

## A LIMIT THEOREM WITH APPLICATION TO BÅTH'S LAW IN SEISMOLOGY

DAVID VERE-JONES,\* *Victoria University of Wellington and Statistical Research Associates Ltd*

### Abstract

In this paper a limit theorem is proved that establishes conditions under which the distribution of the difference between the size of the initial event in a random sequence, modeled as a finite point process, and the largest subsequent event approaches a limiting form independent of the size of the initial event. The underlying assumptions are that the sizes of the individual events follow an exponential distribution, that the expected total number of events increases exponentially with the size of the initial event, and that the structure of the sequence approximates that of a Poisson process. Particular cases to which the results apply include sequences of independent and identically distributed exponential variables, and the epidemic-type aftershock (ETAS) branching process model in the subcritical case. In all these cases the form of the limit distribution is shown to be that of a double exponential (type-I extreme value distribution). In sampling from a family of aftershock sequences, with possibly different underlying parameters, the limit distribution is a mixture of such double exponential distributions. The conditions for the simple limit to exist relate to the approximation of the distribution of the number of events by a Poisson distribution. One such condition requires the coefficient of variation (ratio of standard deviation to mean) of the number of events to converge to 0 as the mean increases. The results provide a statistical background to Båth's law in seismology, which asserts that in an aftershock sequence the magnitude of the main shock is commonly around 1.2 units higher than the magnitude of the largest aftershock.

*Keywords:* Extreme value distribution; ETAS model; Båth's law; Poisson limits; exponential variates

2000 Mathematics Subject Classification: Primary 60F05; 60K40  
Secondary 86A15

### 1. Introduction

The problems addressed in this paper are motivated by the so-called *Båth's law* in seismology. This asserts that in an aftershock sequence the magnitude of the main shock (usually but not always the first shock in the sequence) is commonly around 1.2 magnitude units higher than the magnitude of the largest aftershock, irrespective of the magnitude of the main event. The magnitudes themselves are approximately independently and exponentially distributed (Gutenberg–Richter law).

Båth's law has been observed and widely discussed since the first compilations of large events and their aftershocks (Utsu (1961)). A question of recurrent interest is whether the law has its origin in some physical process underlying the development of aftershock sequences, or has an essentially statistical origin related to the apparent randomness of the magnitudes in the

---

Received 13 April 2007; revision received 27 June 2008.

\* Postal address: School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, PO Box 600, Wellington, 6140, New Zealand. Email address: david.vere-jones@mcs.vuw.ac.nz

sequence. An early attempt at a statistical model was given in Vere-Jones (1969), using the well-known result (see, e.g. Feller (1966, Section I.6)) that if the magnitudes (main shock included) can be treated as an independent and identically distributed (i.i.d.) sequence of exponential variates, the difference between the largest and second largest magnitudes should have the same exponential distribution as the individual variates. This model is plainly inadequate. Among other points, the exponential distribution does not fit the empirical distribution for the difference, the model fails to explain why the main shock nearly always occurs at or near the beginning of the sequence, and it ignores the obvious dependence of the number of terms in the sequence on the size of the main shock.

Nevertheless, the issue has remained of interest. Recent discussions are given, for example, in Utsu *et al.* (1995), Lombardi (2003), and Console *et al.* (2003). A new focus of interest is the possibility of a link between Båth's law and the 'productivity function' of the main shock, meaning the dependence of the expected number of events in the sequence as a function of the magnitude of the main shock. This issue was raised in Felzer *et al.* (2002), (2004) in connection with the further hypothesis that the magnitudes of the events in a sequence are independent of the previous history of the sequence, so that, for example, a foreshock is regarded simply as an initial event which by chance happens to be followed by a larger event. Felzer *et al.* noted that, empirically, the productivity function increases roughly exponentially with the size of the main shock, and that the parameter of the exponential increase roughly equals the parameter in the exponential distribution of magnitudes, and they suggested that such a balance was a crucial factor in the explanation of Båth's law. Helmstetter and Sornette (2003), Saichev and Sornette (2005), and Zhuang and Ogata (2006) considered the same issue in relation to Ogata's epidemic-type aftershock (ETAS) model (Ogata (1988), Daley and Vere-Jones (2003, Chapter 7)), which again assumes exponentiality both for the magnitude distribution and the productivity function. They showed, through simulation and analytical studies, that a form of Båth's law holds in that context also. An alternative approach was suggested by Shcherbakov and Turcotte (2004), namely, to take Båth's law as the primary notion and develop properties of the aftershock sequence from it.

The purpose of the present paper is to find a general probabilistic setting for results of the kind just described. More specifically, we attempt to determine a general class of point process sequences for which a Båth's law phenomenon may be expected. The main result of the paper is a limit theorem for the distribution of the difference between the size (magnitude) of the initial event (whether or not it is the largest event) and that of the largest subsequent event. It is shown that, under assumptions generalizing somewhat those of an exponential magnitude distribution and exponential productivity function, this difference should approach a Gumbel type-I extreme value distribution (see, e.g. Johnson *et al.* (1995, Chapter 22)) whenever the sequence structure has approximately a Poisson form. In essence, the results are variants on results for the recentered maximum for a sequence of i.i.d. random variables; in particular, some cognate ideas are to be found in Chapter 6 of Galambos (1987). The mathematical background and a basic limit theorem are set out in Section 2, while special cases and extensions of the theorem are stated and proved in Section 3. The most important extension is to situations where the number distribution is a mixture of Poisson distributions. The applications to seismology are briefly discussed in Section 4.

