# THE ERGODIC THEOREMS OF DEMOGRAPHY: A SIMPLE PROOF

W.B. Arthur

April 1981
WP-81-52



INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

## ACKNOWLEDGMENTS

I should like to thank Ansley Coale, Andras Por, and James Trussell for useful comments.

ABSTRACT

Standard proofs of the ergodic theorems of demography rely on theorems borrowed from positive matrix theory, tauberian theory, and the theory of time-inhomogeneous Markov matrices. But while these proofs are efficient and expedient, they give little direct insight into the mechanism that causes ergodicity.

This paper proposes a simple and unified proof of the two ergodic theorems. It is shown that the birth dynamics can be decomposed into a *smoothing process* that progressively levels out past fluctuations in the birth sequence and a *reshaping process* that accounts for current period-to-period changes in vital rates. The smoothing process, which causes the birth sequence to lose information on its past shape, is shown to be the ergodic mechanism behind both theorems.

THE ERGODIC THEOREMS OF DEMOGRAPHY:
A SIMPLE PROOF

W.B. Arthur

It is well known to mathematical demographers and population biologists that if the age-specific fertility and mortality patterns of a population remain unchanged over time, its age composition will converge to a fixed form, regardless of its initial shape. This is the Strong Ergodic Theorem of Demography, first proven by Lotka and Sharpe in 1911. And it is well known that if two populations start out with different age compositions, but are subjected to the same sequence of age-specific vital rates, changing over time, their age compositions will become increasingly alike, although changing too, of course, over time. This is the Weak Ergodic Theorem of Demography, conjectured by Coale in 1958 and proven by his student, Lopez, in 1961.

These two theorems stand at the center of mathematical demography. The first theorem makes *stable population theory* possible. Usually there is no clear or simple connection between fertility-mortality behavior and the age composition. But in the special case of unchanging vital rates, the theorem shows that a unique correspondence between age-specific life-cycle behavior and the age composition exists. We can use this correspondence, in demographic analyses, in population projections, and in the estimation of vital rates. The second theorem makes clear *which* vital rates determine the age composition. Only recent vital rates count, the influence of the initial age composition is progressively washed away. We need therefore only know recent demographic behavior if we want to determine the age structure of a population.

Proofs of both theorems are by now routinely available, rigorous, and standard. Strong ergodicity is proven either via positive matrix theory (invoking the Perron-Frobenius theorem) or by asymptotic integral equation theory (invoking tauberian theorems), depending on whether population dynamics are described in discrete or continuous time (see, for example, Leslie (1945), Parlett (1970), or Coale (1972)). Weak ergodicity is proven also by positive matrix theory, or alternatively by appeal to the theory of time-inhomogeneous Markov matrices (see, for example, Lopez (1961) or Cohen (1979)). But while these proofs are not inordinately difficult, they say little directly to our intuition. The mechanism causing ergodicity in both cases tends to lie hidden, obscured by the rather powerful mathematical apparatus needed for proof. Moreover, strong ergodicity appears to describe forces that push the age composition toward a fixed shape; weak ergodicity appears to describe forces that cause the age composition to shed information on its past. To the student unfamiliar with ergodic theory, it is not clear how the two theorems are related.

It turns out that there is a single and simple mechanism behind both types of ergodicity and it can be seen clearly without invocation of powerful outside theorems. This mechanism is the progressive smoothing or averaging of the birth sequence by the fact that both large and small past cohorts act together to produce a given year's crop of births. In this paper I will suggest a simple proof of both theorems based on this smoothing mechanism.

## 1. THE PROBLEM

A single-sex population evolves over time according to the Lotka dynamics

$$(1) \qquad B_t = \sum_x B_{t-x} p_{t,x} m_{t,x} \quad , \qquad t = 0,1,\ldots$$

where $B_t$ is the number of births in year t, $m_{t,x}$ is the proportion of those at age x who reproduce at that age in year t, and $p_{t,x}$ is the proportionate rise or fall in the cohort born in year t-x, due to mortality and migration. The *initial birth history*, $B_{-1},\ldots,B_{-N}$, is assumed given, for ages up to the oldest age N in the population. Summation in this case is understood to run from 1 to M, where M is the upper age limit of childbearing. This year's crop of births in other words is the sum of births born to those born in past years who survive and reproduce.

