(Weighted) Sum of n Correlated Lognormals: convolution integral solution (Jan 2011)

Tamas Nagy

***Abstract*— Probability density function (pdf) for sum of n correlated lognormal variables is deducted as a special convolution integral. Pdf for weighted sums (where weights can be any real numbers) is also presented. The result for four dimensions was checked by Monte Carlo simulation.**

***Index Terms*— amount of fading, cochannel interference, lognormal distribution**

I. Introduction

THE original problem as stated by Fenton [1] in 1960 is the following: "Given several random variables, each with a log-normal probability distribution, what is the probability distribution of the sum of the random variables?"

By mathematical notation: let we have n lognormal variables, defined by the corresponding correlated multidimensional normal distribution (X), where contains expected values and Σ is the covariance matrix. We try to find the probability density function (θ) of sum of (correlated) lognormal variables (S(C)LN):

$$\overline{Y} = \begin{bmatrix} \exp(X_1) \\ \exp(X_2) \\ \dots \\ \exp(X_n) \end{bmatrix} \quad X \sim \mathrm{N}(\overline{\mu}, \Sigma) \quad \theta\left(\sum_{i=1}^{n} Y_i\right) = ?$$

Examples for S(C)LN can be found in almost every science, in wireless communications applications include [2] fading and shadowing modeling, and assessing cochannel interference. In finance the evaluation of exotic (basket, Asian) options [3], portfolio level Value at Risk and optimal portfolio selection are the most relevant S(C)LN problems.

Fenton [1] has written about (without detailed specifications) a numerical integration solution in case of uncorrelated summands. He concentrated on finding an approximate solution for uncorrelated case for avoiding (with his own words) the "tedious" work of numerical integration. He presented an approximation of distribution of uncorrelated SLN by moment matching. After Fenton's initial work almost all of the efforts were done trying to define better approximations [4]-[7], or lower and/or upper bounds for cumulative density function [8-9], [2] and less researches were done on numerical integral solution for problem of S(C)LN.

In the literature there are some direct integral expressions formulating the pdf or cdf of SCLN. For 2 dimensional correlated case Zacks and Tsokos [10] presented a formula of characteristic function. Leipnik [11] demonstrated the pdf of n dimensional uncorrelated SLN.

In the following, an exact, convolution integral solution for probability density function is presented for n-dimensional correlated case, which - based on my best knowledge - was not published before. Pdf of weighted sum of n correlated lognormals is also provided.

II. Convolution Integral Solution

In a general, 2 dimensional independent case, the pdf of two convoluted variables is the following:

$$h(K) = \int_{-\infty}^{\infty} f(x) \cdot g(K - x)dx = f \otimes g \tag{1}$$

In n dimensional case there are n-1 integrals, the argument of last probability density function is $K - \sum_{i=1}^{n-1} x_i$

For getting the right pdf of SCLN, few problems are needed to be handled:

- The random variables are correlated (the correlation structure is defined on the level of corresponding normal factors)
- The integral variable is dx, but the argument in lognormal pdf is ln(x)

Based on the linear transformation property [12] any multidimensional normal distribution (Σ is the covariance matrix) can be built form uncorrelated standard normal distribution (Z):

$$\begin{aligned} &L \cdot Z + \overline{\mu} \sim \mathrm{N}(\overline{\mu}, \mathrm{L} \cdot \mathrm{L}^{\mathrm{T}}) \\ &Z \sim \mathrm{N}(\overline{0}, \mathrm{I}_{\mathrm{n}}) \qquad \Sigma = \mathrm{L} \cdot \mathrm{L}^{\mathrm{T}} \end{aligned} \tag{3}$$

The formula for reversed way transformation is:

$$L^{-1} \cdot \left[\mathrm{N}(\overline{\mu}, \Sigma) - \overline{\mu}\right] = Z \tag{4}$$

Let Y_i denote the i-th dependent lognormal variable, and z_i

Manuscript received January 17, 2011.

T. Nagy is with Corvinus University of Budapest, Budapest, H-1093 Hungary (e-mail: tamasnagy1976@gmail.com).

the i-th independent standard normal variable:

$$N(\overline{\mu},\Sigma)=\begin{bmatrix}\ln Y_1\\ \ln Y_2\\ \cdots\\ \ln Y_n\end{bmatrix} \qquad Z=N(\overline{0},I)=\begin{bmatrix}z_1\\ z_2\\ \cdots\\ z_n\end{bmatrix} \tag{5}$$