## 2. Definitions, assumptions, and a basic limit theorem

Båth's law problem resolves itself into the consideration of a cluster, or sequence,  $\{(t_0, X_0), (t_1, X_1), (t_2, X_2), \dots, (t_N, X_N)\}$  of times  $t_i$  and sizes (magnitudes)  $X_i$ , with random

length  $N$ . Apart from ordering the elements in the sequence, the times play no essential part in the study, and we therefore consider just the sequence of sizes  $\{X_0, X_1, \dots, X_N\}$ , of which  $X_0$  plays a distinguished role.

To provide a framework for studying such a sequence, we consider it as a finite point process. We denote its probability generating functional (PGFL), conditional on (but not including) a given value  $u$  for  $X_0$ , by  $G(h \mid u)$ , so that, for any measurable function  $h(x)$  with  $0 < h(x) \leq 1$ ,

$$G(h \mid u) = E \left[ \prod_{i=1}^N h(X_i) \mid X_0 = u \right].$$

Properties of the PGFL for almost sure finite point processes are set out, for example, in Section 5.4 of Daley and Vere-Jones (2003).

In the earthquake context, it is reasonable to treat the sizes  $X_i$  as identically distributed exponential variables (Gutenberg–Richter law), conditionally independent of each other, but potentially influencing the number of succeeding terms in the sequence. Although we can generalize this requirement to some degree, it underlies much of the conceptual thinking.

We make two basic assumptions in the analysis. The first of these relates to the conditional first moment measure

$$M_1(dx \mid u) = E[N(dx) \mid X_0 = u],$$

where  $N(B)$  counts the number of events in the cluster with sizes falling within the Borel set  $B$ . We assume that, for all  $u$ , this measure exists (in the sense that it allocates finite mass to bounded Borel sets in  $\mathbb{R}$ ) and that, for any bounded Borel set  $B$ ,  $M_1(B \mid u)$  is an increasing, measurable function of  $u$ . The crucial condition, in its simplest form, is that, for some  $A > 0$ ,  $\alpha > 0$ , and  $\beta > 0$ ,  $M_1$  should satisfy the following separability condition.

**Assumption I.**

$$M_1(dx \mid u) = Ae^{\alpha u} \beta e^{-\beta x} dx. \tag{1}$$

This assumption combines the exponential forms both for the productivity function ( $Ae^{\alpha u}$  term) and the size distribution ( $\beta e^{-\beta x} dx$  term).

For many purposes, the condition can be relaxed in the following manner. Let  $S_x$ ,  $x \in \mathbb{R}$ , denote the shift operator acting on  $\sigma$ -finite Borel measures  $\nu$  on  $\mathbb{R}$  in the sense that, for any bounded Borel set  $B$ ,

$$(S_x \nu)(B) = \nu(x + B).$$

Also, let  $M_1^\dagger(dx \mid u)$  denote the measure on  $\mathbb{R}$  equal to  $M_1(dx \mid u)$  on  $\mathbb{R}^+$  and to the empty measure on the negative half-line. For want of a better term, say that a family of totally finite measures on the Borel sets of  $\mathbb{R}$  converges  $h$ -weakly (denoted by ‘ $\xrightarrow{h}$ ’) to a  $\sigma$ -finite limit, with bounded mass on any half-line  $[C, \infty)$ , if their restrictions to any such half-line converge weakly to the corresponding restriction of the limit measure. Then Assumption I can be generalized as follows.

**Assumption I\*.** For some  $A > 0$ , the shifted measures  $S_{(\alpha/\beta)u} M_1^\dagger(dx \mid u)$  converge  $h$ -weakly to the multiple of the extended exponential measure with density  $A\beta e^{-\beta(\cdot)}$  over the whole real line:

$$S_{(\alpha/\beta)u} M_1^\dagger(dx \mid u) \xrightarrow{h} A\beta e^{-\beta x} dx. \tag{2}$$

The condition of the assumption is satisfied, in particular, if  $M_1$  has the multiplicative form of the right-hand side of (1). It is closely related to, but different from, the requirement that

$$e^{-\alpha u} M_1(dx | u) \rightarrow A\beta e^{\beta x} dx,$$

weakly, which is also satisfied by the special case of Assumption I.

The second basic assumption is the requirement that, in the sense below, the point process  $N$ , conditional on  $X(0) = u$ , should approximate a Poisson process as  $u$  increases.

**Assumption II.** For any family of measurable test functions  $h_u(x)$ , with  $0 \leq h_u(x) \leq 1$ , and for which there exists a constant  $C_h$ ,  $0 < C_h < \infty$ , such that

$$I_u = \int_{\mathbb{R}} (1 - h_u(x)) M_1(dx | u) \rightarrow C_h, \tag{3}$$

the PGFL  $G(h | u)$  satisfies

$$|\log G(h_u | u) + I_u| \rightarrow 0, \quad u \rightarrow \infty. \tag{4}$$

Condition (4) is essentially a requirement that, conditionally on  $u$ , the process approximates the finite Poisson process with intensity measure  $M_1(dx | u)$ . Using the terminology of Daley and Vere-Jones (2003, Chapter 5), it requires the contribution  $I_u$  of the first factorial cumulant measure, which is just  $M_1$  itself, to dominate the expansion of the log PGFL. For a Poisson process, the second- and all higher-order factorial cumulant measures vanish, so the condition is trivially satisfied. In the special case that the quantities  $I_u$  tend to a limit  $C_h$ , as required above, then it is clear that in fact

$$\log G(h_u | u) \rightarrow C_h.$$

Much of the argument in the present paper is devoted to identifying some broad classes of finite point process models for which Assumption II holds when  $M_1$  satisfies either Assumption I or Assumption I\*.

