10

Predictability of extreme events in a branching diffusion model

*Andrei Gabrielov, Vladimir Keilis-Borok,
Sayaka Olsen, and Ilya Zaliapin*

10.1 Introduction

Extreme events are the most important yet least understood feature of natural and socioeconomic complex systems. In different contexts these events are also called *critical transitions, disasters, catastrophes*, or *crises*. Among examples are destructive earthquakes, El Niños, heat waves, electric power blackouts, economic recessions, stock-market crashes, pandemics, armed conflicts, and terrorism surges. Extreme events are rare, but consequential: they inflict the lion's share of the damage on population, economy, and environment. This chapter is focused on the theoretical foundations of predicting individual extreme events. The prediction problem is pivotal both for the fundamental understanding of complex systems and for disaster preparedness (e.g., Keilis-Borok and Soloviev, 2003; Sornette, 2004; Albeverio *et al*., 2005).

There exist several well-developed approaches to prediction of extreme events including classical Kolmogorov–Wiener extrapolation of time series (Kolmogorov, 1941a,b; Wiener, 1949), linear (Kalman–Bucy) (Kalman and Bucy, 1961) and non-linear (Kushner–Zakai) (Kushner, 1964; Zakai, 1969; Chow, 2007) filtering, sequential Monte-Carlo methods (Doucet *et al*., 2001), and the extreme-value theory (Embrechts *et al*., 2008). The approach to prediction described in this chapter is complementary to the above methods. The need for an alternative approach is dictated by a non-standard formulation of the prediction problem, where one is particularly interested in the future occurrence times of rare events rather than the complete unobserved state of the system in continuous time. We notice, accordingly, that often the easily observed extreme events cannot be defined as the instants of threshold exceedance by the observed physical or economical fields, like air temperature or asset price. A paradigmatic example is an earthquake initiation time,

which is determined by the complex interplay of stress and strength fields in the heterogeneous Earth lithosphere. The physical theory for spatio-temporal evolution of these fields is still in its infancy, their values can hardly be measured with the existing instruments, or predicted using the available statistical methods. At the same time, earthquakes are readily defined, measured, and studied.

Prediction here is based on analysis of the observable permanent background activity of the complex system. We look for *premonitory patterns*, i.e., particular deviations from long-term averages that emerge more frequently as an extreme event approaches. These patterns might be either *perpetrators* contributing to triggering an extreme event, or *witnesses* merely signalling that the system became unstable, ripe for a disaster. An example of a witness is the proverbial 'straws in the wind' preceding a hurricane.

The following types of premonitory patterns have been established by exploratory data analysis and numerical modelling: (i) increase of background activity; (ii) deviations from self-similarity, i.e. a change of the size distribution of events in favour of relatively strong yet subextreme events; (iii) increase of event clustering; and (iv) emergence of long-range correlations. Solid empirical evidence for the existence of these patterns in seismology and other forms of multiple fracturing has been accumulated since the 1970s (Haberman, 1981; Mogi, 1981; Aki, 1985; Keilis-Borok and Shebalin, 1999; Sykes *et al*., 1999; Keilis-Borok and Soloviev, 2003). Importantly, these patterns are universal, common for complex systems of distinctly different origin. Similar premonitory patterns have been observed in socio-economic systems (Keilis-Borok *et al*., 2000, 2005), dynamic clustering in elastic billiards (Gabrielov *et al*., 2008), hydrodynamics, and hierarchical models of extreme event development (Narkunskaya and

Shnirman, 1990; Newman *et al.*, 1995; Blanter *et al.*, 1997; Gabrielov *et al.*, 2000; Zaliapin *et al.*, 2003a,b). We discuss here a general mechanism that reproduces these universal premonitory patterns.

We focus in particular on *premonitory deviations from self-similarity*. Self-similarity is one of the most prominent features of complex systems. A canonical example is a power-law (self-similar) distribution of system observables, whose remarkable feature is the inevitability of extremely large events that dwarf numerous smaller events. The power-law distribution is well known under different names in such diverse phenomena as inertial-range self-similarity in turbulence (Kolmogorov–Obukhov laws) (Kolmogorov, 1941a; Obukhov, 1941; McWilliams, 1990; Frisch, 1996), energy released in an earthquake (Gutenberg–Richter law) (Gutenberg and Richter, 1944; Ben-Zion, 2003), word usage frequency in a language (Zipf law) (Zipf, 1965), allocation of wealth in a society (Pareto law) (Pareto, 1897; Klass *et al.*, 2006), war casualties (Richardson law) (Richardson, 1960), number of papers published by a given scientist (Lotka law) (Lotka, 1926), mass of a landslide (Malamud *et al.*, 2004; Brunetti *et al.*, 2009), stock price returns (Mandelbrot and Taylor, 1967; Plerou and Stanley, 2008; Gabaix *et al.*, 2003), number of species per genus (Burlando, 1990), and many others (Newman, 1997, 2005; Albert and Barabasi, 2002; Mandelbrot, 1983; Turcotte, 1997). An important paradigm of *self-organised criticality* (Bak *et al.*, 1988; Turcotte, 1999) that is demonstrated by sand-pile (Dhar, 1990), forest-fire (Drossel and Schwabl, 1992), and slider-block (Burridge and Knopoff, 1967; Olami *et al.*, 1992; Rundle and Klein, 1993) models and their numerous ramifications has been introduced in order to understand dynamic processes whose only attractor corresponds to self-similarity (criticality) of the size distribution of appropriately defined *events*.

Exact self-similarity, as well as many other universal properties, however, is only an approximation to (or a mean-field property of) the observed and modelled systems; at each particular time moment the distribution of event sizes deviates from a pure power-law form. We show in this chapter how to use such deviations for understanding the dynamics of a complex system in general and the occurrence of extreme events in particular.

The rest of the chapter is organised as follows. We informally outline our model and the corresponding prediction problem in Section 10.2. A formal model description is given in Section 10.3. Section 10.4 summarises the study's results most relevant to the prediction problem. Section 10.5 derives the spatio-temporal model distribution as a function of the control parameter. Section 10.6 uses these results to find spatio-temporal deviations of the event size distribution from its mean-field form. Results

of numerical experiments are illustrated in Section 10.7. In Section 10.8 we further discuss the relation of our results to prediction of extreme events. Proofs and necessary technical information are collected in Appendices.

10.2 Model outline and prediction problem

The model combines *external driving* ultimately responsible for the occurrence of events, including the extreme ones, a *cascade process* responsible for redistribution of energy (or another appropriate physical quantity such as mass, moment, stress, etc.) within the system, and *spatial dynamics*. We first outline the process of populating a system space Ω with *particles* of discrete ranks and then proceed with the definition of the *observation space* and *events*. We assume that Ω is an *n*-dimensional Euclidean space.

A direct cascade (branching) within a system starts with consecutive injection (immigration) of *particles* of the largest possible rank, r_{max}, into the origin $0 \in \Omega$, which we call the *source*. After injection, each particle diffuses freely and independently of the others across the space Ω. Eventually, it splits into a random number of particles of smaller rank, $r_{max} - 1$, each of which continues to diffuse from the location of the parent and independently of the other particles. These particles split in their turn into even smaller particles, and so on.

At each time instant $t \geq 0$, *observations* can be done on a subspace $\mathcal{R}_t \subset \Omega$. In this chapter we assume that $\mathcal{R}_t$ is an affine subspace of dimension $d < n$. An observed *event* corresponds to an instant when a particle crosses the subspace of observations. Each event is characterised by its occurrence time *t*, spatial location $\mathbf{x} \in \mathcal{R}_t$ within the observation space, and rank *r*. Model observations at instant *t* thus consist of a collection of events $\mathcal{C}_t = (t_i \leq t, \mathbf{x}_i, r_i)$, $i \geq 1$, referred to as the *catalogue*. An *extreme event* is defined as a sufficiently large, although not necessarily the largest, event, $r \geq r_0$, where r_0 is a rank threshold.

Importantly, the location of $\mathcal{R}_t$ within Ω is (a) unknown to an observer, and (b) time-dependent. One can interpret this as movement of the observation space relative to the source, movement of the source relative to the observation space, or a combination of the two. A principal goal of an observer is to assess the likelihood of the occurrence of an extreme event using the catalogue $\mathcal{C}_t$. It is readily seen that the probability of an extreme event increases as the observation space approaches the source and achieves its maximal value when the source belongs to the observation space, $0 \in \mathcal{R}_t$. The distance between the observation subspace and the source thus becomes a natural control parameter and allows one to reduce the prediction problem to estimating the distance to the source. This latter problem is the focus of the chapter.

As the observation subspace approaches the source, the intensity of the observed events increases, larger events become relatively more frequent, and clustering and long-range correlations become more prominent (see Fig. 10.1 and Sections 10.4, 10.8). The emergence of these patterns,

