Lectures on the Coupling Method

TORGNY LINDVALL

Department of Mathematics
Chalmers University of Technology
and Göteborg University
Göteborg, Sweden

DOVER PUBLICATIONS, INC.
Mineola, New York

About the Author

TORGNY LINDVALL teaches at the Department of Mathematics, Chalmers University of Technology and Göteborg University. He has been a visiting scholar at the University of Wisconsin-Madison, Stanford University, Cambridge University, and University of Zurich. He is a member of IMS and the Bernoulli Society. Dr. Lindvall received his PhD in mathematical statistics from Göteborg University.

Copyright

Published in the United Kingdom by David & Charles, Brunel House, Forde Close, Newton Abbot, Devon TQ12 4PU.

Bibliographical Note

This Dover edition, first published in 2002, is a corrected and unabridged republication of the work originally published in 1992 by John Wiley & Sons, Inc., New York.

Library of Congress Cataloging-in-Publication Data

Lindvall, Torgny, 1945–
Lectures on the coupling method / Torgny Lindvall.
p. cm.
"This Dover edition . . . is a corrected and unabridged republication of the work originally published in 1992 by John Wiley & Sons, Inc., New York"—t.p. verso
Includes bibliographical references and index.
ISBN 0-486-42145-7 (pbk.)
1. Probability measures. I. Title.

QA273.6 .L54 2002
519.2'4—dc21

2002022294

Manufactured in the United States of America
Dover Publications, Inc., 31 East 2nd Street, Mineola, N.Y. 11501

Preface

Suppose that some sort of comparison of probability measures on a measurable space is to be carried out. For that purpose it is sometimes possible, and then often rewarding, to construct random elements on a common probability space, with these measures as distributions, in such a way that the comparison may be carried out in terms of the random elements. Such a construction is called a coupling. This method has come to be used primarily for estimates of total variation distances but also works well to establish inequalities, as we shall see.

During the past two decades, the coupling method has developed as an important tool in probability theory. It has been very successful in the study of Markov and renewal process asymptotics, but the range of applications is growing quickly. One purpose of this book is to point out the wide range of uses, including topics from Poisson approximation of weakly dependent 0–1 variables to monotonicity of power functions of statistical tests. So far there has been no comprehensive account of couplings and their possibilities; this book is meant to fill that need.

The method is well suited for presentation in terms of simple examples. The right place for an outline of the book is following these examples, which are given in the Introduction rather than here. But for a quick impression of what is ahead, review the table of contents.

This book is intended to serve graduate courses and seminars in departments of mathematics, statistics, and operations research. Hopefully, it will also find use as a reference. The reader is assumed to be familiar with probability theory based on measures and Lebesgue integrals.

One does not write a book in a mathematical field without much support and encouragement from colleagues and friends. You are all remembered. Rather than providing an embarassingly long list of names, I mention only two: Hermann Thorisson and Chris Rogers. Their influence has been crucial. My grateful thoughts also go to the Academic Zürich for its hospitality during a sabbatical period in 1989.

Thanks also to Yumi Karlsson for help with the typing. Swift, to say the least.

TORGNY LINDVALL

Göteborg in August 1991

Contents

CHAPTER III. CONTINUOUS THEORY 67

Lectures on the Coupling Method

Introduction

1. Three examples. Let $X=(X_n)_0^\infty$ be a Markov chain with a countable state space (equal to $\mathbb{Z}_+$ = the nonnegative integers), governed by a transition matrix $P=(p_{ij})$, making it aperiodic, irreducible, and positive recurrent. It is a classical result that X approaches stationarity as $n\to\infty$, regardless of the initial distribution λ (that of X_0):

$$\mathbf{P}(X_n=j)=\sum_i \lambda_i p_{ij}^{(n)} \to \pi_j \quad \text{as} \quad n\to\infty, \tag{1.1}$$

where $\pi=(\pi_j)_0^\infty$ is the unique stationary distribution, satisfying $\pi=\pi P$, and $(p_{ij}^{(n)})=P^n$ is the n-step transition matrix. The coupling idea for a proof of (1.1) is the following: Introduce a parallel process $X'=(X_n')_0^\infty$, independent of X, governed by P and stationary; the latter property is achieved by letting X_0' have distribution π. Define $X''=(X_n'')_0^\infty$ by

$$X_n''=\begin{cases} X_n & \text{if} \quad n<T \\ X_n' & \text{if} \quad n\ge T, \end{cases}$$

where

$$T=\min\{k;\, X_k=X_k'\}\,.$$

We thus make a coupling of X and X' at T and call that the coupling time. At this instant you should not hesitate to believe that X and X'' are equally distributed (this is due to the strong Markov property). This implies that for each $j\in\mathbb{Z}_+$,

