Environment and Planning A, 1982, volume 14, pages 1449-1467

A dynamic generalization of Zipf's rank-size rule

B Roehner
Laboratoire de Physique Théorique et Hautes Énergies, Université Paris VII, Paris, France
K E Wiese
U.E.R. de Physique, Université Paris VII, Paris, France
Received 26 June 1981

Abstract. A dynamic deterministic model of urban growth is proposed, which in its most simple form yields Zipf's law for city-size distribution, and in its general form may account for distributions that deviate strongly from Zipf's law. The qualitative consequences of the model are examined, and a corresponding stochastic model is introduced, which permits, in particular, the study of zero-growth situations.

1 Introduction

Numerous articles have been concerned with the so-called Zipf's law (or rank-size rule) for city-size distribution. A very full bibliography can be found in a recent study by Pumain (1980). It is indeed of great interest to put this law on a sound basis both for urban planning purposes and for historical demography (see section 2.4.3).

The proposed models fall into two broad classes. First, the static (and usually stochastic) models (Simon, 1955; Steindl, 1965; Hill, 1974; Woodroofe and Hill, 1975), and, second, the dynamic models (Malecki, 1975; Vining, 1977). Vining, in particular, insists on the importance of deterministic, dynamic models. The continual fluctuation in the growth rates of cities is precisely what makes dynamic models essential. The stationary state described by the static model has hardly ever been observed. Zipf's law (even in its general form of the Pareto distribution) is for similar reasons rarely verified with precision at a regional level.

In this article, we shall give a dynamic and enlarged interpretation of Zipf's law, first in a deterministic framework and then in a stochastic framework. Here, the notion of evolution function replaces the usual one of growth rate. Zipf's law will emerge from the proposed model as a particular case.

In its general form the model represents rather closely the observed evolution of the city-size distribution curves. The model allows the study of many ideal cases and thus gives a better understanding of the connection between demographic evolution and change in the distribution curves. Let us emphasize that the aim of the model is to describe, rather than explain, complicated demographic phenomena by means of only a few basic facts. We believe that this description will greatly help a more accurate explanation of urban growth in socioeconomic terms.

The region under study may be defined from two different points of view. If we insist on self-sufficiency, as Zipf did, it would be preferable to consider rather large areas; but if we insist on homogeneity, small regions would be preferable. Even in this case, however, the number of towns in each region should be more than a hundred for a meaningful distribution function to be plotted.

To study the influence of changing growth rates on the distribution function, the second view is preferable. Indeed, in the first case, various growth rates of different areas would be lumped together and partially cancel each other out. This point shows that rank-size curves are as a rule more regular at a national than at a regional level [as has already been stressed by Malecki (1975)].

As an aside, we would like to point out that it is commonly observed that small objects are always more numerous than large ones, whether they are plants, animals, stars, or towns. One may ask whether some general explanation exists, although even it exists, such an explanation would most probably be no more than a formal one, as the following example shows. If one plots the distribution of the stars in our galaxy according to their masses, an almost exact Pareto distribution is obtained. This is a result of the fact that the lifetime of large stars is shorter than that of small ones: large stars burn their fuel proportionally faster. In contrast, the opposite applies for towns: a large town usually has a much better chance of survival through a war or an economic crisis.

This paper is laid out as follows. In section 2 we introduce the model, and establish its theoretical formulae. Some simple examples are presented, followed by some qualitative conclusions. In section 3 we compare the theoretical description with statistical data. In section 4 we study the corresponding stochastic model, the transient behaviour of which is shown to be similar to that obtained with the deterministic model; and, last, we study the evolution of the city-size distribution in the absence of growth.

2 The deterministic model

2.1 *The rank–size representation*

To describe the rank–size distribution at time t, we may use the following cumulated frequency function:

$$\mathrm{G}(x, t) = \frac{\text{number of cities with more than } x \text{ inhabitants at time } t}{\text{total number of cities}} .$$

$$= \int_x^\infty \mathrm{f}(x, t)\,\mathrm{d}x ,$$

where $\mathrm{f}(x, t)$ is the frequency function (normalized to unity), and G is related to the usual cumulated frequency function F by $\mathrm{G} = 1-\mathrm{F}$. The justification of this unusual choice of G is as follows:

1. In accordance with Zipf, the rank–size distribution is generally plotted with population as ordinate, in contrast to the usual practice in statistics to which we shall return in this work.

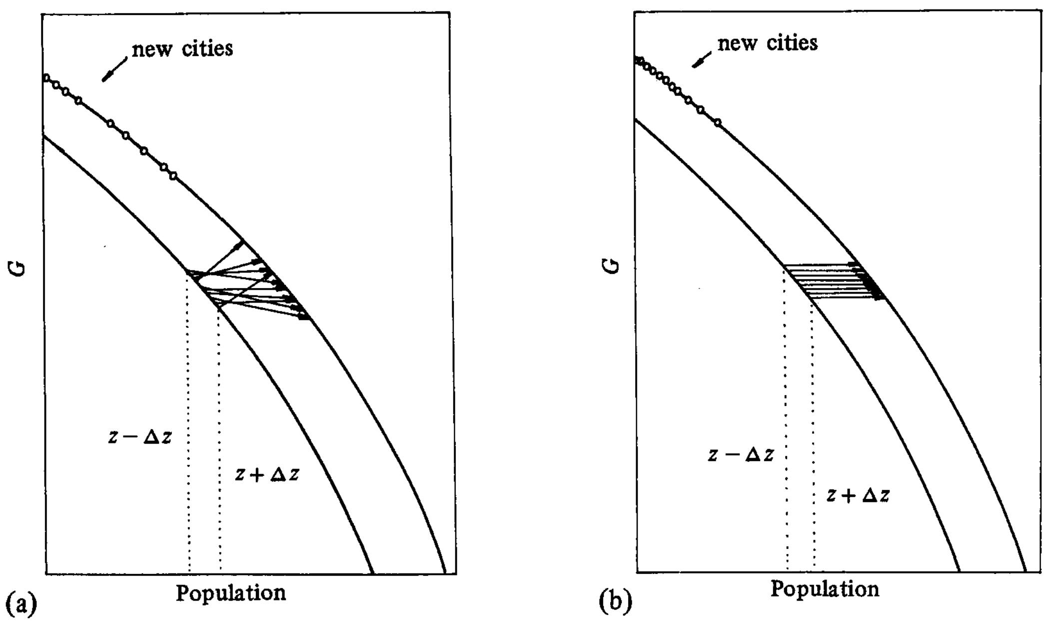

Figure 1. Schematic representation of city evolution (a) in reality and (b) in the model.

2. We shall use a normalized G function in theoretical considerations only. For the representation of statistical data, however, we shall use a nonnormalized G function in order to exhibit the total number of cities which is, of course, a major variable.

2.2 *The model*

Urban evolution is mainly based on three phenomena:
(a) The real growth of each city resulting from the births and deaths occurring in it.
(b) The following migrations: from city to city, that is to say between two urban places; from village or town to city, that is to say between rural and urban places; from foreign country to foreign country, these having been important in the past for countries such as the United States and Australia.
(c) The variation in the number of cities: creation of new cities; absorption of smaller cities into the suburbs of larger ones; or their disappearance by decline of population.
Confronted with the complexity of these phenomena, deliberately simple hypotheses are made as follows:
(1) The variation in the number of cities N will be described by an increasing function $N(t)$ which represents that total number at time t. N should be zero for $t < 0$:

$$N = N(t)y_1(t) , \qquad y_1(t) = \begin{cases} 1, & t \geqslant 0 , \\ 0, & t < 0 . \end{cases}$$

We will assume that the size of all the cities is less or equal to x_0 at the time of their emergence; so that they can only appear in the population range $x \geqslant x_0$, with the population x_0. This kind of hypothesis works if we assume that a city is small before it grows bigger. Nevertheless in section 2.6 a more general hypothesis will be examined.
(2) We assume that a city with population z at time zero will have at time τ, a population x given by

$$x = x(\tau, z) . \tag{1}$$

Obviously, such an equation can only be a first approximation since the development of a city depends not only on its own population but also on its environment (for example, on the size of neighbouring cities).

