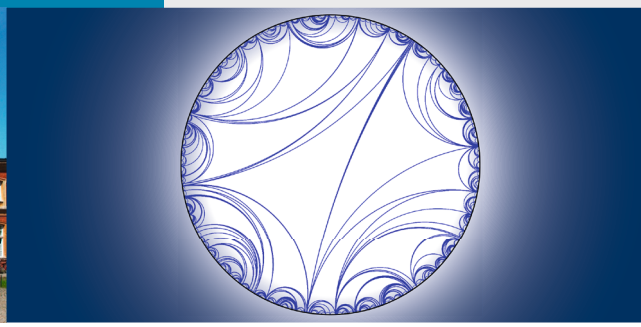
Universität Potsdam

Potsdam University Press



Michael Högele | Sylvie Rœlly | Mathias Rafler (Eds.)

Stochastic processes with applications in the natural sciences

International workshop at Universidad de los Andes, Bogotá, Colombia | 5-9 December 2016

Lectures in Pure and Applied Mathematics | 4

Lectures in Pure and Applied Mathematics

Lectures in Pure and Applied Mathematics | 4

Michael Högele | Sylvie Rœlly | Mathias Rafler (Eds.)

Stochastic processes with applications in the natural sciences

International workshop at Universidad de los Andes,
Bogotá, Colombia | 5-9 December 2016

Potsdam University Press

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Potsdam University Press 2017
http://verlag.ub.uni-potsdam.de/

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: 2292
E-Mail: verlag@uni-potsdam.de

The series **Lectures in Pure and Applied Mathematics** is edited by the Institute of Mathematics of the University of Potsdam.

ISSN (print) 2199-4951
ISSN (online) 2199-496X

Contact:
Institut für Mathematik
Karl-Liebknecht-Str. 24-25
14476 Potsdam OT Golm
Tel.: +49 (0)331 977 1028
WWW: http://www.math.uni-potsdam.de

Cover images:
1. Karla Fritze | Building of the Institute at Campus Neues Palais
2. Nicolas Curien, Wendelin Werner | Random hyperbolic triangulation
Published at: Journal of the European Mathematical Society, Volume 15, Issue 4, 2013, pp. 1309–1341, DOI: 10.4171/JEMS/393

Published online at the Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-401802
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-401802

Simultaneously published in print:
ISBN 978-3-86956-414-2

Preface

Modern natural sciences and even more so biophysics are unthinkable without the understanding of the fundamental nature of the randomness in the time evolution of quantities of interest. The first international and interdisciplinary workshop STOCHASTIC PROCESSES WITH APPLICATIONS IN THE NATURAL SCIENCES held at Universidad de los Andes from December 5th to 9th in 2016 offered four lecture courses to cover a section of this enormous and prolific field. Two of these lectures were centered in biophysics and two of them in the mathematical theory of stochastic processes, ranging from probabilistic modeling of the degradation of cell molecules, biological functional circuits via an introductory mathematics lecture on branching processes and a more advanced mathematics course on the dynamics of Markov processes. The courses were aimed at Master's and Ph.D. students as well as ambitious undergraduates.

The courses in physical biology were focused on the nature and modeling of random effects of particular classes of biological systems. Both lecturers are distinguished researchers in the field, Dr. Angelo Valleriani from Germany, and Prof. Juan Manuel Pedraza from the Physics department of Universidad de los Andes, Colombia. These notes include the chapter **Stochastic modeling of complex macromolecular decay** by Dr. Valleriani, group leader of the team *Stochastic processes in complex and biological systems* from the Max-Planck-Institute of Colloids and Interfaces, Potsdam. They are based on the classes he taught in Bogotá and yield beyond the specific modeling problem a general perspective on the modeling of gene expression.

The first mathematics course offered a profound introduction to a class of stochastic processes which are particularly important in system biology, so-called branching processes. It was held by Prof. Sylvie Rœlly, Vice-Dean of the Faculty of Science at the University of Potsdam and winner of the Itô prize in 2007 with wide expertise in stochastic processes. These lecture notes contain the chapter **Algunas propriedades básicas de procesos de ramificación** (in Spanish) as an introduction to the field of branching

processes. It is well-suited for the self-study in order to penetrate the literature of this wide topic.

The second mathematics course treated the time asymptotic behavior of a large class of Markov chains and, much more generally, of Markov processes. It was delivered by Prof. Alexei Kulik from the Ukrainian National Academy of Sciences, Kiev, Ukraine, an international expert in Markov processes. The chapter **Markov Models: Stabilization and Limit Theorems** of these lecture notes provides an overview in the mathematical theory of ergodic properties of Markov processes, culminating in the respective central limit theorem.

At this point it is our pleasure to thank all participants and all who made this workshop possible. The University of Potsdam showed a major commitment through a KoUP1-Project, which made it possible that a group of four Master's student, Ph.D. student and Postdocs could cross the Atlantic and visit Universidad de los Andes (and Colombia) for the first time. Universidad de los Andes, the Department of Mathematics and the School of Sciences, covered among others the stay of scientists and provided all the facilities and the excellent staff support during the organization, including the support of the website and the posters. It is gratefully acknowledged that the Berlin Mathematical School and Max Planck Institute of Colloids and Interfaces covered the visit of Alexei Kulik and Angelo Valleriani.

Many thanks go to Dr. Mathias Rafler who supervised the edition of this book by pouring the raw texts in its beautiful LaTeX shape.

Potsdam, September 2017 — Sylvie Rœlly
(scientific committee)

Bogotá, September 2017 — Michael Högele
(scientific committee, local organizer)

Contents

Chapter 1

Stochastic modeling of complex macromolecular decay

*Angelo Valleriani**

1 A short model of gene expression

There are many books and articles summarizing what we know about gene expression. Furthermore, the more we study this aspect of life, the more we find out how complex it is and how variable such concepts like gene, genetic code and protein are. Nevertheless, since this article refers to modeling one aspect of the regulation of gene expression, it is perhaps important to provide the reader with at least a very superficial but quick view of the big framework.

Gene expression is a generic term that is commonly related to what is known as the central dogma of molecular biology. According to the commonly accepted version of this dogma, it all starts with the DNA, which is a chain of four nucleotides A, C, G, T, organized in chromosomes. On the chromosomes, the genes occupy specific regions called loci. In humans, there are 23 different chromosomes, each coming in two copies, one inherited from the mother and one from the father, respectively. The sequence of nucleotides of one gene on a certain locus could be different on the two chromosomes

*Max Planck Institute of Colloids and Interfaces, Department of Theory and Bio-Systems; *angelo.valleriani@mpikg.mpg.de*

but in general the number of differences is very small or absent. Population genetics, also a branch of applied stochastic methods, deals with some mechanisms of propagation of gene variants in populations [1, 2]. Certain cells have only one set of chromosomes, e.g. sperm or egg cells. Certain organisms like bacteria have only one single chromosome on which all genes are collected.

The genes found in the DNA are first transcribed into RNA molecules. The majority of the total RNA molecules present in each cell plays a key role in the production of proteins. Some species of RNA molecules become part of ribosomes, which are complex molecular machines made of RNA and proteins. Some RNAs become transport RNA, called also tRNA. Some RNAs are found in the form of small or micro RNA and finally a prominent role in gene expression is played by messenger RNA molecules (mRNA).

Protein synthesis is the final product of gene expression: in this process, the ribosomes read the information encoded in the mRNA and synthesize the proteins using the amino-acid delivered by the tRNAs. This process is called translation. The particular way in which the ribosome reads the mRNA is the basis of what we know as the genetic code. The ribosome reads the nucleotide sequence of the mRNA one triplet of amino-acids per step. To each triplet, called codon, corresponds one amino-acid that will be incorporated into the nascent protein.

The amount of proteins corresponding to a given gene present in the cell will thus depend on several factors. The first factor is the amount of mRNAs of that gene: this is determined by the balance between the synthesis rate of the mRNA (transcription rate) and the degradation rate of the mRNA. The second factor is the amount of ribosomes translating each mRNA molecules, which eventually determines the protein synthesis rate (if we neglect possible translation errors such as ribosome drop-off [3]). The final factor is the degradation rate of the proteins.

The RNA molecules that become part of the ribosome are called rRNA. Both rRNA and tRNA are very stable. Their duty is to provide the machinery of the process of translation, independently of what has to be translated. The mRNA molecules instead are typically not so stable and their lifetime is regulated by some internal cellular mechanisms. Indeed, when the cell needs to change the kind of proteins to be synthesized, due for instance to some stressful condition, it can do that by changing the composition of the cell mRNA population. Beside the important role played by the regulation of transcription, one way to tune the amount of mRNA is to activate or deactivate specific degradation

mechanisms. These degradation mechanisms usually involve several biochemical pathways, which work in parallel or in competition with each other. As an example, micro RNA (miRNA) can bind to a specific mRNA together with some proteins, whose main scope is to recruit other enzymes that start or trigger the digestion of the mRNA [4, 5]. The digestion of the mRNA proceeds usually through several steps until the molecule is not detectable in a fashion that depends on the kind of mRNA and on the organism [6]. This series of steps can be quite complex, since it requires the subsequent participation of several molecules. Since the degradation of mRNA runs over several steps, the lifetime of such molecules is very complex and so we should expect that their decay pattern is not exponential [7, 8, 9, 10]. The degradation of proteins is also a mechanism to regulate gene expression. Also for proteins we know that several factors are needed to degrade them and so we expect their lifetime to be complex and their decay pattern to be non-exponential [11, 12, 13].

Without entering into any of the biochemical mechanisms that determine the degradation of mRNA or proteins, we may still briefly discuss how the decay pattern of such molecules are determined experimentally. There are obviously several experimental techniques. All of them have in common the idea of following the decay in time of the amount of molecules of the species of interest. To achieve this, the experimentalists first cultivate the cells for a certain fixed amount of time during which the molecules of interest are marked, e.g. with radioactive isotopes. This period of labeling with the marker is called the *pulse phase*. Afterwards, the labeling of the molecules is stopped and the experiment proceeds with the harvesting of the molecules at fixed time intervals from the cell extracts of a large amount of cells. This period of harvesting is called the *chase phase*. The measurements at these time points will provide a decaying curve that we call the *decay pattern* of the specific molecule. The curve decays because the amount of labeled molecules cannot increase. In some cases, the pulse phase can be so long that all molecules of interest are labeled. We call this situation the *steady state expression*. In this situation a balanced amount of old and newly synthesized molecules is present in the cell, i.e. we have a stationary age distribution of those molecules. When instead the pulse is extremely short, the labeled molecules are all very young and the age distribution is far away from stationarity. As we shall see in the following, the length of the pulse will make a big difference on how the decay pattern looks like [11, 13]. This is the starting point of our journey towards modeling the process of complex decay.

1.1 Types of decay patterns

Traditionally, decay experiments have been performed under the assumption of an exponential decay. Under this assumption, the decay pattern in a semilog plot looks like a straight line. The fit requires just two point and the liner regression of the data, from which the rate of degradation can be extracted.

As more experimental techniques became available, more datapoints during the chase phase were collected thus revealing that the behavior is often very different from a straight line in a semilog plot (see [9] and references therein). In most cases, the decay pattern could be superficially described as fast at short times and slow at long times thus leading to the idea of a double exponential. Most careful experiments and theory have however shown that the decay pattern changes in shape depending on the length of the pulse phase even if the molecules under examination are the same [11, 13]. From the biochemical point of view, a non-exponential decay pattern means only that the degradation process is complex, i.e. it is composed of several steps [7, 9].

• •

2 Relationship between Poisson processes and uniform distribution

This section states and summarizes a known result concerning the relationship between a Poisson process with constant rate and the uniform distribution [14]. We will show that conditional on a number N of events at a certain time t, the time of origin of the N events is uniformly distributed in $[0,t)$.

Let $X(t)$ be a Poisson process with rate λ and initial condition $X(0)=0$. If $X(t)=n$ we will say that until time t there have been exactly n events. Textbook calculations show that for any fixed t, the number of events n is distributed according to a Poisson distribution

$$\mathsf{P}\{X(t)=n \mid X(0)=0\} = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}, \tag{1.1}$$

which implies also that for $u \leq t$

$$\mathsf{P}\{X(t)=n \mid X(u)=k\} = \frac{[\lambda(t-u)]^{n-k} \exp\big(-\lambda(t-u)\big)}{(n-k)!}, \tag{1.2}$$

holds.

Now we would like to show that under the condition of a fixed number n of events until time t, the time points at which the n events arose are uniformly distributed in $[0,t)$. To proceed with this comparison, we shall first state one of the elementary properties of the uniform distribution. Imagine to throw n points at random on a line of length t such that each throw is independent and the probability density function of the position of each point is uniform. Let now be u an arbitrary point in $[0,t)$ and let Y_u be the number of points, out of the n points thrown in total over the line, that fall into the segment $[0,u)$. Since each point has probability u/t to fall over this segment, the probability distribution for Y_u is given by the binomial distribution

$$\mathsf{P}\{Y_u = k\} = \binom{n}{k}\left(\frac{u}{t}\right)^k\left(1-\frac{u}{t}\right)^{n-k}. \tag{1.3}$$

Let us turn our attention back to the Poisson process and let us consider the ensemble of Poisson processes with rate λ conditioned that $X(t) = n$, namely conditioned that the number of events up to time t is n. Under this condition, we would like to know what is the distribution of the number of events of the process occurring during the interval of time $[0,u)$ for $u \in [0,t)$. This is equivalent to ask about the probability distribution of the stochastic variable $X(u)$. This leads to

$$\begin{aligned}
&\mathsf{P}\{X(u) = k \mid X(0) = 0, X(t) = n\} \\
&\qquad = \frac{\mathsf{P}\{X(u) = k, X(t) = n \mid X(0) = 0\}}{\mathsf{P}\{X(t) = n \mid X(0) = 0\}} \\
&\qquad = \frac{\mathsf{P}\{X(t) = n \mid X(u) = k\}\mathsf{P}\{X(u) = k \mid X(0) = 0\}}{\mathsf{P}\{X(t) = n \mid X(0) = 0\}} \\
&\qquad = \frac{n!}{k!(n-k)!}\left(\frac{u}{t}\right)^k\left(1-\frac{u}{t}\right)^{n-k},
\end{aligned} \tag{1.4}$$

where the last line arises after using (1.1) and (1.2). The comparison between (1.3) and (1.4) should be enough to convince ourselves that the distribution of the number of events falling into any interval of time in the conditioned Poisson process is identical to the distribution of the number of points falling into the same interval. We conclude that the position in time of the events of the conditioned Poisson process follows a uniform distribution.

3 The time dependent distribution of the number of molecules

In this section we will assume that the pulse phase starts at time zero and that the molecules under consideration have an arbitrary life time distribution. We will first derive the distribution of the number of molecules at time t after the start of the pulse. Later we will assume that the pulse phase is interrupted at time t_p and derive the time dependent distribution of the number of molecules at time $t_p + \Delta t$ during the chase phase. Finally, we will derive the time dependent age and rest-life distributions of the molecules as a function of time. To simplify the discussion, in the rest of these notes we will concentrate on the analysis of mRNA decay as an example of application. The extension of the analysis and modeling to protein decay is then easy.

3.1 Basic assumptions about transcription and degradation

As specified earlier, mRNA molecules are synthesized from their DNA template through a process called transcription. For the whole content of this section we will assume that transcription occurs at a fixed constant rate ω_{tc} and that at each transcription event only one mRNA is synthesized. To be precise, in eukaryotic cells the process of transcription does not directly deliver mRNA molecules. In this kind of cells, our ω_{tc} is then the rate of mRNA delivery in the cytoplasm of the cell. Let $R(t)$ be the stochastic variable that gives the number of labeled mRNA found at time t after the start of the pulse phase. We assume that the pulse phase starts at time zero and that thus the initial condition of our process is given by $R(0) = 0$. Thus, in the absence of mRNA degradation, the process of transcription is a Poisson process with constant rate ω_{tc}. Another assumption that we make concerns the life time distribution of each mRNA molecule. Let $U \in [0, \infty)$ be the random life time of an mRNA molecule. We will assume that the probability function $G(t_u)$ of U is given by

$$G(t_u) \equiv \mathsf{P}\{U \leq t_u\} = \int_0^{t_u} \phi_U(\tau) \mathrm{d}\tau, \tag{1.5}$$

where ϕ_U is the probability density function of U. We will not make any special restriction concerning the form of the density ϕ_U apart from the fact of being integrable in $[0, \infty)$ and having well defined moments. In particular, we will not impose that ϕ_U is an exponential

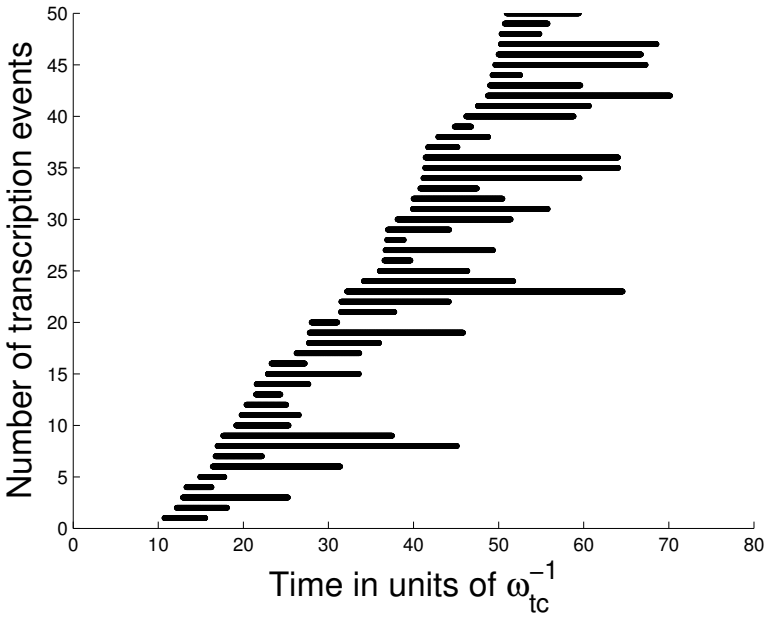


Figure 1.1: This figure shows the process of origination of the mRNA chains at random times starting from an arbitrary point in time, $t = 10$. The units of time have been arbitrarily fixed to be equal to ω_{tc}^{-1}. In this illustration, we have set $\langle U \rangle = 10\omega_{tc}^{-1}$. We see that the mRNA chains are generated at random points and have a random duration. In this example, just to fix the ideas, we have considered the gamma distribution $\phi_U(t) = (\lambda^2 t/2)\exp(-\lambda t)$. At any arbitrary point in time, one can count the number of chains that are still intact at that point. By following this number in time, one obtains the trajectory of the number of mRNA molecules originating from one gene in one cell. This statistics is shown in Figure 1.2.

distribution. As mentioned earlier, the exponential distribution has been considered so far as a suitable approximation but the available experimental data show that for most mRNAs there is no simple exponential fall off and we should consider more complex processes. Therefore, in order to leave the modeling open to any possible functional form of ϕ_U, we will henceforth consider the generic form given in (1.5). On the basis of these assumptions we have therefore that transcription at a fixed rate generates new mRNAs while degradation eliminates them. During the pulse phase, these new mRNAs are labeled. We should expect that after a certain amount of time, typically longer than

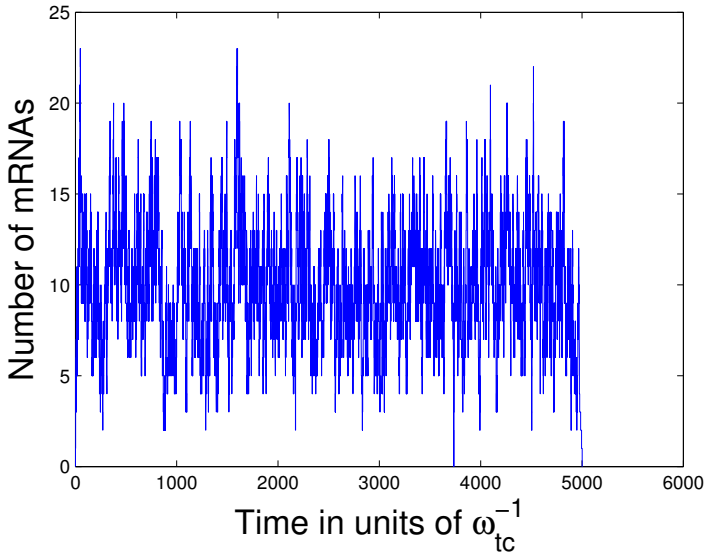


Figure 1.2: Number of mRNA molecules over time. From Figure 1.1 one can derive the number of mRNA chains present in the cell at any given point in time. By following this process over a long period of time, one can obtain a pictorial view of the kind of fluctuations that characterize the number of mRNA molecules. We see in this plot that the number of molecules attains a stationary value until, at the end of the process, the pulse phase is interrupted. Despite the stationarity of the process during a long period of time, the figure shows that the fluctuations are particularly strong. In this work we will provide a detailed analytical and exact description of such fluctuations.

the pulse duration in any common experimental set-up, the processes of synthesis and degradation reach a steady state and that the number of labeled mRNA R attains a stable distribution.

3.2 Formal derivation of the distribution of the amount of mRNA

Consider an experimental set-up in which the pulse duration is t_p. For a time $t \leq t_p$, let $X(t)$ be the underlying Poisson process that describes the amount of labeled mRNA molecules synthesized until time t with transcription rate ω_{tc}. The stochastic variable $R(t)$ is instead the random number of labeled mRNA molecules present in the cell at time

t. Therefore, while $X(t)$ just counts the number of synthesized molecules, $R(t)$ takes into account also the possibility that some molecules synthesized until time t might have been already degraded. Thus, it holds that $0 \leq R(t) \leq X(t)$ for all $t \geq 0$. Under the general assumptions made before, we wish to compute the distribution of $R(t)$, which we can formally write as

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\}, \tag{1.6}$$

for any $k \geq 0$ and $t < t_p$. To approach this problem, we first write this probability by using the law of total probability as [7]

$$\begin{aligned}\mathsf{P}\{R(t) = k \mid X(0) = 0\} \\ = \sum_{n=k}^{\infty} \mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\} \mathsf{P}\{X(t) = n \mid X(0) = 0\}, \end{aligned} \tag{1.7}$$

and consider first $\mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\}$.

We have learned in Section 2 that conditioned on the number of events up to time t, the time position of the events is uniformly distributed in $[0,t)$. Let now be O the origination time of a randomly chosen mRNA and let U be its random lifetime. This mRNA molecule will be present at time t only if the variable $Z = O + U$ satisfies $Z \geq t$. The probability p of this event gives the probability per mRNA to be present at time t. We have therefore

$$p = \mathsf{P}\{Z \geq t\} = \frac{1}{t}\int_0^t \mathsf{P}\{Z \geq t \mid O = s\}\,\mathrm{d}s = \frac{1}{t}\int_0^t (1 - G(u))\,\mathrm{d}u, \tag{1.8}$$

where $G(t_u)$ is defined in (1.5) and we have made use of the fact that $\mathsf{P}\{Z \geq t \mid O = t_o\} = \mathsf{P}\{U \geq t - t_o\}$. Finally, it results that conditioned on $X(t) = n$ the number of labeled mRNA still present at time t is binomially distributed according to

$$\mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\} = \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.9}$$

At this point, we are ready to consider again (1.7) by plugging in both (1.1) and (1.9). This simple computation delivers the time dependent distribution of the number of labeled mRNA at time t after the start of pulse phase, given by

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} A(t)]^k \exp[-\omega_{\mathrm{tc}} A(t)]}{k!}, \tag{1.10}$$

where

$$A(t) = \int_0^t \big(1 - G(u)\big)\,\mathrm{d}u\,.$$

One can notice that (1.10) is a time dependent Poisson distribution with parameter $\omega_{\mathrm{tc}}A(t)$. Nevertheless, one can easily see that $A(t) \to \langle U \rangle$ as $t \to \infty$ with

$$\langle U \rangle = \int_0^\infty u\phi_U(u)\,\mathrm{d}u\,,$$

being the average life time of the mRNA molecules. Thus, for very long pulse durations ($t_p \to \infty$) the limit $t \to \infty$ leads to the stationary distribution

$$\mathsf{P}\{R^{\mathrm{st}} = k\} = \frac{[\omega_{\mathrm{tc}}\langle U \rangle]^k \exp\left[-\omega_{\mathrm{tc}}\langle U \rangle\right]}{k!}\,, \tag{1.11}$$

which depends only on the average life time and not anymore on the details of the degradation process. For the sake of illustration, a comparison between the theoretical distribution and its numerical estimation for a particular choice of $\langle U \rangle$ is shown in Figure 1.3.

Thus we see that the distribution of the amount of mRNA molecules after the start of the pulse phase and before time t_p is Poisson at each time point with a parameter that depends on time. In particular, the *transient time to stationarity* depends critically on the life time distribution of the mRNA [7]. Moreover, the transient time is large if the life distribution is very broad. This must be particularly long in eukaryotic cells, if we interpret ω_{tc} as the rate of delivery of the mRNA. We know indeed that in eukaryotic cells the average life time of the mRNA can be particularly long. In prokaryotic cells instead, the life time of most mRNA molecules is relatively short so that in most of the cases the stationary state is reached in about 20 minutes. The limiting (stationary) distribution given in (1.11) depends instead only on the mean life time of the mRNA chains and not on any other details of the degradation process as long as this process does not modify the average lifetime $\langle U \rangle$.

3.2.1 Time dependent life time distributions

This subsection can be omitted on a first reading. In this section we derive the distribution of the number of mRNA molecules for those cases when the time of origin of the mRNA is relevant. This is usually the case if the experimental conditions change with time.

It is of a certain interest at this point to have a brief look at the special case in which

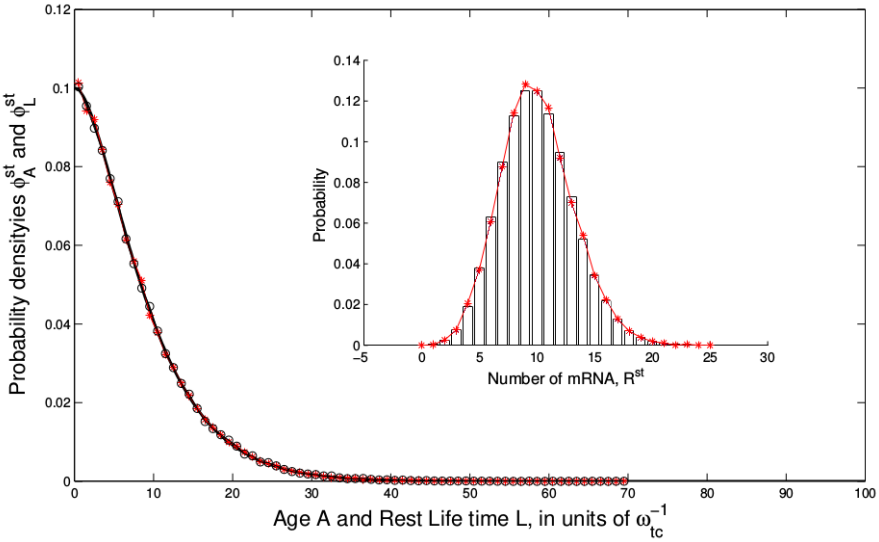


Figure 1.3: Age and rest life time distributions at steady state. Inset: stationary distribution of the number of mRNA. Under the same conditions of Figures 1.1 and 1.2 the main plot shows a comparison between the numerical estimation at steady state of the age of the mRNA taken at any random point in time (circles), the rest life time distribution (stars) and the theoretical predictions given in (1.28) and (1.29), respectively. In the inset, we see the comparison between the statistics from a computer simulation for the number of mRNA at stationary state with $\langle U \rangle = 10\,\omega_{tc}^{-1}$ with the corresponding theoretical distribution derived in (1.11).

the life time distribution of the mRNA depends on time t after the initiation of the experiment. This dependence may be due to changing environmental conditions during the experiments both because of intrinsic and extrinsic origins. Putting it in simple terms, this means the the time of origin of an mRNA cannot be forgotten because this determines the history of that chain.

Formally, let us define the random variable $C(\tau,t)$ that gives the state of a mRNA chain of age τ at time t. This stochastic variable takes values in the state space $\{0,1\}$ such that 0 is the degraded (absorbing) state and 1 is the intact state. We assume to have

$$\mathrm{P}\{C(\tau+h,t+h) = 0 \,|\, C(\tau,t) = 1\} = \omega(\tau,t)h + o(h)\,, \tag{1.12}$$

for $h \to 0$. In this case, it is an easy calculation to find out that

$$\mathsf{P}\{U > t_u \mid O = t_o\} = \exp\left(-\int_0^{t_u} \omega(x, t_o + x)\,\mathrm{d}x\right), \tag{1.13}$$

where t_o is the time at which the mRNA chain was originated. Finally, this leads to the probability function

$$\mathsf{P}\{U \leq t_u \mid O = t_o\} \equiv G(t_u, t_o) = 1 - \exp\left(-\int_0^{t_u} \omega(x, t_o + x)\,\mathrm{d}x\right), \tag{1.14}$$

which depends both on t_o and on t_u and not only on their difference. Therefore, to evaluate p we cannot follow the same route that led to (1.8). In this case, we need to consider that

$$\mathsf{P}\{Z \geq t \mid O = t_o\} = \mathsf{P}\{U \geq t - t_o \mid O = t_o\} = 1 - G(t - t_o, t_o), \tag{1.15}$$

and thus the probability p becomes

$$p = \frac{1}{t}\int_0^t \big(1 - G(t - s, s)\big)\,\mathrm{d}s = \frac{H(t)}{t}, \tag{1.16}$$

with the additional property that the limit of large t for $H(t)$ may not be known or even be unbounded. At this point we impose that

$$\lim_{x \to \infty} G(x, y) = 1 \tag{1.17}$$

for any value of y. Note that Eq. (1.17) is not contained in the definition of the degradation rate given in (1.12). By using the same formalism set in (1.12) we can then rewrite $H(t)$ as

$$H(t) = \int_0^t \exp\left(-\int_0^{t-s} \omega(x, s + x)\,\mathrm{d}x\right)\mathrm{d}s = \int_0^t \exp\left(-\int_s^t \omega(x - s, x)\,\mathrm{d}x\right)\mathrm{d}s, \tag{1.18}$$

which should converge for $t \to \infty$ if $\omega(x, y) \geq \omega_{\min} > 0$ for any $x, y \geq 0$. Finally, the time dependent distribution of the number of mRNA at time t after the start of transcription is given by

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} H(t)]^k \exp(-\omega_{\mathrm{tc}} H(t))}{k!}. \tag{1.19}$$

3.3 The number of mRNA molecules after the pulse

In this section we consider the case in which the pulse is stopped (or interrupted) at some time t_p and we observe the number of mRNA molecules an interval Δt during the chase phase.

Let us assume that pulse phase is terminated at time t_p after having started at time zero. We have seen that the distribution of the number of labeled mRNAs for the transcribed gene is given by (1.10) for $t = t_p$. Since labeling is terminated, we expect that no further labeled mRNAs are produced and that, therefore, there is no increase in the number of labeled mRNAs after time t_p and, rather, we expect to witness a decrease of those mRNAs due to the effect of the degradation mechanisms.

We would like to find out what is the distribution of the number of mRNAs at time $t = t_p + \Delta t$, namely Δt time units after the interruption of labeled synthesis [9]. By using the same notation as in the previous calculations, we shall find

$$\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0,\, \omega_{\text{tc}} > 0 \text{ for } t \in [0,t_p),\, \omega_{\text{tc}} = 0 \text{ for } t \in [t_p,\infty)\}\,, \quad (1.20)$$

where, in the condition, we have explicitly written the change in the synthesis (labeling) rate ω_{tc} underlying the interruption of labeling. To simplify the notation, we will henceforth omit the condition on the synthesis rate ω_{tc} of labeled molecules. In order to compute (1.20) we shall proceed in a fashion similar to what was done previously. The important information that has to be kept in mind is that here the time origin O of the mRNAs is in the interval $[0,t_p)$.

Rephrasing the formalism used in (1.7), we find

$$\begin{aligned}&\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0\}\\ &\quad = \sum_{n=k}^{\infty} \mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0, X(t_p) = n\}\mathsf{P}\{X(t_p) = n \mid X(0) = 0\}\,, \quad (1.21)\end{aligned}$$

and consider first $\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0, X(t_p) = n\}$. From the equivalence between Poisson processes and uniform distribution introduced in Section 2, conditioned on the number of events up to time t_p, the time position of the origin of each mRNA is uniformly distributed in $[0,t_p)$. Let now be O the random origination time of a randomly chosen mRNA and let U be its random life time. This mRNA molecule will be present at time $t_p + \Delta t$ only if the variable $Z = O + U$ satisfies $Z \geq t_p + \Delta t$. The probability p_c of this

event gives the probability per mRNA to be present at time $t_p + \Delta t$. We have therefore

$$p_c = \mathsf{P}\{Z \geq t_p + \Delta t\} = \frac{1}{t_p}\int_0^{t_p} \mathsf{P}\{Z \geq t_p + \Delta t \mid O = s\}\,\mathrm{d}s = \frac{1}{t_p}\int_{\Delta t}^{t_p+\Delta t} \big(1 - G(u)\big)\,\mathrm{d}u\,, \tag{1.22}$$

where $G(t_u)$ is defined in (1.5). Finally, it results that conditioned on $X(t_p) = n$ the number of labeled mRNA still present at time $t_p + \Delta t$ is binomially distributed according to

$$\mathsf{P}\{R(t_p + \Delta t) = k \mid X(0) = 0, X(t_p) = n\} = \binom{n}{k} p_c^k (1 - p_c)^{n-k}\,. \tag{1.23}$$

Using now (1.1) and (1.21) the time dependent distribution of the number of labeled mRNA at time $t_p + \Delta t$ after the start of the pulse and Δt time units after the interruption of the pulse phase and beginning of the chase phase is given by

$$\mathsf{P}\{R(t_p + \Delta t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} A_c(t_p + \Delta t)]^k \exp\big(-\omega_{\mathrm{tc}} A_c(t_p + \Delta t)\big)}{k!}\,, \tag{1.24}$$

where

$$A_c(t_p + \Delta t) = \int_{\Delta t}^{t_p+\Delta t} \big(1 - G(u)\big)\,\mathrm{d}u\,.$$

This relation is similar to (1.10) but not identical. In fact, in (1.24) we have that $A_c(t_p + \Delta t) \to 0$ as $\Delta t \to \infty$, expressing the fact that in the chase phase the number of mRNAs cannot increase. An important point from the experimental point of view is to realize that in (1.24) there is an explicit dependence on the length of the pulse t_p. This means that the distribution will be different for the same target mRNA depending on the value of t_p. This point plays a very important role in the analysis of decay data [13].

To describe the expression of a given gene after a very long pulse phase, it is sufficient to take the limit $t_p \to \infty$ in (1.24). This simple limit gives an expression for the number of mRNA Δt units of time after the start of the chase phase, which is again a Poisson distribution with parameter

$$A_c(\Delta t) = \omega_{\mathrm{tc}} \int_{\Delta t}^{\infty} \big(1 - G(u)\big)\,\mathrm{d}u\,, \tag{1.25}$$

where we see that the life time distribution determines also the behavior of the distribution during the chase phase.

3.4 The decay law of the mean number of mRNA

In experiments aimed at investigating the decay of the number of mRNA in order to deduce its half time t_h, it is not the distribution of the number of mRNA that is measured but the mean number over an ensemble of cells. This measurement is performed by interrupting the transcription and by measuring the relative amount of mRNAs at different time points after interruption.

To compare our results with the experiments we therefore need the average number N_r of mRNA as a function of time after the interruption of transcription [13]. This is simply given by the average performed over the distribution (1.24) and gives

$$N_r(t_p+\Delta t) = \omega_{\mathrm{tc}} A_c(t_p+\Delta t) = \omega_{\mathrm{tc}} \int_{\Delta t}^{t_p+\Delta t} \big(1-G(u)\big)\,\mathrm{d}u\,, \tag{1.26}$$

where the half time t_h measured in the experiments is then given by solving

$$N_r(t_p+t_h) = \frac{1}{2} N_r(t_p)\,, \tag{1.27}$$

as a function of t_h. By substituting (1.26) into (1.27) we find that t_h must satisfy

$$\int_{t_h}^{t_p+t_h} \big(1-G(u)\big)\,\mathrm{d}u = \frac{1}{2}\int_0^{t_p} \big(1-G(u)\big)\,\mathrm{d}u\,,$$

which, in the case in which $t_p \to \infty$ leads to

$$\int_{t_h}^{\infty} \big(1-G(u)\big)\,\mathrm{d}u = \frac{1}{2}\langle U\rangle\,.$$

To fix the ideas about this relationship, we will consider the example in which the life time probability density is a Gamma function of order $n=0$ and $n=1$, where

$$\phi_U(t) = \frac{\lambda(\lambda t)^n \exp(-\lambda t)}{n!}\,,$$

with $n = 0,1,\ldots$ and $\lambda = {}^{(n+1)}/_{\langle U\rangle}$. For $n=0$, namely for an exponentially distributed life time, the half time t_h is related to the mean life time $\langle U\rangle$ through

$$t_h^{(n=0)} = \langle U\rangle \log 2 \sim 0.69\langle U\rangle\,,$$

as is well known. On the other hand, for $n = 1$ the relationship is

$$t_h^{(n=1)} \sim 0.57\langle U \rangle \,,$$

for the same mean value. If one would now ignore that the result for $n = 1$ does not come from an exponential density and one would assume that it comes from an exponential, one would try to derive the mean value $\langle U \rangle'$ such that $t_h = \langle U \rangle' \log 2$. This would lead to

$$\langle U \rangle' = 0.83\,\langle U \rangle \,,$$

namely it would lead to an underestimation of the mean life time by about 20%.

