Chapter 0

A Branching-Process Example

(This Chapter is not essential for the remainder of the book. You can start with Chapter 1 if you wish.)

0.0. Introductory remarks

The purpose of this chapter is threefold: to take something which is probably well known to you from books such as the immortal Feller (1957) or Ross (1976), so that you start on familiar ground; to make you start to think about some of the problems involved in making the elementary treatment into rigorous mathematics; and to indicate what new results appear if one applies the somewhat more advanced theory developed in this book. We stick to one example: a branching process. This is rich enough to show that the theory has some substance.

0.1. Typical number of children, X

In our model, the number of children of a typical animal (see Notes below for some interpretations of 'child' and 'animal') is a random variable X with values in $\mathbf{Z}^+$. *We assume that*

$$\mathsf{P}(X=0)>0.$$

We define the *generating function f of X* as the map $f:[0,1]\to[0,1]$, where

$$f(\theta):=\mathsf{E}(\theta^X)=\sum_{k\in\mathbf{Z}^+}\theta^k\mathsf{P}(X=k).$$

Standard theorems on power series imply that, for $\theta\in[0,1]$,

$$f'(\theta)=\mathsf{E}(X\theta^{X-1})=\sum k\theta^{k-1}\mathsf{P}(X=k)$$

and

$$\mu:=\mathsf{E}(X)=f'(1)=\sum k\mathsf{P}(X=k)\le\infty.$$

Of course, $f'(1)$ is here interpreted as

$$\lim_{\theta\uparrow 1}\frac{f(\theta)-f(1)}{\theta-1}=\lim_{\theta\uparrow 1}\frac{1-f(\theta)}{1-\theta},$$

since $f(1)=1$. *We assume that*

$$\mu<\infty.$$

Notes. The first application of branching-process theory was to the question of survival of family names; and in that context, animal = man, and child = son.

In another context, 'animal' can be 'neutron', and 'child' of that neutron will signify a neutron released if and when the parent neutron crashes into a nucleus. Whether or not the associated branching process is supercritical can be a matter of real importance.

We can often find branching processes embedded in richer structures and can then use the results of this chapter to start the study of more interesting things.

For superb accounts of branching processes, see Athreya and Ney (1972), Harris (1963), Kendall (1966, 1975).

0.2. Size of n^{th} generation, Z_n

To be a bit formal: suppose that we are given a doubly infinite sequence

(a) $$\left\{X_r^{(m)} : m,r\in\mathbf{N}\right\}$$

of independent identically distributed random variables (IID RVs), each with the same distribution as X:

$$\mathsf{P}(X_r^{(m)}=k)=\mathsf{P}(X=k).$$

The idea is that for $n\in\mathbf{Z}^+$ and $r\in\mathbf{N}$, the variable $X_r^{(n+1)}$ represents the number of children (who will be in the $(n+1)^{\text{th}}$ generation) of the r^{th} animal (if there is one) in the n^{th} generation. The fundamental rule therefore is that if Z_m signifies the size of the n^{th} generation, then

(b) $$Z_{n+1}=X_1^{(n+1)}+\cdots+X_{Z_n}^{(n+1)}.$$

We assume that $Z_0=1$, so that (b) gives a full recursive definition of the sequence $(Z_m : m\in\mathbf{Z}^+)$ from the sequence (a). Our first task is to calculate the distribution function of Z_n, or equivalently to find the generating function

(c) $$f_n(\theta) := \mathsf{E}(\theta^{Z_n}) = \sum \theta^k \mathsf{P}(Z_n = k).$$

0.3. Use of conditional expectations

The first main result is that for $n \in \mathbf{Z}^+$ (and $\theta \in [0,1]$)

(a) $$f_{n+1}(\theta) = f_n(f(\theta)),$$

so that for each $n \in \mathbf{Z}^+$, f_n is the n-fold composition

(b) $$f_n = f \circ f \circ \ldots \circ f.$$

Note that the 0-fold composition is by convention the identity map $f_0(\theta) = \theta$, in agreement with – indeed, forced by – the fact that $Z_0 = 1$.