Let us examine the phenomena which can be described by this hypothesis.
(a) Equation (1) may describe the differences in the growth rates of cities according to their population. For example, for a linear equation, $x = \alpha\tau + z$, the growth rate decreases with population: $(1/x)\mathrm{d}x/\mathrm{d}t = \alpha/x$. An exponential function, $x = z\exp(\tau z)$, gives instantaneous growth rates independent of size, according to Gibrat's hypothesis. Yet another function would give growth rates increasing with size, according to Robson's (1973) observation on cities of England and Wales during the nineteenth century.
(b) We shall consider $x(\tau, z)$ mainly as an increasing function of τ for any value of z. However, let us assume for the moment that $x(\tau, z)$ decreases with τ for z belonging to the interval (x_0, x_1). In this case, a decrease in the total number of cities will be described. Such a demographic decline occurred in most of the districts of France between the two World Wars.
(c) According to equation (1), two cities having initially the same population should have the same evolution. Of course, things are not so simple. A detailed statistical analysis of the growth rates (their means, their mean square deviations, their spatial and time correlations) can, for example, be found in Pumain (1980) for the French 'communes' between 1831 and 1911.

However, if we can consider a group of cities corresponding to the size class $(z - \Delta z, z + \Delta z)$, their average evolution will then follow some specific evolution function (1). Our hypothesis substitutes the evolution described in figure 1(b) for the one described in figure 1(a) in the same way that the kinetic theory of gases ignores the chaotic molecular motion and takes only the average motion into account. Our hypothesis holds reasonably well if there are enough cities in the size classes to keep the mean square deviation sufficiently small.

Any choice of the function $x(\tau, z)$ can be made, but it must have some definite interpretation to be interesting. To make the interpretation easier $x(\tau, z)$ can be chosen as the solution of a differential evolution equation:

$$\dot{x} = \mathrm{h}(x) , \qquad x(0) = z , \tag{2}$$

where $\dot{x} = \mathrm{d}x/\mathrm{d}t$. For example, the equation:

$$\dot{x} = a + bx - cx^2 , \qquad b, c > 0 ,$$

which is frequently used in this paper, implies:
an out-migration or an in-migration, according to the sign of a,
a birth and death growth through the term bx,
a saturation effect (through the term $-cx^2$) which can be related for example to the growing service costs of a big city.

We shall not try in this paper to make a precise link between the form of the evolution function and the migration phenomena (for France, see for example Courgeau, 1970; Desplanques, 1975).

2.3 *Evolution of the distribution function*

The reasoning is often easier if we consider x as a spatial coordinate. In that way cities may be represented by points. These points are created at x_0 and then move on the line $x \geqslant x_0$ according to the function $x(\tau, z)$. The total number of points on the line $x \geqslant x_0$ is of course given by $N(t)$. This evolution will be termed an (N, x) process.

We shall also need the inverse functions of $x = x(\tau, z)$, which are $\tau = \tau(x, z)$ and $z = z(x, \tau)$, where $\tau(x, z)$ is the time that a point takes to move from z to x, and $z(x, \tau)$ is the initial position of the point which has position x at time τ. We assume that $x(\tau, z)$ is a smoothly increasing function of τ and z to be sure of the existence of those two functions. Furthermore we assume that

$$\lim_{x \to \infty} z(x, \tau) \to \infty ,$$

which means that the evolution cannot be 'explosive'. It is noted that when $x(\tau, z)$ is a solution of equation (2), then the two inverse functions above have quite a simple expression:

$$\tau(x, z) = \int_z^x \frac{\mathrm{d}x_1}{\mathrm{h}(x_1)} , \qquad z(x, \tau) = x(-\tau, x) .$$

2.3.1 *Evolution with time* The only possible observation is the change with time of an initial distribution $\mathrm{G}_1(x)$. To describe that evolution we proceed in three stages:
(1) compute the distribution resulting from the process (N, x) with $\mathrm{G}_1(x) = y_1(x - x_0)$ which means there is only one size of city at the start;
(2) give the evolution of an arbitrary initial distribution $\mathrm{G}_1(x)$ under a $(0, x)$ process;
(3) combine the preceding cases to give the general one.

Stage (1) At time t the points which are farther than x correspond to those which were created between time 0 and time $t - \tau(x, x_0)$. Therefore,

$$\mathrm{G}(x, t) = \frac{N[t - \tau(x, x_0)]}{N(t)} y_1[x(t, x_0) - x] , \qquad x \geqslant x_0 . \tag{3}$$

Stage (2) Let $f_1(x)$ and $f(x, t)$ be the density functions corresponding to $G_1(x)$ and $G(x, t)$. Then we have

$$f(x, t) = f_1[z(x, t)]\left|\frac{dz}{dx}\right| ,$$

which yields

$$G(x, t) = \int_{z(x, t)}^{\infty} f(z_1)dz_1 = G_1[z(x, t)] .$$

For the moment let $x(\tau, z)$ be a decreasing function of τ for z in the interval (x_0, x_1). The corresponding decrease in the number of cities will be given by:

$$N(t) = \{1-G_1[z(x_0, t)]\}N_1 ,$$

where N_1 is the number of cities corresponding to $G_1(x)$, and the preceding formula can be written as

$$G(x, t) = \frac{G_1[z(x, t)]}{G_1[z(x_0, t)]} .$$

Stage (3) We assume now the creation of new cities [in addition to the ones corresponding to $G_1(x)$] according to a function $N_2(t)$. Hence we have at time t a total number of cities given by $N_1 + N_2(t)$. The total distribution function is the weighted sum of the preceding ones:

$$G(x, t) = \frac{N_2[t-\tau_2(x, x_0)]}{N_1+N_2(t)}y_1[x_2(t, x_0)-x] + \frac{N_1}{N_1+N_2(t)}G_1[z_2(x, t)] . \qquad (4)$$

2.3.2 *Total urban population* By definition, the total urban population $P_u(t)$ is given by

$$P_u(t) = N(t)\int_{x_0}^{\infty} x f(x, t)dx ,$$

with $f(x, t) = -\partial G/\partial x$. Equation (4) shows that $G(x, t)$ has some points of discontinuity, and so the derivative must be calculated with some care. It gives

$$P_u(t) = N(t)x_0 + \int_0^t N(t-\tau)h[x(\tau, x_0)]d\tau .$$

Notice that the present calculation involves an approximation in that we consider the population x as a continuous variable.

2.4 *Examples of transient distributions*

2.4.1 *The time independent distributions and the Pareto distribution* It is interesting to note that if the number of cities grows exponentially, then

$$N = N_0 \exp(\Lambda t) ,$$

and the distribution function will have a time independent form whatever the function $x(\tau, z)$:

$$G(x, t) = \exp[-\Lambda\tau(x, x_0)]y_1[x(t, x_0)-x] .$$

If $x(\tau, z)$ is itself an exponential function, we obtain the usual Pareto distribution:

$$x(\tau, z) = z\exp(\lambda\tau) \Rightarrow G(x, t) = \left(\frac{x_0}{x}\right)^{\Lambda/\lambda} y_1[x_0 \exp(\lambda\tau)-x] .$$

The slope ρ of the corresponding straight line in log-log coordinates is thus given by the ratio of the growth rate of the number of cities with the growth rate of each city,

that is, $\rho = \Lambda/\lambda$. As is well known, the so-called Pareto coefficient ρ is a measure of concentration. Let us state more precisely the relation between this coefficient and two standard characterizations of the concentration:
(a) The existence of the variance σ^2 of a Pareto distribution (extending to infinity) requires $\rho > 2$, which gives

$$\sigma^2 = \frac{\rho x_0^2}{(\rho-2)(\rho-1)} ,$$

so that $\sigma^2 \to 0$ as $\rho \to \infty$.
(b) The Gini coefficient g (Atkinson, 1975, page 45) is such that if $g = 0$, all the cities have the same size x_0; if $g = 1$, all the cities are very small except one which is very large. For $\rho > 1$ we obtain

$$g = \frac{1}{2\rho-1} ,$$

so that $g \to 0$ as $\rho \to \infty$.

