Characteristic Function for the Truncated Triangular Distribution

Jay J. Kim[1], Myron Katzoff[1] and Rahul A. Parsa[2]

[1] National Center for Health Statistics, 3311Toledo Road, Hyattsville, MD. 20782
[2] College of Business and Public Administration, Drake University, Des Moines, IA 50321

Abstract[3]. Two types of procedures for masking microdata can be considered: adding independent noise and multiplying by independent noise. The triangular distribution is one of the options for incorporating multiplicative noise. The triangular distribution truncated around 1 is a better option than the untruncated triangular distribution and is in use. The distributional form of the truncated triangular distribution was developed and is now available. Investigation of the statistical properties of data masked by using the truncated triangular distribution requires knowledge of the moments of the truncated triangular distribution. In this paper, we first develop the characteristic function for the truncated triangular distribution; then, the moments are derived from the characteristic function. We present a general formula for the moments. This will help data users with recovering the estimated moments of the original data.

Key words: masking, multiplicative noise, moment

1. Introduction

Since around 1980, the U.S. Energy Information Administration (EIA) has been using multiplicative noise for masking the number of heating and cooling days in an area, etc., in their public use micro data file from the Residential Energy Consumption Survey. EIA uses noise which follows the truncated normal distribution [Hwang, (1)]. Evans, et al [2] proposed the use of multiplicative noise to mask economic data. They considered noise which follows distributions such as normal and truncated normal distributions. Kim and Winkler [3] considered multiplicative noise which follows the truncated normal distribution. The U.S. Bureau of the Census uses a truncated triangular distribution for masking the Commodity Flow Survey data (Aboud [4]). Kim [5] developed the probability density function (pdf) of the truncated triangular distribution and showed that the estimate from the data masked by the distribution is unbiased if the triangular distribution is symmetric about 1 and truncated symmetrically about 1.

Multiplicative noise has the following form:

---

[3] **Disclaimer**: The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

$$y_i = x_i e_i, \; i = 1, 2, \ldots\ldots, n,$$

where y_i is the masked variable for the i^{th} unit such as person, household, establishment, etc., x_i is the corresponding un-masked variable and e_i (>0) is the noise.

If a public use microdata file is masked by multiplicative noise which follows the truncated triangular distribution, then data users need to know the moments of the truncated triangular distribution in order to estimate the moments of the original data. The best way of deriving moments of a distribution is developing the characteristic function of the distribution. Thus, in this paper, we develop the characteristic function of the truncated triangular distribution. Using the characteristic function, we derive the moments of the distribution. Finally, we generalize the formula for any moment.

2 Truncated Triangular Distribution

2.1 Triangular Distribution

A truncated triangular distribution is a modified form of a triangular distribution, and thus we first consider a triangular distribution which is shown in Figure 1. The triangular distribution is very useful. It can be used for approximating the normal, gamma and beta distributions. The triangular distribution is analytically easier to handle than the normal distribution.

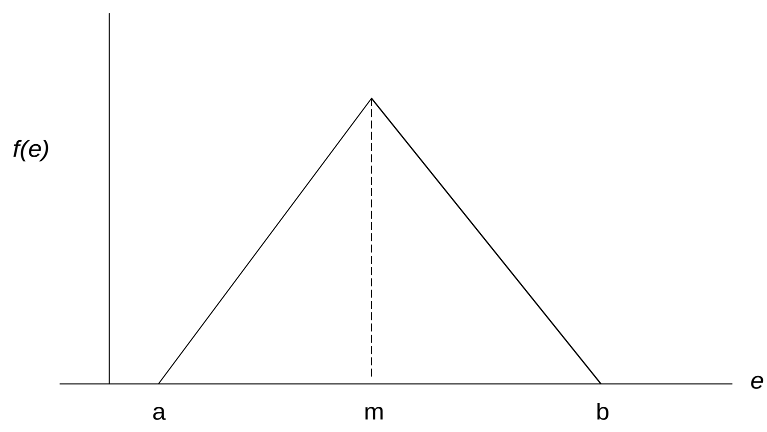

Figure 1. Triangular Distribution

The triangular distribution of a random variable e as shown above has the following form.

$$f(e)=\begin{cases}\dfrac{2}{b-a}\dfrac{1}{m-a}(e-a), & a\le e<m\\[2ex]\dfrac{2}{b-a}\dfrac{1}{b-m}(b-e), & m\le e<b\end{cases}. \tag{1}$$

Here m is the mode and is unique.

