Recursive averages and the renewal theorem

Introduction

Consider the sequence of numbers a_n defined by the iteration

$$a_n = \sum_{r=1}^{k} p_r a_{n-r} \tag{1}$$

for $n \geq k$, where p_r $(1 \leq r \leq k)$ are non-negative numbers with $\sum_{r=1}^{k} p_r = 1$ and the starting values $a_0, a_1, \ldots, a_{k-1}$ are given. So a_n is a weighted average of the previous k terms.

In the *Gazette* paper [1], it was shown that if each p_r is strictly positive, then a_n converges to a limit L, which can be expressed in terms of the given numbers p_r and a_r for $1 \leq r \leq k$ (our notation is slightly different, for reasons that will soon become apparent). The article [2] gives an interesting geometrical illustration of essentially the same method. A completely different proof was given in another *Gazette* note [3] for the special case where $p_r = \frac{1}{k}$ for each r.

In at least one important application (which we describe shortly), some of the p_r will definitely be zero. In the method of [1], the condition that they are non-zero is unavoidable: it cannot be removed by any kind of minor adjustment. In fact, without this condition, the result amounts to a version (though not the most general one) of a deep result known as the "renewal theorem". Although this is purely a result of Analysis, it is recognised as an important topic in Probability Theory. A proof of the full version can be seen in [4, pp. 335–338]: it cannot be described as easy. Here we will present three quite different methods to prove our less general case, all much simpler than the one in [4], each offering its own distinctive perspective on the problem. Two of them use complex numbers in an essential way, one being a development of [3], while the third is a proof avoiding complex numbers, somewhat related to [1]. Of course, no reader is obliged to work through all three!

A trivial example shows that least one further condition must be needed. If $p_1 = 0$ and $p_2 = 1$, then the iteration is $a_n = a_{n-2}$. It is satisfied by $a_n = (-1)^n$, which does not tend to a limit. More generally, let $K(p)$ be the set of r such that $p_r > 0$, and let d be the greatest common divisor of the members of $K(p)$. If $d > 1$, then a_n could simply take different constant values on each congruence class modulo d. We will say that (p_r) satisfies condition (GCD) if $d = 1$.

An application: population dynamics

In a population (human or otherwise), let a_n be the number of births occurring in year n. Of those born, suppose that the proportion surviving to age r is s_r, and that this does not change over time. The values of a_n and s_r, once known, fully determine the total population and its age structure in subsequent years. Suppose that in each year, a proportion b_r of those age r (up to some limit $r = k$) give birth (this is the "age-specific fertility rate", again assumed to remain constant through time). So in year n (for $n \geq k$), those of age r number $a_{n-r}s_r$: they generate $a_{n-r}s_r b_r$ births. Hence a_n satisfies (1), with $p_r = b_r s_r$. It does not matter that in real life fertility rates vary with time. This analysis addresses the question of what would happen in the long term if the rates occurring in a particular year were to persist indefinitely.

Let $\sum_{r=1}^{k} p_r = P$: this is the "net reproduction rate". Our theorem, once proved, will state that if $P = 1$, then a_n will converge to a non-zero limit. It then follows easily that if $P < 1$, then a_n will tend to 0, while if $P > 1$, then a_n will tend to infinity.

In practice, this analysis is actually applied to the female population. The outcome for the male population (possibly with a different survival rate) can then be derived as an afterthought!

Since, even in these uninhibited times, human females aged less than (say) eight do not give birth, it is clear that in this application, p_r will be zero for the first few values of r.

This application also shows why the notation $p_r a_{n-r}$ arises naturally, in contrast to the notation $p_r a_r$ used in [1], [2] and [3].

This analysis was initiated by Lotka in a series of papers, e.g. [5], and continues to be regarded as a central principle in Demography. He formulated the basic theorem and gave heuristic reasoning that does not amount to a watertight proof. His informal reasoning has often been reproduced in later works on the subject, e.g. [6].

Convolutions; products of series

The notion of *convolution* will enable us to rewrite the problem in a way that accommodates a more general result and also opens the way to two of our methods.