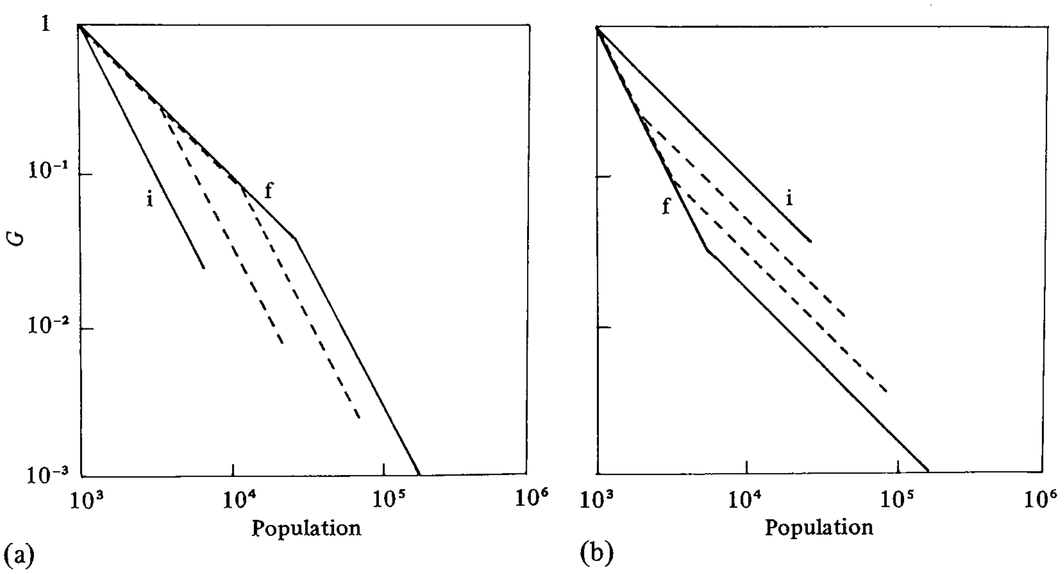

Figure 2. Transient evolution between two Pareto distributions, where i and f are the initial and final distributions, respectively.

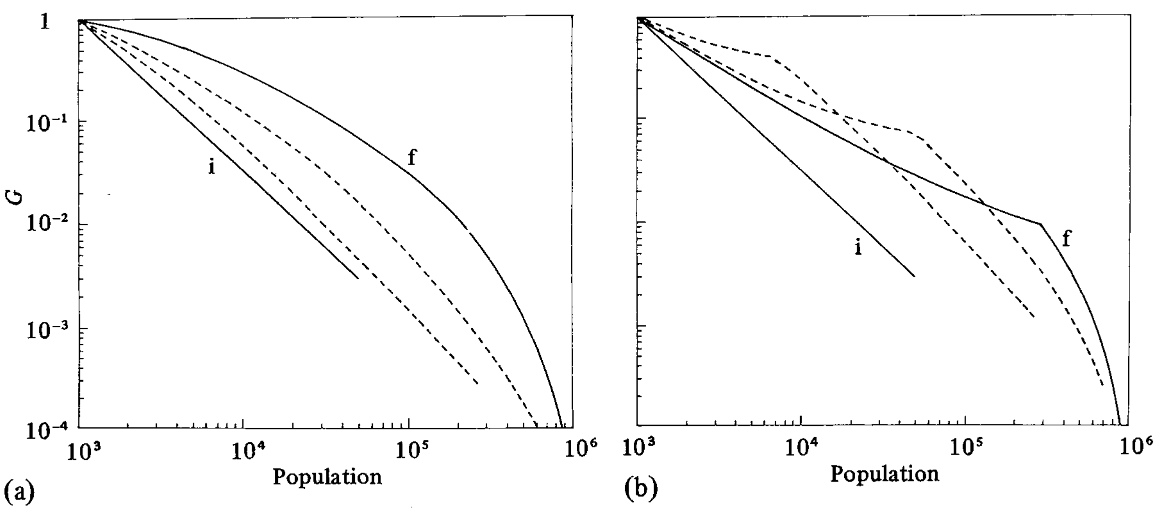

Figure 3. Transient evolution between a Pareto distribution and a distribution resulting from a (N, x) process where N and x are logistic functions in (a) $N_1' = N_1$, and in (b) $N_1' < N_1$; and i and f stand for initial and final distributions, respectively.

For example, in most districts of France a decrease of ρ (that is, a concentration increase) was observed in the last hundred years.

In the present case of a Pareto distribution the expression of the total urban population is given by

$$P_u(t) = N_0 x_0 \frac{\Lambda \exp(\Lambda t) - \lambda \exp(\lambda t)}{\Lambda - \lambda} .$$

2.4.2 *Transient behaviour between two Pareto distributions* Let us consider an initial Pareto distribution with the following characteristics:
(1) N_1 cities with sizes between x_0 and a maximum size x_m^1,
(2) a slope ρ_1.
We study its change under the process

$$N_2(t) = N_1 \exp(\Lambda_2 t) - N_1' , \qquad x_2(\tau, z) = z \exp(\lambda_2 \tau) ,$$

for which equation (4) gives[(1)] for $N_1' = N_1$:

$$G(x, t) = \left(\frac{x_0}{x}\right)^{\rho_2} y_1[x_0 \exp(\lambda_2 t) - x] + \exp[(-\lambda_2 t)(\rho_2 - \rho_1)]\left(\frac{x_0}{x}\right)^{\rho_1} y[x - x_0 \exp(\lambda_2 t)] y_1 [x_m^1 \exp(\lambda_2 t) - x] ,$$

where $\rho_2 = \Lambda_2/\lambda_2$. Figure 2 gives the corresponding curves for the two cases:
(1) $\rho_1 = 2$, $\rho_2 = 1$, $N_1' = N_1$,
(2) $\rho_1 = 1$, $\rho_2 = 2$, $N_1' = N_1$.
Before commenting on them, we shall study another case. The initial distribution and the function $N_2(t)$ remain the same, but for $x(\tau, z)$ we take a logistic function which is the solution of:

$$\dot{x} = \lambda_2 x - \mu_2 x^2 \Rightarrow x(\tau, z) = \frac{\lambda_2}{\mu_2 + [(\lambda_2/z) - \mu_2] \exp(-\lambda_2 t)} .$$

It is similar to the preceding case, but with a more realistic evolution function. Figure 3(a) gives the curves corresponding to $N_1' = N_1$. Such distribution functions which have a downturn are commonly observed in practice. Some authors (Berry, 1969; Ferguson, 1969) have proposed that they should be fitted by a log-normal density function. To allow a precise comparison we give in figure 4 the distribution

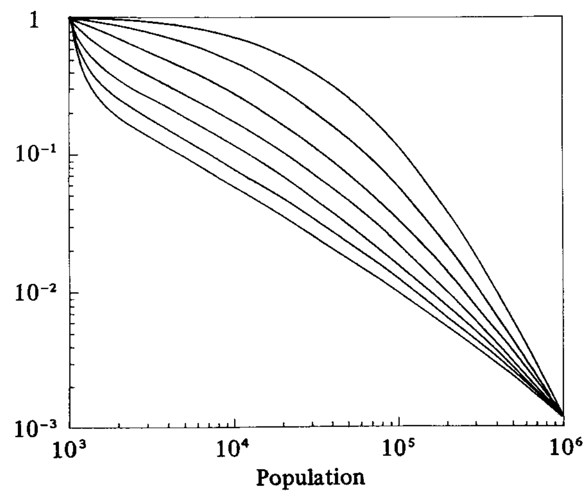

Figure 4. The distribution function for a log-normal density function. Here m increases from 4 to 10 and σ is such that the distribution extends over the whole range of the intervals $(10^3, 10^6)$ and $(1, 10^{-3})$. The lowest curve corresponds to $m = 4$, the highest to $m = 10$.

(1) y is identical to y_1 except at the origin, when $y(0) = \frac{1}{2}$.

functions (in log–log coordinates) relative to the log-normal density function,

$$f(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-x_0)-m}{\sigma}\right)^2\right],$$

with different values of σ and m.

Curves with an upturn as shown in figure 3(b) are rather unusual. Here $N_1' < N_1$ (see, however, section 2.7).

2.4.3 *The memory property of the distribution curves* The former curves suggest two conclusions:

(a) A trace of the initial distribution can be recognized in the large x range, even a long time after the outset of the new process (N_2, x_2). Thus the distribution curves keep track of the past demographic situation, and are thus interesting for historical demography (just as the study of tree rings is useful for historical climatology).

Of course, over a period of several decades, there may be several successive modifications, and it may be difficult to isolate one from another. We plan to study this point further in a later paper.

(b) Second, the transition from an initial to a final state is not a reversible one since the inverse transition does not go through the same intermediate steps. This is another manifestation of the memory property.