2.2 Truncated Triangular Distribution

When triangular distribution-based noise is used, one must avoid using the number close to one (1) for noise, because multiplying by a number very close to 1 does not change the original value that much, and thus the original value does not get any protection. In addition, the probability density for e is the greatest, when e is near 1. This suggests that the largest number of units do not get protection. Hence, it has been suggested [(2)] to truncate the mid-section, or the section near 1 of the triangular distribution. The truncated triangular distribution has the following shape.

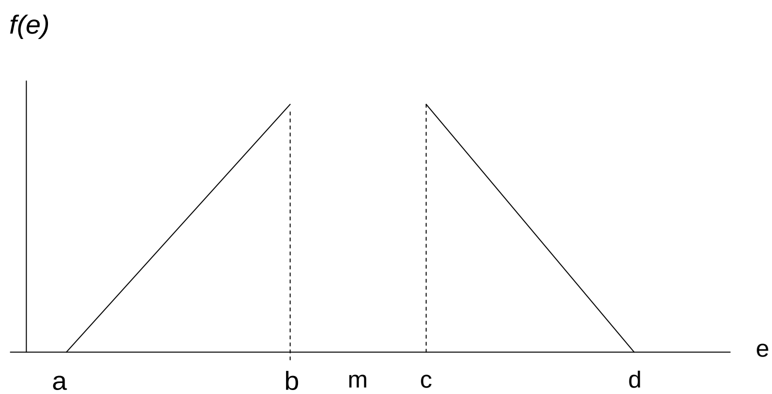

Figure 2. Truncated Triangular Distribution

Suppose the distribution is truncated at b and c, $c>b$, as shown in Figure 2. In this case, the pdf has the following form [Kim, (5)]:

$$f(e)=\begin{cases}\dfrac{2(d-m)}{(b-a)^2(d-m)+(d-c)^2(m-a)}\,(e-a), & a\le e<b\\[2ex] \dfrac{2(m-a)}{(b-a)^2(d-m)+(d-c)^2(m-a)}\,(d-e), & c\le e<d\,.\end{cases} \tag{2}$$

Let r denote the *degree* (or rate) *of truncation*. If we complete the triangle of Figure 2 with base *ad* to form a triangular distribution, the *degree of truncation* is the amount of probability under that density function from b to c. Since the interval of truncation enclosing the mode of that triangular distribution need not have endpoints placed symmetrically about the median m of the triangular distribution with base *ad*, m is now really a convenient symbolic artifact for relating the development of section 2.1 to the case when we truncate the triangular distribution. As we show next, m is an unknown that is the root of a quadratic equation. We have that $r = 1 - p$, where

$$\begin{aligned} p &= \int_a^b \frac{2}{d-a}\frac{1}{m-a}\left(e-a\right)de \;+\; \int_c^d \frac{2}{d-a}\frac{1}{d-m}\left(d-e\right)de \\ &= \frac{1}{d-a}\left[\frac{(b-a)^2}{m-a}+\frac{(d-c)^2}{d-m}\right] \end{aligned} \tag{3}$$

In the statement of the truncated triangular distribution function, m is not ordinarily included (Figure 2 calls this point to the attention of the reader). From equation (3) it follows that

$$m=\frac{-B\pm\sqrt{B^2-4AC}}{2A}\,,$$

where $A=p(d-a)$, $B=\left[(d-c)^2-(b-a)^2-p(d^2-a^2)\right]$ and $C=d(b-a)^2-a(d-c)^2+pad(d-a)$.