$$
\begin{aligned}
(1.2)\quad |\mathbf{P}(X_n = j) - \pi_j| &= |\mathbf{P}(X''_n = j) - \mathbf{P}(X'_n = j)| \\
&= |\mathbf{P}(X''_n = j, T \le n) + \mathbf{P}(X''_n = j, T > n) \\
&\quad - \mathbf{P}(X'_n = j, T \le n) - \mathbf{P}(X'_n = j, T > n)| \\
&\le \mathbf{P}(X_n = j, T > n) + \mathbf{P}(X'_n = j, T > n)
\end{aligned}
$$

since $\mathbf{P}(X''_n = j, T \le n) = \mathbf{P}(X'_n = j, T \le n)$. Hence

$$|\mathbf{P}(X_n = j) - \pi_j| \le 2 \cdot \mathbf{P}(T > n)\,,$$

so (1.1) follows if the coupling is successful (i.e., if $T < \infty$ a.s.). Actually, this inequality holds without that "2" since

$$|\mathbf{P}(X_n = j, T > n) - \mathbf{P}(X'_n = j, T > n)| \le \mathbf{P}(T > n)\,.$$

But (1.2) has more to give; a summation on both sides of that inequality renders

$$\sum_0^\infty |\mathbf{P}(X_n = j) - \pi_j| \le 2 \cdot \mathbf{P}(T > n)$$

(now that "2" must be included!) which we may write in a more compact form as

$$\|\mathbf{P}(X_n \in \cdot\,) - \pi\| \le 2 \cdot \mathbf{P}(T > n) \tag{1.3}$$

or

$$\|\lambda P^n - \pi\| \le 2 \cdot \mathbf{P}(T > n)\,.$$

We identify the absolutely convergent sequences $\nu = (\nu_i)_0^\infty$, corresponding row vectors $(\nu_0, \nu_1, \ldots)$, and bounded signed measures on $\mathbb{Z}_+$, and define

$$\|\nu\| = \sum_0^\infty |\nu_i|\,.$$

So if we can prove that the coupling is successful, uniform convergence toward stationarity is established effortlessly.

A number of questions should now be in your mind. For example:

Is it easy to prove that the coupling is successful?

Does (1.3) yield new results?

Is there an interesting class of Markov chains with general state space to which the coupling method is applicable?

The common answer is: yes.

As the second example, consider a standard birth and death process $X = (X_t)_0^\infty$ with state space $\mathbb{Z}_+$. Assume here that the intensities render X nonexplosive and recurrent. With $p_{ij}(t)$ short for $\mathbf{P}_i(X_t = j)$, where the i indicates that $X_0 = i$ and $P_t = (p_{ij}(t))$ for $t \geq 0$, the distribution of X_t becomes λP_t if that of X_0 is λ.

To investigate the asymptotics of X, introduce the parallel process X', governed by the same intensities as and independent of X, and with initial distribution μ, say. Now letting

$$T = \inf\{t; X_t = X'_t\}$$

and

$$X''_t = \begin{cases} X_t & \text{if } t < T \\ X'_t & \text{if } t \geq T \end{cases} \tag{1.4}$$

we may proceed as in (1.2) to obtain

$$\|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\| = \|\lambda P_t - \mu P_t\| \leq 2 \cdot \mathbf{P}(T > t)\,. \tag{1.5}$$

But since the path of a birth and death process is skip-free and hence two independent paths cannot cross without meeting, it is easy to understand that

$$T \leq \max(\tau_0, \tau'_0)\,,$$

where τ_0, τ'_0 are the times when X, X' hit the state 0. Due to the recurrence assumption, we certainly have $\tau_0, \tau'_0 < \infty$ a.s., so the coupling is successful and (1.5) yields

$$\|\lambda P_t - \mu P_t\| \to 0 \qquad \text{as} \quad t \to \infty \tag{1.6}$$

for any initial distributions λ, μ. Notice that we have not assumed positive recurrence, so a stationary distribution may not exist. If it does, call it π again and put $\mu = \pi$ in (1.6) to obtain

$$\|\lambda P_t - \pi\| \to 0 \quad \text{as} \quad t \to \infty . \tag{1.7}$$

While (1.7) is a result on tendency toward stationarity (or equilibrium, as some prefer to say), the statement (1.6) means that a recurrent birth and death process forgets its initial distribution. The term "ergodic" is used in a number of ways in the literature; in this book we reserve it for results of the type (1.6).

If $X_0 \equiv 0$, then regardless of μ,

$$X''_t \le X'_t \qquad \text{for all} \quad t \ge 0, \tag{1.8}$$

again due to the skip-free paths. Since X and X'' are equally distributed, we may deduce from (1.8) that

$$\mathbf{E}[g(X_t)] \le \mathbf{E}[g(X'_t)] \qquad \text{for all} \quad t \ge 0$$

if $g\colon \mathbb{Z}_+ \to \mathbb{R}$ is nondecreasing. In particular,

$$\mathbf{E}[X_t] \le \mathbf{E}[X'_t] \qquad \text{for all} \quad t \ge 0 .$$

(What alternative proof of that do you support?) Hence the coupling method serves us well also for establishing inequalities: You will see many examples of that use below.