2.5 *How can a slowly decreasing distribution be generated?*

A slowly decreasing distribution (SDD) is similar to what is usually called a skew distribution. More precisely we should say that a distribution is slowly decreasing if it extends both in abscissa and in ordinate inside the intervals $(x_0, 10^3 x_0)$ and $(y_0, 10^3 y_0)$, and if it remains between the two curves

$$y_1 = \left(\frac{x_0}{x}\right)^{\alpha_1}, \qquad y_2 = \left(\frac{x_0}{x}\right)^{\alpha_2}, \qquad \text{with} \quad 0{\cdot}2 < \alpha_1, \alpha_2 < 4 .$$

Interestingly, the class of slowly decreasing distributions includes all the observed distribution curves which are more or less undulating. And it is this very fact which makes the use of bilogarithmic paper essential even if one does not have a straight line.

We now ask the question: what kind of (N, x) process can generate such distribution? We shall only discuss briefly the two kinds of growth (linear and exponential) which have a clear practical interpretation.

As may be expected, an SDD can only be obtained if there is a balance between the number of created cities and their growth:

$$N = \Lambda t + N_0 \quad \text{and} \quad x = z \exp(\lambda\tau)$$

gives

$$G = 1 - \frac{\Lambda}{\lambda(\Lambda t + N_0)} \ln\left(\frac{x}{x_0}\right),$$

$$N = N_0 \exp(\Lambda t) \quad \text{and} \quad x = z(1+\lambda\tau)$$

gives

$$G = \exp\left[-\frac{\Lambda}{\lambda}\left(\frac{x}{x_0} - 1\right)\right].$$

In the first case, G decreases too slowly and in the second, too quickly. But even balanced linear (N, x) pairs do not give an SDD:

$$N = \Lambda t + N_0 \quad \text{and} \quad x = z(1+\lambda\tau)$$

gives

$$G = 1 - \frac{\Lambda}{\lambda x_0} \frac{1}{\Lambda t + N_0} (x - x_0) ,$$

which decreases too slowly.

SDD are obtained for exponential and similar growth (this includes logistic functions far from the threshold) both for N and for x. Thus exponential growth seems to play a very special role.

The previous conclusion fits well with the demographic development of countries such as the United States. In fact both the urban population and the number of cities increased exponentially (figure 5). Older countries such as France, however, have experienced a very slow population increase, with long periods of stagnation; and for these countries our previous conclusion looks at first sight very surprising. We have two complementary explanations:

(a) A slow increase can be interpreted as a succession of periods of fast (that is, exponential) expansion followed by stagnation. And the resulting distribution function will be quite different. Not even an SDD in the first case and a Pareto distribution in the second.

(b) The second explanation results from a more detailed study of the stagnation periods. Of course, in our deterministic model if we have a stagnation both in the number of cities and in the city size, the city-size distribution will remain constant. But then the dispersion of the growth rates becomes essential, and a stochastic model is necessary. In section 4, we shall see that in this case the distribution function converges toward a Pareto distribution with coefficient 1.

2.6 *The distribution of city creation*

Cities do not necessarily emerge with their minimum size as x_0:

(1) Some cities begin with a rather explosive growth rate. For example, the population of Palm Beach Gardens (Florida) leapt from 1 in 1960 to 6102 in 1970.

(2) In East European countries, the creation of a city can result from a political decision.

Let us now assume that among the $N'(t)\mathrm{d}t$ cities created in the interval $(t, t+\mathrm{d}t)$, we have a fraction $N'(t)\mathrm{d}t\,\nu(z)\mathrm{d}z$ which emerge with a size between z and $z+\mathrm{d}z$; where

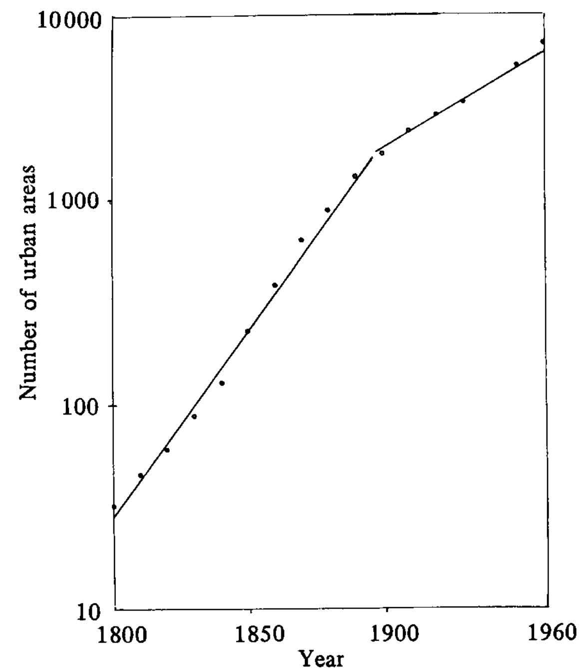

Figure 5. Number of urban areas in the USA (source: Historical Statistics of the USA, 1971).

$\nu(z)$ is a density function with

$$\int_{x_0}^{\infty} \nu(z)\,\mathrm{d}z = 1 \ .$$

The contribution of the interval $(z, z+\mathrm{d}z)$ to $\mathrm{G}(x, t)$ will depend on the position of z in relation to x:

$$\mathrm{G}(x, t) = \int_{x}^{\infty} \nu(z)\,\mathrm{d}z + \frac{1}{N(t)} \int_{x_0}^{x} N[t-\tau(x, z)]\,y[t-\tau(x, z)]\,\nu(z)\,\mathrm{d}z \ ,$$

which can finally be written as

$$\mathrm{G}(x, t) = \mathrm{G}_\nu(x) + \frac{y(x-x_0)}{N(t)} \int_{\max[x_0,\, z(x,\, t)]}^{x} N[t-\tau(x, z)]\,\nu(z)\,\mathrm{d}z \ .$$

2.6.1 *An example: a modification of the Pareto distribution due to a distribution of city creation* We choose:

$$N = N_0 \exp(\Lambda t) \ , \qquad x = z \exp(\lambda\tau) \ , \qquad \nu(z) = \frac{2}{\pi} \frac{\epsilon}{(z-x_0)^2+\epsilon^2} \ .$$

We shall recover our previous hypothesis when ϵ approaches zero since

$$\lim_{\epsilon \to 0} \nu(z) = \delta(z-x_0) \ ,$$

where δ is the usual delta function. Here

$$\mathrm{G}_\nu(x) = 1 - \frac{2}{\pi} \arctan \frac{x-x_0}{\epsilon} \ ,$$

and integration gives

$$\mathrm{G}(x, t) = \mathrm{G}_\nu(x) + \frac{2}{\pi}\frac{x_0}{x} \arctan \frac{x-x_0}{\epsilon} + \frac{1}{\pi}\frac{\epsilon}{x} \ln \frac{(x-x_0)^2+\epsilon^2}{\epsilon^2} \ ,$$

for $x < x_0 \exp(\lambda t)$, and

$$\mathrm{G}(x, t) = \mathrm{G}_\nu(x) + \frac{2}{\pi}\frac{x_0}{x} \left[\arctan \frac{x-x_0}{\epsilon} - \arctan \frac{x\exp(-\lambda t)-x_0}{\epsilon}\right] + \frac{\epsilon}{\pi x} \ln \frac{(x-x_0)^2+\epsilon^2}{[x\exp(-\lambda t)-x_0]^2+\epsilon^2} \ .$$

for $x > x_0 \exp(\lambda t)$.

Figure 6(a) shows the curves $\mathrm{G}_\nu(x)$ for three values of ϵ: viz $\epsilon = 0{\cdot}1$, $\epsilon = 100$, and $\epsilon = 1000$; and figure 6(b) shows the corresponding distribution curves $\mathrm{G}(x, t)$.

The distribution of city creation results in a turning of the distribution curve similar (but of a different appearance) to that due to a logistic growth.