In the expression for m, we have both + and -. Selection of m value will depend on whether m is in the truncated region or not.

Note that the data dissemination agency should let the users know the degree of truncation (r).

3 Characteristic Function

The density function of the truncated triangular distribution is given by equation (2). By letting in that equation

$$k = \frac{2}{(b-a)^2(d-m)+(d-c)^2(m-a)} \ , \tag{4}$$

the truncated triangular distribution can be expressed as

$$f(x) = \begin{cases} k\,(d-m)\,(x-a), & a \le x < b \\ k\ (m-a)(d-x), & c \le x < d\,. \end{cases} \tag{5}$$

Note that the above probability density function has two parts. The first part represents the side to the left of the region of truncation and the second part the side to the right of the region of truncation.

The characteristic function of the truncated triangular distribution due to the left side of the distribution is,

$$k(d-m)\int_a^b (x-a)e^{itx}dx = k(d-m)\int_a^b xe^{itx}dx - k(d-m)\int_a^b ae^{itx}dx \ . \tag{6}$$

Using integration by parts, ignoring $k(d-m)$ for the time being, we let $x = u$ and $dv = e^{itx}dx$. Then $dx = du$ and $v = \frac{e^{itx}}{it}$. Thus the first term in equation (6) without $k(d-m)$ becomes

$$\begin{aligned} \int_a^b xe^{itx}dx &= \frac{x}{it}\ e^{itx}\Big|_a^b - \frac{1}{it}\int_a^b e^{itx}dx \\ &= \frac{x}{it}\ e^{itx}\Big|_a^b - \frac{1}{i^2t^2}e^{itx}\Big|_a^b \\ &= \frac{b}{it}e^{itb} - \frac{a}{it}e^{ita} - \frac{1}{i^2t^2}\left(e^{itb} - e^{ita}\right) \ . \end{aligned} \tag{7}$$

The second term of equation (6) becomes

$$\int_a^b ae^{itx}dx = \frac{a}{it}\ e^{itx}\Big|_a^b = \frac{a}{it}\left(e^{itb} - e^{ita}\right). \tag{8}$$

Putting equations (7) and (8) into equation (6), we have

$$\frac{bk(d-m)}{it}e^{itb}-\frac{ak(d-m)}{it}e^{ita}-\frac{k(d-m)}{i^2t^2}\left(e^{itb}-e^{ita}\right)-\frac{ak(d-m)}{it}\left(e^{itb}-e^{ita}\right)$$

$$=\frac{k(d-m)}{it}\left(be^{itb}-ae^{itb}\right)-\frac{k(d-m)}{i^2t^2}\left(e^{itb}-e^{ita}\right). \qquad (9)$$

Similarly, the characteristic function of the truncated triangular distribution due to the right side of the distribution is,

$$k(m-a)\int_c^d (d-x)e^{itx}dx=k(m-a)\int_c^d de^{itx}dx-k(m-a)\int_c^d xe^{itx}dx$$

$$=\frac{dk(m-a)}{it}\,e^{itx}\Big|_c^d-k(m-a)\frac{x}{it}\,e^{itx}\Big|_c^d+\frac{k(m-a)}{i^2t^2}e^{itx}\Big|_c^d$$

$$=\frac{dk(m-a)}{it}\left(e^{itd}-e^{itc}\right)-\frac{k(m-a)}{it}\left(de^{itd}-ce^{itc}\right)+\frac{k(m-a)}{i^2t^2}\left(e^{itd}-e^{itc}\right)$$

$$=\frac{k(m-a)}{it}\left(ce^{itc}-de^{itc}\right)+\frac{k(m-a)}{i^2t^2}\left(e^{itd}-e^{itc}\right). \qquad (10)$$

