

THE ERGODIC THEOREMS OF DEMOGRAPHY: A SIMPLE PROOF

W. Brian Arthur

Food Research Institute, Stanford University, Stanford, California 94305, USA

RR-83-16

June 1983

Reprinted from *Demography*, volume 19(4) (1982)

**INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria**

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Reprinted with permission from *Demography* 19(4):439–445, 1982.
Copyright © 1982 Population Association of America.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the copyright holder.

FOREWORD

For some years, IIASA has had a keen interest in mathematical demography. In this paper, reprinted from *Demography*, Brian Arthur takes a new look at the two ergodic theorems of demography. He shows that there is a single mechanism behind both types of ergodicity and that this mechanism provides the basis for a simple and unified proof of the two theorems.

THE ERGODIC THEOREMS OF DEMOGRAPHY: A SIMPLE PROOF

W. Brian Arthur

Food Research Institute, Stanford University, Stanford, California 94305

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. These proofs are efficient and expedient, but 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.

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 1957 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. Therefore we need know only 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, Coale, 1972; Leslie, 1945; or Parlett, 1970). Weak ergodicity is proven also by positive matrix theory or alternatively by appeal to the theory of time-inhomogeneous Markov matrices (see, for example, Cohen, 1979; or Lopez, 1961). 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 which 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 we will suggest a simple proof of both theorems based on this smoothing mechanism.

THE PROBLEM

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

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

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 size" at time t of the cohort then aged x (that is, its size in year t relative to its size in its birth-year $t-x$) where size-changes are caused by mortality and migration. The *initial birth history*, B_{-1}, \dots, 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. In other words, this year's crop of births is the sum of births born to those born in past years who survive and reproduce.

Of course, we can set time zero to any year we please, arbitrarily. This will be useful later. A technical condition also is needed later. For certain ages fertility

might well be zero. We will assume that the fertility age pattern $m_{t,x}$ fulfills a *no-common-divisor* condition at each time t : that is, that the fertility rate is strictly positive—greater than ε , some uniform constant—at at least two ages x_1 and x_2 (the same ages each time) which share no common divisor greater than one. (Thus, for example, a pattern of reproduction at ages 9, 10 and 12 fulfills this condition, whereas a pattern of reproduction at ages 10, 12 and 14 does not—there is a common divisor, 2.) Since consecutive integers have no common divisor, human reproduction, which takes place over a block of consecutive ages, fulfills the condition.

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:

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

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

We now state the two theorems we want to prove, assuming the populations we speak of have reproductive patterns that fulfill the no-common-divisor condition.

Weak Ergodic Theorem: Two populations with different age compositions at time zero, if subjected to the same time-changing sequence of 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 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. Therefore we might suspect 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*, and any vital events before time zero, cease 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 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.

ERGODICITY IN THE BIRTH SEQUENCE

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

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

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

$$r_t = \sum_x p_{t,x} m_{t,x} r_{t-x}, \quad (4)$$

from $t = M$ onward, with the initial sequence r_0, r_1, \dots, r_{M-1} given, and set at some arbitrarily chosen positive values. (For example, we can allow the initial r -values to be a sequence of 1's.) Notice that by this definition the r factors are always positive, and in particular they depend neither on the birth history nor on vital events that took place before time zero.¹

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

$$\begin{aligned} B_t/r_t &= \sum_x (B_{t-x}/r_t) p_{t,x} m_{t,x} \\ &= \sum_x (B_{t-x}/r_{t-x}) p_{t,x} m_{t,x} r_t^{-1} r_{t-x}, \\ &\quad \text{for } t = M \text{ onward.} \end{aligned} \quad (5)$$

Also, dividing (4) through by r_t , we have²

$$\begin{aligned} 1 &= \sum_x p_{t,x} m_{t,x} r_t^{-1} r_{t-x}, \\ &\quad \text{for } t = M \text{ onward.} \end{aligned} \quad (6)$$