To prove (a), we use – at the moment in intuitive fashion – the following very special case of the very useful *Tower Property of Conditional Expectation*:

(c) $$\mathsf{E}(U) = \mathsf{E}\mathsf{E}(U|V);$$

to find the expectation of a random variable U, first find the conditional expectation $\mathsf{E}(U|V)$ of U given V, and then find the expectation of *that.* We prove the ultimate form of (c) at a later stage.

We apply (c) with $U = \theta^{Z_{n+1}}$ and $V = Z_n$:

$$\mathsf{E}(\theta^{Z_{n+1}}) = \mathsf{E}\mathsf{E}(\theta^{Z_{n+1}}|Z_n).$$

Now, for $k \in \mathbf{Z}^+$, the conditional expectation of $\theta^{Z_{n+1}}$ given that $Z_n = k$ satisfies

(d) $$\mathsf{E}(\theta^{Z_{n+1}}|Z_n = k) = \mathsf{E}(\theta^{X_1^{(n+1)} + \cdots + X_k^{(n+1)}}|Z_n = k).$$

But Z_n is constructed from variables $X_s^{(r)}$ with $r \le n$, and so Z_n is independent of $X_1^{(n+1)}, \ldots, X_k^{(n+1)}$. The conditional expectation given $Z_n = k$ in the right-hand term in (d) must therefore agree with the absolute expectation

(e) $$\mathsf{E}(\theta^{X_1^{(n+1)}} \ldots \theta^{X_k^{(n+1)}}).$$

But the expression at (e) is a *expectation of the product of independent random variables* and as part of the family of '*Independence means multiply*' results, we know that this expectation of a product may be rewritten as the product of expectations. Since (for every n and r)

$$\mathsf{E}(\theta^{X_r^{(n+1)}}) = f(\theta),$$

we have proved that

$$\mathsf{E}(\theta^{Z_{n+1}} | Z_n = k) = f(\theta)^k,$$

and this is what it means to say that

$$\mathsf{E}(\theta^{Z_{n+1}} | Z_n) = f(\theta)^{Z_n}.$$

[If V takes only integer values, then when $V = k$, the conditional expectation $\mathsf{E}(U|V)$ of U given V is equal to the conditional expectation $\mathsf{E}(U|V = k)$ of U given that $V = k$. (Sounds reasonable!)] Property (c) now yields

$$\mathsf{E}\theta^{Z_{n+1}} = \mathsf{E}f(\theta)^{Z_n},$$

and, since

$$\mathsf{E}(\alpha^{Z_n}) = f_n(\alpha), \qquad \square$$

result (a) is proved.

Independence and *conditional expectations* are two of the main topics in this course.

0.4. Extinction probability, π

Let $\pi_n := \mathsf{P}(Z_n = 0)$. Then $\pi_n = f_n(0)$, so that, by (0.3,b),

(a) $$\pi_{n+1} = f(\pi_n).$$

Measure theory confirms our intuition about the extinction probability:

(b) $$\pi := \mathsf{P}(Z_m = 0 \text{ for some } m) = \uparrow \lim \pi_n.$$

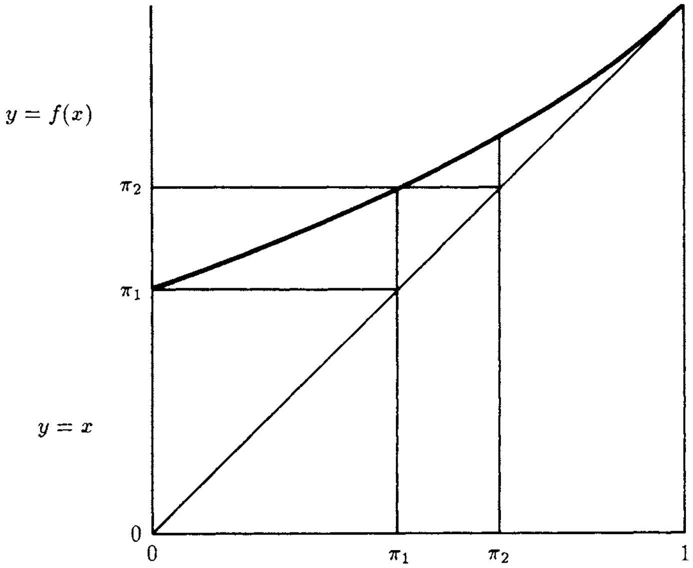
Because f is continuous, it follows from (a) that

