The generalized logarithmic series distribution

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics and Computing Science

Memorandum COSOR 88-03

The generalized logarithmic series distribution

by

B.G. Hansen and E. Willekens

Eindhoven, Netherlands

January 1988

The generalized logarithmic series distribution

B.G. Hansen and E. Willekens
Department of Mathematics and Computing Science
Eindhoven University of Technology
Eindhoven, The Netherlands

ABSTRACT

It is shown that the generalized logarithmic series distribution is log-convex and hence infinitely divisible, and its Levy measure is determined asymptotically up to second order. An application to risk theory is given.

1. Introduction

The generalized logarithmic series distribution (GLSD) is defined by the sequence

$$p_n := \frac{1}{\beta n}\begin{bmatrix}\beta n \\ n\end{bmatrix}\theta^n(1-\theta)^{\beta n-n}/\{-\log(1-\theta)\}, \quad (n = 1,2,3,\ldots), \qquad (1.1)$$

with $\beta \geqslant 1$ and $0<\theta<\beta^{-1}$. Here $[{x \atop y}]$ denotes the generalized binomial coefficients, i.e.

$$\begin{bmatrix}x \\ y\end{bmatrix} = \frac{\Gamma(x+1)}{\Gamma(x-y+1)\Gamma(y+1)} .$$

The GLSD was obtained by Jain and Gupta (1973) through Lagrange's expansion of the ordinary logarithmic series distribution (which is (1.1) with $\beta = 1$). It is also possible to get (1.1) as a limiting form of the zero truncated generalized negative binomial distribution, see Jain (1975). Famoye (1987) showed in a recent "letter" that the GLSD is unimodal but not strongly unimodal, or equivalently, not log-concave. In the next section we prove the stronger result that the GLSD is strictly log-convex. It then follows from Steutel (1970), Theorem 4.2.2, that the GLSD is

(i) infinitely divisible

and

(ii) decreasing and hence unimodal with mode at n=1.

Using the Levy-Khintchine representation for infinitely divisible sequences, we establish in section 3 the asymptotic behaviour of the Levy measure of $(p_n)_n$. Finally in section 4, an application to risk theory is given.

Before stating the results, we note that the GLSD may be used as a model for lifetime distributions. To see this, consider the inverse Gaussian distribution (IGD) which is defined by the density

$$f(x) = (2\pi x^3)^{-1/2}\,\lambda e^{\lambda\gamma} \exp\{-1/2(\gamma^2 x + \lambda^2 x^{-1})\} \quad , x > 0, \gamma > 0, \lambda > 0. \tag{1.2}$$

Now if X is a random variable with an IGD(λ,γ) and if $[X]$ denotes the integer part of X, then it is not hard to show that

$$\mathbb{P}([X]=n) \sim p_n \quad (n \to \infty), \tag{1.3}$$

where p_n is GLSD(θ,β) and

$$\gamma^2 = 2\log\left\{\left(\frac{\beta-1}{\beta(1-\theta)}\right)^{\beta-1}\frac{1}{\beta\theta}\right\},$$

$$\lambda e^{\lambda\gamma}\, 2\gamma^{-2}\,\{1-e^{-1/2\gamma^2}\} = \{-\log(1-\theta)\}^{-1}\,\{\beta(\beta-1)\}^{-1/2}.$$

This shows that the GLSD can be seen, at least asymptotically, as a discretized version of the IGD. The fact that $f(x)$ in (1.2) occurs as the first hitting density of the level λ in a Brownian motion with drift $-\gamma$, suggests its potential use as a lifetime distribution or as a distribution between two renewal points. From (1.3), we may expect that the GLSD is well suited to express similar quantities in discrete models.

2. Log-convexity of GLSD

Theorem 2.1 The GLSD is strictly log-convex.