Florida between 1960 and 1970 gives an example of such a distribution in city creation with the following values:

$$\mathrm{G}_\nu(10000) = 0{\cdot}090 \ , \qquad \mathrm{G}_\nu(5000) = 0{\cdot}346 \ ,$$

$$\mathrm{G}_\nu(3000) = 0{\cdot}504 \ , \qquad \mathrm{G}_\nu(2000) = 0{\cdot}729 \ .$$

2.7 *The form of the distribution curves*

$\mathrm{G}(x, t)$ usually has a complicated analytical expression, and so the study of the curve $y(u) = \ln[\mathrm{G}(\exp u, t)]$ on bilogarithmic paper would not be a straightforward one. Initial information about the curve is confined to its upward or downward turning; and we have the following result:

If the distribution function $G(x, t)$ results from a process (N, x) where x is a solution of $\dot{x} = h(x)$, the curve $y(u) = \ln[G(\exp u, t)]$, with $\ln x_0 \leqslant u$, is downward turning if the following conditions are fulfilled (where primes mean derivatives):
(a) $h(x) \geqslant x h'(x)$, for x in the interval $[x_0, x(t, x_0)]$;
(b) $N'(t)/N(t) \geqslant N''(t)/N'(t)$, for all t;
(c) $x_0 h'(x_0) - h(x_0) + x_0[(N''/N') - (N'/N)] < 0$.
Apart from a constant term we have:

$$y = \ln\{N[t - \tau(\exp u, x_0)]\} \Rightarrow \frac{dy}{du} = -F_1 F_2$$

with

$$F_1 = \frac{\exp u}{h(\exp u)}, \qquad F_2 = \frac{N'(t-\tau)}{N(t-\tau)}.$$

F_1 and F_2 are two positive functions which are steadily increasing if conditions (a) and (b) are satisfied. But the product of two positive increasing functions is also a positive increasing function. Thus y'' cannot vanish. Condition (c) secures precisely that

$$y''(u = \ln x_0) < 0 .$$

The interest of this result comes from the fact that the previous conditions are satisfied in most normal cases: they can easily be verified for the six processes (N, x),

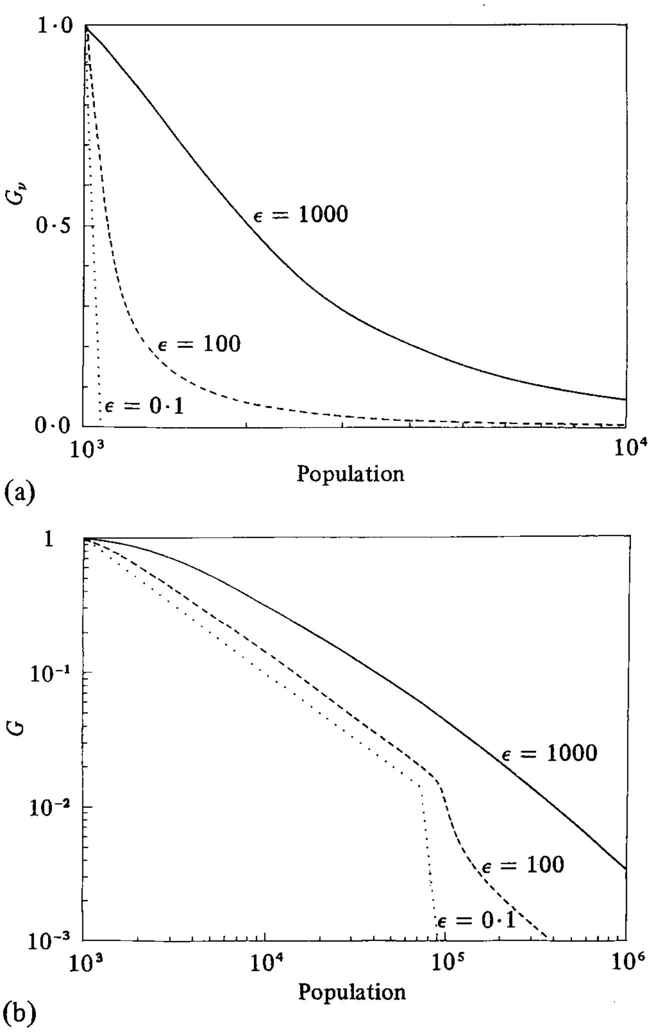

Figure 6. Curves of (a) $G_\nu(x)$ and (b) the corresponding distributions $G(x, t)$, for three values of ϵ. The lowest curve is for $\epsilon = 0 \cdot 1$, the middle curve for $\epsilon = 100$, and the upper curve for $\epsilon = 1000$.

where N and x are one of the following three functions: linear, exponential, and logistic. In fact, most of the observed distribution curves have such a form. The contrary can nevertheless also exist: it was observed for some districts of France in 1876.

3 Experimental verification

Given any initial distribution, equation (4) yields the resulting theoretical evolution of the distribution curve corresponding to a definite choice of $N(t)$ and $x(\tau, z)$. These curves (dashed lines on figure 7) are then compared to the statistical distribution curves. For the determination of the functions $N(t)$ and $x(\tau, z)$ we proceed as follows:
(1) The variation of the observed average growth rates for each size class gives an indication of the sort of function that should be used for $x(\tau, z)$. In any case, it should have a clear practical interpretation as was already indicated in section 2.2.

We have mainly used the solutions of the following differential evolution equation:

$$\dot{x} = \lambda(x-s) - \mu(x-s)^2 , \tag{5}$$

which is of the form discussed in section 2.2, and also of

$$\dot{x} = \lambda(x-s)^\gamma - \mu(x-s) , \qquad 0 < \gamma < 1 . \tag{6}$$

(2) The value of the parameters λ, μ, s, γ have been obtained numerically by a least square method.

Figure 7 shows that the theoretical curves fit the statistically observed distribution rather well, except perhaps for the largest cities, but we know already that our model can hardly be valid in that range. Our example shows the changes in the distribution

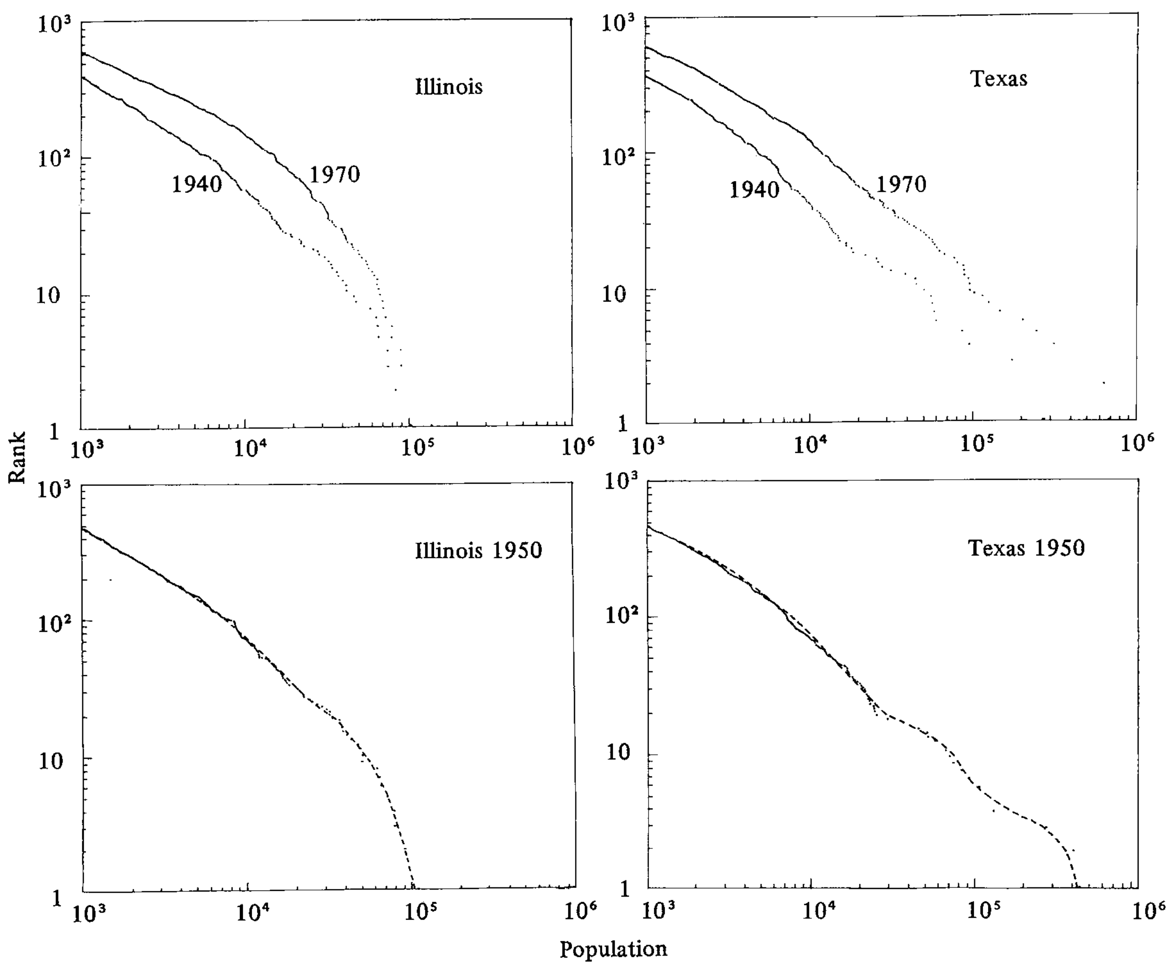

Figure 7. Theoretical and statistical distribution functions for Illinois and Texas.

curves for two American states over a rather short period of time. However, the modifications of the distribution curves are nevertheless noticeable (actually these states have been chosen for their fast demographic growth). In a paper to follow we shall present the same sort of curves for French districts and over a longer period.

