DAVID WILLIAMS

Probability with Martingales

CAMBRIDGE MATHEMATICAL TEXTBOOKS

Probability with Martingales

Probability with Martingales

David Williams
Statistical Laboratory, DPMMS
Cambridge University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo,
Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521406055

First published 1991
15th printing 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-40605-5 Paperback

Contents

PART C: CHARACTERISTIC FUNCTIONS

APPENDICES

Preface – please read!

The most important chapter in this book is *Chapter E: Exercises.* I have left the interesting things for *you* to do. You can start *now* on the 'EG' exercises, but see 'More about exercises' later in this Preface.

The book, which is essentially the set of lecture notes for a third-year undergraduate course at Cambridge, is as lively an introduction as I can manage to the rigorous theory of probability. Since much of the book is devoted to martingales, it is bound to become very lively: look at those Exercises on Chapter 10! But, of course, there is that initial plod through the measure-theoretic foundations. It must be said however that measure theory, that most arid of subjects when done for its own sake, becomes amazingly more alive when used in probability, not only because it is then applied, but also because it is immensely enriched.

You cannot avoid measure theory: an *event* in probability is a measurable set, a *random variable* is a measurable function on the sample space, the *expectation* of a random variable is its integral with respect to the probability measure; and so on. To be sure, one can take some central results from measure theory as axiomatic in the main text, giving careful proofs in appendices; and indeed that is exactly what I have done.

Measure theory for its own sake is based on the fundamental addition rule for measures. Probability theory supplements that with the multiplication rule which describes independence; and things are already looking up. But what really enriches and enlivens things is that we deal with lots of σ-algebras, not just the one σ-algebra which is the concern of measure theory.

In planning this book, I decided for every topic what things I considered just a bit too advanced, and, often with sadness, I have ruthlessly omitted them.

For a more thorough training in many of the topics covered here, see Billingsley (1979), Chow and Teicher (1978), Chung (1968), Kingman and

Taylor (1966), Laha and Rohatgi (1979), and Neveu (1965). As regards measure theory, I learnt it from Dunford and Schwartz (1958) and Halmos (1959). After reading this book, you must read the still-magnificent Breiman (1968), and, for an excellent indication of what can be done with discrete martingales, Hall and Heyde (1980).

Of course, intuition is much more important than knowledge of measure theory, and you should take every opportunity to sharpen your intuition. There is no better whetstone for this than Aldous (1989), though it is a very demanding book. For appreciating the scope of probability and for learning how to think about it, Karlin and Taylor (1981), Grimmett and Stirzaker (1982), Hall (1988), and Grimmett's recent superb book, Grimmett (1989), on percolation are strongly recommended.

More about exercises. In compiling Chapter E, which consists exactly of the homework sheet I give to the Cambridge students, I have taken into account the fact that this book, like any other mathematics book, implicitly contains a vast number of other exercises, many of which are easier than those in Chapter E. I refer of course to the exercises *you* create by reading the statement of a result, and then trying to prove it for yourself, before you read the given proof. One other point about exercises: you will, for example, surely forgive my using expectation E in Exercises on Chapter 4 before E is treated with full rigour in Chapter 6.

Acknowledgements. My first thanks must go to the students who have endured the course on which the book is based and whose quality has made me try hard to make it worthy of them; and to those, especially David Kendall, who had developed the course before it became my privilege to teach it. My thanks to David Tranah and other staff of CUP for their help in converting the course into this book. Next, I must thank Ben Garling, James Norris and Chris Rogers without whom the book would have contained more errors and obscurities. (The many faults which surely remain in it are my responsibility.) Helen Rutherford and I typed part of the book, but the vast majority of it was typed by Sarah Shea-Simonds in a virtuoso performance worthy of Horowitz. My thanks to Helen and, most especially, to Sarah. Special thanks to my wife, Sheila, too, for all her help.

But my best thanks – and yours if you derive any benefit from the book – must go to three people whose names appear in capitals in the Index: J.L. Doob, A.N. Kolmogorov and P. Lévy: without them, there wouldn't have been much to write about, as Doob (1953) splendidly confirms.

Statistical Laboratory,
Cambridge

David Williams
October 1990

A Question of Terminology

Random variables: functions or equivalence classes?