Thus the characteristic function is

$$\Psi(t)=k(d-m)\left(\frac{be^{itb}-ae^{itb}}{it}+\frac{e^{itb}-e^{ita}}{t^2}\right)+k(m-a)\left(\frac{ce^{itc}-de^{itc}}{it}-\frac{e^{itd}-e^{itc}}{t^2}\right)(11)$$

3.1 First Moment

The first derivative of $\Psi(t)$ is,

$$\Psi'(t)=k(d-m)\left(\frac{2ibe^{itb}}{t^2}+\frac{b^2e^{itb}}{t}-\frac{iae^{itb}}{t^2}-\frac{abe^{itb}}{t}+\frac{2e^{itb}}{t^3}+\frac{2e^{itb}}{t^3}-\frac{iae^{ita}}{t^2}\right)$$

$$+k(m-a)\left(\frac{2ice^{itc}}{t^2}+\frac{c^2e^{itc}}{t}-\frac{ide^{itc}}{t^2}-\frac{cde^{itc}}{t}+\frac{2e^{itd}}{t^3}-\frac{2e^{itd}}{t^3}-\frac{ide^{itd}}{t^2}\right) \quad (12)$$

To get the first moment, we normally replace t by 0 in the above. However, by doing so, we get $\frac{a}{0}$ or $\frac{0}{0}$, where a is a non-zero constant. Thus, we must apply L'Hopital's rule first, sometimes more than once for the same term. The results are as follows.

$$ik(d-m)\left(-b^3e^{itb}+b^3e^{itb}+\frac{ab^2e^{itb}}{2}-ab^2e^{itb}+\frac{b^3e^{itb}}{3}-\frac{a^3e^{ita}}{3}+\frac{a^3e^{ita}}{2}\right)$$

$$+ik(m-a)\left(-c^3e^{itc}+c^3e^{itc}+\frac{c^2de^{itc}}{2}-c^2de^{itc}-\frac{d^3e^{itd}}{3}+\frac{d^3e^{itd}}{2}+\frac{c^3e^{itc}}{3}\right)$$

$$=ik(d-m)\left(\frac{b^3e^{itb}}{3}-\frac{ab^2e^{itb}}{2}+\frac{a^3e^{ita}}{6}\right)+ik(m-a)\left(\frac{c^3e^{itc}}{3}-\frac{c^2de^{itc}}{2}+\frac{d^3e^{itd}}{6}\right).$$

By setting $t = 0$ in the above and dividing by i, we get the first moment,

$$\Psi'(0)=\frac{k(d-m)}{6}(2b^3-3ab^2+a^3)+\frac{k(m-a)}{6}(2c^3-3c^2d+d^3). \quad (13)$$

For verification, the first moment of the truncated triangular distribution can be derived by a straightforward integration as follows.

$$E(x)=k(d-m)\int_a^b x(x-a)dx+k(m-a)\int_c^d x(d-x)dx$$

$$=\frac{k(d-m)}{6}\left(2b^3-3ab^2+a^3\right)+\frac{k(m-a)}{6}\left(2c^3-3c^2d+d^3\right). \quad (14)$$

By replacing k in equation (13) with the expression in equation (4),

$$E(x)=\frac{(d-m)(b-a)^2(2b+a)+(m-a)(d-c)^2(2c+d)}{3\left[(b-a)^2(d-m)+(d-c)^2(m-a)\right]}$$

In case $b - a = d - c$ and $m - a = d - m$, the above simplifies further.