Proof. It follows from (1.1) that for $n = 2,3,4,...$

$$\frac{p_{n+1}}{p_n}\frac{p_{n-1}}{p_n} = \frac{n}{(n+1)^2}\frac{n^2}{n-1}\frac{\Gamma(\beta(n+1)+1)\Gamma(\beta(n-1)+1)\Gamma((\beta-1)n+1)^2}{\Gamma((\beta-1)(n+1)+1)\Gamma((\beta-1)(n-1)+1)\Gamma(\beta n+1)^2}$$

$$= \frac{n}{n+1}\prod_{k=1}^{n}((n+1)\beta-k)\prod_{k=1}^{n-2}((n-1)\beta-k)\prod_{k=1}^{n-1}(n\beta-k)^{-2}. \tag{2.1}$$

The proof is finished if we show that (2.1) exceeds one for every $n \geqslant 2$. We therefore show that (2.1), considered as a function in β, takes its minimum in $\beta=1$. Observe that

$$\frac{\partial}{\partial\beta}(\log\frac{p_{n+1}}{p_n}\frac{p_{n-1}}{p_n}) = \sum_{k=1}^{n}\frac{(n+1)}{(n+1)\beta-k} + \sum_{k=1}^{n-2}\frac{(n-1)}{(n-1)\beta-k} - 2\sum_{k=1}^{n-1}\frac{n}{n\beta-k}. \quad (2.2)$$

To prove that (2.2) is positive, it is necessary and sufficient to show that

$$\sum_{k=1}^{n}\frac{(n+1)}{(n+1)\beta-k} = (n+1)\sum_{k=1}^{n}\int_0^\infty \exp\{-((n+1)\beta-k)y\}\,dy$$

$$= \int_0^\infty e^{-\beta x}\,(e^x - e^{x/(n+1)})\,(e^{x/(n+1)}-1)^{-1}\,dx \quad (2.3)$$

is convex in n. This is done by substituting the integer variable $n+1$ by the real variable t in (2.3). Evaluating the second derivative with respect to t easily yields the convexity of (2.3). Hence (2.1) obtains its minimum for $\beta \leqslant 1$, so that

$$\frac{p_{n+1}p_{n-1}}{p_n^2} \geqslant \frac{\theta^{n+1}}{n+1}\frac{\theta^{n-1}}{n-1}\frac{n^2}{\theta^{2n}} = \frac{n^2}{n^2-1} > 1.$$

This completes the proof.

Since $(p_n)_n$ is log-convex, it follows from Steutel (1970) that $(p_n)_n$ is infinitely divisible. The Levy-Khintchine representation of the GLSD takes the form

$$\hat{p}(z) := \sum_{n=1}^{\infty} p_n z^n = z\exp\{-\lambda(1-\sum_{j=1}^{\infty}\alpha_j z^j)\} := z\exp\{-\lambda(1-\hat{\alpha}(z))\}, \quad (2.4)$$

where $\lambda = -\log p_1 > 0$ and $(\alpha_j)_j$ is a probability measure known as the Levy measure of $(p_n)_n$.

We are interested in finding an expression for α_n, $n \geqslant 1$. Obviously, by solving (2.4) we get that

$$\alpha_n = -\lambda^{-1}\sum_{k=1}^{\infty}(-1)^k k^{-1} e^{\lambda k} p_{n+1}^{*k}, \quad n = 1,2,3,\ldots .$$

Here $(p_n^{*k})_n$ denotes the k-th convolution power of the sequence $(p_n)_n$. Although the identity is exact, it is useless in practice since we cannot compute the right hand side. We will therefore concentrate on asymptotic expressions for α_n as $n \to \infty$. The following theorem of Embrechts and Hawkes (1982) illustrates that such expressions are closely related to the convolution behavior of p_n as $n \to \infty$.

Theorem 2.2 (Embrechts and Hawkes). Let $(p_n)_n$ be infinitely divisible such that (2.4) holds and suppose $R \geqslant 1$. the following relations are equivalent.

