

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

WHY A POPULATION CONVERGES TO STABILITY

W.B. Arthur

October 1979
WP-79-85

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

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



WHY A POPULATION CONVERGES TO STABILITY

W.B. Arthur

A large part of mathematical demography is built upon one fundamental theorem, the "strong ergodic theorem" of demography. If the fertility and mortality age-schedules of a population remain unchanged over time, its age distribution, no matter what its initial shape, will converge in time to a fixed and stable form. In brief, when demographic behavior remains unchanged, the population, it is said, converges to stability.

There are two basic ways to prove that this is so, depending on whether demographic behavior is described in discrete time, or in continuous time. For the discrete case, proof amounts to showing that an infinite product of the Leslie transition matrix achieves a limiting constant form. This is the principle behind the proofs of Leslie (1945), Lopez (1961), Parlett (1970), and many others. For the continuous case, proof amounts to solving the Lotka integral renewal equation and studying the asymptotic behavior of its solution terms. This is the principle behind the proofs of Lotka and Sharpe (1911), Lotka (1938), Coale (1972) and again, many others. Some of the modern papers (see, for example Cohen (1979)) probe the abstract, outer reaches of convergence, but the underlying principles remain largely the same.

While neither form of proof is mathematically difficult, neither offers much in the way of direct and ready insight. The problem is that both forms are built on borrowed theory, either on positive matrix theory or on asymptotic integral equation theory. Both forms of proof are not self-contained.¹ The mechanism forcing the age-structure to converge in each case therefore remains partially hidden within the borrowed theory and becomes difficult to see. Those who do not want to steep themselves in the theory of primitive matrices or in k-th order roots of integral equations are therefore left curious. Why should it be that a population converges? What is it about the process of regeneration of population numbers that means the age structure will converge to a stable form, and population growth to a constant rate? What mechanism underlies population convergence?

This short paper presents a new argument for the convergence of the age structure, one that is self-contained, and that brings the mechanism behind convergence into full view. The idea is simple. Looked at directly, the dynamics of the age-distribution say little to our normal intuition. Looked at from a slightly different angle though, population dynamics define a smoothing or averaging process over the generations -- a process comfortable to our intuition. This smoothing and resmoothing turns out to be the mechanism that forces the age structure toward a fixed and final form.

The Problem. The problem can be stated simply enough. Assuming constant fertility and mortality behavior, with no in- or out-migration, a population evolves in discrete time according to the dynamics

$$(1) \quad B_t = \sum_x B_{t-x} p_x m_x ,$$

where B_t is the number of births in year t , m_x is the probability of reproducing at age x , and p_x is the probability of surviving until age x . Present births, in other words, are the

1) Lotka did publish a self-contained proof, in 1922. He sandwiched the initial age-distribution between two boundary curves that close in over time, eventually coinciding to trap the age distribution within a fixed shape. The proof is ingenious but the logic is loose, and the mechanism forcing convergence is difficult to see.

sum of births born to people at childbearing ages who still survive. Summation is taken over the age groups 1 to M (where M is an upper limit to childbearing). And the numbers in the initial "generation", B_M, \dots, B_1 , are assumed given. Neither B , nor p , nor m , of course, is negative.

The age composition, or proportion at age a at time t , is given by

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

the numbers at age a , divided by the total population. We seek to prove that the distribution $c_{a,t}$ converges to a limiting constant function c_a^* .

Two observations will help us to furnish a proof. First, note that it is enough to prove that B_t converges to an exponential form, $B_t \rightarrow B^* e^{rt}$, where B^* and r are constants. For if this is true it follows by substituting for B_{t-a} and B_{t-x} in (2) that the age distribution becomes fixed and unchanging with time:

$$(3) \quad c_{a,t} \rightarrow \frac{e^{-ra} p_a}{\sum_x e^{-rx} p_x} = c_a^*.$$

Armed with this we can confine our attention to why the birth sequence, B_t , should become exponential. Second, note that convergence to an exponential form is hard to prove, the target is moving as it were; but convergence to a fixed value is easy. Therefore we will normalize or redefine the problem to one of convergence to a fixed value.

