

CUMULANTS AND EXPANSIONS FOR SUMS OF LOG-NORMALS

SARALEES NADARAJAH AND CHRISTOPHER S. WITHERS

RÉSUMÉ. Nous obtenons des expressions utiles pour les cumulants d’une loi log-normale. Nous exploitons ces formules pour obtenir les cumulants de la moyenne expérimentale de n variables aléatoires de lois lognormales, possiblement différentes. Enfin, un développement de type Edgeworth-Cornish-Fisher permet d’approcher la fonction de répartition de cette moyenne expérimentale.

ABSTRACT. We give an expression for the cumulants of a log-normal. This is used to obtain the cumulants of the mean of n log-normal random variables with different parameters. The Edgeworth-Cornish-Fisher expansions for the distribution of this mean then follow from those for the mean of an independent and identically distributed sample.

1. Introduction

The log-normal random variable can be defined as

$$(1) \quad X = \exp(\mu + \sigma N),$$

where $N \sim \mathcal{N}(0, 1)$, the unit normal. In Section 2, we give an expression for its general cumulants and list the first seven explicitly. (Until this time only the first four cumulants appeared to have been given; see [1, page 8], and [4, page 115] for the general central moment.) In Section 3, we apply this to obtain the Edgeworth-Cornish-Fisher expansions in powers of $n^{-1/2}$ for the mean of n log-normals, each with its own parameters. A different type of expansion for the density of a sum (and hence for the density of the mean) of n log-normals is given in [5, page 20].

2. The cumulants of the log-normal

The log-normal X defined by (1) is known to have the r -th moment

$$EX^r = \exp(r\mu + r^2\sigma^2/2).$$

Setting $w = \exp(\sigma^2)$, $a_r = w^{r^2/2}$ and $s = te^\mu$, this gives the formal expansion

$$E \exp(tX) \approx \sum_{r=0}^{\infty} a_r s^r / r! = 1 + S,$$

for some S .

This expansion is infinite for t real and positive but finite otherwise. The sign $\approx$ used above and throughout this note should not be confused with the “approximately equal to” sign. It is used in place of an equal sign because the terms on both sides of $\approx$ may not converge all the time. If they both converge then $\approx$ should be interpreted as being an equality sign.

By definition $S^j/j! = \sum_{r=j}^{\infty} B_{rj} s^r/r!$, where $B_{rj} = B_{rj}(a_1, a_2, \dots)$ are the partial exponential Bell polynomials tabled on pages 307-308 of [2]. They satisfy the recurrence relations

$$\binom{j}{j_1} B_{rj} = \sum_{r_1+r_2=r} \binom{r}{r_1} B_{r_1 j_1} B_{r_2 j_2}$$

for $j_1 + j_2 = j$. Thus, the cumulant generating function of X has the formal expansion

$$\log E \exp(tX) = \sum_{j=1}^{\infty} (-1)^{j-1} S^j/j \approx \sum_{r=1}^{\infty} k_r s^r/r!,$$

where

$$(2) \quad k_r = k_r(\sigma) = \sum_{j=1}^r (-1)^{j-1} (j-1)! B_{rj}.$$

That is, X has the r -th cumulant

$$\kappa_r(X) = \exp(r\mu) k_r.$$

Thus, $k_r = \kappa_r(X)$ when $\mu = 0$. The first seven cumulants $\{k_r\}$ in the Appendix are obtained from the formula (2). Factoring these expressions using MAPLE gives

$$(3) \quad k_r = w^{r/2} (w-1)^{r-1} p_r,$$

where $p_1 = p_2 = 1$ and, for $r > 2$, p_r is a polynomial in w of degree $N_r = \binom{r-1}{2}$:

$$\left\{ \begin{array}{l} p_1 = p_2 = 1, \\ p_3 = w + 2, \\ p_4 = w^3 + 3w^2 + 6w + 6, \\ p_5 = w^6 + 4w^5 + 10w^4 + 20w^3 + 30w^2 + 36w + 24, \\ p_6 = w^{10} + 5w^9 + 15w^8 + 35w^7 + 70w^6 + 120w^5 \\ \quad + 180w^4 + 240w^3 + 270w^2 + 240w + 120, \\ p_7 = w^{15} + 6w^{14} + 21w^{13} + 56w^{12} + 126w^{11} + 252w^{10} + 455w^9 \\ \quad + 750w^8 + 1140w^7 + 1610w^6 + 2100w^5 + 2520w^4 \\ \quad + 2730w^3 + 2520w^2 + 1800w + 720. \end{array} \right.$$

It appears that the first two terms of p_r are $w^{N_r} (1 + \binom{r}{2}/w)$ and the last two terms are $(r-2)!((r-2)w/2 + 1)$.

3. Edgeworth-Cornish-Fisher expansions for log-normals

Here, we give the Edgeworth-Cornish-Fisher-expansions for the mean

$$\bar{X} = n^{-1} \sum_{i=1}^n X_{in}$$

of independent random variables $\{X_{in}\}$ with finite cumulants $\{\kappa_r(X_{in})\}$. We then apply this to the case where $\{X_{in}\}$ are log-normals.

The r -th cumulant of $\bar{X}$ is $\kappa_r(\bar{X}) = k_{rn}n^{1-r}$, where $k_{rn} = n^{-1} \sum_{i=1}^n \kappa_r(X_{in})$ is the mean of the r -th order cumulants. Thus, $\bar{X}$ behaves like the mean of n independent and identically distributed random variables, each with cumulants $\{k_{rn}, r \geq 1\}$. Thus, provided that k_{2n} is bounded away from 0, and $\{k_{rn}\}$ is bounded as $n \rightarrow \infty$, the Edgeworth-Cornish-Fisher expansions for

$$P_n(x) = P(Y_n \leq x),$$

where

$$Y_n = (\bar{X} - k_{1n})(n/k_{2n})^{1/2},$$

are

$$(4) \quad P_n(x) \approx \Phi(x) - \phi(x) \sum_{r=1}^{\infty} n^{-r/2} h_r(x),$$

$$(5) \quad \Phi^{-1}(P_n(x)) \approx x - \sum_{r=1}^{\infty} n^{-r/2} f_r(x),$$

$$(6) \quad P_n^{-1}(\Phi(x)) \approx x + \sum_{r=1}^{\infty} n^{-r/2} g_r(x),$$

$$(7) \quad dP_n(x)/dx \approx \phi(x) \left(1 + \sum_{r=1}^{\infty} n^{-r/2} \bar{h}_r(x) \right).$$

Here, Φ and ϕ are the distribution and density of $N \sim \mathcal{N}(0, 1)$. Also $h_r(x)$ and $\bar{h}_r(x)$ have the form

$$\begin{cases} h_r(x) &= \sum_{i=1}^r c_{ri} He_{r-1+2i}(x), \\ \bar{h}_r(x) &= \sum_{i=1}^r c_{ri} He_{r+2i}(x), \\ He_r(x) &= \phi(x)^{-1} (-d/dx)^r \phi(x) \\ &= E(x + iN)^r \end{cases}$$

for $i = \sqrt{-1}$. The polynomial He_r is the r -th Hermite polynomial; see [7]. In fact, h_r , $\bar{h}_r$, f_r and g_r , for $1 \leq r \leq 4$, are polynomials obtainable by setting $a_{r,r-1} = k_{rn}$ and $a_{ri} = 0$ for $i \neq r-1$ in the expressions in [6]. In terms of the standardised cumulants