For Båth's law in particular, the focus of discussion is the behavior of

$$\Delta_u = \max_{1 \leq n \leq N} \{X_n\} - \frac{\alpha}{\beta} u,$$

given  $X_0 = u$ , as  $u \rightarrow \infty$ . In the case in which  $\alpha = \beta$ ,  $-\Delta_u$  can be interpreted as the difference between the size of the initial event and that of the largest subsequent event. The condition  $\alpha = \beta$  is that suggested in Felzer *et al.* (2004) as being necessary for Båth's law to hold; the forms suggested in Assumptions I and II allow extensions to cases where  $\alpha \neq \beta$ .

Now consider the limit, as  $u \rightarrow \infty$ , of the distribution function  $F_u(y) = \Pr\{\Delta_u \leq y | X_0 = u\}$  and, for brevity, write

$$M^* = \max_{1 \leq n \leq N} X_n.$$

We adopt the convention that  $M^* = 0$  if  $N = 0$ , and prove the following result.

**Theorem 1.** Suppose that Assumptions I\* and II hold. Then, as  $u \rightarrow \infty$ ,  $F_u(y)$  converges weakly to the following double exponential distribution:

$$F_u(y) \rightarrow \exp[-Ae^{-\beta y}], \quad u \rightarrow \infty, \quad -\infty < y < \infty, \tag{5}$$

with  $A$  defined as in (2)

*Proof.* The proof is effectively just a particular application of (4). To see this, first observe that the distribution function of the maximum  $M^*$ , for a given value of the index  $u$ , is given, for  $v > 0$ , by

$$\begin{aligned} J_u(v) &= \Pr\{M^* \leq v \mid u\} \\ &= \mathbb{E} \left[ \prod_{n=1}^N H(v - X_n) \mid u \right] \\ &= G(H(v - x) \mid u) \\ &= G(1 - H(x - v) \mid u), \end{aligned}$$

where  $H(\cdot)$  is the Heaviside function. For  $v < 0$ ,  $J_u(v) = 0$ , and since  $J_u(0) = \Pr\{N = 0 \mid X(0) = u\} \equiv p_0(u)$ ,  $J_u$  has an atom at 0 if  $p_0(u) > 0$ . Now taking  $v = (\alpha/\beta)u + y$ , we obtain

$$F_u(y) = J_u\left(\frac{\alpha}{\beta}u + y\right) = G_u\left(H\left(\frac{\alpha}{\beta}u + y - x\right)\right).$$

This corresponds to the choice

$$h_u(x) = H\left(\frac{\alpha}{\beta}u + y - x\right)$$

in the PGFL, (3). The integral in (3) reduces here to

$$I_u = \int_{(\alpha/\beta)u+y}^{\infty} M_1(dx \mid u).$$

From Assumption I\*, as  $u \rightarrow \infty$ , this integral converges to

$$Ae^{\alpha u} e^{-\beta((\alpha/\beta)u+y)} = Ae^{-\beta y}.$$

Assumption II therefore implies that  $\log F_u(y) \rightarrow -Ae^{-\beta y}$ , and so

$$F_u(y) \rightarrow \exp[-Ae^{-\beta y}], \quad u \rightarrow \infty, \quad -\infty < y < \infty. \tag{6}$$

Since the limit holds for all real  $y$ , the distribution functions converge pointwise. Since the limit is a proper distribution function, the distributions converge weakly. This completes the proof.

### 3. Special cases and variations

In this section we examine some special cases of Theorem 1 and some further variations.

#### 3.1. Poisson and mixed Poisson clusters

The paradigm example for the behavior described in Theorem 1 occurs when, given  $X_0 = u$ , the sequence is a finite Poisson process with intensity measure  $\mu^*(dx \mid u)$  having the separable form given in (1), or more generally satisfying Assumption I\*. The left-hand side of (4) is then 0 and no further approximations are needed. This case was noted in Saichev and Sornette (2005).

There is an important extension of this special case to mixed Poisson processes. Such a situation could arise in considering the statistics from a family of aftershock sequences if each individual sequence was of Poisson form, but the parameter  $A$  varied from sequence

to sequence. Let us treat parameter  $A$  in (1) and (2) (Assumption I) as a random variable, say  $W$ . For given  $W$ , the cluster has the Poisson form just discussed, with parameter measure  $W\mu^*(dx | u) = We^{\alpha u} \beta e^{-\beta x} dx$ . Taking expectations over  $W$ , the PGFL  $G(h | u)$  then has the form

$$G(h | u) = E \left[ \exp \left[ W \int_{\mathbb{R}} (h(x) - 1) \mu^*(dx | u) \right] \right] = \ell_W \left( \int_{\mathbb{R}} (1 - h(x)) \mu^*(dx | u) \right),$$

where  $\ell_W(s) = E[e^{-sW}]$  is the Laplace transform of  $W$ .

Giving  $h$  the special form  $h_u(x) = H((\alpha/\beta)u + y - x)$  and assuming that Assumption I holds for  $\mu_u^*$ , it follows, from the arguments used in Theorem 1 and the continuity of the Laplace transform, that

$$F_u(y) = \ell_W \left( \int_{(\alpha/\beta)u+y}^{\infty} \mu_u^*(dx) \right) \rightarrow \ell_W(e^{-\beta y}).$$

We have therefore established the following partial extension to Theorem 1.