Smoothing Process. Begin with the dynamics

$$(4) \quad B_t = \sum_x B_{t-x} p_x m_x,$$

and divide both sides by e^{rt} ,

$$(5) \quad \frac{B_t}{e^{rt}} = \sum_x \frac{B_{t-x}}{e^{rt}} \frac{e^{-rx}}{e^{-rx}} p_x m_x.$$

Renaming Be^{-rt} to be the variable \hat{B}_t -- the "growth-corrected" birth sequence -- the new, but equivalent dynamics become

$$(6) \quad \hat{B}_t = \sum_x \hat{B}_{t-x} e^{-rx} p_x m_x .$$

We will speak somewhat loosely of \hat{B} in what follows as "births", remembering though that these "births" differ from real births by an exponential factor.

We now need only show that for some value of r , \hat{B} eventually becomes constant over time. Allowing ourselves some foresight, we choose r to satisfy

$$(7) \quad 1 = \sum_x e^{-rx} p_x m_x .$$

Finally, renaming $e^{-rx} p_x m_x$ as ψ_x , we may write the new but equivalent dynamics as

$$(8) \quad \hat{B}_t = \sum_x \hat{B}_{t-x} \psi_x$$

where, by virtue of (7),

$$1 = \sum_x \psi_x .$$

The original dynamics have been changed but little; B_t has merely been normalized to the new variable \hat{B}_t . Notice though, in the new system for \hat{B}_t , the coefficients $\psi(x)$ sum to one -- ψ is a weighting function. The new dynamics therefore describe a continuous smoothing process: \hat{B}_t is the weighted average of the M immediate past values of \hat{B} ; \hat{B}_{t+1} is the weighted average of \hat{B}_t and the $M-1$ immediate past values of \hat{B} ; \hat{B}_{t+2} the weighted average of \hat{B}_{t+1} , \hat{B}_t and the $M-2$ immediate past values of \hat{B} . And so on. This constant averaging, then averaging of the averages, we would suspect, will converge \hat{B} to a fixed value B^* (as in Figure 1), and equivalently will converge B to the exponential form $B^* e^{rt}$. Why?

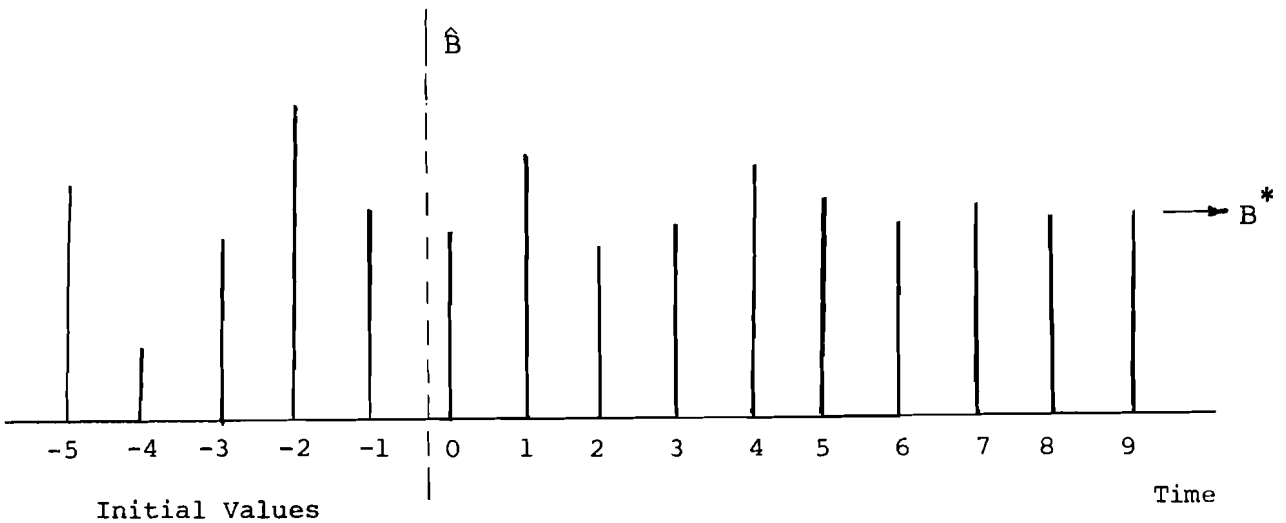


Figure 1

The reason is easy to sketch when all $\psi(x)$ are strictly positive (greater than ϵ say). Mark the largest of the initial birth values as \hat{B}_{-p} , the smallest as \hat{B}_{-q} , the difference between them being d . The initial given values therefore lie within a spread of d units. Now, the first value \hat{B}_0 , generated by the process will fall short of the greatest value \hat{B}_{-p} by at least ϵd units:

$$\begin{aligned} \hat{B}_0 &= \sum_{x \neq q} \hat{B}_{-x} \psi_x + \hat{B}_{-q} \psi_q \\ &= \sum_{x \neq q} \hat{B}_{-x} \psi_x + \hat{B}_{-p} \psi_q - \hat{B}_{-p} \psi_q + \hat{B}_{-q} \psi_q \end{aligned}$$

so that

$$(9) \quad \hat{B}_0 \leq \hat{B}_{-p} - \psi_q (\hat{B}_{-p} - \hat{B}_{-q}) < \hat{B}_{-p} - \epsilon d .$$