each individually and all together, can be therefore used to forecast an approach of a large event; indeed, such a prediction should be understood in a statistical sense. We focus on quantitative description of two of these patterns, intensity increase and deviations from self-similarity,
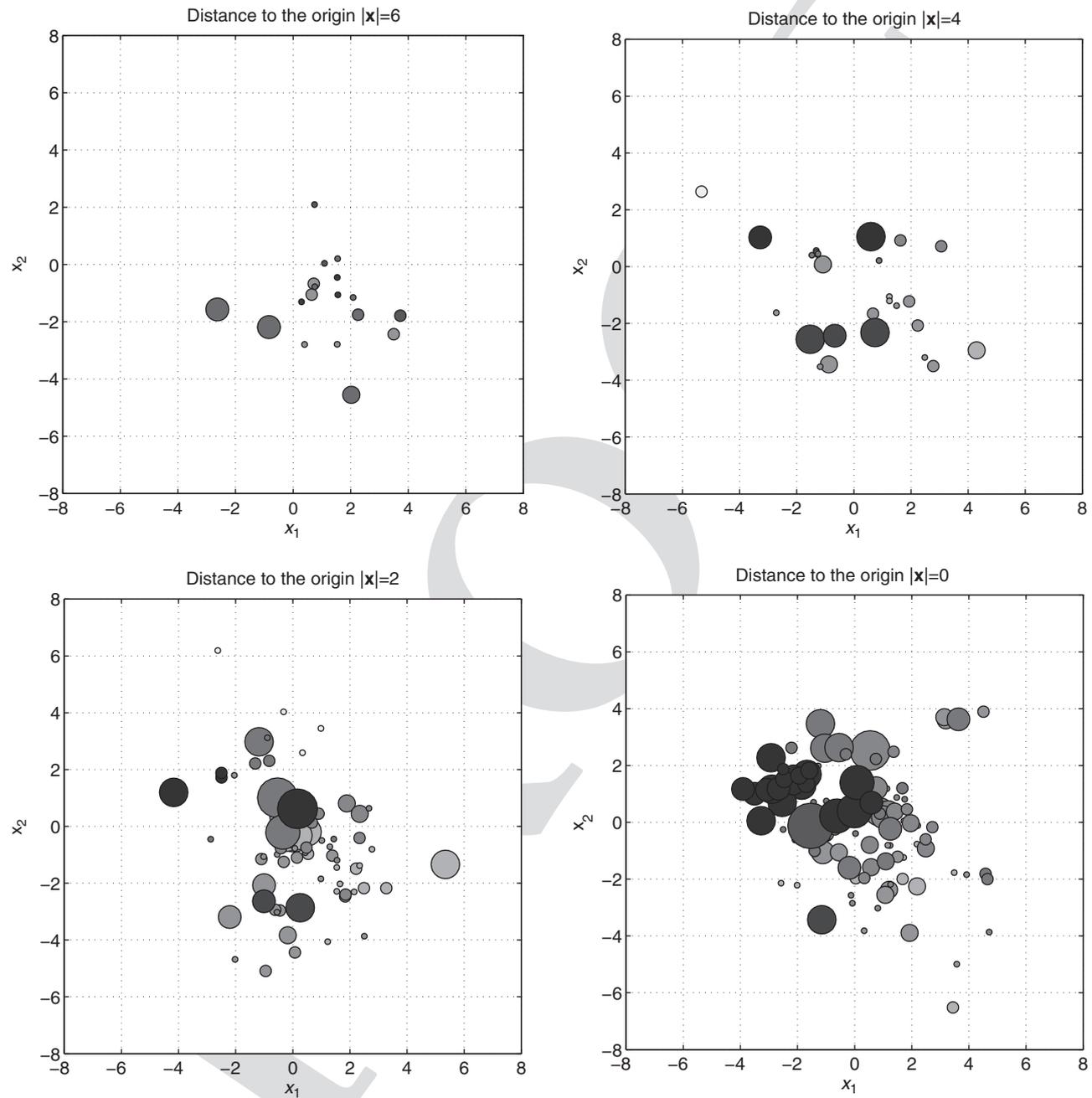
Figure 10.1. Example of a three-dimensional model population. Different panels show two-dimensional subspaces of the model three-dimensional space at different distances $|\mathbf{x}|$ to the origin. Model parameters are $\mu = \lambda = 1, D = 1, B = 2$. Circle size is proportional to the particle rank. Different shades correspond to populations from different immigrants; the descendants of earlier immigrants are a lighter shade. The clustering of particles is explained by the splitting histories. Note that, as the origin approaches, the particle activity significantly changes, indicating the increased probability of an extreme event.

for a classical branching process formally introduced in the next section.

We emphasise that the location and dynamics of the observation space $\mathcal{R}_t$ within Ω depend on details of a particular system of interest and may be hard to estimate or model. Important points of this chapter are that (i) the information about this unknown dynamics can be summarised by a scalar value of the control parameter (distance between the observational subspace and the origin); and (ii) knowledge of the control parameter is sufficient to solve the prediction problem.

Finally, it is important to mention that we do not use direct cascade as a dynamical model of event formation, which would imply that large events *cause* smaller ones. We merely use this analytically tractable approach to create a *hierarchical network* of spatially distributed particles. A dynamic interpretation of the latter will depend on a particular application, and may include inverse cascading or other physically relevant processes.

10.3 Model formulation

We consider an age-dependent multi-type branching diffusion process with immigration in $\mathbb{R}^n$. The system consists of particles, each of which belongs to a *generation* $k = 0, 1,\ldots$. Particles of zero generation (the largest ones) appear in a system as a result of external driving (forcing); we will refer to them as *immigrants*. Particles of any other generation $k > 0$ are produced as a result of the splitting of the particles of generation $k - 1$. Immigrants ($k = 0$) are born at the origin $\mathbf{x}:=(x_1,\ldots,x_n) = \mathbf{0}$ at a constant rate μ; that is, the probability for a new immigrant to appear within the time interval of length Δt is $\mu \Delta t + o(\Delta t)$ as $\Delta t \to 0$. Accordingly, the birth instants form a homogeneous Poisson process with intensity μ. Each particle lives for some random time τ and then transforms (splits) into a random number β of particles of the next generation. The probability laws of the lifetime τ and branching β are generation-, time-, and space-independent. We assume that new particles are born at the location of their parent at the moment of splitting.

The particle lifetime has an exponential distribution:

$$G(t) := \mathrm{P}\{\tau < t\} = 1 - \mathrm{e}^{-\lambda t}, \; \lambda > 0. \tag{10.1}$$

The conditional probability that a particle transforms into $k \geq 0$ new particles (0 means that it disappears) given that the transformation took place is denoted by p_k. The probability generating function (pgf) for the number β of new particles is thus

$$h(s) = \sum_k p_k \, s^k. \tag{10.2}$$

The expected number of offspring (also called the *branching number*) is $B := E(\beta) = h'(1)$ (e.g., Athreya and Ney (2004) Chapter 1).

Each particle diffuses in $\mathbb{R}^n$ independently of other particles. This means that the density $p(\mathbf{x}, \mathbf{y}, t)$ of a particle that was born at instant 0 at point $\mathbf{y}$ solves the equation

$$\frac{\partial p}{\partial t} = D \left(\sum_i \frac{\partial^2}{\partial x_i^2} \right) p \equiv D \triangle_{\mathbf{x}} p \tag{10.3}$$

with the initial condition $p(\mathbf{x}, \mathbf{y}, 0) = \delta(\mathbf{x} - \mathbf{y})$. The solution of (10.3) is given by Evans (1998):

$$p(\mathbf{x}, \mathbf{y}, t) = (4 \pi D t)^{-n/2} \exp\left\{ -\frac{|\mathbf{x} - \mathbf{y}|^2}{4 D t} \right\}, \; |\mathbf{x}|^2 = \sum_i x_i^2. \tag{10.4}$$

Accordingly, the density of each particle, given that it is alive at the instant t, is $\phi(\mathbf{x}, t) := p(\mathbf{x}, 0, t)$. Naturally, the positions of the particles produced by the same immigrant are correlated. This can be reflected by the joint distribution of pairs, triplets, etc.

The model is specified by the following parameters: immigration intensity $\mu > 0$, branching intensity $\lambda > 0$, diffusion constant $D > 0$, and branching distribution $\{p_k\}$, which will be often represented by its pgf $h(z)$ or simply by the branching number B. An appropriate choice of the temporal and spatial scales allows one to assume $\mu = 1$ and $D = 1$.

It is convenient to introduce particle *rank* $r := r_{max} - k$ for an arbitrary integer r_{max} and thus consider particles of ranks $r \leq r_{max}$. Particle rank can be considered a logarithmic measure of the size. Similar to the analysis of the real-world systems, we sometime only consider particles of the first several generations $0 \leq k \leq r_{max} - 1$, which corresponds to the largest ranks $1 \leq r \leq r_{max}$. Figure 10.1 illustrates the model population.

10.4 Prediction: summary

This section summarises important findings that are most relevant to the prediction problem. Recall that the prediction problem consists of assessing the likelihood of an extreme event; the latter corresponds to an instant when a sufficiently large particle crosses the observation space. The likelihood of an extreme event is thus directly related to the distance between the space of observations and the origin. Accordingly, the prediction problem is reduced to the estimation of this distance from available data. For that, one should look for increase in the intensity of medium-to-large-sized events, as well as upward deviations in the event size distribution. We believe that this general idea

Extreme Natural Hazards, Disaster Risks and Societal Implications

can be useful in a wide range of models and observed systems, not necessarily based on a branching diffusion mechanism. Statistical assessment of particular prediction schemes based on this idea is left for a future study.

All statements below refer to a steady state of the model (dynamics after a transient). All asymptotic statements have been confirmed numerically in finite models.

(1) *Meanfield self-similarity* Particle ranks, averaged over time and space, have an exponential distribution; this is equivalent to a power-law distribution of particle sizes; see (10.33) and Fig. 10.2.

(2) *Small-size self-similarity* The particle rank distribution at *any* spatial point is asymptotically exponential as rank decreases, with the exponent index $-B$; see (10.38) and Figs. 10.3 and 10.4. This is equivalent to a power-law distribution of particle sizes with power-law index $-B$. Furthermore, this implies that deviations from self-similarity, if any, can be only seen at large ranks (large particle sizes).

(3) *Upward deviations close to the origin* At any point sufficiently close to the origin, the particle size distribution deviates from a self-similar power-law form so as to have a larger number of medium-to-large-sized events. The magnitude of this deviation increases with the event size, as well as with dimension of the model space; see (10.36) and the upper lines in Figs. 10.3 and 10.4.

(4) *Downward deviations away from the origin* At any point sufficiently far from the origin, the particle size distribution deviates from a self-similar power-law
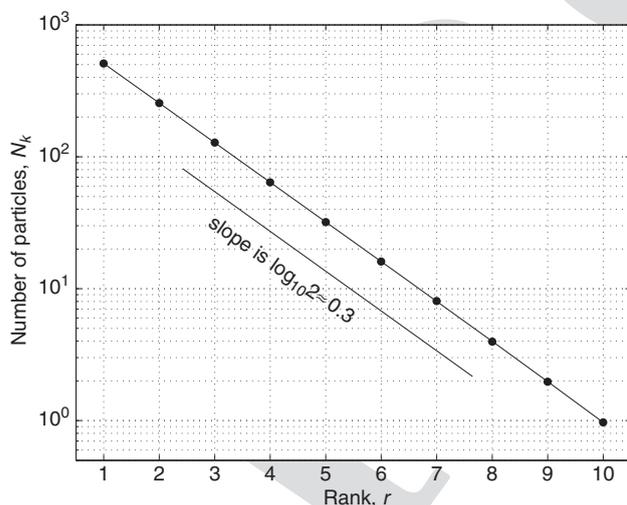

Figure 10.2. Spatially averaged particle rank distribution at $t = 30$. The distribution is averaged over 4000 independent realisations of a three-dimensional model with parameters $\mu = \lambda = 1$, $D = 1$, $B = 2$, $r_{\max} = 10$. One can clearly see the exponential rank distribution of Eq. (10.33).

form so as to have a smaller number of medium-to-large-sized events. The magnitude of this deviation increases with the event size and is independent of the model's dimension; see (10.37) and the lower lines in Figs. 10.3 and 10.4.

(5) *Exponential decay of event intensity* The intensity of events of any fixed size is exponentially decaying away from the origin; see (10.28).

(6) *Divergence of event intensity at the origin* For models with a spatial dimension larger than 1, the intensity of sufficiently large events diverges at the origin in a power-law fashion; see (10.28),(10.30) and Fig. 10.3(b)-(d).

10.5 Model solution: moment generating functions

The model introduced in Section 10.3 is a superposition of independent branching processes generated by individual immigrants. Sections 10.5.1 and 10.5.2 analyse, respectively, the one-point and two-point moments of a particle distribution produced by a single immigrant. Then we expand these results to the case of multiple immigrants in Section 10.5.3.