Illinois was modelled by equation (6) with the values:

Period	Λ	λ	μ	s	γ
1940-1950	0·012	0·133	$1{\cdot}9 \times 10^{-7}$	474	0·889
1950-1960	0·018	0·301	$6{\cdot}1 \times 10^{-7}$	758	0·893
1960-1970	0·011	0·173	$4{\cdot}1 \times 10^{-8}$	810	0·904

Texas was modelled by equation (5) with the values:

Period	Λ	λ	μ	s
1940-1950	0·020	0·0417	$4{\cdot}6 \times 10^{-9}$	195
1950-1960	0·020	0·0352	$3{\cdot}5 \times 10^{-10}$	328
1960-1970	0·087	0·0172	$1{\cdot}7 \times 10^{-10}$	−162

Notice that we have an almost exponential growth (μ is very small) between 1950 and 1970. And in both cases, the influence of s is almost negligeable (except perhaps at the very beginning of the curves, that is for populations close to 1000).

Of course the deviation between the theoretical and statistical curves can become larger in some cases. But this situation is interesting in itself because it may lead (a) to the introduction of a new evolution equation which may be associated with a new kind of growth; or (b) to the conclusion that some other parameters (such as economic variables or central-place hierarchy) play an important role.

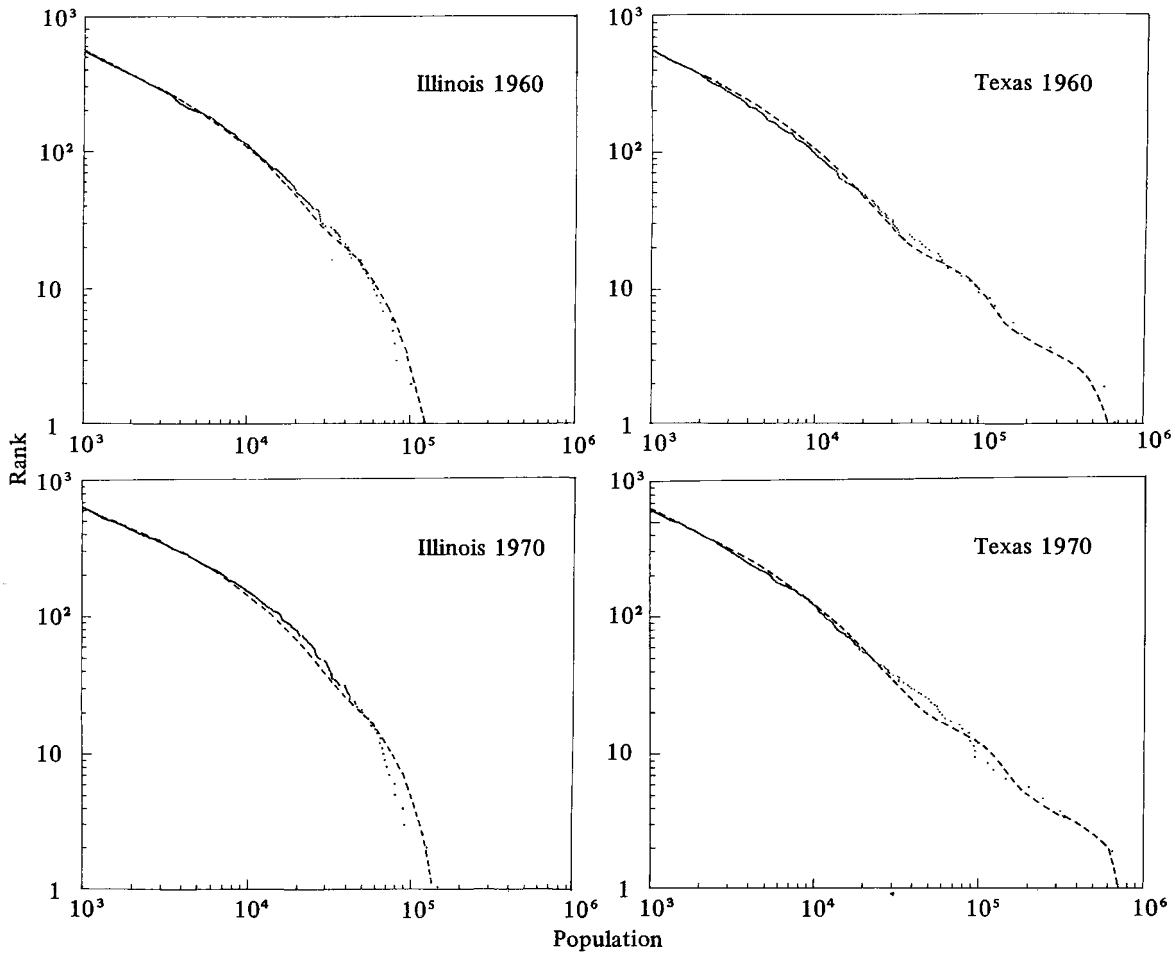

Figure 7 (continued)

4 The stochastic model

The transformation of the previous deterministic model into a stochastic one may be performed in two ways: viz

(a) by the introduction of random variables in the evolution equations, or

(b) by a description of the evolution of the city as a birth and death process.

We shall use here the second possibility and we want to see what modifications and further information it will entail.

4.1 *Definition of the stochastic model*

This model has already been used by Yule (1924) and Steindl (1965). Steindl used a non-Markovian recurrent process (Steindl, 1965, page 178) and studied only the stationary states.

The present description is Markovian and, we believe, best suited for the study of the transient states.

The population of each city will take the values 0, 1, 2, ...; and we take into account all states from zero to infinity because the state zero will be a natural limit for the birth and death processes we shall use. Nevertheless the distribution function could be observed above a certain minimal size, if desired. Also, let us denote by i a city of size i, by n_i the number of such cities, and by $\boldsymbol{n}$ the vector with components $n_0, n_1, n_2, \ldots$.

The following transitions will occur between these states.

(a) An increase of one unit in population of a city of size i in the time interval $(t, t+\Delta t)$:

$$(z \ldots n_i, n_{i+1} \ldots) \rightarrow (\ldots n_i - 1, n_{i+1} + 1 \ldots) ,$$

with probability $n_i \lambda_i \Delta t + \mathrm{O}(\Delta t)$, where λ_i is the birth rate of the birth and death process occurring in each city, and $\mathrm{O}(\Delta t)$ is an infinitesimal such that

$$\frac{\mathrm{O}(\Delta t)}{\Delta t} \underset{\Delta t \to 0}{\rightarrow} 0 .$$

(b) A loss of one unit in the population of a city of size i:

$$(\ldots n_{i-1}, n_i \ldots) \rightarrow (\ldots n_{i-1} + 1, n_i - 1 \ldots) ,$$

with probability $n_i \mu_i \Delta t + \mathrm{O}(\Delta t)$, where μ_i is the death rate of the birth and death process.

(c) Creation of a city of size i:

$$(\ldots n_i \ldots) \rightarrow (\ldots n_i + 1 \ldots) ,$$

with probability $\Lambda_i(\boldsymbol{n}) \Delta t + \mathrm{O}(\Delta t)$.

For the birth and death processes that we shall consider, increases and losses will be in fact births and deaths. Nevertheless, the consideration of birth and death processes with in-migration (Bailey, 1964, page 37) will make possible the description of migration phenomena.

We define the probability $p(t; \boldsymbol{n})$ to have at time t, n_i cities of kind i; and $p(t; \boldsymbol{n})$ will satisfy the evolution equations of a multidimensional birth and death process:

$$\begin{aligned} \dot{p}(t; \boldsymbol{n}) = & -\sum_{i \geqslant 0} (\Lambda_i + n_i \lambda_i + n_i \mu_i) p(t; \boldsymbol{n}) + \sum_{i \geqslant 0} p(t; \ldots n_i + 1, n_{i+1} - 1 \ldots)(n_i + 1) \lambda_i \\ & + \sum_{i \geqslant 0} p(t; \ldots n_{i-1} - 1, n_i + 1 \ldots)(n_i + 1) \mu_i \\ & + \sum_{i \geqslant 0} p(t; \ldots n_i - 1 \ldots) \Lambda_i(\ldots n_i - 1 \ldots) , \end{aligned}$$

with $p(0; \boldsymbol{n}) = \mathrm{f}(\boldsymbol{n})$ as the initial condition.

