International Journal of Statistika and Mathematika, ISSN: 2277- 2790 E-ISSN: 2249-8605, Volume 3, Issue 2, 2012 pp 77-81

On Half-Cauchy Distribution and Process

Elsamma Jacob[1] , K Jayakumar[2]

[1]Malabar Christian College, Calicut, Kerala, INDIA.

[2]Department of Statistics, University of Calicut, Kerala, INDIA.

Corresponding addresses:

elsammartin@gmail.com, jkumar19@rediffmail.com

Research Article

Abstract: A new form of half- Cauchy distribution using Marshall-Olkin transformation is introduced. The properties of the new distribution such as density, cumulative distribution function, quantiles, measure of skewness and distribution of the extremes are obtained. Time series models with half-Cauchy distribution as stationary marginal distributions are not developed so far. We develop first order autoregressive process with the new distribution as stationary marginal distribution and the properties of the process are studied. Application of the distribution in various fields is also discussed.

Keywords: *Autoregressive Process, Geometric Extreme Stable, Half-Cauchy Distribution, Skewness, Stationarity, Quantiles.*

1. Introduction:

The half Cauchy (HC) distribution is derived from the standard Cauchy distribution by folding the curve on the origin so that only positive values can be observed. A continuous random variable X is said to have the half Cauchy distribution if its survival function is given by

$$\bar{F}(x)=1-\frac{2}{\pi}\tan^{-1}x, \qquad x>0 \qquad (1.1)$$

As a heavy tailed distribution, the HC distribution has been used as an alternative to exponential distribution to model dispersal distances (Shaw (1995)) as the former predicts more frequent long distance dispersed events than the latter. Paradis et.al (2002) used the HC distribution to model ringing data on two species of tits (*Parus caeruleus and Parus major*) in Britain and Ireland.

From (1.1), we get the probability density function (pdf) *f (x)* and cumulative distribution function (cdf) *F(x)* of the HC distribution as

$$f(x)=\frac{2}{\pi}\frac{1}{1+x^2} \qquad x>0 \qquad (1.2)$$

and

$$F(x)=\frac{2}{\pi}\tan^{-1}x \qquad x>0 \qquad (1.3)$$

respectively, see Johnson et al. (2004). The Laplace transform of (1.2) is

$$\emptyset(t)=\int_0^\infty e^{-tx}f(x)dx = -\sin(t)\,ci(t)\text{-}\cos(t)\,\mathrm{si}(t) \qquad t\geq 0 \qquad (1.4)$$

where

$$si(t)=-\int_t^\infty \frac{\sin\xi}{\xi}d\xi, ci(t)=\int_t^\infty\frac{\cos\xi}{\xi}d\xi, t\geq 0$$

Remark 1.1. *For the HC distribution the moments do not exist.*

Remark 1.2. *The HC distribution is infinitely divisible (Bondesson (1987)) and self decomposable (Diedhiou (1998)).*

Relationship with other distributions:

1. Let Y be a folded *t* variable with pdf given by

$$f(y) = \frac{2\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}}\left(1+\frac{y^2}{\nu}\right)^{(-\frac{\nu+1}{2})}, \qquad y>0, \quad \nu\in N, \qquad (1.5)$$

When $\nu = 1$, (1.5) reduces to

$$f(y) = \frac{2}{\pi}\frac{1}{1+y^2}, \qquad y>0$$

Thus, HC distribution coincides with the folded t distribution with $\nu = 1$ degree of freedom.

2. Let Z_1 and Z_2 be two independent non negative, real valued rvs having the folded standard normal distribution. Then $Y = \frac{Z_1}{Z_2}$ has the HC distribution.

3. It is known that the folded standard normal distribution coincides with the chi-square distribution with one degree of freedom. Therefore, if Z_1 and Z_2 are two independent chi-square variables with parameter 1, then $Y = \frac{Z_1}{Z_2}$ has the HC distribution.

According to Gaver and Lewis (1980), a self decomposable distribution can be the marginal distribution of a stationary first order autoregressive (AR(1)) process of the form

$$X_n=\rho X_{n-1}+\varepsilon_n, \qquad (1.6)$$

Where $\{\varepsilon_n\}$ is a sequence of independent identically distributed (i.i.d) random variables, independent of Xn. The usual procedure to develop AR(1) models of the form (1.6) for self decomposable distributions is by considering the generating functions. Here since the Laplace transform of HC distribution is not in closed form, we use the minification model introduced by Pillai et al. (1995) to develop first order autoregressive process with HC distribution as stationary marginal distribution.