We write just a to denote a sequence (a_n) (regarding a as a function defined on the non-negative integers, which is technically correct; we need this kind of notation). Given sequences $a = (a_n)$ and $b = (b_n)$ (defined for $n \geq 0$), their *convolution* is the sequence $a * b$ is defined by

$$(a * b)_n = \sum_{r=0}^{n} a_r b_{n-r}.$$

Of course, this can equally be written as $\sum_{r=0}^{n} a_{n-r} b_r$ or $\sum_{r+s=n} a_r b_s$. Convolutions arise naturally in the context of multiplication of polynomials or power series. Let $A(z) = \sum_{n=0}^{\infty} a_n z^n$ and $B(z) = \sum_{n=0}^{\infty} b_n z^n$ (in which the variable z could be real or complex). For the moment, assume that a and b have only finitely many non-zero terms (we will say that such sequences are "finitely non-zero"). Then the sums are finite (but we will continue to write them as infinite sums) and $A(z)$, $B(z)$ are polynomials. Collecting the z^n terms in the product, we have at once

$$A(z)B(z) = \sum_{n=0}^{\infty} (a * b)_n z^n \tag{2}$$

for all z. The case $z = 1$ says:

If $\sum_{n=0}^{\infty} a_n = A$ and $\sum_{n=0}^{\infty} b_n = B$, then $\sum_{n=0}^{\infty} (a * b)_n = AB$. (*)

In the case when a and b have infinitely many non-zero terms, statement (*) still holds provided that the series are absolutely convergent: this is a standard result in Analysis courses [7, p. 376–377]. $A(z)$ and $B(z)$ are now power series, with respective radii of convergence R_1, R_2 (possibly ∞). Let $R = \min(R_1, R_2)$. Power series are absolutely convergent within their radius of convergence, so (*) now applies to show that (2) holds for $|z| < R$.

The converse also holds, in the following sense. If a, b are two sequences such that $A(z) = B(z)$ within their common radius of convergence, then $a_n = b_n$ for all n: this is the uniqeness theorem for power series. Consequently, if we have three sequences a, b, c such that $A(z)B(z) = C(z)$ within the common radius of convergence, it follows that $c = a * b$.

(*Note in passing.* The *Dirichlet* convolution of sequences (a_n), (b_n) (defined for $n \geq 1$) has nth term $\sum_{rs=n} a_r b_s$. This arises from the multiplication of Dirichlet series $\sum_{n=1}^{\infty} a_n / n^s$. By contrast, our $a * b$ is sometimes called the *Cauchy* convolution.)

We will need a few more facts about convolutions. First, the sequence $e = (1, 0, 0 \ldots)$ is the identity for convolution: $a * e = a$ for all a. If $a_0 \neq 0$, then a has an inverse b: take $b_0 = \frac{1}{a_0}$ and define b_n recursively by: $a_0 b_n = -\sum_{r=0}^{n-1} a_{n-r} b_r$. For example, the inverse of $(1, 1, 0, 0, \ldots)$ is $(1, -1, 1, -1, \ldots)$, corresponding to the series identity $(1 + z)(1 - z + z^2 - \cdots) = 1$.

Lemma 1: If $a_n = 0$ for $n > k_1$ and $b_n = 0$ for $n > k_2$, then $(a * b)_n = 0$ for $n > k_1 + k_2$.

Proof: If $n > k_1 + k_2$ and $r \leq n$, then either $r > k_1$ or $n - r > k_2$. In either case, $a_r b_{n-r} = 0$.

Lemma 2: Write $B_n = \sum_{r=0}^{n} b_r$. Then $\sum_{r=0}^{n} (a * b)_r = \sum_{r=0}^{n} a_r B_{n-r}$.

Proof: Reversing the order of summation, we have

$$\sum_{r=0}^{n} (a * b)_r = \sum_{r=0}^{n} \sum_{s=0}^{r} a_s b_{r-s} = \sum_{s=0}^{n} a_s \sum_{r=s}^{n} b_{r-s} = \sum_{s=0}^{n} a_s \sum_{t=0}^{n-s} b_t.$$

By another routine exercise in reversing the order of summation (which we leave to any reader who is so inclined), one can show that convolution is associative: $(a * b) * c = a * (b * c)$

for sequences a, b, c. Consequently, we can write $a*b*c$ (and indeed longer strings) without brackets. In particular, it makes sense to consider the convolution of m copies of a, which we denote by a^{m*}. Of course, the notation $(a^{m*})_n$ simply means the nth term of this sequence.

Lemma 3: If $a_n \to L$ as $n \to \infty$ and $\sum_{n=0}^{\infty} b_n$ is absolutely convergent, with $\sum_{n=0}^{\infty} b_n = B$, then $(a*b)_n \to LB$ as $n \to \infty$.