**Theorem 2.** *Suppose that the cluster process has a mixed Poisson form with random parameter measure  $W\mu^*(dx | u)$ , where  $\mu^*$  satisfies the separability condition (I) of Assumption I, and  $W$  has Laplace transform  $\ell_W(s) = E[e^{-sW}]$ . Then the distribution  $F_u$  is a mixture of double exponential distributions with distribution function  $\ell_W(e^{-\beta y})$ .*

Assuming that second moments exist, it is only in the case that the distribution of  $W$  is degenerate (reduces to a single fixed atom) that the distribution  $F_u(y)$  reduces to the double exponential form given in Theorem 1. To see this, first note that the overall mean is  $m(u) = Ce^{\alpha u}$ , where  $C = E[W]$ . Then, looking at the second-order term in the expansion of  $\log G(1 - h | u)$ , we see that its leading term is of the form  $De^{\alpha u}$ , where  $D$  is the standard deviation of the mixing distribution. Thus, Assumption I cannot hold unless  $D = 0$ , which is possible only if the mixing distribution is degenerate.

Supposing that parameter  $\beta$  is fixed, the role of parameter  $A$  in the double exponential (5) is to fix the location (mean or median) of the distribution. Thus, a distribution of the form  $\ell_W(e^{-\beta Y})$  can be roughly regarded as a mixture of the simple double exponentials with varying locations. Commonly, this will still result in a unimodal distribution, but the mixing will broaden the modal peak.

### 3.2. Magnitudes independent of cluster structure

Another simple context relates to situations where the magnitudes are i.i.d. with common exponential distribution with parameter  $\beta$  and, with the exception of the initial magnitude  $u$ , do not affect the subsequent sequence structure.

In this case the behavior of the system is controlled by the dependence on  $u$  of the probability distribution for the total number of events, say  $p_n(u) = \Pr\{N = n | X_0 = u\}$ , with mean

$$m(u) = E[N | X_0 = u] = \sum_0^{\infty} n p_n(u).$$

Here

$$M_1(dx | u) = m(u) \beta e^{-\beta x} dx,$$

while

$$G(h_u | u) = \sum_1^{\infty} p_n(u) \left( \int_0^{\infty} h_u(x) \beta e^{-\beta x} dx \right)^n = \Pi(\bar{h}_u | u),$$

where

$$\Pi(z | u) = E[z^N | X_0 = u] = \sum_1^\infty p_n(u)z^n$$

is the probability generating function for the total number  $N$  in the sequence, given  $X_0 = u$ , and

$$\bar{h}_u = \int_0^\infty h_u(x)\beta e^{-\beta x} dx.$$

Assumption I\* now reduces to the requirement that

$$m(u) \sim Ae^{au}, \tag{7}$$

while to satisfy Assumption II, we need

$$|\log\{\Pi(\bar{h}_u | u)\} + I_u| \rightarrow 0.$$

Since  $I_u$  is given by  $m(u)(1 - \bar{h}_u)$  here, this reduces to

$$\left| \log \left\{ \Pi \left( \left( 1 - \frac{I_u}{m(u)} \right) \middle| u \right) \right\} + I_u \right| \rightarrow 0.$$

We are now in a position to derive a more transparent sufficient version of Assumption II for the present context, in terms of the first absolute moment of the number distribution about its mean:

$$v(u) = E[|N - m(u)| | X_0 = u] = \sum_0^\infty |n - m(u)|p_n(u).$$

**Theorem 3.** *Suppose that the magnitudes  $X_i$  are i.i.d. with common exponential distribution with parameter  $\beta$ , and that the conditional mean  $m(u)$  and first absolute moment  $v(u)$  of the total number  $N$  exist. Then Assumption II holds, and hence also Theorem 1, provided that  $m(u) \rightarrow \infty$  and*

$$\frac{v(u)}{m(u)} \rightarrow 0, \quad u \rightarrow \infty. \tag{8}$$

*If, in particular, the second moment of the number of points  $N$  in the sequence is finite, so that its standard deviation  $\sigma(u)$  exists, then (8) is satisfied whenever*

$$\frac{\sigma(u)}{m(u)} \rightarrow 0, \quad u \rightarrow \infty.$$

*Proof.* First suppose that condition (8) holds and, for given  $A > 0$ , consider the difference

$$\begin{aligned} & \left| \Pi \left( \left( 1 - \frac{z}{m(u)} \right) \middle| u \right) - E \left[ \left( 1 - \frac{z}{N} \right)^N \middle| X(0) = u \right] \right| \\ &= \left| E \left[ \left( 1 - \frac{z}{m(u)} \right)^N - \left( 1 - \frac{z}{N} \right)^N \right] \right| \\ &\leq \sum_0^\infty p_n(u) \left| \left( 1 - \frac{z}{m(u)} \right)^n - \left( 1 - \frac{z}{n} \right)^n \right|. \end{aligned}$$

Using the factorization

$$x^n - y^n = (x - y)(x^{n-1} + x^{n-2}y + \dots + y^{n-1})$$

and the fact that  $x$  and  $y$  are both bounded by unity here, the last sum is bounded by

$$z \sum_0^\infty np_n(u) \left| \frac{n - m(u)}{nm(u)} \right| = \frac{z}{m(u)} \sum_0^\infty p_n(u) |n - m(u)| = \frac{z\nu(u)}{m(u)}.$$

Writing  $z = I_u$  and recalling that  $I_u$  is bounded, the right-hand side converges to 0 by the assumption of the theorem, and, hence,

$$\left| \Pi_u \left( \left( 1 - \frac{I_u}{m(u)} \right) \middle| u \right) - E \left[ \left( 1 - \frac{I_u}{N} \right)^N \middle| X(0) = u \right] \right| \rightarrow 0.$$

On the other hand, since, for any finite  $n$ ,  $\Pr\{N > n \mid X_0 = u\} \rightarrow 1$ , and, hence, since  $[1 - z/n]^n \approx e^{-z}$  for large  $n$ , while  $I_u$  remains bounded,

$$\left| E \left[ \left( 1 - \frac{I_u}{N} \right)^N \middle| X(0) = u \right] - \exp[-I_u] \right| \rightarrow 0.$$

Hence, also

$$\left| G \left( \left( 1 - \frac{I_u}{m(u)} \right) \middle| u \right) - \exp[-I_u] \right| \rightarrow 0,$$

from which Assumption II follows by continuity of the logarithmic function and the fact that  $I_u$  is also bounded away from 0.