Adding parameters to an existing distribution will give extended forms of the distribution and these distributions are more flexible to model real data. Marshall and Olkin (1997) introduced a general method of adding a parameter to a family of distributions. According to them, if *F(x)* denote the cdf of a continuous rv X, then

$$G(x) = \frac{F(x)}{\alpha + (1-\alpha)F(x)}, \qquad \alpha > 0 \quad (1.7)$$

is also a proper cdf. In this paper, we introduce a new family of HC distribution by applying transformation (1.7) to the HC distribution and using this new class, we develop an autoregressive maximum process with the new form of HC distribution as stationary marginal distribution and study its properties.

The paper is organized as follows: The generalized half Cauchy distribution is introduced and some of its properties are given in section 2. Section 3 deals with the estimation of the parameter. First order autoregressive process with the new distribution as stationary marginal distribution is introduced
and its properties are studied in section 4. Simulated sample path behavior of the process and autoregressive process of order k are given in this section.
Section 5 gives the concluding remarks and the possible area of application of the new model.

2. Generalized Half-Cauchy Distribution:

A generalization of the HC distribution, named Beta Half-Cauchy distribution obtained through beta transformation was introduced by Cordeiro and Lemonte (2011). Here, we introduce another generalization of the HC distribution, which has simple closed form expression for the cdf, using the Marshall-Olkin transformation. By substituting the cdf (1.3) in transformation (1.7), we get a new family of HC distribution with cdf

$$G(x) = \frac{2\tan^{-1}x}{\pi\alpha + 2(1-\alpha)\tan^{-1}x}, \qquad \alpha > 0 \quad (2.1)$$

When $\alpha = 1$, (2.1) reduces to (1.3). The pdf is obtained as

$$g(x) = \frac{2\pi\alpha}{(1+x^2)(\pi\alpha + 2(1-\alpha)\tan^{-1}x)^2}, \quad \alpha > 0, x > 0 \qquad (2.2)$$

and the survival function is

$$\overline{G}(x) = \frac{\propto(\pi - 2\tan^{-1}x)}{\pi\alpha + 2(1-\alpha)\tan^{-1}x}, \qquad (2.3)$$

We call the distribution with cdf given by (2.1) as the Generalized half-Cauchy distribution with parameter α, denoted as *GHC*(α).

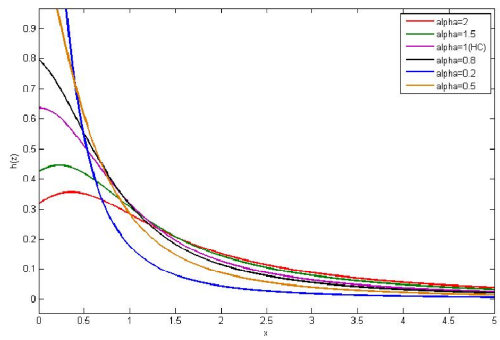


Fig.1 pdf of GHC(α) for α = 2, 1:5, 1, 0.8, 0.2, 0.5

The hazard rate function is

$$h_g(x) = \frac{2\pi}{(1+x^2)(\pi - 2\tan^{-1}x)(\pi\propto + 2(1-\propto)\tan^{-1}x)}, \qquad (2.3)$$

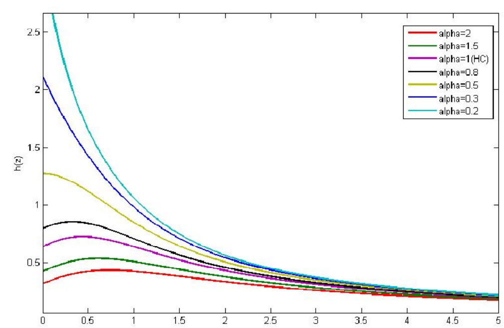


Fig.2 hazard rate function of GHC($\propto$) for $\propto$ = 2, 1,5, 1, 0.8, 0.5, 0.3, 0.2

The shapes of the density function for various values of the parameter are given in figure 1. Figure 2 shows the shapes of the hazard rate function for selected values of $\propto$. The new model is very simple and can be easily simulated as follows. If U is an Uniform (0,1) random variable, then $X = tan\left(\frac{\pi}{2}\frac{\propto U}{1-(1-\propto)U}\right)$ has *GHC*($\propto$) distribution.

