normal distribution

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

The following graph shown in **Figure 1** represents the error function for x between -3 and $+3$.
 Linked to the function of errors, we have the complementary error function

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x).$$

It can be easily seen that

$$\operatorname{erf}(-x) = -\operatorname{erf}(x).$$

The calculation of the mean difference of the lognormal distribution requires the utilization of integrals involving various functions, one of these being the error function. These integrals, not available for consideration in standard mathematical calculation software, are presented here in forms rendered highly adaptable to various needs. In our adaptation, in order to obtain the mean difference of the lognormal model, these integrals have been listed as follows

$$\int_0^\infty \frac{1}{2} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right] dy = \frac{1}{2} \left[1 - e^{-\frac{\lambda^2}{2}} \operatorname{erf} \left(\frac{\lambda}{\sqrt{2}} \right) \right]$$

$$\int_0^\infty \frac{1}{2} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right] dy = \frac{1}{2} \left[-1 + e^{\frac{\lambda^2}{2}} \left(1 + \operatorname{erf} \left(\frac{\lambda}{\sqrt{2}} \right) \right) \right]$$

$$\int_0^\infty \frac{1}{2} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right]^2 dy = -\frac{1}{2} + \frac{e^{\frac{\lambda^2}{2}}}{2} + e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(-\frac{\lambda}{2} \right) - e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(-\frac{\lambda}{2} \right)^2 - e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(-\frac{\lambda}{\sqrt{2}} \right)$$

$$\int_0^\infty \frac{1}{2} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right]^2 dy = \frac{1}{2} - \frac{e^{\frac{\lambda^2}{2}}}{2} - e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(\frac{\lambda}{2} \right) + e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(\frac{\lambda}{2} \right)^2 + e^{\frac{\lambda^2}{2}} \operatorname{erf} \left(\frac{\lambda}{\sqrt{2}} \right)$$

$$\int_0^\infty \frac{1}{2} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right]^2 dy = \frac{1}{2} \left[1 - e^{-\frac{\lambda^2}{2}} \operatorname{erfc} \left(\frac{\lambda}{2} \right)^2 \right]$$

$$\int_0^\infty \frac{1}{2} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right) \right]^2 dy = \frac{1}{2} \left\{ e^{\frac{\lambda^2}{2}} \left[1 + \operatorname{erf} \left(\frac{\lambda}{2} \right) \right]^2 - 1 \right\}$$

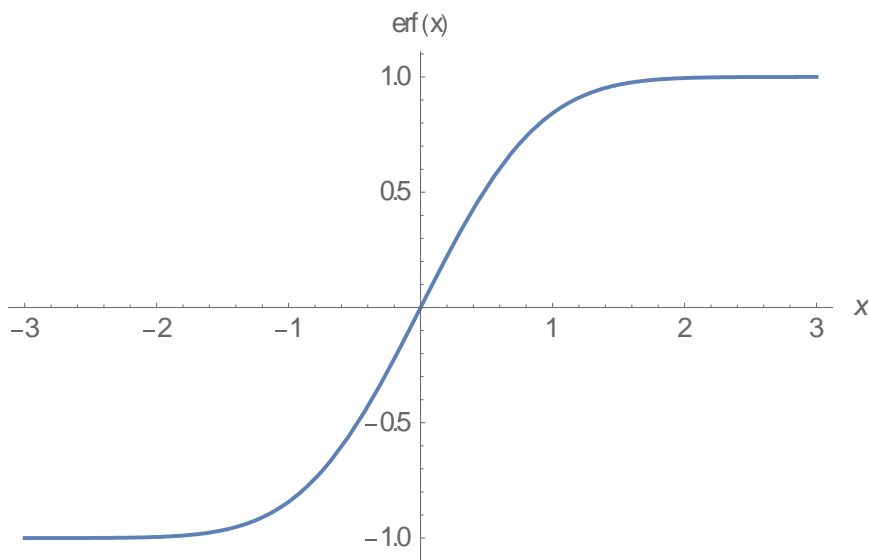


Figure 1. Function of error for x between -3 and 3 .