(c) $$\pi = f(\pi).$$

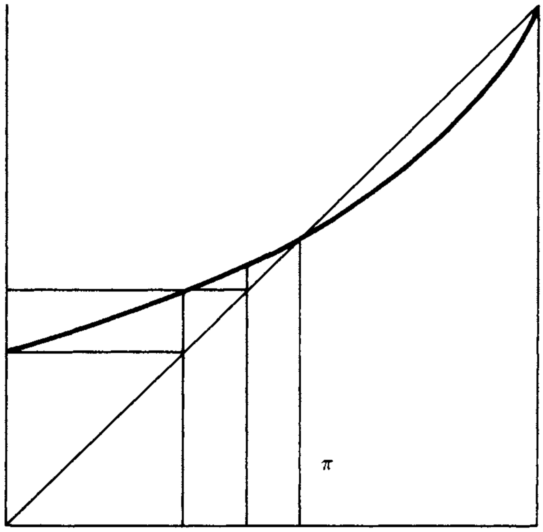
The function f is analytic on (0,1), and is non-decreasing and convex (of non-decreasing slope). Also, $f(1) = 1$ and $f(0) = \mathsf{P}(X = 0) > 0$. The slope $f'(1)$ of f at 1 is $\mu = \mathsf{E}(X)$. The celebrated pictures opposite now make the following Theorem obvious.

THEOREM

> *If* $\mathsf{E}(X) > 1$, *then the extinction probability* π *is the unique root of the equation* $\pi = f(\pi)$ *which lies strictly between 0 and 1. If* $\mathsf{E}(X) \le 1$, *then* $\pi = 1$.



Case 1: subcritical, $\mu = f'(1) < 1$. Clearly, $\pi = 1$.
The *critical case* $\mu = 1$ has a similar picture.



Case 2: supercritical, $\mu = f'(1) > 1$. Now, $\pi < 1$.

0.5. Pause for thought: measure

Now that we have finished revising what introductory courses on probability theory say about branching-process theory, let us think about why we must find a more precise language. To be sure, the claim at (0.4,b) that

(a) $$\pi = \uparrow \lim \pi_n$$

is intuitively plausible, but how could one *prove* it? We certainly cannot prove it at present because we have no means of stating with pure-mathematical precision what it is supposed to mean. Let us discuss this further.

Back in Section 0.2, we said 'Suppose that we are given a doubly infinite sequence $\{X_r^{(m)} : m, r \in \mathbf{N}\}$ of independent identically distributed random variables each with the same distribution as X'. What does this mean? A random variable is a (certain kind of) function on a sample space Ω. We could follow elementary theory in taking Ω to be the set of all outcomes, in other words, taking Ω to be the Cartesian product

$$\Omega = \prod_{r,s} \mathbf{Z}^+,$$

the typical element ω of Ω being

$$\omega = (\omega_s^{(r)} : r \in \mathbf{N}, s \in \mathbf{N}),$$

and then setting $X_s^{(r)}(\omega) = \omega_s^{(r)}$. Now Ω is an uncountable set, so that we are outside the 'combinatorial' context which makes sense of π_n in the elementary theory. Moreover, if one assumes the Axiom of Choice, one can *prove* that it is impossible to assign to *all* subsets of Ω a probability satisfying the 'intuitively obvious' axioms and making the X's IID RVs with the correct common distribution. So, we have to know that the set of ω corresponding to the event 'extinction occurs' *is* one to which one can uniquely assign a probability (which will then provide a definition of π). Even then, we have to prove (a).

Example. Consider for a moment what is in some ways a bad attempt to construct a 'probability theory'. Let $\mathcal{C}$ be the class of subsets C of $\mathbf{N}$ for which the 'density'

$$\rho(C) := \lim_{n\uparrow\infty} \sharp\{k : 1 \le k \le n; k \in C\}$$

exists. Let $C_n := \{1, 2, \ldots, n\}$. Then $C_n \in \mathcal{C}$ and $C_n \uparrow \mathbf{N}$ in the sense that $C_n \subseteq C_{n+1}, \forall n$ and also $\bigcup C_n = \mathbf{N}$. However, $\rho(C_n) = 0, \forall n$, but $\rho(\mathbf{N}) = 1$.