Remark 2.1. *The mode of the distribution is the solution of the equation*

$$2(1-\propto) + x[\pi\alpha + 2(1-\alpha)\tan^{-1}x] = 0$$

Remark 2.2. *For the GHC($\propto$) distribution the moments do not exist. (Theorem 1 Rubio and Steel (2012)).*

The q^{th} quantile of the *GHC*($\propto$) distribution is given by

$x_q = F^{-1}(q) = tan\left[\frac{\pi q \propto}{2(1-q(1-\propto))}\right]$, where $0 \leq q \leq 1$ and $F^{-1}(.)$ is the inverse distribution function. For $q = \frac{1}{2}, \frac{1}{4}$ *and* $\frac{3}{4}$; the median and quartiles are respectively

$$Median(X) = tan\left(\frac{\pi\alpha}{2(1+\alpha)}\right)$$

$$Q_1 = tan\left(\frac{\pi\alpha}{2(3+\alpha)}\right)$$

$$Q_3 = tan\left(\frac{3\pi\alpha}{2(1+3\alpha)}\right),$$

The quartile measure of skewness is obtained as

$$\gamma(\alpha) = \frac{tan\left(\frac{3\pi\alpha}{2(1+3\alpha)}\right) + tan\left(\frac{\pi\alpha}{2(3+\alpha)}\right) - 2tan\left(\frac{\pi\alpha}{2(1+\alpha)}\right)}{tan\left(\frac{3\pi\alpha}{2(1+3\alpha)}\right) - tan\left(\frac{\pi\alpha}{2(3+\alpha)}\right)}$$

Definition 2.3 (Marshall and Olkin (1997)). *Let $\{X_i, i \geq 1\}$ be a sequence of independent identically distributed random variables with common cdf F (x) and let N be a geometric random variable with parameter p such that P (N = n) = p(1-p)$^{n-1}$, n=1,2,….; 0<p<1, which is independent of X_i for all $i \geq 1$. Also, let U=min($X_1,X_2,,,,, X_N$) and V =* max*($X_1,X_2,,,,, X_N$). If $F \epsilon \Im$ implies that the distribution of U (V) is in $\Im$, then $\Im$ is said to be geometric-minimum stable (geometric-maximum stable). If $\Im$ is both geometric-minimum and geometric-maximum stable, then $\Im$ is said to be geometric-extreme stable.*

Theorem 2.4. *Let $\{X_i, i \geq 1\}$ and N are defined as in definition (2.3). Then,*

(i) min*($X_1,X_2,,,,, X_N$) has GHC($\propto$ p) distribution,*

(ii) max*($X_1,X_2,,,,, X_N$) has GHC$\left(\frac{\propto}{p}\right)$ distribution.*

Proof. *Let* U = min*(X_1,X_2, …. X_N)* and V=max*(X_1,X_2,X_N)*. The survival function of U is given by

$$P(U \geq x) = P(\min(X_1,X_2,,,,, X_N) \geq x)$$

$$= \sum_{n=1}^{\infty} p(1-p)^{n-1} (\overline{F_x}(x))^n$$

$$= \frac{\propto p(\pi - 2\tan^{-1} x)}{\pi\alpha p + 2(1-\alpha p)\tan^{-1} x}$$

That is, *GHC($\propto$)* distribution is geometric minimum stable. Similarly, the cdf of V is given by

$$P(V \leq x) = P(\max(X_1,X_2,,,,, X_N) \leq x)$$

$$= \sum_{n=1}^{\infty} p(1-p)^{n-1} F_X^n(x)$$

$$= \frac{2p\tan^{-1} x}{\pi\alpha + 2(p-\alpha)\tan^{-1} x},$$

which shows that the distribution is geometric maximum stable.

Remark 2.5. *It follows from definition (2.3) and theorem (2.4) that GHC($\propto$) distribution is geometric extreme stable.*

3. Estimation of Parameters:

The log-likelihood of the sample is given by

$$Log\ L = nlog(2\pi) + nlog\alpha - \sum_{i=1}^{n}\log(1+x_i^2) - 2\sum_{i=1}^{n}\log(\pi\alpha + 2(1-\alpha)\tan^{-1} x_i)$$

The normal equation is

$$2\alpha\sum_{i=1}^{n}\frac{\pi - 2\tan^{-1} x_i}{\pi\alpha + 2(1-\alpha)\tan^{-1} x_i} = n \qquad 3.1$$

The maximum likelihood estimate of α is the solution of equation (3.1). It can be solved numerically by using the function nlm in the statistical software R.