$K_r = k_{rn} k_{2n}^{-r/2}$, this gives

$$\left\{ \begin{array}{l} c_{11} = K_3/6, \\ c_{21} = K_4/24, \\ c_{22} = K_3^2/72, \\ c_{31} = K_5/120, \\ c_{32} = K_3 K_4/144, \\ c_{33} = K_3^3/1296, \\ c_{41} = K_6/720, \\ c_{42} = c'_{42} + c''_{42}, \end{array} \right.$$

where

$$\left\{ \begin{array}{l} c'_{42} = K_4^2/1152, \\ c''_{42} = K_3 K_5/720, \\ c_{43} = K_3^2 K_4/1728, \\ c_{44} = K_3^4/31104, \\ h_1(x) = f_1(x) = g_1(x) = c_{11}(x^2 - 1), \\ f_r(x) = c_{r1} H e_{r+1}(x) + \bar{f}_r(x), \\ g_r(x) = c_{r1} H e_{r+1}(x) + \bar{g}_r(x), \end{array} \right.$$

with

$$\left\{ \begin{array}{l} \bar{f}_2(x) = 2c_{22}(-4x^3 + 7x), \\ \bar{f}_3(x) = -c_{32}(11x^4 - 42x^2 + 15) + 2c_{33}(69x^4 - 187x^2 + 52), \\ \bar{f}_4(x) = -3c'_{42}(5x^5 - 32x^3 + 35x) - 2c''_{42}(7x^5 - 48x^3 + 51x) \\ \quad + 2c_{43}(111x^5 - 547x^3 + 456x) - 4c_{44}(948x^5 - 3628x^3 + 2473), \\ \bar{g}_2(x) = 2c_{22}(-2x^3 + 5x), \\ \bar{g}_3(x) = -6c_{32}(x^4 - 5x^2 + 2) + 4c_{33}(12x^4 - 53x^2 + 17), \\ \bar{g}_4(x) = -3c'_{42}(3x^5 - 24x^3 + 29x) - 4c''_{42}(2x^5 - 17x^3 + 21x) \\ \quad + 6c_{43}(14x^5 - 103x^3 + 107x) - 4c_{44}(232x^5 - 1688x^3 + 1511x). \end{array} \right.$$

The polynomials h_r , f_r , g_r and $\bar{h}_r$ have degree $3r - 1$, $r + 1$, $r + 1$ and $3r$ respectively and can be written in terms of $K_2, \dots, K_{r+2}$. The expansions (4)–(7) typically diverge, that is, terms typically decrease in magnitude as r increases and then increase. So, they must be truncated. In practice, they will only be calculated up to $r = 4$. If desired, methods of accelerating convergence, like that of Aitken, may be applied.

Now suppose that $X_{in} = \exp(\mu_{in} + \sigma_{in} N_i)$, with $\{N_i\}$ independent and identically distributed $\mathcal{N}(0, 1)$. Then the above expansions hold with

$$k_{rn} = n^{-1} \sum_{i=1}^n \exp(r\mu_{in}) k_r(\sigma_{in}),$$

where $k_r(\sigma) = k_r$, with k_r defined as in (3).

In [3] and [5], the authors gave classes of distributions with the same moments as the log-normal. The one given by Leipnik in [5] is a discrete distribution with support $\{aL^n \mid n \text{ is an integer}\}$, so is not covered by the Edgeworth-Cornish-Fisher expansions. However Heyde [3] gives a continuous distribution, namely that of $X = \exp(\mu + \sigma Y)$, where Y has density $\phi(y)(1 + \epsilon \sin(2\pi ky/\sigma))$ for ϵ in $(0, 1)$ and k a positive integer.

It could be that the Cornish-Fisher expansions only apply when the moments uniquely determine the distribution, or that the regularity conditions implicitly rule out this family of distributions. A more likely explanation is that both the log-normal and Heyde's distribution satisfy the Cornish-Fisher expansions viewed as (possibly divergent) asymptotic expansions in $n^{-1/2}$. That is, the error obtained in taking only the first r terms has magnitude $n^{-r/2}$ as n increases.