3.5 Age and rest life time distributions

Given that there is a turn-over of the mRNA, there is an age distribution of the molecules. This age distribution depends on the length of the pulse phase and on the time point of measurement after the start of the chase phase. The same holds also for the rest life time. There is a connection with renewal theory here. In renewal theory the age is called *current life* and the rest life time is called *excess life*.

3.5.1 The age distribution of one chain

We shall start by considering the age distribution of one mRNAs at any arbitrary point in time $t + \Delta t$ after the start of the pulse phase[1]. At this point in time, the only age distribution that can be computed is the one of mRNAs that at this time are not yet degraded. In other words, one can compute the age distribution of an mRNA only under the condition that this molecule still exists at the moment of measurement. Using the same notation as before, let us call O the time of origination of a randomly chosen labeled mRNA and U its random life time. The random time of origination, O, must be contained in the interval $[0,t)$. The given mRNA will be present at time $t + \Delta t$ only if the variable $Z = O + U$ satisfies $Z > t + \Delta t$. Then, the age distribution of the mRNA is given by the distribution of $A = t + \Delta t - O$ under the condition $Z > t + \Delta t$.

[1]The choice of $t + \Delta t$ for the measurement time may seem to be a bit unusual at this point. We shall see that this choice allows for a natural derivation of a number of useful limits

In order to compute this quantity we shall first compute the probability density for O conditional that $Z > t + \Delta t$. Thus, we have

$$\begin{aligned}
\mathsf{P}\{O \le x \mid Z > t + \Delta t\} &= \frac{\mathsf{P}\{O \le x, Z > t + \Delta t\}}{\mathsf{P}\{Z > t + \Delta t\}} \\
&= \int_{t+\Delta t - x}^{\infty} \frac{\mathsf{P}\{O \le x, Z > t + \Delta t \mid U = u\}}{\mathsf{P}\{Z > t + \Delta t\}} \phi_U(u)\,\mathrm{d}u \\
&= \int_{t+\Delta t - x}^{\infty} \frac{\mathsf{P}\{t + \Delta t - u < O \le x\}}{\mathsf{P}\{Z > t + \Delta t\}} \phi_U(u)\,\mathrm{d}u
\end{aligned}$$

for $0 \le x < t$ by making a simple use of the definition of conditional probability and of the law of total probability. We shall now use the fact that the random variable O is uniformly distributed in $[0,t)$ (because we have conditioned that there is one mRNA alive at time $t + \Delta t$) and thus that $\mathsf{P}\{Z > t + \Delta t\}$ is given by a formula similar to (1.22). After canceling the normalization factors $t + \Delta t$ in the numerator and denominator, this leads to

$$\begin{aligned}
&\int_{t+\Delta t - x}^{\infty} \frac{\mathsf{P}\{t + \Delta t - u < O \le x\}}{\mathsf{P}\{Z > t + \Delta t\}} \phi_U(u)\,\mathrm{d}u \\
&\qquad = \left[\int_{\Delta t}^{t+\Delta t} \big(1 - G(y)\big)\,\mathrm{d}y\right]^{-1} \int_{t+\Delta t - x}^{\infty} \big[x - \max(0, t + \Delta t - u)\big]\,\phi_U(u)\,\mathrm{d}u\,,
\end{aligned}$$

from which we can compute the distribution of the age $A = t + \Delta t - O$ under the condition $Z > t + \Delta t$. This distribution is given by

$$\begin{aligned}
\mathsf{P}\{A \le a \mid Z > t + \Delta t\} &= 1 - \mathsf{P}\{O \le t + \Delta t - a \mid Z > t + \Delta t\} \\
&= 1 - \left[\int_{\Delta t}^{t+\Delta t} \big(1 - G(y)\big)\,\mathrm{d}y\right]^{-1} \int_{a}^{\infty} \big[\min(u, t + \Delta t) - a\big]\,\phi_U(u)\,\mathrm{d}u\,,
\end{aligned}$$

which finally leads to the probability density function after differentiation

$$\phi_A(a \mid t + \Delta t) = \left[\int_{\Delta t}^{t+\Delta t} \big(1 - G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1 - G(a)\big)$$

for $\Delta t \le a < t + \Delta t$ and zero otherwise. One can discuss the following limiting cases:

- $\Delta t > 0$ $t = t_p > 0$. In this case $\phi_A(a \mid t_p + \Delta t)$ gives the age distribution of the mRNAs at time $t_p + \Delta t$ after the start of labeling given that the pulse phase has

been interrupted at time t_p.

- ◇ $\Delta t = 0$ and $t \leq t_p$. In this case $\phi_A(a \mid t)$ gives the age distribution of the mRNAs at time t after the start of pulse.

- ◇ $\Delta t = 0$ and $t_p \to \infty$. In the case

$$\phi_A^{\mathrm{st}}(a) = \lim_{t\to\infty} \phi_A(a \mid t) = \frac{1-G(a)}{\langle U \rangle} \tag{1.28}$$

gives the stationary age distribution of the mRNAs assuming that labeling was active since very long time. A comparison between this result and its numerical estimation is given in Figure 1.3.

- ◇ $\Delta t > 0$ $t_p \to \infty$. In this case

$$\lim_{t\to\infty} \phi_A(a \mid t+\Delta t) = \left[\int_{\Delta t}^{\infty} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1-G(a)\big)$$

gives the age distribution of the mRNAs Δt time units after the beginning of the chase phase, assuming that the pulse duration was so long that all the mRNAs are labeled.

- ◇ $\Delta t = 0$ and $t \ll 1$ leads to $\phi_A(a \mid t) = t^{-1}$ and thus gives an age distribution that is asymptotically flat between 0 and t.

3.5.2 The rest life time distribution

The determination of the distribution of the rest life time L proceeds in a similar fashion. Let $Z = O + U$ be the sum of the time origin and of the life time of a given mRNA and let $t + \Delta t$ the time of observation or measurement after the start of the pulse phase. The rest life time is clearly given by $L = Z - (t + \Delta t)$. We start then by looking at the statistics of Z. We have

$$\begin{aligned}\mathsf{P}\{Z \leq t+\Delta t+x \mid Z > t+\Delta t\} &= \frac{\mathsf{P}\{t+\Delta t < Z \leq t+\Delta t+x\}}{\mathsf{P}\{Z > t+\Delta t\}} \\ &= \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \left(\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y - \int_{\Delta t+x}^{t+\Delta t+x} \big(1-G(y)\big)\,\mathrm{d}y\right),\end{aligned}$$

where one can easily verify that this expression satisfies the conditions to be a probability function. Using this result, one can easily find out that the following relationship holds

$$\mathsf{P}\{L \leq r \mid Z > t + \Delta t\} = \mathsf{P}\{Z \leq t + \Delta t + r \mid Z > t + \Delta t\},$$

which leads to the probability density for L upon derivation by r, thus leading to

$$\phi_L(r \mid t + \Delta t) = \left[\int_{\Delta t}^{t+\Delta t} \big(1 - G(y)\big)\,\mathrm{d}y\right]^{-1} \big(G(t + \Delta t + r) - G(\Delta t + r)\big),$$

for $r \geq 0$. One can discuss the following limiting cases:

- $\Delta t > 0$ $t = t_p > 0$. In this case $\phi_L(r \mid t_p + \Delta t)$ gives the rest life time distribution of the mRNAs at time $t_p + \Delta t$ after the start of labeling given that the pulse duration was t_p.

- $\Delta t = 0$. In this case $\phi_L(r \mid t)$ gives the rest life time distribution of the mRNAs at time t after the start of the pulse phase.

- $\Delta t = 0$ and $t \to \infty$. In the case

$$\phi_L^{\mathrm{st}}(r) = \lim_{t\to\infty} \phi_L(r \mid t) = \frac{1 - G(r)}{\langle U \rangle} \tag{1.29}$$

  gives the stationary rest life time distribution of the mRNAs assuming that the pulse duration was so long to reach steady state. A comparison between the theoretical prediction and its numerical estimation is given in Figure 1.3. One can notice that the stationary age and stationary rest life time distributions are identical.

- $\Delta t > 0$ and $t \to \infty$. In this case

$$\lim_{t\to\infty} \phi_L(r \mid t + \Delta t) = \left[\int_{\Delta t}^{\infty} \big(1 - G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1 - G(\Delta t + r)\big),$$

  gives the rest life time distribution of the mRNAs Δt time units after the start of the chase phase assuming that labeling was active since very long time and the number of labeled mRNAs had reached steady state.

- ⋄ $\Delta t = 0$ and $t \to 0$ leads to $\phi_L(r \mid t \to 0) = \phi_U(r)$ as it is intuitively clear since the rest life of a new born mRNA is given by its whole life.

3.5.3 Heuristic derivation of the age and rest life time distributions

At any given moment of observation far away from the initial condition we will detect an mRNA with a probability that is proportional to its life time U. This means that the probability to catch an mRNA of total life time between u and $u + \mathrm{d}u$ is proportional to $u\phi_U(u)\,\mathrm{d}u$. Thus, normalizing accordingly we obtain that this probability is governed by the density $u\phi_U(u)/\langle U\rangle$.

Now, conditional that the life time of an mRNA is $U = u$, both the age and the rest life of this mRNA are uniformly distributed in $[0, u)$ with density $1/u$, namely

$$\phi_A(x \mid U = u) = \phi_L(x \mid U = u) = \frac{1}{u},$$

with $0 \leq x < u$. Finally,

$$\phi_A(x) = \phi_L(x) = \int_x^\infty \frac{1}{u}\frac{u\phi_U(u)}{\langle U\rangle}\,\mathrm{d}u = \frac{1 - G(x)}{\langle U\rangle},$$

in agreement with the previous calculations.

• •

4 Time dependent transcription rates

In this section we abandon the experimental set-up of pulse and chase and we describe more closely a situation in which the synthesis (or transcription) rate changes with time. This situation is biologically meaningful because the life of a single cell is made of different phases (cell-cycle phases) in which the expression of certain genes is turned on or off. Here we will assume that the changes of the transcription rate follow a known temporal, deterministic pattern. As before, we will describe the distribution of the number of mRNA molecules and other statistical properties of the system under the condition that the transcription rate is not constant. Here, we will consider the simple problem of a deterministic series of changes in transcription rate.

4.1 Two consecutive transcription phases

In the previous sections we have considered always the initial condition in which the number of mRNA chains present at time zero (or at any initial arbitrary time) was zero. This was useful in order to single out the pattern of growth of the number of mRNA chains without the complication of a different initial condition. It is perfectly possible to perform the computation with a different initial condition but it is certainly more convenient to deal with this question according to the following scheme. We start by considering here the case in which at time $t_i^{(1)}$ starts a transcription process with rate $\omega_{\text{tc}}^{(1)}$ which is concluded at time $t_s^{(1)}$ followed by a second transcription process starting at time $t_i^{(2)} = t_s^{(1)}$ with rate $\omega_{\text{tc}}^{(2)}$ and concluded at time $t_s^{(2)}$. We assume here that the life time distribution ϕ_U is unaffected by these changes and that it is not dependent on the time at which the measurements are made. We also assume that at time $t_i^{(1)}$ there are no mRNA chains of this kind, so that $X(t_i^{(1)}) = 0$.

We consider now a measurement time t at some point in the middle of the second transcription phase. Thus, $t_i^{(2)} < t < t_s^{(2)}$ and therefore $t > t_s^{(1)}$. We want to know what is the distribution of the number of mRNA chains at this time t. The first observation to make is that this distribution has two contributions. The first contribution comes from the mRNAs that originated in the first period and did not degrade before time t. The second contribution comes from the mRNAs that originated in the second period, before time t, and did not degrade before time t. These two contributions are independent of each other and thus the number of mRNA that we observe at time t is made of the sum of two independent random numbers.

The contribution from the mRNAs that originated in the first period leads to a Poisson distribution that was derived in Section 3.3 Eq. (1.24). In the present contest, this is a Poisson distribution with parameter

$$\lambda_1(t) = \omega_{\text{tc}}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \big(1 - G(u)\big)\, \mathrm{d}u\,, \tag{1.30}$$

where we have substituted the time intervals in (1.25) in explicit time differences. The contribution arising from the second period, instead, can be derived using the results of Section 3.2 Eq. (1.10). This leads to a Poisson distribution with parameter

$$\lambda_2(t) = \omega_{\text{tc}}^{(2)} \int_{0}^{t-t_i^{(2)}} \big(1 - G(u)\big)\, \mathrm{d}u\,, \tag{1.31}$$

where the lower integration limit is zero because $t < t_s^{(2)}$. Although the present status of the derivation may look discouraging, especially if one thinks about its generalization to more than two periods, here it follows a result that saves the day.

Let X_1 and X_2 two independent Poisson distributed random numbers, whose distributions have parameters λ_1 and λ_2, respectively. Let Y be the random number given by the sum of X_1 and X_2,

$$Y = X_1 + X_2 \,,$$

then the probability distribution of Y is also Poisson with parameter $\lambda_1 + \lambda_2$. The simplest way to see this, is by means of the generating functions. Indeed, the generating function $g_i(s)$ of the variables X_i is given by

$$g_i(s) = \exp\left[-\lambda_i(1-s)\right]$$

and thus the generating function $g(s)$ of their sum is the product of the g_i, given by

$$g(s) = g_1(s)g_2(s) = \exp\left[-(\lambda_1+\lambda_2)(1-s)\right] \,,$$

which concludes the demonstration.

Therefore, the distribution of the number of mRNA at time t at any point in the middle of the second transcription period is a Poisson distribution with parameter

$$\lambda(t) = \omega_{\mathrm{tc}}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \big(1-G(u)\big)\,\mathrm{d}u + \omega_{\mathrm{tc}}^{(2)} \int_0^{t-t_i^{(2)}} \big(1-G(u)\big)\,\mathrm{d}u \,, \tag{1.32}$$

with the first term from (1.30) and the second from (1.31).

4.1.1 Change of transcription and degradation

The switching point between two different transcription rates can be the time point at which also the degradation changes. This would imply that before the switching point the mRNA are transcribed with rate $\omega_{\mathrm{tc}}^{(1)}$ and have a lifetime probability density $\phi_U^{(1)}$ while after the switching they are generated with a rate $\omega_{\mathrm{tc}}^{(2)}$ and have a lifetime probability density $\phi_U^{(2)}$. Under these conditions, the parameter of the Poisson distribution for the

mRNAs generated after the switching point $t_i^{(2)}$ is given by (1.31) with the modification

$$\lambda_2(t) = \omega_{\mathrm{tc}}^{(2)} \int_0^{t-t_i^{(2)}} \big(1 - G_2(u)\big)\,\mathrm{d}u\,, \tag{1.33}$$

where $G_2(t)$ is the probability function. The fate of the mRNAs generated before the switching, i.e. before time $t_i^{(2)}$ is more complicated and could be characterized in different ways depending on the biological processes of aging. The simplest assumption is that the age accumulated by each mRNA during the first period is carried over or recognized by the environment in the second period as well (for instance, only some of the microscopic rates are changed). Under this scenario, we have that the conditional probability given in Eq. (1.10) must be modified as follows

$$\mathsf{P}\{U \geq t - s\} = \mathsf{P}\{U_1 \geq t_s^{(1)} - s\}\mathsf{P}\{U_2 \geq t - s \mid U_2 \geq t_i^{(2)} - s\}\,,$$

where the conditional probability in the rhs expresses the fact that also according to the degradation process that determines U_2 the mRNAs have the same age as under the process that determines U_1. Notice that in all these equations the identity $t_s^{(1)} \equiv t_i^{(2)}$ holds. Thus, now one obtains

$$\lambda_1(t) = \omega_{\mathrm{tc}}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \frac{1 - G_1\big(t_s^{(1)} - s\big)}{1 - G_2\big(t_i^{(2)} - s\big)} \big(1 - G_2(t-s)\big)\,\mathrm{d}s\,, \tag{1.34}$$

which obeys the obvious property that it goes back to (1.30) when the two degradation processes are identical. Finally, at any time $t > t_i^{(2)}$ the number of mRNA is distributed according to a Poisson distribution with parameter

$$\lambda(t) = \lambda_1(t) + \lambda_2(t)\,,$$

given by (1.34) and (1.33).

4.2 A long series of transcription phases

It is not difficult to generalize the result derived in (1.32) to accomplish a row of $n+1$ such phases, each characterized by its own transcription rate $\omega_{\mathrm{tc}}^{(j)}$. Indeed, also the sum

on many independent Poisson distributed random numbers is Poisson distributed with parameter given by the sum of the parameters.

Thus let us assume that there are $n+1$ phases. Each phase j starts at time $t_i^{(j)}$ and terminates at time $t_s^{(j)}$ with transcription rate $\omega_{\mathrm{tc}}^{(j)}$. We make a measurement at time t such that it is in the middle of the $n+1$ phase, namely $t_i^{(n+1)} < t < t_s^{(n+1)}$. Thus, the distribution of the number of mRNA molecules at time t, will be a Poisson distribution with parameter

$$\lambda(t) = \sum_{j=1}^{n} \omega_{\mathrm{tc}}^{(j)} \int_{t-t_s^{(j)}}^{t-t_i^{(j)}} \big(1-G(u)\big)\,\mathrm{d}u + \omega_{\mathrm{tc}}^{(n+1)} \int_0^{t-t_i^{(n+1)}} \big(1-G(u)\big)\,\mathrm{d}u\,, \tag{1.35}$$

which in the limit of large n would give a distribution that does not depend on the initial conditions at time $t_i^{(1)}$ and where we still assume that the sequence of all $\omega_{\mathrm{tc}}^{(j)}$ is known and deterministic.

Now we notice that we can rewrite (1.35) by using the following definition. Let

$$\omega_{\mathrm{tc}}(\tau) = \omega_{\mathrm{tc}}^{(j)}\,,$$

if $t_i^{(j)} \leq \tau < t_s^{(j)}$. Then from (1.35) we obtain

$$\lambda(t) = \sum_{j=1}^{n} \int_{t-t_s^{(j)}}^{t-t_i^{(j)}} \omega_{\mathrm{tc}}(t-u)\big(1-G(u)\big)\,\mathrm{d}u + \int_0^{t-t_i^{(n+1)}} \omega_{\mathrm{tc}}(t-u)\big(1-G(u)\big)\,\mathrm{d}u\,, \tag{1.36}$$

which then, taking into consideration that $t_i^{(j+1)} = t_s^{(j)}$ leads to

$$\lambda(t) = \int_0^{t-t_i^{(1)}} \omega_{\mathrm{tc}}(t-u)\big(1-G(u)\big)\,\mathrm{d}u\,. \tag{1.37}$$

If we, without loss of generality, arbitrarily set now $t_i^{(1)} = 0$ and apply a rather intuitive change of variable, we obtain that the distribution of the number of mRNA during a process with any arbitrary number of transcription phases can be expressed as a Poisson distribution with parameter

$$\lambda(t) = \int_0^{t} \omega_{\mathrm{tc}}(u)\big(1-G(t-u)\big)\,\mathrm{d}u\,, \tag{1.38}$$

where we see that the major contribution arises from the recent events, close to t, if we remind that $\big(1-G(x)\big) \to 0$ as $x \to \infty$. For the present and future reference we write then the full form of the distribution as

$$\mathsf{P}\{R(t)=k \mid X(0)=0,\, \omega_{\mathrm{tc}}(u)\; 0 \le u < t\} = \frac{[\lambda(t)]^k \exp\big(-\lambda(t)\big)}{k!}\,, \tag{1.39}$$

with $\lambda(t)$ given in (1.38). We shall notice that until one does not know the complete behavior of $\omega_{\mathrm{tc}}(t)$ for large t it is not possible to discuss the question of the stationary distribution.

• • • • • • • • • •

5 Conclusions

The previous sections have provided a general framework to describe any kind of non-exponential decay pattern. We have found out that as long as the synthesis rate is either constant or varies deterministically in time, the distribution of the number of molecules in the cell is a time-dependent Poisson distribution. Nevertheless, when dealing with experimental data we would like to find out some properties associated to the molecule that has been measured, such as the mean lifetime. The framework outlined above, thus, has to be complemented with a parametric way to obtain the lifetime distribution of the molecules.

An important hint in this direction comes from biochemistry. The degradation mechanisms are characterized by the interaction of the target molecule with other molecules, complexes or enzymes present in the cell environment. For instance, in miRNA mediated degradation, the target mRNA is first recruited by a complex made of miRNA and other molecules. Afterwards, once the miRNA and its partners are bound to the target mRNA, more enzymes are recruited that start modifying the structure of the mRNA and start degrading the mRNA one nucleotide per step. In general, thus, the degradation process can be described at the single molecule level as a network of biochemical interactions. In this network, the vertices are biochemical state of the target mRNA. One or more of these states then lead to the degradation, which is obviously irreversible. If we take this idea, we can generalize it by describing the target molecule as visiting a network of states during its life until a final absorbing state is reached, which is the complete degradation [15]. For our purposes, a molecule is degraded when it is no longer detectable by the

experimental technique. In an even more abstract step, we assume that it is possible to describe the single-molecule process of degradation as a continuous time Markov chain which has a well defined initial condition (the newly synthesized mRNA) and has one well defined absorbing state (the degraded state). In this way, the life time probability density $\phi_U(t)$ is just the probability density of the absorption time on this network.

When the details of the biochemical process of degradation are known, the network of states, or at least its starting structure, can be taken from it. In that case, the data are used to estimate the parameters (the transition rates) of the network and the analysis can be used to validate the hypothesized biochemical process [15]. When the biochemical network is unknown, the network of states has to be chosen among the most parsimonious networks that are able to fit the data in a satisfactorily manner. Here, the analysis provides important kinematic properties of the degradation process and some hint to further experimental clarification of the underlying biochemistry.

Alternatively, one can define an age dependent degradation rate by using results from hazard rate theory [8], whenever there is a mechanistic model of degradation that cannot be cast into the framework of a stochastic process with a finite number of discrete states.

• • • • • • • • • • • •

Acknowledgments

I would like to thank all those people who contributed to the understanding of this topic, both from the theoretical and from the biological point of view, by providing fruitful contributions, useful criticisms, and inspiring comments: J. Belasco, G. Blanchard, D. Chiarugi, C. Deneke, Z. Ignatova, R. Lipowsky, E. McShane, J. Pedraza, S. Rœlly, S. Rudorf, M. Selbach, C. Sin, C. Vogel.

• • • • • • • • •

Bibliography

[1] Gillespie, J. H.: *Population Genetics: A concise guide*. John Hopkins University Press (2004).

[2] Ewens, W. J.: *Mathematical population genetics: I. Theoretical introduction*. Springer, (2004).

[3] Sin, C., Chiarugi, D., Valleriani, A.: *Quantitative assessment of ribosome drop-off in E. coli*, Nucl. Acid Res. **44**(6), 2528–2537 (2016).

[4] Braun, J. E., Huntzinger, E., Fauser, M., Izaurralde, E.: *GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets*, Mol. Cell. **44**(1), 120–133 (2011).

[5] Huntzinger, E., Izaurralde, E.: *Gene silencing by microRNAs: Contributions of translational repression and mRNA decay*, Nat. Rev. Gen. **12**(2), 99–110 (2011).

[6] Belasco, J. G.: *All things must pass: Contrasts and commonalities in eukaryotic and bacterial mRNA decay*, Mat. Rev. Mol. Cell Biol. **11**, 467–478 (2010).

[7] Deneke, C., Rudorf, S., Valleriani, A.: *Transient phenomena in gene expression after induction of transcription*, PLoS one **7**(4), e35044 (2012).

[8] Deneke, C.: *Theory of mRNA degradation*, PhD Thesis University of Potsdam (2012).

[9] Deneke, C., Lipowsky, R., Valleriani, A.: *Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA*, PLoS one **8**(2), e55442 (2013).

[10] Deneke, C., Lipowsky, R., Valleriani, A.: *Effect of ribosome shielding on mRNA stability*, Phys. Biol. **10**, 046008 (2013).

[11] Wheatley, D. N., Giddings, M. R., Inglis, M. S.: *Kinetics of degradation of "short-" and "long-lived" proteins in cultured mammalian cells*, Cell Biol. Int. Rep. **4**(12), 1081–1090 (1980).

[12] McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J.A., Valleriani, A., Selbach, M.: *Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation*, Cell **167**, 803–815 (2016).

[13] Sin, C., Chiarugi, D., Valleriani, A.: *Degradation parameters from pulse-chase experiments*, PLoS one **11**(5), e0155028 (2016).

[14] Taylor, H. M., Karlin, S.: *An introduction to stochastic modeling*. Academic Press, San Diego (1998).

[15] Sin, C., Chiarugi, D., Valleriani, A.: *Single-molecule modeling of mRNA degradation by miRNA: Lessons from data*, BMC Sys. Biol. **9**(S3), S2 (2015).

Chapter 2

Algunas propiedades básicas de procesos de ramificación

*Sylvie Rœlly**

Introducción

La teoría de procesos de ramificación surge a principios del siglo XX: el origen se atribuye a los estudios realizados, independientemente, por Bienaymé en Francia y por Galton y Watson en Inglaterra, con objeto de investigar las razones de la extinción de determinadas líneas familiares de la aristocracia.

Se considera como punto de partida el siguiente *Problema 4001* planteado por F. Galton en la revista *Educational Times*, en 1874:

> *Consideremos una población de la que sólo nos interesan los varones adultos, de cantidad i que representan apellidos distintos e independientes. Su ley de población es tal que, en cada generación, un r_0 por ciento de los varones adultos tienen hijos varones de los cuales ninguno alcanza la edad adulta, un r_1 por ciento tienen un sólo hijo varón que alcanza la edad adulta, un r_2 por ciento tiene dos, y así sucesivamente hasta un r_5 por ciento que tienen cinco. Encontrar* (1) *¿qué proporción de apellidos se extinguirán después de n generaciones?* (2) *¿Cuántos individuos habrá del mismo apellido en una generación formada por j personas?*

*Institut für Mathematik der Universität Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam OT Golm, Germany; *roelly@math.uni-potsdam.de*; www.math.uni-potsdam.de/professuren/proba

La primera respuesta a este problema fue enviada a Galton por el reverendo H. W. Watson. La respuesta aportada por Watson no era totalmente satisfactoria y el problema de la probabilidad de extinción permaneció, al menos parcialmente, sin solución durante otro medio siglo. En 1922, R. A. Fisher consideró dicho problema en un contexto genético, trasladándolo al estudio de la progenie de un gen mutante. La importancia de su trabajo reside en que por primera vez los Procesos de Galton-Watson fueron aplicados en la genética. A partir de entonces, los procesos de ramificación han sido utilizados como modelos matemáticos para describir procesos empíricos, no sólo relacionados con la biología, sino también con la física nuclear, la medicina, las ciencias de la computación, la demografía, etc. Durante el siglo XX se ha desarollado ampliamente, ocupando en la actualidad un lugar destacado dentro del contexto general de la teoría de procesos estocásticos, con numerosas monografías publicadas sobre su teoría y aplicaciones, entre las que citaremos por su especial significación las de Harris [15], Athreya & Ney [1], Jagers [17], Sewastjanow [25], Pakes [24], Axel & Kimmel [2], Duquesne & Le Gall [6], Haccou, Jagers & Vatutin [14].

Las notas se estructuran en 3 capítulos. El primer capítulo incluye varias propiedades (Markov, martingala, extinción, cuasi-estacionaridad, condicionamiento a la no extinción) del proceso de ramificación simple de Bienaymé-Galton-Watson (BGW), un proceso de tiempo discreto. En el capítulo dos se introduce el proceso de ramificación BGW de tiempo continuo y se estudian sus propiedades como proceso de Markov, y como martingala. También se analiza lo que resulta cuando se condiciona la dinámica a la no extinción en el futuro lejano. Se discute la similitud con el caso de tiempo discreto. El tercer capítulo incluye límites de procesos de BGW para densidad de población grande. Así se costruye -entre otras cosas- el proceso de difusión de Feller. Se mencionarán unas propiedades cualitativas típicas.

Deseo expresar mi agradecimiento especialmente a mis compañeros Sylvie Méléard, Maria Emilia Caballero, Begoña Fernández, Luis Gorostiza, Manuel Molina Fernández y Miguel González Velasco por su material [22], [4], [10], [12], [21] y [23] que me prestaron muy generosamente y a Sophie Pénisson, Jessica Erl y Dennis Gemballa por haber dibujado algunas figuras.

Este curso se dictó por primera vez al ICIMAF, La Habana, Cuba, en abril 2009 y marzo 2010. Agradezco la oportunidad de presentar una version mejorada en la Universidad de los Andes en diciembre 2016, y espero haber podido despertar el interés de los jóvenes científicos para este dominio profundo y rico de matemáticas que lleva todavía numerosos secretos!

Potsdam, septiembre 2017

. . . . . .

1 Los procesos de Bienaymé-Galton-Watson en tiempo discreto (BGW): algunas propiedades

Consideramos una población sujeta a reproducción. En el instante $n = 0$ hay X_0 individuos. Al transcurrir el tiempo, después de un tiempo de vida igual a 1, los individuos se reproducen independientemente según una ley de ramificación $\mathbf{r} := (r_j)_{j\in\mathbb{N}}$. Los nuevos individuos – después de un tiempo de vida igual a 1 – se reproducen independientemente con las mismas reglas.

El proceso de BGW es la sucesión de variables aleatorias $(X_n)_{n\in\mathbb{N}}$ – con espacio de estados $\mathbb{N}$ – representando el número total de individuos de las generaciónes sucesivas.

1.1 Definición

El proceso de BGW se define de modo siguiente:

$$X_{n+1} := \begin{cases} \sum_{k=1}^{i} Y_{n,k} & \text{si } X_n = i \geq 1 \\ 0 & \text{si } X_n = 0. \end{cases}$$

Las variables aleatorias $(Y_{n,k})_{n,k}$ son independientes e idénticamente distribuidas con distribución $\mathbf{r}$. Intuitivamente, $Y_{n,k}$ representa el número de descendientes del k-ésimo individuo de la n-ésima generación.

En estas notas, supondremos siempre que la ley $\mathbf{r}$ no es degenerada, es decir: $r_0 > 0$ y $r_0 + r_1 < 1$.

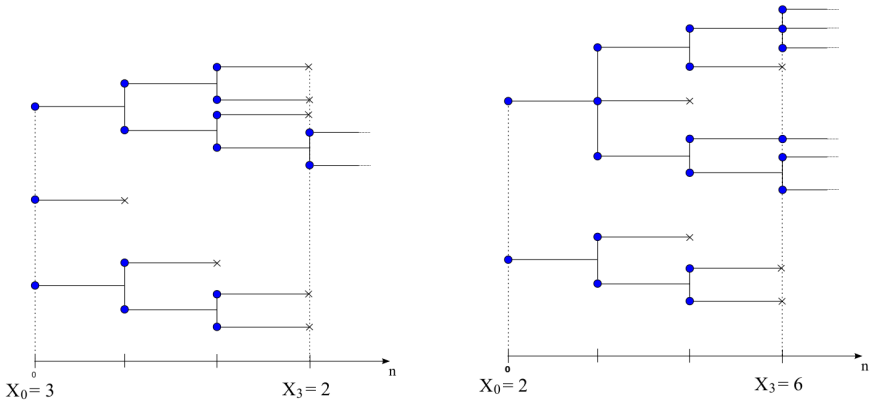


Figura 2.1: Una realización del BGW binario, Ej. 2.1 1), y BGW **(W2)** propuesto por Watson, Ej. 2.1 3) (S. Pénisson).

Ejemplo 2.1

1) BGW con reproducción binária: ver figura 2.1
$r_2 = 1 - r_0$, $r_j = 0$ por $j \notin \{0,2\}$. Cuando $r_0 = r_2 = 1/2$ se llama proceso de BGW *crítico*.

2) Reproducción de segmentos de ADN de tamaño N:
$r_0 = b$, $r_1 = (1-b)(1-a^N)$, $r_2 = (1-b)a^N$, $r_j = 0$ para todas $j \geq 3$, $a, b \in]0,1[$.

3) Modelos propuestos por Watson:

(W1) $\mathbf{r}$ es la distribución uniforme: $U_{\{0,1,2\}}$, así $r_0 = r_1 = r_2 = 1/3$ ver [11], p. 141

(W2) $\mathbf{r}$ es una distribución binomial: $\mathbf{r} = \mathscr{B}_{5,1/4}$, ver [11], p.142 y figura 2.1. Así que

r_0	r_1	r_2	r_3	r_4	r_5
$3^5/4^5$	$5\ 3^4/4^5$	$10\ 3^3/4^5$	$10\ 3^2/4^5$	$5\ 3/4^5$	$1/4^5$
$\approx 0{,}237$	0,396	0,264	0,088	0,014	0,001

4) Modelo geométrico de parámetro $a \in [0,1]$: $\mathbf{r} = \mathscr{G}_a$
$r_j = a(1-a)^j$ para $j \geq 0$.

5) Modelo cuasi-geométrico (a,b) – o geométrico modificado en 0:
$r_0 = b$, $r_j = (1-b)a(1-a)^{j-1}$ para $j \geq 1$. (Si $a = b$, es geométrico.)
Ver figura 2.2 para un ejemplo de A. Lotka [20].

	r_0	r_1	r_2	r_3	r_4	r_5	r_{10}
estimadas	0,4981	0,2103	0,1270	0,0730	0,0418	0,0241	0,0005
aproximados	0,4828	0,2289	0,12794	0,0714	0,0399	0,0233	0,0012

Se interpretan como un modelo cuasi-geométrico con los parámetros $a \approx 0,4414$ y $b \approx 0,4828$.

1.2 La propiedad de ramificación

1) Los descendientes de un individuo de cualquier generación forman un proceso equivalente al total.

2) El proceso iniciado en $X_0 = i$ es la suma de i copias independientes de X iniciadas en 1:

$$\forall i,k \in \mathbb{N} \qquad \mathsf{P}\Big((X_{n+k})_{n\geq 0} \in \cdot \,\Big|\, X_k = i\Big) = \mathsf{P}\left(\left(\sum_{j=1}^{i} X_n^{(j)}\right)_{n\geq 0} \in \cdot\right)$$

donde $\mathsf{P}\left(X^{(j)} \in \cdot\right) = \mathsf{P}(X \in \cdot \,|\, X_0 = 1)$.

1.3 La propiedad de Markov y la clasificación de los estados

Un proceso de BGW X es una cadena de Markov dado que su comportamiento futuro sólo depende del presente independientemente del pasado:

$$\begin{aligned}
&\forall n \in \mathbb{N}, i,j \geq 1, x_0, x_1, \ldots \in \mathbb{N}, \\
&\qquad \mathsf{P}\Big(X_{n+1} = j \,\Big|\, X_0 = x_0, X_1 = x_1, \ldots, X_n = i\Big) \\
&\qquad\qquad = \mathsf{P}\left(\sum_{k=1}^{i} Y_{n,k} = j \,\middle|\, X_0 = x_0, X_1 = x_1, \ldots, X_n = i\right)
\end{aligned}$$

$$= \mathsf{P}\left(\sum_{k=1}^{i} Y_{n,k} = j\right) = \sum_{k_1+\ldots+k_i=j} r_{k_1}\cdots r_{k_i}$$
$$= (\mathbf{r}^{\star i})_j = \mathsf{P}(X_{n+1} = j \,|\, X_n = i)$$

además tenemos

$$\mathsf{P}(X_{n+1} = j \,|\, X_0 = x_0, X_1 = x_1, \ldots, X_n = 0) = \begin{cases} 0 & \text{si } j \neq 0 \\ 1 & \text{si no.} \end{cases}$$

Su matriz de transición $P = (P_{ij})_{i,j\geq 0}$ satisface $P_{ij} := \mathsf{P}(X_{n+1} = j \,|\, X_n = i) = (\overbrace{\mathbf{r}\star\cdots\star\mathbf{r}}^{i})_j$. El estado 0 es absorbente, como $X_n = 0$ implica $X_m = 0$ para todos $m \geq n$.

Teorema 2.2 (Clasificación de estados) Sea X un proceso de BGW no degenerado. El estado 0 es el único estado recurrente, es decir

$$\mathsf{P}(X_n = 0 \text{ para alguna } n \in \mathbb{N} \,|\, X_0 = 0) = 1.$$

El resto de estados son transitorios: sea $i \geq 1$;

$$\mathsf{P}(X_n = i \text{ para algún } n \in \mathbb{N} \,|\, X_0 = i) < 1.$$

La única distribución estacionaria (o de equilibrio) π es trivial: $\pi = \delta_0$.