Proof: It is enough to prove this statement for the case where $L = 0$, for then if $a_n \to L$, we have

$$\sum_{r=0}^{n} a_{n-r}b_r = \sum_{r=0}^{n}(a_{n-r} - L)b_r + LB_n \to 0 + LB \quad \text{as } n \to \infty.$$

So suppose that $a_n \to 0$. Then (a_n) is certainly bounded: there exists M such that $|a_n| \le M$ for all n. Also, let $\sum_{n=0}^{\infty} |b_n| = B^*$. Choose $\varepsilon > 0$. There exists N such that $|a_n| \le \varepsilon$ for $n > N$ and also $\sum_{n=N+1}^{\infty} |b_n| \le \varepsilon$. For $n > 2N$, we then have

$$\left|\sum_{r=0}^{N} a_{n-r}b_r\right| \le \varepsilon \sum_{r=0}^{N} |b_r| \le \varepsilon B^*,$$

$$\left|\sum_{r=N+1}^{\infty} a_{n-r}b_r\right| \le M \sum_{r=N+1}^{\infty} |b_r| \le \varepsilon M.$$

Another way in which convolutions arise has to be mentioned, though it will not be used in our proofs (so readers may ignore it). Let X be a random variable taking the values r (for integers $r \ge 0$) with probability p_r, so that $\sum_{r=0}^{\infty} p_r = 1$. Let Y be another random variable taking the value r with probability q_r. If X and Y are independent, then the probability that $X + Y = n$ is $(p*q)_n$. Further, the "expectation" $E(X)$ is $\sum_{r=1}^{\infty} rp_r$: denote it by μ. Let $E(Y) = \nu$. Then $E(X+Y) = \sum_{n=1}^{\infty} n(p*q)_n$. One can verify (again by reversing the order of summation) that this equals $\mu + \nu$: this is a standard result in Probability Theory. So it follows, for example, that $\sum_{r=1}^{\infty} r(p^{m*})_r = m\mu$.

Identification of a possible limit, and statements of the theorems

Returning to our problem, let us restate it in convolution notation. Extend the definition of p_r by setting $p_r = 0$ for $r > k$, also $p_0 = 0$. Then (1) says that $a_n = (p*a)_n$ for $n \ge k$. Meanwhile for $n < k$, we have

$$a_n - (p*a)_n = a_n - \sum_{r=1}^{n} p_r a_{n-r}.$$

This is a combination of the given values as far as a_n, and there is no reason why it should be zero. Denoting it by g_n for $0 \le n \le k-1$, and setting $g_n = 0$ for $n \ge k$, we can state

$$a_n = \sum_{r=1}^{n} p_r a_{n-r} + g_n \tag{3}$$

for all $n \geq 0$, in other words, $a = p * a + g$.

This reproduces all the original information. It also gives us a useful alternative formulation of the problem. Regard g_n as given, rather than a_n (with $g_n = 0$ for $n \geq k$). Then define a_n recursively by (3) for all $n \geq 1$, starting with $a_0 = g_0$. This opens the possibility of generalisation to the case where (p_r) and (g_n) are not finitely non-zero. (In the population example, g_n equates to the contribution to a_n of the population present in year 0.)

Now let

$$q_r = p_{r+1} + \cdots + p_k$$

for $0 \leq r \leq k-1$, also $q_r = 0$ for $r \geq k$. Then $q_0 = 1,\ 0 \leq q_r \leq 1$ for all r and $q_r = 1 - P_r$, where $P_r = p_1 + \cdots + p_r$. Also,

$$\sum_{r=0}^{k-1} q_r = \sum_{r=0}^{k-1} \sum_{s=r+1}^{k} p_s = \sum_{s=1}^{k} p_s \sum_{r=0}^{s-1} 1 = \sum_{s=1}^{k} s p_s.$$

Denote this by μ (as seen above, it is the expectation of the probability distribution (p_r).)

Also, write $G_n = \sum_{r=0}^{n} g_r$. Of course, $G_n = G_{k-1}$ for all $n \geq k$: denote this by γ. The next Lemma is pivotal in our development. It is similar to the reasoning in [1], in different notation.

Lemma 4: With q and G defined this way, we have $q * a = G$.

Proof: Write also $A_n = \sum_{r=0}^{n} a_r$. By (3) and Lemma 2,

$$\begin{aligned} G_n = A_n - \sum_{r=0}^{n} (p * a)_r &= A_n - \sum_{r=0}^{n} a_r P_{n-r} \\ &= A_n - \sum_{r=0}^{n} a_r (1 - q_{n-r}) \\ &= \sum_{r=0}^{n} a_r q_{n-r}. \end{aligned}$$

This algebra is reversible: $q * a = G$ is actually equivalent to $a = p * a + g$.

By Lemma 3, if a_n tends to a limit L, then G_n tends to $L \sum_{r=0}^{k-1} q_r = L\mu$. But G_n has the constant value γ for all $n \geq k$. So the only possible candidate for L is γ/μ.