The final statement follows from the inequality, valid for  $E[X] = 0$ ,

$$\text{var}(X) = E[|X|^2] \geq E[|X|]^2.$$

This theorem has the general character of an Abelian theorem, with condition (8) as the direct statement and condition (4) as the averaged form. This suggests that the two statements are close to equivalent, with the converse true under some additional constraints. Indeed, if we keep the same assumptions, and suppose in addition that

$$\left| \frac{\log\{G_u(1 - \eta)\}}{\eta m(u)} + 1 \right| \rightarrow 0$$

uniformly in  $u$  as  $\eta \rightarrow 0$ , then it is not difficult to show that the distribution  $F_u(y)$  of  $\Delta_u$  converges as in (6) if and only if (7) holds.

It is of interest to see how Theorem 3 works for some special classes of distribution. Suppose, for example, that the number distributions are negative binomial with PGFLs of the general form

$$\Pi_u(z) = \left( \frac{1 - p(u)}{1 - p(u)z} \right)^{r(u)},$$

where  $p(u)$  and  $r(u)$  are parameters depending on the initial magnitude  $u$ . The mean and the variance are respectively given by

$$m(u) = \frac{p(u)r(u)}{1 - p(u)}, \quad \sigma^2(u) = \frac{p(u)r(u)^2}{1 - p(u)},$$

so that  $\sigma(u)/m(u) = 1/\sqrt{p(u)r(u)}$ . Thus, both conditions of Theorem 3 are satisfied provided that  $p(u)r(u) \rightarrow \infty$ . Since  $p(u)$  is bounded above by 1, this requires that  $r(u) \rightarrow \infty$ , which cannot be achieved by merely requiring that  $p(u) \rightarrow 1$ .



### 3.3. Clusters in the form of a Poisson cluster process

Here we consider the application of Theorem 1 to situations where the overall cluster is itself a Poisson cluster process; thus, it is made up of subclusters along the magnitude axis (and possibly also in time, although we do not consider the time dimension explicitly). Such a process is characterized by a PGFL of the form

$$G_u(h) = \exp \left[ \int_0^\infty (G^m(h | y) - 1) \mu^\dagger(dy | u) \right], \tag{9}$$

where  $\mu^\dagger(dy | u)$  governs the Poisson process of cluster centers, while an individual cluster, with center at  $y$ , has PGFL  $G^m(h | y)$  (cf. Daley and Vere-Jones (2003, Section 6.3)). The overall first moment measure in this case is given by

$$M_1(dx | u) = \int_0^\infty M^m(dx | y) \mu^\dagger(dy | u),$$

where  $M^m(dx | y)$  is the first moment measure of the cluster member process with center at  $y$ .

Now take logarithms of (9) and expand  $G^m(h | x)$  about  $h(x) \equiv 1$  in terms of the factorial moment measures of the cluster member process. The first two of these,  $M^m(dy | x)$  and  $M_{[2]}^m(dy_1 \times dy_2 | x)$ , respectively correspond to the probabilities of finding a cluster member at  $y$ , and two cluster members at  $y_1$  and  $y_2$ , given that the cluster center is at  $x$ . A wide range of possible behaviors can be obtained, depending on the relation of these cluster points to the cluster center. Rather than trying to formulate a specific theorem, we indicate some of the possibilities which can arise.

1. If there is a positive probability that the cluster will contain two or more events in the vicinity of the cluster center, then the clustering tendency will be carried forward to events of arbitrarily large magnitude and the Poisson approximation embodied in Assumption II, as well as Theorem 1 itself, will fail.
2. If the locations (magnitudes) of cluster members are independent of the location of the cluster center, then we are essentially back in the situation considered in the previous subsection where the cluster structure and the magnitude distribution are separated. A Båth's law can hold under suitable conditions on the total size of the cluster as a function of the initial value  $u$ .
3. In general, the inequality (see Daley and Vere-Jones (2007, Corollary 9.5.VII))

$$\left| \log\{G^m((1 - h) | x)\} - \int h(y)M_1^m(dy | x) \right| \leq \frac{1}{2!} \int \int h(y_1)h(y_2)M_{[2]}^m(dy_1 \times dy_2 | x),$$

valid provided that the second moment measure  $M_{[2]}^m$  of the cluster member process exists, can be used to frame conditions for the validity of a Båth's law in terms of the second-order properties of the cluster member process. For example, versions of Theorem 1 can be obtained by requiring that the double integral on the right-hand side above converges to 0 faster than  $I_u$  when  $h_u$  satisfies (3). We have not found a simple and natural condition of this kind, however, and therefore we desist from considering this option in more detail.

### 3.4. The ETAS model

The most interesting example of a Poisson cluster process is the self-exciting or ETAS model, which has a branching process interpretation. However, the approach suggested in the previous section fails, as we shortly show, although a direct argument based on the branching structure is still possible.

