SUMS OF LOGNORMALS

Daniel Dufresne
Centre for Actuarial Studies
University of Melbourne

Abstract

The problem of finding the distribution of sums of lognormally distributed random variables is discussed. References going back to the 1930's are given, as well as some possible solutions. A formula for the characteristic function of one lognormal is stated, and then the moments and distribution of the logarithm of sums of lognormals are considered.

1. Introduction

Finance: In financial mathematics, the most popular model for a stock's price is the lognormal distribution: if P is the stock price, then $\log P$ has a normal distribution.

Suppose there are two such stock prices, P_1 and P_2, and that they both have a lognormal distribution. What is the probability that the sum of the prices will be greater than y? Making the simplifying assumption that the stocks are stochastically independent, then this probability is (for $y > 0$):

$$\int_{\ell_1+\ell_2>y} \frac{d\ell_1\, d\ell_2}{2\pi\sigma_1\sigma_2\ell_1\ell_2} \exp\left\{-\frac{[(\log \ell_1)-\mu_1]^2}{2\sigma_1^2} - \frac{[(\log \ell_2)-\mu_2]^2}{2\sigma_2^2}\right\}.$$

This integral can be evaluated numerically, but it would seem that nothing can be done explicitly for the distribution of the sum of lognormals. This is one of the most surprising facts of elementary probability theory: almost nothing is known of the distribution of the sum of lognormals, although the lognormal distribution is a simple transformation of the very pleasant normal distribution.

Option pricing: Same problem with "Asian" and "basket" options, which involve sums of two or more lognormals.

Actuarial science: Individual claims are often well represented by a lognormal distribution; what is the distribution of total claims?

Engineering: The oldest and widest literature on the sum of lognormals is in engineering. Amplitudes of signals are modelled as lognormals. In telecommunications, engineers talk of the "power sum", or the logarithm of a sum of signals. This is of importance in wireless systems.

Other applications: The lognormal distribution has been used in many other fields: in economics, finance, reliability, biology, ecology, atmospheric sciences, geology. Even to model the duration of marriage (Aitchison & Brown, 1957).

Historical Summary

- Weber (1834). Said to be one of the first studies of the properties of the lognormal distributions (cited by Limpert *et al.* (2001)).
- Dixon, J.T. (1932, unpublished, cited by Marlow (1967)): method for approximating the sum of lognormals
- Wilkinson, R.I. (1934, Bell Telephone Labs, unpublished, cited by Marlow (1967)): possibly the first to use the "lognormal approximation":

$$\text{if} \quad S = \sum e^{N_j} \quad \text{then} \quad \log S \approx N(\mu, \sigma^2).$$

No mathematical justification was apparently given.

- Fenton, L.F. (1960). Takes up Wilkinson's (1934) idea of a lognormal approximation for sums of lognormals based on moment-matching, henceforth called the *Fenton-Wilkinson approximation*. (Later used in pricing Asian and basket options.)
- Mitchell, R.L. (1968). Applies a modified Gram-Charlier series to approximate sums of lognormals. Pursued by Schleher (1977) and others. Similar formulas ("Edgeworth series") used in finance by Jarrow & Rudd (1982) and Turnbull & Wakeman (1991). Resulting series has not been shown to converge.
- Leipnik, R.B. (1991). Derives an exact integral expression for the characteristic function of the lognormal distribution.
- Dufresne, D. (2004). Proves that in the limit as volatilities (σ_j) tend to 0 a sum of lognormals tends both to a lognormal and to a normal, depending on the normalization used.
- Wu *et al.* (2005). Sum of lognormals approximated by a single lognormal, based on two approximate values of Laplace transform. Used in wireless systems.

The only exact results known to date are, as far as I know:

1. Convolution integrals such as:

$$f_{L_1+L_2}(w) \;=\; \int_0^w \frac{d\ell}{2\pi\sigma_1\sigma_2\ell(w-\ell)} e^{-\frac{1}{2\sigma_1^2}(\log\,\ell-\mu_1)^2-\frac{1}{2\sigma_2^2}(\log(w-\ell)-\mu_2)^2}$$

(or $(n-1)$-fold integrals for the sum of n lognormals). The case $n=2$ is not bad numerically, but $n \geq 3$ is a problem.