Further, $\gamma = (q * a)_{k-1} = \sum_{r=0}^{k-1} q_r a_{k-1-r}$, a combination of the values $a_0, a_1, \ldots, a_{k-1}$. For our original version of the problem, it is natural to express our candidate limit in this way. The fact that $G_n = \gamma$ for $n > k$ expresses the fact that the limit is the same if $a_{n-k}, \ldots, a_{n-1}$ are regarded as the starting values. Obviously, any combination of values that is supposed to represent the limit has to pass this test.

We are now ready to set out the exact statement of our Theorems.

Theorem 1: Suppose that $p_r \geq 0$ for $1 \leq r \leq k$, that $\sum_{r=1}^k p_r = 1$ and that (p_r) satisfies condition (GCD). Let $\sum_{r=1}^n rp_r = \mu$. Suppose also that numbers g_n are given, with $g_n = 0$ for $n \geq k$ and $\sum_{n=0}^{k-1} g_n = \gamma$. Let a_n be defined recursively by $a_0 = g_0$ and

$$a_n = \sum_{r=1}^n p_r a_{n-r} + g_n$$

for all n. Then

$$a_n \to \frac{\gamma}{\mu} \quad \text{as } n \to \infty.$$

As we have seen, this implies the following variant, in line with our original version of the problem.

Theorem 2: Suppose that (p_r) is as in Theorem 1. Write $q_r = p_{r+1} + \cdots + p_k$ for $0 \leq r \leq k-1$. Suppose also that $a_0, a_1, \ldots, a_{k-1}$ are given and $a_n = \sum_{r=1}^k p_r a_{n-r}$ for $n \geq k$. Then $a_n \to L$ as $n \to \infty$, where

$$L = \frac{\sum_{r=0}^{k-1} q_r a_{k-1-r}}{\sum_{r=0}^{k-1} q_r}.$$

For the special case where $p_r = \frac{1}{k}$ for $1 \leq r \leq k$, one can easily deduce the well-known formula $L = [2/k(k+1)] \sum_{r=1}^k r a_{r-1}$.

The formulation of Theorem 1, unlike Theorem 2, points to the full version of the renewal theorem. For this, (p_r) and (g_n) are given *infinite* sequences satisfying $\sum_{r=1}^\infty p_r = 1$, $\sum_{r=1}^\infty rp_r = \mu$ and $\sum_{n=0}^\infty g_n = \gamma$. The conclusion is the same as before. As we progress, we will consider the extent to which our methods stretch to this case.

The sub-unital case and an easy special case

Let $\sum_{r=1}^k p_r = P$. We call the case $P = 1$ "unital", and the case where $P < 1$ "sub-unital". Of course, the unital case is our main concern, but we can make good use of the sub-unital one. Observe next that in the unital case, if $|a_n| \leq M$ for $0 \leq n \leq k-1$, then, by an obvious induction, $|a_n| \leq M$ for all n. We can deduce the following very easy result for the sub-unital case (condition (GCD) is not needed):

Lemma 5: Suppose that a_n satisfies (1) for $n \geq k$, with $\sum_{r=1}^k p_r = P < 1$. Then for some M and $\lambda > 1$, we have $|a_n| \leq M/\lambda^n$ for all n, hence $a_n \to 0$ as $n \to \infty$.

Proof: Let $c_n = \lambda^n a_n$, where λ is to be chosen. Then for $n \geq k$,

$$c_n = \lambda^n \sum_{r=1}^k p_r a_{n-r} = \lambda^n \sum_{r=1}^k p_r \lambda^{r-n} c_{n-r} = \sum_{r=1}^k p_r \lambda^r c_{n-r}.$$

By the intermediate value theorem, there exists $\lambda > 1$ such that $\sum_{r=1}^{k} p_r \lambda^r = 1$. By the preceding remark, (c_n) is bounded, say $|c_n| \leq M$ for all n. So $|a_n| \leq M/\lambda^n$ for all n.

Of course, it follows, not only that $a_n \to 0$, but that $\sum_{n=0}^{\infty} a_n$ is convergent: let its sum be A. We can actually identify A. Recall our basic equation $a - p * a = g$. By (*), we have $\sum_{n=0}^{\infty} (p * a)_n = PA$. Hence $A - PA = \gamma$, so $A = \gamma/(1 - P)$.