We can of course set time zero to any year we please, arbitrarily. This will be useful later. Also needed later is a technical condition. For certain ages fertility might well be zero. I assume that the net fertility age pattern is *non-periodic* in the age dimension: that is, that for all times t, there are at least two ages $a_1$ and $a_2$ (the same ages each time) which share no common divisor and have strictly positive fertility rates (greater than $\varepsilon$ say, some uniform constant).

The age composition, or proportion of the population at age a at time t, is given by the numbers at age a divided by the total population:

$$c_{t,a} = \frac{B_{t-a}p_{t,a}}{\sum_x B_{t-x}p_{t,x}} \quad . \tag{2}$$

Summation in this case is over all ages 1 to N in the population.

I now state the two theorems we want to prove.

*Weak Ergodic Theorem:* Two populations with different age compositions at time zero, if subjected to the same time-changing sequence of (non-periodic) fertility and mortality patterns, tend asymptotically to have identical but time-changing age compositions.

*Strong Ergodic Theorem:* The age composition of a population subjected to time constant patterns of (non-periodic) fertility and mortality tends asymptotically to a fixed form.

In looking for a proof of these theorems, we might start by noticing that the age composition of a population, once the vital rates are given, depends only on the birth sequence. We might therefore suspec' that strong and weak ergodicity reside somehow in the birth sequence itself. Looking further at both theorems we see that what is common in them is that the initial age composition before time zero eventually ceases to count. In the weak version it is progressively reshaped by events after time zero, identical events for two populations producing identical reshaping. In the strong version it is also progressively reshaped, but this time into a fixed form that we know and can predict. Translated to birth sequence terms, what we must show then is that the shape of the birth sequence before time zero, the *birth history*, ceases to determine the future course of the birth sequence as time passes. This is ergodicity.

In one special case, ergodicity in the birth sequence would be easy to show. This is where the net reproductive probabilities taken across all cohorts in each period sum to one. The size of any given year's birth cohort would then be a weighted average of the size of the reproductive cohorts. The birth sequence, under these circumstances, would "average its past"; it would smooth over time to a constant level; and it would therefore forget its initial shape.

In general things are not so simple. Reproductive levels vary from period to period, usually conforming to no particular level or trend. But the special case does suggest a strategy for proving ergodicity in general. Suppose we adjust the birth sequence by factors chosen carefully so that it smooths, as in the special case, to a constant level. We choose these factors to depend only on vital rates after time zero. Thus adjusted, the birth sequence must forget its initial shape. We now recover the actual birth sequence by the reverse adjustment process. By doing this we will reshape the smoothed adjusted sequence, but note that we will reshape it only according to the dictates of vital events *after time zero*. The initial birth history remains forgotten, smoothed away, and reshaping determines the future course of the actual birth sequence. If these operations are possible, ergodicity will be straightforward to show.

## 2. ERGODICITY IN THE BIRTH SEQUENCE

Following the strategy just outlined, we adjust the birth variable $B_t$ by a factor $r_t$, so that "adjusted births", $\hat{B}_t$, are

$$\hat{B}_t = B_t/r_t \quad . \tag{3}$$

We want to show first that for careful choice of the factors $r_t$, the adjusted birth sequence, $\hat{B}_t$, iterates to a constant level. Allowing ourselves some foresight, we choose the factors $r_t$ so that they evolve according to the dynamics

$$r_t = \sum_x p_{t,x} m_{t,x} r_{t-x} \tag{4}$$

from t = 0 onward, with the initial condition that $r_{-j} = g^{-j}$ where g is the real root of $1 = \sum p_{0,x} m_{0,x} g^{-x}$.