Demostración. 0 es absorbente entonces es recurrente:
Sea T_i el primer tiempo positivo de visita de X al estado $i \in \mathbb{N}$. $\mathsf{P}(T_0 < +\infty \,|\, X_0 = 0) = \mathsf{P}(T_0 = 1 \,|\, X_0 = 0) = 1$, de donde 0 es recurrente. El tiempo de primera vuelta a 0 es casi seguramente finito.
Sea $i \geq 1$.

$$\mathsf{P}(T_i = +\infty \,|\, X_0 = i) \geq \mathsf{P}(X_1 = 0 \,|\, X_0 = i) = (r_0)^i > 0,$$

de donde i es transitorio. El tiempo de vuelta a i puede ser infinito.
Sea π una distribución estacionaria. Es una ley que es invariante sobre la dinámica de reproducción: Si la ley de X_0 es π, entonces para cada tiempo n la ley de X_n se queda

igual a π; en particular,

$$\pi_0 = \mathsf{P}(X_1 = 0) = \sum_{i\geq 0} \mathsf{P}(X_1 = 0 \,|\, X_0 = i)\mathsf{P}(X_0 = i) = \sum_{i\geq 0} (r_0)^i \pi_i > \pi_0 \quad \text{(que es absurdo!)}$$

excepto si $\pi_i = 0$ para todas las $i \geq 1$. □

1.4 La función generatriz y los momentos

Se toma la notación f para la función generatriz de la ley $\mathbf{r}$:

$$f(s) := \mathsf{E}\big(s^{Y_{n,k}}\big) = \mathsf{E}\big(s^{X_1} \,\big|\, X_0 = 1\big) = \sum_{j\geq 0} r_j s^j, \qquad 0 \leq s \leq 1.$$

En el caso no degenerado la función f es continua, creciente y estríctamente convexa.

Ejemplo 2.3

1) Reproducción binaria: $f(s) = r_0 + r_2 s^2 = 1 + r_2(s^2 - 1)$.
   En el caso crítico $r_0 = r_2 = 1/2$, $f(s) = \frac{1}{2}(1+s^2)$ tenemos según los casos

2) **(W1)** $f(s) = \dfrac{1}{3}(1+s+s^2) = \dfrac{1}{3}\dfrac{1-s^3}{1-s}$

   (W2) $f(s) = \left(\dfrac{3+s}{4}\right)^5$.

3) Modelo geométrico de parámetro a: $f(s) = \dfrac{a}{1-(1-a)s}$.

4) Reproducción cuasi-geométrica (a,b): $f(s) = b + (1-b)\dfrac{as}{1-(1-a)s}$. (Ver figura 2.2.)

El valor esperado m de la ley $\mathbf{r}$ representa el número promedio de individuos generados en una reproducción:

$$m = \sum_{j\geq 0} j r_j = \mathsf{E}(X_1 \,|\, X_0 = 1) = f'(1^-).$$

TABLE . THE PROBABILITY $P_{1j}^{(n)}$ PER 1,000 THAT A NEWBORN MALE SHALL HAVE j MALE DESCENDANTS IN THE nTH GENERATION

j / n	None	One or More	1	2	3	4	5	6	7	8	9	10
$(r_j) \rightarrow$ 1	482.80	517.20	228.95	127.89	71.44	39.91	22.29	12.45	6.95	3.88	2.17	1.21
2	634.16	365.84	98.29	72.12	52.92	38.83	28.49	20.90	15.33	11.25	8.25	6.05
3	707.65	292.35	53.97	44.19	36.18	29.62	24.25	19.85	16.25	13.30	10.89	8.92
4	750.74	249.26	33.80	29.36	25.50	22.15	19.24	16.71	14.51	12.60	10.94	9.50
5	778.83	221.17	22.95	20.68	18.63	16.79	15.13	13.63	12.28	11.07	9.97	8.98
6	798.45	201.55	16.48	15.22	14.06	12.99	12.00	11.09	10.24	9.46	8.74	8.07
7	812.80	187.20	12.29	11.56	10.87	10.22	9.61	9.04	8.50	7.99	7.51	7.06
8	823.68	176.32	9.46	9.01	8.59	8.19	7.80	7.43	7.08	6.75	6.43	6.13
9	832.13	167.87	7.41	7.13	6.86	6.60	6.35	6.11	5.88	5.66	5.45	5.25
10	838.81	161.19	5.90	5.73	5.56	5.40	5.24	5.08	4.93	4.78	4.64	4.50

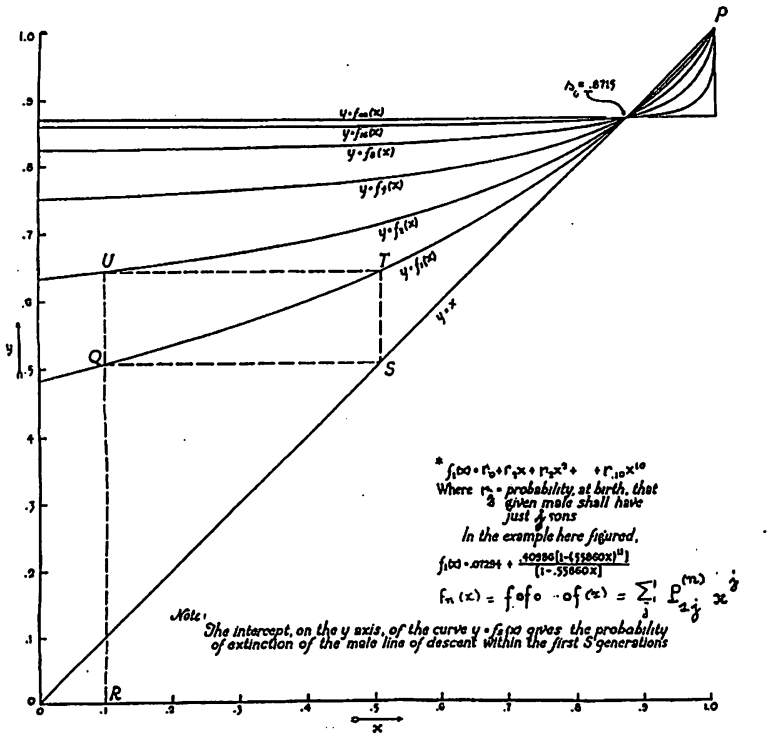


Figura 2.2: Distribucíon del modelo cuasi-gemetrico de Lotka, Ej. 2.1 5). Su función generatriz f y sus iteraciónes, ver [20].

Supondremos siempre en estas notas que $m < +\infty$.
Se consideran tres casos:

$$\begin{cases} m < 1: & \text{ramificación subcrítica} \\ m = 1: & \text{ramificación crítica} \\ m > 1: & \text{ramificación supercrítica} \end{cases}$$

Supondremos también que existe el segundo momento de la ley de reproducción:

$$\sigma^2 := \sum_{j\geq 0} (j-m)^2 r_j = \mathsf{Var}(X_1 \,|\, X_0 = 1) < +\infty.$$

Ejemplo 2.4

1) Reproducción binaria: $m = 2r_2$.
   Es crítica solo si $r_2 = 1/2$ (y por tanto $r_0 = 1/2$).

2) **(W1)** $m = 1$ y $\sigma^2 = 2/3$.
   (W2) $m = 5/4$ y $\sigma^2 = 15/16$.

3) Modelo geométrico de parámetro a: $m = \dfrac{1-a}{a}$ y $\sigma^2 = \dfrac{1-a}{a^2}$.

4) Reproducción cuasi-geométrica (a,b): $m = \dfrac{1-b}{a}$.
   En el caso de Lotka $m \approx \frac{0{,}5172}{0{,}4414} = 1{,}17$. El modelo es supercrítico.

Proposición 2.5 (Estructura iterativa de la función generatriz de BGW) Si denotamos por $f_n(s) := \mathsf{E}\big(s^{X_n} \,\big|\, X_0 = 1\big)$ la función generatriz de X_n cuando $X_0 = 1$, entonces

$$f_n = f \circ f \circ \cdots \circ f \quad (n \text{ veces }) \qquad \text{y} \qquad \mathsf{E}\big(s^{X_n} \,\big|\, X_0 = i\big) = \big(f_n(s)\big)^i, \quad i \geq 1.$$

Demostración. Por inducción tenemos para cada $n \in \mathbb{N}$

$$\begin{aligned} \mathsf{E}\big(s^{X_{n+1}} \,\big|\, X_0 = 1\big) &= \sum_{i\geq 0} \mathsf{E}\big(\mathbb{1}_{\{X_n = i\}} s^{X_{n+1}} \,\big|\, X_0 = 1\big) = \sum_{i\geq 0} \mathsf{E}\big(\mathbb{1}_{\{X_n = i\}} s^{\sum_{k=1}^{i} Y_{n,k}} \,\big|\, X_0 = 1\big) \\ &= \sum_{i\geq 0} \mathsf{E}\big(\mathbb{1}_{\{X_n = i\}} \,\big|\, X_0 = 1\big) \Big(\mathsf{E}\big(s^Y \,\big|\, X_0 = 1\big)\Big)^i \\ &= \sum_{i\geq 0} \mathsf{P}(X_n = i \,|\, X_0 = 1) f(s)^i = f_n \circ f(s). \end{aligned}$$

□

Así la respuesta a la pregunta (2) de Galton es

$$\mathsf{P}(X_n = j \,|\, X_0 = i) = \frac{1}{j!}\frac{\partial^j}{\partial s^j}\big(f_n(s)^i\big)\Big|_{s=0}.$$

Ejemplo 2.6

1) Para el modelo geométrico con parámetro a se verifica que

$$f_n(s) = \begin{cases} \dfrac{m^n(1-s)+ms-1}{m^{n+1}(1-s)+ms-1} & \text{si } a \neq 1/2, \\ \dfrac{n(1-s)+s}{(n+1)-ns} & \text{si } a = 1/2 \text{ (caso crítico).} \end{cases}$$

con $m = (1-a)/a$.

2) La reproducción cuasi-geométrica (b,a) (cf. figura 2.2):
Se muestra por inducción que en el caso no crítico con $a \neq 1-b$

$$f_n(s) = b_n + (1-b_n)\frac{a_n s}{1-(1-a_n)s},$$

con $b_n := 1 - m^n a_n j$, $a_n := \dfrac{a+b-1}{b-(1-a)m^n}$, y $m = (1-b)/a$.

Los momentos

Se verifican por inducción las siguientes igualdades:

$$\mathsf{E}(X_n) = m^n\,\mathsf{E}(X_0) \qquad \text{y} \qquad \mathsf{Var}(X_n \,|\, X_0 = 1) = \begin{cases} n\sigma^2 & \text{si } m = 1 \\ m^{n-1}\dfrac{m^n-1}{m-1}\sigma^2 & \text{si } m \neq 1. \end{cases} \tag{2.1}$$

Demostración. Por inducción obtenemos para todas las $n \in \mathbb{N}$ que

$$\mathsf{E}(X_{n+1} \,|\, X_0 = 1) = f'_{n+1}(1) = \big(f_n \circ f\big)'(1) = f'(1) f'_n \circ f(1) = m\,\mathsf{E}(X_n \,|\, X_0 = 1). \qquad \square$$

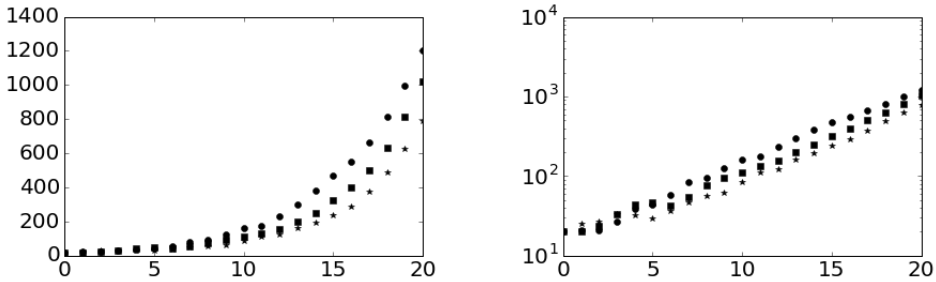


Figura 2.3: Tres realizaciones del tamaño de la población hasta $n = 20$ generaciónes del modelo BGW **(W2)** propuesto por Watson: $\mathbf{r} = \mathscr{B}_{5,1/4}$ y $X_0 = i = 20$, $m = 5/4$ (J. Erl y D. Gemballa).

Ejemplo 2.7

(W1) $\mathsf{E}(X_n) = \mathsf{E}(X_0)$ y $\mathsf{Var}(X_n) = \frac{2}{3}\, n\, \mathsf{E}(X_0)$

(W2) $\mathsf{E}(X_n) = (5/4)^n\, \mathsf{E}(X_0)$ y $\mathsf{Var}(X_n) = \frac{15}{4}(5/4)^{n-1}\big((5/4)^n - 1\big)\, \mathsf{E}(X_0)$

1.5 Martingalas asociadas al proceso BGW

Normalizando el proceso BGW (con $m \in]0, +\infty[$), se define la sucesión $W_n := \dfrac{1}{m^n} X_n$.

Proposición 2.8 (Primera propiedad de martingala) Sea $\mathscr{F}_n$ la σ-álgebra definida por $\mathscr{F}_n := \sigma(X_0, \ldots, X_n)$. Entonces, el proceso normalizado $W_n := \dfrac{X_n}{m^n}$ es una martingala no negativa con respecto a $(\mathscr{F}_n)_n$. Por lo tanto $W := \lim_n W_n$ existe casi seguramente y $0 \leq W < +\infty$.

Demostración. Una martingala debe verificar:

$$\forall n \in \mathbb{N}, \quad \mathsf{E}(W_{n+1} \,|\, \mathscr{F}_n) = W_n.$$

Sin embargo $\mathsf{E}(X_{n+1} \,|\, \mathscr{F}_n) = \mathsf{E}(X_{n+1} \,|\, X_n)$. Ademas sabemos que $\mathsf{E}\big(s^{X_{n+1}} \,\big|\, X_n = i\big) = f(s)^i$ y

$$\frac{\partial}{\partial s} \mathsf{E}\big(s^{X_{n+1}} \,\big|\, X_n = i\big) = \mathsf{E}\big(X_{n+1} s^{X_{n+1}-1} \,\big|\, X_n = i\big) = i f'(s) f(s)^{i-1}.$$

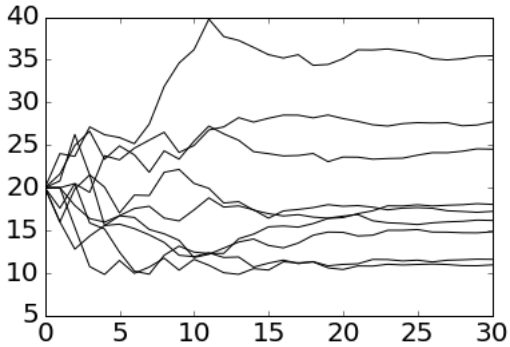


Figura 2.4: Convergencia de unas realizaciones de $W_n = X_n/m^n$ por el modelo de Watson **(W2)**, Ej. 2.1 3) (J. Erl y D. Gemballa).

Para $s = 1$ tenemos que $\mathsf{E}(X_{n+1} \,|\, X_n = i) = if'(1) = m \cdot i$ y por tanto

$$\mathsf{E}(W_{n+1} \,|\, \mathscr{F}_n) = \frac{1}{m^{n+1}} \mathsf{E}(X_{n+1} \,|\, \mathscr{F}_n) = \frac{1}{m^{n+1}} mX_n = W_n \quad \text{c.s.}$$

Se sabe que cada martingala no negativa converge c.s. (ver [26] Teorema 2.21 o [30]) y por consiguente existe W.
En el caso $m \leq 1$, $X_n \to 0$ c.s. y $W \equiv 0$. En el caso $m > 1$, $\{W > 0\} = \{X_n \to +\infty\}$. $\square$

Proposición 2.9 (Una familia de martingalas) Para cada función h acotada, el proceso $\left(h(X_n) - \sum_{k=0}^{n-1} (P - \mathrm{Id}) h(X_k) \right)_n$ es una martingala con respecto a $(\mathscr{F}_n)_n$.

Demostración. Sea $Z_n := h(X_n) - \sum_{k=0}^{n-1} (P - \mathrm{Id}) h(X_k)$.

$$\mathsf{E}(Z_{n+1} - Z_n \,|\, \mathscr{F}_n) = \mathsf{E}\big(h(X_{n+1}) \,\big|\, X_n\big) - h(X_n) - \big(P - \mathrm{Id}\big) h(X_n).$$

Sin embargo

$$\mathsf{E}\big(h(X_{n+1}) \,\big|\, X_n = i\big) = \mathsf{E}\big(h(X_1) \,\big|\, X_0 = i\big) = \sum_{j=0}^{+\infty} P_{ij} h(j) =: Ph(i)$$

por tanto $\mathsf{E}\big(h(X_{n+1}) \,\big|\, X_n\big) = Ph(X_n)$ y $\mathsf{E}(Z_{n+1} - Z_n \,|\, \mathscr{F}_n) = 0 \quad \forall n \in \mathbb{N}$. $\square$

Comentario 2.10 En el caso crítico $m = 1$, la función $h = \mathrm{Id}$ es armónica en el sentido de que $(P - \mathrm{Id})h \equiv 0$ dado que

$$\big((P - \mathrm{Id})h\big)(i) = \sum_{j \geq 0} P_{ij} h(j) - h(i) = \sum_{j \geq 0} jP_{ij} - i = \mathsf{E}(X_1 \,|\, X_0 = i) - i = mi - i = 0.$$

Entonces, verificamos una vez más que X_n es una martingala.

1.6 La probabilidad de extinción

La probabilidad de extinción de la n-ésima generación $\mathsf{P}(X_n = 0 \,|\, X_0 = 1)$ vale $f_n(0)$: Respuesta a la pregunta (1) de Watson. La probabilidad de extinción es el límite de $f_n(0)$ y representa la probabilidad del suceso de $\{X_n \to 0\}$. Ella constituyó el punto de arranque del estudio del BGW. Se buscan condiciones para que el proceso no se extinga. Dichas condiciones dependen en gran medida del valor experado m de la ley de reproducción. Esto originó la clasificación de los BGW en los tres tipos supercrítico, crítico y subcrítica, según $m > 1$, $m = 1$ o $m < 1$.

Teorema 2.11 (Probabilidad de extinción)

1) **Caso (sub)crítico**, $m \leq 1$:
   La población se extingue casi seguramente, es decir,

$$X_\infty = \lim_n X_n = 0 \quad \text{c.s.} \qquad \Longleftrightarrow \qquad \forall i \in \mathbb{N}, \quad \mathsf{P}(X_n = 0 \text{ para algún } n \,|\, X_0 = i) = 1.$$

2) **Caso supercrítico**, $m > 1$:
   La población sobrevive con probabilidad positiva, es decir,

$$X_\infty = \lim_n X_n \in \{0, +\infty\} \quad \text{existe c.s. y} \qquad \mathsf{P}(X_\infty = +\infty \,|\, X_0 = i) = 1 - q^i$$

   donde q es la unica raíz en $[0, 1[$ de la ecuación $f(s) = s$.

Demostración. El tiempo aleatorio de extinción T_0 es el primer tiempo positivo de visita a 0, $T_0 = \inf\{n \in \mathbb{N} : X_n = 0\}$. Entonces $\{T_0 < +\infty\} = \cup_{\nearrow n}\{X_n = 0\} = \{X_\infty = 0\}$. Así

$$\mathsf{P}(T_0 < +\infty \,|\, X_0 = i) = \lim_n \mathsf{P}(X_n = 0 \,|\, X_0 = i) = \lim_n \mathsf{P}(X_n = 0 \,|\, X_0 = 1)^i = \lim_n \big(f_n(0)\big)^i.$$

La prueba del teorema es inmediata si se toma en cuenta el siguiente lema. □

Lema 2.12 $\lim_n f_n(0) = q$ donde q es la raíz la mas pequeña en [0,1] de la ecuación $f(s) = s$. Cuando $m \leq 1$, $q = 1$ y cuando $m > 1$, $q < 1$.

Demostración.

- Si $m \leq 1$, $s \mapsto f'(s) - 1 \leq 0$ crece para todo $s \in [0,1]$, luego $s \mapsto f(s) - s$ decrece estríctamente desde el valor $r_0 > 0$ hasta el valor 0. Es decir $q = 1$. Dado que f crece,

$$0 < f(0) \leq f_2(0) \leq \ldots \leq f_n(0) \leq \ldots,$$

  entonces $f_n(0) \in [0,1]$ crece y tiende a q cuando $n \to \infty$, donde $q = 1$ es el único punto fijo de f en [0,1].

- Si $m > 1$, $s \mapsto f'(s) - 1$ toma valores negativos y después positivos, entonces $s \mapsto f(s) - s$ decrece a partir de r_0 y después crece hasta el valor 0. De ahí existen valores negativos y (por el teorema del valor intermedio) un valor $q \in]0,1[$ tal que $f(q) - q = 0$. Dado que f crece,

$$0 < r_0 = f(0) \leq f_2(0) \leq \ldots \leq f_n(0) \leq \ldots,$$

  entonces $f_n(0)$ crece y tiende a la raíz más pequeña de la ecuación $f(s) = s$, a saber exacto al valor q. □

Ejemplo 2.13

1) Reproducción binaria supercrítica, $r_2 > 1/2$:
   q es la raíz la mas pequeña de $r_2 q^2 - q + 1 - r_2 = 0$. Entonces $q = (1-r_2)/r_2$.

2) **(W1)** BGW crítico.

$f_1(0)$	$f_2(0)$	$f_3(0)$	$f_4(0)$	$f_5(0)$
$\approx 0{,}333$	0,481	0,571	0,641	0,675

   $q = \lim_n f_n(0)$ es raíz de $\frac{1}{3}(1 - 2x + x^2) = 0$, entonces $q = 1$.

   (W2)

$f_1(0)$	$f_2(0)$	$f_3(0)$	$f_4(0)$	$f_5(0)$	$f_6(0)$	$f_7(0)$	$f_8(0)$	$f_9(0)$	$f_{10}(0)$
$\approx 0{,}237$	0,346	0,410	0,450	0,477	0,496	0,510	0,520	0, 527	0,533

   BGW supercrítico, $q = \lim_n f_n(0)$ es la única raíz en $[0,1)$ de

$$x^5 + 15x^4 + 90x^3 + 270x^2 - 619x + 243 = 0.$$

Se calcula: $q \approx 0,553$.

3) Modelo geométrico de parámetro a: La probabilidad de extinción de la n-ésima generación vale

$$f_n(0) = \begin{cases} \dfrac{m^n - 1}{m^{n+1} - 1} & \text{si } a \neq 1/2, \\ \dfrac{n}{n+1} & \text{si } a = 1/2 \text{ (caso critico)} . \end{cases}$$

En el caso supercrítico $m > 1 \Leftrightarrow a < 1/2$, la probabilidad de sobrevivir de la n-ésima generación es

$$\mathsf{P}(X_n > 0 \,|\, X_0 = i) = 1 - \left(\frac{m^n - 1}{m^{n+1} - 1} \right)^i .$$

Entonces

$$q = \lim_n \frac{m^n - 1}{m^{n+1} - 1} = \frac{1}{m} = \frac{a}{1-a}$$

En el caso subcrítico $m < 1$,

$$\mathsf{P}(X_n = 0 \,|\, X_0 = i) = \left(1 - m^n \frac{1-m}{1-m^{n+1}} \right)^i$$

que converge a 1 cuando $n \to \infty$ con velocidad exponencial.
Ver la figura 2.5, con $m = 0,95$ y $i = 10$.

4) Reproducción cuasi-geométrica (b,a): $m = {}^{(1-b)}/_{a}$.
La probabilidad de extinción de la n-ésima generación es

$$f_n(0) = b_n = 1 - \frac{(a+b-1)m^n}{b-(1-a)m^n} .$$

En el caso supercrítico $1 - a > b$ tenemos

$$q = 1 - \lim_n \frac{m^n(a+b-1)}{b-(1-a)m^n} = 1 - \frac{a+b-1}{a-1} = \frac{b}{1-a} .$$

En la aplicación al modelo de Lotka aparece el valor $q \approx 0,86$ (ver la figura 2.2).

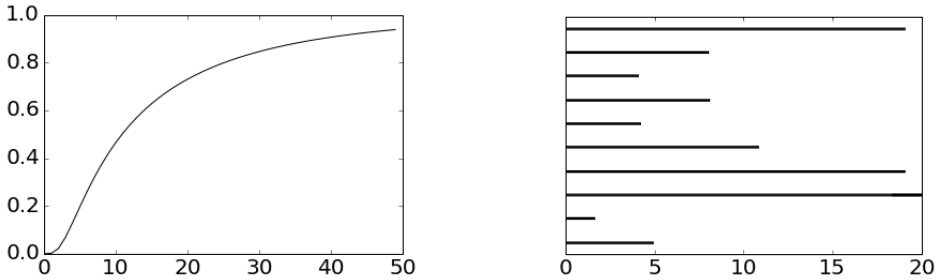


Figura 2.5: **r** geométrica de parámetro $0,51$ y $X_0 = i = 10$ ($m = 0,95$). Probabilidad de extinción de las generaciones hasta $n = 50$ (izquierda) y tiempo de sobreviviencia de cada familia (derecha) (J. Erl y D. Gemballa).

1.7 Estimación práctica del tiempo de extinción

Se puede estimar, solo con los dos primeros momentos de la ley **r**, la probabilidad de extinción de la n-ésima generación cuando la población inicial es $i \geq 1$.

1.7.1 Los cálculos

Teorema 2.14 Sea X un proceso de BGW donde la ley de reproducción tiene m como valor medio y σ^2 como varianza.

1) Para cada $i \geq 1$ tenemos

$$i\,\mathsf{P}(X_n > 0\,|\,X_0 = 1)\mathsf{P}(X_n = 0\,|\,X_0 = 1)^{i-1} \leq \mathsf{P}(X_n > 0\,|\,X_0 = i) \leq i\,\mathsf{P}(X_n > 0\,|\,X_0 = 1).$$

2) En particular, en el caso subcrítico $m < 1$ y cuando $n \to +\infty$,

$$i\,(1-m)\frac{m^{n+1}}{\sigma^2} \leq \mathsf{P}(X_n > 0\,|\,X_0 = i) = \mathsf{P}(T_0 > n\,|\,X_0 = i) \leq i\,m^n.$$

Demostración. Primera desigualdad:

$$\begin{aligned}\mathsf{P}(X_n > 0\,|\,X_0 = i) = 1 - f_n(0)^i &= \big(1 - f_n(0)\big)\big(1 + f_n(0) + f_n(0)^2 + \ldots + f_n(0)^{i-1}\big)\\ &\geq \big(1 - f_n(0)\big)\,i\,f_n(0)^{i-1}\\ &= i\,\mathsf{P}(X_n > 0\,|\,X_0 = 1)\mathsf{P}(X_n = 0\,|\,X_0 = 1)^{i-1}.\end{aligned}$$

Segunda desigualdad:

$$\begin{aligned}\mathsf{P}(X_n > 0 \,|\, X_0 = i) = 1 - f_n(0)^i &= \big(1 - f_n(0)\big)\big(1 + f_n(0) + f_n(0)^2 + \ldots + f_n(0)^{i-1}\big) \\ &\leq \big(1 - f_n(0)\big)\, i \\ &= i\,\mathsf{P}(X_n > 0 \,|\, X_0 = 1).\end{aligned}$$

Cuarta desigualdad:

$$\begin{aligned}\mathsf{P}(X_n > 0 \,|\, X_0 = 1) &= \sum_{j \geq 1} \mathsf{P}(X_n = j \,|\, X_0 = 1) \leq \sum_{j \geq 1} j\,\mathsf{P}(X_n = j \,|\, X_0 = 1) \\ &= \mathsf{E}(X_n \,|\, X_0 = 1) = m^n.\end{aligned}$$

Tercera desigualdad:

$$\mathsf{P}(X_n > 0 \,|\, X_0 = 1)\mathsf{P}(X_n = 0 \,|\, X_0 = 1)^{i-1} \geq \mathsf{P}(X_n > 0 \,|\, X_0 = 1)(1 - m^n)^{i-1}.$$

Para n grande y $m < 1$ obtenemos $(1 - m^n)^{i-1} \approx 1$. Entonces queda por probar el siguiente lema. □

Lema 2.15 Para n grande tenemos la desigualdad $\mathsf{P}(X_n > 0 \,|\, X_0 = 1) \geq \dfrac{1-m}{\sigma^2} m^{n+1}$.

Demostración. Asumimos aquí la condición inicial $X_0 = 1$. Entonces primero,

$$\mathsf{E}^2\big(X_n \mathbb{1}_{\{X_n > 0\}}\big) \leq \mathsf{E}\big(X_n^2\big)\mathsf{E}\big(\mathbb{1}^2_{\{X_n > 0\}}\big) = \mathsf{E}\big(X_n^2\big)\mathsf{P}(X_n > 0).$$

Entonces $\mathsf{P}(X_n > 0) \geq \mathsf{E}^2(X_n)/\mathsf{E}(X_n^2) = m^{2n}/\mathsf{E}(X_n^2)$. La forma del segundo momento $\mathsf{E}(X_n^2)$ através de la función generatriz nos da $\mathsf{E}(X_n^2) = f_n''(1) + \mathsf{E}(X_n) = f_n''(1) + m^n$. Además

$$\begin{aligned}f_n''(1) &= f''(1)\big(f_{n-1}'(1)\big)^2 + f'(1) f_{n-1}''(1) = f''(1) m^{2(n-1)} + m f_{n-1}''(1) \\ &= f''(1)\big(m^{2n-2} + m^{2n-3} + m^{2n-4} + \ldots + m^{n-1}\big) \\ &= \big(\sigma^2 + m^2 - m\big) m^{n-1} \frac{1 - m^n}{1 - m},\end{aligned}$$

luego

$$f_n''(1) + m^n = m^{n-1}\left(\sigma^2 \frac{1 - m^n}{1 - m} + m^{n+1}\right).$$

Entonces

$$\mathsf{P}(X_n > 0 \,|\, X_0 = 1) \geq \frac{m^{2n}}{m^{n-1}\left(\sigma^2 \frac{1-m^n}{1-m} + m^{n+1}\right)} \approx \frac{1-m}{\sigma^2} m^{n+1}. \qquad \square$$

1.7.2 Aplicación a la extinción de las ballenas negras en el Atlántico Norte

Nos inspira [5], [14], [27, Lecture 3] y http://www.rightwhale.ca. La unidad de tiempo representa un año. La ley de reproducción en este modelo es:

$$r_0 = b, \qquad r_1 = (1-b)(1-a), \qquad r_2 = (1-b)a,$$

donde b es la probabilidad que una ballena hembra se muera durante el año siguiente, en el año 1994 esta probabilidad fue estimada en 0,06 [5]; y a es la probabilidad que una ballena engendre una ballena hembra pequeña durante el año siguiente, esta probabilidad fue estimada en 0,038. Entonces,

$$m = r_1 + 2r_2 = (1-b)(1+a) \approx 0,976 < 1: \text{ reproducción subcrítica!}$$

Cálculo de σ^2: $\sigma^2 = r_1 + 4r_2 - m^2 = m(1-m) + 2(1-b)a \approx 0,095$.

En el año 1994, la población de ballenas negras hembras en el Atlántico Norte era de $i \approx 150$. Si sus condiciones de vida no cambian, aplicando el teorema 2.14, esta población no se extinguirá antes de n años con probabilidad superior a 0,99, si

$$i(1-m)\frac{m^{n+1}}{\sigma^2} \geq 0,99 \quad \Longleftrightarrow \quad 150 \cdot 0,024\, \frac{(0,976)^{n+1}}{0,095} \geq 0,99 \quad \Longleftrightarrow \quad n \leq 150,$$

entonces no antes del año 2144.

Al contrario, esta población se extinguirá antes de n años con probabilidad superior a 0,99 si

$$i m^n \leq 0,01 \quad \Longleftrightarrow \quad 150 \cdot (0,976)^n \leq 0,01 \quad \Longleftrightarrow \quad n \geq 395\,,$$

entonces antes del año 2389, con probabilidad superior a 0,99, no habrá más ballenas negras hembras en el Atlántico Norte.

1.7.3 El valor medio del tiempo de extinción T_0 en el caso (sub)crítico

Ya se sabe que en este caso $T_0 < +\infty$ c.s. Nos interesa estimar el valor medio de T_0, en particular para valores iniciales i grandes.

Teorema 2.16

1) En el caso crítico $m = 1$ tenemos para cada $i \geq 1$ que

$$\mathsf{E}(T_0 \,|\, X_0 = i) = +\infty.$$

2) Sin embargo en el caso subcrítico $m < 1$ obtenemos para cada $i \geq 1$ que

$$\mathsf{E}(T_0 \,|\, X_0 = i) < +\infty.$$

Además, si la ley $\mathbf{r}$ satisface $\sum_j j \log(j+1) r_j < +\infty$ tenemos para i grande la asintótica

$$\mathsf{E}(T_0 \,|\, X_0 = i) \approx \frac{\log i}{|\log m|}.$$

Demostración.

$$\mathsf{E}(T_0 \,|\, X_0 = i) = \sum_{n \geq 0} \mathsf{P}(T_0 > n \,|\, X_0 = i) = \sum_n \mathsf{P}(X_n > 0 \,|\, X_0 = i) = \sum_n 1 - f_n(0)^i.$$

Caso crítico: Se puede verificar que para n grande,

$$f_n(0) \approx 1 - \frac{2}{\sigma^2 n}.$$

Luego

$$1 - f_n(0)^i \approx 1 - (1 - \frac{2}{\sigma^2 n})^i \approx \frac{2i}{\sigma^2 n}.$$

La serie $\sum_n {}^1\!/_n$ diverge, entonces $\mathsf{E}(T_0 \,|\, X_0 = i) = +\infty$.
Caso subcrítico: Hemos visto en el Teorema 2.14 que $f_n(0) \geq 1 - m^n$. Entonces

$$\mathsf{E}(T_0 \,|\, X_0 = i) \leq \sum_n 1 - (1 - m^n)^i.$$

Para n grande obtenemos $1-(1-m^n)^i \approx im^n$, además la serie $\sum_n m^n$ converge. Luego $\mathsf{E}(T_0 \,|\, X_0 = i) < +\infty$.

En [14] Theorem 5.4 los autores demostraron (con las notaciones N en lugar de i y τ en lugar de T_0) las desigualdades

$$\left(\frac{\log i - \log\log i}{|\log m|} - 1\right)\left(1 - \frac{1}{ic_1}\right) \leq \mathsf{E}(T_0 \,|\, X_0 = i) \leq \frac{\log i}{|\log m|} + \frac{2-m}{1-m},$$

con $c_1 > 0$. Completemos la prueba del teorema con la observación de que los términos de izquierda y de derecha son equivalentes a $\log i / |\log m|$ para grandes valores de i. □

Ejemplo 2.17

1) Reproducción geométrica crítica:

$$f_n(0) = 1 - \frac{1}{n+1}$$
$$\Longrightarrow \quad \mathsf{E}(T_0 \,|\, X_0 = i) = \sum_n 1 - \left(1 - \frac{1}{n+1}\right)^i \approx \sum_n \frac{i}{n+1} = +\infty$$

2) Tiempo medio de extinción de las ballenas negras:

$$i = 150 \quad \text{grande y} \quad m \approx 0{,}976 < 1$$
$$\Longrightarrow \quad \mathsf{E}(T_0 \,|\, X_0 = i) \approx -\frac{\log 150}{\log 0{,}976} \approx 206 \in [150, 395].$$

1.8 Distribución cuasi-estacionaria de BGW

Antes de extinguirse, a menudo, el proceso subcrítico muestra un comportamiento cuasi-estacionario. Su distribución es invariante sobre la dinámica de reproducción condicionada a la no extinción.

