LETTER TO THE EDITOR

Dear Editor,

Remarks on the Luria–Delbrück distribution

Ma et al. (1992) obtain some asymptotic properties of the Luria–Delbrück (LD) distribution, i.e. the discrete law whose probability generating function (p.g.f.) is

$$(1) \qquad F(s) = \sum p_n s^n = (1 - s)^{m(1-s)/s}$$

where, for our purposes, m is a positive constant. Properties of this distribution are very useful in helping to resolve some controversial issues about bacterial mutagenesis. By writing $F(s)$ as a composite function and using successive differentiation, Ma et al. derive the computationally efficient recursion

$$(2) \qquad p_0 = e^{-m}, \quad \text{and for } n \geqq 1, \quad p_n = (m/n) \sum_{i=0}^{n-1} \frac{p_i}{n - i + 1}.$$

A formally equivalent, but much easier, approach is just to logarithmically differentiate (1) and equate coefficients of powers of s. Probabilists and statisticians will recognise (2) as an instance of the general identity connecting the masses of a compound Poisson law with p.g.f. $F(s) = \exp[-m(1 - \phi(s))]$ and the masses of the p.g.f. ϕ. Here

$$(3) \qquad \phi(s) = 1 + [(1 - s)/s]\log(1 - s),$$

and it follows that the LD law is infinitely divisible. The general identity referred to here seems to arise first in the statistical literature of the early 1960s. For references see Ord (1972), p. 130, and Douglas (1980), p. 86, who cites this relation as '... perhaps the most generally useful formula ...'. In a slightly different form, the identity is a quite well known criterion for infinite divisibility of p_n; see van Harn (1978), pp. 15–18, for discussion and references.

Next, Ma et al. seek the asymptotic behaviour of p_n as $n \rightarrow \infty$ using an elementary theory of discrete convolution powers. The masses corresponding to (3) are $f_0 = 0$ and $f_n = 1/n(n + 1)$, $n \geqq 1$. Letting f_n^{*k} denote the masses of the p.g.f. $(\phi(s))^k$, $(k \in \mathbb{N})$, Ma et al. give arguments supporting the asymptotic estimates

$$(4) \qquad f_n^{*k} \sim 2^{k-1} c_k / n^2 \quad (n \rightarrow \infty), \qquad c_2 \approx 1, \qquad c_3 \approx 0.75 \quad \text{and} \quad c_4 \approx 0.5;$$

the approximations were found by numerical means. On the basis of a representation which we express as

(5)
$$p_n = \exp(-m) \sum_{k=1}^{n} (m^k/k)\, f_n^{*k},$$

they assert that

(6)
$$p_n \approx c/n^2$$

where (correcting what appears to be an error)

(7)
$$c = \exp(-m) \sum_{k=1}^{n} c_k 2^{k-1} m^k/k.$$

Ma et al. do not separate m and ϕ as we do but work with $\alpha(s) = m\phi(s)$ and its weights $a_n = mf_n$. Hence the terms in their version of (5) are a_n^{*k}/k, and they derive (5) analytically (and with a small error). Again, probabilists will recognise (5) as the familiar expression for a compound Poisson law — the sum is finite here because $\phi(0) = 0$.

Observe that the coefficients $2^{k-1}c_k$ are approximately 2, 3 and 4 for $k = 2$, 3 and 4, respectively. If this is a general pattern then from (7) we have $c \approx m$. All of this should look familiar since it springs fully developed from the long-known and rich theory of convolution powers of subexponential laws; see Chover et al. (1973) and Embrechts and Hawkes (1982) for the discrete case we need here.

Indeed, a tabulated power series representation for $[\log(1-s)]^2$ (see Jolly (1961), p. 21) gives the exact representation

$$f_n^{*2} = \frac{2}{n(n+1)} + \frac{4H(n) - 3 - 2/n}{n(n+1)(n+2)}$$

where $H(n) = \sum_{j=1}^{n} 1/j$ is the nth harmonic number. Consequently, as $n \to \infty$,

(8)
$$f_n^{*2} \sim 2/n^2 \quad \text{and} \quad f_{n+1} \sim f_n,$$

and it now follows from subexponential theory (Chover et al. (1973), Lemma 5) that

$$f_n^{*k} \sim k/n^2 \quad (n \to \infty).$$

Hence the above approximations for the c_k are exact. Again from subexponential theory (Embrechts and Hawkes (1982), Theorem 1), (8) is equivalent to

(9)
$$p_n \sim m/n^2 \quad \text{and} \quad p_{n+1} \sim p_n.$$

This 'remarkable' (Ma et al. (1992), p. 262) result is thus explained (and vastly generalised) through subexponential theory. Indeed, this theory applies to the more general model of Mandelbrot (1974) where (3) is replaced by

$$\phi(s) = \alpha s \int_0^1 \frac{u^\alpha}{1 - s(1-u)}\, du.$$

Consequently $f_n = \alpha B(\alpha + 1, n)$ where $B(\cdot, \cdot)$ is the beta function. The parameter α is interpreted as the ratio of growth rates of normal and mutant bacteria. These formulae are equivalent to those listed by Mandelbrot (1974) on p. 439. It follows that $\{f_n\}$ is

strictly decreasing and $f_n \sim \alpha\Gamma(\alpha + 1)n^{-1-\alpha}$ (which contradicts the expression at the top of p. 440 in Mandelbrot (1974) which obviously is wrong since it implies that $Y(1)$ has a finite mean). Hence the conditions of Chover et al. (1973), p. 259, are fulfilled, giving $f_n^{*2} \sim 2f_n$ and hence $p_n \sim mf_n$.