10.5.1 Single immigrant: one-point properties

10.5.1.1 *Moment generating functions*

Let $p_{k,i}(G, \mathbf{y}, t)$ be the conditional probability that at time $t \geq 0$ there exist $i \geq 0$ particles of generation $k \geq 0$ within spatial region $G \subset \mathbb{R}^n$ given that at time 0 a single immigrant was injected at point $\mathbf{y}$. The corresponding *moment generating function* is

$$M_k(G, \mathbf{y}, t; s) = \sum_i p_{k,i}(G, \mathbf{y}, t) e^{si}. \qquad (10.5)$$

Proposition 5.1 *The moment generating functions M_k (G, $\mathbf{y}$, t; s) solve the following recursive system of non-linear partial differential equations:*

$$\frac{\partial}{\partial t} M_k(G, \mathbf{y}, t; s) = D\Delta_{\mathbf{y}} M_k - \lambda M_k + \lambda\, h(M_{k-1}), \quad k \geq 1, \qquad (10.6)$$

with initial conditions M_k (G, $\mathbf{y}$, 0; s) $\equiv 1$, $k \geq 1$, and

$$M_0(G, \mathbf{y}, t; s) = (1 - P) + Pe^s, \; P := e^{-\lambda t}\int_G p(\mathbf{x}, \mathbf{y}, t)d\mathbf{x}. \qquad (10.7)$$

Here $h(s)$ is defined by (10.2) and $\Delta_{\mathbf{y}} = \sum_i \partial^2/\partial y_i^2$.

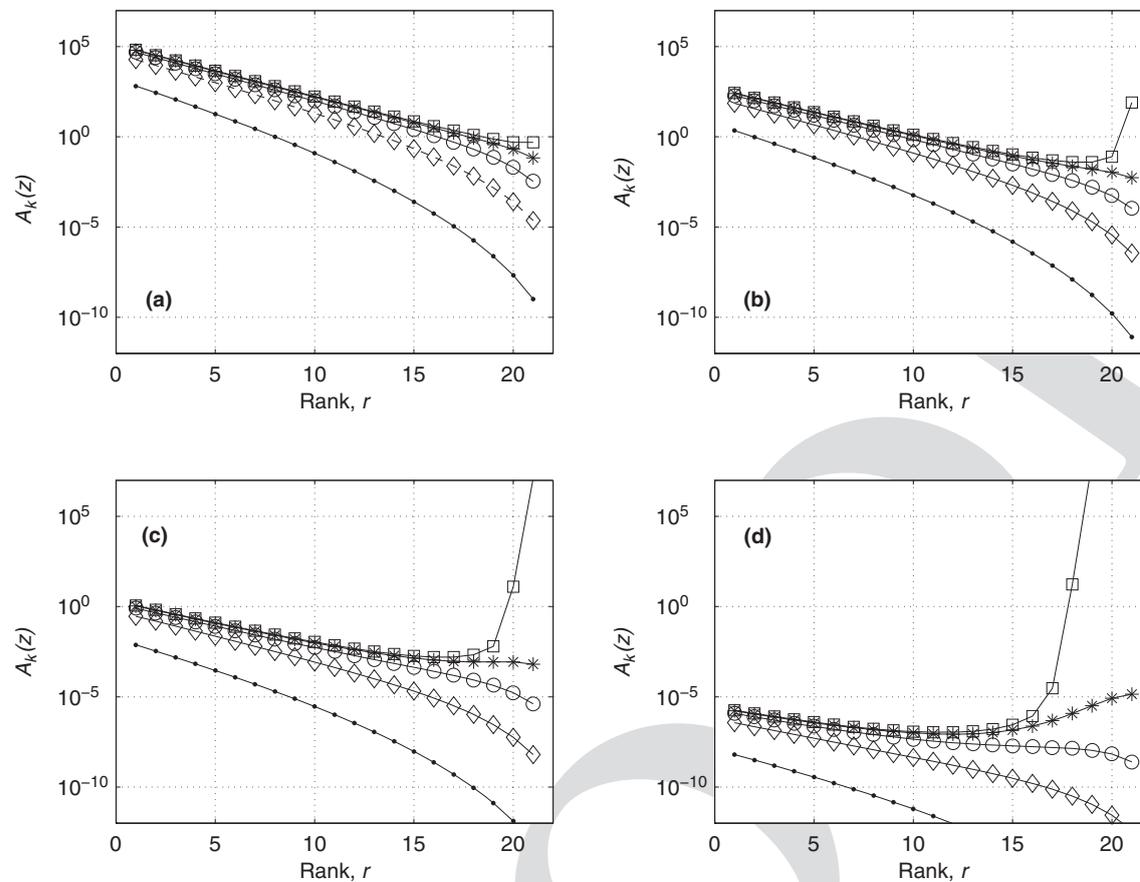**Proof** This is given in Appendix 10.A.

Figure 10.3. Expected number A_k (z) of generation-k particles at distance z from the origin (cf. Proposition 10.6.2). The distance z is increasing (from top to bottom line in each panel) as $z = 10^{-3}$, 2, 5, 10, 20. Model dimension is (a) $n = 1$, (b) $n = 3$, (c) $n = 5$, and (d) $n = 10$. Other model parameters: $\mu = \lambda = 1$, $D = 1$, $B = 2$, $r_{\max} = 21$.

10.5.1.2 *The first moment densities*

Let $\overline{A}_k(G, \mathbf{y}, t)$ be the expected number of generation-k particles at instant t within the region G, produced by a single immigrant injected at point $\mathbf{y}$ at time $t = 0$. It is given by the following partial derivative (e.g., Athreya and Ney 2004 Chapter 1):

$$\overline{A}_k(G, \mathbf{y}, t) := \left. \frac{\partial M_k(G, \mathbf{y}, t; s)}{\partial s} \right|_{s=0}. \tag{10.8}$$

Consider also the expectation density $A_k(\mathbf{x}, \mathbf{y}, t)$ that satisfies, for any $G \subset \mathbb{R}^n$,

$$\overline{A}_k(G, \mathbf{y}, t) = \int_G A_k(\mathbf{x}, \mathbf{y}, t) d\mathbf{x}. \tag{10.9}$$

Corollary 10.5.2. *The first moment densities* $A_k(\mathbf{x}, \mathbf{y}, t)$ *solve the following recursive system of partial differential equations:*

$$\frac{\partial A_k(\mathbf{x}, \mathbf{y}, t)}{\partial t} = D\Delta_{\mathbf{x}} A_k - \lambda A_k + \lambda B A_{k-1}, \ k \geq 1, \tag{10.10}$$

with the initial conditions $A_k(\mathbf{x}, \mathbf{y}, 0) \equiv 0, \ k \geq 1,$

$$A_0(\mathbf{x}, \mathbf{y}, 0) = \delta(\mathbf{y} - \mathbf{x}), \ A_0(\mathbf{x}, \mathbf{y}, t) = e^{-\lambda t} p(\mathbf{x}, \mathbf{y}, t), \ t > 0. \tag{10.11}$$

The solution to this system is given by

$$A_k(\mathbf{x}, \mathbf{y}, t) = \frac{(\lambda B t)^k}{k!} A_0(\mathbf{x}, \mathbf{y}, t)$$

$$= \frac{(\lambda B)^k}{k!(4\pi D)^{n/2}} t^{k-n/2} \exp\left\{-\lambda t - \frac{|\mathbf{x} - \mathbf{y}|^2}{4Dt}\right\}. \tag{10.12}$$

Proof This is given in Appendix 10.C. It follows from a general result for the higher moments obtained in Appendix 10.B.
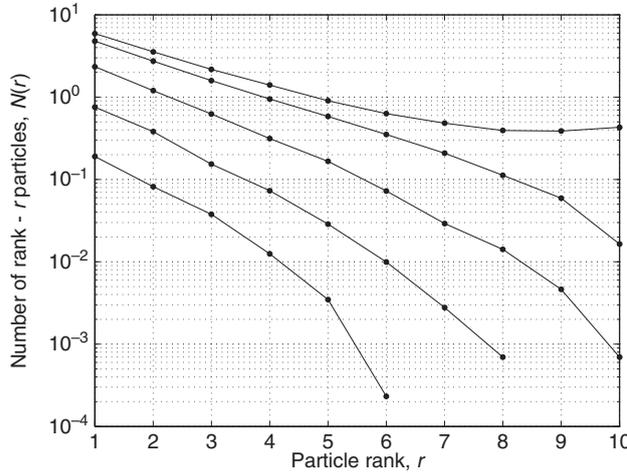
Figure 10.4. Particle rank distribution at $t = 30$ and fixed distance z from the origin (cf. Proposition 10.6.2). The distribution is averaged over 4000 independent realisations of a three-dimensional model with parameters $\mu = \lambda = 1$, $D = 1$, $B = 2$, $r_{\max} = 10$. Different lines correspond to different distances (from top to bottom): $z = 0, 2, 4, 6, 8$. One can clearly see that the rank distribution deviates from the pure exponential form, which corresponds to a straight line in the semilogarithmic scale used here. One observes *downward deviations* at large distances from the origin, and *upward deviations* close to the origin.

The system (10.10) has a transparent intuitive meaning. The rate of change of the expectation density $A_k(\mathbf{x}, \mathbf{y}, t)$ is affected by three processes: diffusion of the existing particles of generation k (the first term in the right hand side of (10.10)), splitting of the existing particles of generation k at the rate λ (the second term), and splitting of the generation $k - 1$ particles that produce on average B new particles of generation k (the third term).

To obtain the solution for the entire population, we sum up the contributions from all generations:

$$A(\mathbf{x}, 0, t) = \sum_{k=0}^{\infty} A_k(\mathbf{x}, 0, t) = e^{-\lambda t(1-B)} \, p(\mathbf{x}, 0, t)$$
$$= \frac{e^{-\lambda t(1-B)}}{(4 \pi D t)^{n/2}} \exp\left(-\frac{|\mathbf{x}|^2}{4 D t}\right). \quad (10.13)$$

This formula emphasises the role of the branching parameter B: in the subcritical case, $B < 1$, the population goes extinct exponentially; in the supercritical case, $B > 1$, the population grows exponentially; in the critical case, $B = 1$, the expected number of particles remains the same (steady state) and is given by the diffusion density $p(\mathbf{x}, 0, t)$.