More importantly, the case of Theorem 1 where $p_r > 0$ for $1 \leq r \leq k$ now follows very easily (in fact, a little more easily than in [1]). Let $b_n = a_n - L$, where $L = \gamma/\mu$. Then for $n \geq k$,

$$\sum_{r=1}^{k} p_r b_{n-r} = \sum_{r=1}^{k} p_r a_{n-r} - L = a_n - L = b_n \tag{4}$$

and

$$\sum_{r=1}^{k} q_{r-1} b_{n-r} = \sum_{s=0}^{k-1} q_s a_{n-1-s} - L\mu = G_{n-1} - L\mu = \gamma - L\mu = 0. \tag{5}$$

For some $\delta > 0$, we have $p_r \geq \delta$ for $1 \leq r \leq k$. Then $p_r - \delta q_{r-1} \geq 0$ for each r, and $\sum_{r=1}^{k} (p_r - \delta q_{r-1}) = 1 - \delta\mu$. But by (4) and (5),

$$b_n = \sum_{r=1}^{k} (p_r - \delta q_{r-1}) b_{n-r}.$$

for $n > k$. So by Lemma 5, $b_n \to 0$ as $n \to \infty$.

Without further refinement, this reasoning does not adapt to the case where some p_r are zero. For example, if $p_5 = p_6 = \frac{1}{2}$, then q has a block of six non-zero terms: no translation of them can be covered by non-zero p_r. The refinement needed is substantial: we describe it in Method 3 below.

Method 1: complex power series

This certainly the slickest of our methods. It is a rather striking example of Complex Analysis being applied to prove a result in Real Analysis. The scheme is as follows. Recall that $q * a = G$. Since $q_0 = 1$, we know that q has an inverse with respect to convolution: denote it by c, so $c * q = e$, the identity. Hence $a = (c * q) * a = c * (q * a) = c * G$. Now $G_n = \gamma$ for all $n \geq k$, so if we can show that $\sum_{n=0}^{\infty} c_n$ is absolutely convergent, then Lemma 3 shows that a_n tends to $\gamma \sum_{n=0}^{\infty} c_n$.

Let $P(z) = \sum_{r=1}^{k} p_r z^r$ and $Q(z) = \sum_{n=0}^{k-1} q_r z^r$. Note that $P(1) = 1$ and $Q(1) = \mu$.

Lemma 6: We have $(1 - z)Q(z) = 1 - P(z)$.

Proof: Since $q_0 = 1$ and $q_{r-1} - q_r = p_r$ for $r \geq 1$,

$$(1 - z)Q(z) = q_0 + \sum_{r=1}^{k} (q_r - q_{r-1}) z^r = 1 - \sum_{r=1}^{k} p_r z^r = 1 - P(z).$$

We now deploy the assumption that (p_r) satisfies (GCD): this is to be understood in the following Lemmas.

Lemma 7: $P(z) \neq 1$ for complex $z \neq 1$ with $|z| = 1$.

Proof: Recall that $K(p)$ denotes the set of r such that $p_r > 0$. Suppose that $|z| = 1$ and $P(z) = 1$, so also $\mathrm{Re}\, P(z) = 1$. If $\mathrm{Re}\, z^r < 1$ for some r in $K(p)$, then $\mathrm{Re}\, P(z) < \sum_{r=1}^{k} p_r = 1$. So for each r in $K(p)$, we have $\mathrm{Re}\, z^r = 1$, so in fact $z^r = 1$.

By the generalised Bezout identity in number theory, there exist $r_1, \ldots, r_j$ in $K(p)$ and integers $n_1, \ldots, n_j$ (some positive, some negative) such that $\sum_{i=1}^{j} n_i r_i = 1$. Then $z = \prod_{i=1}^{j} z^{n_i r_i} = 1$.

Lemma 8: $Q(z) \neq 0$ for all complex z with $|z| \leq 1$.

Proof: First, $Q(1) = \mu > 0$. For $z \neq 1$, Lemma 6 shows that $Q(z) = 0$ if and only if $P(z) = 1$. By Lemma 7, this does not occur for $z \neq 1$ with $|z| = 1$. If $|z| < 1$, then $|z^r| < 1$ for all r, so $|P(z)| < \sum_{r=1}^{k} p_r = 1$.

Lemmas 6,7 and 8 appear in [4, chap. 13], but our final step does not. We apply the following theorem of Complex Analysis: *if a function $f(z)$ is differentiable (alias holomorphic) for $|z| < R$, then it is given by a power series for $|z| < R$.*

Lemma 9: There exist $R > 1$ and a sequence $c = (c_n)$ such that $\sum_0^\infty c_n z^n = 1/Q(z)$ for $|z| < R$. Further, $c * q = e$ and $\sum_{n=0}^\infty c_n$ is absolutely convergent, with sum $1/\mu$.