Hence the logic which will allow us correctly to deduce (a) from the fact that

$$\{Z_n = 0\} \uparrow \{\text{extinction occurs}\}$$

fails for the $(\mathbf{N}, \mathcal{C}, \rho)$ set-up: $(\mathbf{N}, \mathcal{C}, \rho)$ is not 'a probability triple'. □

There *are* problems. Measure theory resolves them, but provides a huge bonus in the form of much deeper results such as the Martingale Convergence Theorem which we now take a first look at – at an intuitive level, I hasten to add.

0.6. Our first martingale

Recall from (0.2,b) that

$$Z_{n+1} = X_1^{(n+1)} + \cdots + X_{Z_n}^{(n+1)},$$

where the $X_{\cdot}^{(n+1)}$ variables are independent of the values $Z_1, Z_2, \ldots, Z_n$. It is clear from this that

$$\mathsf{P}(Z_{n+1} = j | Z_0 = i_0, Z_1 = i_1, \ldots, Z_n = i_n) = \mathsf{P}(Z_{n+1} = j | Z_n = i_n),$$

a result which you will probably recognize as stating that the process $Z = (Z_n : n \geq 0)$ *is a Markov chain.* We therefore have

$$\begin{aligned} \mathsf{E}(Z_{n+1} | Z_0 = i_0, Z_1 = i_1, \ldots, Z_n = i_n) &= \sum_j j \mathsf{P}(Z_{n+1} = j | Z_n = i_n) \\ &= \mathsf{E}(Z_{n+1} | Z_n = i_n), \end{aligned}$$

or, in a condensed and better notation,

(a) $$\mathsf{E}(Z_{n+1} | Z_0, Z_1, \ldots, Z_n) = \mathsf{E}(Z_{n+1} | Z_n).$$

Of course, it is intuitively obvious that

(b) $$\mathsf{E}(Z_{n+1} | Z_n) = \mu Z_n,$$

because each of the Z_n animals in the n^{th} generation has on average μ children. We can confirm result (b) by differentiating the result

$$\mathsf{E}(\theta^{Z_{n+1}} | Z_n) = f(\theta)^{Z_n}$$

with respect to θ and setting $\theta = 1$.

Now define

(c) $$M_n := Z_n/\mu^n, \qquad n \geq 0.$$

Then

$$\mathsf{E}(M_{n+1}|Z_0, Z_1, \ldots, Z_n) = M_n,$$

which exactly says that

(d) *M is a martingale relative to the Z process.*

Given the history of Z up to stage n, the next value M_{n+1} of M is on average what it is now: M is 'constant on average' in this very sophisticated sense of conditional expectation given 'past' and 'present'. The true statement

(e) $$\mathsf{E}(M_n) = 1, \quad \forall n$$

is of course infinitely cruder.

A statement $\mathcal{S}$ is said to be true **almost surely** (a.s.) or **with probability** 1 if (surprise, surprise!)

$$\mathsf{P}(\mathcal{S} \text{ is true}) = 1.$$

Because our martingale M is *non-negative* ($M_n \geq 0, \forall n$), the **Martingale Convergence Theorem** implies that *it is almost surely true that*

(f) $$M_\infty := \lim M_n \quad \textit{exists.}$$

Note that if $M_\infty > 0$ for some outcome (which can happen with positive probability only when $\mu > 1$), then the statement

$$Z_n/\mu^n \to M_\infty \qquad \text{(a.s.)}$$

is a precise formulation of 'exponential growth'. A particularly fascinating question is: *suppose that $\mu > 1$; what is the behaviour of Z conditional on the value of M_∞?*

0.7. Convergence (or not) of expectations

We know that $M_\infty := \lim M_n$ exists with probability 1, and that $\mathsf{E}(M_n) = 1$, $\forall n$. We might be tempted to believe that $\mathsf{E}(M_\infty) = 1$. However, we already know that if $\mu \leq 1$, then, almost surely, the process dies out and M_n is eventually 0. Hence