Teorema 2.18 (Existencia y cuasi-estacionaridad de la distribución de Yaglom) Sea $X = (X_n)_n$ un proceso de BGW subcrítico y no degenerado, de ley de reproducción $\mathbf{r}$. Existe una ley de probabilidad sobre $\mathbb{N}^*$, que denotamos $\bar{\mathbf{r}} = (\bar{r}_i)_{i\geq 1}$, llamada *distribución de Yaglom*, definida por

$$\forall i \geq 1, \quad \bar{r}_i = \lim_{n\to+\infty} \mathsf{P}(X_n = i \,|\, X_n > 0);$$

además, si $\bar{f}$ es la función generadora de la ley $\bar{\mathbf{r}}$,

$$\forall s \in [0,1], \quad 1 - \bar{f} \circ f(s) = m\big(1 - \bar{f}(s)\big). \tag{2.2}$$

Esta ley es cuasi-estacionaria, es decir

$$X_0 \sim \bar{\mathbf{r}} \quad \Longrightarrow \quad \forall n \geq 1,\ \mathsf{P}(X_n = \cdot \,|\, X_n > 0) = \bar{\mathbf{r}}.$$

Demostración. En primer lugar, se muestra que existe un límite de la función generadora condicionada $\bar{f}_n$, definida como: $\bar{f}_n(s) := \mathsf{E}(s^{X_n} \,|\, X_n > 0)$.

$$\begin{aligned}\bar{f}_n(s) &= \frac{\mathsf{E}\big(\mathbb{1}_{X_n>0}\, s^{X_n}\big)}{\mathsf{P}(X_n > 0)} = \frac{\mathsf{E}\big((1 - \mathbb{1}_{X_n=0})\, s^{X_n}\big)}{1 - \mathsf{P}(X_n = 0)} = \frac{f_n(s) - f_n(0)}{1 - f_n(0)} \\ &= 1 - \frac{1 - f_n(s)}{1 - f_n(0)} \in [0,1].\end{aligned}$$

Es obvio que $\frac{1-f_n(s)}{1-f_n(0)} \leq 1$. Mostramos ahora que $n \mapsto \frac{1-f_n(s)}{1-f_n(0)}$ es creciente:

$$\frac{1 - f_{n+1}(s)}{1 - f_{n+1}(0)} = \frac{1 - f\big(f_n(s)\big)}{1 - f_n(s)} \cdot \frac{1 - f_n(0)}{1 - f\big(f_n(0)\big)} \cdot \frac{1 - f_n(s)}{1 - f_n(0)} \geq \frac{1 - f_n(s)}{1 - f_n(0)},$$

ya que la función $u \mapsto \frac{1-f(u)}{1-u}$ crece hasta el valor m y entonces tiene un valor más grande para $u = f_n(s)$ que para $u = f_n(0)$. Así, para cada $s \in [0,1]$ existe el límite de la sucesión $n \mapsto \frac{1-f_n(s)}{1-f_n(0)}$. Definimos $\bar{f}(s)$ igual a 1 en el complemento de este límite.

Para mostrar la ecuación (2.2) notamos que

$$1 - \bar{f}(s) = \lim_n \frac{1 - f_{n+1}(s)}{1 - f_{n+1}(0)} = \lim_n \frac{1 - f_n\big(f(s)\big)}{1 - f_n(0)} \cdot \lim_n \frac{1 - f_n(0)}{1 - f\big(f_n(0)\big)},$$

cuyo segundo límite es $1 - \bar{f} \circ f(s)$ y $\lim_n \dfrac{1 - f_n(0)}{1 - f\big(f_n(0)\big)} = \lim_{u \to 1} \dfrac{1-u}{1 - f(u)} = \dfrac{1}{m}$.

Mostramos ahora que $\bar{\mathbf{r}}$ es cuasi-estacionaria: Sea $\mathsf{P}_{\bar{\mathbf{r}}}$ la disribución $\mathsf{P}(\,\cdot\,|\,X_0 \sim \bar{\mathbf{r}})$. Fíjese que

$$\mathsf{P}_{\bar{\mathbf{r}}}(X_1 = 0) = \sum_{i \geq 1} \mathsf{P}(X_1 = 0 \,|\, X_0 = i)\ \bar{r}_i = \sum_{i \geq 1} (r_0)^i\ \bar{r}_i = \bar{f}(r_0),$$

$$\mathsf{E}_{\bar{\mathbf{r}}}\big(\, s^{X_1}\big) = \sum_{i\geq 1} f(s)^i\, \bar{r}_i = \bar{f}\circ f\ (s).$$

Ahora,

$$\mathsf{E}_{\bar{\mathbf{r}}}\big(s^{X_1}\,\big|\,X_1>0\big) = \frac{\mathsf{E}_{\bar{\mathbf{r}}}\big(\mathbb{1}_{X_1>0}\, s^{X_1}\big)}{\mathsf{P}_{\bar{\mathbf{r}}}(X_1>0)} = \frac{\mathsf{E}_{\bar{\mathbf{r}}}\big(\, s^{X_1}\big) - \mathsf{P}_{\bar{\mathbf{r}}}(X_1=0)}{1-\mathsf{P}_{\bar{\mathbf{r}}}(X_1=0)} = \frac{\bar{f}\circ f(s) - \bar{f}(r_0)}{1-\bar{f}(r_0)}.$$

Calculamos $\bar{f}(r_0)$ usando la ecuación (2.2):

$$1-\bar{f}\circ f(0) = m\big(1-\bar{f}(0)\big) \quad\Longrightarrow\quad \bar{f}(r_0) = 1-m(1-0) = 1-m.$$

Entonces

$$\mathsf{E}_{\bar{\mathbf{r}}}\big(s^{X_1}\,\big|\,X_1>0\big) = \frac{\bar{f}\circ f(s)-1+m}{m} = 1-\frac{1-\bar{f}\circ f(s)}{m} =_{(2.2)} \bar{f}(s),$$

es decir $\mathsf{P}(X_1=\cdot\,|\,X_1>0, X_0\sim\bar{\mathbf{r}}) = \bar{\mathbf{r}}$. La prueba para $n>1$ es análoga. □

Ejemplo 2.19

1) Reproducción geométrica $\mathscr{G}_a$ subcrítica: $a > 1/2$.
   Sea $\bar{f}_n(s) := \mathsf{E}\big(s^{X_n}\,\big|\,X_n>0\big)$. Entonces

$$\bar{f}_n(s) = \frac{f_n(s)-f_n(0)}{1-f_n(0)} = \frac{a_n s}{1-(1-a_n)s}$$
$$\Longrightarrow\quad \lim_n \bar{f}_n(s) = \left(1-\frac{1-a}{a}\right)\frac{s}{1-\frac{1-a}{a}s} =: \bar{f}(s)$$

   ya que $\lim_n a_n := 1-{}^{(1-a)}/_a$.
   En este caso, la distribucíon de Yaglom $\bar{\mathbf{r}}$ resulta ser una ley geométrica de parámetro $2-{}^1/_a = 1-m$ sobre $\mathbb{N}^*$ definida por

$$\bar{r}_i = \left(1-\frac{1-a}{a}\right)\left(\frac{1-a}{a}\right)^{i-1}, \quad i\geq 1.$$

2) Reproducción cuasi-geométrica subcrítica, $1-b<a$. (Por ejemplo $a=0,65$ y $b=0,4$).

$$\bar{f}_n(s) = \frac{f_n(s) - f_n(0)}{1 - f_n(0)} = \frac{a_n s}{1 - (1 - a_n)s}$$
$$\Longrightarrow \quad \lim_n \bar{f}_n(s) = \left(1 - \frac{1-a}{b}\right) \frac{s}{1 - \frac{1-a}{b} s} =: \bar{f}(s)$$

ya que $\lim_n a_n := 1 - {}^{(1-a)}/_{b}$.

En este caso, la distribucíon de Yaglom $\bar{\mathbf{r}}$ es una ley geométrica de parámetro ${}^{(a+b-1)}/_{b}$ concentrada sobre $\mathbb{N}^*$ y definida por

$$\bar{r}_i = \left(1 - \frac{1-a}{b}\right)\left(\frac{1-a}{b}\right)^{i-1}, \quad i \geq 1.$$

1.9 El proceso condicionado a la no extinción en el futuro lejano

En lugar de condicionar la dinámica de reproducción a la no extinción en el tiempo presente, se puede condicionar a la no extinción en un tiempo futuro lejano. ¿Como se comporta $\mathsf{P}(X_n = \cdot \,|\, X_{n+k} > 0)$ cuando k es grande?

Teorema 2.20 (Existencia del proceso condicionado) Sea $X = (X_n)_n$ un proceso de BGW subcrítico y no degenerado, de ley de reproducción $\mathbf{r}$ verificando $\sum_j j \log j \, r_j < +\infty$. Existe un proceso de Markov $X^* = (X_n^*)_n$ definido por

$$\mathsf{P}(X_n^* = j) = \lim_{k \to +\infty} \mathsf{P}(X_n = j \,|\, X_{n+k} > 0).$$

Su matriz de transición P^* verífica $(P_{ij}^*)^n := \mathsf{P}(X_n^* = j \,|\, X_0^* = i) = \frac{j}{i m^n} P_{ij}^n$, $\forall i, j \geq 1$.

A veces, este proceso se llama Q-proceso en la literatura.

Demostración.

$$\begin{aligned}
\mathsf{P}(X_n = j \,|\, X_0 = i, X_{n+k} > 0) &= \frac{\mathsf{P}(X_0 = i, X_n = j, X_{n+k} > 0)}{\mathsf{P}(X_{n+k} > 0 \,|\, X_0 = i)\mathsf{P}(X_0 = i)} \\
&= \frac{\mathsf{P}(X_{n+k} > 0 \,|\, X_n = j, X_0 = i)\mathsf{P}(X_n = j \,|\, X_0 = i)\mathsf{P}(X_0 = i)}{\left(1 - f_{n+k}(0)^i\right)\mathsf{P}(X_0 = i)} \\
&= \frac{\mathsf{P}(X_{n+k} > 0 \,|\, X_n = j)\mathsf{P}(X_n = j \,|\, X_0 = i)}{1 - f_{n+k}(0)^i} = \frac{1 - f_k(0)^j}{1 - f_{n+k}(0)^i} P_{ij}^n.
\end{aligned}$$

Kolmogorov demostró la siguiente estimación:

$$\text{Si } \sum_j (j\log j)\, r_j < +\infty, \quad \exists c>0, \mathsf{P}(X_n=0) = f_n(0) \approx_n 1 - c\,m^n.$$

Ver [27] Chapter 3, Theorem 1 por una demostración. (De hecho, en el ejemplo $\mathbf{r} = \mathscr{G}_a$, $f_n(0) = \frac{1-m^n}{1-m^{n+1}} \approx_n 1-(1-m)m^n$.)

Entonces,

$$\frac{1-f_k(0)^j}{1-f_{n+k}(0)^i} \approx_k \frac{1-(1-c\,m^k)^j}{1-(1-c\,m^{n+k})^i} \approx_k \frac{1-(1-jc\,m^k)}{1-(1-ic\,m^{n+k})} = \frac{j}{i\,m^n}.$$

Se puede verificar la propiedad de Markov: $\mathsf{P}(X^*_{n+l} = j \,|\, X^*_0, \ldots, X^*_{l-1}, X^*_l = i) = (P^*_{ij})^n$. □

Ademas, el proceso condicionado muestra un comportamiento estable cuando n es grande.

Proposición 2.21 Sobre las mismas hipótesis que en el Teorema anterior, existe una distribución estacionaria $\mathbf{r}^*$ para X^* igual a la distribución de Yaglom sesgada: $r^*_j = c\,j\,\bar{r}_j$.

Demostración. $\lim_n (P^*_{1j})^n = j \lim_n {}^{P^n_{1j}}/_{m^n}$. Pero

$$\bar{r}_j = \lim_n \mathsf{P}(X_n = j \,|\, X_n > 0) = \lim_n \frac{1}{1-f_n(0)} P^n_{1j} = \lim_n \frac{1}{1-(1-cm^n)} P^n_{1j} = \frac{1}{c} \lim_n \frac{P^n_{1j}}{m^n}$$

$$\Longrightarrow \quad \lim_n (P^*_{1j})^n = c\,j\,\bar{r}_j.$$ □

. . . . . .

2 Unas propiedades del proceso BGW de tiempo continuo

En el modelo de Bienaymé-Galton-Watson de tiempo discreto cada individuo vive una unidad de tiempo, y por lo tanto todos los individuos que coexisten pertenecen a la misma generación. Al contrario, en el proceso de ramificación BGW de tiempo continuo (BGWc) $X = (X_t)_t$ cada uno de esos individuos tiene una duración de vida aleatoria, al término de la cual se muere o se produce un número aleatorio de individuos decendientes. Así coexisten siempre varias generaciónes. Como en el modelo discreto, los

nuevos individuos viven, mueren y se reproducen independientemente con las mismas reglas. Cada individuo inicial genera un árbol aleatorio, como se muestra en la Figura 2.6. La ramificación se refiere a la estructura de árbol. Las ramas representan las líneas de descendencia y su evolución temporal. El proceso estocástico al tiempo t, X_t, es el número de individuos presentes al tiempo t.

Ademas, supongamos que las leyes probabilísticas de vida de cada individuo son las mismas, exponenciales de parámetro α – que notamos $\mathscr{E}xp(\alpha)$; la reproducción de cada individuo tiene la misma ley $\mathbf{r}$ (supongamos que $r_1 = 0$).

2.1 El BGWc como proceso markoviano de salto puro

Definición 2.22 Un proceso $X = (X_t)_{t\geq 0}$ a valores en un espacio $\mathbf{E}$ numerable es un proceso de salto puro si existe una succesión $(Z_n)_n \in \mathbf{E}$ con $Z_{n+1} \neq Z_n$ y una succesión de tiempos aleatorios $\tau_0 = 0 < \tau_1 < \ldots < \tau_n < \ldots$ tal que

$$\forall t > 0, \quad X_t = \sum_{n\geq 0} \mathbb{1}_{\{\tau_n \leq t < \tau_{n+1}\}} Z_n.$$

De esta forma τ_n es el tiempo en el que ocurre el n-ésimo salto, y Z_n es el $(n+1)$-ésimo estado en $\mathbf{E}$ visitado por el proceso X.

Además, X satisface la propriedad de Markov: Para cada $k \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_k < t_{k+1}$ y $x_1, \ldots, x_{k+1} \in \mathbf{E}$ tenemos

$$\mathsf{P}(X_{t_{k+1}} = x_{k+1} \,|\, X_{t_1} = x_1, \ldots, X_{t_k} = x_k) = \mathsf{P}(X_{t_{k+1}} = x_{k+1} \,|\, X_{t_k} = x_k).$$

El proceso es homogéneo si $\mathsf{P}(X_t = \cdot \,|\, X_s) = \mathsf{P}(X_{t-s} = \cdot \,|\, X_0)$.

La propriedad de Markov dice que la probabilidad condicional de un futuro estado en el tiempo t_{k+1}, dado el estado presente en el tiempo t_k y todos los estados pasados, solo depende del presente estado y es independiente del pasado. Al proceso de tiempo discreto $Z_n = X_{\tau_n}$ se llamara el **esqueleto** del proceso de tiempo continuo X. Si X es markoviano, su esquelto $(Z_n)_n$ es una cadena de Markov.

Teorema 2.23 Sea $X = (X_t)_{t\geq 0}$ un proceso BGW de tiempo continuo. Es un proceso markoviano homogéneo de salto puro a valores en $\mathbb{N}$. El proceso esqueleto $(Z_n)_n$ de estados visitados es un BGW y los tiempos aleatorios $(\tau_{n+1} - \tau_n)_{n\geq 0}$ son independientes.

Además, si $Z_n = i$, la distribución del tiempo $\tau_{n+1} - \tau_n$ transcurrido en i es una distribución exponencial $\mathscr{E}xp(i\alpha)$.

Ver la figura 2.6.

Demostración. Por la construcción hecha, el esqueleto $(Z_n)_n$ del proceso X es un BGW. Si Z_0 vale i, es decir que el número de individuos en la población al tiempo iniciál es i. El primero salto de la población ocurre cuando il primero de estos i individuos muere y se reproduce, entonces en el tiempo $\min\{V_1, \ldots, V_i\}$ donde V_k es la duración aleatoria de la vida del k-ésimo individuo. Se sabe que para toda $k \geq 1$, $V_k \sim \mathscr{E}xp(\alpha)$.

Lema 2.24 Sean V_1 y V_2 dos v.a.i. con distribución exponencial con parámetros α_1 y α_2 respectivamente. Entonces $\min\{V_1, V_2\} \sim \mathscr{E}xp(\alpha_1 + \alpha_2)$.

Demostración. Observemos que para $x > 0$

$$\begin{aligned} &\big\{\min\{V_1, V_2\} > x\big\} = \{V_1 > x\} \cap \{V_2 > x\} \\ &\qquad\qquad \Longrightarrow \quad \mathsf{P}\big(\min\{V_1, V_2\} > x\big) = \mathrm{e}^{-\alpha_1 x}\mathrm{e}^{-\alpha_2 x} = \mathrm{e}^{-(\alpha_1+\alpha_2)x}. \end{aligned}$$

Ver también Proposición 3.3.2 en [28]. □

Esto implica que il primero salto del proceso de ramificación BGWc dado $Z_0 = i$, sará en el instante

$$\tau_1 = \min\{V_1, \ldots, V_i\} \sim \mathscr{E}xp(\alpha + \ldots + \alpha) = \mathscr{E}xp(i\alpha).$$

La propriedad de Markov es consecuencia del hecho que $(Z_n)_n$ es una cadena de Markov y de la falta de memoria de la distribución exponencial: si V es una v.a. exponencial con parámetro α, para $x, y > 0$,

$$\begin{aligned} &\frac{\mathsf{P}\big(\{V > x+y\} \cap \{V > x\}\big)}{\mathsf{P}(V > x)} = \frac{\mathrm{e}^{-\alpha(x+y)}}{\mathrm{e}^{-\alpha x}} = \mathrm{e}^{-\alpha y} \\ &\qquad\qquad \Longrightarrow \quad \mathsf{P}(V > x+y \,|\, V > x) = \mathsf{P}(V > y). \end{aligned}$$

Luego,

$$\mathsf{P}(\tau_{n+1} - \tau_n > x \,|\, X_{\tau_n} = i) = \mathrm{e}^{-i\alpha x} \quad \Longrightarrow \quad \mathsf{P}(\tau_{n+1} - \tau_n \in \cdot \,|\, Z_n = i) = \mathscr{E}xp(i\alpha). \quad \square$$

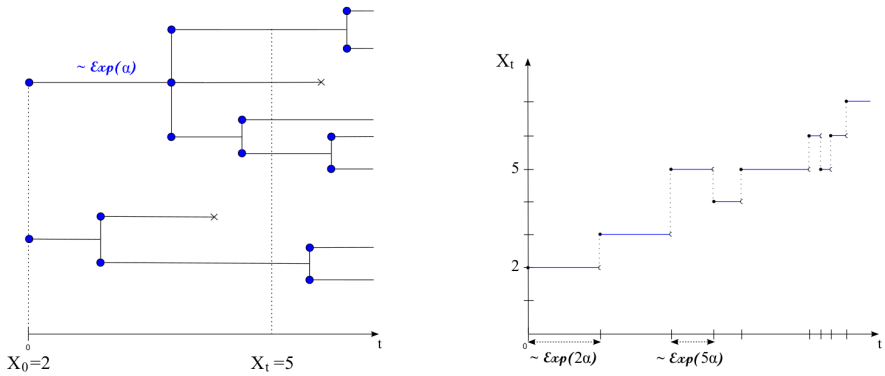


Figura 2.6: El proceso BGWc: La genealogía y el proceso de salto puro (S. Pénisson).

2.2 Las ecuaciones de Chapman-Kolmogorov

A partir de ahora en adelante $P(t)$ denotará la matriz infinita cuyas entradas son las probabilidades de transición del proceso BGWc:

$$P_{ij}(t) = \mathsf{P}(X_t = j \,|\, X_0 = i), \quad i, j \in \mathbb{N}.$$

Teorema 2.25 Sea $X = (X_t)_{t \geq 0}$ un proceso BGW de tiempo continuo y $\big(P(t)\big)_t$ sus matrices de transición. Entonces verifican la propriedad de semigrupo por la ecuación de Chapman-Kolmogorov

$$\forall s, t > 0, \quad P(t+s) = P(s)P(t) = P(t)P(s). \tag{2.3}$$

Demostración. (ver también Teorema 3.4.1 en [28]) Utilizando las propriedades de uno proceso de Markov homogéneo de tiempo continuo tenemos que:

$$\begin{aligned} P_{ij}(t+s) = \mathsf{P}(X_{t+s} = j \,|\, X_0 = i) &= \sum_{k \in \mathbb{N}} \mathsf{P}(X_{t+s} = j, X_t = k \,|\, X_0 = i) \\ &= \sum_{k \in \mathbb{N}} \mathsf{P}(X_{t+s} = j \,|\, X_t = k) \mathsf{P}(X_t = k \,|\, X_0 = i) \\ &= \sum_{k \in \mathbb{N}} P_{kj}(s) P_{ik}(t) = \big(P(t)P(s)\big)_{ij} \end{aligned}$$

□

2.2.1 El generador infinitesimal

Proposición 2.26 Existen las derivadas respecto de t por $t=0$ de la matriz de transición $P(t)$:

1) $\frac{\partial}{\partial t}P_{ii}(t)\big|_{t=0} = -\lim_{t\to 0}\frac{1-P_{ii}(t)}{t} =: Q_{ii}, \quad i\in\mathbb{N}.$
$q_i := -Q_{ii}$ es la tasa (o velocidad) de salida del proceso X del estado i. En el caso del BGWc $q_i = i\alpha$. En particular, $q_0 = 0$ y el estado 0 es absorbente.

2) Para cada $j\neq i\in\mathbb{N}$, $\frac{\partial}{\partial t}P_{ij}(t)\big|_{t=0} = \lim_{t\to 0}\frac{P_{ij}(t)}{t} =: Q_{ij}.$
Q_{ij} es la tasa (o velocidad) de salto del proceso X de un estado i a otro j. En el caso del BGWc

$$Q_{ij} = \begin{cases} 0 & \text{si } j < i-1, \\ -i\alpha & \text{si } j = i, \\ i\alpha\, r_{j-i+1} & \text{si } j \geq i-1, j\neq i. \end{cases}$$

La matriz $\frac{\partial}{\partial t}P(t)|_{t=0} =: Q$, cuyas entradas son $(Q_{ij})_{i,j\in\mathbb{N}}$, se llama generador infinitesimal del proceso markoviano X. Ella verífica para cada $i\in\mathbb{N}$, $\sum_j Q_{ij} = 0$.

Demostración. Podemos hacer ríguroso el cálculo siguiente:

$$1-P_{ii}(t) = \mathsf{P}(X_t\neq i\,|\,X_0=i) \overset{t\to 0}{\approx} \mathsf{P}(\tau_1 < t\,|\,X_0=i) = 1-\mathrm{e}^{-i\alpha t} \overset{t\to 0}{\approx} i\alpha t$$
$$\Longrightarrow \quad \lim_{t\to 0}\frac{1-P_{ii}(t)}{t} = \lim_{t\to 0}\frac{i\alpha t}{t} = i\alpha.$$

Sea $j\neq i$.

$$\begin{aligned} P_{ij}(t) &= \mathsf{P}(X_t=j\,|\,X_0=i) \overset{t\to 0}{\approx} \mathsf{P}(\tau_1<t, Z_1=j\,|\,X_0=i) \\ &= \mathsf{P}(\tau_1<t)\,\mathsf{P}(Z_1=j\,|\,Z_0=i) = \big(1-\mathrm{e}^{-i\alpha t}\big)\,\mathsf{P}\big(Y=j-(i-1)\big) \\ &\Longrightarrow \quad \lim_{t\to 0}\frac{P_{ij}(t)}{t} = \lim_{t\to 0}\frac{1-\mathrm{e}^{-i\alpha t}}{t}\, r_{j-i+1} = i\alpha\, r_{j-i+1}. \end{aligned}$$

y

$$\sum_j Q_{ij} = \sum_j \frac{\partial}{\partial t}P_{ij}(t)\big|_{t=0} = \frac{\partial}{\partial t}\sum_j P_{ij}(t)\big|_{t=0} = \frac{\partial}{\partial t}1 = 0. \qquad \square$$

Forma de la matriz Q, generador infinitesimal de un BGWc general.

$$Q = \alpha \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ r_0 & -1 & r_2 & r_3 & \dots \\ 0 & 2r_0 & -2 & 2r_2 & \dots \\ 0 & 0 & 3r_0 & -3 & \dots \\ \vdots & & & \ddots & \dots \end{pmatrix}$$

En particular para cada función h acotada tenemos

$$Qh(i) = i\alpha \left(\sum_{j \geq i-1, j \neq i} r_{j-i+1} h(j) - h(i) \right) = i\alpha \sum_{k \geq 0} r_k \big(h(i+k-1) - h(i) \big). \tag{2.4}$$

Ejemplo 2.27 Reproducción binária crítica:

$$Q = \alpha \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 1/2 & -1 & 1/2 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & 3/2 & -3 & \dots \\ \vdots & & & \ddots & \dots \end{pmatrix}$$

2.2.2 Las ecuaciones de Kolmogorov hacia adelante y hacia atrás

Teorema 2.28 Se cumplen las siguientes ecuaciones diferenciales que describen la evolución temporal de la ley de probabilidad del proceso.

1) Ecuación de Kolmogorov hacia atrás:

$$\forall t \geq 0 \quad \frac{\partial}{\partial t} P(t) = QP(t) \quad \text{y} \quad P(0) = \mathrm{Id}. \tag{2.5}$$

Además, para cada función h acotada, la función $u(i,t) := \mathsf{E}(h(X_t) \,|\, X_0 = i)$ resuelve la ecuación diferencial

$$\frac{\partial}{\partial t} u(\,\cdot\,,t) = Qu\,(\,\cdot\,,t), \quad u(0) = h. \tag{2.6}$$

2) Ecuación de Kolmogorov hacia adelante:

$$\forall t \geq 0 \quad \frac{\partial}{\partial t} P(t) = P(t)Q \quad \text{y} \quad P(0) = \mathrm{Id}. \tag{2.7}$$

Además, la ley de probabilidad de X_t, notada μ_t, resuelve la ecuación diferencial de Fokker-Planck

$$\frac{\partial}{\partial t} \mu(t) = \mu(t)Q. \tag{2.8}$$

Demostracíon (ver también Teorema 3.4.4 en [28]). Probaremos (2.5) ya que (2.7) se demuestra razonando de manera análoga. Utilizando la ecuación de Chapman-Kolmogorov (2.3):

$$\frac{\partial}{\partial t} P(t) = \lim_{h \to 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \to 0} \frac{P(h)P(t) - P(t)}{h} = QP(t).$$

Por otro lado,

$$\begin{aligned} \frac{\partial}{\partial t} \mu_j(t) = \frac{\partial}{\partial t} \sum_i \mu_i(0) P_{ij}(t) &= \sum_i \mu_i(0) \frac{\partial}{\partial t} P_{ij}(t) \overset{(2.7)}{=} \sum_i \mu_i(0) \sum_k P_{ik}(t) Q_{kj} \\ &= \sum_k \mu_k(t) Q_{kj} = \big(\mu(t)Q\big)_j \end{aligned}$$

□

Corolario 2.29 La matriz de transición $P(t)$ del proceso BGWc resuelve las ecuaciónes diferenciales de Kolmogorov:

$$\forall i, j \in \mathbb{N}, \quad \frac{\partial}{\partial t} P_{ij}(t) = i\alpha r_0 P_{i-1,j}(t) - i\alpha P_{ij}(t) + i\alpha \sum_{k \geq 1} r_{k+1} P_{i+k,j}(t), \quad P_{ij}(0) = \delta_{ij}. \tag{2.9}$$

Por ejemplo, si la reproducción es binaria crítica ($r_0 = r_2 = 1/2$) tenemos

$$\forall i, j \in \mathbb{N}, \quad \frac{\partial}{\partial t} P_{ij}(t) = \frac{i\alpha}{2} \big(P_{i+1,j}(t) + P_{i-1,j}(t) - 2P_{ij}(t)\big).$$

2.3 La ley del proceso al tiempo t

La ley del proceso BGWc X_t al tiempo t se define por su función generatriz:

$$F(s,t) := \mathsf{E}\big(s^{X_t} \,\big|\, X_0 = 1\big) = \sum_{j \geq 0} \mathsf{P}(X_t = j \,|\, X_0 = 1)\, s^j = \sum_{j \geq 0} P_{1j}(t)\, s^j.$$

La serie converge uniformemente con t, al menos para $|s| < 1$ (Ver [28], Apéndice B).

2.3.1 La función generatriz del proceso al tiempo t

Proposición 2.30 (Estructura diferencial de la función generatriz de BGWc) Para todos $t > 0$ la función generatriz $F(\cdot,t)$ verifica la siguiente ecuación diferencial:

$$\frac{\partial}{\partial t}F(s,t) = \tilde{f}\big(F(s,t)\big), \quad F(s,0) = s, \qquad 0 \le s \le 1, \tag{2.10}$$

con $\tilde{f}(s) := \alpha\big(f(s) - s\big)$.

Demostración. De la ecuación (2.6) por $u(i,t) = \mathsf{E}\big(s^{X_t} \,\big|\, X_0 = i\big)$ $(h(j) = s^j)$, es fácil ver que

$$\frac{\partial}{\partial t}F(s,t) = \frac{\partial}{\partial t}u(1,t) = \big(Qu(\cdot,t)\big)_1 = \sum_{j\ge 0} Q_{1j}u(j,t) = \sum_{j\ge 0} Q_{1j}u(1,t)^j.$$

Utilizando la Proposición 2.26, $Q_{1j} = \alpha r_j$, $j \neq 1$ y $Q_{11} = -\alpha$. Entonces

$$\frac{\partial}{\partial t}F(s,t) = \alpha\left(\sum_{j\ge 0} r_j F(s,t)^j - F(s,t)\right) = \tilde{f}\big(F(s,t)\big). \qquad \square$$

Ejemplo 2.31

1) Reproducción binaria crítica, $m = 1$: $\tilde{f}(s) = \frac{\alpha}{2}(1-s)^2$.
   F verifica $\dfrac{\partial}{\partial t}F(s,t) = \dfrac{\alpha}{2}\big(1 - F(s,t)\big)^2$, $F(s,0) = s$.
   Entonces $F(s,t) = 1 - \dfrac{1-s}{1 + \frac{\alpha}{2}t(1-s)}$.

2) Reproducción binaria no crítica:
   $\tilde{f}(s) = \alpha\big(r_2(1-s)^2 + (1-2r_2)(1-s)\big) = \alpha(1-s)\big(\frac{m}{2}(1-s) + 1 - m\big)$
   Entonces $F(s,t) = 1 - \dfrac{1-s}{\mathrm{e}^{-\alpha(m-1)t} + \frac{m}{2(m-1)}\big(1 - \mathrm{e}^{-\alpha(m-1)t}\big)(1-s)}$.

3) **(W1)** $\tilde{f}(z) = \frac{\alpha}{3}(1-z)^2$ y $F(s,t) = 1 - \dfrac{1-s}{1 + \frac{\alpha}{3}t(1-s)}$.

(W2) $\tilde{f}(z) = \alpha\Big(\big(\frac{3+z}{4}\big)^5 - z\Big)$ y F verifica

$$\frac{\partial}{\partial t}F(s,t) = \alpha\left(\left(\frac{3+F(s,t)}{4}\right)^5 - F(s,t)\right), \qquad F(s,0) = s.$$

4) Reproducción crítica de tipo γ-estable: $\tilde{f}(s) = 1 - s + c(1-s)^\gamma$ con $\gamma \in]1,2]$ y $c \geq 0$.
Entonces $F(s,t) = 1 - \dfrac{1-s}{\big(1 + c(\gamma-1)t(1-s)^{\gamma-1}\big)^{1/\gamma-1}}$.

2.3.2 Los primeros momentos

Se sabe que el valor esperado de la variable X_t, si existe, está dado por $\mathsf{E}(X_t) = \frac{\partial}{\partial s}F(s,t)\big|_{s=1}$. Calculando la derivada respecto de s en la ecuación (2.10), concluimos que $m(t) := \mathsf{E}(X_t)$ es solución de la ecuación

$$\frac{\partial}{\partial t}m(t) = \frac{\partial}{\partial s}\tilde{f}\big(F(s,t)\big)\Big|_{s=1} = \tilde{f}'\big(F(1,t)\big)m(t) = \tilde{f}'(1)\, m(t) = \alpha(m-1)\, m(t)$$
$$\Longrightarrow \quad \mathsf{E}(X_t) = \mathrm{e}^{\alpha(m-1)t}\, \mathsf{E}(X_0).$$

La constante $\rho := \alpha(m-1)$ se llama *parámetro de Malthus*.

Así, de manera semejante al modelo discreto BGW asociado, se encuentran tres casos distintos:

$$\begin{cases} \rho < 0 & \Rightarrow \quad (X_t)_t \text{ es } \textbf{subcrítico} \text{ y } \lim_{t\to+\infty}\mathsf{E}(X_t) = 0; \\ \rho = 0 & \Rightarrow \quad (X_t)_t \text{ es } \textbf{crítico} \text{ y } \mathsf{E}(X_t) \equiv \mathsf{E}(X_0); \\ \rho > 0 & \Rightarrow \quad (X_t)_t \text{ es } \textbf{supercrítico} \text{ y } \lim_{t\to+\infty}\mathsf{E}(X_t) = +\infty. \end{cases}$$

Además, cuando $X_0 = 1$, $\mathsf{E}(X_t^2 - X_t) = \frac{\partial^2}{\partial s^2}F(s,t)\big|_{s=1}$, y calculando la segunda derivada respecto de s en la ecuación (2.10), concluimos que $v(t) def eq \mathsf{E}(X_t^2 \,|\, X_0 = 1)$ es solución de la ecuación

$$\begin{cases} \dfrac{\partial}{\partial t}v(t) = \tilde{f}''\big(F(1,t)\big)\, m^2(t) + \rho v(t) = (\alpha\sigma^2 + m\rho)\mathrm{e}^{2\rho t} + \rho v(t) \\ v(0) = 1 \end{cases}$$

Entonces,

$$\mathsf{Var}(X_t \,|\, X_0 = 1) = \begin{cases} \big(\alpha\sigma^2 + \rho(m-1)\big)\,\frac{\mathrm{e}^{2\rho t} - \mathrm{e}^{\rho t}}{\rho} & \text{si } \rho \neq 0, \\ \alpha\sigma^2 t & \text{si } \rho = 0. \end{cases}$$

2.4 El *problema* de martingalas

Proposición 2.32 (Una familia de martingalas) Para cada función h acotada, el proceso $\left(h(X_t) - h(X_0) - \int_0^t \big(Qh\big)(X_\tau)\,\mathrm{d}\tau\right)_{t>0}$ es una martingala respecto a la filtración $(\mathscr{F}_t)_t$, donde $\mathscr{F}_t := \sigma(X_s, 0 \leq s \leq t)$.

Demostración. (Ver [7], Chapter 9-1) Se debe mostrar que, si

$$Z_t := h(X_t) - h(X_0) - \int_0^t \big(Qh\big)(X_\tau)\,\mathrm{d}\tau$$

y $s < t$, $\mathsf{E}(Z_t \,|\, \mathscr{F}_s) = Z_s$. Sin embargo

$$\begin{aligned} \mathsf{E}(Z_t \,|\, \mathscr{F}_s) &= \mathsf{E}\big(h(X_t) \,\big|\, \mathscr{F}_s\big) - h(X_0) - \int_0^s \mathsf{E}\big(Qh(X_\tau) \,\big|\, \mathscr{F}_s\big)\,\mathrm{d}\tau - \int_s^t \mathsf{E}\big(Qh(X_\tau) \,\big|\, \mathscr{F}_s\big)\,\mathrm{d}\tau \\ &= \big(P(t-s)h\big)(X_s) - h(X_0) - \int_0^s Qh(X_\tau)\,\mathrm{d}\tau - \int_s^t \big(P(\tau - s)Q\big)(h)(X_s)\,\mathrm{d}\tau \\ &= \left(P(t-s) - \int_0^{t-s} P(\tau)Q\,\mathrm{d}\tau\right)(h)(X_s) - h(X_0) - \int_0^s Qh(X_\tau)\,\mathrm{d}\tau \\ &\overset{(2.7)}{=} h(X_s) - h(X_0) - \int_0^s Qh(X_\tau)\,\mathrm{d}\tau. \end{aligned}$$

□

Comentario 2.33

1) Ademas, esta familia de martingalas caracteriza el proceso $(X_t)_t$ como un proceso markoviano asociado al generador infinitesimal Q.

2) Existe una version un poco mas general: si $h : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{R}$ es acotada y diferenciable en t, $\left(h(t, X_t) - h(0, X_0) - \int_0^t \left(\frac{\partial h}{\partial t}(\tau, X_\tau) + Qh(\tau, X_\tau)\right)\mathrm{d}\tau\right)_{t>0}$ es una martingala. Así se puede mostrar que $\mathrm{e}^{-\rho t} X_t$ es una martingala: se toma $h(t,i) := \mathrm{e}^{-\rho t} i$.

De un lado $\frac{\partial h}{\partial t} = -\rho h$. Pero, de otro lado,

$$Qh(t,i) = \mathrm{e}^{-\rho t}\left(i\alpha \sum_{j\geq i-1, j\neq i} r_{j-i+1} j - i^2\alpha\right) = \mathrm{e}^{-\rho t} i\alpha(m-1) = \rho h(i).$$

Entonces $\frac{\partial h}{\partial t} + Qh \equiv 0$.