Dividing (4) through by $r_t$ we have

$$(5) \qquad 1 = \sum_x p_{t,x} m_{t,x} r_t^{-1} r_{t-x} \quad .$$

I shall call this the *generalized characteristic equation*[1]. Note that the r factors, thus chosen, act at each time to adjust total period fertility to one, and that they depend only on vital events after time zero.

Now rewrite the dynamics (1) by dividing through by $r_t$:

$$(6) \qquad B_t/r_t = \sum_x (B_{t-x}/r_t) p_{t,x} m_{t,x}$$

$$= \sum (B_{t-x}/r_{t-x}) p_{t,x} m_{t,x} r_t^{-1} r_{t-x} \quad .$$

Writing the terms $p_{t,x}\ m_{t,x}\ r_t^{-1}\ r_{t-x}$ as $\psi_{t,x}$ enables us to rewrite (7) simply as

$$(7) \qquad \hat{B}_t = \sum_x \hat{B}_{t-x} \psi_{t,x} \quad , \qquad \text{with } \sum_x \psi_{t,x} = 1$$

the last condition following from the generalized characteristic equation.

The original dynamics have been adjusted merely by dividing through by the variable factor $r_t$. But notice that for the adjusted birth sequence, for $\hat{B}_t$, we have a new dynamic process which is a weighted-averaging or smoothing process. $\hat{B}_t$ is a weighted average, with weights $\psi_{t,x}$, of the M immediately past $\hat{B}$-values. In turn $\hat{B}_{t+1}$ is a weighted average, with new weights $\psi_{t+1,x}$, of $\hat{B}_t$ and M-1 past $\hat{B}$-values. $\hat{B}_{t+2}$ is a weighted average of $\hat{B}_{t+1}$, $\hat{B}_t$ and M-2 past $\hat{B}$-values. And so on. This repeated averaging of the $\hat{B}$ sequence -- of averaging, then of averaging the averages -- we would expect intuitively, will converge $\hat{B}$ to a limiting constant value $\bar{B}$. I shall not

[1] In this form it is not obvious that this is a more general form of the familiar characteristic equation. If we put $r_t = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_t$ however, (5) becomes $1 = \sum p_{t,x} m_{t,x} \lambda_t^{-1}, \lambda_{t-1}^{-1}, \ldots, \lambda_{t-x+1}^{-1}$ where the summation is over reproductive ages. This reduces to the familiar form when there are no time variations.

give the full argument for this here.[1] Suffice it to say that at any time the next $\hat{B}$ value, providing the weights are positive and non-periodic, must lie within the spread or dispersion of the M past values it averages. Hence the $\hat{B}$ process progressively narrows its dispersion, eventually becoming trapped at a constant level.[2]

In the limit then, as time t tends toward infinity,

$$(8) \qquad \hat{B}_t = \bar{B} \quad , \text{ a constant} \quad .$$

Since $B_t = \hat{B}_t \cdot r_t$ we may recover the actual birth sequence quite simply by multiplying through by $r_t$. Hence in the limit

$$(9) \qquad B_t = r_t \bar{B} \quad .$$

In sum, the argument shows the birth dynamics to be a composite of two processes, one a process that smooths away the initial birth history to a constant $\bar{B}$, and the other a process that progressively reshapes this smoothed, adjusted birth sequence according to current vital events. This smoothing and reshaping of the birth sequence is illustrated in Figure 1.