2. Series for $\mathsf{E}(e^{N_1}+\cdots+e^{N_k})^{-r}$.

3. Series for $\mathsf{E}\log(e^{N_1}+\cdots+e^{N_k})$ and $\mathsf{E}[\log(e^{N_1}+\cdots+e^{N_k})]^2$.

4. Epression for the characteristic function of a single lognormal (Leipnik, 1991).

2. The characteristic function of the lognormal distribution

If $N \sim \mathbf{N}(0,1)$, then $L = e^{\sigma N} \sim \mathbf{Lognormal}(0,\sigma^2)$. Define

$$H(t) \;=\; \mathsf{E}\, e^{ie^t L} \;=\; \mathsf{E}\, e^{ie^t e^{\sigma N}}, \qquad t \in \mathbb{R}.$$

Then:

$$\begin{aligned} H'(t) \;&=\; \mathsf{E}\, ie^{t+\sigma N} e^{ie^t e^{\sigma N}} \\ &=\; ie^{t+\frac{\sigma^2}{2}}\, \mathsf{E}\, e^{\sigma N-\frac{\sigma^2}{2}} e^{ie^t e^{\sigma N}} \\ &=\; ie^{t+\frac{\sigma^2}{2}}\, \mathsf{E}\, e^{ie^t e^{\sigma(N+\sigma)}} \qquad \text{(Cameron-Martin)} \\ \Rightarrow \qquad H'(t) &=\; ie^{t+\frac{\sigma^2}{2}} H(t+\sigma^2). \end{aligned}$$

This is a delay-differential equation, and Leipnik (1991) uses it as a starting point to find an expression for the characteristic function of the lognormal distribution. His derivation uses de Bruijn's method for delay-differential equations. After some algebra, Leipnik's result is:

"if $L \sim \mathbf{Lognormal}(0,\sigma^2)$, then ($u \in \mathbb{R}-\{0\}$, $0<c<1$):

$$\mathsf{E}e^{iuL} \;=\; \frac{1}{2\pi}\int_{c-i\infty}^{c+i\infty} dz\, e^{\frac{\sigma^2}{2}z^2-z(\log\,u+\pi i/2)} \sin(\pi z)\Gamma(z)." \tag{1}$$

Proceeding differently, this author gets a slightly different expression: if $L_{\mu,\sigma^2} \sim \mathbf{Lognormal}(\mu,\sigma^2)$, then

$$\mathsf{E}\, e^{iuL_{\mu,\sigma^2}} \;=\; \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} dz\, e^{-\left(\log|u|+\mu-\frac{\pi i}{2}\mathrm{sgn}(u)\right)z+\frac{\sigma^2 z^2}{2}}\,\Gamma(z). \tag{2}$$

Here $c > 0$ is arbitrary, and the integration path is to the right of the origin. It has not been possible to reconcile those two expressions. Numerically, (2) does agree with the direct integral

$$\int_0^\infty \frac{dy}{\sigma\sqrt{2\pi}y}\exp\left[iuy - \frac{1}{2\sigma^2}(\mu-\log y)^2\right]$$

computed using Mathematica, for many values of μ and σ^2. However, in all those cases formula (1) gave vastly incorrect numbers. For instance, if $\mu = 0$, $\sigma = 1$, then the direct formula yields (using Mathematica)

$$0.340301 + 0.507190i$$

to six decimal places. Formula (2) yields the same, $0.340301+0.507190i$. Formula (1) yields $13.2399+19.8899i$. Observe that the latter violates the condition that the characteristic function has norm at most equal to 1. However, this is not a rigorous proof that (1) is incorrect, because I could have used incorrect Mathematica code, nor does itprove that (2) is correct.