2.5 Tiempo de extinción y probabilidad de extinción

Asumamos en esta sección que $X_0 = 1$. La probabilidad $e(t)$ que la población sea ya extinguida en el tiempo t es:

$$e(t) := \mathsf{P}(X_t = 0) = F(0,t).$$

Proposición 2.34 (Relación implícita de la probabilidad de extinción) Sea $e(t), t > 0$, la función de distribución del tiempo aleatorio de extinción T_0, $e(t) := \mathsf{P}(T_0 \leq t)$. En otras palabras, $e(t)$ es la probabilidad que la población sea ya extinguida en el tiempo t. La función $e(\cdot)$ verifica la relación implícita: $\displaystyle\int_0^{e(t)} \frac{1}{\tilde{f}(s)}\,\mathrm{d}s = t$, $t > 0$.

Demostración. Dada la definición de la función generatriz F

$$e(t) = \mathsf{P}(X_t = 0) = F(0,t), \quad e(0) = 0 \quad \text{y} \quad \lim_{t\to+\infty} e(t) = s_0.$$

Ya que F verifica la ecuación (2.10), $\displaystyle\frac{\partial}{\partial t} e(t) = \tilde{f}\big(e(t)\big)$, $e(0) = 0$. Sea Φ una acumulada de la función $\frac{1}{\tilde{f}}$: $\displaystyle\Phi(t) = \int_0^t \frac{1}{\tilde{f}(s)}\,\mathrm{d}s$. Es bien definida por $t < s_0$, porque $f(s) > s \Rightarrow \tilde{f}(s) \neq 0$.

$$\frac{\partial}{\partial t}\Big(\Phi\big(e(t)\big) - t\Big) = \Phi'\big(e(t)\big)\frac{\partial}{\partial t} e(t) - 1 = \frac{1}{\tilde{f}\big(e(t)\big)}\frac{\partial}{\partial t} e(t) - 1 = 0.$$

Entonces la función $t \mapsto \Phi\big(e(t)\big) - t$ es constante y $\Phi\big(e(t)\big) - t \equiv \Phi\big(e(0)\big) - 0 = 0$. $\square$

Ejemplo 2.35

1) Reproducción binaria crítica, $r_0 = r_2 = \frac{1}{2}, \rho = 0$:

$$F(s,t) = 1 - \frac{1-s}{1+\frac{\alpha}{2}t(1-s)} \quad \Longrightarrow \quad e(t) = 1 - \frac{1}{1+\frac{\alpha}{2}t} \overset{t \text{ grande}}{\approx} 1 - \frac{2}{\alpha t}.$$

En particular, análogamente a lo establecido en la sección 1.7.31.7.3, el valor medio del tiempo de extinción es infinito:

$$\mathsf{E}(T_0) = \int_0^{+\infty} \mathsf{P}(T_0 > t)\,\mathrm{d}t = \int_0^{+\infty} \big(1 - e(t)\big)\,\mathrm{d}t \approx \int^{+\infty} \frac{2}{\alpha t}\,\mathrm{d}t = +\infty.$$

2) Reproducción binaria no crítica:

$$F(s,t) = 1 - \frac{1-s}{\mathrm{e}^{-\rho t} + \frac{m}{2(m-1)}(1-\mathrm{e}^{-\rho t})(1-s)}$$

$$\Longrightarrow \quad e(t) = 1 - \frac{2(m-1)}{(m-2)\mathrm{e}^{-\rho t} + m}$$

a) Caso subcrítico, $\rho < 0$: $\lim_{t\to+\infty} e(t) = 1$, es decir que la problación se extingue casi seguro. Ahora, el valor medio del tiempo de extinción es finito:

$$\mathsf{E}(T_0) = \int_0^{+\infty} \big(1 - e(t)\big)\,\mathrm{d}t \approx \int^{+\infty} \frac{2(m-1)}{m-2}\mathrm{e}^{\rho t}\,\mathrm{d}t < +\infty.$$

b) Caso supercrítico, $\rho > 0$: $\lim_{t\to+\infty} e(t) = \frac{2}{m} - 1 = q$;
la problación sobrevive con una probabilidad igual a $\frac{2(m-1)}{m}$.

3) Reproducción crítica de tipo γ-estable:

$$F(s,t) = 1 - \frac{1-s}{\big(1 + c(\gamma-1)t(1-s)^{\gamma-1}\big)^{1/\gamma-1}}$$

$$\Longrightarrow \quad e(t) = 1 - \frac{1}{\big(1 + c(\gamma-1)t\big)^{1/\gamma-1}},$$

$$\mathsf{E}(T_0) \approx \left(\frac{1}{c(\gamma-1)}\right)^{1/\gamma-1} \int^{+\infty} \frac{1}{t^{1/(\gamma-1)}}\,\mathrm{d}t < +\infty$$

dado que $1/(\gamma-1) > 1$. Entonces este proceso se extingue mas rápidamente que el binario crítico.

2.6 El proceso BGWc condicionado

Nos interesamos a la situación del BGWc (sub)crítico antes de su c.s. extinción. Como en el caso discreto, se puede condicionar la dinámica a la no extinción en un futuro lejano. Así, para cada tiempo $t > 0$ y cada suceso $B \in \mathscr{F}_t$ que depende de los valores del proceso hasta el tiempo t, mostramos que $\mathsf{P}(B \,|\, X_0 = i, X_{t+\theta} > 0)$ converge cuando θ es grande.

Teorema 2.36 (Existencia del proceso condicionado a la no extinción en un tiempo futuro cuasi-infinito.) Sea $X = (X_t)_t$ un proceso de BGWc subcrítico y no degenerado, de ley de reproducción $\mathbf{r}$ con momento segundo finito. Existe un proceso de Markov $X^* = (X_t^*)_t$ definido por

$$\mathsf{P}(X_t^* = \cdot \,|\, X_0^* = i) = \lim_{\theta \to +\infty} \mathsf{P}(X_t = \cdot \,|\, X_0 = i, X_{t+\theta} > 0).$$

Ademas, su ley tiene una densidad sobre $\mathscr{F}_t$ con respecto a la ley de X:

$$\mathsf{P}(X_{t_1}^* = i_1, \ldots, X_{t_k}^* = i_k \,|\, X_0^* = i) = \mathsf{E}\left(\mathrm{e}^{-\rho t} \frac{X_t}{i} \mathbb{1}_{\{X_{t_1} = i_1, \ldots, X_{t_k} = i_k\}} \,\middle|\, X_0 = i \right), \qquad t_1 < \ldots < t_k < t \quad (2.11)$$

Demostración.

$$\begin{aligned}
\mathsf{P}(X_t = j \,|\, X_0 = i, X_{t+\theta} > 0) &= \frac{\mathsf{P}(X_t = j, X_{t+\theta} > 0 \,|\, X_0 = i)}{\mathsf{P}(X_{t+\theta} > 0 \,|\, X_0 = i)} \\
&= \frac{\mathsf{P}(X_{t+\theta} > 0 \,|\, X_t = j)\mathsf{P}(X_t = j \,|\, X_0 = i)}{1 - \mathsf{P}(X_{t+\theta} = 0 \,|\, X_0 = i)} \\
&= \frac{\big(1 - \mathsf{P}(X_\theta > 0)^j\big)\mathsf{P}(X_t = j \,|\, X_0 = i)}{1 - F(0, t+\theta)^i} \\
&= \frac{1 - F(0,\theta)^j}{1 - F(0, t+\theta)^i} \mathsf{P}(X_t = j \,|\, X_0 = i).
\end{aligned}$$

De manera semejante al caso tiempo discreto, se busca una estimación de $F(0,t)$ para t grande.

$$\sum_j j^2 r_j < +\infty \quad \Longrightarrow \quad \exists c > 0,\ \mathsf{P}(X_t = 0) \approx_t 1 - c \exp(\rho t).$$

(Ver [15] p. 109.) Entonces,

$$\frac{1-F(0,\theta)^j}{1-F(0,t+\theta)^i} \approx_\theta \frac{1-\left(1-cje^{\rho\theta}\right)}{1-\left(1-cie^{\rho(t+\theta)}\right)} = e^{-\rho t}\frac{j}{i}$$

y

$$\lim_{\theta\to+\infty} \mathsf{P}(X_t = j \,|\, X_0 = i, X_{t+\theta} > 0) = e^{-\rho t}\frac{j}{i}\mathsf{P}(X_t = j \,|\, X_0 = i).$$

Para verificar (2.11) es simple de mostrar que, para $t_1 < \ldots < t_k < t$,

$$\mathsf{P}(X^*_{t_1} = i_1, \ldots, X^*_{t_k} = i_k \,|\, X^*_0 = i) = \mathsf{E}\left(e^{-\rho t_k}\frac{X_{t_k}}{i}\mathbb{I}_{\{X_{t_1}=i_1\}}\cdots\mathbb{I}_{\{X_{t_k}=i_k\}}\middle| X_0 = i\right)$$

y de utilizar la propiedad de martingala de $e^{-\rho t}X_t$. □

• •

3 Limites de procesos de BGW cuando la población es numerosa

Se muestran varias ventajas del estudio del sistema límite.

- ◇ Primero, a menudo la estructura del límite es más sencilla y se puede interpretar mejor el comportamiento qualitativo.
- ◇ Muchas veces el número de parámetros que determinan el proceso límite se reduce en relación al proceso original.

Cuando una población es muy numerosa, en lugar de considerar el número de individuos conviene estudiar la *densidad* de individuos. Este enfoque, que se usa en física estadística, se emplea para obtener los procesos con estado continuo.

3.1 Limite determinista

Consideremos el proceso de Bienaymé-Galton-Watson, pero en lugar de uno solo individuo inicial, ahora habrá muchos individuos iniciales, $X_0^{(N)}$, con $X_0^{(N)} \approx Nx$, $x \in \mathbb{R}^+$.

Antes supusimos implícitamente que cada individuo tiene una masa unitaria, y ahora los individuos tendrán masas muy pequeñas $1/N$.

3.1.1 BGW de tiempo discreto renormalizado

Sea $\left(\check{X}_n^{(N)}\right)_n := \left(\frac{1}{N}X_n^{(N)}\right)_n$ el nuevo proceso de densidad de poblacíon. El proceso $\check{X}^{(N)}$ toma sus valores en $\frac{1}{N}\mathbb{N} \subset \mathbb{R}^+$.

Suponemos que los primero y segundo momentos de la ley de reproducción existen y estudiamos en primer lugar el comportamiento de los dos primeros momentos de $\check{X}_n^{(N)}$. Según (2.1) tenemos

$$\begin{aligned}
\mathsf{E}\left(\check{X}_n^{(N)}\right) &= \frac{1}{N}\mathsf{E}\left(X_n^{(N)}\right) = \frac{X_0^{(N)}}{N} m^n \overset{N\to+\infty}{\longrightarrow} x m^n \\
\mathsf{Var}\left(\check{X}_n^{(N)} \,\middle|\, X_0^{(N)}\right) &= \frac{1}{N^2}\mathsf{Var}\left(X_n^{(N)} \,\middle|\, X_0^{(N)}\right) = \frac{X_0^{(N)}}{N^2}\mathsf{Var}(X_n \,|\, X_0 = 1) \\
&\approx_N \frac{x}{N}\,\mathsf{Var}(X_n \,|\, X_0 = 1) \overset{N\to+\infty}{\longrightarrow} 0
\end{aligned}$$

Así, el límite de $\check{X}_n^{(N)}$ para N grande parece ser determinista, entonces igual a su esperanza, y tiene un desarollo en el tiempo dado por $x m^n$. No es muy interesante.

3.1.2 BGWc renormalizado

Sea $\left(\check{X}_t^{(N)}\right)_{t\geq 0} := \left(\frac{1}{N}X_t^{(N)}\right)_{t\geq 0} \in \frac{1}{N}\mathbb{N} \subset \mathbb{R}^+$ el nuevo proceso de densidad de poblacíon.

Se puede estudiar el comportamiento en N de los primeros momentos, como en el caso discreto. Proponemos este enunciado como *ejercicio*.

Se puede también estudiar el comportamiento de la transformada de Laplace de la variable aleatoria renormalizada $\check{X}_t^{(N)}$, definida por

$$\begin{aligned}
\check{L}_t^N(\lambda) &:= \mathsf{E}\Big(\exp\left(-\lambda \check{X}_t^{(N)}\right)\Big) \\
&= \mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X_t^{(N)}}\right) = \left(\mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X_t} \,\middle|\, X_0 = 1\right)\right)^{X_0^{(N)}} = \left(F\left(\mathrm{e}^{-\lambda/N}, t\right)\right)^{X_0^{(N)}}
\end{aligned}$$

Porque

$$F\left(\mathrm{e}^{-\lambda/N},t\right) = \mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X_t}\right) \approx_N \mathsf{E}\left(\left(1-\frac{\lambda}{N}\right)^{X_t}\right) \approx_N \mathsf{E}\left(1-\frac{\lambda X_t}{N}\right) \approx_N 1-\frac{\lambda\mathsf{E}(X_t)}{N}$$
$$= 1-\frac{\lambda\mathrm{e}^{\rho t}}{N},$$

$$X_0^{(N)} \approx_N Nx \quad \Rightarrow \quad \breve{L}_t^N(\lambda) \approx_N \left(1-\frac{\lambda\mathrm{e}^{\rho t}}{N}\right)^{Nx} \quad \Rightarrow \quad \lim_{N\to+\infty} \breve{L}_t^N(\lambda) = \exp\left(-\lambda x\mathrm{e}^{\rho t}\right),$$

que se interprete como la transformada de Laplace de la función determinista $\breve{X}_t^{(\infty)} \equiv x\mathrm{e}^{\rho t}$. Es solución de la ecuación diferencial de Malthus:

$$\frac{\partial}{\partial t}\breve{X}_t^{(\infty)} = \rho\,\breve{X}_t^{(\infty)}, \qquad \breve{X}_0^{(\infty)} = x. \tag{2.12}$$

Si el parámetro de Malthus ρ es positivo, $\breve{X}_t^{(\infty)}$ explode asintóticamente, si $\rho = 0$, $\breve{X}_t^{(\infty)}$ es constante, y si $\rho < 0$, $\breve{X}_t^{(\infty)}$ converge a 0 por t grande.

3.2 Limite difusión

Para evitar un límite determinista, se toma muchos individuos iniciales, cada uno de masa pequeña, pero se acelera además el tiempo de modo que en una unidad de tiempo ocurrirán muchas ramificaciones. También se hace que las leyes de ramificación dependen de N (tienen primero y segundo momento finidos, notados $m^{(N)}$ y $(\sigma^{(N)})^2$).

Ver [9] cuando la leye de ramificación no depende de N, [14] Capitulo 4 para explicaciónes heurísticas y [19] para una descripción completa de las condiciones suficientes y necesarias para la convergencia de BGW renormalizados hacia procesos de ramificación muy generales con estado continuo.

3.2.1 BGW de tiempo discreto renormalizado

Para cada N definimos el proceso estocástico en tiempo continuo

$$\tilde{X}_t^{(N)} := \frac{1}{N}X_{[Nt]}^{(N)} \qquad \text{con} \quad X_0^{(N)} \approx_N Nx, \qquad ([y] := \sup\{n\in\mathbb{N}, n\le y\}).$$

Así, las ramificaciones ocurren en los instantes i/N, $i = 1,2,\ldots$, y cada individuo tiene masa $1/N$.

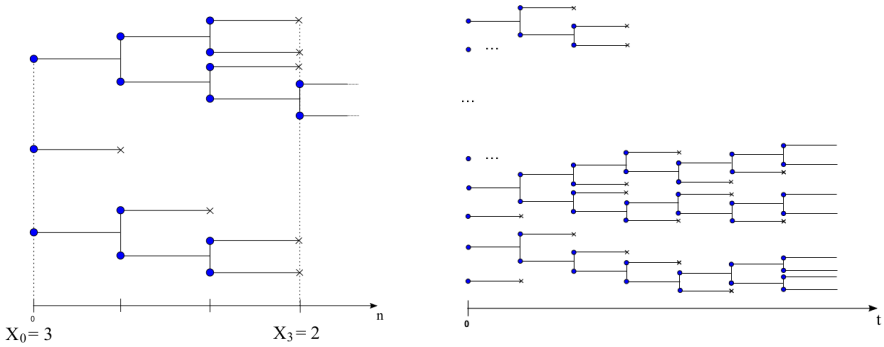


Figura 2.7: Unas realizaciones del BGW binario y binario renormalizado (S. Pénisson).

Ejemplo 2.37 Caso geométrica no crítico con parámetro $a^{(N)}$:
Como antes, se estudia la convergencia de la transformada de Laplace del proceso renormalizado.

$$\tilde{L}_t^N(\lambda) := \mathsf{E}\Big(\exp\big(-\lambda\tilde{X}_t^{(N)}\big)\Big) = \mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X_{[Nt]}^{(N)}}\right)$$
$$= \left(\mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X_{[Nt]}^{(N)}}\,\middle|\, X_0^{(N)} = 1\right)\right)^{X_0^{(N)}} = \left(f_{[Nt]}^{(N)}\big(\mathrm{e}^{-\lambda/N}\big)\right)^{X_0^{(N)}}.$$

Se sabe que, en el caso no crítico,

$$f_{[Nt]}^{(N)}(s) = \frac{(m^{(N)})^{[Nt]}(1-s) + m^{(N)}s - 1}{(m^{(N)})^{[Nt]+1}(1-s) + m^{(N)}s - 1}$$

donde $m^{(N)} = {}^{(1-a^{(N)})}/_{a^{(N)}}$. Verifiquen como *ejercicio* que cuando $m^{(N)} = 1 + \frac{b}{N}$ la función $\tilde{L}_t^N(\lambda)$ converge. Se identifica la forma de la transformada de Laplace en el límite.

Volvemos al caso general. Aquí también estudiamos en primer lugar el comportamiento del primero momento:

$$\mathsf{E}\big(\tilde{X}_t^{(N)}\big) = \frac{X_0^{(N)}}{N}(m^{(N)})^{[Nt]} \approx_{N\to+\infty} x\,(m^{(N)})^{[Nt]}.$$

Una condición necesaria para la convergencia de $(m^{(N)})^{[Nt]}$ es que $\mathbf{r}^{(N)}$ sea asintóticamente crítica: $m^{(N)} \overset{N\to+\infty}{\approx} 1+\dfrac{b}{N}$, $b\in\mathbb{R}$. En este caso $\lim\limits_{N\to+\infty} \mathsf{E}\big(\tilde{X}_t^{(N)}\big) = x\mathrm{e}^{bt}$.

Analizando el segundo momento:

$$\mathsf{Var}\big(\tilde{X}_t^{(N)}\big) = \frac{1}{N^2}\mathsf{Var}\big(X_{[Nt]}^{(N)}\big) = \frac{X_0^{(N)}}{N^2}\mathsf{Var}\big(X_{[Nt]}^{(N)}\,\big|\,X_0^{(N)}=1\big) \approx_N \frac{x}{N}\mathsf{Var}\big(X_{[Nt]}^{(N)}\,\big|\,X_0^{(N)}=1\big).$$

Sin embargo, si $b\neq 0$,

$$\begin{aligned}\mathsf{Var}\big(X_{[Nt]}^{(N)}\,\big|\,X_0^{(N)}=1\big) &= \big(\sigma^{(N)}\big)^2\,(m^{(N)})^{[Nt]-1}\,\frac{(m^{(N)})^{[Nt]}-1}{m^{(N)}-1}\\ &\approx_N \big(\sigma^{(N)}\big)^2\left(1+\frac{b}{N}\right)^{[Nt]}\frac{\big(1+\frac{b}{N}\big)^{[Nt]}-1}{\big(1+\frac{b}{N}\big)-1} \approx_N N\,\big(\sigma^{(N)}\big)^2\,\mathrm{e}^{bt}\,\frac{\mathrm{e}^{bt}-1}{b}.\end{aligned}$$

Así, si $\lim_N \sigma^{(N)} =: \sigma$, $\lim\limits_{N\to+\infty}\mathsf{Var}\big(\tilde{X}_t^{(N)}\big) = x\sigma^2\mathrm{e}^{bt}\,\dfrac{\mathrm{e}^{bt}-1}{b}$. Si $b=0$ (que es equivalente a que $\lim_N N(m^{(N)}-1)=0$) obtenemos que

$$\mathsf{Var}\big(\tilde{X}_t^{(N)}\big) \approx_N \frac{1}{N}x\big(\sigma^{(N)}\big)^2\,[Nt] \to_N x\,\sigma^2\,t.$$

Teorema 2.38 (Construcción del proceso límite via su transformada de Laplace) Si las tres condiciónes

$$\text{i) } m^{(N)} \approx_N 1+{}^{b}\!/_{N}, \qquad \text{ii) } \lim_N \sigma^{(N)} = \sigma \text{ y} \qquad \text{iii) } \lim_N \textstyle\sum_{j\geq tN} j^2 r_j^{(N)} = 0$$

son realizadas, y si $\lim_N \tilde{X}_0^{(N)} = x$, entonces por cada $t>0$, $\tilde{X}_t^{(N)}$ converge hacia la variable aleatoria $\tilde{X}_t^{\infty}$ definido por su transformada de Laplace

$$\tilde{L}_t(\lambda) := \mathsf{E}\Big(\exp\big(-\lambda\tilde{X}_t^{\infty}\big)\,\Big|\,\tilde{X}_0^{\infty}=x\Big) = \exp\big(-xu_t(\lambda)\big)$$

donde la función u_t verifica (2.15).

Demostración. (Ver [9], Appendix II o [13] Theorem 3.2)

$$\tilde{L}_t^N(\lambda) := \mathsf{E}\Big(\exp\big(-\lambda\tilde{X}_t^{(N)}\big)\Big) = \mathsf{E}\left(\big(\mathrm{e}^{-\lambda/N}\big)^{X_{[Nt]}^{(N)}}\right)$$

$$= \left(\mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X^{(N)}_{[Nt]}} \middle| X^{(N)}_0 = 1 \right) \right)^{X^{(N)}_0} = \left(f_{[Nt]}\left(\mathrm{e}^{-\lambda/N}\right) \right)^{X^{(N)}_0} . \quad (2.13)$$

Como

$$\begin{aligned} f\left(\mathrm{e}^{-\lambda/N}\right) &\approx_N 1 + \left(\mathrm{e}^{-\lambda/N} - 1\right)\left(1 + \frac{b}{N}\right) \\ &\quad + \frac{1}{2}\left(\mathrm{e}^{-\lambda/N} - 1\right)^2 \left(\left(\sigma^{(N)}\right)^2 - m^{(N)} + \left(m^{(N)}\right)^2 \right) + O(1/N^3) \\ &\approx_N 1 - \frac{\lambda}{N} - \left(b\lambda - \frac{1}{2}(\sigma^2+1)\lambda^2 \right) \frac{1}{N^2} + O(1/N^3), \end{aligned}$$

y $\log f\left(\mathrm{e}^{-\lambda/N}\right) \approx_N -\frac{\lambda}{N} - \frac{b\lambda - \frac{\sigma^2}{2}\lambda^2}{N^2} + O(1/N^2)$, luego

$$\begin{aligned} \tilde{L}^N_{t+1/N}(\lambda) \approx_N \left(f_{[Nt]}\left(f\left(\mathrm{e}^{-\lambda/N}\right) \right) \right)^{Nx} &\approx \tilde{L}^N_t\left(-N \log f\left(\mathrm{e}^{-\lambda/N}\right) \right) \\ &\approx \tilde{L}^N_t \left(\lambda + \frac{b\lambda - \frac{\sigma^2}{2}\lambda^2}{N} + O(1/N^2) \right). \end{aligned}$$

Si $\lim_N \tilde{L}^N_t(\lambda) =: \tilde{L}_t(\lambda)$ se verifica $\frac{\partial \tilde{L}_t}{\partial t}(\lambda) = (b\lambda - \frac{\sigma^2}{2}\lambda^2)\frac{\partial \tilde{L}_t}{\partial \lambda}(\lambda)$. Denotamos por $\tilde{L}_t(\lambda) =: \exp\left(-x u_t(\lambda)\right)$ la función u_t verifica también

$$\frac{\partial}{\partial t} u_t = \left(b\lambda - \frac{\sigma^2}{2}\lambda^2 \right) \frac{\partial}{\partial \lambda} u_t, \qquad u_0(\lambda) = \lambda. \quad (2.14)$$

La unica solución de este ecuación es

$$u_t(\lambda) = \begin{cases} \dfrac{\lambda \mathrm{e}^{bt}}{1 + \lambda\sigma^2(\mathrm{e}^{bt} - 1)/2b} & \text{si } b \neq 0, \text{ y} \\ \dfrac{\lambda}{1 + \lambda\sigma^2 t/2} & \text{si } b = 0 \text{ (caso critico)} . \end{cases} \quad (2.15)$$

□

La función u_t se llama a veces cumulante y es también solución de la ecuación diferencial no lineal:

$$\frac{\partial}{\partial t} u_t = b u_t - \frac{\sigma^2}{2} u_t^2.$$

Utilizando la propiedad de Markov, se puede también mostrar que las distribuciones finito-dimensionales $\left(\tilde{X}^{(N)}_{t_1},\ldots,\tilde{X}^{(N)}_{t_k}\right)$ convergen.

En general, la convergencia débil de las distribuciones finito-dimensionales no es una condición suficiente para la convergencia de los procesos, considerados como funciones aleatorias a valores en $\mathbf{D}([0,T];\mathbb{R})$, el espacio de todos los mapas de $[0,T]$ en $\mathbb{R}$, continuos por la derecha, con límites por la izquierda. En $\mathbf{D}([0,T];\mathbb{R})$, se define la métrica

$$d(f,g) := \inf_{\theta\in\Theta}\left(\sup_{t\in[0,T]}\left|f(t)-g\big(\theta(t)\big)\right| + \sup_{s,t\in[0,T],s\neq t}\left|\log\frac{\theta(t)-\theta(s)}{t-s}\right|\right)$$

donde Θ es la clase de todas las funciones continuas, estrictamente crecientes, del intervalo $[0,T]$ sobre sí mismo. La métrica d induce una topología llamada de Skorohod. El espacio $\mathbf{D}([0,T];\mathbb{R})$ con esta topología es un espacio métrico, separable y completo. Ver [3].

Una condición adicional suficiente para la convergencia de la sucesión de los procesos es su compacidad relativa. El milagro con los procesos de BGW renormalizados consiste en que su convergencia en $\mathbf{D}([0,T];\mathbb{R})$ ya es una consecuencia de la convergencia de sus distribuciones uni-dimensionales. Ver [13] Theorem 3.4.

3.2.2 BGWc renormalizado

Para cada N definimos el proceso estocástico en tiempo continuo $\hat{X}^{(N)}_t := \frac{1}{N}X^{(N)}_t$ donde $X^{(N)}_t$ es un BGWc con parametros $\alpha^{(N)}$ y $\mathbf{r}^{(N)}$. Cada individuo tiene masa $1/N$. Acortamos la duración de vida aleatoria de cada individuo en función de N para crecer el numero medio de ramificaciones en una unidad de tiempo (es decir: $\lim_N \alpha^{(N)} = +\infty$).

Como antes, se controla la convergencia de $\hat{X}^{(N)}_t$ utilizando su transformada de Laplace.

$$\begin{aligned}\hat{L}^N_t(\lambda) = \mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X^{(N)}_t}\right) &= \left(\mathsf{E}\left(\left(\mathrm{e}^{-\lambda/N}\right)^{X^{(N)}_t}\middle| X^{(N)}_0 = 1\right)\right)^{X^{(N)}_0} \\ &= \left(F^{(N)}\left(\mathrm{e}^{-\lambda/N},t\right)\right)^{X^{(N)}_0} \approx_N \left(F^{(N)}\left(\mathrm{e}^{-\lambda/N},t\right)\right)^{Nx}\end{aligned}$$

En primar lugar analisamos un

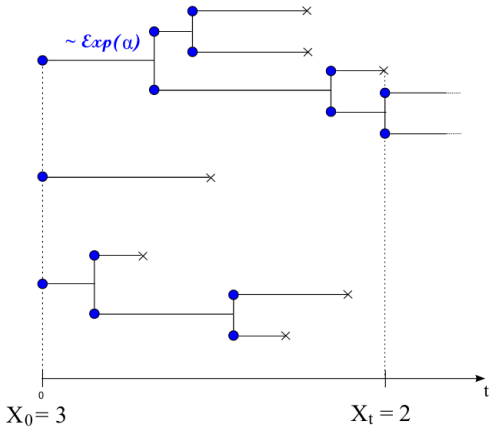


Figura 2.8: Una realización del BGWc binario (S. Pénisson).

Ejemplo 2.39 (reproducción binaria no crítica) Ver figura 2.8. Se calcula

$$F^{(N)}(s,t) = 1 - \frac{1-s}{\mathrm{e}^{-\alpha^{(N)}(m^{(N)}-1)t} + \frac{m^{(N)}}{2(m^{(N)}-1)}\left(1-\mathrm{e}^{-\alpha^{(N)}(m^{(N)}-1)t}\right)(1-s)}$$

para $s = \mathrm{e}^{-\lambda/N}$.

Ejercicio 2.40 Mostrar que, cuando $m^{(N)} = 1 + \frac{b}{N}$ ($\Rightarrow \mathbf{r}^{(N)}$ es asintoticamente crítica: $r_0^{(N)} = \frac{1}{2} - \frac{b}{N}$, $r_2^{(N)} = \frac{1}{2} + \frac{b}{N}$, $\sigma^{(N)} \to 1$ y cuando $\alpha^{(N)} = N\alpha$, entonces

$$\lim_N \hat{L}_t^N(\lambda) = \begin{cases} \exp\left(-x\dfrac{\lambda \mathrm{e}^{\alpha b t}}{1+\lambda\,(\mathrm{e}^{\alpha b t}-1)/2b}\right) & \text{si } b \neq 0, \\ \exp\left(-x\dfrac{\lambda}{1+\lambda\,\alpha t/2}\right) & \text{si } b = 0 \text{ (caso critico).} \end{cases}$$

De hecho, las condiciones $\rho^{(N)} = \alpha^{(N)}(m^{(N)}-1) \xrightarrow{N} \rho\,(=\alpha b)$ y $\sigma^{(N)} \xrightarrow{N} \sigma$ son cuasi necesarias y suficientes para la convergencia de los procesos $\hat{X}^{(N)}$ en general.

Teorema 2.41 (Construcción de la difusión de Feller como proceso límite) Si las tres condiciónes

$$\text{i) } m^{(N)} \approx_N 1 + {}^{b}\!/_{N}, \qquad \text{ii) } \lim_N \sigma^{(N)} = \sigma \text{ y} \qquad \text{iii) } \sup_N \sum_{j\geq 0} j^3 r_j^{(N)} < +\infty$$

son realizadas, si $\alpha^{(N)} = N\alpha$ y $\lim_N \hat{X}_0^{(N)} = x$, entonces los procesos $\hat{X}^{(N)}$ convergen hacia un proceso markoviano Y de generador infinitesimal A donde $Ah(x) = \alpha b x h'(x) + \alpha \frac{\sigma^2}{2} x h''(x)$, $h \in \mathscr{C}_b^2(\mathbb{R}^+)$. Llamaremos al proceso Y *difusión de Feller*, con parámetros $(\alpha b, \alpha\sigma^2)$.

Demostración. (Ver Theorem 1.3 in [7])
Se hace en tres pasos:

(i) Se muestra la convergencia de los generadores infinitesimales $\hat{Q}^{(N)}$ asociados a $\hat{X}^{(N)}$ hacia A

(ii) Se muestra que las leyes de los procesos $\hat{X}^{(N)}$ son relativamente compactas.

(iii) Se verifica que el problema de martingala asociado a A tiene una única solución.

Parte (i): Sea $h \in \mathscr{C}_b^2(\mathbb{R}^+)$ y $x = \frac{i}{N} \in \frac{1}{N}\mathbb{N}$.

$$\begin{aligned}
\hat{Q}^{(N)}h(x) &= \frac{\partial}{\partial t}\bigg|_{t=0} \left(P^{(N)}(t)h\left(\frac{\cdot}{N}\right)\right)(i) = \alpha^{(N)} i \sum_{k\geq 0} r_k^{(N)} \left(h\left(\frac{i+k-1}{N}\right) - h\left(\frac{i}{N}\right)\right) \\
&= N^2 \alpha x \sum_{k\geq 0} r_k^{(N)} \left(\frac{k-1}{N} h'(x) + \frac{1}{2}\left(\frac{k-1}{N}\right)^2 h''(x) + O\left(\frac{k^3}{N^3}\right)\right) \\
&= \alpha x N \big(m^{(N)} - 1\big) h'(x) + \frac{1}{2}\alpha x \sum_{k\geq 0} r_k^{(N)} \big(k - m^{(N)} + m^{(N)} - 1\big)^2 h''(x) + O\left(\frac{1}{N}\right) \\
&\xrightarrow{N} \alpha b x h'(x) + \frac{\sigma^2}{2}\alpha x h''(x) = Ah(x).
\end{aligned}$$

Parte (ii): Las leyes de los procesos $\hat{X}^{(N)}$ son relativamente compactas como probabilidades sovre $\mathbf{D}([0,T];\mathbb{R})$ si (Ver en [7], el criterio de Aldous)

a) $\forall t \leq T, \forall \varepsilon > 0, \exists C_\varepsilon \subset \mathbb{R}$ compacto: $\inf_N \mathsf{P}\big(\hat{X}_t^{(N)} \in C_\varepsilon\big) \geq 1 - \varepsilon$

b) Para cada familia $(\tau_N)_N$ de tiempos de paro limitados por T,
$\forall \varepsilon > 0, \exists \delta > 0$ y N_0 así que $\sup_{N\geq N_0} \sup_{\theta\leq\delta} \mathsf{P}\big(\big|\hat{X}_{\tau_N+\theta}^{(N)} - \hat{X}_{\tau_N}^{(N)}\big| > \varepsilon\big) \leq \varepsilon$.

Para mostrar esta ultima desigualdad, se verifica para la variación cuadrática de la parte martingala de $\hat{X}^{(N)}$ y para su parte a variación limitada.

Sobre (iii) (Ver [26] Section 6.2). Primero, verificamos que cada solución del problema de martingalas asociado con A es también solución debil de la ecuación diferencial estocástica

$$Y_t = x + \int_0^t \sigma\sqrt{\alpha Y_s}\,\mathrm{d}B_s + \int_0^t \alpha b Y_s\,\mathrm{d}s, \tag{2.16}$$

donde B es un movimiento Browniano.

Aplicando el problema de martingala a la función $h = \mathrm{Id}$ obtenemos: $\bar{Y}_t := Y_t - \int_0^t A\big(\mathrm{Id}\big)(Y_s)\,\mathrm{d}s = Y_t - \int_0^t \alpha b Y_s\,\mathrm{d}s$ es una martingala locale. Como es además integrable, es una martingala. Se quiere probar que la variación cuadrática de la martingala $\bar{Y}_t$ es igual a $\int_0^t \alpha\sigma^2 Y_s\,\mathrm{d}s$, para poder representarla como $\int_0^t \sigma\sqrt{\alpha Y_s}\,\mathrm{d}B_s$. Probamos de mañera equivalente que $\bar{Y}_t^2 - \int_0^t \alpha\sigma^2 Y_s\,\mathrm{d}s = Y_t^2 - 2Y_t\int_0^t \alpha b Y_s\,\mathrm{d}s + \Big(\int_0^t \alpha b Y_s\,\mathrm{d}s\Big)^2 - \int_0^t \alpha\sigma^2 Y_s\,\mathrm{d}s$ es una martingala; se sabe ya, aplicando el problema de martingala a la función $h(x) = x^2$, que

$$Y_t^2 - \int_0^t A\big(h\big)(Y_s)\,\mathrm{d}s = Y_t^2 - \int_0^t (2\alpha b Y_s^2 + \alpha\sigma^2 Y_s)\,\mathrm{d}s$$

es una martingala. Entonces, es suficiente probar que la diferencia $2Y_t\int_0^t \alpha b Y_s\,\mathrm{d}s - \Big(\int_0^t \alpha b Y_s\,\mathrm{d}s\Big)^2 - 2\int_0^t \alpha b Y_s^2\,\mathrm{d}s$ es una martingala. Esto se muestra utilizando la formula de integración por partes:

$$\begin{aligned} Y_t\int_0^t Y_s\,\mathrm{d}s &= \int_0^t Y_s Y_s\,\mathrm{d}s + \int_0^t\int_0^s Y_r\,\mathrm{d}r\,\mathrm{d}Y_s = \int_0^t Y_s^2\,\mathrm{d}s + \int_0^t\int_0^s Y_r(\alpha b Y_s)\,\mathrm{d}r\,\mathrm{d}s + \text{mart.} \\ &= \int_0^t Y_s^2\,\mathrm{d}s + \frac{1}{2}\int_0^t\int_0^t \alpha b Y_r Y_s\,\mathrm{d}r\,\mathrm{d}s + \text{mart.} = \int_0^t Y_s^2\,\mathrm{d}s + \frac{1}{2}\alpha b\Big(\int_0^t Y_s\,\mathrm{d}s\Big)^2 + \text{mart.} \end{aligned}$$

Ahora, como el coeficiente de difusión $x \mapsto \sigma\sqrt{\alpha x}$ de la ecuación (2.16) es una función Hölder-continua, los resultados clasicos permiten de concluir que la solución es unica. □

3.2.3 Unas propiedades de la difusión de Feller

Proposición 2.42 (Universalidad de la difusión de Feller) El proceso $\tilde{X}^\infty$ costruido como límite de BGW renormalizados es también una difusión de Feller Y con parámetros (b, σ^2).