Proof: Let the complex zeros of the polynomial $Q(z)$ be $z_1, z_2, \ldots, z_k$. By Lemma 7, $|z_i| > 1$ for each i, so if R is the smallest $|z_i|$, then $R > 1$. Then for all z with $|z| < R$, we have $Q(z) \neq 0$, so $1/Q(z)$ is well defined and differentiable: denote it by $C(z)$. By the theorem just quoted, $C(z)$ is given by a power series $\sum_{n=0}^\infty c_n z^n$ for such z. So $C(z)Q(z) = 1$: as explained earlier, this implies that $c * q = e$. Also, $\sum_{n=0}^\infty c_n = C(1) = 1/Q(1) = 1/\mu$. Convergence is absolute, as always with power series.

By Lemma 3, it follows that $a_n \to \gamma/\mu$ as $n \to \infty$, completing the proof of Theorem 1.

One might be tempted to suppose that this method will extend easily to the case where (p_r) and (g_n) are infinite sequences. Lemmas 6, 7 and 8 do indeed extend with no trouble. The snag is that $Q(z)$ may well have radius of convergence 1, and fail to be defined for $|z| > 1$. For example, this happens with $q_r = 2/(r+1)(r+2)$. In this situation, we can only conclude that $\sum_{n=0}^\infty c_n z^n$ converges for $|z| < 1$, giving no information about the case $z = 1$, which is what we want. This might seem like a small distinction, but it is critical. In fact, convergence of $\sum_{n=0}^\infty |c_n|$ equates to a deep result known as "Wiener's Tauberian theorem": which is best proved using the spectral theory of Banach algebras (e.g. see [8, p. 333]).

Method 2: an explicit expression for a_n

This is a familiar method in the context of difference equations. We consider the scenario of Theorem 2 rather than Theorem 1. The recurrence (1) will be satisfied by $a_n = z^n$ if $f(z) = z^k$, where

$$f(z) = \sum_{r=1}^{k} p_r z^{k-r} = p_1 z^{k-1} + \cdots + p_{k-1} z + p_k.$$

With $q_r = p_{r+1} + \cdots + p_k$, we have $z^k - f(z) = (z-1)g(z)$, where

$$g(z) = \sum_{r=0}^{k-1} q_r z^{k-1-r} = q_0 z^{k-1} + \cdots + q_{k-2} z + q_{k-1}.$$

Let the complex factorisation of $g(z)$ be $\prod_{i=1}^{k-1}(z - z_i)$. Also, let $z_0 = 1$. Then any linear combination of the form $\sum_{i=0}^{n} c_i z_i^n$ satisfies (1). As we now show, if the z_i are all distinct, then a_n can be expressed as such a combination.

Theorem 3: Let a_n satisfy (1) for $n \geq k$, with $a_0, a_1, \ldots, a_{k-1}$ given. Suppose that $z_1, z_2, \ldots, z_k$ are distinct. Let $\sum_{r=0}^{k-1} q_r = \mu$. Then there exist c_i $(0 \leq i \leq k-1)$ such that

$$a_n = c_0 + \sum_{i=1}^{k-1} c_i z_i^n \tag{6}$$

for all n. Also, $c_0 = \frac{1}{\mu} \sum_{r=0}^{k-1} q_r a_{k-1-r}$.

Proof: We need to choose the c_i so that (6) holds for $0 \leq n \leq k-1$, matching the given starting values a_n. This is possible, because the vectors $(1, z_i, z_i^2, \ldots, z_i^{k-1})$ for $0 \leq i \leq k-1$ are linearly independent: they form the "van der Monde matrix". Also,

$$\sum_{r=0}^{k-1} q_r a_{k-1-r} = c_0 \mu + \sum_{i=1}^{k-1} c_i g(z_i) = c_0 \mu.$$

Of course, non-real z_i occur in conjugate pairs z_i, $\overline{z_i}$. The sum in (6) becomes real when $c_i z_i^n$ is combined with $\overline{c_i z_i}^n$. Also, any real z_i is negative, because $g(x) > 0$ for $x > 0$.

Lemma 10: If (p_r) satisfies (GCD), then $|z_i| < 1$ for $1 \leq i \leq k-1$.

Proof. First, observe that $g(1) > 0$, so $z_i \neq 1$. If $|z| > 1$, then $|z|^r < |z|^k$ for $r < k$, so $|f(z)| < |z|^k$. Hence $|z_i| \leq 1$. If $|z| = 1$ and $z \neq 1$, then, as in Lemma 7, $z^r = 1$ for each r in $K(p)$, and hence $z = 1$.

Theorem 2 now follows for the case where the z_i are distinct: each z_i^n tends to 0, so $a_n \to c_0$.