Final comments

1. The derivation of (4) for $k = 2$ in Appendix B of Ma et al. (1992) yields the precise value $c_2 = 1$, but this seems to have been lost in the main text.

2. Ma et al. observe that it has long been known that the moments of $\{p_n\}$ are infinite. Certainly this is explicitly stated by Bailey (1964), p. 129 and Bartlett (1978), p. 134. But Ma et al. make assertions about rates of divergence of the moments. Presumably they (should) have in mind the asymptotic behaviour of truncated moments

$$\mu_k(n) = \sum_{j=1}^{n} j^k p_j$$

as $n \to \infty$. It is clear that

$$\mu_1(n) \sim m \log n$$

and for $k \geqq 1$,

$$\mu_{k+1}(n) \sim (m/k)n^k.$$

3. In the context of bacterial mutation the parameter m is an exponentially increasing function of time; it is the rate of formation of mutant clones. If Z_m is the mutant population size, whose p.g.f. is $F(s)$, then it is clear that $Z_m \to_p \infty$ as $m \to \infty$. Elementary calculation shows that norming and centering gives a non-trivial limit law: as $m \to \infty$,

$$(Z_m/m) - \log m \Rightarrow S,$$

a stable random variable whose characteristic function is

$$\exp[-|\theta|(\pi/2 + (\text{sgn}(\theta)))\log|\theta|],$$

but whose distribution does not have an elementary closed form. However, this result shows that the mutant population size normed by clonal formation rate grows linearly in time. In particular we have the cruder result $Z_m/(m \log m) \to_p 1$. Thus we have a pedagogically interesting example of a family of random variables (the Z's) having accessible distributional properties (via (2)) but having an intractable limit law. Mandelbrot (1974) asserts something like the above limit theorem (see p. 442) where he gives what seems to be a bilateral Laplace transform of the density of the limit law. But this transform does not exist! Also, Mandelbrot works with a random sum which has the same law as the mutant population size, but differs from it.

4. Sarkar et al. (1992) looked at an extension of the LD law due to Bartlett (1978) which allows for the finite number of cell divisions between seeding and harvesting the culture used in a fluctuation experiment. Generalising (1) we have $F_\alpha(s) = (1 - s + \alpha s)^{m(1-s)/s}$ which is a p.g.f. iff $0 \leqq \alpha < 1$. Sarkar et al. (1992) obtain a recursion

generalising (2) which can easily be derived by logarithmic differentiation, as suggested above. When $\alpha \geqq 0$ the asymptotic behaviour of the masses $p_n(\alpha)$ is quite different from the LD case: as $n \to \infty$,

$$p_n(\alpha) \sim (1 - \alpha)^n n^{\alpha m - 1}/\Gamma(\alpha m).$$

Hence the upper tail is proportional to that of a negative binomial law having parameters $1 - \alpha$ and αm. Note that this law also is infinitely divisible. I thank the referee for directing me to this reference.

5. The author is investigating more general mutation models.

Acknowledgement

References

BAILEY, N. T. J. (1964) *The Elements of Stochastic Processes*. Wiley, New York.

BARTLETT, M. S. (1978) *An Introduction to Stochastic Processes*, 3rd edn. Cambridge University Press.

CHOVER, J., NEY, P. AND WAINGER, S. (1973) Functions of probability measures. *J. D'Analyse Math.* **XXVI**, 255–302.

DOUGLAS, J. B. (1980) *Analysis with Standard and Contagious Distributions*. Intl. Co-operative Pub. House, Fairland, MD, USA.

EMBRECHTS, P. AND HAWKES, J. (1982) A limit theorem for tails of discrete infinitely divisible laws with applications to fluctuation theory. *J. Austral. Math. Soc. Series A* **32**, 412–422.

HARN, K. VAN (1978) *Classifying Infinitely Divisible Distributions by Functional Equations*. Math. Centrum Tract 103, Amsterdam.

JOLLY, L. B. W. (1961) *Summation of Series*, 2nd revised edn. Dover, New York.

MA, W. T., SANDRI, G. VH. AND SARKAR, S. (1992) Analysis of the Luria–Delbrück distribution using convolution powers. *J. Appl. Prob.* **29**, 255–267.

MANDELBROT, B. (1974) A population birth-mutation process, I: Explicit distributions for the number of mutants in an old culture of bacteria. *J. Appl. Prob.* **11**, 437–444.

ORD, J. K. (1972) *Families of Frequency Functions*. Griffin, London.

SARKAR, S., MA, W. T. AND SANDRI, G. VH. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173–179.

Yours sincerely,
ANTHONY G. PAKES

Department of Mathematics
University of Western Australia
Nedlands
WA 6009
Australia