4. First order AR process with *GHC($\propto$)* as marginal distribution:

In this section, we develop stationary autoregressive maximum process with *GHC($\propto$)* marginal distribution. The autoregressive sequence $\{X_n\}$ of *GHC($\propto$)*distributed random variables are related in the following manner:

$$X_n = \begin{cases} \varepsilon_n, & w.p\ p, \\ \max(X_{n-1}, \varepsilon_n), & w.p.\ 1-p \end{cases} \qquad (4.1)$$

where 0 < p < 1 and $\{\varepsilon_n\}$ is a sequence of i.i.d HC random variables, independent of $\{X_n\}$.

Theorem 4.1. *$\{X_n\}$ as defined by (4.1) is a stationary AR(1) process with GHC (1/p) marginal distribution if and only if $\{\epsilon_n\}$ is distributed as HC, provided $X_0 \overset{d}{=} GHC(1/\text{p})$.*

Proof. From (4.1),

$$F_{X_n}(x) = pF_{\varepsilon_n}(x) + (1-p)F_{X_{n-1}}(x)F_{\varepsilon_n}(x)$$

$$= \frac{pF_{\varepsilon_n}(x)}{1-(1-p)F_{\varepsilon_n}(x)},$$

assuming stationarity. Let $\varepsilon_n \overset{d}{=} HC$ with cdf (1.3). Then,

$$F_{X_n}(x) = \frac{2\tan^{-1} x}{\frac{\pi}{p} + \left(1-\frac{1}{p}\right)2\tan^{-1} x}$$

Conversely, Let $X_n \overset{d}{=} GHC(1/\text{p})$

$$F_{\varepsilon_n}(x) = \frac{F_{X_n}(x)}{p+(1-p)F_{X_n}(x)}$$

$$= \frac{2}{\pi}\tan^{-1} x$$

To prove stationarity, let $X_0 \overset{d}{=} GHC(1/\text{p})$ and X_1 is as defined by (4.1).

$$F_{X_1}(x) = [p + (1-p)F_{X_0}(x)]F_{\varepsilon_n}(x)$$

$$= \frac{2\tan^{-1} x}{\frac{\pi}{p} + \left(1-\frac{1}{p}\right)2\tan^{-1} x}$$

Now, let $X_{n-1} \overset{d}{=} GHC(1/\text{p})$ and X_n is as defined by (4.1). Following similar steps, it can be shown that $X_n \overset{d}{=} GHC(1/\text{p})$. Hence the proof.

Remark 4.2. *Even if X_0 is distributed arbitrary with cdf F_{X_0}, the process is asymptotically stationary with $GHC(1/\text{p})$ marginal distribution.*

Proof. Let X_0 is distributed arbitrary with cdf F_{X_0} and $\varepsilon_n \stackrel{d}{=} HC$. Using the autoregressive structure (4.1), the cdf of *Xn* can be expressed as

$$F_{X_n}(x) = pF_{\varepsilon_n}(x)\sum_{i=0}^{n-1}(1-p)^i F_{\varepsilon_n}^{\,i}(x) + (1-p)^n F_{X_0}(x)F_{\varepsilon_n}^{\,n}(x) \to \frac{pF_{\varepsilon_n}(x)}{1-(1-p)F_{\varepsilon_n}(x)},$$

as n → ∞ and since 0 < p < 1.

Now, if $\varepsilon_n \stackrel{d}{=} HC$, the cdf of X_n is obtained as

$$\lim_{n\to\infty} F_{X_n}(x) = \frac{2\tan^{-1}x}{\frac{\pi}{p}+\left(1-\frac{1}{p}\right)2\tan^{-1}x}$$

Hence X_n converges in distribution to GH C (1/p) as n → ∞.