It was shown in [3] that the z_i are distinct in the special case where $p_r = \frac{1}{k}$ for each r. But in general they may well fail to be distinct. An actual example is easily constructed by choosing the z_i first and deducing the p_r: let $k = 3$ and $z_1 = z_2 = -\frac{1}{3}$. Then $g(z) = (z + \frac{1}{3})^2 = z^2 + \frac{2}{3}z + \frac{1}{9}$, given by $q_1 = \frac{2}{3}$ and $q_2 = \frac{1}{9}$, so by $p_3 = \frac{1}{9},\ p_2 = \frac{5}{9},\ p_1 = \frac{1}{3}$.

In [1, p. 220], it is stated that the method of [3] can be "easily modified" to prove the general result, apparently overlooking the possibility of repeated z_i. In fact, this case requires quite a bit more work. First, we describe the solution when one z_i (say z_1) is repeated just once.

Lemma 11: If z_1 is a repeated zero of $g(z)$, then (1) is satisfied by $a_n = nz_1^n$.

Proof: Write $F(z) = z^k - f(z) = (z-1)g(z)$. Since $F(z_1) = 0$, we have $f(z_1) = \sum_{r=1}^k p_r z_1^{k-r} = z_1^k$. Also, $g'(z_1) = 0$, so $F'(z_1) = 0$, hence

$$0 = z_1 F'(z_1) = kz_1^k - \sum_{r=1}^k (k-r)p_r z_1^{k-r},$$

so

$$\begin{aligned}
\sum_{r=1}^k p_r a_{n-r} &= \sum_{r=1}^k (n-r)p_r z_1^{n-r} \\
&= z_1^{n-k}\sum_{r=1}^k [(n-k)+(k-r)]p_r z_1^{k-r} \\
&= (n-k)z_1^{n-k}\sum_{r=1}^k p_r z_1^{k-r} + z_1^{n-k}\sum_{r=1}^k (k-r)p_r z_1^{k-r} \\
&= (n-k)z_1^{n-k}z_1^k + kz_1^{n-k}z_1^k \\
&= nz_1^n = a_n.
\end{aligned}$$

Suppose now that $z_2 = z_1$, while the other z_i are distinct. In (6), replace z_2^n by nz_1^n. The resulting set of vectors is still linearly independent, so can still be matched with the starting values $a_0, a_1, \ldots, a_{k-1}$.

The reader will not be surprised to hear that if z_i occurs three times, then $n^2 z_i^n$ satisfies (1), and so on. The verification becomes increasingly complicated as we progress to higher numbers of repetitions, and we will refrain from spelling it out in detail. If z_i is repeated m times, then we obtain m_i corresponding vectors, which are inserted into (6). By a fairly simple extension of the van der Monde result, the resulting set of vectors is still linearly independent. Theorem 2 still follows, because for any $r \geq 1$, $n^r z_i^n \to 0$ as $n \to \infty$.

Of course, there is no possibility of this method extending to infinite sequences.

Method 3: repeated convolutions, reduction to the sub-unital case

We have seen two proofs of this theorem of Real Analysis using complex numbers in an essential way. When this happens, it is a natural challenge to find a proof avoiding complex numbers. A particularly famous instance where this occurs is the prime number theorem.

The idea of this method (which, I think, is not well known) was already seen in the "easy special case" described earlier. With $b_n = a_n - L$, the aim is to show that b satisfies a condition like the one in Lemma 5, so tends to zero.

Given our iteration $a_n = \sum_{r=1}^{k} p_r a_{n-r} + g_n$, a natural idea is to substitute for a_{n-r} using the same identity again, thereby obtaining a double sum expressing a_n in terms of certain previous values. Convolution notation describes this step very neatly: given $a = g + p * a$, apply p again and substitute to find $a = g + p * g + p * p * a$. Repeating this, we obtain:

Lemma 12: For each $m \geq 1$,

$$a = g + p * g + \cdots + p^{(m-1)*} * g + p^{m*} * a. \tag{7}$$

Proof. Induction on m. Assuming the statement for m, apply p again to obtain

$$p * a = p * g + p * p * g + \cdots + p^{m*} * g + p^{(m+1)*} * a.$$

Since $a = g + p * a$, this implies the identity for $m + 1$.

We remark that any attempt to express this identity without convolution notation would involve increasingly complicated multiple sums, a nightmare even to contemplate!

Write formula (7) as $a = t + p^{m*} * a$, where $t = g + p * g + \cdots + p^{(m-1)*} * g$. This is an identity of the same sort as the original $a = g + p * a$, with p^{m*} replacing p. By Lemma 1, $(p^{j*} * g)_n = 0$ for $n > (j+1)k$, so for $n > mk$, we have $t_n = 0$, hence $(p^{m*} * a)_n = a_n$.

Now let us come back to $b_n = a_n - L$. We can restate (4) as $b - p * b = h$, where all that matters is that $h_n = 0$ for $n \geq k$ (actually, $h = g - Lq$, but we do not need this.) By Lemma 12, applied to b and h instead of a and g, we see that $(p^{m*} * b)_n = b_n$ for $n > mk$.