10.5.2 Single immigrant: two-point properties

10.5.2.1 *Moment generating functions*
Let $p_{k_1,k_2,i,j}(G_1, G_2, \mathbf{y}, t)$ be the conditional probability that at instant $t \geq 0$ there exist $i \geq 0$ particles of generation $k_1 \geq 0$ within region $G_1 \subset \mathbb{R}^n$ and $j \geq 0$ particles of generation $k_2 \geq 0$ within region $G_2 \subset \mathbb{R}^n$ given that at time 0 a single immigrant was injected at point $\mathbf{y}$. Assume that G_1 and G_2 do not overlap. The corresponding moment generating function is

$$M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$= \sum_{i,j \geq 0} p_{k_1,k_2,i,j}(G_1, G_2, \mathbf{y}, t) e^{i\,s_1 + j\,s_2}. \quad (10.14)$$

Proposition 10.5.3 *The moment generating functions* $M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$ *solve the following recursive system of non-linear partial differential equations:*

$$\frac{\partial}{\partial t} M_{k_1,k_2} = D\Delta_{\mathbf{y}} M_{k_1,k_2} - \lambda M_{k_1,k_2} + \lambda h(M_{k_1-1,k_2-1}), k_1,k_2 \geq 1, \quad (10.15)$$

with the initial conditions

$$M_{k_1,k_2}(G_1, G_2, \mathbf{y}, 0; s_1, s_2) \equiv 1, \ k_1, k_2 \geq 1, \quad (10.16)$$
$$M_{0,0}(G_1, G_2, \mathbf{y}, t; s_1, s_2) = P_1 e^{s_1} + P_2 e^{s_2} + 1 - P_1 - P_2, \quad (10.17)$$

$$M_{0,k}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$= (M_k(G_2, \mathbf{y}, t; s_2) - e^{-\lambda t}) + (e^{-\lambda t} - P_1) + P_1 e^{s_1}, \quad (10.18)$$

where $P_i := e^{-\lambda t} \int_{G_i} p(\mathbf{x}, \mathbf{y}, t) d\mathbf{x}$, $i = 1,2$. *Here, as before,* $h(s)$ *is defined by* (10.2) *and* $\Delta_{\mathbf{y}} = \sum_i \partial^2/\partial y_i^2$.

Proof This is given in Appendix 10.D.

10.5.2.2 *Moments*
Consider the expected value $\overline{A}_{k_1,k_2}(G_1, G_2, \mathbf{y}, t)$ of the product of the number of generation-k_1 particles in region G_1 and the number of generation-k_2 particles in region G_2 at instant t, produced by a single immigrant injected at point $\mathbf{y}$ at time $t = 0$. It is given by the following partial derivative:

$$\overline{A}_{k_1,k_2}(G_1, G_2, \mathbf{y}, t) := \frac{\partial^2 M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)}{\partial s_1 \partial s_2}\bigg|_{s_1=s_2=0}. \quad (10.19)$$

We notice that the expectations $\overline{A}_{k_1}(G_1, \mathbf{y}, t)$ and $\overline{A}_{k_2}(G_2, \mathbf{y}, t)$ of (10.8) can be represented as

$$\overline{A}_{k_1}(G_1, \mathbf{y}, t) := \left.\frac{\partial M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)}{\partial s_1}\right|_{s_1=s_2=0}$$
(10.20)

and

$$\overline{A}_{k_2}(G_2, \mathbf{y}, t) := \left.\frac{\partial M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)}{\partial s_2}\right|_{s_1=s_2=0}.$$
(10.21)

Consider also the expectation density $A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)$ that satisfies, for any non-overlapping $G_1, G_2 \subset \mathbb{R}^n$,

$$\overline{A}_{k_1,k_2}(G_1, G_2, \mathbf{y}, t) = \int_{G_2} \int_{G_1} A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)\mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2.$$
(10.22)

Corollary 10.5.4 *The moment densities $A_{k_1,k_2} \equiv A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)$ solve the following recursive system of partial differential equations:*

$$\frac{\partial A_{k_1,k_2}}{\partial t} = D\,\Delta_{\mathbf{y}}A_{k_1,k_2} - \lambda\,A_{k_1,k_2} + \lambda\,B\,A_{k_1-1,k_2-1}$$

$$+\lambda\,h''(1)\,A_{k_1-1}(\mathbf{x}_1)\,A_{k_2-1}(\mathbf{x}_2),\quad k_1, k_2 \geq 1,$$
(10.23)

with the initial conditions

$$A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, 0) \equiv 0,\quad k_1, k_2 \geq 1,$$
(10.24)

$$A_{0,k}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t) \equiv 0,\quad k \geq 0,\ t \geq 0,$$
(10.25)

and $A_k(\mathbf{x}) \equiv A_k(\mathbf{x}, \mathbf{y}, t)$ given by (10.12).

Proof This is given in Appendix 10.E.

10.5.3 Multiple immigrants
Here we expand the results of the Section 10.5.1 to the case of multiple immigrants that appear at the origin according to a homogeneous Poisson process with intensity μ. The expectation $\mathcal{A}_k$ of the number of particles of generation k is given by

$$\mathcal{A}_k(\mathbf{x}, t) = \int_0^t A_k(\mathbf{x}, 0, s)\,\mu\,\mathrm{d}s$$

$$= \frac{\mu\,(\lambda\,B)^k}{k!(4\,\pi\,D)^{n/2}}\int_0^t s^{k-n/2}\,\exp\left\{-\lambda\,s - \frac{|\mathbf{x}|^2}{4\,D\,s}\right\}\mathrm{d}s.$$
(10.26)

The steady-state spatial distribution corresponds to the limit $t \to \infty$:

$$\mathcal{A}_k(\mathbf{x}) := \mathcal{A}_k(\mathbf{x}, \infty) = \frac{2\,\mu\,(\lambda\,B)^k}{k!(4\,\pi\,D)^{n/2}}\left(\frac{|\mathbf{x}|^2}{4\,D\,\lambda}\right)^{v/2}K_v\left(|\mathbf{x}|\sqrt{\frac{\lambda}{D}}\right).$$
(10.27)

Here $v = k - n/2 + 1$ and K_v is the modified Bessel function of the second kind (see Appendix 10.G). Introducing the normalised distance from the origin $z := |\mathbf{x}|\sqrt{\lambda/D}$ we obtain

$$\mathcal{A}_k(z) = \frac{\mu}{\lambda\,k!}\left(\frac{B}{2}\right)^k\left(\frac{2\,\pi\,D}{\lambda}\right)^{-n/2}z^v\,K_v(z).$$
(10.28)

For odd n, there are explicit expressions for $K_v(z)$ (Appendix 10.G, (10.G2),(10.G3)). In particular, we have

$$\mathcal{A}_0(z) = \frac{\mu}{\sqrt{4\,D\,\lambda}}\mathrm{e}^{-z}, \text{ for } n = 1,$$

$$\mathcal{A}_0(z) = \sqrt{\frac{\lambda}{D^3}}\,\frac{\mu}{4\,\pi\,z}\mathrm{e}^{-z}, \text{ for } n = 3.$$
(10.29)

From (10.28) and the asymptotic behaviour of $K_v(z)$ as $z \to 0$ (Appendix 10.G, (10.G5)) it follows that

$$\lim_{z\to 0} A_k(z) = \begin{cases} \infty, & \text{for } v \leq 0,\ \text{i.e.,}\ k \leq n/2 - 1 \\ \text{const} < \infty, & \text{for } v > 0,\ \text{i.e.,}\ k > n/2 - 1. \end{cases}$$
(10.30)

Thus, in a model with spatial dimension $n \geq 2$, the elements of several of the lowest generations ($k \leq n/2 - 1$) have an infinite concentration at the origin.

10.5.4 Alternative model representation
In this section we derive a system of equations for the steady-state expectations $\mathcal{A}_k(\mathbf{x})$ using the radial symmetry of the problem. By integrating (10.10) from $t = 0$ to ∞, we obtain

$$D\Delta_{\mathbf{x}}\mathcal{A}_k(\mathbf{x}) - \lambda\mathcal{A}_k(\mathbf{x}) + \lambda B\mathcal{A}_{k-1}(\mathbf{x}) = 0,$$

since $A_k(\mathbf{x}, \mathbf{y}, \infty) = 0$. We now rewrite this equation in terms of the normalised distance from the origin, $z := |\mathbf{x}|\sqrt{\lambda/D}$, using the fact that $\mathcal{A}_k(\mathbf{x}) \equiv \mathcal{A}_k(z)$ as soon as $|\mathbf{x}| = |z|$:

$$\mathcal{A}_k''(z) + \frac{n-1}{z}\,\mathcal{A}_k'(z) - \mathcal{A}_k(z) + B\mathcal{A}_{k-1}(z) = 0.\quad (10.31)$$

We notice, furthermore, that one can rewrite the expectation densities (10.12) as a function of z, which results in $A_k(z) \equiv A_k(\mathbf{x}, 0, t)$. It is then readily seen that

$$A_k'(z) = -\frac{B}{2k}\,z\,A_{k-1}(z).$$
(10.32)

The same recursive system holds for $\mathcal{A}_k(z)$, which is shown by integrating the last equation with respect to time.

10.6 Particle rank distribution

We analyse here the particle rank distribution; recall that the rank is defined as $r = r_{max} - k$, where k is the particle's generation. A self-similar branching mechanism that governs the model suggests an exponential distribution of particle ranks. Indeed, the spatially averaged steady-state rank distribution is a pure exponential law with index B:

$$A_k := \int_{\mathbb{R}^n} \int_0^\infty A_k(\mathbf{x}, 0, t) \mu \, \mathrm{d}t \, \mathrm{d}\mathbf{x}$$

$$= \frac{\mu B^k}{k!} \int_0^\infty (\lambda t)^k \mathrm{e}^{-\lambda t} \, \mathrm{d}t = \frac{\mu}{\lambda} B^k \propto B^{-r}. \quad (10.33)$$

Remark 10.6.1 *The use of the term 'self-similar' with respect to the exponential distribution, often seen in physical literature, requires some explanation. As we mentioned earlier, the particle rank serves as a logarithmic measure of its size. Thus, the exponential distribution of ranks corresponds to the power-law distribution of sizes; hence the term 'self-similarity'.*

To analyse rank- and space-dependent deviations from the pure exponential distribution, we consider the ratio $\gamma_k(\mathbf{x})$ between the number of particles of two consecutive generations:

$$\gamma_k(\mathbf{x}) := \frac{\mathcal{A}_k(\mathbf{x})}{\mathcal{A}_{k+1}(\mathbf{x})}. \quad (10.34)$$

For the purely exponential rank distribution, $A_k(\mathbf{x}) = c\, B^k$, the value of $\gamma_k(\mathbf{x}) = 1/B$ is independent of k and $\mathbf{x}$; while deviations from the pure exponential distribution cause γ_k to vary as a function of k and/or $\mathbf{x}$. Plugging (10.28) into (10.34) we find

$$\gamma_k(\mathbf{x}) = \frac{2(k+1)}{B z} \frac{K_v(z)}{K_{v+1}(z)}, \quad (10.35)$$

where, as before, $z := |\mathbf{x}| \sqrt{\lambda/D}$ and $v = k - n/2 + 1$.