At the level of this book, the theory would be more 'elegant' if we regarded a random variable as an *equivalence class* of measurable functions on the sample space, two functions belonging to the same equivalence class if and only if they are equal almost everywhere. Then the conditional-expectation map

$$X \mapsto \mathsf{E}(X|\mathcal{G})$$

would be a truly well-defined contraction map from $L^p(\Omega, \mathcal{F}, \mathbf{P})$ to $L^p(\Omega, \mathcal{G}, \mathbf{P})$ for $p \geq 1$; and we would not have to keep mentioning versions (representatives of equivalence classes) and would be able to avoid the endless 'almost surely' qualifications.

I have however chosen the 'inelegant' route: firstly, I prefer to work with *functions*, and confess to preferring

$$4 + 5 = 2 \bmod 7 \qquad \text{to} \qquad [4]_7 + [5]_7 = [2]_7.$$

But there is a substantive reason. I hope that this book will tempt you to progress to the much more interesting, and more important, theory where the parameter set of our process is uncountable (e.g. it may be the time-parameter set $[0, \infty)$). There, the equivalence-class formulation just will not work: the 'cleverness' of introducing quotient spaces loses the subtlety which is essential even for formulating the fundamental results on existence of continuous modifications, etc., unless one performs contortions which are hardly elegant. Even if these contortions allow one to *formulate* results, one would still have to use genuine functions to *prove* them; so where does the reality lie?!

A Guide to Notation

► signifies something important, ►► something very important, and ►►► the Martingale Convergence Theorem.

I use ':=' to signify 'is defined to equal'. This Pascal notation is particularly convenient because it can also be used in the reversed sense.

I use analysts' (as opposed to category theorists') conventions:

► $$\mathbf{N} := \{1, 2, 3, \ldots\} \subseteq \{0, 1, 2, \ldots\} =: \mathbf{Z}^+.$$

Everyone is agreed that $\mathbf{R}^+ := [0, \infty)$.

For a set B contained in some universal set S, I_B denotes the indicator function of B: that is $I_B : S \to \{0, 1\}$ and

$$I_B(s) := \begin{cases} 1 & \text{if } s \in B, \\ 0 & \text{otherwise.} \end{cases}$$

For $a, b \in \mathbf{R}$,

$$a \wedge b := \min(a, b), \qquad a \vee b := \max(a, b).$$

CF:characteristic function; DF: distribution function; pdf: probability density function.

σ-algebra, $\sigma(\mathcal{C})$ (1.1); $\sigma(Y_\gamma : \gamma \in \mathcal{C})$ (3.8, 3.13). π-system (1.6); d-system (A1.2).

---

a.e.:	almost everywhere (1.5)
a.s.:	almost surely (2.4)
bΣ:	the space of bounded Σ-measurable functions (3.1)

$\mathcal{B}(S)$:	the Borel σ-algebra on S, $\mathcal{B} := \mathcal{B}(\mathbf{R})$ (1.2)
$C \bullet X$:	discrete stochastic integral (10.6)
$d\lambda/d\mu$:	Radon-Nikodým derivative (5.14)
$d\mathbf{Q}/d\mathbf{P}$:	Likelihood Ratio (14.13)
$\mathsf{E}(X)$:	expectation $\mathsf{E}(X) := \int_\Omega X(\omega)\mathbf{P}(d\omega)$ of X (6.3)
$\mathsf{E}(X;F)$:	$\int_F X d\mathbf{P}$ (6.3)
$\mathsf{E}(X\|\mathcal{G})$:	conditional expectation (9.3)
(E_n, ev):	$\liminf E_n$ (2.8)
$(E_n, \text{i.o.})$:	$\limsup E_n$ (2.6)
f_X:	probability density function (pdf) of X (6.12).
$f_{X,Y}$:	joint pdf (8.3)
$f_{X\|Y}$:	conditional pdf (9.6)
F_X:	distribution function of X (3.9)
lim inf:	for sets, (2.8)
lim sup:	for sets, (2.6)
$x = \uparrow \lim x_n$:	$x_n \uparrow x$ in that $x_n \leq x_{n+1}$ $(\forall n)$ and $x_n \to x$.
log:	natural (base e) logarithm
$\mathcal{L}_X$, Λ_X:	law of X (3.9)
$\mathcal{L}^p$, L^p:	Lebesgue spaces (6.7, 6.13)
Leb:	Lebesgue measure (1.8)
$\mathrm{m}\Sigma$:	space of Σ-measurable functions (3.1)
M^T:	process M stopped at time T (10.9)
$\langle M \rangle$:	angle-brackets process (12.12)
$\mu(f)$:	integral of f with respect to μ (5.0, 5.2)
$\mu(f;A)$:	$\int_A f d\mu$ (5.0, 5.2)
φ_X:	CF of X (Chapter 16)
φ:	pdf of standard normal N(0,1) distribution
Φ:	DF of N(0,1) distribution
X^T:	X stopped at time T (10.9)