Uncorrelated standard normal variables can be created by first decreasing the logarithm of correlated lognormal variables by their expected values and then creating linear combination by corresponding elements of L-1 matrix. The i-th independent standard normal variable can be calculated as follows (whereas L^{-1} matrix is lower triangular, only k ≤ i elements have non zero weights):

$$z_i=\sum_{k=1}^{n}\left[L^{-1}(i,k)\cdot(\ln Y_k-\mu_k)\right]=\sum_{k=1}^{i}\left[L^{-1}(i,k)\cdot(\ln Y_k-\mu_k)\right] \tag{6}$$

A. Joint probability density function

In the multidimensional dependent lognormal factor space the probability density related to a given point equals to the probability density related to the corresponding point in the multidimensional independent standard normal factor space. Joint probability density function of multidimensional correlated lognormal variables (f) can be expressed by one dimensional standard normal distributions (denoting h multivariate standard normal pdf, φ is one dimensional standard normal pdf, $L^{-1}_{i\_}$ the ith row of matrix L^{-1}):

$$f(\overline{Y})=h\left(L^{-1}\cdot\left[\ln\overline{Y}-\overline{\mu}\right]\right)=\prod_{i=1}^{n}\varphi\left(L^{-1}_{i\_}\cdot\left[\ln(\overline{Y})-\overline{\mu}\right]\right) \tag{7}$$

B. Modifications related to integral variables

For integrating function f, we have to make modifications related to integral variables. The relevant question is the following: if we change the correlated lognormal variables infinitesimally, what will be the n-volume of the subspace determined by changing of corresponding uncorrelated standard normal variables?

Let us change the lognormal variables one by one, let $\overline{\Delta Y_i}$ denote the change vector, where all of items are 0 but i-th element is ΔY. Vector $\overline{g_i}$ means the corresponding change vector in uncorrelated standard normal factor space caused by $\overline{\Delta Y_i}$, it can be calculated as follows:

$$\overline{g_i}=L^{-1}\cdot\left(\ln\left(\overline{Y}+\overline{\Delta Y_i}\right)-\overline{\mu}\right)-L^{-1}\cdot\left(\ln(\overline{Y})-\overline{\mu}\right) \tag{8}$$

In continuous case (ΔY→0), if $L^{-1}_{\_i}$ is the i-th column of matrix, and $*$ denoting element wise multiplication, then:

$$\overline{g_i}=L^{-1}_{\_i}*\overline{\left(\frac{dY}{Y}\right)} \tag{9}$$

Based on g vectors an M matrix can be built, storing the changes in space of independent standard normal variables caused by one by one changing of correlated lognormal variables. M determines an n-dimensional hyper-parallelepiped, its n-volume (nV) can be calculated as absolute value of M's determinant. Whereas M is triangular matrix, its determinant is the product of diagonal elements:

$$nV=\left|\prod_{i=1}^{n}\frac{L^{-1}_{ii}}{Y_i}\cdot dY_i\right| \tag{10}$$

C. Probability density function of SCLN

Based on the result we can construct the pdf of SCLN in n dimensional case. First in 3 dimensions (for the sake of illustration):

$$\begin{aligned}&\theta_{SCLN}(K)=\\&\int_0^\infty\int_0^\infty\frac{L^{-1}_{11}\cdot L^{-1}_{22}\cdot L^{-1}_{33}}{Y_1\cdot Y_2\cdot(K-Y_1-Y_2)}\cdot\varphi\left(L^{-1}_{11}\cdot(\ln(Y_1)-\mu_1)\right)\cdot\varphi\left(L^{-1}_{21}\cdot(\ln(Y_1)-\mu_1)+L^{-1}_{22}\cdot(\ln(Y_2)-\mu_2)\right)\cdot\\&\cdot\varphi\left(L^{-1}_{31}\cdot(\ln(Y_1)-\mu_1)+L^{-1}_{32}\cdot(\ln(Y_2)-\mu_2)+L^{-1}_{33}\cdot(\ln(K-Y_1-Y_2)-\mu_3)\right)\cdot dY_1\cdot dY_2\\&for\ K-Y_1-Y_2>0\end{aligned} \tag{11}$$

In n dimensional case the pdf of SCLN will be the following (for getting simpler structure let us use a b vector containing notations of variables):

$$\begin{aligned}&\theta_{SCLN}(K)=\int_0^\infty\int_0^\infty\ldots\int_0^\infty\prod_{i=1}^{n}\left[\frac{L^{-1}_{ii}}{b_i}\cdot\varphi\left(\sum_{k=1}^{i}L^{-1}_{ik}\cdot(\ln(b_k)-\mu_k)\right)\right]\cdot\prod_{i=1}^{n-1}dY_i\\&where\quad b=\left[Y_1,\ \ Y_2,\ \ \ldots,\ \ Y_{n-1},\ \ K-\sum_{i=1}^{n-1}Y_i\right]\\&for\quad K-\sum_{i=1}^{n-1}Y_i>0\end{aligned} \tag{12}$$