We now incorporate condition (GCD). From the definition of convolution, it is obvious that if u, v are non-negative sequences with $r \in K(u)$ and $s \in K(v)$, then $r + s \in K(u * v)$. Applying this repeatedly, we see that if $r_i \in K(p)$ and c_i are positive integers for $1 \leq i \leq I$, with $\sum_{i=1}^{I} c_i = c$, then $\sum_{i=1}^{I} c_i r_i \in K(p^{c*})$.

Lemma 13: If (p_r) satisfies (GCD), then there exist positive integers c, d, N such that if $u = \frac{1}{2}(p^{c*} + p^{d^*})$, then both N and $N + 1$ are in $K(u)$. Further, $n \in K(u^{k*})$ for $kN \leq n \leq k(N + 1)$.

Proof: Separating the positive and negative terms in the Bezout identity, we see that there exist r_i, s_j in $K(p)$ and positive integers c_i, d_j such that $\sum_{i=1}^{I} c_i r_i = N$ and $\sum_{j=1}^{J} d_j s_j = N+1$ for some N. Let $\sum_{i=1}^{I} c_i = c$ and $\sum_{j=1}^{J} d_j = d$. Then $N \in K(p^{c^*})$ and $N+1 \in K(p^{d*})$. Hence both are in $K(u)$. Now let $n = kN+r$, where $0 \le r \le k$. Then $n = (k-r)N + r(N+1)$, hence $n \in K(u^{k*})$.

(Of course, if $K(p)$ itself contains two consecutive integers, we just take $u = p$.)

Clearly, $(u * b)_n = b_n$ for all large enough n. Write $u^{k*} = v$. Then, as with a and p, we have $(v * b)_n = b_n$ for all large enough n. Also, the non-zero values of v occur for $r \le R$ (for some R) and $\sum_{r=1}^{R} v_r = 1$. Our final step is to use a translation of q to replace this by a sum less than 1. Recall that (5) says, in convolution notation, that $(q * b)_n = 0$ for $n \ge k$. Let $\tilde{q}_n = q_{n-kN}$ for $n \ge kN$, also $\tilde{q}_n = 0$ for $n < kN$. Then for $n \ge k(N+1)$,

$$(\tilde{q} * b)_n = \sum_{r=kN}^{n} \tilde{q}_r b_{n-r} = \sum_{s=0}^{n-kN} q_s b_{n-kN-s} = (q * b)_{n-kN} = 0.$$

Completion of the proof of Theorem 1. By Lemma 13, there exists $\delta > 0$ such that $v_r \ge \delta$ for $kN \le r \le kN + 1$. Let $w = v - \delta\tilde{q}$. Then $w_r = 0$ for $r > R$ and $(w * b)_n = b_n$ for large enough n. To conclude that $b_n \to 0$ using Lemma 5, we need to know that $w_r \ge 0$ and $\sum_{r=1}^{R} w_r < 1$. Now $\tilde{q}_r \le 1$ for all r and $\tilde{q}_r$ is only non-zero for $kN \le r \le k(N+1)$, so $w_r \ge 0$ for all r. Since $\sum_{r=1}^{R} \tilde{q}_r = \mu$, we have $\sum_{r=1}^{R} w_r = 1 - \delta\mu$.

With a little extra effort, this method can be modified to deal with infinite sequences. Wherever we previously had finitely non-zero sequences, we now have sequences that have to be shown to tend to 0, usually using Lemma 3. This applies also to differences between two sequences. Some proofs, especially Lemma 5 and the final step, become a little more delicate. Even with these further refinements, the method is still arguably more elementary than the proof in [4].

References

1. H. Flanders, Averaging sequences again, *Math. Gaz.* **80** (1996), 219–222.
2. N. Lord, Sequences of averages revisited, *Math. Gaz.* **95** (2011), 314–317.
3. P. Sanders, Averaging sequences, *Math. Gaz.* **78** (1994), 326–328.
4. W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley (1971).
5. A. J. Lotka, A contribution to the theory of self-renewing aggregates, with special reference to industrial replacement, *Ann. Math. Stats.* **10** (1939), 1–25.
6. N. Keyfitz, *Introduction to the Mathematics of Population*, Addison-Wesley (1968).
7. T. M. Apostol, *Mathematical Analysis*, Addison-Wesley (1957).
8. G. Bachman and L. Narici, *Functional Analysis*, Academic Press (1966).

13 Sandown Road, Lancaster LA1 4LN, UK
e-mail: *pgjameson@talktalk.net*