Demostración. Verificamos que, por $\alpha = 1$, la transformada de Laplace de Y_t es también

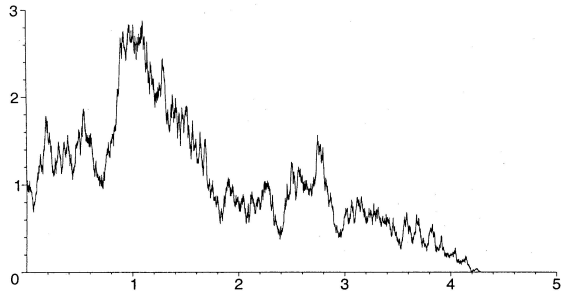


Figura 2.9: Simulación de $t \mapsto Y_t$ con el parámetro $b = 0$ (S. Pénisson).

solución de la ecuación (2.13).

Aplicando el problema de martingala a la función $h(x) = \mathrm{e}^{-\lambda x}$ inferimos

$$\mathrm{e}^{-\lambda Y_t} - \int_0^t A\big(h\big)(Y_s)\,\mathrm{d}s = \mathrm{e}^{-\lambda Y_t} - \int_0^t \left(\frac{\sigma^2}{2}\lambda^2 - b\lambda\right) Y_s \mathrm{e}^{-\lambda Y_s}\,\mathrm{d}s$$

es una martingala. Entonces

$$\mathsf{E}\big(\mathrm{e}^{-\lambda Y_t}\big) = \mathrm{e}^{-\lambda x} + \int_0^t \left(\frac{\sigma^2}{2}\lambda^2 - b\lambda\right) \mathsf{E}\big(Y_s \mathrm{e}^{-\lambda Y_s}\big)\,\mathrm{d}s$$

$$\Longrightarrow \quad \frac{\partial}{\partial t}\mathsf{E}\big(\mathrm{e}^{-\lambda Y_t}\big) = \left(b\lambda - \frac{\sigma^2}{2}\lambda^2\right) \frac{\partial}{\partial \lambda}\mathsf{E}\big(\mathrm{e}^{-\lambda Y_t}\big)$$

que es exactamente la ecuación diferencial (2.13). Como la transformada de Laplace caracteriza la ley de una variable aleatoria, se termina la demostración. □

Comentario 2.43

1) Así, los processos BGW y BGWc renormalizados convergen hacia el mismo proceso límite (si se toma $\alpha = 1$).

2) La difusión de Feller verifica la propiedad de ramificación: El proceso Y iniciado en $Y_0 = x_1 + x_2$ es la suma de dos copias independientes de Y iniciadas en x_1 y en x_2, ya que la transformada de Laplace es multiplicativa con respecto a Y_0.

3) ($\alpha = 1$) El parámetro b, coeficiente de la deriva del proceso, se interprete como un coeficiente de velocidad condicionada: $\lim_{\varepsilon\to 0} \mathsf{E}\left(\frac{Y_{t+\varepsilon}-Y_t}{\varepsilon}\,\middle|\, Y_t\right) = bY_t$, y el segundo

parámetro σ^2 aparece como el coeficiente de la fluctuación cuadtrática condicionada del proceso: $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{Var}\big(Y_{t+\varepsilon}-Y_t\,\big|\,Y_t\big)=\sigma^2 Y_t$.
Así se presente que las trajectorias del proceso Y ya que tienen una variación cuadrática finita, no tienen variación finita. (ver figura 2.9).

4) Cuando $b=0$ (resp. $b\leq 0$) el proceso se llama crítico (resp. subcrítico). En estos casos, c.s. el proceso toma el valor 0 y después, se queda en 0. Es la extinción del proceso.

5) En el caso subcrítico, se puede también condicionar la difusión Y a la no extinción en un futuro lejano, ver [18]. Se obtiene una otra difusión Y^*, de generador infinitesimal $A^*h(x)=Ah(x)+\sigma^2h'(x)=(bx+\sigma^2)h'(x)+\frac{\sigma^2}{2}xh''(x)$.
Y^* es una solución debil de la ecuación diferencial estocástica

$$Y_t=x+\int_0^t\sigma\sqrt{\alpha Y_s}\,\mathrm{d}B_s+\int_0^t(\alpha bY_s+\sigma^2)\,\mathrm{d}s. \tag{2.17}$$

• • • • • • • • • • •

Agradecimientos

Quiero hacer patente mi especial reconocimiento a Michael Högele por la feliz idea de organizar un workshop a la Universidad de los Andes, por haberme invitado y por haberme ayudado a concevir y escribir correctamente estas notas.

• • • • • • • • •

Referencias

[1] Athreya, K. B., Ney, P.: *Branching processes*. Springer, New York (1972).

[2] Axelrod, D. E., Kimmel, M.: *Branching processes in biology*. Springer, New York (2002).

[3] Billingsley, P.: *Convergence of probability measures*, second Edition, Wiley & Sons (1999).

[4] Caballero, M. E., Rivero, V. M. , Uribe Bravo, G., Velarde, C.: *Cadenas de Markov. Un enfoque elemental*. Aportaciones Matemáticas **29**, Sociedad Matematica Mexicana (2004).

[5] Caswell, H., Fujiwara, M., Brault, S.: *Declining survival probability threatens the North Atlantic right whale*. Proc. Nat. Acad. Sci. USA **96**, Pop. Biology, 3308–3313 (1999).

[6] Duquesne, T., Le Gall, J.-F.: *Random trees, Lévy processes and spatial branching processes*, Astérisque **281**, (2002).

[7] Ethier, S. N., Kurtz, T. G.: *Markov processes. Characterization and convergence*. Wiley, New York (1986).

[8] Euler, L.: *Recherches générales sur la mortalité et la multiplication du genre humain.* Mémoires de l'Académie des Sciences de Berlin **16**, (1760).

[9] Feller, W.: *Diffusion processes in genetics*. Proceedings of the second Berkeley symposium on mathematical statistics and probability. The Regents of the University of California, 227–246 (1951).

[10] Fernandez, B.: *Teoremas límites de alta densidad para campos aleatorios ramificados*. Aportaciones Matemáticas Serie Comunicaciones **2**, Sociedad Matematica Mexicana (1986).

[11] Galton, F., Watson, H. W.: *On the probability of the extinction of families*. The Journal of the Anthropological Institute of Great Britain and Ireland **4**, 138–144 (1874).

[12] Gorostiza, L. G.: *Ramificación y superprocesos*, Notas (2009) http://www.cimat.mx/Eventos/mesd09/gorostiza.pdf (sept. 2017).

[13] Grimvall, A.: *On the convergence of sequences of branching processes*. The Annals of Probability **2**(6), 1027–1045 (1974).

[14] Haccou, P., Jagers, P., Vatutin P.: *Branching processes. Variation, growth and extinction of populations*, Cambridge Univ. Press (2005).

[15] Harris, T. E.: *The theory of branching processes*, Springer-Verlag (1963).

[16] Heyde, C. C., Seneta, E.: *Statistical theory anticipated, I. J. Bienaymé*. Springer (1977).

[17] Jagers, P.: *Branching processes with biological applications*, Wiley (1975).

[18] Lambert, A.: *Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct*. Electronic Journal of Probability **12**, 420–446 (2007).

[19] Lamperti, J.: *The limit of a sequence of branching processes*. Zeit. für Wahrschein. Geb. **7**, 271-288 (1967).

[20] Lotka, A. J.: *The extinction of families*. Journal of the Washington Academy of Sciences **21**(18), 453–459 (1931).

[21] Martinez Qin tana, R.: *Proceso de ramificación de Galton-Watson multitipo*. Tesis doctoral, Universidad de Extremadura (2001).

[22] Méléard, S.: *Quasi-stationary distributions for population processes*, VI Escuela de probabilidad y procesos estocasticos, Guanajuato (2009) Notas http://www.cmap.polytechnique.fr/IMG/pdf/Guana01122009final.pdf (sept. 2017).

[23] Mota Medina, M.: *Aportaciones a la teoría de los Procesos de Ramificación Modificados*. Tesis doctoral, Universidad de Extremadura (1998).

[24] Pakes, A. G.: *Biological applications of branching processes*. in: Stochastic Processes: Modelling and Simulation. Eds: Shanbhag, Rao. Handbook of Statistics **21**, Elsevier (2003), 693–773.

[25] Sewastjanow, B. A.: *Verzweigungsprozesse*, Akademie Verlag Berlin (1974).

[26] Tudor, C.: *Procesos estocásticos*. Aportaciones Matemáticas Serie Textos **2**, Sociedad Matemática Mexicana, 3a edición (2002).

[27] Vatutin, V. A.: *Branching processes and their applications* (2005), Notas http://www.math.uni-frankfurt.de/ ismi/vatutin/Lecture2.pdf (sept. 2017).

[28] Vega, M. V.: *Cadenas de Markov de tiempo continuo y aplicaciones,* Trabajo monográfico, Licenciatura en Matemática, Universidad de la Republica Montevideo, Uruguay (2004).

[29] Wakolbinger, A.: *Poblaciones aleatorias ramificadas y sus equilibrios.* Notas de investigación **9**, Sociedad Matemática Mexicana (1994).

[30] Williams, D.: *Probability with martingales*, Cambridge University Press (1991).

Chapter 3

Markov models: stabilization and limit theorems

*Alexei Kulik**

Preface

The lecture course is focused on the study of long-time behavior of Markov chains and processes. A detailed introduction will be given to general methods for proving quantitative *ergodic rates* for the principal *stabilization* feature; that is, convergence of the transition kernels to the invariant probability measure when the observation horizon tend to ∞. This topic is quite diverse: depending on an actual long-time behavior of a Markov model, different notions and tools appear to be most natural. We explain the genealogy of these ideas and methods, starting from the classical case of a Markov chain with a finite state space, and ending up with much less studied *weak* ergodic rates, which appear to be very efficient for Markov models with rich (and thus complicated) local structure. We will discuss one particularly important field where the ergodic rates for Markov models find their natural application: the limit theorems for functionals of Markov models, which are of a principal importance for statistical inference of such models, their simulation, etc. These two parts of the lecture course are closely related; in particular, the form in which the ergodic rates are given is strongly motivated by their further applications.

*Ukrainian National Academy of Sciences, Institute of Mathematics Institute of Mathematics, Ukrainian National Academy of Sciences, 01601 Tereshchenkivska str. 3, Kyiv, Ukraine, and Technische Universität Berlin, Fakultät II, Institut für Mathematik, Str. 17. Juni 136, D-10623 Berlin, Germany; *kulik@imath.kiev.ua*

• • • • • • • • • •

1 Introduction

1.1 Elementary Markov chains

Let $\mathbb{X} = \{i, j, \dots\}$ be a *finite* set. An $\mathbb{X}$-valued sequence $X = \{X_n, n \geq 0\}$ is called a *Markov Chain* (MC) if for every time moments $m < n$ and every $m_1, \dots, m_k < m$,

$$\mathsf{P}(X_n = j | X_m = i, X_{m_1} = i_1, \dots, X_{m_k} = i_k) = \mathsf{P}(X_n = j | X_m = i), \qquad j, i, i_1, \dots, i_k \in \mathbb{X}.$$

The matrix $\Pi^{m,n} = \{p_{ij}^{m,n}\}_{i,j \in \mathbb{X}}$,

$$p_{ij}^{m,n} = \mathsf{P}(X_n = j | X_m = i),$$

is called the *transition probability matrix* of the MC X on the time interval $[m, n]$.

The matrix $\Pi^{m,n}$ has the following basic properties:

- ◇ each $\Pi^{m,n}$ is a stochastic matrix:

$$p_{ij}^{m,n} \geq 0, \quad i, j \in \mathbb{X}, \qquad \sum_j p_{ij}^{m,n} = 1, \quad i \in \mathbb{X};$$

- ◇ $p_{ij}^{m,m} = 1_{i=j}$;
- ◇ (Chapman-Kolmogorov equation): for each $m \leq r \leq n$, $\Pi^{m,n} = \Pi^{m,r}\Pi^{r,n}$; that is,

$$p_{ij}^{m,n} = \sum_{k \in \mathbb{X}} p_{ik}^{m,r} p_{kj}^{r,n}, \qquad i, j \in \mathbb{X}.$$

The initial distribution $\mu := \{\mu_i = \mathsf{P}(X_0 = i)\}_{i \in \mathbb{X}}$ and the transition probability matrices completely define the law of the sequence X,

$$\mathsf{P}(X_0 = i_0, X_1 = i_1, \dots, X_m = i_m) = \mu_{i_0} p_{i_0 i_1}^{0,1} \dots p_{i_{m-1} i_m}^{m-1,m}.$$

In what follows, we consider the case of time-homogeneous chain X, which by definition means that $\Pi^{m,n}$ depends on $n - m$ (the length of the corresponding time interval) only. In this case

$$\Pi^{m,n} = \Pi^{0,n-m} =: \Pi^{n-m}$$

(notation), and if we denote $\Pi := \Pi^1$ (the on step transition probability matrix), then

$$\Pi^n = \underbrace{\Pi \cdots \Pi}_{n};$$

that is, Π^n is just the (matrix) n-th power of Π. In what follows, p_{ij}^n denote the entries of Π^n. The matrix $\Pi = \{p_{ij}\}$ completely defines the "stochastic dynamics" related to the MC X.

Example 3.1 (Ehrenfest model of a gas diffusion) Let one have two boxes, each of the size N, and N balls be distributed within these $2N$ places. Each time moment, one ball is chosen at random and is placed from its current box to the other one.

Denote by X_n the number of balls in the first box. Then $\mathbb{X} = \{0, \dots, N\}$,

$$p_{i(i+1)} = \frac{N-i}{N}, \qquad p_{i(i-1)} = \frac{i}{N},$$

all the other entries are 0. *Typical questions are:* What is the limit of the distribution of X_n when $n \to \infty$? How quickly the limit is reached?

If μ is the initial distribution for X, then

$$\mu_j^n := \mathsf{P}(X_n = j) = \sum_{i \in \mathbb{X}} \mu_i p_{ij}^n, \qquad \mu^n = \mu \Pi^n.$$

Example 3.2 ("Toy" example with i.i.d. $\{X_n\}$) Let $\{X_n\}$ take values in $\mathbb{X}$ and be *independent and identically distributed.* Then

$$p_{ij}^n = \mathsf{P}(X_n = j | X_0 = i) = \mathsf{P}(X_n = j) =: \pi_j, \qquad j \in \mathbb{X}, \quad n \geq 1,$$

$$\Pi^n = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$$

Then for any μ,

$$\mu^n = \mu \Pi = \pi.$$

That is, after just *one* step X stabilizes its distribution.

Theorem 3.3 Let the MC X be *irreducible* and *aperiodic*. Then there exists $C, \rho > 0$ such that

$$\max_i \sum_j |p_{ij}^n - \pi_j| \le Ce^{-\rho n}, \tag{3.1}$$

where $\pi = \{\pi_i\}$ is the unique probability distribution on $\mathbb{X}$ which satisfies

$$\pi = \pi\Pi. \tag{3.2}$$

A probability distribution which satisfies (3.2) is called an *invariant probability measure* (IPM) for X. An IPM is characterized by the property that it is preserved under the action of MC: If $\mathrm{Law}(X_0) = \pi$ (i.e. an IPM), then for each n,

$$\mathrm{Law}(X_n) = \pi\Pi^n = \pi.$$

In this case the sequence $\{X_n\}$ is *strictly stationary*.

1.2 Limit theorems: The classical framework and Markov's breakthrough

In the classical framework, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) are formulated for the sums of *independent* random variables. Namely, if $\{\xi_n\}$ is a sequence of i.i.d. random variables with $\mathsf{E}\xi_1 = a$, $\mathsf{Var}\xi = \sigma^2$, then

$$\frac{1}{n}\sum_{k=1}^n \xi_k \to^{\mathsf{P}} a \tag{3.3}$$

(the LLN) and

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n (\xi_k - a) \Rightarrow \mathsf{N}(0, \sigma^2) \tag{3.4}$$

(the CLT).

Question (Markov, 1900ies): Is the independence assumption really necessary? Are there any realistic settings where this assumption fails but the main limit theorems still hold?

The answer is *positive*: For a MC the stabilization (in the sense of Theorem 3.3) leads to the LLN and the CLT. This was demonstrated by Markov himself:

- ◇ 1906: LLN, general setting;

◇ 1908: CLT, a complicated version of the *method of moments*, $\#\mathbb{X} \le 3$.

Markov's method to prove the LLN is remarkably simple and general. By the Chebyshev inequality

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathsf{E}\xi_k)\right| > \varepsilon\right) \le \varepsilon^{-2}\mathsf{E}\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathsf{E}\xi_k)\right)^2.$$

For any (square integrable) sequence $\{\xi_n\}$

$$\mathsf{E}\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathsf{E}\xi_k)\right)^2 = \frac{1}{n^2}\sum_{k,l=1}^{n}\mathsf{E}(\xi_k - \mathsf{E}\xi_k)(\xi_l - \mathsf{E}\xi_l) = \frac{1}{n^2}\sum_{k,l=1}^{n}\mathrm{Cov}(\xi_k, \xi_l).$$

Assume that

$$\mathrm{Cov}(\xi_k, \xi_l) \le C, \qquad v_M := \sup_{|k-l|>M}\mathrm{Cov}(\xi_k, \xi_l) \to 0, \quad M \to \infty. \tag{3.5}$$

Take $M = M(n)$ such that $M \to \infty$, $M/n \to 0$ (e.g. $M = [\sqrt{n}]$), then

$$\frac{1}{n^2}\sum_{k,l=1}^{n}\mathrm{Cov}(\xi_k, \xi_l) \;=\; \left(\sum_{\substack{k,l\le n\\ |k-l|\le M}} + \sum_{\substack{k,l\le n\\ |k-l|>m}}\right)\mathrm{Cov}(\xi_k, \xi_l) \;\le\; C\frac{M^2}{N^2} + v_M \;\to\; 0.$$

For Markov Chains the assumption (3.5) is easy to verify. Namely, let X be such that the assertion of Theorem 3.3 holds true, and let $\xi_n = f(X_n)$, where $f : \mathbb{X} \to \Re$ is a given function ("an observable"). Then for (say) $l \ge k$,

$$\begin{aligned}\mathrm{Cov}(\xi_k, \xi_l) &= \mathsf{E}f(X_k)f(X_l) - \mathsf{E}f(X_k)\mathsf{E}f(X_l)\\ &= \sum_{i,j} f(i)f(j)\left[p_{ij}^{l-k} - \mathsf{P}(X_l = j)\right]\mathsf{P}(X_k = i)\\ &\le 2\mathrm{e}^{-\rho(l-k)}\max_i f^2(i).\end{aligned}$$

Corollary 3.5 Under the conditions of Theorem 3.3, for any $f : \mathbb{X} \to \Re$,

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \to \langle f, \pi\rangle := \sum_i f(i)\pi_i \tag{3.6}$$

in mean square (and hence in probability).

Proof. It remains to prove that

$$\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\xi_k \to \langle f,\pi\rangle.$$

But

$$\mathsf{E}\xi_n = \sum_{i,j}\mu_i p_{ij}^n f(j) \to \sum_{i,j}\mu_i\pi_j f(j) = \sum_j \pi_j f(j) = \langle f,\pi\rangle,$$

which together with the Ćesaro theorem completes the proof. □

1.3 Coupling and stabilization for a Markov chain

Here we explain, in the simplest setting, one very efficient method to prove convergence of transition probabilities for a MC, which was developed by W. Döblin in 1936, and is based on the notion of *coupling*.

By the definition, a *coupling* for a pair of probability measures μ and ν is a pair of random variables (ξ,η) on a common probability space $(\Omega,\mathscr{F},\mathsf{P})$ such that

$$\mathrm{Law}(\xi)=\mu, \qquad \mathrm{Law}(\eta)=\nu.$$

We use the notation $(\xi,\eta)\in C(\mu,\nu)$.

The core of the coupling method is provided by the following simple observation: If μ and ν are two probability measures on finite state space $\mathbb{X}$, then for each $(\xi,\eta)\in C(\mu,\nu)$,

$$\sum_i |\mu_i-\nu_i| \le 2\mathsf{P}(\xi\neq\eta). \tag{3.7}$$

To prove (3.7), we just note that for the joint law $\kappa=\{\kappa_{ij}\}$ of the pair (ξ,η) we have

$$\sum_j \kappa_{ij}=\mu_i, \qquad \sum_i \kappa_{ij}=\nu_j,$$

hence

$$\sum_i|\mu_i-\nu_i| = \sum_i\left|\sum_j\kappa_{ij}-\sum_j\kappa_{ji}\right| \le \sum_{i,j}|\kappa_{ij}-\kappa_{ji}| \le \sum_{i\neq j}(\kappa_{ij}+\kappa_{ji}) = 2\mathsf{P}(\xi\neq\eta).$$

Let the MC X satisfy the following additional assumption: For each i, j,

$$p_{ij} > 0.$$

Define the new MC $Z = (X, Y)$ on $\mathbb{X} \times \mathbb{X}$ with the transition probabilities

$$p_{(i_1,i_2)(j_1,j_2)} = \begin{cases} p_{i_1,j_1} p_{i_2,j_2}, & i_1 \neq i_2, \\ p_{i_1,j_1} 1_{j_1=j_2}, & i_1 = i_2. \end{cases}$$

That is:

- ◇ if the current positions of X and Y are different, X and Y perform one step independently with the prescribed transition probabilities;
- ◇ if the current positions of X and Y coincide, then X and Y perform one step together with the prescribed transition probability.

The construction and the assumption made above yield that

- ◇ if $X_n = Y_n$ then $X_N = Y_N$, $N \geq n$ (once coupled, the trajectories stay coupled);
- ◇ the (conditional) probability for uncoupled trajectories to become coupled at each step is
$$\geq \min_{i_1 \neq i_2} \sum_j p_{i_1 j} p_{i_2 j} =: c > 0.$$

Then for each initial distribution of (X, Y) one has

$$\mathsf{P}(X_n \neq Y_n) \leq (1-c)^n.$$

Taking $X_0 = i$ and $Y_0 \sim \pi$, we get that (X_n, Y_n) is a coupling for the pair of distributions $\{p_{ij}^n, j \in \mathbb{X}\}$ and $\{\pi_j, j \in \mathbb{X}\}$, and thus

$$\sum_j |p_{ij}^n - \pi_j| \leq 2(1-c)^n.$$

This proves Theorem 3.3 with $C = 2$ and $\rho = -\log(1-c)$.

Remark 3.6 Ny a more careful choice of the Markov chain Z one weakens the above assumptions and provides a quicker convergence under the *Markov condition*

$$\sum_k \min(p_{ik}, p_{jk}) > 0, \qquad i, j \in \mathbb{X}, \tag{3.8}$$

see below.

For an arbitrary chain which satisfies the conditions of Theorem 3.3 there exists M such that, for some j_0,

$$p^M_{ij_0} > 0.$$

The proof remains mainly the same, with the "one-step coupling attempt" replaced by the "M-step coupling attempt".

1.4 General state space: general Markov chains, coupling lemma and the uniform convergence rate

By the definition, a *probability kernel* on a given measurable space $(\mathbb{X}, \mathscr{X})$ is a function $P(x,A)$, $x \in \mathbb{X}$, $A \in \mathscr{X}$, which is a measurable function in x and a probability measure in A. A general (time-homogeneous) Markov Chain on a given *filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}\}_n, \mathsf{P})$ with the state space $\mathbb{X}$ is a process such that, for a certain family of probability kernels $\{P^n, n \geq 0\}$,

$$\mathsf{P}(X_{n+m} \in A | \mathscr{F}_m) = P^n(X_m, A), \qquad n, m \geq 0, \quad A \in \mathscr{X},$$

and the family $\{P^n, n \geq 0\}$ satisfies

- $P^0(x,A) = 1_A(x) = \delta_x(A)$;
- (Chapman-Kolmogorov equation): for $r \leq n$

$$P^n(x,A) = \int_{\mathbb{X}} P^r(x, \mathrm{d}y) P^{n-r}(y,A).$$

By the definition, the total variation (TV) distance between two measures μ and ν on $(\mathbb{X}, \mathscr{X})$ equals to the total variation of the signed measure $\mu - \nu$; notation $\|\mu - \nu\|_{\mathrm{TV}}$. Alternative definitions are:

- $\|\mu-\nu\|_{\mathrm{TV}} := 2\sup_{A\in\mathscr{X}}|\mu(A)-\nu(A)|$;

- if $\mu \ll \lambda$ and $\nu \ll \lambda$, then $\|\mu-\nu\|_{\mathrm{TV}} := \int_{\mathbb{X}}\left|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}-\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right|\mathrm{d}\lambda$.

Example 3.7 (Finite state space) Let $\mathbb{X}$ be finite, then $\mu=\{\mu_i\}$, $\nu=\{\nu_i\}$. Take λ equal to the counting measure, then the corresponding Radon-Nikodym derivatives at a point i just equal μ_i, ν_i, hence

$$\|\mu-\nu\|_{\mathrm{TV}} = \sum_i |\mu_i-\nu_i|.$$

That is, (3.1) actually gives a bound for the TV distance between the transition probability $P^n(j,\cdot)$ and the IPM π. The left hand side term in (3.7) is just the TV distance between μ and ν.

The following result gives an important two-way extension of the inequality (3.7): Firstly, it is given for a general state space $\mathbb{X}$; secondly, it is shown that under a proper choice of the coupling $(\xi,\eta)\in C(\mu,\nu)$ the inequality turns into identity.

Lemma 3.8 (The Coupling Lemma)

$$\|\mu-\nu\|_{\mathrm{TV}} = 2\min_{(\xi,\eta)\in C(\mu,\nu)} \mathsf{P}(\xi\neq\eta). \tag{3.9}$$

Proof. Let $\mathbb{X}=A_+\cup A_-$ be the *Hahn decomposition* for $\mu-\nu$, then

$$\begin{aligned}\|\mu-\nu\|_{\mathrm{TV}} &= \big(\mu(A_+)-\nu(A_+)\big)-\big(\mu(A_-)-\nu(A_-)\big)\\ &= \big(2\mu(A_+)-1\big)-\big(2\nu(A_+)-1\big)\\ &= 2\big(\mu(A_+)-\nu(A_+)\big).\end{aligned}$$

Then for any $(\xi,\eta)\in C(\mu,\nu)$,

$$\begin{aligned}\|\mu-\nu\|_{\mathrm{TV}} = 2\big(\mu(A_+)-\nu(A_+)\big) &= 2\mathsf{E}\big[1_{\xi\in A_+}-1_{\eta\in A_+}\big]\\ &\le 2\mathsf{P}(\xi\in A_+,\eta\notin A_+)\le 2\mathsf{P}(\xi\neq\eta).\end{aligned} \tag{3.10}$$

This inequality is an abstract version of (3.7). Below we construct the joint law κ of $(\xi,\eta)\in C(\mu,\nu)$ explicitly, which turns inequality (3.10) into an equality.

Consider a probability measure λ such that $\mu \ll \lambda$ and $\nu \ll \lambda$ and put

$$f = \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}, \qquad g = \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}, \qquad h = f \wedge g.$$

If $p = \int h \, \mathrm{d}\lambda = 1$, then $\mu = \nu$, and we put $\kappa(A_1 \times A_2) = \mu(A_1 \cap A_2)$; that is, we set the components ξ and η equal to one another, with the law $\mu = \nu$. Otherwise, we decompose

$$\mu = p\theta + (1-p)\sigma_1, \qquad \nu = p\theta + (1-p)\sigma_2, \qquad \mathrm{d}\theta = \frac{1}{p} h \, \mathrm{d}\lambda, \tag{3.11}$$

with the convention that $\theta = \lambda$ if $p = 0$. We note that

$$\kappa(A_1 \times A_2) = p\theta(A_1 \cap A_2) + (1-p)\sigma_1(A_1)\sigma_2(A_2)$$

and observe that κ is the distribution of some $(\xi, \eta) \in C(\mu, \nu)$ with

$$\kappa\big(\{(x,y) : x \neq y\}\big) \leq 1 - p.$$

To explain the above inequality, recall that in the above construction of κ, we first represent μ and ν as mixtures (3.11) with the same first part. Then we toss a coin with probability of success p, and in the case of a success, we take both components equal to the law θ; otherwise the components are chosen independently with the laws $\sigma_{1,2}$. Clearly, under such a construction the probability for the components to be different does not exceed $1-p$. Hence

$$\begin{aligned} 2\mathsf{P}(\xi \neq \eta) = 2\kappa\big(\{(x,y) : x \neq y\}\big) &\leq 2(1-p) = 2\int (1-h) \, \mathrm{d}\lambda \\ &= \int \big(f + g - 2(f \wedge g)\big) \, \mathrm{d}\lambda = \int |f - g| \, \mathrm{d}\lambda = \|\mu - \nu\|_{\mathrm{TV}}. \quad \square \end{aligned}$$

Theorem 3.9 (Dobrushin's theorem) Let a general MC X satisfy the following *(Markov-)Dobrushin condition:* For some $c > 0$,

$$\sup_{x,y} \|P(x, \cdot\,) - P(y, \cdot\,)\|_{\mathrm{TV}} \leq 2(1-c). \tag{3.12}$$

Then there exists a unique IPM π for X and

$$\sup_{x} \|P^n(x, \cdot\,) - \pi\|_{\mathrm{TV}} \leq 2(1-c)^n.$$

The proof is essentially the same as for Theorem 3.3, though the coupling construction should be changed. Now the two-component chain (X,Y) is constructed in such a way that the joint law of (X_1,Y_1) conditioned on $X_0=x$ and $Y_0=y$ has the distribution which gives the minimum value in the righthand side of (3.9) for $\mu = P(x,\cdot);\nu = P(y,\cdot)$. Under such a choice,

$$\mathsf{P}(X_n \neq Y_n) \leq 2(1-c)^n,$$

which provides the required convergence rate. □

Remark 3.10 One should take care of making the law of (X_1,Y_1) (which now depends on (x,y)) to be a measurable function of (x,y). This technicality can be resolved easily; see [23] or [13], Lemma 1.

Further References: [7] (for the elementary theory of Markov chains); [15], [22] (for the general coupling method); [16], [17], [3],[4] (for the origins of the theory); [8], [9], [12] (lecture notes).

• •

2 Intrinsic memory, coupling distances, weak ergodicity

2.1 Motivating examples

Example 3.12 (Finite-dimensional diffusions) Consider a *Stochastic Differential Equation* (SDE) in $\mathbb{R}^d$:

$$\mathrm{d}X(t) = a\big(X(t)\big)\,\mathrm{d}t + \sigma\big(X(t)\big)\,\mathrm{d}W(t). \tag{3.13}$$

Background: Any "mechanical" dynamical system, perturbed by a "Gaussian white noise".

A solution of (3.13) is a Markov process in $\mathbb{R}^d$; $X^h := \{X^h_n = X(hn)\}$ is the corresponding *skeleton chain*. Under proper conditions on the coefficients (Hölder continuity and non-degeneracy of $\sigma\sigma^*$), this Markov process has a continuous *transition probability density*

$$p_t(x,z) = \frac{P_t(x,\mathrm{d}z)}{\mathrm{d}z},$$

and hence the Dobrushin condition holds true *locally*. A special case is provided by *periodic coefficients*, then the process can be projected on the corresponding d-dimensional

torus, and the Dobrushin condition holds true globally.

Example 3.13 (Delay Equations) Consider a *Stochastic Differential Delay Equation* (SDDE) in $\mathbb{R}^d$:

$$\mathrm{d}X(t) = a\big(X(t), X(t-r)\big)\,\mathrm{d}t + \sigma\big(X(t), X(t-r)\big)\,\mathrm{d}W(t). \tag{3.14}$$

Background:

- ◇ *Population dynamics:* an offspring requires the time period $r > 0$ to reach maturity.
- ◇ *Climate models:* El Niño/La Niña Southern Oscillation (ENSO); is triggered by warm equatorial ocean masses which need a certain concentration period after a previous El Niño effect.

A solution to (3.14) is NOT a Markov process: To define $X(\,\cdot\,)$ after time instant t, together with $X(t)$, the value $X(t-r)$ is required, as well. Define

$$X_t = \{X(t+s), s \in (-r,0)\} \in C(-r,0);$$

the *segment* of the trajectory of $X(t)$. The *segment process* $\{X_t, t \geq 0\}$ is a Markov process with the state space $\mathbb{X} = C(-r,0)$.

For the segment process, the *intrinsic memory* property can be naturally observed. Recall that, given a trajectory of a stochastic integral $I(t) := \int_{t_0}^t \beta_s\,\mathrm{d}W(s)$, $t \in [t_0,t_1]$, we can obtain its *quadratic variation* $\int_{t_0}^t (\beta_s)^2\,\mathrm{d}s$, $t \in [t_0,t_1]$, as a mean square limit of the sums $\sum_k \Big(I(t_{k,n}) - I(t_{k-1,n})\Big)^2 1_{t_{k,n} \leq t}$, where $\{t_{k,n}\}$ is a properly chosen partition of the interval $[t_0,t_1]$. That is, observing one value X_t of the segment process we actually observe the family

$$\Sigma_t(s) := \sigma^2\big(X(t+s), X(t+s-r)\big), \qquad s \in [-r,0].$$

Assume that the function $z \mapsto \sigma^2(x,z)$ has an inverse for each x, then we also observe

$$X(t+s-r) = \Big[\sigma^2(X(t+s),\cdot)\Big]^{-1}\big(\Sigma_t(s)\big), \qquad s \in [-r,0].$$

That is, observing ONE value X_t we actually have the complete information about ALL previous values. In particular, for different starting values $f, g \in C(-r,0)$ the transition probabilities $P_t(f,\,\cdot\,)$ and $P_t(g,\,\cdot\,)$ are mutually singular. The Dobrushin condition fails.

2.2 Coupling probability distances

Let $\mathbb{X}$ be a *Polish* space (a *metric* space, which is *separable* and *complete*) with the metric ρ. We call a *distance-like function* any function $d : \mathbb{X} \times \mathbb{X} \to [0, \infty)$, which is symmetric, satisfies

$$d(x,y) = 0 \quad \Longleftrightarrow \quad x = y,$$

and is lower semi-continuous; that is, for any sequences $x_n \to x$, $y_n \to y$

$$d(x,y) \le \liminf_n d(x_n, y_n).$$

Typical choices for d are

- $d = \rho^p$ with some $p > 0$;
- $d(x,y) = 1_{x \neq y}$, the *discrete* metric.

Denote by $\mathscr{P}(\mathbb{X})$ the class of all probability measures on $\mathbb{X}$. The *coupling (probability) distance* on $\mathscr{P}(\mathbb{X})$, corresponding to d, is denoted by the same letter d, and is defined by

$$d(\mu, \nu) = \inf_{(\xi,\eta) \in C(\mu,\nu)} \mathsf{E} d(\xi, \eta), \qquad \mu, \nu \in \mathscr{P}(\mathbb{X}). \tag{3.15}$$

We use the term "coupling distance" instead of "coupling metric" because, in general, d may fail to satisfy the triangle inequality. However, the following statement shows that the coupling distance may inherit the (weak) triangle inequality property from the initial distance-like function.

Proposition 3.14 Assume that the distance-like function d is a quasi-metric; that is, there exists $c \ge 1$ such that

$$d(x,z) \le c\big(d(x,y) + d(y,z)\big), \qquad x, y, z \in \mathbb{X}.$$

Then the same property holds true for the respective coupling distance d:

$$d(\mu, \lambda) \le c\big(d(\mu, \nu) + d(\nu, \lambda)\big), \qquad \mu, \nu, \lambda \in \mathscr{P}(\mathbb{X}).$$

Proof. Fix $\varepsilon > 0$ and choose $(\xi_\varepsilon, \eta_\varepsilon) \in C(\mu, \nu)$, $(\xi'_\varepsilon, \eta'_\varepsilon) \in C(\nu, \lambda)$ such that

$$\mathsf{E}d(\xi_\varepsilon, \eta_\varepsilon) \le d(\mu, \nu) + \varepsilon, \qquad \mathsf{E}d(\xi'_\varepsilon, \eta'_\varepsilon) \le d(\nu, \lambda) + \varepsilon.$$

The following useful fact is well known ([6, Problem 11.8.8]; see also [12, Lemma 2.6]).

Lemma 3.15 Let (ξ, η) and (ξ', η') be two pairs of random elements valued in a Borel measurable space $(\mathbb{X}, \mathscr{X})$ such that η and ξ' have the same distribution. Then on a proper probability space, there exist three random elements ζ_1, ζ_2, ζ_3 such that the law of (ζ_1, ζ_2) in $(\mathbb{X} \times \mathbb{X}, \mathscr{X} \otimes \mathscr{X})$ coincides with the law of (ξ, η) and the law of (ζ_2, ζ_3) coincides with the law of (ξ', η').