The cluster here consists of the total offspring from all generations. Since each offspring generates further total offspring according to the same probability law, the PGFL for the total offspring, excluding the initial ancestor, satisfies the relation

$$G(h \mid u) = G_1(h(y)G(h \mid y) \mid u), \tag{10}$$

where  $G_1(h \mid u)$  is the PGFL for the first generation offspring, which in the ETAS model is a Poisson process with intensity measure having exactly the separable form of Assumption I. Equation (10) therefore defines a Poisson cluster process, the cluster center process having the identical structure to the total cluster with the addition of the initial point.

To keep the discussion as simple as possible, suppose, as in the ETAS model, that  $G_1$  corresponds to a Poisson process with factorizable intensity measure

$$\mu(dx \mid u) = Ae^{\alpha u} \beta e^{-\beta x} dx,$$

so that

$$G_1(h \mid u) = \exp\left[-Ae^{\alpha u} \int_{\mathbb{R}^+} (1 - h(x))\beta e^{-\beta x} dx\right].$$

Let us first consider the moment measures of such a process. The first moment measure of the total population (all generations accumulated) generated by a single ancestor of size  $u$  is then given by

$$M_1(dx \mid u) = 1 + \mu(dx \mid u) + \mu * \mu(dx \mid u) + \mu * \mu * \mu(dx \mid u) + \dots,$$

where the symbol “\*” denotes the operation

$$\mu * \nu(dx \mid u) = \int \mu(dx \mid w)\nu(dw \mid u).$$

In the special case considered,  $\mu * \mu(dx \mid u)$  exists and equals  $\rho\mu(dx \mid u)$ , provided that  $\alpha < \beta$ , where

$$\rho = \frac{A}{1 - \alpha/\beta}.$$

Similar simplifications apply to the higher-order convolutions. In this case, therefore, the series defining the first moment measure converges if and only if  $\rho < 1$ , the latter also being the condition for the process to be subcritical. When it is satisfied,

$$M_1(dx \mid u) = \frac{\mu(dx \mid u)}{1 - \rho},$$

and therefore it retains the separable form required for Assumption I.

The second moment measure can also be obtained by summing the contributions from one, two, three, ... generations. Even the contribution from two generations raises a further

constraint, for it includes a term corresponding to the situation where one initial offspring at  $z$  gives rise to a pair of offspring at  $x$  and  $y$ , leading to an integral of the form

$$\int \mu(dz | u)\mu(dx | z)\mu(dy | z) = A^3 \beta^3 e^{\alpha u} e^{-\beta(x+y)} \int e^{2\alpha z} e^{-\beta z} dz,$$

which converges only if  $\beta < \alpha/2$ . Similarly, on considering a chain of three offspring, we find terms which converge only if  $\alpha < \beta/3$ , so that, overall, the series defining the second moment will converge only in the trivial case,  $\alpha = 0$ .

Arguments based on the existence of second moments therefore fail for the ETAS and similar Hawkes' processes. However, Zhuang and Ogata (2006) (see also Saichev and Sornette (2005)) have developed a more explicit analysis for the ETAS model, which can be adapted for our purposes. Relation (10) for the PGFL here takes the form of the functional equation

$$G(h | u) = \exp \left[ -Ae^{\alpha u} \int_{\mathbb{R}^+} (1 - h(x)G(h | x))\beta e^{-\beta x} dx \right]. \tag{11}$$

Write  $\Phi = \Phi(h | u)$  for the integral

$$\int_{\mathbb{R}^+} (1 - h(x)G(h | x))\beta e^{-\beta x} dx$$

appearing in the exponential term, so that, on substituting for  $G(h | x)$  inside the integral, we see that, for given  $u$  and  $h$ ,  $\Phi$  satisfies the recurrence relation

$$\Phi = \int_{\mathbb{R}^+} (1 - h(x) \exp[-Ae^{\alpha x} \Phi])\beta e^{-\beta x} dx.$$

We rewrite this in the form

$$\Phi = \int_{\mathbb{R}^+} (1 - h(x))\beta e^{-\beta x} dx + \int_{\mathbb{R}^+} h(x)(1 - \exp[-Ae^{\alpha x} \Phi])\beta e^{-\beta x} dx$$

and seek conditions under which the second integral, say  $I_2$ , converges to 0 with  $u$  more rapidly than the first integral, say  $I_1$ . Then only  $I_1$  needs to be considered as we let  $u \rightarrow \infty$ .

In the application to Båth's law we take  $h(x) = H(x - (\alpha/\beta)u - y)$  and, for brevity, write  $r = \alpha/\beta$  and assume that  $r < 1$ . Then  $I_1$  reduces to

$$\int_{ru+y}^{\infty} \beta e^{-\beta x} dx = e^{-\beta(ru+y)} = e^{-\alpha u - \beta y}.$$

Also, using the inequality  $1 - e^{-M} \leq M$ ,  $I_2$  is bounded by

$$\int_0^{ru+y} Ae^{\alpha x} \Phi \beta e^{-\beta x} dx \leq \Phi A \int_0^{\infty} \beta e^{-(\beta-\alpha)x} dx = \Phi \frac{A}{1-r},$$

the integrals converging provided that  $\beta > \alpha$  and leading to an initial bound of the form

$$\Phi \leq \frac{e^{-\alpha u + y}}{1 - \rho}$$

provided that the criticality parameter  $\rho = A/(1-r) < 1$ .