Proposition 10.6.2 *The asymptotic behaviour of the function $\gamma_k(z)$ is given by*

$$\lim_{z \to 0} \gamma_k(z) = \begin{cases} \infty, & v \le 0, \\ \dfrac{1}{B}\left(1 + \dfrac{n}{2v}\right), & v > 0, \end{cases} \quad (10.36)$$

$$\gamma_k(z) \sim \frac{2(k+1)}{B z}, \ z \to \infty, \ \text{fixed } k, \quad (10.37)$$

$$\gamma_k(z) \sim \frac{1}{B}\left(1 + \frac{n}{2v}\right), \ k \to \infty, \ \text{fixed } z. \quad (10.38)$$

Proof This and the explicit rates of divergence in (10.6.4) are given in Appendix 10.F.

Proposition 10.6.2 describes the spatio-temporal deviations of the particle rank distribution from the pure exponential law (10.33). We interpret below each of (10.36)–(10.38) in some detail. Equation (10.38) implies that at any spatial point, the distribution asymptotically approaches the exponential form as generation k increases (rank r decreases). In other words, the distribution of small ranks (large generation numbers) is close to the exponential with index $-B$; thus the deviations can only be observed at the largest ranks (small generation numbers). Analysis of the large-rank distribution is done using (10.36) and (10.37). Near the origin, where the immigrants enter the system, (10.36) implies that $\gamma_k(z) > \gamma_{k+1}(z) > 1/B$ for $v > 0$. Hence, one observes the *upward deviations* from the pure exponential distribution: for the same number of rank r particles, the number of rank $r + 1$ particles is larger than predicted by the exponential law. The same behaviour is in fact observed for $v \le 0$ (see Appendix 10.F, (10.F5)). In addition, for $v \le 0$ the ratios $\gamma_k(z)$ do not merely deviate from $1/B$, but diverge to infinity at the origin. Away from the origin, according to (10.37), we have $\gamma_k(z) < \gamma_{k+1}(z) < 1/B$, which implies *downward deviations* from the pure exponent: for the same number of rank r particles, the number of rank $r + 1$ particles is smaller than predicted by the exponential law.

Figure 10.3 illustrates the above findings; it shows the distribution of particles for the largest ranks at different distances from the origin. One can clearly see the transition from downward to upward deviation of the rank distributions from the pure exponential form as we approach the origin. Notably, the magnitude of the upward deviation close to the origin (the upper line in all panels) strongly increases with the model dimension n.

10.7 Numerical analysis

Our analytical results and asymptotics are closely reproduced in numerical experiments with a finite number of generations, limited spatial extent, and spatial averaging (unavoidable when working with observations). Here, to mimic the ensemble averaging, the numerical results have been averaged over 4000 independent realisations of a three-dimensional model with parameters $\mu = \lambda = 1$, $D = 1$, and $B = 2$.

First, we check the exponential rank distribution of (10.33). Figure 10.2 shows the observed spatially averaged particle rank distribution. The exponential form (10.33) is indeed well reproduced.

Next, we see how the spatial averaging affects the rank distribution. Figure 10.4 shows the rank distribution at $t = 30$ at various distances to the origin. The spatial averaging has been done within spherical shells (space between two concentric spheres) of a constant volume $V = 5$ with inner radius z. Thus, here we see an observable counterpart of the theoretical distributions shown in Fig. 10.3(b). Although the spatial averaging somewhat tapers off the upward bend at the largest ranks close to the origin, the predicted transition from the downward to upward bend is clearly seen.

Figure 10.5 illustrates in more detail how the spatial averaging affects the upward bend in a three-dimensional model. It shows the particle rank distributions at $t = 30$ spatially averaged over spheres of different volumes centred at the origin. The upward bend is prominent for the spheres with volumes $V \leq 5$; and it gradually disappears within larger spheres in favour of an exponential distribution observed after a complete spatial averaging. Notably, the pure exponential distribution can only be achieved by averaging over *all* events in the model ($V = \infty$).
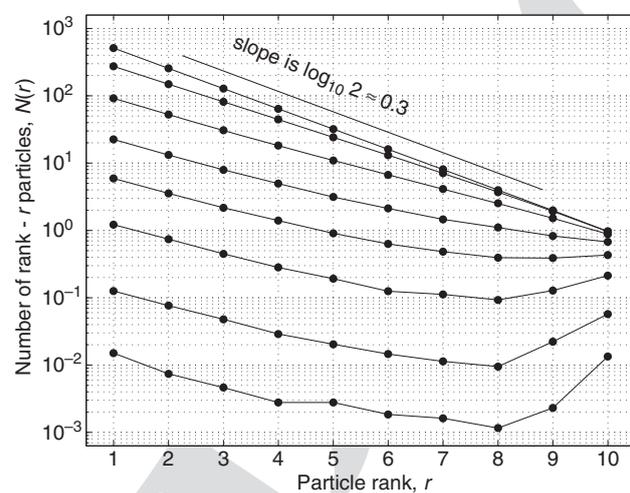

Figure 10.5. Particle rank distribution at $t = 30$ in a three-dimensional model. The distribution is spatially averaged over spheres of volume V centred at the origin with (from top to bottom): $V = \infty$, 500, 100, 200, 5, 1, 0.01. Model parameters: $\mu = \lambda = 1$, $D = 1$, $B = 2$, $r_{max} = 10$. The upward deviations from the exponential distribution (a straight line) are fading away with the extent of the spatial averaging.

10.8 Discussion

The problem of predicting extreme events in complex systems is challenging. This chapter proposes a simple mechanism and a single control parameter for premonitory patterns that has been reported in the literature.

Quantitative analysis is performed here for a classical model of a spatially distributed population of particles of different sizes governed by direct cascade of branching and external driving (see Section 10.3). In probability theory this model is known as the age-dependent multi-type branching diffusion process with immigration (Athreya and Ney, 2004). This chapter discusses a new scope of problems for this model. We assume that observations (detection of particles) are only possible on a subspace of the system space while the source of external driving (origin) remains unobservable, as is the case in many real-world systems. The natural question under this approach concerns the dependence of the process statistics on the distance to the source. A complete analytical solution to this problem, in terms of the moments with respect to the particle density, is given by Proposition 10.5.1. In addition, the correlation structure of the particle field can be found using Proposition 10.5.3.

It is natural to consider rank as a logarithmic measure of the particle size. The exponential rank distribution derived in (10.33) corresponds to a self-similar, power-law distribution of particle sizes, characteristic for many complex systems. The self-similarity in the considered model, as well as in the real-world systems, is only observed after global spatial averaging in a steady state. Proposition 10.6.2 and Fig. 10.3 describe space-dependent deviations from the self-similarity. Recall that an extreme event in the examined system is defined as an observation of a particle of sufficiently large size. As the source approaches the observation subspace, the probability of an extreme event increases. These results are thus directly connected to prediction: when the location of the source changes in time and approaches the subspace of observation (or vice versa), the increase of event intensity and the downward bend in the event size distribution becomes premonitory to an extreme event. The numerical experiments confirm the validity of the analytical results and asymptotics in a finite model.

The examined model exhibits very rich and intriguing premonitory behaviour. Figure 10.1 shows several two-dimensional snapshots of a three-dimensional model at different distances from the source. One can see that, as the source approaches, the following changes in the background activity emerge: (a) the intensity (total number of particles) increases; (b) particles of larger size become relatively more numerous; (c) particle clustering becomes more prominent; (d) the correlation radius

increases. All these premonitory changes have been independently observed in natural and socioeconomic systems. Here they are all determined by a single control parameter – the distance between the source and the observation space.

The above-mentioned premonitory patterns closely resemble universal properties of models of statistical physics in the vicinity of second order phase transition (Stanley, 1971; Ma, 2000; Kadanoff, 2000), percolation models near the percolation threshold (Stauffer and Aharony, 1994; Grimmett, 1999), and random graphs prior to the emergence of a giant cluster (Bollobás, 2001; Durrett, 2006; Newman *et al.*, 2006). In these models, the approach of an extreme event, usually referred to as a critical point, and the emergence of premonitory patterns, called critical phenomena, correspond to an instant when a control parameter crosses its critical value. In statistical physics a typical control parameter is temperature or magnetisation; in percolation it is the site or bond occupation density; in a random graph it is the probability for two vertices to be connected. The theory of critical phenomena (Ma, 2000) quantifies a system's behaviour at the critical value of the corresponding control parameter. The remarkable power of this theory is connected to the fact that very different systems demonstrate similar behaviour near to criticality. More precisely, when the control parameter is close to its critical value, the system sticks to one of just a few types of possible limit behaviours, each being described by an appropriate scale-invariant statistical field theory. In particular, each limit behaviour corresponds to the asymptotic power-law size distribution of system observables with a characteristic value of critical exponent.

We have focused in this chapter on a problem inverse to that considered by the critical phenomena theory: estimating the deviation of a control parameter from the critical value using the observed system behavior. The motivation for this comes from environmental, geophysical, and other applied fields where one faces the problem of assessing the likelihood of the occurrence of an extreme event associated with a critical point. We have formulated and solved such a prediction problem for a spatially embedded cascade process, which enjoys both mean-field self-similarity and realistic premonitory time- and space-dependent deviations from the latter. The methods presented in this chapter may provide a framework for studying predictability of extreme events in complex systems of arbitrary nature.

Appendix 10.A

PROOF OF PROPOSITION 10.5.1

We will need the following calculus lemma that is readily proven by using the definition of derivative:

Lemma 10.A.1 *Let $f(z)$, $g(z)$, $z \in \mathbb{R}$ be continuous functions such that the definite integral $G(t) = \int_0^t f(z) \, g(t-z)\mathrm{d}z$ exists. We also assume that $g(z)$ is differentiable. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}t}G(t) = \int_0^t f(z) \, g'(t-z)\mathrm{d}z + f(t) \, g(0).$$

There are two possible scenarios for the model development up to time t. In the first one, the initial immigrant will not split; the probability for this is $P = e^{-\lambda t}$. In the second one, the initial immigrant will split at instant $0 \le u \le t$; the probability of the first split within the time interval $[u, u + \mathrm{d}u]$ is $\lambda e^{-\lambda t} \, du + o(du)$ as $\mathrm{d}u \to 0$. The spatial position of the split is given by the diffusion density $p(\mathbf{x}, \mathbf{y}, u)$. If the immigrant splits, the composition property of generating functions gives $M_k = h[M_{k-1}]$. Integrating over all possible split instants and locations, we obtain

$$M_k(G, \mathbf{y}, t; s) = e^{-\lambda t} + \int_{\mathbb{R}^n} \mathrm{d}\mathbf{y}' \int_0^t \mathrm{d}u \, \lambda e^{-\lambda u} p(\mathbf{y}', \mathbf{y}, u)$$
$$h[M_{k-1}(G, \mathbf{y}', t-u; s)]. \qquad (10.\text{A}1)$$