Now we consider the joint distribution of the random variables X_n and X_{n-1}. We have

$$F_{X_{n-1},X_n}(x,y) = P(X_{n-1}\le x, X_n \le y)$$

$$= pP(X_{n-1}\le x)P(\varepsilon_n \le y) + (1\text{-p})\, P(X_{n-1}\le \min(x,y))P(\varepsilon_n \le y)$$

$$= F_{\varepsilon_n}(y)[pF_{X_{n-1}}(x) + (1\text{-p})\, F_{X_{n-1}}(\min(x,y))]$$

$$= \begin{cases} F_{\varepsilon_n}(y)F_{X_{n-1}}(x), & x<y \\ F_{\varepsilon_n}(y)[pF_{X_{n-1}}(x) + (1\text{-p})\, F_{X_{n-1}}(y), & x>y \end{cases}$$

$$= \begin{cases} \frac{4p}{\pi}\frac{\tan^{-1}x\,\tan^{-1}y}{[\pi-2(1-p)\tan^{-1}x]}, & x<y \\ \frac{4p}{\pi}\frac{\tan^{-1}y\,(\pi[p\tan^{-1}x+(1-p)\tan^{-1}y]}{[\pi-2(1-p)\tan^{-1}x][\pi-2(1-p)\tan^{-1}y]} - \frac{8p(1-p)\tan^{-1}x\,(\tan^{-1}y)^2}{\pi[\pi-2(1-p)\tan^{-1}x][\pi-2(1-p)\tan^{-1}y]}, & x>y \end{cases}$$

The joint distribution of the random variables X_n and X_{n-1} is not absolutely continuous since

$$P(X_n = X_{n-1}) = (1-p)P(\varepsilon_n \le X_{n-1})$$

$$= 4p(1-p)\int_0^\infty \frac{\tan^{-1}x\,dx}{(1+x^2)\,(\pi-2(1-p)\tan^{-1}x)^2}$$

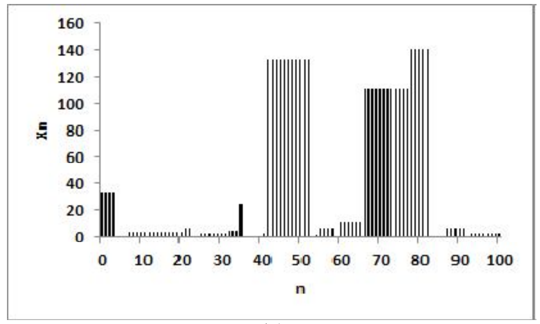
$$= \frac{1-p+plogp}{1-p} \quad \epsilon\ (0,1)$$

On the other hand, we have that

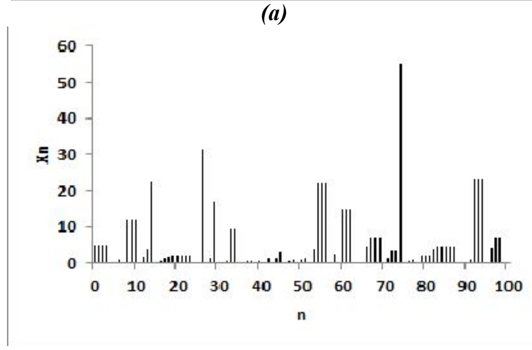
$$P(X_n < X_{n-1}) = pP(\varepsilon_n < X_{n-1})$$

$$= 4p^2\int_0^\infty \frac{\tan^{-1}x\,dx}{(1+x^2)\,(\pi-2(1-p)\tan^{-1}x)^2}$$

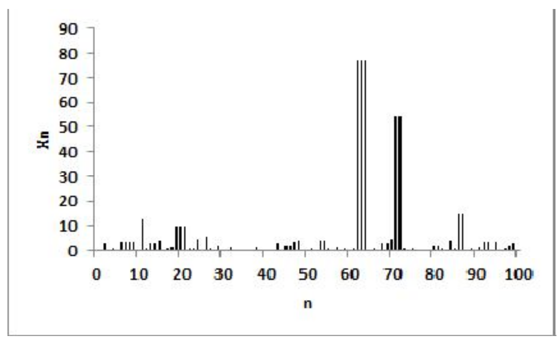
$$= \frac{p(1-p+plogp)}{(1-p)^2}\ \epsilon\ (0,1/2)$$



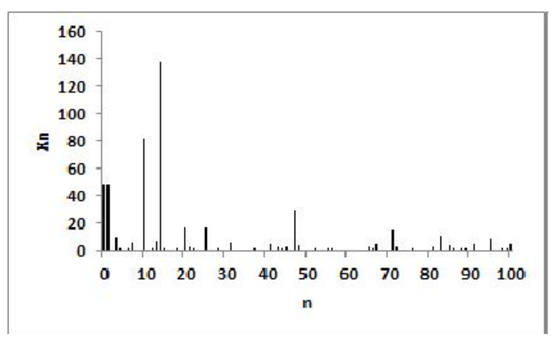
(a)



(b)



(c)



(d)

fig. 3 Simulated Sample path of the Process GHC(∝)

The simulated sample path for the $GHC(\propto)$ process for $\propto = 0.1, 0.3, 0.6, 0.9$ is given in figure 3, (a) – (d). The inferences can be verified by referring to the expression for $P\,(X_n < X_{n-1})$.