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*

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

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

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

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

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

and

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

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:

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

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 $r \leq n$, 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

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

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

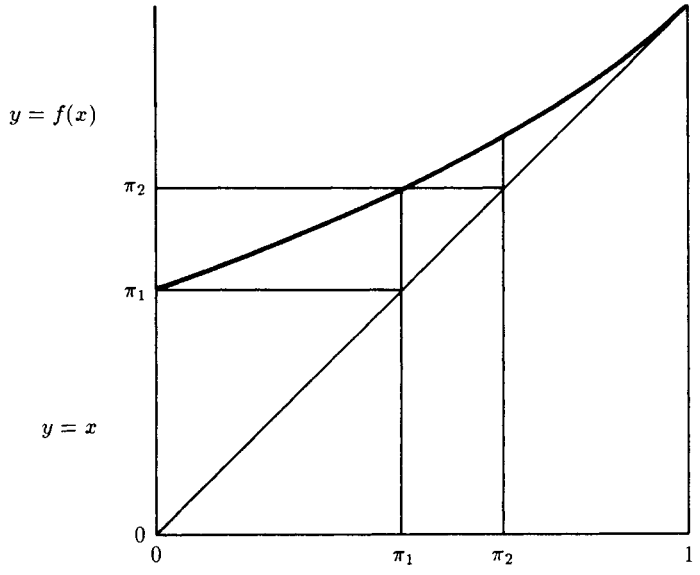
Because f is continuous, it follows from (a) that

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

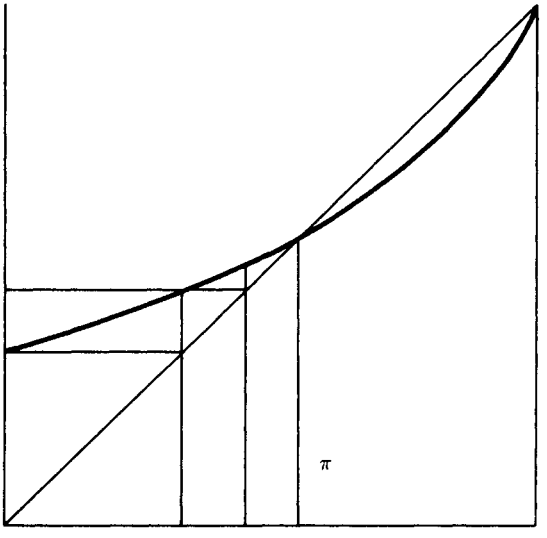
The function f is analytic on (0,1), and is non-decreasing and convex (of non-decreasing slope). Also, $f(1) = 1$ and $f(0) = \mathsf{P}(X = 0) > 0$. The slope $f'(1)$ of f at 1 is $\mu = \mathsf{E}(X)$. The celebrated pictures opposite now make the following Theorem obvious.

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 $\mathsf{E}(X) \leq 1$, 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'

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

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,

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

Of course, it is intuitively obvious that

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

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:

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

What is 'going wrong' at (a) is that (when $\mu \leq 1$) 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) $$\mathsf{E}(M_\infty) = 1 \textit{ if and only if } \text{both } \mu > 1 \textit{ and } \mathsf{E}(X \log X) < \infty,$$

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 $M_n \geq 0$, 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* $y \mapsto \mathsf{P}(Y \leq y)$ *is completely determined by the map*

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