Applying this fact, we get a triple ζ_1, ζ_2, ζ_3, defined on the same probability space, such that

$$\mathsf{E}d(\zeta_1, \zeta_2) \le d(\mu, \nu) + \varepsilon, \qquad \mathsf{E}d(\zeta_2, \zeta_3) \le d(\nu, \lambda) + \varepsilon.$$

In addition, the laws of ζ_1, ζ_3 are equal to μ, λ, respectively. Hence

$$d(\mu, \lambda) = \inf_{(\xi,\eta) \in C(\mu,\lambda)} \mathsf{E}d(\xi, \eta) \le \mathsf{E}d(\zeta_1, \zeta_3) \le c\Big(d(\mu, \nu) + d(\nu, \lambda) + 2\varepsilon\Big).$$

Because $\varepsilon > 0$ is arbitrary, this gives the required inequality. □

The definition of the coupling distance is strongly related to the classical Monge-Kantorovich *mass transportation problem* (e.g. [21]). Namely, given two "mass distributions" μ, ν and the "transportation cost" $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$, the coupling distance $d(\mu, \nu)$ introduced above represents exactly the "minimal cost to transport μ to ν". An important fact is that the "optimal transportation plan" exists. In terms of couplings (which is just another name for transportation plans) and coupling distances, this fact can be formulated as follows.

Proposition 3.16 Let $\mathbb{X}$ be a Polish space and denote by d a distance-like function. Then for any $\mu, \nu \in \mathscr{P}(\mathbb{X})$, there exists a coupling $(\xi_*, \eta_*) \in C(\mu, \nu)$ such that

$$d(\mu, \nu) = \mathsf{E}d(\xi_*, \eta_*). \tag{3.16}$$

In other words, "inf" in (3.15) in fact can be replaced by "min". We call any pair $(\xi_*, \eta_*) \in C(\mu, \nu)$ which satisfies (3.16) an *optimal coupling* and denote the class of optimal couplings $C_{d,\mathrm{opt}}(\mu, \nu)$. The same terminology and notation is also used when

we deal with laws on $\mathbb{X}\times\mathbb{X}$ instead of pairs of random elements. By $C(\mu,\nu)$, we will denote the class of measures on $\mathbb{X}\times\mathbb{X}$ such that their projections on the first and second coordinates equal μ and ν respectively and by $C_{d,opt}(\mu,\nu)$ the subclass of measures $\kappa\in C(\mu,\nu)$ such that

$$d(\mu,\nu)=\int_{\mathbb{X}\times\mathbb{X}} d(x,y)\,\kappa(\mathrm{d}x,\mathrm{d}y).$$

Proof of Proposition 3.16. Observe that the family of measures $C(\mu,\nu)$ is *tight* (e.g. [2]): to construct a compact set $K\subset\mathbb{X}\times\mathbb{X}$ such that, for a given ε, $\kappa(K)\geq 1-\varepsilon$, one can simply choose two compact sets $K_1,K_2\subset\mathbb{X}$ such that

$$\mu(K_1)\geq 1-\frac{\varepsilon}{2},\qquad \mu(K_2)\geq 1-\frac{\varepsilon}{2}$$

and then put $K=K_1\times K_2$.

Consider a sequence of pairs $\{(\xi_n,\eta_n)\}\subset C(\mu,\nu)$ such that

$$\mathsf{E}d(\xi_n,\eta_n)\leq d(\mu,\nu)+\frac{1}{n}.$$

Then by the Prokhorov theorem, there exists a subsequence $\{(\xi_{n_k},\eta_{n_k})\}$ which converges in law to some pair (ξ_*,η_*). Then both sequences of components $\{\xi_{n_k}\}$, $\{\eta_{n_k}\}$ also converge in law to ξ_*,η_* respectively, and hence $(\xi_*,\eta_*)\in C(\mu,\nu)$. Next, by Skorokhod's "common probability space" principle (e.g. [6, Theorem 11.7.2]), there exists a sequence $\{(\tilde{\xi}_k,\tilde{\eta}_k)\}$ and a pair $(\tilde{\xi}_*,\tilde{\eta}_*)$, defined on the same probability space, such that the laws of respective pairs (ξ_{n_k},η_{n_k}) and $(\tilde{\xi}_k,\tilde{\eta}_k)$ coincide and

$$\big(\tilde{\xi}_k,\tilde{\eta}_k\big)\to\big(\tilde{\xi}_*,\tilde{\eta}_*\big)$$

with probability 1. Then by the lower semi-continuity of d, one has

$$d\big(\tilde{\xi}_*,\tilde{\eta}_*\big)\leq\liminf_k d\big(\tilde{\xi}_k,\tilde{\eta}_k\big),$$

and hence the Fatou lemma gives

$$\mathsf{E}d\big(\tilde{\xi}_*,\tilde{\eta}_*\big)\leq\liminf_k\mathsf{E}d\big(\tilde{\xi}_k,\tilde{\eta}_k\big)=\liminf_k\mathsf{E}d(\xi_k,\eta_k)=d(\mu,\nu).$$

Because $\big(\tilde{\xi}_*,\tilde{\eta}_*\big)$ has the same law as $(\xi_*,\eta_*)\in C(\mu,\nu)$, this completes the proof. $\square$

Below we give several typical examples.

Example 3.17 Let $d(x,y) = \rho(x,y)$; we denote the respective coupling distance with $W_{\rho,1}$ and discuss its properties. First, since ρ satisfies the triangle inequality, so does $W_{\rho,1}$. Next, $W_{\rho,1}$ is symmetric and non-negative. Finally, it possesses the identification property:

$$W_{\rho,1}(\mu,\nu) = 0 \quad \Longleftrightarrow \quad \mu = \nu.$$

The part "$\Leftarrow$" of this statement is trivial; to prove the "$\Rightarrow$" part we just notice that there exists an optimal coupling (ξ_*, η_*) for μ, ν: Because d is continuous, we can apply Proposition 3.16. For this coupling, we have

$$\mathsf{E}d(\xi_*, \eta_*) = W_{\rho,1}(\mu,\nu) = 0,$$

and because d has the identification property this means that in fact, $\xi_* = \eta_*$ a.s. Hence their laws coincide.

We have just seen that for the coupling distance $W_{\rho,1}$ all the axioms of a metric hold true; the one detail which may indeed cause $W_{\rho,1}$ to not be a metric is that $W_{\rho,1}$, in general, may take value ∞. If ρ is *bounded*, this does not happen, and $W_{\rho,1}$ is a metric on $\mathscr{P}(\mathbb{X})$.

Example 3.18 Let $p > 1$ and $d(x,y) = \rho^p(x,y)$; we denote the respective coupling distance is a quasi-metric: By Proposition 3.14 this is an easy consequence of the triangle inequality for ρ and the elementary inequality $(a+b)^p \le 2^{p-1}(a^p + b^p)$, $a, b \ge 0$. Moreover, similarly to Proposition 3.14 one can prove using the triangle inequality for L_p-distance that

$$W_{\rho,p}(\mu,\lambda) := \Big(d(\mu,\nu)\Big)^{1/p}$$

satisfies the triangle inequality. The probability distance $W_{\rho,p}$ is often called the p-Wasserstein distance.

Example 3.19 Let $d(x,y) = 1_{x\neq y}$ be the *discrete metric* on $\mathbb{X}$. Then, by the Coupling Lemma,

$$d(\mu,\nu) = \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}.$$

Hence the total variation distance is a particular representative of the class of coupling distances. Note that the function $d(x,y) = 1_{x\neq y}$ is lower semi-continuous; thus, Proposition 3.16 gives an alternative way to obtain the statement of the Coupling Lemma.

2.3 Uniform weak ergodicity

Theorem 3.20 (General Dobrushin-type theorem) Let a general MC X satisfy the following *condition:* For some distance-like function and some $c > 0$,

$$d\big(P(x,\cdot),P(y,\cdot)\big) \le (1-c)\,d(x,y), \qquad x,y \in \mathbb{X}. \tag{3.17}$$

Then for every x, $y \in \mathbb{X}$,

$$d\big(P_n(x,\cdot),P_n(y,\cdot)\big) \le (1-c)^n\,d(x,y), \tag{3.18}$$

and there exists at most one IPM for X. If IPM π exists, then

$$d\big(P_n(x,\cdot),\pi\big) \le (1-c)^n \int_{\mathbb{X}} d(x,y)\,\pi(\mathrm{d}y), \qquad n \ge 1. \tag{3.19}$$

Proof. Consider a Markov chain $Z = \big(Z^1, Z^2\big)$ with the transition probability Q such that for every x, $y \in \mathbb{X}$ the measure $Q\big((x,y),\cdot\big)$ belongs to the set of d-optimal couplings of $P(x,\cdot)$ and $P(y,\cdot)$; such a kernel exists due to Proposition 3.16. For given x, $y \in \mathbb{X}$ consider the law $\mathsf{P}^Z_{(x,y)}$ of this process with $Z_0 = (x,y)$.

Then by Condition (3.17),

$$\mathsf{E}^Z_{x,y} d\big(Z^1_n, Z^2_n\big) \le (1-c)\,\mathsf{E}^Z_{x,y} d\big(Z^1_{n-1}, Z^2_{n-1}\big) \le \ldots \le (1-c)^n\,d(x,y).$$

Since for every $n \ge 1$ the pair Z^1_n, Z^2_n gives a coupling for $P_n(x,\cdot)$, $P_n(y,\cdot)$, this proves (3.18). The same argument, but with y replaced by a random variable with the law π, proves (3.19). □

Remark 3.21 Under certain topological conditions (of a technical kind), one can show that (3.18) yields existence of the IPM; see Proposition 2.23 in [12].

Remark 3.22 Note that in the above proof one has to take care of the following delicate technical question: The choice of the optimal coupling $Q((x,y),\cdot)$ should be made in the *measurable* way w.r.t. (x,y). For a solution of this *measurable selection problem* we refer to Proposition 2.19 in [12] in the case of continuous d and Proposition 4.2 in [14] in the general case.

If $d = \rho^p$ and ρ is bounded, d-convergence of $P_n(y,\cdot)$ to π is actually equivalent to the weak convergence. That is why we call a Markov process which satisfies (3.19) (exponentially) *weakly ergodic*.

2.4 Estimates for means and covariances

For given d, $\gamma \in (0,1]$ and a function $W : \mathbb{X} \to \mathfrak{R}^+$, we introduce the *weighted Hölder class $H_{\gamma,W}(\mathbb{X},d)$ w.r.t. d with the index γ and the weight W*. By definition, this is the set of functions $f : \mathbb{X} \to \mathfrak{R}$ such that

$$\|f\|_{d,\gamma,W} = \sup_{x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{d^\gamma(x_1,x_2)\big(W(x_1)+W(x_2)\big)^{1-\gamma}} < \infty.$$

Here and below we use the convention $a^0 = 1$, $a \in \mathfrak{R}^+$; hence for $\gamma = 1$ the weight W is inessential, and $H_{1,W}(\mathbb{X},d) = H_1(\mathbb{X},d)$ is just the Lipschitz class w.r.t. d.

The following statement is very simple, but it gives an efficient tool to derive the stabilization of the expectations from the weak ergodic bounds such as (3.19).

Proposition 3.23 Let a function f belong to $H_{\gamma,W}(\mathbb{X},d)$ for some $\gamma \in (0,1]$. Then for any $\mu,\nu \in \mathscr{P}(\mathbb{X})$,

$$\left|\int_{\mathbb{X}} f \,\mathrm{d}\mu - \int_{\mathbb{X}} f \,\mathrm{d}\nu\right| \le \|f\|_{d,\gamma,W}\big(d(\mu,\nu)\big)^\gamma \left(\int_{\mathbb{X}} W \,\mathrm{d}\mu + \int_{\mathbb{X}} W \,\mathrm{d}\nu\right)^{1-\gamma}. \tag{3.20}$$

Proof. For any coupling $(\xi,\eta) \in C(\mu,\nu)$ we have

$$\begin{aligned}\left|\int_{\mathbb{X}} f \,\mathrm{d}\mu - \int_{\mathbb{X}} f \,\mathrm{d}\nu\right| &= \big|\mathsf{E}f(\xi) - \mathsf{E}f(\eta)\big| \le \mathsf{E}\big|f(\xi) - f(\eta)\big| \\ &\le \|f\|_{d,\gamma,W}\,\mathsf{E}\left[d^\gamma(\xi,\eta)\big(W(\xi)+W(\eta)\big)^{1-\gamma}\right].\end{aligned}$$

If $\gamma < 1$, we apply the Hölder inequality with $p = 1/\gamma$:

$$\begin{aligned}\left|\int_{\mathbb{X}} f \,\mathrm{d}\mu - \int_{\mathbb{X}} f \,\mathrm{d}\nu\right| &\le \|f\|_{d,\gamma,W}\big(\mathsf{E}d(\xi,\eta)\big)^\gamma\big(\mathsf{E}W(\xi)+\mathsf{E}W(\eta)\big)^{1-\gamma} \\ &= \|f\|_{d,\gamma,W}\big(\mathsf{E}d(\xi,\eta)\big)^\gamma\left(\int_{\mathbb{X}} W \,\mathrm{d}\mu + \int_{\mathbb{X}} W \,\mathrm{d}\nu\right)^{1-\gamma}.\end{aligned}$$

For $\gamma = 1$ the same bound holds true directly. Taking the infimum in this bound w.r.t. all the couplings $(\xi, \eta) \in C(\mu, \nu)$, we get the required statement. $\square$

As a corollary, we obtain the following.

Proposition 3.24 Let X be stationary and satisfy (3.17). Let $f \in H_{\gamma,W}(\mathbb{X}, d)$ and

$$\int_{\mathbb{X}} f^2 \, \mathrm{d}\pi < \infty, \qquad \int_{\mathbb{X}} W^2 \, \mathrm{d}\pi < \infty.$$

Then

$$\begin{aligned} &\Big|\mathrm{Cov}\big(f(X_j), f(X_k)\big)\Big| \\ &\qquad \le 2^{1-\gamma}(1-c)^{\gamma(k-j)} \|f\|_{d,\gamma,W} \left(\int_{\mathbb{X}} f^2 \, \mathrm{d}\pi\right)^{1/2} \left(\int_{\mathbb{X}} W^2 \, \mathrm{d}\pi\right)^{(1-\gamma)/2}. \end{aligned} \tag{3.21}$$

Proof. Using the bound (3.20) with $\mu(\mathrm{d}y) = P_n(x, \mathrm{d}y)$, $\nu(\mathrm{d}y) = \pi(\mathrm{d}y)$, we get

$$\left|\mathsf{E}_x f(X_n) - \int_{\mathbb{X}} f \, \mathrm{d}\pi\right| \le (1-c)^{n\gamma} \|f\|_{d,\gamma,W} \left(\mathsf{E}_x W(X_n) + \int_{\mathbb{X}} W \, \mathrm{d}\pi\right)^{1-\gamma}. \tag{3.22}$$

Then, by the Hölder inequality,

$$\begin{aligned} \Big|\mathrm{Cov}\big(f(X_j), f(X_k)\big)\Big| &= \left|\mathsf{E} f(X_j) \left(\mathsf{E}_{X_j} f(X_{k-j}) - \int_{\mathbb{X}} f \, \mathrm{d}\pi\right)\right| \\ &\le (1-c)^{(k-j)\gamma} \|f\|_{d,\gamma,W} \mathsf{E}\left[\big|f(X_j)\big| \left(\mathsf{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, \mathrm{d}\pi\right)^{1-\gamma}\right] \\ &\le (1-c)^{(k-j)\gamma} \|f\|_{d,\gamma,W} \big(\mathsf{E} f^2(X_j)\big)^{1/2} \left(\mathsf{E}\left(\mathsf{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, \mathrm{d}\pi\right)^2\right)^{(1-\gamma)/2}. \end{aligned}$$

Recall that X is stationary, hence

$$\mathsf{E} f^2(X_j) = \int_{\mathbb{X}} f^2 \, \mathrm{d}\pi,$$

$$\mathsf{E}\left(\mathsf{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, \mathrm{d}\pi\right)^2 \le 2\mathsf{E} W^2(X_k) + 2\int_{\mathbb{X}} W^2 \, \mathrm{d}\pi = 4\int_{\mathbb{X}} W^2 \, \mathrm{d}\pi,$$

which completes the proof of (3.21). $\square$

Now we can repeat literally the proof of Corollary 3.5, and get the following version of the LLN.

Corollary 3.25 Under the conditions of Theorem 3.20 and Proposition 3.24,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \int_{\mathbb{X}} f \, \mathrm{d}\pi \tag{3.23}$$

in mean square (and hence in probability).

Further References: [6], [24] (for the details on coupling (minimal) probability metrics); [5] (of the origins of the theory); [12] (lecture notes).

• •

3 Recurrence and Harris-type theorems

3.1 Prelude 1: Recurrence and transience for discrete Markov chains

If $\mathbb{X}$ is infinite but countable, the following classification is very instructive. Denote for $i \in \mathbb{X}$

$$\tau_i = \{n : X_n = i\}$$

with the convention that $\inf \varnothing = \infty$; that is, τ_i is an extended random variable which can take the value ∞. State i is called *recurrent* if

$$\mathsf{P}_i(\tau_i < \infty) = 1;$$

otherwise it is called *transient*. A recurrent state is called *positively recurrent* if $\mathsf{E}_i \tau_i < \infty$ and *zero recurrent* otherwise.

Example 3.27 (Bernoulli random walk on a half-line) Let $\mathbb{X} = \{0, 1, \dots\}$ with the one step transition probabilities

$$p_{00} = p_{i(i-1)} = q, \quad p_{01} = p_{i(i+1)} = p, \qquad i = 1, 2, \dots,$$

where $p + q = 1$. The chain is

- ⋄ transient for $p > q$;

- ◇ zero recurrent for $p = q = 1/2$;
- ◇ positive recurrent for $p < q$ (for the proof, see Example 3.30 below).

3.2 Prelude 2: Exponential recurrence for a solution to an SDE

Let X be the solution to the SDE

$$\mathrm{d}X(t) = a\big(X(t)\big)\,\mathrm{d}t + \sigma\big(X(t)\big)\,\mathrm{d}W(t) \tag{3.24}$$

with values in $\mathbb{R}^d$. Assume the following:

- ◇ σ is bounded;
- ◇ a is locally bounded and satisfies for some $R, c > 0$,

$$\langle a(x), x\rangle \le -c|x|^2, \qquad |x| > R.$$

Take $V(x) = |x|^2$ and put

$$Y(t) = \mathrm{e}^{\lambda t} V\big(X(t)\big),$$

where $\lambda > 0$ is a constant which will be defined later.

By the Itô formula,

$$\mathrm{d}Y(t) = \Big[\lambda \mathrm{e}^{\lambda t} V\big(X(t)\big) + \mathrm{e}^{\lambda t} AV\big(X(t)\big)\Big]\,\mathrm{d}t + V'\big(X(t)\big)\sigma\big(X(t)\big)\,\mathrm{d}W(t)$$

with

$$AV(x) := a(x)V'(x) + \frac{1}{2}\operatorname{Tr}\sigma(x)V''(x)\sigma^*(x) = 2\langle a(x), x\rangle + \operatorname{Tr}\sigma(x)\sigma^*(x).$$

Take $\lambda < 2c$, then because σ is bounded

$$\lambda V(x) + AV(x) \le (\lambda - 2c)\,|x|^2 + C \le 0, \qquad |x| \ge Q$$

for Q large enough.

Define

$$\tau^Q := \inf\big\{t : |X(t)| \le Q\big\}.$$

By the *Doob optional sampling theorem*

$$\begin{aligned}
\mathsf{E}_x Y(\tau_Q \wedge t) = Y(0) & & \big(= V(x)\big) \\
+ \mathsf{E}_x \int_0^{\tau_Q \wedge t} \Big[\lambda \mathrm{e}^{\lambda s} V\big(X(s)\big) + \mathrm{e}^{\lambda s} AV\big(X(s)\big)\Big]\, \mathrm{d}s & & (\le 0) \\
+ \mathsf{E}_x \int_0^{\tau_Q \wedge t} V'\big(X(s)\big)\sigma(X(s))\, \mathrm{d}W(s) & & (= 0).
\end{aligned}$$

That is,

$$\mathsf{E}_x \mathrm{e}^{\lambda \tau^Q} \le x^2.$$

3.3 Geometric recurrence for Markov chains: the Lyapunov function approach

Denote by

$$Af(x) = \mathsf{E}_x f(X_1) - f(x),$$

the "discrete generator" for a Markov chain.

Theorem 3.28 Let for some function V

$$AV(x) \le c_1 - c_2 V(x), \qquad x \in \mathbb{X}. \tag{3.25}$$

For a fixed $\rho > {}^{c_1}/_{c_2}$ denote

$$B := \{x : V(x) \le \rho\}, \qquad \tau := \min\{n : X_n \in B\}.$$

Then

$$\mathsf{E}_x \Big[1 - c_2 + c_1 \rho^{-1}\Big]^{-n\wedge\tau} V(X_{n\wedge\tau}) \le V(x). \tag{3.26}$$

Proof. For $x \notin B$, we have

$$1 \le \rho^{-1} V(x),$$

and thus

$$\mathsf{E}_x V(X_1) = V(x) + AV(x) \le V(x) + c_1 - c_2 V(x) \le (1 - c_2 + c_1 \rho^{-1}) V(x).$$

This inequality can be re-written as follows:

$$1_{x\notin B}\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-1}V(X_1)\le V(x). \tag{3.27}$$

Now we prove (3.26) using the induction by n. We have

$$\begin{aligned}\mathsf{E}_x&\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)\wedge\tau}V(X_{(n+1)\wedge\tau})\\ &=\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)\wedge\tau}V(X_{(n+1)\wedge\tau})(1_{\tau\le n}+1_{\tau>n})\\ &=\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-n\wedge\tau}V(X_{n\wedge\tau})1_{\tau\le n}\\ &\qquad+\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)}V(X_{n+1})1_{\tau>n}.\end{aligned}$$

On the other hand, by the *strong Markov property* of X,

$$\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)}V(X_{n+1})1_{\tau>n}=\mathsf{E}_x\left(\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)}1_{\tau>n}\mathsf{E}_{X_n}V(X_1)\right).$$

On the set $\{\tau>n\}$ we have $X_n\notin B$, hence applying (3.27) we get

$$\begin{aligned}\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)}V(X_{n+1})1_{\tau>n}&\le\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-n}V(X_n)1_{\tau>n}\\ &=\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-n\wedge\tau}V(X_{n\wedge\tau})1_{\tau>n}.\end{aligned}$$

That is,

$$\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-(n+1)\wedge\tau}V(X_{(n+1)\wedge\tau})\le\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-n\wedge\tau}V(X_{n\wedge\tau}).$$

Since (3.26) holds true for $n=0$ trivially, this proves the required inequality for all n. □

Corollary 3.29 The following *recurrence bounds* hold true:

- ◇ $\mathsf{E}_x\left[1-c_2+c_1\rho^{-1}\right]^{-\tau}\le V(x)$
- ◇ $\mathsf{E}_xV(X_n)1_{\tau\ge n}\le\left[1-c_2+c_1\rho^{-1}\right]^{n}V(x)$.

Example 3.30 Let X be a Bernoulli random walk on $\{0,1,\dots\}$ with $p<q$. Take $V(x)=$

λ^x with $\lambda > 1$ which is yet to be defined, then

$$\mathsf{E}_x V(X_1) - V(x) = p\lambda^{x+1} + \lambda^{x-1} - \lambda^x = \Big(p\lambda + q\lambda^{-1} - 1\Big)V(x), \qquad x \neq 0,$$
$$\mathsf{E}_0 V(X_1) - V(0) = p\lambda + q - 1.$$

Recall that $p < q$, hence the function

$$\lambda \mapsto p\lambda + q\lambda^{-1} - 1$$

takes negative values on the interval $(1, \sqrt{q/p})$. Taking any λ from this interval we get that the condition of the theorem holds true with

$$c_1 = q(1-\lambda^{-1}), \qquad c_2 = 1 - p\lambda - q\lambda^{-1}.$$

3.4 The Harris type theorem

Recall that the chain exhibits an exponential stabilization rate if it satisfies a *general Dobrushin-type condition*

$$d\big(P(x,\cdot), P(y,\cdot)\big) \leq (1-c)\, d(x,y), \qquad x,y \in \mathbb{X}. \tag{3.28}$$

Harris type theorems extend the domain where similar effect can be observed and proved by reducing the assumption (3.28) from the uniform-in-space form to a local form.

We call the distance-like function d *contracting* on a set B if for some $c = c_B > 0$

$$d\big(P(x,\cdot), P(y,\cdot)\big) \leq (1-c)\, d(x,y), \qquad x,y \in B. \tag{3.29}$$

This, certainly, is a local version of the Dobrushin condition (3.28). In addition, we call d *non-expanding* if

$$d\big(P(x,\cdot), P(y,\cdot)\big) \leq d(x,y), \qquad x,y \in \mathbb{X}.$$

Next, for any $p > 1$ define a new distance-like function

$$d_p(x,y) := \big(d(x,y)\big)^{1/p},$$

and denote by the same letter the corresponding coupling distance on $\mathscr{P}(\mathbb{X})$. Using the Hölder inequality one can easily see that if d is non-expanding, then d_p is non-expanding, as well, and if d satisfies (3.29), then for d_p the similar condition holds with $(1-c)$ changed to $(1-c)^{1/p}$.

Theorem 3.31 Let a function $V \geq 1$ be such that (3.25) holds true and a non-expanding distance-like function $d(x,y)$ is contracting on each of the sets $\{x : V(x) \leq r\}$, $r \geq 1$. Then for each $p > 1$ there exist $C = C_p$ and $\kappa = \kappa_p < 1$ such that for each $x, y \in \mathbb{X}$,

$$d_p\big(P_n(x,\cdot), P_n(y,\cdot)\big) \leq C\kappa^n d_p(x,y)\,\big(V(x)+V(y)\big)^{1/q}, \qquad n \geq 1,$$

here $1/p + 1/q = 1$. As a corollary, an IPM, if exists, is unique, and in this case

$$d_p\big(P_n(x,\cdot), \pi\big) \leq C\kappa^n d_p(x,y)\, V^{1/q}(x), \qquad n \geq 1.$$

Proof. Define the coupling Z as a Markov chain in $\mathbb{X}\times\mathbb{X}$ with the transition kernel Q such that for each (x,y) the measure $Q\big((x,y),\cdot\big)$ belongs to the set of d_p-optimal couplings of $P(x,\cdot)$ and $P(y,\cdot)$. We note that, for any such chain, the function $V^Z(x,y) = V(x)+V(y)$ satisfies

$$\mathsf{E}^Z_{(x,y)} V^Z(Z_1) \leq 2c_1 - c_2 V^Z(x,y). \tag{3.30}$$

We fix $\rho > 2c_1/c_2$ and put $K = \{x : V(x) \leq \rho\}$, $B = K \times K$. Then, by Corollary 3.29, there exists $\chi \in (0,1)$ such that

$$\mathsf{E}^Z_{(x,y)} \chi^{-\tau} \leq V(x)+V(y), \qquad \tau := \min\{n : Z_n \in B\}. \tag{3.31}$$

For a given $x, y \in \mathbb{X}$ consider the law $\mathsf{P}^Z_{(x,y)}$ of this process with $Z_0 = (x,y)$. Then for every $n \geq 1$, the pair Z^1_n, Z^2_n gives a coupling for $P_n(x,\cdot)$, $P_n(y,\cdot)$, and therefore

$$d_p\big(P_n(x,\cdot), P_n(y,\cdot)\big) \leq \mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big).$$

Denote by S_k, $k \geq 1$, the time moments of consecutive visits of Z to B. Then

$$d_p\big(P_n(x,\cdot), P_n(y,\cdot)\big) \leq \mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{S_1 \geq n} + \sum_{k=1}^{\infty} \mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{n \in (S_k, S_{k+1}]}. \tag{3.32}$$

We have simply

$$\mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{S_1 \ge n} \le \Big(\mathsf{E}^Z_{(x,y)} d\big(Z^1_n, Z^2_n\big)\Big)^{1/p} \Big(\mathsf{P}^Z_{(x,y)}(S_1 \ge n)\Big)^{1/q}.$$

Recall that d is non-expanding and $S_1 = \tau$ satisfies (3.31), hence

$$\mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{S_1 \ge n} \le \Big(d(x,y)\Big)^{1/p} \Big(V(x) + V(y)\Big)^{1/q} \chi^{n/q}.$$

Subsequent terms in (3.32) are estimated in a similar way, but some new technicalities arise. Namely, we have

$$\mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{n \in (S_k, S_{k+1}]} \le \Big(\mathsf{E}^Z_{(x,y)} d\big(Z^1_n, Z^2_n\big) 1_{S_k > n}\Big)^{1/p} \Big(\mathsf{P}^Z_{(x,y)}(S_{k+1} \ge n)\Big)^{1/q}.$$

The chain Z visits the set B, where d is contracting, at least k times on the set $\{S_k < n\}$, hence

$$\mathsf{E}^Z_{(x,y)} d\big(Z^1_n, Z^2_n\big) 1_{S_k < n} \le (1-c)^k d(x,y).$$

On the other hand, there exists C such that

$$\mathsf{E}^Z_{(x,y)} \chi^{-S_{k+1}} \le C^k \big(V(x) + V(y)\big).$$

One can deduce this from (3.31), the strong Markov property, and the simple observation that

$$\mathsf{E}\big[V^Z(Z_{S_k+1}) \big| \mathscr{F}_{S_k}\big] \le C.$$

Then for any $\gamma \in (0,1)$

$$\mathsf{P}^Z_{(x,y)}(S_{k+1} \ge n) \le C^{k\gamma} \big(V(x) + V(y)\big)^{\gamma} \chi^{-\gamma n}.$$

Taking γ small enough, we can guarantee that $(1-c)^{1/p} C^{\gamma/q} < 1$, and then

$$\sum_{k=1}^{\infty} \mathsf{E}^Z_{(x,y)} d_p\big(Z^1_n, Z^2_n\big) 1_{n \in (S_k, S_{k+1}]} \le \chi^{-\gamma n/q} d_p(x,y) \underbrace{\big(V(x) + V(y)\big)^{\gamma/q}}_{\le \left(V(x)+V(y)\right)^{1/q}} \underbrace{\sum_{k=1}^{\infty} (1-c)^{k/p} C^{k\gamma/q}}_{<\infty},$$

which proves the required bound with $\kappa = \chi^{-\gamma/q}$. □

Further References: [19], [20] (the general theory of irreducible Markov chains); [10] (Harris-type theorems); [12] (lecture notes).

4 The central limit theorem for Markov systems

4.1 Main problem: Formulation, motivation, and preliminaries

In this lecture we explain several practical methods to prove the CLT

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} f(X_j) \Rightarrow \mathsf{N}(0,\sigma_f^2), \tag{3.33}$$

where X is a fixed Markov chain and $f : \mathbb{X} \to \Re$ is a certain function ("observable"). Such a statement is a cornerstone to various critically important steps in the statistical analysis, simulation methods, in particular

- ◇ asymptotical normality of statistical estimators;
- ◇ confidence intervals for unknown parameters;
- ◇ accuracy bounds for Monte-Carlo simulation.

We plug in (3.33) into a more general "auto-regressive" model

$$\zeta_n^\varepsilon = \sum_{k=1}^{n} \xi_k^\varepsilon, \tag{3.34}$$

with a sequence $\{\xi_k^\varepsilon\}$ adapted to a filtration $\mathbb{F}^\varepsilon = \{\mathscr{F}_k^\varepsilon, k \geq 0\}$ and the main conditions formulated in the terms of dependence of ξ_k^ε w.r.t. $\mathscr{F}_j^\varepsilon$ for $k > j$.
Important particular cases are:

- ◇ Independent summands: ξ_k^ε is independent of $\mathscr{F}_j^\varepsilon$, $k > j$.
- ◇ Martingale setting: $\{\xi_k^\varepsilon\}$ is a *martingale difference,*

$$\mathsf{E}[\xi_k^\varepsilon | \mathscr{F}_{k-1}] = 0, \quad k \geq 0.$$

- ◇ Markov setting: taking

$$\xi_k^\varepsilon = \sqrt{\varepsilon} f(X_k), \qquad \varepsilon = n^{-1}, \quad \mathscr{F}_k^\varepsilon = \mathscr{F}_k = \sigma(X_j, j \leq k),$$

  we plug the original model (3.4) into the AR model (3.34).

We take in (3.34) $n \sim \varepsilon^{-1}$ and enrich the model by considering a family

$$Y^\varepsilon(t) := \zeta^\varepsilon_{[t/\varepsilon]} = \sum_{k \le \varepsilon^{-1} t} \xi^\varepsilon_k, \qquad t \ge 0.$$

4.2 Prelude 1: The martingale CLT

Theorem 3.33 Let the family $\{\xi^\varepsilon_k\}$ satisfy the following conditions:

- ◇ (asymptotic martingale property):

$$\mathsf{E}\big[\xi^\varepsilon_k \big| \mathscr{F}_{k-1}\big] = \varepsilon \gamma^\varepsilon_k, \qquad k \ge 0$$

  and

$$\sup_k \mathsf{E}|\gamma^\varepsilon_k|^2 \to 0;$$

- ◇ (LLN for conditional variances): The random variables $\beta^\varepsilon_k := \mathsf{E}\big[(\xi^\varepsilon_k)^2 \big| \mathscr{F}_{k-1}\big]$ are well defined and for some $\sigma \ge 0$

$$\sum_{k \le \varepsilon^{-1} t} \beta^\varepsilon_k \to t\sigma^2, \qquad t \ge 0$$

  in mean sense;

- ◇ (Lyapunov condition): For some $p > 2$,

$$\mathsf{E}|\xi^\varepsilon_k|^p \le C\varepsilon^{p/2}, \qquad k \ge 1.$$

Then for $t_1, \dots, t_m$,

$$\big(Y^\varepsilon(t_1), \dots, Y^\varepsilon(t_m)\big) \Rightarrow \sigma\big(W(t_1), \dots, W(t_m)\big).$$

Proof. Since

$$\mathsf{E}\big(Y^\varepsilon(t) - Y^\varepsilon(s)\big)^2 = \mathsf{E} \sum_{\varepsilon^{-1} s < k \le \varepsilon^{-1} t} \Big(\beta^\varepsilon_k + (\gamma^\varepsilon_k)^2\Big) \to (t-s)\sigma^2,$$

the family $\{Y^\varepsilon\}$ is weakly compact (in the sense of convergence of finite-dimensional distributions). That is, for every sequence there exists a subsequence such that for any

$t_1,\dots,t_m$,

$$\big(Y^{\varepsilon_r}(t_1),\dots,Y^{\varepsilon_r}(t_m)\big) \Rightarrow \big(Y(t_1),\dots,Y(t_m)\big).$$

Hence the main problem is to identify uniquely a limit point Y as σW.

For a fixed $f \in C_b^3$, write

$$f(\zeta_k^\varepsilon) - f(\zeta_{k-1}^\varepsilon) = f'(\zeta_{k-1}^\varepsilon)\xi_k^\varepsilon + \frac{1}{2}f''(\zeta_{k-1}^\varepsilon)(\xi_k^\varepsilon)^2 + \chi_k^\varepsilon. \tag{3.35}$$

Applying, if necessary, the Hölder inequality, we can assume that p in the Lyapunov condition is ≤ 3. Since f'', f''' are bounded, we have that f'' is Hölder continuous with the index $p-2 \le 1$. Then

$$\left|f(y+z) - f(y) - f'(y)z - \frac{1}{2}z^2 f''(z)\right| = \left|z^2 \int_0^1 \int_0^t \big(f''(y+sz) - f''(y)\big)\,\mathrm{d}s\,\mathrm{d}t\right| \le C|z|^p. \tag{3.36}$$

That is, $|\chi_k^\varepsilon| \le C|\xi_k^\varepsilon|^p$, and the Lyapunov condition gives

$$\mathsf{E}\sum_{k=1}^{\varepsilon^{-1}t} |\chi_k^\varepsilon| \to 0.$$

Combining this with the first two assumptions we get

$$\mathsf{E}f\big(Y^\varepsilon(t)\big) = f(0) + \frac{1}{2}\varepsilon\sigma^2 \sum_{k\le\varepsilon^{-1}t} \mathsf{E}f\big(Y^\varepsilon(k\varepsilon)\big)\beta_k^\varepsilon + r_f^\varepsilon(t) \tag{3.37}$$

with $r_f^\varepsilon(t) \to 0$. Passing to the limit w.r.t. the subsequence ε_r which corresponds to the limit process Y, we get

$$\mathsf{E}f\big(Y(t)\big) = f(0) + \frac{1}{2}\sigma^2 \int_0^t \mathsf{E}f''\big(Y(s)\big)\,\mathrm{d}s. \tag{3.38}$$

Take $f(y) = \mathrm{e}^{i\lambda y}$ with some $\lambda \in \Re$, then $f'' = -\lambda^2 f$ and the function $m(t) = \mathsf{E}f\big(Y(t)\big)$ satisfies the linear ODE

$$m'(t) = -\frac{1}{2}\sigma^2\lambda^2 m(t), \qquad m(0) = 1,$$

and thus

$$\mathsf{E}\mathrm{e}^{i\lambda Y(t)} = \mathrm{e}^{-\sigma^2\lambda^2/2},$$

which proves that $Y(t) \sim \mathsf{N}(0, t\sigma^2)$. The proof for m-tuple $\big(Y(t_1),\dots,Y(t_m)\big)$ is similar and omitted. $\square$

4.3 Prelude 2: The LLN in the weakly ergodic Markov setting

From now on, we assume a MC X with a unique IPM π be fixed, and the following "stabilization" property to hold true:

$$d\big(P_n(x,\mathrm{d}y),\pi(\mathrm{d}y)\big) \le R_n V(x), \qquad n \ge 0. \tag{3.39}$$

Here d is a distance on $\mathscr{P}(\mathbb{X})$ which corresponds to a distance-like function d, $\{R_n\}$ is a sequence which has the meaning of the stabilization rate, V is a function which has the meaning of a "penalty" brought to the stabilization bound by an initial position of the process. We assume $R_n \to 0$.