Writing the terms $p_{t,x} m_{t,x} r_t^{-1} r_{t-x}$ as $\psi_{t,x}$ enables us to rewrite (5) and (6) together

simply as

$$\hat{B}_t = \sum_x \hat{B}_{t-x} \psi_{t,x}, \text{ with } \sum_x \psi_{t,x} = 1,$$

for $t = M$ onward. (7)

The original dynamics have been adjusted merely by dividing through by the variable factor r_t . But notice that the adjusted birth sequence, \hat{B}_t , "reproduces" itself with its own "net fertility function," ψ_t , that sums to one. In other words, the new dynamic process, in (7), for adjusted births is a weighted-averaging or repeated-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} as long as the weights $\psi_{t,x}$ are spread over several ages (the ages at which reproduction takes place in the original dynamics). Large and small past values of \hat{B} , thrown together by the averaging process, will determine the current \hat{B} at some value intermediate and any dispersion in the \hat{B} values will tend to disappear.

Exact conditions under which repeated-averaging processes smooth to a constant have been examined in Feller (1968) and in Arthur (1981).³ We will not repeat the analysis here. Suffice it to say that two outcomes are possible. Where the weights are positive at ages which are multiples of some integer d , then \hat{B} could oscillate indefinitely with period d . But where there is no common divisor d , as assumed earlier, smoothing to a constant *must* take place. In this case, at any time all \hat{B} values sufficiently far in the future must lie within the spread or dispersion of the M values just past. Hence the \hat{B} process must progressively narrow

its dispersion, eventually becoming trapped at a constant level.⁴

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

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

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 we have

$$\lim_{t \rightarrow \infty} (B_t - r_t \bar{B}) = 0. \quad (9)$$

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 for an illustrative birth sequence is shown in Figure 1.

We now can see clearly the ergodic mechanism at work within the birth sequence. All the information on the initial birth history and vital rates before time zero 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 in the past. 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.

THE WEAK ERGODIC THEOREM: PROOF

Recalling the definition of the age composition in (2), we now have

$$\lim_{t \rightarrow \infty} \left(\frac{B_{t-a} p_{t,a}}{\sum_x B_{t-x} p_{t,x}} - \frac{\bar{B} r_{t-a} p_{t,a}}{\sum_x \bar{B} r_{t-x} p_{t,x}} \right) = 0$$