Here the first and the second terms correspond to the first and second scenarios, respectively. Using the new integration variable $z = t - u$, we write

$$M_k(G, \mathbf{y}, t; s) = e^{-\lambda t} + e^{-\lambda t} \int_{\mathbb{R}^n} \mathrm{d}\mathbf{y}' \int_0^t \mathrm{d}u \, \lambda e^{\lambda(t-u)} p(y', \mathbf{y}, u)$$
$$h[M_{k-1}(G, y', t-u; s)]$$
$$= e^{-\lambda t} \left(1 + \int_{\mathbb{R}^n} \mathrm{d}y' \int_0^t \mathrm{d}z \, \lambda e^{\lambda z} p(y', \mathbf{y}, t-z) \right.$$
$$\left. h[M_{k-1}(G, y', z; s)] \right).$$

Now we take the derivative with respect to t of both sides and apply Lemma 10.A.1 using the fact that $p(\mathbf{y}', \mathbf{y}, 0) = \delta(\mathbf{y}' - \mathbf{y})$ and $(\partial/\partial t - D\Delta_{\mathbf{y}})p = 0$:

$$\frac{\partial}{\partial t}M_k(G, \mathbf{y}, t; s) = -\lambda M_k(G, \mathbf{y}, t; s)$$

$$+ e^{-\lambda t} \left[\int_{\mathbb{R}^n} \mathrm{d}y' \int_0^t \mathrm{d}z \, e^{\lambda z} \, h[M_{k-1}(G, y', z; s)] \right.$$
$$\left. D\Delta_{\mathbf{y}} p(y', \mathbf{y}, t-z) + \lambda e^{\lambda t} \, h[M_{k-1}(G, \mathbf{y}, t; s)] \right].$$

Taking the operator $\Delta_{\mathbf{y}}$ out of the integration signs, we find

$$\frac{\partial}{\partial t} M_k(G, \mathbf{y}, t; s) = D\Delta_{\mathbf{y}} M_k - \lambda M_k + \lambda\, h[M_{k-1}].$$

It is left to establish the initial conditions. Since we start the model with a particle of generation $k = 0$ and the distribution of splitting is continuous, at $t = 0$ there are no other particles with probability 1. Hence, $M_k(G, \mathbf{y}, 0; s) = 1$ for all $k \geq 1$. For generation $k = 0$, we can only have 1 or no particles at time $t > 0$. The probability to have 1 particle is given by the product of probabilities that there was no split up to time t and that the particle happens to be within region G at time t: $P = \mathrm{e}^{-\lambda t} \int_G p(\mathbf{x}, 0, t)\mathrm{d}\mathbf{x}$. The probability to have no particles is then $(1 - P)$. This implies (10.7).

Appendix 10.B

MOMENTS IN ONE-POINT SYSTEM

For any natural number j, consider the jth moment $\overline{A}_k^{(j)}(G, \mathbf{y}, t)$ of the number of generation-k particles at instant t within the region G, produced by a single immigrant injected at point $\mathbf{y}$ at time $t = 0$. It is given by the following partial derivative (e.g., Athreya and Ney, 2004, Chapter 1):

$$\overline{A}_k^{(j)}(G, \mathbf{y}, t) := \frac{\partial^j M_k(G, \mathbf{y}, t; s)}{\partial s^j}\bigg|_{s=0}. \qquad (10.B1)$$

Corollary 10.B.1 *The moments* $\overline{A}_k^{(j)}(G, \mathbf{y}, t)$ *solve the following recursive system of partial differential equations:*

$$\frac{\partial}{\partial t}\overline{A}_k^{(j)}(G, \mathbf{y}, t) = D\Delta_{\mathbf{y}}\overline{A}_k^{(j)} - \lambda\overline{A}_k^{(j)}$$
$$+ \lambda\left[\sum \frac{j!}{m_1! m_2! \ldots m_j!} h^{(m)}(1) \prod_{i=1}^{j}\left(\frac{\overline{A}_{k-1}^{(i)}}{i!}\right)^{m_i}\right], \qquad (10.B2)$$

where $m = m_1 + \cdots + m_j$ *and the sum is over all partitions of* j, *i.e., values of* $m_1, \ldots, m_j$ *such that* $m_1 + 2m_2 + \cdots + jm_j = j$, *with the initial conditions*

$$\overline{A}_k^{(j)}(G, \mathbf{y}, 0) \equiv 0, k \geq 1, \qquad (10.B3)$$

$$\overline{A}_0^{(j)}(G, \mathbf{y}, 0) = \int_G \delta(\mathbf{y} - \mathbf{x})\mathrm{d}\mathbf{x}, \qquad (10.B4)$$

$$\overline{A}_0^{(j)}(G, \mathbf{y}, t) = \mathrm{e}^{-\lambda t}\int_G p(\mathbf{x}, \mathbf{y}, t)\mathrm{d}\mathbf{x}, \ t > 0, \qquad (10.B5)$$

and

$$h^{(i)}(1) := \frac{d^i}{ds^i} h(s)\bigg|_{s=1} = \sum_{n=i}^{\infty} \frac{n!}{(n-i)!}\, p_n.$$

Proof The validity of (10.B2) follows from Proposition 10.5.1. Namely, applying the operator $\partial^j/\partial s^j(\cdot)|_{s\,=\,0}$ to both sides of (10.6), changing the order of differentiation, and using Faà di Bruno's formula (Faá di Bruno, 1855) for the jth derivative of a composition function, one finds, for each $k \geq 1$,

$$\frac{\partial^j}{\partial s^j}\left[\frac{\partial M_k(G, \mathbf{y}, t; s)}{\partial t}\right]\bigg|_{s=0}$$
$$= \frac{\partial^j}{\partial s^j}\left[D\Delta_{\mathbf{y}} M_k - \lambda M_k + \lambda\, h(M_{k-1})\right]\bigg|_{s=0},$$

$$\frac{\partial}{\partial t}\left[\frac{\partial^j M_k(G, \mathbf{y}, t; s)}{\partial s^j}\bigg|_{s=0}\right]$$
$$= \left[D\Delta_{\mathbf{y}}\frac{\partial^j M_k}{\partial s^j} - \lambda\frac{\partial^j M_k}{\partial s^j} + \lambda\,\frac{\partial^j}{\partial s^j} h(M_{k-1})\right]\bigg|_{s=0},$$

$$\frac{\partial}{\partial t}\overline{A}_k^{(j)}(G, \mathbf{y}, t) = \left[D\Delta_{\mathbf{y}}\frac{\partial^j M_k}{\partial s^j} - \lambda\frac{\partial^j M_k}{\partial s^j}\right.$$
$$+ \lambda\left(\sum\frac{j!}{m_1! m_2! \cdots m_j!} h^{(m)}(M_{k-1})\right.$$
$$\left.\left.\prod_{i=1}^{j}\left(\frac{M_{k-1}^{(i)}}{i!}\right)^{m_i}\right)\right]\bigg|_{s=0}$$
$$= D\Delta_{\mathbf{y}}\overline{A}_k^{(j)} - \lambda\overline{A}_k^{(j)} + \lambda\left[\sum\frac{j!}{m_1! m_2! \cdots m_j!} h^{(m)}(1)\right.$$
$$\left.\prod_{i=1}^{j}\left(\frac{\overline{A}_{k-1}^{(i)}}{i!}\right)^{m_i}\right],$$

where $m = m_1 + \cdots + m_j$ and the sum is over all partitions of j, i.e., values of $m_1, \ldots, m_j$ such that $m_1 + 2m_2 + \cdots + jm_j = j$. The initial conditions are established by applying the operator $\partial^j/\partial s^j(\cdot)|_{s=0}$ to both sides of (10.7) and using the definition of $\overline{A}_k^{(j)}(G, \mathbf{y}, t)$ in (10.8).

Appendix 10.C

PROOF OF COROLLARY 10.5.2

For $j = 1$, the equation in Corollary 10.B.1 simplifies to

$$\frac{\partial}{\partial t}\overline{A}_k(G, \mathbf{y}, t) = D\Delta_\mathbf{y}\overline{A}_k - \lambda\overline{A}_k + \lambda B\overline{A}_{k-1}.$$

Using the definition of A_k $(\mathbf{x}, \mathbf{y}, t)$ given in (10.9), one obtains for each $k \geq 1$,

$$\frac{\partial}{\partial t}A_k(\mathbf{x}, \mathbf{y}, t) = D\Delta_\mathbf{y}A_k - \lambda A_k + \lambda BA_{k-1}.$$

It is left to use the translation property A_k $(\mathbf{x}, \mathbf{y}, t) = A_k$ $(\mathbf{x} - \mathbf{y}, 0, t)$ to change $\Delta_\mathbf{y}$ to $\Delta_\mathbf{x}$.

The validity of general solution (10.12) is proven by induction using the fact that

$$\left[\frac{\partial}{\partial t} - D\Delta_\mathbf{x} + \lambda\right]A_0 = 0.$$

The last equality in (10.12) follows from (10.B5) and (10.4).

Appendix 10.D

PROOF OF PROPOSITION 10.5.3

The proof of Proposition 10.5.3 follows the line of the proof of Proposition 10.5.1. There are two possible scenarios for the model development up to time t. In the first one, the initial immigrant will not split; the probability for this is $P = e^{-\lambda t}$. In the second one, the initial immigrant will split at instant $0 \leq u \leq t$; the probability of the first split within the time interval $[u, u + du]$ is $\lambda e^{-\lambda t}\, du + o(du)$ as $du \to 0$. The spatial position of the split is given by the diffusion density $p(\mathbf{x}, \mathbf{y}, u)$. If the immigrant splits, the composition property of generating functions gives $M_{k_1,k_2} = h[M_{k_1-1,k_2-1}]$. Integrating over all possible split instants and locations, we obtain

$$M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$= e^{-\lambda t} + \int_{\mathbb{R}^n} d\mathbf{y}' \int_0^t du\; \lambda e^{-\lambda u}p(\mathbf{y}', \mathbf{y}, u)$$
$$h[M_{k_1-1,k_2-1}(G_1, G_2, \mathbf{y}', t - u; s_1, s_2)].$$

Here the first and second terms correspond to the first and second scenarios, respectively. Using the new integration variable $z = t - u$, we write

$$M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$= e^{-\lambda t} + e^{-\lambda t}\int_{\mathbb{R}^n} d\mathbf{y}' \int_0^t du\; \lambda e^{\lambda(t-u)}p(\mathbf{y}', \mathbf{y}, u)$$
$$h[M_{k_1-1,k_2-1}(G_1, G_2, \mathbf{y}', t - u; s_1, s_2)]$$
$$= e^{-\lambda t}\left(1 + \int_{\mathbb{R}^n} d\mathbf{y}' \int_0^t dz\; \lambda e^{\lambda z}p(\mathbf{y}', \mathbf{y}, t - z)\right.$$
$$\left. h[M_{k_1-1,k_2-1}(G_1, G_2, \mathbf{y}', z; s_1, s_2)]\right).$$