D. Probability density function of weighted SCLN (wSCLN)

We can extend the problem of SCLN to weighted case (where weights (α) can be any real numbers). The weighted sum of correlated lognormals (wSCLN) problem is typical in finance (e.g. a portfolio where we can hold negative quantity of asset, and the prices are lognormally distributed):

$$\theta\left(\sum_{i=1}^{n}\alpha_iY_i\right)=?$$

Pdf of wSCLN can be expressed as follows:

$$\begin{aligned}&\theta_{wSCLN}(K)=\int_0^\infty\int_0^\infty\ldots\int_0^\infty\prod_{i=1}^{n}\left[\frac{L^{-1}_{ii}}{c_i}\cdot\varphi\left(\sum_{k=1}^{i}L^{-1}_{ik}\cdot(\ln(b_k)-\mu_k)\right)\right]\cdot\prod_{i=1}^{n-1}dY_i\\&for\quad\left[K-\sum_{i=1}^{n-1}\alpha_i\cdot Y_i\right]\Big/\alpha_n>0\\&where\quad b=\left[Y_1,\ \ Y_2,\ \ \ldots,\ \ Y_{n-1},\ \ \frac{K-\sum_{i=1}^{n-1}\alpha_i\cdot Y_i}{\alpha_n}\right]\\&where\quad c=\left[Y_1,\ \ Y_2,\ \ \ldots,\ \ Y_{n-1},\ \ \left|K-\sum_{i=1}^{n-1}\alpha_i\cdot Y_i\right|\right]\end{aligned} \tag{13}$$

III. NUMERICAL VALIDATION

The numerical validation was done for wSCLN (weighted SLCN). The convolution integral was computed by own built algorithm. When designing such an algorithm significant amount of computational time can be spared by taking the condition in (13) into account.

For potential values of underlying lognormal variables (Y_i) symmetric intervals were supposed, defined by d times the given standard deviation:

$$\exp(\mu_i - d \cdot \sigma_i) \le Y_i \le \exp(\mu_i + d \cdot \sigma_i)$$

An initial lower (ilbi) and upper bound (iubi) for the given αiYi can be calculated (the min and max functions are required for handling the potential negative sign of α):

$$\begin{aligned} ilb_i &= \min[\alpha_i \cdot \exp(\mu_i - d \cdot \sigma_i), \alpha_i \cdot \exp(\mu_i + d \cdot \sigma_i)] \\ iub_i &= \max[\alpha_i \cdot \exp(\mu_i - d \cdot \sigma_i), \alpha_i \cdot \exp(\mu_i + d \cdot \sigma_i)] \end{aligned} \tag{15}$$

For further purposes a reversed-way cumulated sums of the iubi and iubi vectors were used:

$$\begin{aligned} rclb &= \left[\sum_{i=1}^{n} ilb_i, \quad ...., \quad \sum_{i=1}^{2} ilb_i, \quad ilb_1 \right] \\ rcub &= \left[\sum_{i=1}^{n} iub_i, \quad ...., \quad \sum_{i=1}^{2} iub_i, \quad iub_1 \right] \end{aligned} \tag{16}$$

Based on these vectors the potential values of K can be restricted (by increasing d, the intervals widen):

$$rclb(1) \le K \le rcub(1) \tag{17}$$

If we calculate value of θ(K) by multi-level nested loops, and the order of calculation is determined by i, conditional (narrowed) bounds for Y_i can be determined based on:

• Known variables (denoting with underlyings): K, Y_1, Y_2, …, Y_{i-1}

• Potential maximum and minimum values of yet unknown variables (Y_{i+1}, Y_{i+2}, …. Y_{n-1}), their cumulated sum stored by rclb and rcub vectors

• Initial bounds for Y_i (ilb_i and iub_i)

The conditional (narrowed) bounds for a given Y_i will be the following:

$$\begin{aligned} &\max\left[ilb_i, \quad \min\left(\left[\underline{K} - \sum_{j=1}^{i-1} \alpha_j \cdot \underline{Y_j} - rclb_{i+1} \right] \Big/ \alpha_i, \quad \left[\underline{K} - \sum_{j=1}^{i-1} \alpha_j \cdot \underline{Y_j} - rcub_{i+1} \right] \Big/ \alpha_i \right) \right] \\ &\le Y_i \le \\ &\min\left[iub_i, \quad \max\left(\left[\underline{K} - \sum_{j=1}^{i-1} \alpha_j \cdot \underline{Y_j} - rclb_{i+1} \right] \Big/ \alpha_i, \quad \left[\underline{K} - \sum_{j=1}^{i-1} \alpha_j \cdot \underline{Y_j} - rcub_{i+1} \right] \Big/ \alpha_i \right) \right] \end{aligned} \tag{18}$$