3.2 Second Moment

By taking the second derivative of $\Psi(t)$ in equation (10), we get

$$\Psi''(t) = k(d-m)\left(-\frac{6ibe^{itb}}{t^3} - \frac{b^2e^{itb}}{t^2} + \frac{ib^3e^{itb}}{t} + \frac{2iae^{itb}}{t^3} + \frac{2abe^{itb}}{t^2} - \frac{iab^2e^{itb}}{t} + \frac{6e^{itb}}{t^4} - \frac{6e^{ita}}{t^4} + \frac{4iae^{ita}}{t^3} + \frac{a^2e^{ita}}{t^2}\right) + k(m-a)\left(-\frac{6ice^{itc}}{t^3} - \frac{3c^2e^{itc}}{t^2} + \frac{ic^3e^{itc}}{t} + \frac{2ide^{itc}}{t^3} + \frac{2cde^{itc}}{t^2} - \frac{ic^2de^{itc}}{t} - \frac{6e^{itd}}{t^4} + \frac{6e^{itc}}{t^4} + \frac{4ide^{itd}}{t^3} + \frac{d^2e^{itd}}{t^2}\right). \quad (15)$$

Applying L'Hopital's rule repeatedly, we get

$$k(d-m)\left(\frac{2i^2b^4e^{itb}}{3} - i^2b^4e^{itb} - \frac{i^2b^4e^{itb}}{2} + i^2b^4e^{itb} - \frac{i^2ab^3e^{itb}}{3} - \frac{i^2b^4e^{itb}}{4} + \frac{i^2b^4e^{itb}}{3} + \frac{i^2a^4e^{ita}}{4} - \frac{2i^2a^4e^{ita}}{3} + \frac{i^2a^4e^{ita}}{2}\right) + k(m-a)\left(i^2c^4e^{itc} - \frac{3i^2c^4e^{itc}}{2} - \frac{i^2c^3de^{itc}}{3} + i^2c^3de^{itc} - \frac{2i^2d^4e^{itd}}{3} + \frac{i^2d^4e^{itd}}{2} + i^2c^4e^{itc} - i^2c^3de^{itc} + \frac{i^2d^4e^{itd}}{4} - \frac{i^2c^4e^{itc}}{4}\right).$$

Setting $t = 0$ in the above and dividing by i^2, we get the second moment around 0,

$$\Psi''(0) = \frac{k(d-m)}{12}\left(3b^4 - 4ab^3 + a^4\right) + \frac{k(m-a)}{12}\left(3c^4 - 4c^3d + d^4\right). \quad (16)$$

3.3 Third Moment

By taking the third derivative of $\Psi(t)$ in equation (11) and pulling the same terms together, we get

$$\Psi'''(t)=k(d-m)\left(\frac{24ibe^{itb}}{t^4}+\frac{12ib^2e^{itb}}{t^3}-\frac{4ib^3e^{itb}}{t^2}-\frac{b^4e^{itb}}{t}-\frac{6iae^{itb}}{t^4}+\frac{18iae^{ita}}{t^4}\right.$$

$$\left.-\frac{6abe^{itb}}{t^3}+\frac{3iab^2e^{itb}}{t^2}-\frac{ab^3e^{itb}}{t}+\frac{24e^{ita}}{t^5}-\frac{24e^{itb}}{t^5}-\frac{6a^2e^{ita}}{t^3}+\frac{ia^3e^{ita}}{t^2}\right)$$

$$+k(m-a)\left(\frac{24ice^{itc}}{t^4}+\frac{12c^2e^{itc}}{t^3}-\frac{4ic^3e^{itc}}{t^2}-\frac{c^4e^{itc}}{t}-\frac{6ide^{itc}}{t^4}-\frac{18ide^{itd}}{t^4}\right.$$

$$\left.-\frac{6cde^{itc}}{t^3}+\frac{3ic^2de^{itc}}{t^2}+\frac{c^3de^{itc}}{t}+\frac{24e^{itd}}{t^5}-\frac{24e^{itc}}{t^5}-\frac{6d^2e^{itd}}{t^3}+\frac{id^3e^{td}}{t^2}\right). \quad (17)$$

Applying L'Hopital's rule repeatedly, setting $t = 0$ in $\Psi'''(t)$ and dividing the combined terms by i^3, we get the third moment around 0 as follows:

$$\Psi'''(0)=\frac{k(d-m)}{20}\left(4b^5-5ab^4+a^5\right)+\frac{k(m-a)}{20}\left(4c^5-5c^4d+d^5\right). \qquad (18)$$