Holgate (1989) finds approximations for the characteristic function of the lognormal

3. Moments of the log of the sum of lognormals

The logonormal approximation naturally leads to the problem of finding the first two moments of the logarithm of such a sum. Apart from the direct multiple integral there is no simple, explicit expression for those moments. Series are known in some simple cases. What follows is in the literature (see Crow & Shimizu, 1988 for references). Consider two independent normal variables $X_j \sim \mathbf{N}(0,\sigma^2)$, $j = 1,2$, and let $Y_1 = \min(X_1,X_2), Y_2 = \max(X_1,X_2)$. Then

$$e^{X_1} + e^{X_2} \;=\; e^{Y_2}(1+e^{Y_1-Y_2})$$

and so

$$\mathsf{E}[\log(e^{X_1}+e^{X_2})]^k \;=\; \mathsf{E}[Y_2 + \log(1+e^{Y_1-Y_2})]^k, \qquad k = 1,2,\ldots.$$

The appearance of the order statitics of the normal vector yields some simplification, especially for the first moment. In that case the problem reduces to computing two single integrals. The first one has a well-known explicit expression, while the second integral may be expressed as a series. Letting $N \sim \mathbf{N}(0,1)$, one finds:

$$\mathsf{E}Y_2 \;=\; -\mathsf{E}Y_1; \qquad \mathsf{E}(Y_2-Y_1) \;=\; 2\mathsf{E}Y_2 \;=\; \mathsf{E}|X_1-X_2| \;=\; \mathsf{E}(\sigma\sqrt{2}|N|) \;=\; \frac{2\sigma}{\sqrt{\pi}},$$

and thus $\mathsf{E}Y_2 = \sigma/\sqrt{\pi}$. Next,

$$\mathsf{E}\log(1+e^{Y_1-Y_2}) \;=\; \mathsf{E}\log(1+e^{-|X_1-X_2|}) \;=\; \mathsf{E}\log(1+e^{-\sigma\sqrt{2}|N|}),$$

which can be expanded using the Taylor expansion about 0 of $\log(1+z)$, yielding

$$\mathsf{E}\log(1+e^{Y_1-Y_2}) \;=\; \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\mathsf{E}e^{-n\sigma\sqrt{2}|N|} \;=\; 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}e^{n^2\sigma^2}\Phi(-n\sigma\sqrt{2}),$$

where $\Phi(x) \;=\; \int_{-\infty}^{x}\frac{dy}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$. Finally,

$$\mathsf{E}\log(e^{X_1}+e^{X_2}) \;=\; \frac{\sigma}{\sqrt{\pi}} + 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}e^{n^2\sigma^2}\Phi(-n\sigma\sqrt{2}).$$

A similar formula is also known for the second moment. This series converges slowly. This is because

$$e^{n^2\sigma^2}\Phi(-n\sigma\sqrt{2}) \;\sim\; \frac{1}{\sqrt{2\pi}}\left[\frac{1}{n\sigma\sqrt{2}} - \frac{1}{(n\sigma\sqrt{2})^3} + \cdots\right]$$

as $n \to \infty$ (Feller, 1968, p.193). It also converges faster for larger σ. For instance, the relative error of the ten-term truncated series is 6% when $\sigma = .01$, while it is .05% when $\sigma = 3$. Convergence is significantly improved using Richardson's extrapolation.

4. Density of the sum of two lognormals

It is possible to find series for the density of the logarithm of the sum of two lognormals. Each term is a polynomial times the normal density times the normal distribution function.

Example. Suppose $X_1, X_2 \sim \mathbf{N}(0,1)$ are independent, and let $Y = \log(e^{X_1} + e^{X_2})$. Figures 1 to 4 compare the exact density of Y (found by numerical integration using Mathematica) with n-term approximations. The truncated series is very well behaved, and gets closer to the exact density as n increases.

References

Aitchison, J., and Brown, J.A.C. (1957). *The Lognormal Distribution: With Special Reference to Its Use in Economics.* Cambridge University Press.

Crow, E.L., and Shimizu, K. (1988). *Lognormal Distributions: Theory and Applications*. New York, Marcel Dekker.

Dufresne, D. (2004). The log-normal approximation in financial and other applications. *Adv. Appl. Prob.* **36**: 747-773.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Vol.1.* (Third Edition.) Wiley, New York.