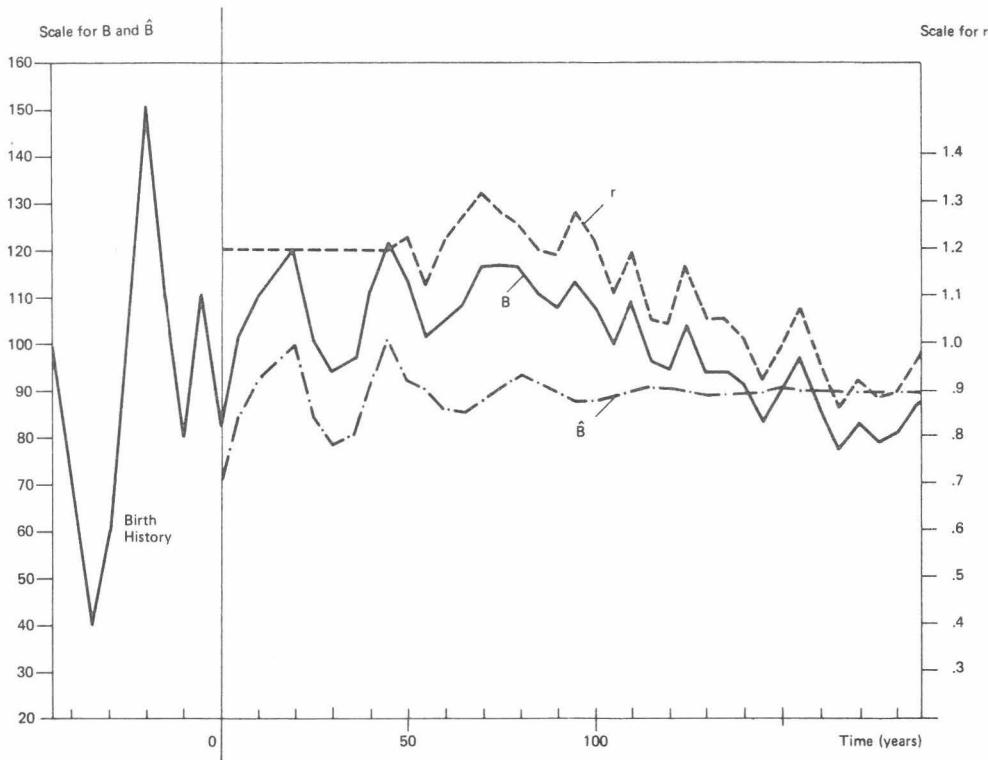


Figure 1.

that is,

$$\lim_{t \rightarrow \infty} \left(c_{t,a} - \frac{r_{t-a} p_{t,a}}{\sum_x r_{t-x} p_{t,x}} \right) = 0. \quad (10)$$

As time becomes large, the age composition becomes arbitrarily close to a composition that depends only on vital rates after time zero. Given two populations with different initial age compositions but subjected to identical vital events after time zero, we may choose their accompanying initial r values identically; then their reshaping sequences will be the same. By equation (10), their age compositions will become arbitrarily close.

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 λ be the unique real root of the equation

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

Since we may choose the initial r sequence, let it be given by $r_t = \lambda^t$ for times 0 to $M - 1$. It is then easy to show from (4) that these constant vital rates cause r_t to equal λ^t for all 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

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

which is a function constant in time and uniquely determined.

REMARKS

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 75 to 100 years of vital data determine the age composition to a fair degree of accuracy.

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.⁵ 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.

CONCLUSION

This paper has 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 gradually to 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.

NOTES

¹ Note that although certain "cohort lifetable" rates $p_{t,x}$ (those that pertain to cohorts born before time zero, that is, those for which $t < x$) are affected by events before time zero, equation (4) connects the r values with lifetable rates only *after* times t greater than M , the maximum age x occurring in (4). Hence r , as defined, is independent of vital events before time zero.

² This equation is in fact a *generalized characteristic equation*. If we put $r_t = \lambda_0 \cdot \lambda_1 \dots \lambda_t$ it becomes $1 = \sum p_{t,x} m_{t,x} \lambda_t^{-1} \cdot \lambda_{t-1}^{-1} \dots \lambda_{t-x+1}^{-1}$ where the summation is over reproductive ages. This reduces to the familiar characteristic equation when there are no time variations.

³ Feller uses renewal theory arguments. Arthur argues from first principles.

⁴ It is easy to show that this smoothing process for \hat{B} converges within geometrically narrowing bounds.

⁵ 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. If we identify, from time zero onward, the sequence 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.

ACKNOWLEDGMENTS

This paper was written while the author was a Research Scholar at the International Institute for Applied Systems Analysis, Laxenburg, Austria. The author thanks Ansley Coale, Joel Cohen, Omer Galle, Andras Por, James Trussell and an anonymous referee for valuable comments on a previous draft.

REFERENCES

- Arthur, W. B. 1981. Why a Population Converges to Stability. *American Mathematical Monthly* 88:557–563.
- Coale, A. J. 1957. How the Age Distribution of a Human Population is Determined. *Cold Spring Harbor Symposia in Quantitative Biology* 22:83–89.
- Coale, A. J. 1972. The Growth and Structure of Human Populations. Princeton, N.J.: Princeton University Press.
- Cohen, J. E. 1979. Ergodic Theorems in Demography. *Bulletin of the American Mathematical Society* 1:275–295.
- Feller, W. 1968. Introduction to Probability Theory and its Applications. Vol. 1, 3rd Ed. New York: Wiley.
- 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, N.J.: Office of Population Research.
- Lopez, A. 1967. Asymptotic Properties of a Human Age Distribution Under a Continuous Net Maternity Function. *Demography* 4:680–687.
- Lotka, A. J. and Sharpe, F. R. 1911. A Problem in Age Distribution. *Philosophical Magazine*, Ser. 6, 21:339–345.
- 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.