For comparing the theoretical and simulation result the calculation was done in 4 dimensions with the following parameters:

$$\alpha = \begin{bmatrix} -1.2 \\ -0.8 \\ 1.8 \\ 0.3 \end{bmatrix} \quad \mu = \begin{bmatrix} 0.10 \\ 0.08 \\ 0.22 \\ 0.07 \end{bmatrix} \quad \sigma = \begin{bmatrix} 0.10 \\ 0.2 \\ 0.08 \\ 0.1 \end{bmatrix} \quad \rho = \begin{bmatrix} 1.0 & 0.4 & 0.3 & 0.5 \\ 0.4 & 1.0 & -0.2 & 0.2 \\ 0.3 & -0.2 & 1.0 & -0.1 \\ 0.5 & 0.2 & -0.1 & 1.0 \end{bmatrix}$$

For calculating the theoretical pdf, own algorithm was used (d = 5, number of steps for K = 400, number of steps for Y1, Y2, Y3 = 80). The simulation method was Monte Carlo (with 6.000.000 repetitions).

The simulated histogram and theoretical frequencies were the followings:

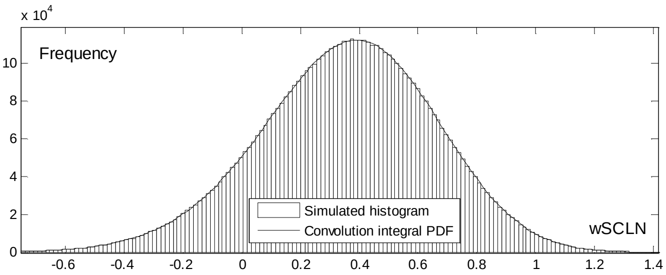

Fig. 1. Comparison of theoretical (convolution integral) and simulated frequencies

As we can see the simulated and theoretical frequencies are very close to each other.

IV. Conclusion

The exact form of pdf of SCLN and wSCLN can be calculated based on convolution integrals.

In numerical validation a 4 dimension weighted case was presented (which can not be handled by simple lognormal approximation (logarithm of negatives are not valid)). By special designing of numerical integral algorithm significant computational time can be spared, the total computation by suggested algorithm required 11.5 minutes in a single laptop. By using more sophisticated algorithm and powerful computer, the exact convolution solution can be a feasible way for also high dimensions.

References

[1] L. F. Fenton, "The sum of lognormal probability distributions in scatter transmission systems," *IRE Trans. Commun. Syst.*, vol. CS-8, pp. 57–67, Mar. 1960.

[2] C. Tellambura, "Bounds on the Distribution of a Sum of Correlated Lognormal Random Variables and Their Application," *IEEE Trans. Commun.*, vol. 56, no. 8, pp. 1241 – 1248, Aug. 2008.

[3] M. A. Milevsky and S. E. Posner, "Asian options, the sum of lognormals and the reciprocal gamma distribution," *J. Financ. Quant. Anal.*, vol. 33, no. 3, pp. 409–422, Sep. 1998.

[4] S. C. Schwartz and Y. S.Yeh, "On the distribution function and moments of power sums with lognormal components," *Bell Syst. Tech. J.*, vol. 61, pp. 1441–1462, Sept. 1982.

[5] N. C. Beaulieu and Q. Xie, "An optimal lognormal approximation to lognormal sum distribution," *IEEE Trans. Veh. Technol.*, vol. 53, no. 2, pp. 479–489, Mar. 2004.

[6] N. B. Mehta, J. Wu, A. F. Molisch, and J. Zhang, "Approximating a Sum of Random Variables with a Lognormal," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2690–2699, Jul. 2007.

[7] Q. T. Zhang and S. H. Song, "A Systematic Procedure for Accurately Approximating Lognormal-Sum Distributions," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 663–666, Jan. 2008.

[8] Ben Slimane, "Bounds on the distribution of a sum of independent lognormal random variables," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 975–978, Jun. 2001.

[9] F. Berggren and S. Slimane, "A simple bound on the outage probability with lognormally distributed interferers," *IEEE Commun. Lett.*, vol. 8, no. 5, pp. 271–273, May. 2004.

[10] S. Zacks and C. P. Tsokos, "The distribution of sums of dependent log-normal variables," *Technical Report No 31*, Defense Technical Information Center, Apr. 1978.

[11] R. B. Leipnik, "On lognormal random variables-I: The characteristic function," *J. Aust. Math. Soc. Ser. B*, no. 3, pp. 327-347, 1991.

[12] P. Glasserman, "Generating Random Numbers and Random Variables" in *Monte Carlo Methods in Financial Engineering*, 1st ed. New York, Springer-Verlag, 2003, ch. 2, sec. 2.3.1, pp. 65.