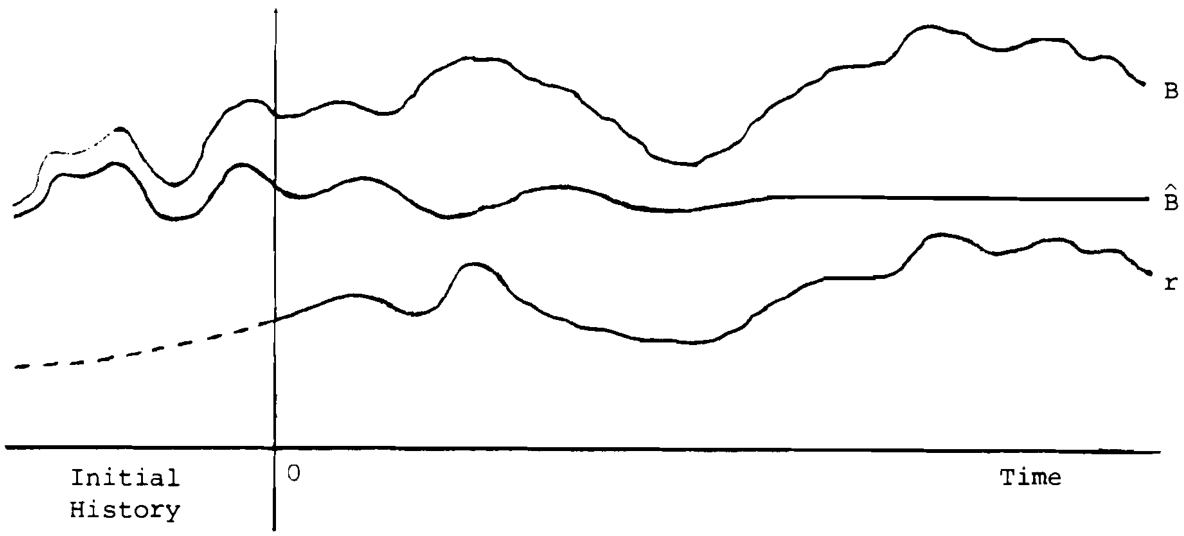


Figure 1

[1]For a renewal-theory proof that non-periodic smoothing processes iterate to a constant see Feller (1968 ; Vol. 1, Chapt. XIII). For an alternative proof, from first principles, see Arthur (1981). Periodic smoothing processes in general oscillate. They do not usually settle down to a limit.

[2]It is easy to show that this smoothing process for $\hat{B}$ converges within geometrically narrowing bounds.

We can now see clearly the ergodic mechanism at work within the birth sequence. All the information on the initial birth history is contained in the $\hat{B}$ smoothing process. But this information is repeatedly averaged away into a single constant so that the birth sequence "forgets" the shape it had before time zero. The r sequence reshapes this constant into the actual future birth sequence, but this reshaping sequence depends only on vital rates, and by definition only on these *after time zero*. Since the age composition is a simple transformation of the present birth sequence the two theorems follow immediately.

## 3. THE WEAK ERGODIC THEOREM: PROOF

By the decomposition, the age composition as $t \to \infty$, can be written

$$(10) \quad \lim_{t\to\infty} c_{t,a} = \frac{\bar{B} r_{t-a} p_{t,a}}{\sum_x \bar{B} r_{t-x} p_{t,x}} = \frac{r_{t-a} p_{t,a}}{\sum_x r_{t-x} p_{t,x}} .$$

Any two populations with different initial age compositions, but with identical time-changing vital rates will have the same reshaping sequence. Hence their age compositions, given by (10), will, in the limit, be identical. □

## 4. THE STRONG ERGODIC THEOREM: PROOF

In this case the vital rates are constant over time, if not over age: that is, $p_{t,x} = p_x$ and $m_{t,x} = m_x$. Let $\lambda$ be the real root of the equation

$$1 = \sum_x p_x m_x \lambda^{-x} .$$

We then see from the initial conditions for r and from (4) that $r_t$ equals $\lambda^t$. Thus r grows geometrically, and in turn so does the asymptotic birth sequence. Any population subjected to these unchanging vital rates will therefore, by (10), tend to the fixed age composition

$$(11) \qquad \lim_{t\to\infty} c_{t,a} = \frac{p_a}{\sum_x p_x \lambda^{a-x}} ,$$

which is a function constant in time and uniquely determined. □

## 5. REMARKS

1. Ergodicity, as shown in both theorems, would seem to be more a once and for all phenomenon than a continual shedding of past information. This of course is not the case. By shifting the arbitrary time zero reference point forward at will in the above proof we can show that the past is continually forgotten. Another way to see this is to notice that the r sequence is itself governed by exactly the same dynamics as the birth sequence. Therefore it too is ergodic. Therefore events *after* time zero, which determine r and equivalently the future movements of the birth sequence, progressively cease to count too. As time travels forward ergodicity follows behind. Just how fast the birth sequence forgets its past is an empirical question; Kim and Sykes (1976) have shown in a series of simulation experiments that in practical cases 50 to 75 years of vital data determines the age composition to a fair degree of accuracy.