The introduced maximum autoregressive process of the first order can be easily generalized to high-order process. Namely, we can introduce maximum autoregressive process of order k as following

$$X_n = \begin{cases} \varepsilon_n, & w.p.\ p_0 \\ \max(X_{n-1}, \varepsilon_n), & w.p.\ p_1 \\ \max(X_{n-2}, \varepsilon_n), & w.p.\ p_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ \max(X_{n-k}, \varepsilon_n), & w.p.\ p_k \end{cases}$$

Where $0 \le p_i < 1$ for $i = 0,1,\ldots,k$ and $\sum_{i=1}^{k} p_i = 1 - p_0$. Then, $\{X_n\}$ is a stationary process with *GHC* $(1/p_0)$ marginal distribution if and only if $\{\varepsilon_n\}$ is distributed as HC. We have that

$$P(X_n \le x) = p_0 P(\varepsilon_n \le x) + \sum_{i=1}^{k} p_i P(\max(X_{n-i}, \varepsilon_n) \le x)$$
$$= p_0 F_{\varepsilon_n}(x) + \sum_{i=1}^{k} p_i\, F_{X_{n-i}}(x)\ F_{\varepsilon_n}(x)$$

On assuming stationarity, we get

$$F_X(x) = \frac{p_0 F_{\varepsilon_n}(x)}{1-(1-p_0)F_{\varepsilon_n}(x)} = \frac{2p_0 \tan^{-1} x}{\pi + 2(p_0 - \pi)\tan^{-1} x}$$

Thus, $X_n \stackrel{d}{=} GHC(1/p_0)$. This shows that the first order model can be easily extended to the k^th^ order case and all results derived above are valid here also.

5. Conclusion

In this paper, we use the transformation introduced by Marshall and Olkin (1997) to define a new model called Generalized half-Cauchy distribution, which extends the half-Cauchy distribution. We study some properties of the model and discuss the maximum likelihood estimation of its parameters. The proposed model is more flexible than the half-Cauchy distribution and can be used effectively for modeling lifetime data. First order autoregressive process with half-Cauchy distribution as stationary marginal distribution is developed for the first time and the properties of the process are studied.

References:

[1] Bondesson, L. On the infinite divisibility of the half-Cauchy and other densities and probability functions on the nonnegative line. Scandinavian Actuarial Journal 31, 225-247, 1987.

[2] Cordeiro, G.M and Lemonte, A.J. The beta-half-Cauchy distribution. Journal of Probability and Statistics, doi: 10.1155/2011/904705, 2011.

[3] Diedhiou, A. On the self decomposability of the half-Cauchy distribution. Journal of Mathematical Analysis and Applications 220, 42-64, 1998.

[4] Gaver, D.P and Lewis, P.A.W. First order autoregressive gamma sequences and point processes. Advances in Applied Probability 12, 724-745, 1980.

[5] Johnson N.L.,Kotz S., Balakrishnan N. Continuous univariate distributions, Vol.1; John Wiley and Sons, Inc., New York, 2004.

[6] Marshall, A.W. and Olkin, I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika 84, 642-652, 1997.

[7] Paradis, E, Baillie, S.R and Sutherland, W.J. Modeling large-scale dispersal distances. Ecological Modeling 151, 279-292, 2002.

[8] Pillai R.N, Jose K.K. Jayakumar K. Autoregressive minification processes and the class of universal geometric minima. Journal of the Indian Statistical Association, 33, 53-61, 1995.

[9] Rubio, F.J. and Steel, M.F.J. On the Marshall-Olkin transformation as a skewing mechanism. Computational Statistics and Data Analysis, 56, 2251-2257, 2012.

[10] Shaw, M.W. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. Proceedings of the Royal Society B 259, 243-248, 1995.