(i) $\hat{\alpha}(R) < \infty$, $\alpha_n^{*2} \sim 2\hat{\alpha}(R)\alpha_n$ and $R\alpha_{n+1} \sim \alpha_n$ $(n \to \infty)$.

(ii) $\hat{p}(R) < \infty$, $p_n^{*2} \sim 2\hat{p}(R)p_n$ and $Rp_{n+1} \sim p_n$ $(n \to \infty)$.

(iii) $p_n \sim \lambda\hat{p}(R)\alpha_n$ and $R\alpha_{n+1} \sim \alpha_n$ $(n \to \infty)$.

We apply this theorem to the GLSD. This is done in the next section.

3. The Levy measure of GLSD

It follows from Stirlings formula (see Abramowitz and Stegun, 1965) that for p_n in (1.1),

$$p_n \sim \frac{1}{(2\pi)^{1/2} n^{3/2}} \frac{1}{(\beta(\beta-1))^{1/2}(-\log(1-\theta))} \left[\left(\frac{\beta(1-\theta)}{\beta-1}\right)^{\beta-1} \beta\theta \right]^n \quad (n \to \infty). \qquad (3.1)$$

Chover et al (1973) showed that if p_n has an expansion as in (3.1), it satisfies Theorem 2.2 with

$$R = \frac{1}{\beta\theta} \left[\frac{\beta-1}{\beta(1-\theta)} \right]^{\beta-1} \geq 1 .$$

Hence as an immediate consequence of Theorem 2.2 we have

Corollary 3.1 Let p_n be given as in (1.1) and let $(\alpha_n)_n$ be defined as in (2.4). Then

$$\alpha_n \sim \lambda^{-1} \frac{\log(1-\theta)}{\log(1-\beta^{-1})} p_n \quad (n \to \infty). \qquad (3.2)$$

Equation (3.2) provides an easy relation for α_n as $n \to \infty$. In order to estimate the accuracy in (3.2) we now estimate

$$p_n - \lambda \frac{\log(1-\beta^{-1})}{\log(1-\theta)} \alpha_n \quad (n \to \infty).$$

This result is stated in

Theorem 3.2 Let $(p_n)_n$ be GLSD and let $(\alpha_n)_n$ be given as in (2.4). Then

$$\Delta_n := \frac{p_n}{\alpha_n} - \lambda \frac{\log(1-\beta^{-1})}{\log(1-\theta)} = o(np_n) \quad (n \to \infty)$$

and np_n is the right rate of convergence in the sense that for any sequence $r_n \to \infty$, $r_n \Delta_n (np_n)^{-1} \to \infty$ as $n \to \infty$.

Proof. Defining $q_n := R^n p_n (\hat{p}(R))^{-1}$,$n = 1,2,3,\ldots$ it follows from (2.4) that

$$\hat{q}(z) = Rz(\hat{p}(R))^{-1}\exp\{-\lambda(1-\sum_{j=1}^{\infty}\alpha_j(Rz)^j)\}. \tag{3.3}$$

Hence $(q_n)_n$ is infinitely divisible with corresponding Levy measure $\beta_n := \alpha_n R^n$, $n=1,2,\ldots$. Let $\hat{c}(z)$ be the power series defined by

$$\hat{c}(z) = (1-z^{-1}\hat{q}(z))\, z(\hat{q}(z))^{-1} = \sum_{n=0}^{\infty} c_n z^n \ . \tag{3.4}$$

As in Embrechts and Hawkes (1982), we have that $\sum c_n$ is absolutely convergent, $\sum c_n = 0$ and $c_n \sim -q_n$, $(n \to \infty)$. It follows from (3.3) and (3.4) that

$$\lambda(\hat{\beta}(z))' = (z^{-1}\hat{q}(z))' + \hat{c}(z)(z^{-1}\hat{q}(z))' ,$$

so that

$$\lambda n\, \beta_n - nq_{n+1} = \sum_{j=0}^{n-1}(n-j)q_{n-j+1}c_j \quad , n=1,2,3,\ldots . \tag{3.5}$$