To proceed further, we examine  $I_2$  more carefully; following Zhuang and Ogata (2006) we write  $v = e^{\alpha x}$  and bound the integral by

$$I_2 \leq \frac{1}{r} \int_0^\infty (1 - e^{-A\Phi v})v^{-(1+1/r)} dv.$$

Then making the further substitution  $w = A\Phi v$ , we find that

$$I_2 \leq (\text{constant})\Phi^{1/r} \int_0^\infty (1 - e^{-w})w^{-(1+1/r)} dw.$$

Since  $0 < r < 1$ , the integral converges and  $\Phi^{1/r} = o(\Phi)$  as  $u \rightarrow \infty$  and  $\Phi \rightarrow 0$ . Thus, the first integral dominates, as required, and we have established the following theorem.

**Theorem 4.** *Let  $\{X_0, X_1, \dots, X_N\}$  be a finite cluster (parent and total offspring) from an ETAS process with  $\alpha < \beta$  and  $\rho < 1$ . Then, using the notation of Theorem 1, as  $X_0 = u \rightarrow \infty$ ,*

$$F_u(y) \rightarrow \exp[-\alpha e^{-\beta y}], \quad -\infty < y < \infty.$$

Assumption I fails in the cases in which  $\rho \geq 1$ , because the overall first moment  $M_1$  for the total cluster is then infinite. An interesting question is whether a form of Båth's law continues to hold in the critical case with  $\alpha < \beta$ . The functional equation (11) continues to hold, but the first-order terms in  $\Phi$  then cancel, so that the solution corresponds to branch-point behavior, and further analysis is required.

#### 4. Seismological background and interpretation

We briefly consider the applicability and interpretation of the preceding results in the seismological context. A fuller discussion is in preparation for the companion paper Vere-Jones *et al.* (2008).

Aftershock data from a global or regional catalogue is subject to important limitations: the data on very large earthquakes is limited by the time frame of the study, and the data on small earthquakes is limited by network coverage and sensitivity. The definition of an aftershock sequence, or other cluster, is fraught with difficulties and disagreements; commonly, some arbitrary elements enter into the definitions.

Even bearing in mind the above difficulties, it is generally possible to group the events in a shallow earthquake catalogue into nonoverlapping clusters, or sequences, with one or more members. The majority of sequences with at least two members have a well-defined initial event and a well-defined largest event (often but not always the initial event), which is commonly followed by a sequence of smaller events with frequency decreasing in time until the terms of the sequence are indistinguishable from the general background activity. Some other types of sequences, such as swarms and multiplets, are also observed, particularly in volcanic regions.

Such empirically determined aftershock sequences then form the data to which the results of the preceding section may be applicable. Both the number of events in the sequence  $N_u$  and the value of the difference  $\Delta_u$  may be recorded for each sequence and plotted against the magnitude  $u$  of the initial event, or (more commonly) of the largest event. Empirical studies of this kind form the basis of papers cited previously, such as Utsu (1961), Vere-Jones (1969), Lombardi (2002), and many others.

The exponential form of the productivity function is generally easy to confirm empirically, whether the size of the sequence is plotted against the magnitude of the initial event or that

of the largest event. However, there may be some outliers in the form of extremely long aftershock sequences, often caused by the occurrence of unusually large later events in the original sequence, which then set off their own supplementary aftershock sequences.

Determining the empirical distribution of  $\Delta_u$  presents somewhat greater difficulties. Firstly, it is necessary to firmly adhere to a common lower magnitude threshold, say  $M_c$ , and to include all sequences with  $u$  above the threshold. Selecting just the sequences initiated by large events, and neglecting the others, leads to significant bias; see, e.g. Vere-Jones (1969) and Lombardi (2002) for further discussion of this point.

In the seismological context it is traditionally

$$B_u = -\Delta_u = u - M^*$$

which is studied, at least when  $u \geq M^*$ . Because of the sign change, the limiting distribution for  $B_u$ , assuming that Theorem 1 holds and  $\alpha = \beta$ , has the form

$$\Pr\{B_u \leq y\} = 1 - \exp[-Ae^{\beta y}].$$

Moreover, because of the existence of the lower threshold  $M_c$ , for any finite  $u$ , the distribution of  $B_u$  is bounded above by  $u - M_c$ , where it has an atom taking up (approximately) the remaining mass in the limit distribution. The empirical distribution usually displayed corresponds to the renormalized form after removing this atom. Empirical results to date have not generally discriminated between contexts where  $\alpha = \beta$  or otherwise.

When  $u < M_u$ , the initial event is commonly considered a foreshock and  $B_u$  is replaced by  $M_u - M_u^{(2)}$ , where  $M^{(2)}$  is the second largest magnitude of the events in the sequence. Thus, the study in Vere-Jones *et al.* (2006) breaks with tradition in taking the magnitude  $u$  of the initial event as the classifying variable, even when the initial event is not the largest event. Another difficulty is that, unless  $\alpha = \beta$ , we should expect the distribution of  $B_u$  to be shifted a distance

$$\delta_u = \left(\frac{\alpha}{\beta} - 1\right)(u - M_c) \quad (12)$$

from the double exponential form predicted in the case in which  $\alpha = \beta$ . Unless there is enough data to examine individual values of  $u$ , the observed distribution for  $B_u$  will therefore be a mixture of shifted versions of the double exponential, leading to a broadening of the modal peak, as in the discussion of Theorem 2.

Issues of this kind are discussed briefly in Vere-Jones *et al.* (2006) and will be treated at greater length in Vere-Jones *et al.* (2008). The broad conclusions from those studies may be summarized as follows.