Similarly, \hat{B}_0 must exceed the smallest value \hat{B}_{-q} by at least ϵd units. Therefore \hat{B}_0 will lie strictly inside the initial spread of birth values -- inside by a fixed factor $1 - 2\epsilon$. The same argument applies all the more so to \hat{B}_1 , and again to \hat{B}_2 , and so on until \hat{B}_{M-1} . The spread of the entire new generation of \hat{B} values therefore lies strictly within that of the old one, and by a specified uniform factor. Repeating the argument over the generations, the generational spread in \hat{B} diminishes geometrically to zero. \hat{B}_t therefore converges to a fixed value B^* and

B_t therefore converges to exponential growth, $B^* e^{rt}$.

So far so good. But what if some of the ψ_x values are zero as in real populations where no reproduction takes place at certain ages? Will the process always converge? The answer is no.

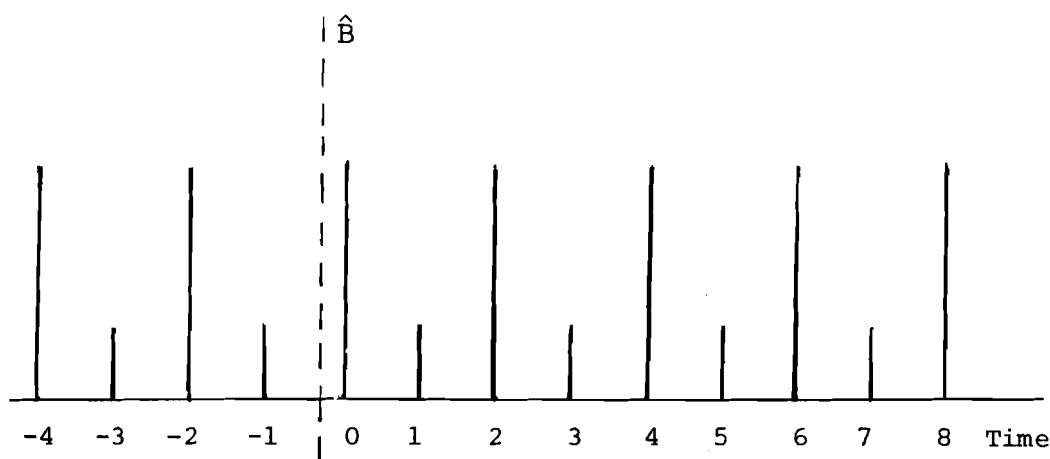


Figure 2

Consider the four-age-group population in Figure 2, with $\psi_1 = \psi_3 = 0$, and $\psi_2 = \psi_4 = \frac{1}{2}$. Childbearing occurs only in the second and fourth age-groups. This population will oscillate indefinitely. Here the smoothing process does not smooth: something is wrong. To see what, we need to look at smoothing more closely.

Smoothing - A Closer Look. In general, assume that some, or several, of the ψ values are zero. The value \hat{B}_0 then depends directly on only certain of the original \hat{B} values. Similarly, \hat{B}_1 depends directly only on the neighbors of these values. For the system

$$\hat{B}_t = \hat{B}_{t-3}\psi_3 + \hat{B}_{t-4}\psi_4 \quad ; \quad \psi_1, \psi_2 = 0 \quad ,$$

we can graph this dependence as in Figure 3a, picturing each birth cohort as a point, with a directed arrow drawn between them if dependent.