(a) *if $\mu \leq 1$, then $M_\infty = 0$ (a.s.) and*

$$0 = \mathsf{E}(M_\infty) \neq \lim \mathsf{E}(M_n) = 1.$$

This is an excellent example to keep in mind when we come to study *Fatou's Lemma*, valid for any sequence (Y_n) of non-negative random variables:

$$\mathsf{E}(\liminf Y_n) \le \liminf \mathsf{E}(Y_n).$$

What is 'going wrong' at (a) is that (when $\mu \le 1$) for large n, the chances are that M_n will be large if M_n is not 0 and, very roughly speaking, this large value times its small probability will keep $\mathsf{E}(M_n)$ at 1. See the concrete examples in Section 0.9.

Of course, it is very important to know when

(b) $$\lim \mathsf{E}(\cdot) = \mathsf{E}(\lim \cdot),$$

and we do spend quite a considerable time studying this. The best general theorems are rarely good enough to get the best results for concrete problems, as is evidenced by the fact that

(c) $\mathsf{E}(M_\infty) = 1$ *if and only if* both $\mu > 1$ *and* $\mathsf{E}(X \log X) < \infty$,

where X is the typical number of children. Of course $0 \log 0 = 0$. If $\mu > 1$ and $\mathsf{E}(X \log X) = \infty$, then, even though the process may not die out, $M_\infty = 0$, a.s.

0.8. Finding the distribution of M_∞

Since $M_n \to M_\infty$ (a.s.), it is obvious that for $\lambda > 0$,

$$\exp(-\lambda M_n) \to \exp(-\lambda M_\infty) \qquad \text{(a.s.)}$$

Now since each $M_n \ge 0$, the whole sequence $(\exp(-\lambda M_n))$ is bounded in absolute value by the constant 1, independently of the outcome of our experiment. The *Bounded Convergence Theorem* says that we *can* now assert what we would wish:

(a) $$\mathsf{E}\exp(-\lambda M_\infty) = \lim \mathsf{E}\exp(-\lambda M_n).$$

Since $M_n = Z_n/\mu^n$ and $\mathsf{E}(\theta^{Z_n}) = f_n(\theta)$, we have

(b) $$\mathsf{E}\exp(-\lambda M_n) = f_n(\exp(-\lambda/\mu^n)),$$

so that, in principle (if very rarely in practice), we can calculate the left-hand side of (a). However, for a non-negative random variable Y, *the distribution function* $y \mapsto \mathsf{P}(Y \le y)$ *is completely determined by the map*

$$\lambda \mapsto \mathsf{E}\exp(-\lambda Y) \quad \textit{on} \quad (0, \infty).$$

Hence, in principle, we can find the distribution of M_∞.

We have seen that the real problem is to calculate the function

$$L(\lambda) := \mathsf{E}\exp(-\lambda M_\infty).$$

Using (b), the fact that $f_{n+1} = f \circ f_n$, and the continuity of L (another consequence of the Bounded Convergence Theorem), you can immediately establish the functional equation:

(c) $$L(\lambda\mu) = f(L(\lambda)).$$

0.9. Concrete example

This concrete example is just about the only one in which one can calculate everything explicitly, but, in the way of mathematics, it is useful in many contexts.

We take the 'typical number of children' X to have a *geometric distribution*:

(a) $$\mathsf{P}(X = k) = pq^k \quad (k \in \mathbf{Z}^+),$$

where

$$0 < p < 1, \quad q := 1 - p.$$

Then, as you can easily check,

(b) $$f(\theta) = \frac{p}{1 - q\theta}, \quad \mu = \frac{q}{p},$$

and

$$\pi = \begin{cases} p/q & \text{if } q > p, \\ 1 & \text{if } q \le p. \end{cases}$$

To calculate $f \circ f \circ \ldots \circ f$, we use a device familiar from the geometry of the upper half-plane. If

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

is a non-singular 2×2 matrix, define the fractional linear transformation:

(c) $$G(\theta) = \frac{g_{11}\theta + g_{12}}{g_{21}\theta + g_{22}}.$$