Fenton, L.F. (1960). The sum of log-normal probability distibutions in scattered transmission systems. *IRE Trans. Commun. Systems* **8**: 57-67.

Holgate, P. (1989). The lognormal characterisitic function. *Communications in Statistics – Theory and Methods* **18**: 4539-4548.

Jarrow, R., and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *J. of Financial Economics* **10**: 347-369.

Leipnik, R.B. (1991). On lognormal random variables: I – The characteristic function. *J. Australian Math. Soc. Ser. B* **32**: 327-347.

Limpert, E., Stahel, W.A., and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience* **51**: 341-352.

Mitchell, R.L. (1968). Permanence of the log-normal distribution. *J. Optical Society of America.* **58**: 1267-1272.

Marlow, N.A. (1967). A normal limit theorem for power sums of independent random variables. *Bell System Technical J.* **46**: 2081-2089.

Schleher (1977). Generalized Gram-Charlier series with applications to sums of log-normal variates. *IEEE Trans. Inform. Theory*: 275-280.

Turnbull, S., and Wakeman, L. (1991). A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis* **26**: 377-389.

Wu, J, Mehta, N., and Zhang, J. (2005). A flexible lognormal sum approximation. *Proceedings of IEEE Global Telecommunications Conference GLOBECOM 2005* **6**: 3413-3417.

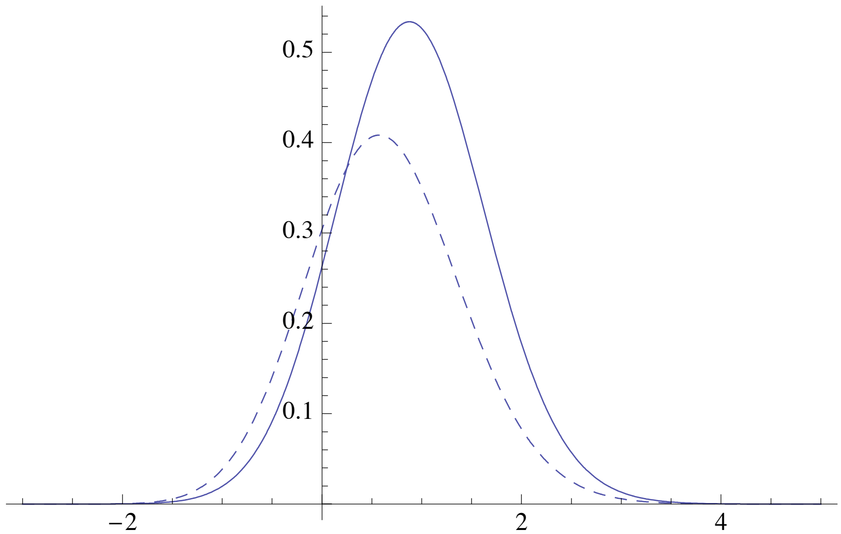

Figure 1. **Exact density of** $\log(e^{X_1} + e^{X_2})$ **and 1-term approximation (dotted line)**