The previous equation cannot in general be solved exactly. The functions $p(t; \boldsymbol{n})$ are in any case too detailed in comparison with the existing statistical data. Such a situation occurs also in statistical mechanics and leads to the introduction of the so-called reduced distribution functions. We shall therefore introduce the average number of cities of kind k:

$$m_k(t) = \sum_{n_0, n_1, \ldots \geqslant 0} n_k p(t; \boldsymbol{n}) ,$$

$m_k(t)$ is the coefficient of θ_k in the development of the moment generating function:

$$M(t; \theta_0, \theta_1 \ldots) = \sum_{n_0, n_1 \ldots \geqslant 0} \exp(\theta_0 n_0) \exp(\theta_1 n_1) \ldots p(t; n_0, n_1 \ldots) .$$

The evolution equation of the moment generating function is a consequence of the evolution equation for $p(t, \boldsymbol{n})$:

$$\frac{\partial M}{\partial t} = \sum_{i \geqslant 0} \left\{ (\exp\theta_i - 1)\Lambda_i\left(\frac{\partial}{\partial\theta_0}, \frac{\partial}{\partial\theta_1} \ldots\right) + [\exp(\theta_{i+1} - \theta_i) - 1]\lambda_i \frac{\partial}{\partial\theta_i} + [\exp(\theta_{i-1} - \theta_i) - 1]\mu_i \frac{\partial}{\partial\theta_i} \right\} M .$$

By expanding to the first order we obtain the evolution equation of the $m_k(t)$:

$$\dot{\boldsymbol{m}} = \mathrm{A}\boldsymbol{m} + \boldsymbol{\Lambda}([\boldsymbol{m}]) , \qquad \boldsymbol{m}(t = 0) = \boldsymbol{m}^0 , \tag{7}$$

where A denotes the transition matrix of the birth and death process (λ_i, μ_i), and the notation $\boldsymbol{\Lambda}([\boldsymbol{m}])$ signifies:

$$\boldsymbol{\Lambda}([\boldsymbol{m}]) = a_0 + \sum a_k m_k + \sum a_{ij} m_{ij} + \ldots ,$$

if

$$\boldsymbol{\Lambda}(n_0, n_1, \ldots) = a_0 + \sum a_k n_k + \sum a_{ij} n_i n_j + \ldots .$$

In the two cases studied in the following sections $\boldsymbol{\Lambda}([\boldsymbol{m}])$ reduces to $\boldsymbol{\Lambda}(\boldsymbol{m})$.

Equation (7) is a linear one whenever $\boldsymbol{\Lambda}(\boldsymbol{m})$ is independent of m_k, which happens if (a) $\boldsymbol{\Lambda}(\boldsymbol{m})$ is a constant vector,

$$\Lambda_k(\boldsymbol{m}) = \Lambda_k ;$$

(b) $\Lambda_k(\boldsymbol{m})$ is proportional to the total number of cities existing at time t, which we denote by $M(t)$ [not to be confused with moment generating function $M(t; \theta_0, \theta_1, \ldots)$]

$$\Lambda_k(\boldsymbol{m}) = \Lambda_k[m_0(t) + m_1(t) + \ldots] .$$

In both cases, by addition of equation (7) we obtain easily the expression

$$M(t) = m_0(t) + m_1(t) + \ldots .$$

In the first case

$$M(t) = \Lambda t + M^0 , \qquad \Lambda = \Lambda_0 + \Lambda_1 + \ldots ,$$

where M^0 is the initial number of cities, and in the second case

$$M(t) = M^0 \exp(\Lambda t) ,$$

and $\Lambda_k(\boldsymbol{m})$ becomes independent of $\boldsymbol{m}$:

$$\Lambda_k(\boldsymbol{m}) = \Lambda_k M(t) .$$

The solution of equation (7) takes the form (Cartan, 1965, page 139)

$$\boldsymbol{m}(t) = \int_0^t \exp[\mathrm{A}(t-\tau)]\boldsymbol{\Lambda}(\tau)\mathrm{d}\tau + \exp(\mathrm{A}t)\boldsymbol{m}^0 .$$

The elements of the matrix $\exp A(t-\tau)$ are the transition probabilities of the birth and death process:

$$\exp[A(t-\tau)]_{nk} = p_{k\to n}(t-\tau) ,$$

where $p_{k\to n}(t-\tau)$ is the probability that a city with an initial population of k inhabitants has a population of n inhabitants at time $t-\tau$. This gives the general expression for the average number of cities having, at time t, k inhabitants:

$$m_k(t) = \sum_{k\geqslant 0}\int_0^t p_{k\to n}(t-\tau)\Lambda_k(\tau)\mathrm{d}\tau + \sum_{k\geqslant 0} p_{k\to n}(t)m_k^0 . \tag{8}$$

4.2 *Application to a Pareto distribution*

4.2.1 *Average exponential growth* First we consider the case that led to the Zipf–Pareto law in the deterministic model (see section 2.4.1):

(a) cities are created in a single state s, according to the exponential function

$$\Lambda_k(t) = \Lambda M^0 \exp(\Lambda t)\delta_{ks} ;$$

(b) the birth and death process is a pure birth process, so that

$$\lambda_n = \lambda n , \qquad \mu_n = 0 .$$

We shall be interested in the proportion $x_n(t)$ of the cities having n inhabitants:

$$x_n(t) = \frac{m_n(t)}{m_0(t)+m_1(t)+\ldots} .$$

Applying equation (8), we find that

$$x_n(t) = \Lambda\int_0^t p_{s\to n}(\theta)\exp(-\Lambda\theta)\mathrm{d}\theta + \exp(-\Lambda t)\sum_{k=s}^{n} x_k^0 p_{k\to n}(t)$$

with

$$p_{k\to n}(\theta) = \binom{n-1}{k-1}X^k(1-X)^{n-k} , \qquad n\geqslant k\geqslant 1 , \quad X = \exp(-\lambda t) .$$

This can be written by introducing the incomplete Beta function, $\mathrm{B}(x, y; \alpha)$,

$$x_n(t) = \rho\binom{n-1}{s-1}\mathrm{B}(n-s+1, s+\rho;\ 1-X) + \exp(-\Lambda t)\sum_{k=1}^{n} x_k^0\binom{n-1}{k-1}X^k(1-X)^{n-k} ,$$

where $\rho = \Lambda/\lambda$. For $t\to\infty$, we have

$$x_n(t) \to \rho\binom{n-1}{s-1}\mathrm{B}(n-s+1, s+\rho) ,$$

where $\mathrm{B}(x, y)$ is the ordinary Beta function. For large n we have the asymptotic distribution given by

$$x_n(\infty) \underset{n\to\infty}{\sim} \rho\frac{\Gamma(s+\rho)}{\Gamma(s)}\frac{1}{n^{1+\rho}} .$$

We obtain a Pareto distribution, with a Pareto coefficient which is again the ratio of the two growth rates; that is, the ratio of the average number of cities and of the average population of each city. But in contrast to the deterministic case, the Pareto distribution is obtained only asymptotically after transient behaviour, and figures 8(a) and 8(b) show that transient behaviour for two initial distributions, figure 8(a) when there is only one size of city initially, and figure 8(b) when there is an initial Pareto distribution. We conclude that the random character of city growth does not introduce here a considerable modification of the deterministic model.

4.2.2 *Average zero growth* Next we consider the evolution of the distribution functions when the average growth is zero. Using the Champernowne model, Vining has already studied the consequences of an absence of growth, in particular for urban planning.

In this case, the effects of the growth rate dispersions become fundamental and we can no longer avoid a stochastic analysis. We consider thus a linear birth and death process with equal birth and death rates:

$$\lambda_n = \lambda n \,, \qquad \mu_n = \lambda n \,.$$

Its stochastic mean remains constant

$$\bar{n}(t) = \sum_{n \geqslant 0} n p_{n_0 \rightarrow n}(t) = \bar{n}_0 \,.$$

The practical consequence for cities is that in each size class, the average population of the cities remains constant with an increase for some and a decrease for others.