Now we take the derivative with respect to t of both sides and apply Lemma 10.A.1 using the fact that $p(\mathbf{y}', \mathbf{y}, 0) = \delta(\mathbf{y}' - \mathbf{y})$ and $(\partial/\partial t - D\Delta_\mathbf{y})p = 0$:

$$\frac{\partial}{\partial t}M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$= -\lambda M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2)$$
$$+ e^{-\lambda t}\left[\int_{\mathbb{R}^n} d\mathbf{y}' \int_0^t dz\, \lambda e^{\lambda z}\right.$$
$$h[M_{k_1-1,k_2-1}(G_1, G_2, \mathbf{y}', z; s_1, s_2)]D\Delta_\mathbf{y}p(\mathbf{y}', \mathbf{y}, t - z)$$
$$\left. + \lambda e^{\lambda t}\, h[M_{k_1-1,k_2-1}(G_1, G_2, \mathbf{y}, t; s_1, s_2)]\right].$$

Taking the operator $\Delta_\mathbf{y}$ out of the integration signs, we find

$$\frac{\partial}{\partial t}M_{k_1,k_2}(G_1, G_2, \mathbf{y}, t; s_1, s_2) = D\Delta_\mathbf{y}M_{k_1,k_2} - \lambda M_{k_1,k_2}$$
$$+ \lambda\; h[M_{k_1-1,k_2-1}].$$

It is left to establish the initial conditions. Since we start the model with a particle of generation $k = 0$ and the distribution of splitting is continuous, at $t = 0$ there are no other particles with probability 1. Hence, $M_{k_1,k_2}(G_1, G_2, \mathbf{y}, 0; s_1, s_2) = 1$ for all $k_1, k_2 \geq 1$. For generation $k_1 = k_2 = 0$, we have three possibilities: the initial immigrant has not split and is in G_1 $(i = 1, j = 0)$, the initial immigrant has not split and is in G_2 $(i = 0, j = 1)$, and neither $(i = 0, j = 0)$, with corresponding probabilities of P_1, P_2, and $1 - P_1 - P_2$, respectively. This implies (10.17).

For generation $k_1 = 0$ and $k_2 = k \geq 1$, we again have three possibilities: the initial immigrant has not split and is in G_1 $(i = 1, j = 0)$, the initial immigrant has not split and is not in G_1 $(i = 0, j = 0)$, and the initial immigrant has split $(i = 0, j \geq 0)$, with corresponding probabilities of P_1, $e^{-\lambda t} - P_1$, and $1 - e^{-\lambda t}$, respectively. In the last case, the number of the zeroth generation particles in G_1 is 0 with probability 1 while the information on the kth generation particles in G_2 is given by

$$\int_{R^n} d\mathbf{y}' \int_0^t du\; \lambda e^{-\lambda u}p(\mathbf{y}', \mathbf{y}, u)\; h[M_{k-1}(G_2, \mathbf{y}', t - u; s_2)].$$

From (10.A1), we see that the above expression equals $M_k(G_2, \mathbf{y}, t; s_2) - e^{-\lambda t}$. This implies (10.18). We notice

that setting $s_2 = 0$ in (10.17) and (10.18) each yields $(1 - P_1) + P_1 e^{s_1}$ as it should (cf. (10.7)).

Appendix 10.E

PROOF OF COROLLARY 10.5.4

The validity of (10.13) follows from Proposition 10.5.3 and the definition of $\overline{A}_{k_1,k_2}(G_1, G_2, \mathbf{y}, t)$, $A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)$. Formally, applying the operator $\partial^2/\partial s_1 \partial s_2 (\cdot)|_{s_1=s_2=0}$ to both sides of (10.15) and changing the order of differentiation, one finds, for each $k_1, k_2 \geq 1$,

$$\frac{\partial^2}{\partial s_1 \partial s_2} \left[\frac{\partial M_{k_1,k_2}}{\partial t} \right] \Bigg|_{s_1=s_2=0}$$

$$= \frac{\partial^2}{\partial s_1 \partial s_2} \left[D\Delta_{\mathbf{y}} M_{k_1,k_2} - \lambda M_{k_1,k_2} + \lambda h(M_{k_1-1,k_2-1}) \right] \Big|_{s_1=s_2=0},$$

$$\frac{\partial}{\partial t} \left[\frac{\partial^2 M_{k_1,k_2}}{\partial s_1 \partial s_2} \Big|_{s_1=s_2=0} \right]$$

$$= \left[D\Delta_{\mathbf{y}} \frac{\partial^2 M_{k_1,k_2}}{\partial s_1 \partial s_2} - \lambda \frac{\partial^2 M_{k_1,k_2}}{\partial s_1 \partial s_2} \right.$$

$$+ \lambda h'(M_{k_1-1,k_2-1}) \frac{\partial^2 M_{k_1-1,k_2-1}}{\partial s_1 \partial s_2}$$

$$\left. + \lambda h''(M_{k_1-1,k_2-1}) \frac{\partial M_{k_1-1,k_2-1}}{\partial s_1} \frac{\partial M_{k_1-1,k_2-1}}{\partial s_2} \right] \Bigg|_{s_1=s_2=0},$$

$$\frac{\partial}{\partial t} = D\Delta_{\mathbf{y}} \overline{A}_{k_1,k_2} - \lambda \overline{A}_{k_1,k_2} + \lambda B \overline{A}_{k_1-1,k_2-1}$$

$$+ \lambda h''(1) \overline{A}_{k_1-1}(G_1) \overline{A}_{k_2-1}(G_2).$$

The system (10.23) readily follows now from the definition of $A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)$. The initial conditions (10.24)–(10.25) are established by applying the operator $\partial^2/\partial s_1 \partial s_2 (\cdot)|_{s_1=s_2=0}$ to both sides of (10.16)–(10.18) and using again the definition of $A_{k_1,k_2}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t)$.

Appendix 10.F

PROOF OF PROPOSITION 10.6.2

The asymptotic (10.37) readily follows from (10.G4). To prove (10.38), let $r_v(z) := K_v(z)/K_{v+1}(z)$. From (10.G.1) one finds that

$$\frac{K_{v+1}(z)}{K_v(z)} = \frac{K_{v-1}(z)}{K_v(z)} + \frac{2v}{z} \tag{10.F1}$$

and furthermore

$$\frac{z}{2v} \frac{1}{r_v(z)} = \frac{z}{2v} r_{v-1}(z) + 1. \tag{10.F2}$$

From the monotonicity of $K_v(z)$ with respect to the index $v > 0$ it follows that $r_v(z) < 1$ for $v > 0$. Accordingly, the first term on the right hand side of (10.F2) goes to zero as $k \to \infty$. Hence,

$$\lim_{k \to \infty} \frac{z}{2v} \frac{1}{r_v(z)} = 1, \text{ or } r_v(z) \sim \frac{z}{2v}, \ k \to \infty. \tag{10.F3}$$

To complete the proof of (10.38), we use this asymptotic in (10.35). Finally, we prove (10.36). In fact, we will derive a stronger result showing the asymptotics of $r_v(z)$ and $\gamma_v(z)$ as $z \to 0$. To find the asymptotics for $r_v(z)$, we use (10.G5) for all possible combinations of signs for v and $v + 1$. We take into account that by definition v can only take values $(i, i + 1/2)_{i \in \mathbf{Z}}$.

$$r_v(z) = \begin{cases} \dfrac{K_v(z)}{K_{v+1}(z)} \sim \dfrac{\Gamma(-v)}{\Gamma(-v-1)} \left(\dfrac{2}{z}\right)^{-v-(-v-1)} \sim 2(-v-1)/z, & v \leq -3/2, \\[2ex] \dfrac{K_{-1}(z)}{K_0(z)} \sim \quad [z(ln(2/z)-\gamma)]^{-1} \quad \sim -(zlnz)^{-1}, & v=-1, \\[2ex] \dfrac{K_{-1/2}(z)}{K_{1/2}(z)} \quad\quad\quad\quad\quad = \quad\quad 1, & v=-1/2, \\[2ex] \dfrac{K_0(z)}{K_1(z)} \sim \quad z(ln(2/z)-\gamma) \quad \sim -zlnz, & v=0, \\[2ex] \dfrac{K_v(z)}{K_{v+1}(z)} \sim \dfrac{\Gamma(v)}{\Gamma(v+1)} \left(\dfrac{2}{z}\right)^{v-(v+1)} = z/(2v), & v>0. \end{cases} \tag{10.F4}$$

Combining this with (10.35) we find

$$\gamma_v(z) = \frac{2(k+1)}{Bz} r_v(z) \sim \begin{cases} \dfrac{4}{Bz^2}(v+n/2)(-v-1), & v \leq -3/2, \\[2ex] -\dfrac{(n-2)}{Bz^2 \ln z}, & v = -1, \\[2ex] \dfrac{n-1}{Bz}, & v = -1/2, \\[2ex] -\dfrac{n \ln z}{B}, & v = 0, \\[2ex] \dfrac{1}{B}\left(1 + \dfrac{n}{2v}\right), & v > 0. \end{cases} \tag{10.F5}$$

One can see that for $v \leq 0$ the ratio $\gamma_v(z)$ diverges at the origin. The rate of divergence increases monotonously from $\ln z$ to z^{-2} with the absolute value of v.

Appendix 10.G

PROPERTIES OF K_N

Here we summarise some essential facts about the modified Bessel function of the second kind $K_v(z)$. The sources of this as well as further information about $K_v(z)$ are handbooks (Abramowitz and Stegun, 1965, Chapters 9, 10) and (Gradshtein and Ryzhik, 2007, Section 8.4). The function K_v can be defined as a decreasing solution of the modified Bessel differential equation

$$x^2\, y'' + x\, y' - \left(x^2 + v^2\right) y = 0.$$

The function $K_v(z)$ exponentially decreases as $z \to \infty$ and diverges at $z = 0$. In addition, $K_{-v}(z) = K_v(z)$ and

$$K_{v+1}(z) = K_{v-1}(z) + \frac{2\,v}{z}\, K_v(z). \qquad (10.\text{G1})$$

For integer $k \geq 0$ we have

$$K_{k+1/2}(z) = \sqrt{\frac{\pi}{2z}}\mathrm{e}^{-z}\sum_{m=0}^{k} \frac{(k+m)!}{m!(k-m)!(2z)^m}, \qquad (10.\text{G2})$$

and in particular

$$K_{\pm 1/2}(z) = \sqrt{\frac{\pi}{2\,z}}\,\mathrm{e}^{-z};\ K_{3/2}(z) = \sqrt{\frac{\pi}{2\,z^3}}\,\mathrm{e}^{-z}. \qquad (10.\text{G3})$$

For arbitrary fixed v and $z \gg v$

$$K_v(z) \sim \sqrt{\frac{\pi}{2\,z}}\,\mathrm{e}^{-z},\ z \to \infty. \qquad (10.\text{G4})$$

The asymptotic behaviour at $z = 0$ is given by

$$K_v(z) \sim \begin{cases} \dfrac{\Gamma(|v|)}{2} \left(\dfrac{2}{z}\right)^{|v|}, & |v| \neq 0, \\[2ex] \log\left(\dfrac{2}{z}\right) - \gamma, & v = 0, \end{cases} \qquad (10.\text{G5})$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant (Euler, 1734).