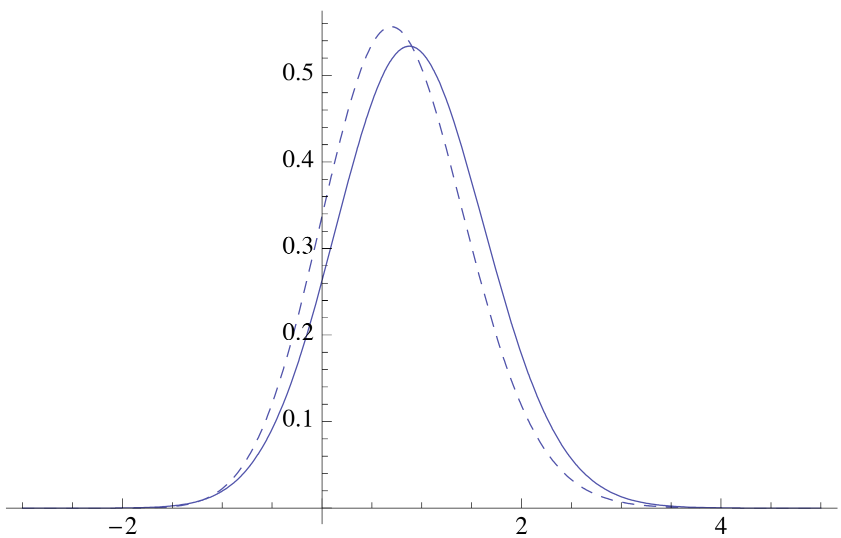

Figure 2. **Exact density of** $\log(e^{X_1} + e^{X_2})$ **and 3-term approximation (dotted line)**

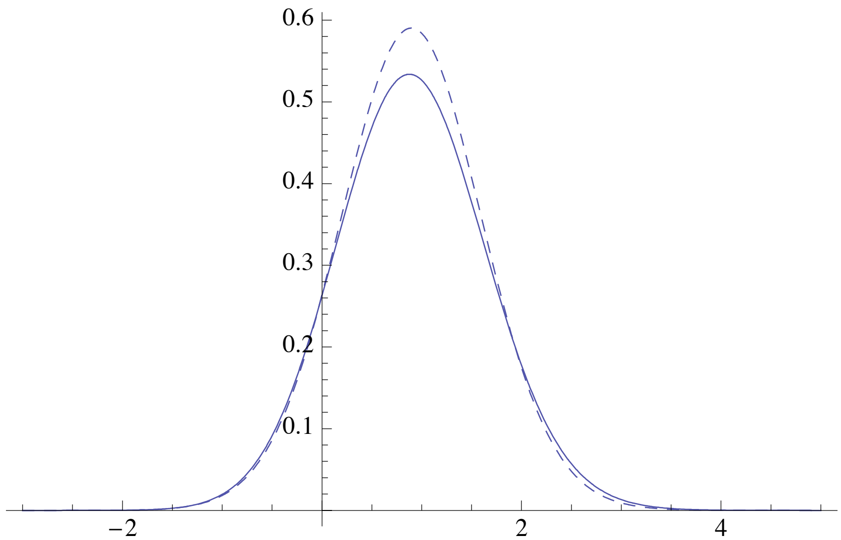

Figure 3. Exact density of $\log(e^{X_1} + e^{X_2})$ and 10-term approximation (dotted line)

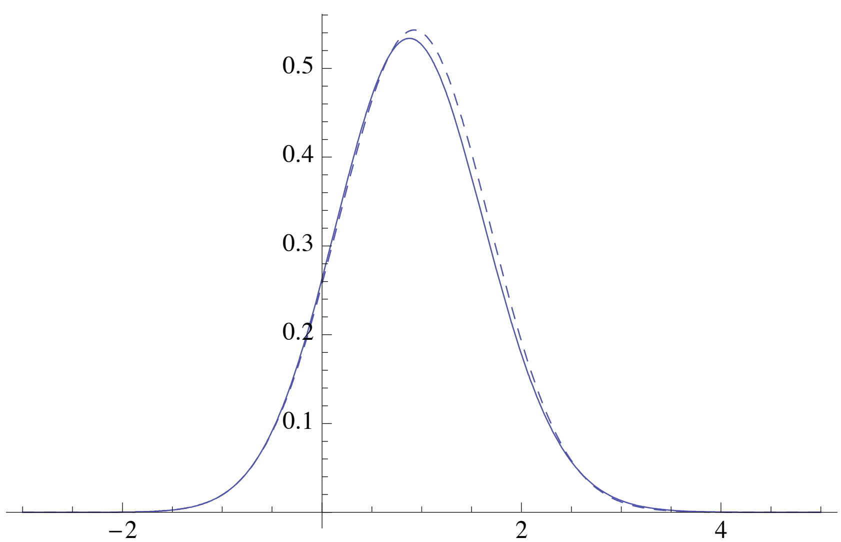

Figure 4. Exact density of $\log(e^{X_1} + e^{X_2})$ and 20-term approximation (dotted line)