Choose $0<\delta<\epsilon<1$. We write the right hand side of (3.5) as

$$\begin{aligned}
& -\sum_{j=0}^{n-1} q_{n-j+1}\, c_j \\
& + \sum_{j=0}^{[n\delta]}\{(n-j+1)q_{n-j+1}-nq_n\}\, c_j \\
& + \sum_{j=[n\delta]+1}^{[n\epsilon]}\{(n-j+1)q_{n-j+1}-nq_n\}\, c_j \\
& + \sum_{j=[n\epsilon]+1}^{n-1}(n-j+1)q_{n-j+1}\, c_j \\
& + nq_n \sum_{j=0}^{[n\epsilon]} c_j \\
& := \ -(\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV}) + (\mathrm{V}).
\end{aligned}$$

By definition of q_n and (3.1), we have that $q_n \sim \mu n^{-3/2}$ $(n \to \infty)$ where μ may be found explicitly from (3.1). Hence the sequence $(q_n)_n$, and therefore also the sequence $-c_n$, is regularly varying with index $-3/2$ (see e.g. Seneta, 1976).

We now estimate the terms (I) to (V). Since $\sum c_j = 0$, we have that

$$(\mathrm{V}) = -nq_n \sum_{j=[n\epsilon]+1}^{\infty} c_j$$

and from Karamata's theorem (see Seneta, 1976)

$$(\mathrm{V}) \sim (nq_n)^2 2\epsilon^{-1/2} \ (n \to \infty).$$

As to (IV), it follows from the regular variation of c_n and the local uniformity (see Seneta, 1976) that there exists $1>\nu>0$ such that for $n>n_0(\nu)$,

$$-(1+\nu)\epsilon^{-3/2}q_n \sum_{j=2}^{n-[n\epsilon]} jq_j \leqslant \text{(IV)} \leqslant -(1-\nu)q_n \sum_{j=2}^{n-[n\epsilon]} jq_j \quad .$$

Furthermore by Karamata's theorem,

$$\sum_{j=2}^{n-[n\epsilon]} jq_j \sim 2(1-\epsilon)^{1/2} n^2 q_n \quad (n \to \infty).$$

We write (III) as

$$\text{(III)} = \int_{j=[n\delta]+1}^{[n\epsilon]} \{(n-[s]+1)q_{n-[s]+1} - nq_n\}\, c_{[s]+1}\, ds \ .$$

Then again from the regular behaviour of q_n and c_n,

$$\text{(III)} \sim n^2 q_n c_n \int_\delta^\epsilon ((1-s)^{-1/2}-1)\, s^{-3/2}\, ds \ .$$

A similar treatment for (II) gives that

$$\text{(II)} \sim n^2 q_n \int_0^\delta ((1-s)^{-1/2}-1)\, c_{[ns]+1}\, ds$$

$$= n^2 q_n \int_0^\delta ((1-u)^{-3/2} \{ \int_u^\delta c_{[ns]+1} ds \}\, du \ .$$

Clearly for n large enough,

$$|\text{(II)}| \leqslant n^2 q_n (1-\delta)^{-3/2} \int_0^\delta s\, |c_{[ns]+1}|\, ds \ ,$$

while by the regular variation of $|c_n|$,

$$\int_0^\delta s\, |c_{[ns]+1}|\, ds \sim 2\delta^{1/2} |c_n| \ .$$

Finally, as to (I), it is easily seen by splitting up the summation from $j=0$ to $[n/2]$ and from $[n/2]+1$ to $n-1$, that for n sufficiently large

$$|\text{(I)}| \leqslant cq_n$$

where $c > 0$ is some constant.