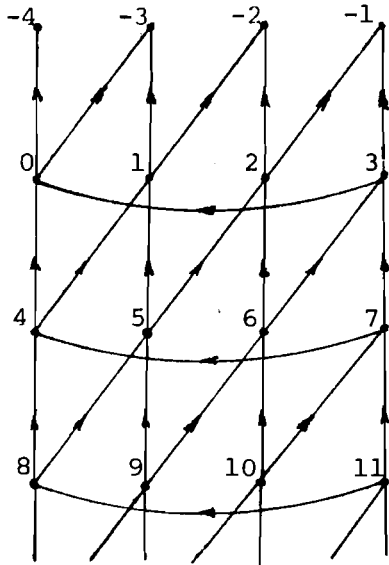


Figure 3a

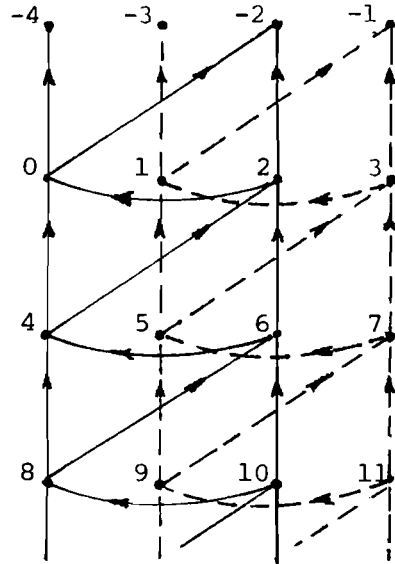


Figure 3b

The graph extends indefinitely downward. Notice though that while \hat{B}_0 depends directly on only two of the initial values, \hat{B}_4 depends on three of them, and \hat{B}_8 on all four of them. If we so chose, we could therefore write the dynamics with present \hat{B} values specifying \hat{B} eight steps ahead:

$$\hat{B}_{t+8} = \hat{B}_{t-1}\psi'_1 + \hat{B}_{t-2}\psi'_2 + \hat{B}_{t-3}\psi'_3 + \hat{B}_{t-4}\psi'_4 .$$

This process, with new weights ψ' , describes the evolution of \hat{B} perfectly well, moreover it remains a smoothing process as we can see by following the weights backward from \hat{B}_8 : they divide up but continue to sum to one. Most important, it is a function of *all* the initial values and is strictly positive in all its coefficients. We could therefore apply the convergence argument above, showing that the spread in any four consecutive values must be reduced eight steps ahead by a fixed factor. Taking \hat{B} values now twelve at a time (the original initial four plus the intervening eight), generational spread once again reduces geometrically; \hat{B} converges, even though we started with some ψ weights as zero.

What then went wrong with the case where ψ_1 and ψ_3 were zero? Forming its graph (Figure 3b) we see there is no future \hat{B}

value that is a function of all the original given values. Even-indexed \hat{B} 's depend on even-indexed \hat{B} 's; odd ones depend on odd ones. Here two separate but identical processes are going on: the even process never "sees" the initial values of the odd process and vice-versa. Both processes iterate their initial values to a limit: but there is an even limit and an odd limit. The process oscillates indefinitely between them.

We can therefore deduce a general principle. If the graph of the process, drawn one stage down and ignoring the directedness, is connected, some \hat{B} value can be written as an all-positive function of the initial values and \hat{B} will converge. If it is unconnected, two or more separate processes are going on. \hat{B} will have cluster points but no single limit; the process will oscillate. It is not too difficult to show that if B_0 depends on at least two initial points, $-j$ and $-k$, that are relatively prime (no common divisor other than 1), then the graph must connect. Sufficient for our purposes though is the more easily checked observation that if there are two consecutive positive reproductive ages, the graph must connect. Only one smoothing process then happens; hence \hat{B}_t converges to a limiting value B^* ; hence B_t converges to steady exponential growth $B^* e^{rt}$; hence the age composition converges to a fixed and stable form.

The Limiting Coefficient. One question remains. How can we determine B^* , the limiting coefficient of the exponential birth sequence? One possibility is to look for a quantity that is invariant, that is carried along unchanged over the generations. Such a quantity would enable us to relate B^* at the end of the process to the \hat{B} values at the beginning. Now, each generation at any time can donate to the future a certain number of direct descendants. On these direct descendants all future population must be built -- they are the system's "reproductive potential" or "reproductive value" as it were. We might suspect this reproductive potential, *in the growth-corrected dynamics* we have defined, to be invariant. A little algebra shows that this turns out to be the case.