Theorem 3.34

1) Let X be stationary and satisfy (3.39). Let $f \in H_{\gamma,W}(\mathbb{X},d)$ and

$$\int_{\mathbb{X}} f^2 \,\mathrm{d}\pi < \infty, \qquad \int_{\mathbb{X}} V^2 \,\mathrm{d}\pi < \infty, \qquad \int_{\mathbb{X}} W^2 \,\mathrm{d}\pi < \infty.$$

Then

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \int_{\mathbb{X}} f \,\mathrm{d}\pi \tag{3.40}$$

in mean square.

2) Let, for some $\gamma > 0$, $W \in L_2(\mathbb{X},\pi)$ the class $H_{\gamma,W}(\mathbb{X},d)$ be dense in $L_1(\mathbb{X},\pi)$. Then the LLN (3.40) holds true in L_1 for any $L_1(\mathbb{X},\pi)$.

Proof. Statement 1) can be proved using essentially the same argument with the one used in Section 2.4. Namely, using (3.20) and (3.39) we get the following analogue of (3.21):

$$\begin{aligned}&\Big|\mathrm{Cov}\big(f(X_j),f(X_k)\big)\Big| \\ &\qquad \le 2^{1-\gamma} R^{\gamma}_{|k-j|} \|f\|_{d,\gamma,W} \left(\int_{\mathbb{X}} f^2 \,\mathrm{d}\pi\right)^{1/2} \left(\int_{\mathbb{X}} V^2 \,\mathrm{d}\pi\right)^{\gamma/2} \left(\int_{\mathbb{X}} W^2 \,\mathrm{d}\pi\right)^{(1-\gamma)/2}.\end{aligned}$$

Since $R_n^{\gamma} \to 0$, $n \to \infty$, this yields (3.40).

To prove 2), recall some notation used in the Birhkoff theorem. Let X be a strictly stationary sequence, and random variable ξ have the form

$$\xi = F(X_0,\dots,X_m), \tag{3.41}$$

where $m \geq 1$ is arbitrary and F is such that $\mathsf{E}|\xi| < \infty$. Denote the shift operator θ_n on ξ by

$$\theta_n \xi = F(X_n, \dots, X_{n+m}),$$

then θ_n is both, an L_1 and L_2-isometry, and thus is uniquely extended to the entire $L_1\big(\Omega, \sigma(X), \mathsf{P}\big)$. *The Birkhoff theorem* states that, for any $\xi \in L_1\big(\Omega, \sigma(X), \mathsf{P}\big)$,

$$\frac{1}{n}\sum_{k=1}^{n} \theta_k \xi \to \mathsf{E}[\xi \,|\, \mathscr{J}_X], \qquad n \to \infty \tag{3.42}$$

a.s. and in mean sense, where $\mathscr{J}_X$ is the invariant σ-algebra for the sequence X, which consists of all $A \in \sigma(X)$ such that

$$\theta_n 1_A = 1_A \quad \text{a.s.}, \qquad n \in \mathbb{Z}.$$

In addition, the convergence holds true in L_2 if $\mathsf{E}\xi^2 < \infty$. A strictly stationary sequence X is called *ergodic*, if its invariant σ-algebra $\mathscr{J}_X$ is degenerate:

$$A \in \mathscr{J}_X \Rightarrow \mathsf{P}(A) \in \{0, 1\}.$$

The ergodicity of X means that the right hand side of (3.42) equals to $\mathsf{E}\xi$, and to prove the ergodicity it is sufficient to prove such a statement for $\xi \in \mathscr{K}$ for some class $\mathscr{K}$, dense in $L_2\big(\Omega, \sigma(X), \mathsf{P}\big)$.

We will do that for the class of square integrable random variables which have representation (3.41). To do that, it is enough to prove that

$$\mathrm{Cov}(\theta_n \xi, \xi) \to 0, \qquad n \to \infty. \tag{3.43}$$

If $\xi = f(X_0)$ with $f \in H_{\gamma,W}(\mathbb{X}, d)$, this follows directly from the estimate for the covariance from the first part of the proof. For $f \in L_2(\mathbb{X}, \pi)$, consider an L_2-approximating sequence of functions $\{f_k\} \subset H_{\gamma,W}(\mathbb{X}, d)$, and denote $\xi_k = f_k(X_0)$. Then

$$\big|\mathrm{Cov}(\theta_n \xi, \eta)\big| \leq \big|\mathrm{Cov}(\theta_n \xi - \theta_n \xi_k, \eta)\big| + \|f - f_k\|_{L_2(\mathbb{X},\mu)} \|g\|_{L_2(\mathbb{X},\mu)},$$

hence

$$\limsup_{n\to\infty} \big|\mathrm{Cov}(\theta_n \xi, \eta)\big| \leq \|f - f_k\|_{L_2(\mathbb{X},\mu)} \|g\|_{L_2(\mathbb{X},\mu)}.$$

Taking $k \to \infty$ we complete the proof of (3.43) in this case. The general case is essentially the same, since for any ξ of the form (3.41) we have

$$\mathrm{Cov}(\theta_n \xi, \eta) = \mathrm{Cov}\big(\theta_{n-m} \xi^{(m)}, \eta\big), \qquad n \geq 2m,$$

where

$$\xi^{(m)} = \mathsf{E}[\xi | \mathscr{F}_0] = f^{(m)}(X_0).$$

Summarizing all the above, we see that X is ergodic, which competes the proof of Statement 2). □

4.4 Reduction of the Markov setting to the martingale CLT: the corrector term approach

Theorem 3.35 Let a stationary Markov chain X satisfy (3.39), and a function $f \in H_{\gamma,W}(\mathbb{X}, d)$ be centered:

$$\int_{\mathbb{X}} f(x) \pi(\mathrm{d}x) = 0. \tag{3.44}$$

Assume that $\sum_k r_k^\gamma < \infty$, and for some $p > 2$ $V \in L_p(\mathbb{X}, \mu)$, $W \in L_p(\mathbb{X}, \mu)$. Then the CLT (3.4) holds true with

$$\sigma^2 = \mathsf{E} f^2(X_0) + 2 \sum_{k=1}^{\infty} \mathsf{E} f(X_k) f(X_0). \tag{3.45}$$

Remark 3.36 The centering condition (3.44) is a natural pre-requisite to the CLT (3.4). Note that this condition does not imply that

$$\xi_k^\varepsilon = \sqrt{\varepsilon} f(X_k)$$

is a martingale difference: we have

$$\mathsf{E}[\xi_k^\varepsilon | \mathscr{F}_{k-1}] = \sqrt{\varepsilon} \int_{\mathbb{X}} f(x) P(X_{k-1}, \mathrm{d}x),$$

and the martingale property requires

$$\int_{\mathbb{X}} f(x) P(x_0, \mathrm{d}x) = 0, \qquad x_0 \in \mathbb{X}, \tag{3.46}$$

which is much more restrictive than (3.44).

The proof of Theorem 3.35 is based on the following auxiliary construction. Define the *(extended) potential* of the function f by

$$\mathscr{R}f(x) = \sum_{k=0}^{\infty} \mathsf{E}_x f(X_k), \qquad x \in \mathbb{X}. \tag{3.47}$$

Proposition 3.37 Let X satisfy (3.39) and a function $f \in L_2(\mathbb{X}, \pi)$ be centered. Assume also that for some γ, W, $f \in H_{\gamma,W}(\mathbb{X}, d)$,

$$\sum_n R_n^\gamma < \infty,$$
$$\int_{\mathbb{X}} V^2 \, \mathrm{d}\pi < \infty, \qquad \int_{\mathbb{X}} W^2 \, \mathrm{d}\pi < \infty.$$

Then the series (3.47) converges in $L_2(\mathbb{X}, \pi)$ sense, and if $\{X_k\}$ is stationary the sequence,

$$M_k = \sum_{j=0}^{k-1} f(X_j) + \mathscr{R}f(X_k), \qquad k \geq 0,$$

is a martingale.

For the proof we refer to Proposition 3.9 in [12].

Proof of Theorem 3.35. Denote

$$\widetilde{\mathscr{R}}f(x) = \mathscr{R}f(x) - f(x) = \sum_{k=1}^{\infty} \mathsf{E}_x f(X_k),$$

then

$$\zeta_k^\varepsilon = \sqrt{\varepsilon} \sum_{j \leq k} f(X_k) = \sqrt{\varepsilon} M_k - \sqrt{\varepsilon} \widetilde{\mathscr{R}} f(X_k).$$

Since $\widetilde{\mathscr{R}} f \in L_2(\mathbb{X}, \pi)$, if X is stationary then

$$\sqrt{\varepsilon} \widetilde{\mathscr{R}} f(X_k) \to 0$$

in mean square, hence the limit behavior of ζ^ε is the same as for the martingale sequence

$$\widetilde{\zeta}_k^\varepsilon = \sqrt{\varepsilon} M_k - \sqrt{\varepsilon}\widetilde{\mathscr{R}} f(X_0) = \sum_{j=1}^k \widetilde{\xi}_j^\varepsilon,$$
$$\widetilde{\xi}_k^\varepsilon = \sqrt{\varepsilon}\big(f(X_k) + \widetilde{\mathscr{R}} f(X_k) - \widetilde{\mathscr{R}} f(X_{k-1})\big) = \sqrt{\varepsilon}\big(\mathscr{R} f(X_k) - \widetilde{\mathscr{R}} f(X_{k-1})\big).$$

We have

$$\mathsf{E}\big[(\widetilde{\xi}_k^\varepsilon)^2 \big| \mathscr{F}_{k-1}^\varepsilon\big] = \varepsilon g(X_{k-1}),$$
$$g(x) = \int_{\mathbb{X}} \big(\mathscr{R} f(y)\big)^2 P(x, \mathrm{d}y) - \big(\widetilde{\mathscr{R}} f(x)\big)^2.$$

By Theorem 3.34,

$$\varepsilon \sum_{k \le \varepsilon^{-1} t} g(X_k) \to t\sigma^2, \qquad t \ge 0.$$

Hence the required statement follows by the martingale CLT; note that

$$\sigma^2 = \int_{\mathbb{X}} \big(\mathscr{R} f(y)\big)^2 \pi(\mathrm{d}y) - \int_{\mathbb{X}} \big(\widetilde{\mathscr{R}} f(x)\big)^2 \pi(\mathrm{d}x) = \mathsf{E} f^2(X_0) + 2\sum_{k \ge 1} \mathsf{E} f(X_0) f(X_k). \qquad \square$$

4.5 Alternative method: Blocks and delay

Theorem 3.35 is formulated under the very mild set of assumptions, which is a clearly seen advantage. A hidden disadvantage of this theorem is that the method of its proof is rather implicit: For the martingale CLT to be applied, one needs to prove the LLN for conditional variances, and for that purpose we used the Birkhoff theorem and the density argument. Such an implicit proof may be inconvenient, for instance, one is interested in further details, rates of convergence, etc. With this purpose in mind, here we present an alternative method of proof.

Assume the general autoregressive model (3.34) satisfies the following principal condition.

$$\mathsf{E}\big[\xi_{k+n}^\varepsilon \big| \mathscr{F}_k^\varepsilon\big] = \varepsilon^{1/2} r_n \upsilon_{k,n}^\varepsilon, \qquad k \ge 0, \quad n \ge 1, \tag{$\mathbf{H}_0$}$$

where for some $\kappa > 1$,

$$r_n \le C n^{-\kappa}. \tag{3.48}$$

and for some $p > 2$

$$\mathsf{E}\left|\xi_k^\varepsilon\right|^p \le C\varepsilon^{p/2}, \qquad \mathsf{E}\left|\upsilon_{k,n}^\varepsilon\right|^p \le C. \tag{3.49}$$

This condition is well designed to cover the Markov models which satisfy the centering condition (3.44), but fail the conditional centering condition (3.46), and thus can not be included into the martingale CLT framework immediately.

Indeed, for $\xi_\varepsilon = \sqrt{\varepsilon} f(X_k)$, with Markov X,

$$\mathsf{E}\left[\xi_{k+n}^\varepsilon \middle| \mathscr{F}_k^\varepsilon\right] = \sqrt{\varepsilon} \int_{\mathbb{X}} f(y) P_n(X_n, \mathrm{d}y),$$

and $(\mathbf{H}_0)$ holds true with

$$\upsilon_{k,n}^\varepsilon = (r_n)^{-1} \int_{\mathbb{X}} f(y) P_n(X_n, \mathrm{d}y).$$

Let

- ◇ X be stationary such that (3.39) holds;
- ◇ $f \in H_{\gamma,W}$ be centered.

Then we take $r_n = R_n^\gamma$ and by Proposition 3.23,

$$\left|\upsilon_{k,n}^\varepsilon\right| \le C \left(\int_{\mathbb{X}} V(y)\, P_n(X_n, \mathrm{d}y)\right)^\gamma \left(\int_{\mathbb{X}} W(y)\, P_n(X_n, \mathrm{d}y) + \int_{\mathbb{X}} W \,\mathrm{d}\pi\right)^{1-\gamma}.$$

Then

$$\int_{\mathbb{X}} (V^p + W^p)\,\mathrm{d}\pi < \infty \tag{3.50}$$

provides the second bound in (3.49):

$$\begin{aligned}
&\mathsf{E}\left|\upsilon_{k,n}^\varepsilon\right|^p \\
&\le C \left(\mathsf{E}\int_{\mathbb{X}} V^p(y)\, P_n(X_n, \mathrm{d}y)\right)^\gamma \left(2^{p-1}\mathsf{E}\int_{\mathbb{X}} W^p(y)\, P_n(X_n, \mathrm{d}y) + 2^{p-1}\int_{\mathbb{X}} W^p \,\mathrm{d}\pi\right)^{(1-\gamma)} \\
&= 2^{p(1-\gamma)} C \left(\int_{\mathbb{X}} W^p \,\mathrm{d}\pi\right)^\gamma \left(\int_{\mathbb{X}} W^p \,\mathrm{d}\pi\right)^{(1-\gamma)}.
\end{aligned}$$

The first bound in (3.49) follows directly from the assumption

$$\int_{\mathbb{X}} |f|^p \, \mathrm{d}\pi < \infty, \tag{3.51}$$

and (3.48) holds true if

$$R_n \le C n^{-\kappa/\gamma}. \tag{3.52}$$

In addition to ($\mathbf{H}_0$), we assume the following.

$$\text{for each } j,k,n \ge 0, \quad \mathsf{E}\big[\xi^\varepsilon_{k+n}\xi^\varepsilon_{k+n+j}\big|\mathscr{F}^\varepsilon_k\big] = \varepsilon\beta_j + \varepsilon\varkappa^\varepsilon_{k,j,n}, \tag{$\mathbf{H}_1$}$$

where

$$\sum_j |\beta_j| < \infty$$

and for each j

$$\sup_k \mathsf{E}|\varkappa^\varepsilon_{k,j,n}| \to 0, \qquad n \to \infty.$$

In the Markov setting, this assumption can be verified easily, as well.

Theorem 3.38 If ($\mathbf{H}_0$) and ($\mathbf{H}_1$) hold true, then for $t_1,\dots,t_m$,

$$\big(Y^\varepsilon(t_1),\dots,Y^\varepsilon(t_m)\big) \Rightarrow \sigma\big(W(t_1),\dots,W(t_m)\big)$$

with

$$\sigma^2 = \beta_0 + 2\sum_{j\ge 1}\beta_j.$$

Proof. We put the summands to the blocks of the size $D_\varepsilon = [\varepsilon^{-\rho}]+1$ with $\rho > 1/2$:

$$\widetilde{\xi}^\varepsilon_k = \sum_{l=(k-1)D_\varepsilon+1}^{kD_\varepsilon} \xi^\varepsilon_l.$$

Now we have $\zeta^\varepsilon_{\varepsilon^{-1}t}$ represented as a sum of $\sim \varepsilon^{-1}/D_\varepsilon \sim \varepsilon^{-1+\rho} \ll \varepsilon^{-1/2}$ summands. For each of these summands, denoting $\widetilde{\mathscr{F}}_{k-1} = \mathscr{F}_{(k-1)D_\varepsilon}$, we have

$$\mathsf{E}\big[\widetilde{\xi}^\varepsilon_k\big|\widetilde{\mathscr{F}}_{k-1}\big] = \sqrt{\varepsilon}\sum_{l=(k-1)D_\varepsilon+1}^{kD_\varepsilon} r_{l-(k-1)D_\varepsilon}\,\upsilon^\varepsilon_{(k-1)D_\varepsilon,l-(k-1)D_\varepsilon} = \sqrt{\varepsilon}\,\widetilde{\upsilon}^\varepsilon_k,$$

and

$$\mathsf{E}\big|\widetilde{\upsilon}_k^\varepsilon\big| \le C.$$

That is, the first assertion of Theorem 3.33 holds true, with the obvious changes in the notation, e.g. ε therein should be replaced by $\widetilde{\varepsilon} = \varepsilon^{1-\rho}$. The second assertion in this theorem easily follows from $(\mathbf{H}_1)$. Namely,

$$\begin{aligned}\mathsf{E}\big[\big(\widetilde{\xi}_k^\varepsilon\big)^2\big|\widetilde{\mathscr{F}}_{k-1}\big] &= \sum_{i,j=1}^{D_\varepsilon} \mathsf{E}\big[\xi^\varepsilon_{(k-1)D_\varepsilon+i}\xi^\varepsilon_{(k-1)D_\varepsilon+i}\big|\widetilde{\mathscr{F}}_{k-1}\big] \\ &= \underbrace{\varepsilon D_\varepsilon}_{\sim\widetilde{\varepsilon}}\left(\beta_0 + 2\sum_{j=1}^{D_\varepsilon}\beta_j\right) + \underbrace{\varepsilon\sum_{i,j=1}^{D_\varepsilon}\mathsf{E}\big|\varkappa^\varepsilon_{k,j,n}\big|}_{\ll\widetilde{\varepsilon}}.\end{aligned}$$

The third assertion is the most difficult one to verify here. It is guaranteed by the following statement, whose proof we will give later

Proposition 3.39 Under conditions of Theorem (3.38), for every $T > 0$ there exists a constant C_T such that

$$\mathsf{E}\big|Y^\varepsilon(t) - Y^\varepsilon(s)\big|^p \le C_T\big(|t-s| + \varepsilon\big)^{p/2}, \qquad |s-t| \le T. \tag{3.53}$$

By this proposition, for $k \le \widetilde{\varepsilon}^{-1}T$

$$\mathsf{E}\big|\widetilde{\xi}_k^\varepsilon\big|^p \le C_T\widetilde{\varepsilon}^{p/2},$$

which gives the third assumption of Theorem 3.33. Now the required statement follows directly from Theorem 3.33. □

We note that the "blocking" construction presented in the proof has a different spirit from the classical "Bernstein Blocks"; cf. [1] and [11], Chapter 18.

Proof of Proposition 3.39. Since Y^ε is piece-wise constant, it is sufficient to prove (3.53) only for s, t of the form $i\varepsilon$, $k\varepsilon$. To simplify the notation, we prove the required inequality only for $i = 0$ assuming $\zeta_0^\varepsilon = 0$. That is, our aim is to prove the inequality

$$\mathsf{E}\big|\zeta_k^\varepsilon\big|^p \le C_T(\varepsilon k)^{p/2}, \qquad \varepsilon k \le T. \tag{3.54}$$

Denote

$$f(x) = |x|^p, \qquad g(x) = p|x|^{p-1}\operatorname{sign} x, \qquad h(x) = f''(x) = p(p-1)|x|^{p-2}.$$

It follows from the elementary inequality

$$|1+t|^p \le 1 + pt + \frac{p(p-1)}{2}t^2 + C|t|^p, \qquad t \in \Re$$

that

$$f(\zeta_k^\varepsilon) \le f(\zeta_{k-1}^\varepsilon) + g(\zeta_{k-1}^\varepsilon)\xi_k^\varepsilon + \frac{1}{2}h(\zeta_{k-1}^\varepsilon)(\xi_k^\varepsilon)^2 + C|\xi_k^\varepsilon|^p.$$

Therefore

$$\mathsf{E} f(\zeta_k^\varepsilon) \le \sum_{j=1}^k \mathsf{E} g(\zeta_{j-1}^\varepsilon)\xi_j^\varepsilon + \frac{1}{2}\sum_{j=1}^k \mathsf{E} h(\zeta_{j-1}^\varepsilon)(\xi_j^\varepsilon)^2 + C\sum_{j=1}^k \mathsf{E}|\xi_j^\varepsilon|^p. \tag{3.55}$$

We perform a further transformation of the first sum on the right hand side of (3.55) using the "time delay" trick. To clarify the exposition, we first explain the idea this transformation is based on. A direct estimation of the expectations involved into the first sum is hardly possible, because under our principal Assumption ($\mathbf{H}_0$) the decomposition of the conditional expectation for ξ_j^ε w.r.t. $\mathscr{F}_{j-1}^\varepsilon$ contains the term $\varepsilon^{1/2} r_1 \upsilon_{j-1,1}^\varepsilon$, which has the order $\varepsilon^{1/2} \gg \varepsilon$. Our aim would be to change ζ_{j-1}^ε to $\zeta_{(j-D_\varepsilon)_+}^\varepsilon$; that is, we to perform a delay of the time variable in ζ^ε. Because $\varepsilon^{1/2} r_{D_\varepsilon} \le C\varepsilon^{1/2+\rho\kappa} \ll \varepsilon$, the (new) conditional expectation of ξ_j^ε will be well manageable; this is actually our main tool to utilize the "long run stabilization" property of the model.

Let us develop in details the procedure outlined above. We have

$$g(x+\delta) - g(x) = h(x)\delta + \rho(x,\delta),$$

where

$$\rho(x,\delta) = \delta\big(h(x+\theta) - h(x)\big)$$

with θ being an intermediate point between x and $x+\delta$. Recall that $h(x) = C|x|^{p-2}$ and $p-2 \in (0,1]$. Then $\big|h(x+\theta) - h(x)\big| \le C|\theta|^{p-2}$, which yields $\big|\rho(x,\delta)\big| \le C|\delta|^{p-1}$. That is,

$$g(\zeta_i^\varepsilon) = g(\zeta_{i-1}^\varepsilon) + h(\zeta_{i-1}^\varepsilon)\xi_i^\varepsilon + \rho_i^\varepsilon, \qquad \big|\rho_i^\varepsilon\big| \le C\big|\xi_i^\varepsilon\big|^{p-1},$$

and thus

$$g(\zeta^\varepsilon_{j-1}) = g(\zeta^\varepsilon_{(j-D_\varepsilon)_+}) + \sum_{i=(j-D_\varepsilon)_+ +1}^{j-1} \big(h(\zeta^\varepsilon_{i-1})\xi^\varepsilon_i + \rho^\varepsilon_i\big).$$

We re-arrange (3.55) and then analyze separately each sum which appear in the right hand side. We have

$$\begin{aligned}\mathsf{E}f(\zeta^\varepsilon_k) \le \sum_{j=1}^k \mathsf{E}g\big(\zeta^\varepsilon_{(j-D_\varepsilon)_+}\big)\xi^\varepsilon_j + \sum_{j=1}^k \sum_{i=(j-D_\varepsilon)_+ +1}^{j-1} \mathsf{E}h\big(\zeta^\varepsilon_{i-1}\big)\xi^\varepsilon_i\xi^\varepsilon_j \\ + \sum_{j=1}^k \sum_{i=(j-D_\varepsilon)_+ +1}^{j-1} \mathsf{E}\rho^\varepsilon_i\xi^\varepsilon_j + \frac{1}{2}\sum_{j=1}^k \mathsf{E}h\big(\zeta^\varepsilon_{j-1}\big)\big(\xi^\varepsilon_j\big)^2 + C\sum_{j=1}^k \mathsf{E}|\xi^\varepsilon_j|^p \\ =: \sum_{l=1}^5 S^{l,\varepsilon}_k. \quad (3.56)\end{aligned}$$

Recall that $f(0) = 0$ and ζ^ε_0 is assumed to be equal 0. That is, each term in the sum $S^{1,\varepsilon}_k$ with $j \le D_\varepsilon$ equals 0. For all the other terms, we have $\zeta^\varepsilon_{(j-D_\varepsilon)_+} = \zeta^\varepsilon_{j-D_\varepsilon}$ and thus by $(\mathbf{H}_0)$,

$$S^{1,\varepsilon}_k = \sum_{j=D_\varepsilon+1}^k \mathsf{E}g\big(\zeta^\varepsilon_{j-D_\varepsilon}\big)\Big(\varepsilon^{1/2}\upsilon^\varepsilon_{j-D_\varepsilon,D_\varepsilon}r_{D_\varepsilon} + \varepsilon\gamma^\varepsilon_{j-D_\varepsilon,D_\varepsilon}\Big).$$

Recall that $|g(x)|^{p/(p-1)} = Cf(x)$, and denote $\mu^\varepsilon_j = \mathsf{E}f(\zeta^\varepsilon_j)$. Then by (3.49) and the Hölder inequality,

$$S^{1,\varepsilon}_k \le C\sum_{j=D_\varepsilon+1}^k \big(\mu^\varepsilon_{j-D_\varepsilon}\big)^{p/(p-1)}\big(\varepsilon^{1/2}r_{D_\varepsilon} + \varepsilon\big) \le C\varepsilon\sum_{j=D_\varepsilon+1}^k \big(\mu^\varepsilon_{j-D_\varepsilon}\big)^{(p-1)/p};$$

the latter inequality holds true because $\varepsilon^{1/2}r_{D_\varepsilon} \le C\varepsilon^{1/2+\rho p_0} \ll \varepsilon$.

Next, by $(\mathbf{H}_0)$

$$\mathsf{E}h\big(\zeta^\varepsilon_{i-1}\big)\xi^\varepsilon_i\xi^\varepsilon_j = \mathsf{E}h\big(\zeta^\varepsilon_{i-1}\big)\xi^\varepsilon_i\big(\varepsilon^{1/2}\upsilon^\varepsilon_{i,j-i}r_{j-i} + \varepsilon\gamma^\varepsilon_{i,j-i}\big).$$

We have $|h(x)|^{p/(p-2)} = Cf(x)$, hence by the Lyapunov condition and (3.49) we get

$$\begin{aligned}\mathsf{E}h\big(\zeta^\varepsilon_{i-1}\big)\xi^\varepsilon_i\upsilon^\varepsilon_{i,j-i} &\le C\varepsilon^{1/2}\big(\mu^\varepsilon_{i-1}\big)^{(p-2)/p}, \\ \mathsf{E}h\big(\zeta^\varepsilon_{i-1}\big)\xi^\varepsilon_i\gamma^\varepsilon_{i,j-i} &\le C\varepsilon^{1/2}\big(\mu^\varepsilon_{i-1}\big)^{(p-2)/p}.\end{aligned}$$

Taking into account that $\sum_l r_l < \infty$ and $D_\varepsilon \le C\varepsilon^{-1/2}$, we deduce that

$$S_k^{2,\varepsilon} \le C\varepsilon \sum_{i=1}^k \left(\mu_{i-1}^\varepsilon\right)^{(p-2)/p}.$$

Similarly,

$$\mathsf{E}\rho_i^\varepsilon \xi_j^\varepsilon = \mathsf{E}\rho_i^\varepsilon \left(\varepsilon^{1/2} \upsilon_{i,j-i}^\varepsilon r_{j-i} + \varepsilon \gamma_{i,j-i}^\varepsilon\right),$$

and because $|\rho_i^\varepsilon| \le C|\xi_i^\varepsilon|^{p-1}$,

$$\mathsf{E}\rho_i^\varepsilon \upsilon_{i,j-i}^\varepsilon \le \left(\mathsf{E}|\rho_i^\varepsilon|^{p/(p-1)}\right)^{(p-1)/p} \left(\mathsf{E}\left|\upsilon_{i,j-i}^\varepsilon\right|^p\right)^{1/p} \le C\varepsilon^{(p-1)/2},$$
$$\mathsf{E}\rho_i^\varepsilon \gamma_{i,j-i}^\varepsilon \le C\varepsilon^{(p-1)/2}.$$

Taking into account that $\sum_l r_l < \infty$ and $D_\varepsilon \le C\varepsilon^{-1/2}$, we deduce

$$S_k^{3,\varepsilon} \le C \sum_{i=1}^k \varepsilon^{p/2}.$$

We also have

$$S_k^{4,\varepsilon} \le C\varepsilon \sum_{i=1}^k \left(\mu_{i-1}^\varepsilon\right)^{(p-2)/p}$$

(the proof is omitted since it similar and simpler to the proof of the inequality for $S_k^{2,\varepsilon}$), and

$$S_k^{5,\varepsilon} \le C \sum_{i=1}^k \varepsilon^{p/2}$$

(which follows directly from the Lyapunov condition). We denote $M_k^\varepsilon = \max_{i\le k} \mu_i^\varepsilon$ and summarize the above estimates as follows:

$$M_k^\varepsilon \le C_* \sum_{j=1}^k \left(\varepsilon \left(M_{j-1}^\varepsilon\right)^{(p-1)/p} + \varepsilon \left(M_{j-1}^\varepsilon\right)^{(p-2)/p} + \varepsilon^{p/2}\right)$$

(we fix the constant C_* for the further reference needs). Now we use the induction to prove that, under a proper choice of $C_T \ge 1$, (3.54) holds true. Assume (3.54) to hold for

all $k' \le k$ $(\le \varepsilon^{-1}T)$, then by the above inequality

$$\begin{aligned} M_{k+1}^{\varepsilon} &\le C_* \sum_{j=1}^{k} \Big(\varepsilon C_T^{(p-1)/p} (\varepsilon j)^{(p-1)/2} + \varepsilon C_T^{p-2/p} (\varepsilon j)^{(p-2)/2} + \varepsilon^{p/2} \Big) \\ &\le C_* \sum_{j=1}^{k} \Big(\varepsilon C_T^{(p-1)/p} T^{1/2} (\varepsilon j)^{(p-2)/2} + \varepsilon C_T^{p-2/p} (\varepsilon j)^{(p-2)/2} + \varepsilon^{p/2} \Big) \\ &\le C_* C_T^{(p-1)/p} \big(T^{1/2} + 1 \big) \varepsilon^{p/2} \sum_{j=1}^{k} \Big(j^{(p-2)/2} + 1 \Big) \end{aligned}$$

(recall that we assumed $C_T \ge 1$). It is elementary to prove that for some $C_p > 0$

$$\sum_{j=1}^{k} \Big(j^{(p-2)/2} + 1 \Big) \le C_p (k+1)^{p/2}.$$

If $C_T \ge 1$ was chosen $\ge \big(C_* C_p (T^{1/2} + 1) \big)^{1/p}$, then (3.54) holds true for $k+1$, as well. This completes the entire proof. □

Further References: [11] (comprehensive presentation of limit theorems for dependent summands); [1] (the origin of the Bernstein block method); [12] (lecture notes).

• • • • • • • • •

Bibliography

[1] Bernstein, S. N.: *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann. **97**, 1–59 (1926).

[2] Billingsley, P.: *Convergence of probability measures*, 2nd ed., Wiley, New York (1999).

[3] Döblin, W.: *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples*, Thèse de doctorat: Sciences mathematiques, Paris (1938).

[4] Döblin, W.: *Éléments d'une théorie générale des chaînes simples constantes de Markoff*, Ann. Sci. E.N.S. **57**, 61–111 (1940).

[5] Dobrushin, R. L.: *Central limit theorem for nonstationary Markov chains*, Theor. Prob. Appl. **1**, 65–80; 329–383 (1956).

[6] Dudley, R. M.: *Real analysis and probability*. Cambridge University Press, Cambridge (2002).

[7] Feller, W.: *An introduction to probability theory and its applications*, Wiley, New York, Vol. 1, 3rd ed. (1968), Vol. 2, 2nd ed. (1971).

[8] Hairer, M.: *Ergodic properties of Markov processes*, Lecture Notes (2006). http://www.hairer.org/notes/Markov.pdf (Aug. 2017).

[9] Hairer, M.: *P@W course on the convergence of Markov processes*, Lecture Notes (2010). http://www.hairer.org/notes/Convergence.pdf (Aug. 2017).

[10] Hairer, M., Mattingly, J., Scheutzow, M.: *Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations*, Prob. Theory Rel. Fields **149**(1-2), 223–259 (2011).

[11] Ibragimov, I. A., Linnik, Yu.V.: *Independent and stationary sequences of random variables*, Wolters-Noordhoff, Groningen (1971).

[12] Kulik, A. M.: *Introduction to ergodic rates for Markov chains and processes, with applications to limit theorems*, Potsdam University Press, Potsdam (2015). https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-79360 (Aug. 2017).

[13] Kulik, A., Scheutzow, M.: *A coupling approach to Doob's theorem*, Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **26**, 83–92 (2015).

[14] Kulik, A., Scheutzow, M.: *Generalized couplings and convergence of transition probabilities*, arXiv:1512.06359 (Aug. 2017).

[15] Lindvall, T.: *Lectures on the coupling method*, Wiley, New York (1992).

[16] Markov, A. A.: *Rasprostranenie zakona bolshikh chisel na velichiny, zavisyashchie drug ot druga (The extension of the law of large numbers on mutually dependent variables)*, Isvestia Fiz.-mat. obsch. Kazan univ. **15**(4), 135–156 (1906). See also [18], pp. 339–361.

[17] Markov, A. A.: *Rasprostranenie predel'nyh teorem ischisleniya veroyatnostej na summu velichin svyazannyh v cep' (Extension of the limit theorems of probability theory to a sum of variables connected in a chain)*, Zapiski Akademii Nauk po Fiziko-matematicheskomu otdeleniyu, VIII seriya **25**(3) (1908). Translated into English in: R. A. Howard (ed.): *Dynamic probabilities systems*, vol. 1, Wiley, New York (1971), pp. 552–576. See also [18], pp. 364–397.

[18] Markov, A. A.: *Izbrannye trudy. Teoria chisel. Teoria veroyatnostei (Selected Works. Number Theory. Probability Theory)*, Yu. V. Linnik (ed.), Moscow-Leningrad (1951).

[19] Meyn, S. P.: Tweedie, R. L.: *Markov chains and stochastic stability*, Springer, London (1993).

[20] Numellin, E.: *General irreducible Markov chains and non-negative operators*, 2nd ed, Cambridge University Press, Cambridge (2004).

[21] Rachev, S. T., Rüschendorf, L.: *Mass transportation problems: Volume I: Theory*, Springer, Berlin (1998).

[22] Thorisson, H.: *Coupling, stationarity, and regeneration*. Springer, New York (2000).

[23] Veretennikov, A. Yu.: *Coupling method for Markov chains under integral Doeblin type condition*, Theory Stoch. Processes **8**(3–4), 383–391 (2002).

[24] Zolotarev, V. M.: *Sums of independent random variables: a modern approach*, Wiley, New York (1995).

The interdisciplinary workshop *Stochastic processes with applications in the natural sciences* was held in Bogotá, at Universidad de los Andes from December 5 to December 9, 2016. It brought together researchers from Colombia, Germany, France, Italy, Ukraine, who communicated recent progress in the mathematical research related to stochastic processes with application in biophysics. The present volume collects three of the four courses held at this meeting by Angelo Valleriani, Sylvie Rœlly and Alexei Kulik.
A particular aim of this collection is to inspire young scientists in setting up research goals within the wide scope of fields represented in this volume.

Alexei Kulik, Doctor of Sciences, is a Leading researcher at the Institute of Mathematics of Ukrainian National Academy of Sciences.

Sylvie Rœlly, Docteur en Mathématiques, is the head of the chair of Probability at the University of Potsdam.

Angelo Valleriani, PhD in high energy physics, is group leader of the team "Stochastic processes in complex and biological systems" from the Max-Planck-Institute of Colloids and Interfaces, Potsdam.

ISSN 2199-4951
ISBN 978-3-86956-414-2
9 783869 564142