1. In general, the double exponential form in Theorem 1, or more properly the generalized form described in Theorem 2, provides a plausible candidate for the distribution of the Båth's law difference  $B_u$ . For fixed  $u$ , or at least restricted ranges of  $u$ , the visual appearance is convincing, and can be extended, with suitable adjustments, even to cases where  $\alpha < \beta$ .
2. A key role is played by the probability that the initiating event is followed by a larger event, or, more loosely, is a foreshock. For the double exponential distribution, the limit form of this probability is given by  $F_u(0) = 1 - e^{-A}$  and, therefore, uniquely determines parameter  $A$ . Provided that  $u$  here is interpreted as the number of magnitude units above the threshold, this probability is independent of the magnitude  $u$  of the initiating event

if  $\alpha = \beta$ , but otherwise has to be adjusted as in (12). Typical values of this probability (assuming that  $\alpha \approx \beta$ ) lie around 5% (see, e.g. Felzer *et al.* (2004)), so that  $A \approx 0.051$ .

3. The mode and other measures of location of the limit distribution for  $F_u(y)$  in Theorem 1 are also determined by parameter  $A$ . For example, the mode occurs when

$$\frac{d}{dx} \log f(x) = -\beta + A\beta e^{-\beta x} = 0,$$

that is, when  $x = \log A/\beta$ . Parameter  $\beta$  is commonly around  $\log_e 10 = 2.3$ , corresponding to a Gutenberg–Richter  $b$ -value of around 1. Thus, we should expect the mode to occur at approximately

$$\Delta \approx \frac{3}{2.3} \approx 1.29,$$

with similar values for the median and mean. Broadly similar behavior is retained even after mixing over  $A$ , providing that the mixture distribution is not too extreme. The quoted values are in satisfactory agreement with observed values for Båth's law.

4. Observed values of  $\alpha$  and  $\beta$  are often similar, but not necessarily equal. Dependence of foreshock probabilities on  $u$  do show the linear relation implicit in (12). Most large-scale studies suggest that  $\beta > \alpha$ . However, in fitting the ETAS model to individual sequences, situations where  $\alpha > \beta$  are not uncommon, even though this implies that the resulting model must be unstable if applied on a large scale.

## 5. Concluding remarks

The paper shows that a form of the extreme value theorem can be used to explain the appearance of a Båth's law feature under a wide range of conditions. Two features, the exponential distribution for magnitudes and the exponential productivity law for the numbers of events, seem crucial for the appearance of such a feature. The implication is that it is to these underlying features, rather than to Båth's law itself, that physical theories should be addressed.

## Acknowledgements

I am grateful to Annemarie Christofferson and Junko Murakami for their help and comments, and to Eugene Seneta, Kostya Borovkov, and an attentive anonymous referee for helping me to avoid at least some of the errors in an earlier draft.

## References

- CONSOLE, R., LOMBARDI, A. M. AND MURRU, M. (2003). Båth's law and the self-similarity of earthquakes. *J. Geophys. Res.* **108**, 2128.
- DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I, 2nd edn. Springer, New York.
- DALEY, D. J. AND VERE-JONES, D. (2007). *An Introduction to the Theory of Point Processes*, Vol. II, 2nd edn. Springer, New York.
- FELLER, W. (1966). *Introduction to Probability Theory and Its Applications*, Vol. 2. John Wiley, New York.
- FELZER, K. R., ABERCROMBIE, R. E. AND EKSTRÖM, G. (2004). A common origin for aftershocks, foreshocks and multiplets. *Bull. Amer. Seismological Soc.* **94**, 88–98.
- FELZER, K. R. *et al.* (2002). Triggering of the 1999  $M_W$  7.1 Hector Mine earthquake by aftershocks of the 1992  $M_W$  7.3 Landers earthquake. *J. Geophys. Res.* **107**, 1029.
- GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. Krieger, Florida.
- HELMSTETTER, A. AND SORNETTE, D. (2003). Båth's law derived from the Gutenberg–Richter law and from aftershock properties. *Geophys. Res. Lett.* **30**, 2069.

- JOHNSON, N., KOTZ, S. AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edn. John Wiley, New York.
- LOMBARDI, A. M. (2002). Probabilistic interpretation of Båth's law. *Ann. Geophys.* **45**, 455–472.
- OGATA, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9–27.
- SAICHEV, A. AND SORNETTE, D. (2005). Distribution of the largest aftershock in branching models of triggered seismicity: theory of Båth's law. *Phys. Rev. E* **71**, 056127.
- SHCHERBAKOV, R. AND TURCOTTE, D. L. (2004). A modified form of Båth's law. *Bull. Seismological Soc. Amer.* **94**, 1968–1975.
- UTSU, T. (1961). A statistical study on the properties of aftershocks. *Geophys. Mag.* **30**, 521–605.
- UTSU, T., OGATA, Y. AND MATU'URA, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth* **43**, 1–33.
- VERE-JONES, D. (1969). A note on the statistical interpretation of Båth's law. *Bull. Seismological Soc. Amer.* **59**, 1535–1541.
- VERE-JONES, D., MURAKAMI, J. AND CHRISTOPHERSEN, A. (2006). A further note on Båth's law. In *Proc. 4th International Workshop Statist. Seismology* (Zushi, Japan, January 2006), pp. 244–247.
- VERE-JONES, D., MURAKAMI, J. AND CHRISTOPHERSEN, A. (2008). A further note on Bath's law. (In preparation.)
- ZHUANG, J. AND OGATA, Y. (2006). Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. *Phys. Rev. E* **73**, 046134.