We take $M(t) = M^0 \exp(\Lambda t)$, and for the sake of simplicity $s = 1$. Applying equation (8), we obtain

$$x_n(t) = \Lambda \int_0^t p_{1 \rightarrow n}(\theta) \exp(-\Lambda\theta) \mathrm{d}\theta + \exp(-\Lambda t) \sum_{k \geqslant 0} x_k^0 p_{k \rightarrow n}(t) \,.$$

In the limit of large time periods, we have

$$x_n(\infty) = \Lambda \int_0^\infty p_{1 \rightarrow n}(\theta) \exp(-\Lambda\theta) \mathrm{d}\theta \,.$$

With the $p_{1 \rightarrow n}(\theta)$ given in Bailey (1964, page 95) we obtain

$$x_n(\infty) = \rho \int_0^\infty \frac{u^{n-1}}{(1+u)^{n+1}} \exp(-\rho u) \mathrm{d}u \,, \qquad \text{for } n \geqslant 1 \,,$$

$$x_0(\infty) = \rho \int_0^\infty \frac{u}{1+u} \exp(-\rho u) \mathrm{d}u \,.$$

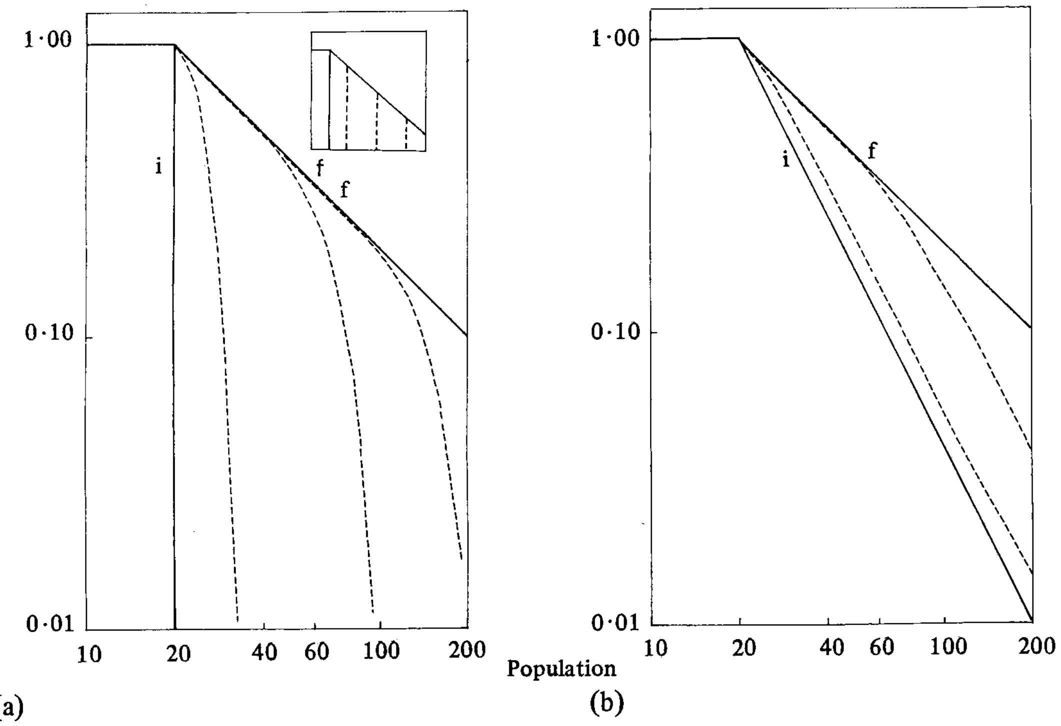

Figure 8. (a) Transient evolution leading to a Pareto distribution, with λt taking the values 0, 0·3, 1·3, 2·1, and ∞; and (b) transient evolution between two Pareto distributions with coefficient $\rho = 2$ and $\rho = 1$, and λt taking the values 0, 0·3, 1·3, and ∞.

The first integral does not have a simple expression. But an approximate expression may be obtained, if ρ is small enough so that the factor $\exp(-\rho u)$ can be omitted[(2)].

Hence we obtain the interesting result

$$x_n(\infty) \simeq \rho \int_0^\infty \frac{u^{n-1}}{(1+u)^{n+1}} \mathrm{d}u = \frac{\rho}{n}, \qquad n \geqslant 1 .$$

Proposition: *If the average population of the cities remains constant (in each size class) and if the number of emerging cities is small enough, the rank-size distribution, whatever the initial distribution, becomes closer to a Pareto distribution with coefficient 1.*

This result differs from Vining's result. He assumed a rather small decrease in average population of a city and a constant number of cities; and his Pareto coefficient was dependent on the variance of the growth rates.

The present result can explain why, for slowly increasing populations, a slope close to value 1 is so often observed.

Two points are now added:

(1) A linear evolution of the city number

$$M(t) = \Lambda t + M^0$$

leads again to a Pareto distribution with coefficient 1, that is,

$$x_n(T) = \frac{\rho}{\Lambda T + M^0} \frac{1}{n} ,$$

where T is a time large enough so that the second term in equation (8) can be neglected.

(2) Since the state 0 corresponds to cities with a zero population, the actual city number $N(t)$ will be given by

$$N(t) = m_1(t) + m_2(t) + \ldots = M(t)[1 - x_0(t)] ,$$

and $M(t) = \Lambda t + M^0$ yields

$$N(t) = \ln(1 + \lambda t) + M^0 + \rho .$$

The actual city number is a very slowly increasing function because of a balance between the newly created cities and the cities absorbed in the state 0. Thus the situation described corresponds indeed to an absence of growth.

5 Conclusion

The evolution of the rank-size distribution is a characterization of the urban growth which gives detailed information about the variations in growth rate with size, but smoothes over the individual irregularities. We showed that this evolution may often be described by three functions: the city number evolution function, the city-size evolution function, and the distributed creation function. A next step concerns the connection between those functions, migration phenomena, and socioeconomic activities.

Even in the cases where other parameters besides population are necessary, the present analysis permits an estimation of the proportion of the deviation that should be attributed to phenomena such as, for example, the influence of a central-place hierarchy.

In this paper, we stressed that a stochastic framework is necessary for the zero-growth situation. Also, some qualitative consequences of the model are that

(2) Such an approximation is, of course, not valid for $x_0(\infty)$, but $x_0(\infty)$ can be expressed in terms of the exponential integral function.

(a) the memory property of the distribution curves, (b) the most usual form of the distribution curves, and (c) the effect of a distribution of city creation.

Acknowledgements. The authors wish to thank G Bordes, D Courgeau, M Froissart, and D Pumain for their help and valuable suggestions.

References

Atkinson A B, 1975 *The Economics of Inequality* (Clarendon Press, London)

Bailey N T J, 1964 *The Elements of Stochastic Processes* (John Wiley, Chichester, Sussex)

Berry B J L, 1969 "City size and economic development: conceptual synthesis and policy problems, with special reference to South and South-East Asia" unpublished paper, Department of Geography, University of Chicago, Chicago, IL

Cartan H, 1965 *Calcul Différentiel* (Hermann, Paris)

Courgeau D, 1970 "Les champs migratoires en France" INED, 58, PUF

Desplanques G, 1975 "Les migrations intercensitaires de 1962 à 1968" *Les Collections de l'INSEE, Série D, No 39*

Ferguson R I, 1969 "The rank-size rule" *Geographical Articles* **12** 36-42; **13** 25-44; Department of Geography, University of Cambridge, Cambridge, England

Hill B, 1974 "The rank frequency form of Zipf's law" *Journal of the American Statistical Association* **69** 1017-1026

Historical Statistics of the USA, 1971 *Historical Statistics of the USA. Colonial Times to 1970* (US Government Printing Office, Washington, DC)

Malecki E, 1975 "Examining change in rank-size systems of cities" *Professional Geographer* **27** 43-47

Pumain D, 1982 *La Dynamique des Villes* (Economica, Paris)

Robson B T, 1973 *Urban Growth, an Approach* (Methuen, London)

Simon H, 1955 "On a class of skew distributions" *Biometrika* **42** 425-440

Steindl J, 1965 *Random Processes and the Growth of Firms* (Charles Griffin, High Wycombe, Bucks)

Vining D R, 1977 "The rank-size in the absence of growth" *Journal of Urban Economics* **4** 15-29

Woodroofe H B, 1975 "On Zipf's law" *Journal of Applied Probability* **12** (3) 425

Yule G U, 1924 "A mathematical theory of evolution based on the conclusions of Dr J C Willis FRS" *Philosophical Transactions of the Royal Society, Series B* **213** 21

Zipf G K, 1972 *Human Behavior and the Principle of Least Effort* (Hafner Press, New York)