Combining the above estimates and letting $n \to \infty$, $\epsilon \uparrow 1$ and $\delta \downarrow 0$ then gives that

$$\frac{\lambda n \beta_n - nq_{n+1}}{(nq_n)^2} \to -\int_0^1 \{(1-s)^{-1/2}-1\}\, s^{-3/2}\, ds + 2 \ , \qquad (n \to \infty) \ . \tag{3.6}$$

The theorem now follows since $q_n - q_{n+1} = o(nq_n^2)$ as $n \to \infty$, and since the right hand side in (3.6) is 0 (see e.g. Omey and Willekens, 1986).

4. Application

The context of our application is risk theory. Suppose that an insurance company has a portfolio in which claims X_i ,$(i=1,2,...)$ occur at consecutive timepoints Y_i ,$(i=1,2,...)$. We assume that $(Y_i)_i$ are negative exponential (λ) and that $(X_i)_i$ is a sequence of independent random variables with the same distribution function F, and independent of $(Y_i)_i$. A quantity of great interest to the insurance company is the total claim size distribution up to time t, i.e.

$$F_t(x) := \mathbb{P}\Big(\sum_{i=1}^{N(t)} X_i \leqslant x\Big),$$

where $N(t)=\max\{n : \sum_1^n Y_i \leqslant t\}$. Since $\mathbb{E}Y_i=\lambda^{-1}$ it follows by standard methods that

$$F_t(x)=e^{-\lambda t}\sum_{k=0}^{\infty}(\lambda t)^k (k!)^{-1} F^{*k}(x), \quad x \geqslant 0. \tag{4.1}$$

In order to analyze the portfolio or to make predictions, we are interested in the individual claimsize distribution F. However, it is often the case in practise that the administration of an insurance company only provides aggregate information, or that one has an uneasy feeling about the data on the tails of the claims distribution (see Kaas (1987)). Therefore one often concentrates on F_t instead of F. However, since (4.1) shows that F_t is compound Poisson, we may get information on the behavior of F if we make suitable assumptions on the distribution F_t (cf. Corollary 3.1). As indicated by Hogg and Klugman (1984) it is not unreasonable to assume that F_t is GLSD. Doing so, we may estimate the parameters θ and β as in Jain and Gupta (1973), and we get the following result as a direct consequence of Corollary 3.1.

Corollary 4.1 Let F_t and F be given as in (4.1). If F_t is GLSD(θ,β), then

$$\mathbb{P}(X_1=n) \sim \frac{1}{t\lambda}\frac{\log(1-\theta)}{\log(1-\beta^{-1})}p_n \quad (n\to\infty). \tag{4.2}$$

with p_n as in (1.1).

The accuracy in (4.2) may be estimated from Theorem 3.2.

References

Abramowitz, M. and Stegun, I. (1965), *Handbook of Mathematical Functions* (Dover Publications, Inc., New York).

Chover, J., Ney, P. and Wainger, S. (1973), Functions of probability measures, *J. Analyse Math.* **26**, 255-302.

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.

Famoye, F. (1987), A short note on the generalized logarithmic series distribution, *Statistics and Probability Letters* **5**, 315-316.

Hogg, R.V. and Klugman, S.A. (1984), *Loss Distributions*, (Wiley, New York).

Jain, G.C. (1975), On power series distributions associated with Lagrange expansion, *Biom. Z. Bd.* **17**, 85-97.

Jain, G.C. and Gupta, R.P. (1973), A logarithmic type distribution, *Trab. Estadist.* **24**, 99-105.

Kaas, R. (1987) *Bounds and Approximations for some Risk Theoretical Quantities*, Ph.D. Dissertation, University of Amsterdam.

Omey, E. and Willekens, E. (1986), Second order behaviour of the tail of a subordinated probability distribution, *Stoch. Proc. Appl.* **21**, 339-353.

Seneta, E. (1976), *Regularly Varying Functions*, Lecture Notes in Mathematics, 508, (Springer Verlag, Berlin).

Steutel, F.W. (1970), *Preservation of Infinite Divisibility under Mixing and Related Topics*, Mathematical Center Tracts 33, Mathematisch Centrum, Amsterdam.