The first two integrals are easily calculated using mathematical software. The following two are very hard and were calculated considering particular cases of an integral given by Prudnikov, Brychkov and Marichev, 1986. The last two integrals were obtained by subtracting the first integral from the third, twice, and the second integral from the fourth, twice.

5. Calculation Procedure for the Mean Difference of Lognormal Distribution

To calculate the mean difference of lognormal distribution, it is possible to start from one of the four formulas indicated in paragraph 2. After several attempts, we opted for the formula based solely on the cumulative distribution function

$$\Delta = \int_a^b F(x)[1 - F(x)] dx.$$

The cumulative distribution function of lognormal distribution, as indicated in paragraph 2, is

$$F(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\log x}{\sqrt{2\lambda}} \right) \right].$$

Utilizing the said cumulative distribution function in the Δ formula, after some simplifications, we obtain

$$\Delta = \frac{1}{2} \int_0^\infty \left[1 - \operatorname{erf} \left(\frac{\log x}{\sqrt{2\lambda}} \right)^2 \right] dx,$$

introducing the transformation of the variable

$$Y = \log X,$$

we achieve the formula:

$$\Delta = \frac{1}{2} \int_{-\infty}^{\infty} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy,$$

which, due to the symmetry of the error function, can also be expressed by the negative argument of the exponential function

$$\Delta = \frac{1}{2} \int_{-\infty}^{\infty} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy,$$

splitting the two integrals, we obtain the formulas

$$\Delta = \frac{1}{2} \int_{-\infty}^0 e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy + \frac{1}{2} \int_0^{\infty} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy,$$

$$\Delta = \frac{1}{2} \int_{-\infty}^0 e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy + \frac{1}{2} \int_0^{\infty} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy.$$

Considering the presence of equality

$$\frac{1}{2} \int_{-\infty}^0 e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy = \frac{1}{2} \int_0^{\infty} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy,$$

the formula for the mean difference can be expressed by utilizing only the integrals defined by the positive semi-axis

$$\Delta = \frac{1}{2} \int_0^{\infty} e^y \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy + \frac{1}{2} \int_0^{\infty} e^{-y} \left[1 - \operatorname{erf} \left(\frac{y}{\sqrt{2\lambda}} \right)^2 \right] dy.$$

The two integrals in this formula were obtained in the previous paragraph, introducing them into the said formula and simplifying, we obtain the extremely simple expression for the mean difference of the lognormal distribution:

$$\Delta = 2e^{\frac{\lambda^2}{2}} \operatorname{erf}(\lambda/2).$$

6. Conclusion

Through a painstaking process utilizing selected hard integrals with functions including the error function, we have obtained an extremely simple formula of the mean difference of the lognormal distribution. This result can be added to the other findings regarding other distribution models. This result is useful for elaborating in-depth characterizations of the lognormal model and facilitating comparisons with other variability indexes of the said model.

Attributions of Work

Girone Sections 2, 4 and 6, Manca Sections 1, 3 and 5.

References

- [1] Girone, G. and Mazzitelli, D. (2007) La differenza media nei principali modelli distributivi continui, Annali del Dipartimento di Scienze Statistiche "Carlo Cecchi" dell'Università degli Studi di Bari, Cacucci Editore.
- [2] Girone, G., Massari, A. and Mazzitelli, D. (2015) More on the Mean Difference of Continuous Distributive Models. *Spontaneous Contribution in the Proceedings/Book of Abstracts of the SIS Conference "Statistics and Demography: The Legacy of Corrado Gini"*, Treviso, 9-11 Settembre 2015.
- [3] Girone, G., Manca, F. and D'Uggento, A.M. (2015) The Mean Difference of Discrete Distribution Models. *Spontaneous Contribution in the Proceedings/Book of Abstracts of the SIS Conference "Statistics and Demography: The Legacy of Corrado Gini"*, Treviso, 9-11 Settembre 2015.
- [4] Prudnikov, A.P., Bychkov, Yu.A. and Marichev, O.I. (1986) Integrals and Series, Vol. 2: Special Functions. Gordon and Breach Science Publishers, New York.