At time t , age-groups $\hat{B}_{t-M}, \dots, \hat{B}_{t-1}$, taken together, contribute V_t direct descendants to the future -- to the period from t onward:

$$V_t = \hat{B}_{t-M}\psi_M + \hat{B}_{t-M+1}(\psi_{M-1} + \psi_M) + \dots + \hat{B}_{t-1}(\psi_1 + \psi_2 + \dots + \psi_M) .$$

(10)

Similarly age-groups $\hat{B}_{t+1-M}, \dots, \hat{B}_t$ contribute V_{t+1} to the period from $t+1$ onwards:

$$V_{t+1} = \hat{B}_{t-M+1}\psi_M + \hat{B}_{t-M+2}(\psi_{M-1} + \psi_M) + \dots + \hat{B}_t(\psi_1 + \psi_2 + \dots + \psi_M) .$$

Noting that the coefficient of \hat{B}_t is one, and using (8) to replace \hat{B}_t , we find

$$V_{t+1} = \hat{B}_{t-M+1}\psi_M + \hat{B}_{t-M+2}(\psi_{M-1} + \psi_M) + \dots + \hat{B}_{t-1}\psi_1 + \dots + \hat{B}_{t-M+1}\psi_{M-1} + \hat{B}_{t-M}\psi_M .$$

(11)

Comparing (10) with (11) term for term, we see that $V_t = V_{t+1}$; V_t is indeed an invariant quantity V .

At the start

$$\dot{V}_0 = V = \hat{B}_{-M}\psi_M + \hat{B}_{-M+1}(\psi_{M-1} + \psi_M) + \dots + \hat{B}_{-1}(\psi_1 + \psi_2 + \dots + \psi_M) .$$

(12)

And in the limit

$$V = B^*\psi_M + B^*(\psi_{M-1} + \psi_M) + \dots + B^*(\psi_1 + \dots + \psi_M)$$

(13)

$$V = B^*(\psi_1 + 2\psi_2 + \dots + M\psi_M) .$$

Since ψ is the distribution of childbearing in the population, the coefficient of B^* is the mean age of childbearing, denoted A_r . Putting (12) and (13) together yields the result we seek:

$$B^* = \frac{1}{A_r} \{ \hat{B}_{-M}\psi_M + \hat{B}_{-M+1}(\psi_{M-1} + \psi_M) + \dots + \hat{B}_{-1}(\psi_1 + \dots + \psi_M) \}$$

$$B^* = \frac{1}{A_r} \left\{ B_{-M} e^{rM} \psi_M + B_{-M+1} e^{r(M-1)} (\psi_{M-1} + \psi_M) + \dots + B_{-1} e^r (\psi_1 + \dots + \psi_M) \right\}$$

$$(14) \quad B^* = \frac{1}{A_r} \sum_{j=0}^{M-1} e^{rj} \sum_{x=-(M-j)}^{-1} B_{-x} p_{x+j} m_{x+j} .$$

The value B^* is directly determined by the initial birth sequence and the fertility and mortality age patterns.

Conclusion. To go back to the original question, why, in plain words, does a population converge? The argument presented here is both simple and new. Once the population's tendency to grow is eliminated, by dividing growth out of the dynamics, the process of population replacement, barring bizarre reproductive patterns, literally smoothes the generations out. Childbearing and hence the function ψ is not concentrated at one age but is spread over several years. Hence past humps and hollows in the birth sequence are thrown in together in the replacement process. They are averaged together -- they smooth out.

Adding growth back means that a smooth exponential increase is reached in the long run -- an exponential that is fully fixed given information from the initial birth sequence and the net fertility pattern. And once the birth sequence reaches exponential increase, the age-distribution must assume its stable shape, no matter what it started as.

References

- Coale, A.J., *The Growth and Structure of Human Populations*, Princeton University Press, Princeton, 1972.
- Cohen, J.E. "Ergodic Theorems in Demography," *Bulletin of the American Mathematical Society* 1:275-295, March 1979.
- Leslie, P.H., "On the Use of Matrices in certain Population Mathematics," *Biometrika* 33:183-212, 1945.
- Lotka, A.J. and Sharpe, F.R., "A Problem in Age-Distribution," *Philosophical Magazine*, Ser.6,21:435-438, 1911.
- Lotka, A.J., "The Stability of the Normal Age Distribution," *Proceedings of the National Academy of Sciences* 8:339-345, November 1922.
- Lotka, A.J., "A Contribution to the Theory of Self-Renewing Aggregates, with Special Reference to Industrial Replacement," *Annals of Mathematical Statistics* 10:1-25, March 1939.
- Lopez, A., *Problems in Stable Population Theory*, Office of Population Research, Princeton, 1961.