Appendix: The cumulants of the log-normal when $\mu = 0$

Here we derive the first seven values k_r of formula (2) in terms of $w = \exp(\sigma^2)$:

$$\left\{ \begin{array}{l} k_1 = B_{11} = a_1 = w^{.5}, \\ k_2 = B_{21} - B_{22} = a_2 - a_1^2 = w^2 - w, \\ k_3 = B_{31} - B_{32} + 2B_{33} = a_3 - 3a_1a_2 + 2a_1^3 \\ \quad = w^{1.5}(w^3 - 3w + 2), \\ k_4 = B_{41} - B_{42} + 2B_{42} + 2B_{43} - 6B_{44} \\ \quad = a_4 - 4a_1a_3 - 3a_2^2 + 12a_1^2a_2 - 6a_1^4 \\ \quad = w^2(w^6 - 4w^3 - 3w^2 + 12w - 6), \\ k_5 = a_5 - 5a_1a_4 - 10a_2a_3 + 20a_1^2a_3 + 30a_1a_2^2 \\ \quad = w^{2.5}(w^{10} - 5w^6 - 10w^4 + 20w^3 + 30w^2 - 60w + 24), \\ k_6 = a_6 - 6a_1a_5 - 15a_2a_4 - 10a_3^2 + 30a_1^2a_4 + 120a_1a_2a_3 + 30a_2^3 \\ \quad - 120a_1^3a_3 - 270a_1^2a_2^2 + 360a_1^4a_2 - 120a_1^6 \\ \quad = w^3(w^{15} - 6w^{10} - 15w^7 + 20w^6 + 120w^4 - 90w^3 \\ \quad - 270w^2 + 360w - 120), \\ k_7 = a_7 - 7a_1a_6 - 21a_2a_5 - 35a_3a_4 + 42a_1^2a_5 + 210a_1a_2a_4 \\ \quad + 140a_1a_3^2 + 210a_2^2a_3 - 210a_1^3a_4 - 1260a_1^2a_2a_3 - 630a_1a_2^3 \\ \quad + 24(35a_1^4a_3 + 105a_1^3a_2^2) - 2520a_1^5a_2 + 720a_1^7 \\ \quad = w^{3.5}(w^{21} - 7w^{15} - 21w^{11} + 42w^{10} - 35w^9 + 210w^7 - 70w^6 \\ \quad + 210w^5 - 1260w^4 + 210w^3 + 2520w^2 - 2520w + 720). \end{array} \right.$$

REFERENCES

- [1] J. Aitchison and J. A. C. Brown, *The lognormal distribution, with special reference to its uses in economics*, Cambridge University Press, 1957, xviii+176 pp.
- [2] L. Comtet, *Advanced combinatorics*, The art of finite and infinite expansions, D. Reidel Publishing Co., Dordrecht, 1974, xi+343 pp.
- [3] C. C. Heyde, *On a property of the lognormal distribution*, J. Roy. Statist. Soc. Ser. B **25** (1963), 392–393.
- [4] N. L. Johnson and S. Kotz, *Distributions in statistics. Continuous univariate distributions. 2*, Houghton Mifflin Co., Boston, Mass., 1970, xiii+306 pp.
- [5] R. B. Leipnik, *On lognormal random variables. I. The characteristic function*, J. Austral. Math. Soc. Ser. B **32** (1991), no. 3, 327–347.
- [6] C. S. Withers, *Asymptotic expansions for distributions and quantiles with power series cumulants.*, J. Roy. Statist. Soc. Ser. B **46** (1984), no. 3, 389–396.
- [7] C. S. Withers, *A simple expression for the multivariate Hermite polynomials*, Statist. Probab. Lett. **47** (2000), no. 2, 165–169.

S. NADARAJAH, SCHOOL OF MATHEMATICS, U. OF MANCHESTER, MANCHESTER, M60 1QD, UNITED KINGDOM
mbsssn2@manchester.ac.uk

C. S. WITHERS, APPLIED MATHEMATICS GROUP, INDUSTRIAL RESEARCH LIMITED, LOWER HUTT, NEW ZELAND
C.Withers@irl.cri.nz