2. Standard proofs of the weak ergodic theorem work by showing that the age compositions of two initially different populations become "closer" as defined by some norm, over time. This proves ergodicity of course, but indirectly in the sense that if two populations approach each other their different pasts *must* no longer count.[1] The above proof is different. It shows directly the ergodic mechanism operating within the single population as the progressive forgetting of the past birth sequence due to the natural spreading and smoothing out of reproduction. And it shows how the asymptotic age composition can be constructed from knowledge only of vital events after time zero.

---

[1]Among these two-population proofs is one due to McFarland (1969) that discusses the mechanism causing the approach of the two populations in some detail, and one by Lopez (1967) which uses a smoothing argument. The Lopez argument turns out to be closely related to the one given here. (To see this, note first that the initial conditions for $r_t$ were chosen to expedite the strong ergodic case, and that any initial history $r_{-1}, \ldots, r_{-M}$ would allow the proof to go through. If we identify $r_t$ with $B_2(t)$, the birth sequence of Lopez's second population which has an arbitrary initial history, then by the argument in this paper the two birth sequences $B(t)$ and $B_2(t)$ tend to a constant ratio. Resemblance of the two age compositions follows. This is the essence of Lopez's argument.)

3. In a series of remarkable and sophisticated theorems Cohen (1979) has recently shown that ergodicity extends to the stochastic case where vital rates are drawn from a sample set governed by a Markov process. From our viewpoint though, it would be surprising if ergodicity did not carry over to the stochastic case. What is important in the above smoothing argument is not the level of fertility rates but the fact that reproduction is spread over several, non-periodic ages. Providing this spreading property is preserved, actual levels can be chosen by a stochastic mechanism and we should still expect ergodicity to take place, although now with the degree of forgetting and reshaping subject to probability.

## 6. CONCLUSION

In this paper I have attempted to show a simple and unified proof of the two central theorems of demography. The proof relies on a simple decomposition of the birth sequence into a smoothing part inherent in spreading the replacement of population over several age groups and a reshaping part due to period-to-period changes in reproductive levels after time zero. It is the process of smoothing that averages out past humps and hollows in the birth sequence and this is the ergodic property -- the tendency to lose information on the past shape of the birth sequence -- that lies behind both theorems. In the Strong Ergodic case it causes the birth sequence to forget its initial shape and converge to geometric growth, and hence the age composition to assume a fixed form. In the Weak Ergodic case it causes the birth sequence to gradually lose information on its past shape, and to follow the period-to-period relative change in vital rates, and hence the age composition to be uniquely determined by recent demographic history.

## REFERENCES

Arthur, W.B. 1981. Why a population converges to stability. American Mathematical Monthly, forthcoming, June-July 1981.

Coale, A.J. 1972. The Growth and Structure of Human Populations. Princeton, New Jersey: Princeton University Press.

Cohen, J.E. 1979. Ergodic theorems in demography. Bulletin of the American Mathematical Society 1:275-295.

Kim, Y.J. and Z.M. Sykes. 1976. An experimental study of weak ergodicity in human populations. Theoretical Population Biology 10:150-172.

Leslie, P.H. 1945. On the use of matrices in certain population mathematics. Biometrika 33:183-212.

Lopez, A. 1961. Problems in Stable Population Theory. Princeton, New Jersey: Office of Population Research.

Lopez, A. 1967. Asymptotic properties of a human age distribution under a continuous net maternity function. Demography 4:680-687.

McFarland, D.D. 1969. On the theory of stable populations: a new and elementary proof of the theorems under weaker assumptions. Demography 6:301-322.

Parlett, B. 1970. Ergodic properties of populations I: the one sex model. Theoretical Population Biology 1:191-207.