From equations (13), (16) and (18), one can infer that the fourth moment around 0 is

$$\Psi''''(0)=\frac{k(d-m)}{30}\left(5b^6-6ab^5+a^6\right)+\frac{k(m-a)}{30}\left(5c^6-6c^5d+d^6\right), \qquad (19)$$

and that the fifth moment around 0 is,

$$\Psi'''''(0)=\frac{k(d-m)}{42}\left(6b^7-7ab^6+a^7\right)+\frac{k(m-a)}{42}\left(6c^7-7c^6d+d^7\right). \qquad (20)$$

3.4 p^{th} Moment

Theorem.

The p^{th} moment of the truncated triangular distribution is

$$\Psi^p(0)=\frac{k(d-m)}{(p+1)(p+2)}\left[(p+1)b^{p+2}-(p+2)ab^{p+1}+a^{p+2}\right]$$

$$+\frac{k(m-a)}{(p+1)(p+2)}\left[(p+1)c^{p+2}-(p+2)c^{p+1}d+d^{p+2}\right]. \qquad (21)$$

Proof.

The result can be proved by mathematical induction.

When $p = 1$,

$$\Psi'(0)=\frac{k(d-m)}{2\cdot 3}(2b^3-3ab^2+a^3)+\frac{k(m-a)}{2\cdot 3}(2c^3-3c^2d+d^3).$$

The above formula is the same as what we observed in equations (13) and (14) for $p = 1$.

Suppose it is true for $p = h$. That is,

$$\Psi^{h}(0)=\frac{k(d-m)}{(h+1)(h+2)}\left[(h+1)b^{h+2}-(h+2)ab^{h+1}+a^{h+2}\right]$$

$$+\frac{k(m-a)}{(h+1)(h+2)}\left[(h+1)c^{h+2}-(h+2)c^{h+1}d+d^{h+2}\right].$$

Then when $p = h + 1$,

$$\Psi^{h+1}(0)=\frac{k(d-m)}{[(h+1)+1][(h+1)+2]}\left\{[(h+1)+1]b^{(h+1)+2}-[(h+1)+2]ab^{(h+1)+1}+a^{(h+1)+2}\right\}$$

$$+\frac{k(m-a)}{[(h+1)+1][(h+1)+2]}\left\{[(h+1)+1]c^{(h+1)+2}-[(h+1)+2]c^{(h+1)+1}d+d^{(h+1)+2}\right\}.$$

The above shows the theorem holds when $p = h + 1$.

4. Concluding Remarks

If a public use microdata file is masked by multiplicative noise which follows the truncated triangular distribution, then data users need to know the moments of the truncated triangular distribution in order to estimate the moments of the original data. The best way of deriving moments of a distribution is developing the characteristic function of the distribution. Thus, in this paper, we developed the characteristic function of the truncated triangular distribution. Using the characteristic function, we derived the moments of the distribution. Finally, we generalized the formula for the moments.

5. References

1. Hwang, J.T. (1986) Multiplicative Errors-in-Variables Models with Applications to Recent Data Released by the U.S. Department of Energy, Journal of the American Statistical Association, Vol. 81, No. 395, 680-688.

2. Evans, T., Zayatz, L., and Slanta, J. (1998) Using Noise for Disclosure Limitation of Establishment Tabular Data, Journal of Official Statistics, Vol. 14, No. 4, pp537-551.

3. Kim, J..J. and Winkler, W. E. (2001) Multiplicative Noise for Masking Continuous Data, Proceedings of the Survey Methods Research Section, American Statistical Association, CD Rom.

4. Personal communication in 2007 with Aboud, J., Cornell University.

5. Kim, J..J. (2007) Application of Truncated Triangular and Trapezoidal Distributions for Developing Multiplicative Noise, Proceedings of the Survey Methods Research Section, American Statistical Association, CD Rom.