References

Abramowitz, M. and Stegun, I. A. eds., (1965) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover.

Aki, A. (1985) Theory of earthquake prediction with special reference to monitoring of the quality factor of lithosphere by the coda method *Earthquake Prediction Research*, **3**, 219–230.

Albert, R. and Barabasi, A. L. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74** (1), 47–97.

Albeverio, S., Jentsch, V. and Kantz, H. (eds) (2005) *Extreme Events in Nature and Society*. Heidelberg: Springer.

Allegre, C. J., Lemouel, J. L., and Provost, A. (1982) Scaling rules in rock fracture and possible implications for earthquake prediction. *Nature*, **297**, 5861, 47–49.

Athreya, K. B. and Ney, P. E. (2004) *Branching Processes*. New York: Dover Publications.

Bak, P., Tang, C. and Wiesenfeld, K. (1988) Self-organized criticality. *Physical Review A*, **38**, 1, 364–374.

Ben-Zion, Y. (2003) Appendix 2: Key formulas in earthquake seismology, In *International Handbook of Earthquake and Engineering Seismology*, Part B, 1857–1875, Academic Press.

Blanter, E. M., Shnirman, M. G. and LeMouel, J. L. (1997) Scaling laws in blocks dynamics and dynamic self-organized criticality, *Physics of the Earth and Planetary Interiors*, **103** (1–2), 135–150.

Bollobás, B. (2001) *Random Graphs*, Cambridge: Cambridge University Press, second edn.

Brunetti, M. T., Guzzetti, F. and Rossi, M. (2009) Probability distributions of landslide volumes. *Nonlinear Processes in Geophysics*, **16**, 2, 179–188.

Burlando, B. (1990) The fractal dimension of taxonomic systems. *Journal of Theoretical Biology*, **146**, 99–114.

Burridge, R. and Knopoff, L. (1967) Model and theoretical seismicity. *Bulletin of the Seismological Society of America*, **57**, 3, 341–371.

Chow, P. L. (2007) *Stochastic Partial Differential Equations* Boca Raton, FL: Chapman Hall/CRC Press.

Dhar, D. (1990) Self-organized critical state of sandpile automaton models. *Physical Review Letters*, **64**, 14, 1613–1616.

Doucet, A., de Freitas, N. and Gordon, N. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. Springer.

Drossel, B. and Schwabl, F. (1992) Self-organized critical forest-fire model. *Physical Review Letters*, **69**, 11, 1629–1632.

Durrett, R. (2006) *Random Graph Dynamics*. Cambridge: Cambridge University Press.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2008) *Modelling Extremal Events for Insurance and Finance (Stochastic Modelling and Applied Probability)*. Springer.

Euler, L. (1734) De Progressionibus harmonicis observationes, *Commentorii: Academiae Scientiarum Imperialis Petroplitanae*, **7** (1734/35), 150–161.

Evans, L. C. (1998) *Partial Differential Equations*, American Mathematical Society, Providence.

Faá di Bruno, F. (1855), Sullo sviluppo delle Funzioni, *Annali di Scienze Matematiche e Fisiche* (in Italian), **6**: 479–480.

Frisch, U. (1996) *Turbulence: The Legacy of A. M. Kolmogorov*, Cambridge: Cambridge University Press.

Gabaix, X., Gopikrishnan, P. Plerou, V. and Stanley, H. E. (2003) A theory of power-law distributions in financial market fluctuations. *Nature*, **423**, (6937), 267–270.

Gabrielov, A., Keilis-Borok, V., Sinai, Y. and Zaliapin, I. (2008) Statistical properties of the cluster dynamics of the systems of statistical mechanics. in *Boltzmann's Legacy*, ESI Lecture Notes in Mathematics and Physics: G. Gallavotti, W. Reiter and J. Yngvason. Zurich: European Mathematical Society, pp. 203–216.

Gabrielov, A. M., Zaliapin, I. V., Keilis-Borok, V. L. and Newman W. I. (2000) Colliding cascades model for earthquake prediction. *Geophysical Journal International*, **143**, 427–437.

Gradshtein, I. S. and Ryzhik, I. M., (2007) *Tables of Integrals, Series and Products* eds., A. Jeffrey and D. Zwillinger Academic Press, Seventh edn.

Grimmett, G. R. (1999) *Percolation*, Springer, second edn.

Gutenberg, B. and Richter, C. F. (1944) Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**, 185–188.

Haberman, R. E. (1981) Precursory seismicity pattern: Stalking the mature seismic gap. in *Earthquake Prediction: An International Review*, Maurice Ewing Series, 4. American Geophysical Union. Washington, DC: 1981, pp. 29–42.

Kadanoff L. P. (2000) *Statistical Physics: Statics, Dynamics and Renormalization*, World Scientific Publishing Company.

Kalman, R. E. and Bucy, R. S. (1961) New results in linear filtering and prediction theory. Transactions of the ASME, Series D *Journal of Basic Engineering*, **83**, 95–108.

Keilis-Borok, V. I. and Shebalin, P. N. (eds.) (1999) Dynamics of lithosphere and earthquake prediction (special issue), *Physics of the Earth and Planetary Interiors*, **111**.

Keilis-Borok, V. I. and Soloviev, A. A. (eds) (2003) *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*. Heidelberg: Springer.

Keilis-Borok, V. I., Soloviev, A. A., Allegre, C. B., *et al.* (2005) Patterns of macroeconomic indicators preceding the unemployment rise in Western Europe and the USA. *Pattern Recognition*, **38**, (3), 423–435.

Keilis-Borok, V. I., Stock, J. H., Soloviev, A. and Mikhalev, P. (2000) Pre-recession pattern of six economic indicators in the USA. *Journal of Forecasting*, **19**, 65–80.

Klass, O. S., Biham, O., Levy, M., Malcai, O. and Soloman, S. (2006) The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, **90**, 2, 290–295.

Kolmogorov, A. N. (1941a) The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akademii Nauk. USSR*, **30**, 299–303 (in Russian). English translation: Proceedings of the Royal Society of London. Series A, **434**, 9–13 (1991).

Kolmogorov, A. N. (1941) Interpolation, extrapolation of stationary random sequences. *Izv. Akad. Nauk. SSSR, Ser. Mat.*, **5**, 3–14. (English translation: W. Doyle, RAND corporation memorandum RM–3090–PR, April 1962).

Kushner H. J. (1964) On the differential equations satisfied by conditional probability densities of Markov processes, with applications. *SIAM Journal on Control, Ser. A*, **2**(1), 106–119.

Lotka, A. J. (1926) The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, **16** (12), 317–324.

Ma, S.-K. (2000) *Modern Theory of Critical Phenomena*. Boulder, CO: Westview Press.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., and Reichenbach, P. (2004) Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, **29**(6), 687–711.

Mandelbrot, B. (1983) *The Fractal Geomery of Nature*, W. H. Freeman.

Mandelbrot, B. and Taylor, H. M. (1967) On the distribution of stock price differences. *Operations Research*, **15**, 6, 1057–1062.

McWilliams, J. C. (1990) The vortices of two-dimensional turbulence, *Journal of Fluid Mechanics*, **219**, 361–385.

Mogi, K. (1981) Seismicity in Western Japan and long-term forecasting. in *Earthquake Prediction: An International Review*, Maurice Ewing Series, 4. Washington, DC: American Geophysical Union, pp. 43–51.

Narkunskaya, G. S. and Shnirman, M. G. (1990) Hierarchical model of defect development and seismicity. *Physics of the Earth and Planetary Interiors*, **61**, 29–35.

Newman, M. E. J. (1997) Evidence for self-organized criticality in evolution. *Physica D*, **107**, 293–296.

Newman, M. E. J. (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46** (5), 323–351.

Newman, M. E. J., Barabasi, A.-L. and Watts, D. J. (2006) *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Newman, W. I., Turcotte, D. L. and Gabrielov, A. M. (1995) Log-periodic behavior of a hierarchical failure model with applications to precursory seismic activation. *Physical Review E*, **52**, 4827–4835.

Obukhov, A. M. (1941) Spectral energy distribution in a turbulent flow. *Doklady Akademii Nauk. USSR*, **1**, 22–24.

Olami, Z., Feder, H. J. S. and Christensen, K. (1992) Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical Review Letters*, **68**, 8, 1244–1247.

Pareto, V. (1897) *Cours d'economie Politique* Lausanne: F. Rouge.

Plerou, V. and Stanley, H. E. (2008) Stock return distributions: Tests of scaling and universality from three distinct stock markets. *Physical Review E*, **77**, 3, Art. No. 037101.

Richardson, L. F. (1960) *Statistics of Deadly Quarrels*. Pacific Grove, CA: Boxwood Press.

Rundle, J. B. and Klein, W. (1993) Scaling and critical phenomena in a cellular automaton slider-block model for earthquakes. *Journal of Statistical Physics*, **72**, (1–2), 405–412.

Rundle, J., Turcotte, D. and Klein, W. eds. (2000) *Geocomplexity and the Physics of Earthquakes*. Washington, DC: American Geophysical Union.

Sornette, D. (2004) *Critical Phenomena in Natural Sciences*. second edition, Heidelberg: Springer-Verlag.

Stanley, H. E. (1971) *Introduction to Phase Transitions and Critical Phenomena*. Oxford: Oxford University Press.

Stauffer, D. and A. Aharony (1994) *Introduction to Percolation Theory*. Baton Roca, FL: CRC.

Sykes, L. R., Shaw, B. E. and Scholz, C. H. (1999) Rethinking earthquake prediction. *Pure and Applied Geophysics*, **155**, 2–4, 207–232.

Turcotte, D. L. (1997) *Fractals and Chaos in Geology and Geophysics*. second edition Cambridge: Cambridge University Press.

Turcotte, D. L. (1999) Self-organized criticality. *Reports on Progress in Physics*, **62**, 10, 1377–1429.

Extreme Natural Hazards, Disaster Risks and Societal Implications

Wiener, N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*. Wiley.

Zakai, M. (1969) On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**, (3), 230–243.

Zaliapin, I., Keilis-Borok, V. and Ghil, M. (2003a) A Boolean delay equation model of colliding cascades. Part I: Multiple seismic regimes *Journal of Statistical Physics*, **111**, (3–4), 815–837.

Zaliapin, I., Keilis-Borok, V. and Ghil, M. (2003b) A Boolean delay equation model of colliding cascades. Part II: Prediction of critical transitions *Journal of Statistical Physics*, **111**, (3–4), 839–861.

Zipf, G. K. (1965) *Psycho-Biology of Languages*, Houghton-Mifflin, MIT Press.