How far-reaching are the consequences of the observation that crossing paths produce a coupling? Very far. In fact, it has made the coupling method an indispensable tool in the study of the Markov processes of birth and death type with state space, for example, $\{0, 1\}^S$ (S countable), that are used to model interacting particle systems.

For the final example, let $Y_1, \ldots, Y_n$ be independent 0–1 variables with

$$\mathbf{P}(Y_i = 1) = p_i$$

and let $X = \Sigma_1^n Y_i$. When all p_i values are small, X is approximately

Poisson distributed with parameter $\Sigma_1^n p_i$, as is well known. To illuminate that, we search for a bound of

$$\sum_0^\infty |\mathbf{P}(X=i)-p_\lambda(i)| = \|\mathbf{P}(X\in\cdot)-p_\lambda\|,$$

where $p_\lambda(i)=\exp(-\lambda)\cdot\lambda^i/i!$ for $i\geq 0$ and $\lambda=\Sigma_1^n p_i$. Now suppose that we can find a variable X' with distribution p_λ such that $X=X'$ with high probability. Then

(1.9)

$$\begin{aligned}|\mathbf{P}(X=i)-p_\lambda(i)| &= |\mathbf{P}(X=i)-\mathbf{P}(X'=i)|\\ &= |\mathbf{P}(X=i, X=X')+\mathbf{P}(X=i, X\neq X')\\ &\quad -\mathbf{P}(X'=i, X=X')-\mathbf{P}(X'=i, X\neq X')|\\ &\leq \mathbf{P}(X=i, X\neq X')+\mathbf{P}(X'=i, X\neq X'),\end{aligned}$$

hence

$$\|\mathbf{P}(X\in\cdot)-p_\lambda\|\leq 2\cdot\mathbf{P}(X\neq X'), \tag{1.10}$$

so if $\mathbf{P}(X\neq X')$ is small, we get a good Poisson approximation.

But how is such a variable X' produced? For one possible answer, let (Y_i, Y_i') be independent for $1\leq i\leq n$ and such that

(1.11) $\mathbf{P}((Y_i, Y_i')=(j,k))$

$$=\begin{cases}1-p_i & \text{if } (j,k)=(0,0),\\ \exp(-p_i)\cdot p_i^k/k! & \text{if } j=1 \text{ and } k\geq 1, \text{ and}\\ \exp(-p_i)-(1-p_i) & \text{if } (j,k)=(1,0).\end{cases}$$

Then Y_i is a 0–1 variable with $\mathbf{P}(Y_i=1)=p_i$, and Y_i' is Poisson distributed with parameter p_i. Due to independence, $X'=\Sigma_1^n Y_i'$ has the required distribution, and (1.10) renders

$$\begin{aligned}\|\mathbf{P}(X\in\cdot)-p_\lambda\| &\leq 2\cdot\mathbf{P}\left(\sum_1^n Y_i\neq\sum_1^n Y_i'\right)\leq 2\cdot\mathbf{P}(Y_i\neq Y_i' \text{ for some } i)\\ &\leq 2\cdot\sum_1^n \mathbf{P}(Y_i\neq Y_i').\end{aligned}$$

But

$$\begin{aligned}\mathbf{P}(Y_i \neq Y_i') &= \exp(-p_i) - (1-p_i) + \mathbf{P}(Y_i' \geq 2)\\ &= p_i \cdot (1 - \exp(-p_i)) \leq p_i^2\end{aligned}$$

and we have proved that

$$\|\mathbf{P}(X \in \cdot\,) - p_\lambda\| \leq 2 \cdot \sum_1^n p_i^2\,, \tag{1.12}$$

which is good enough for many applications.

The use of couplings for approximations such as (1.12) will be only a minor theme of this book. However, there are topics that should not be withheld; among them is a result on Poisson approximation for sums of weakly dependent 0–1 variables, sharp enough to generalize the remarkable Le Cam's theorem.

2. An outline. To know a method is to have learned how it works. What we have ahead of us is essentially a collection of applications of a few basic ideas consisting largely of topics of wide common interest with an attempt to maximize diversity. References to specialized studies are given in the Notes at the ends of chapter parts.

In Chapter I a number of cornerstones are laid. The term "coupling" is defined, and the coupling inequality, of which (1.3), (1.5), and (1.12) are applications, is settled and elaborated.

Renewal theory is a main theme of the book. Part 1 of Chapter II is dedicated to a detailed account of coupling of discrete-time renewal processes. The ideas and results here are used repeatedly, with consequences of type (1.3) for Markov chains discussed in the second parts of Chapters II and III, so this is basic. Skill is gained from learning how to couple discrete random walks, shuffle cards, and make Poisson approximations in the remaining parts of Chapter II.

New challenges face us as soon as we go beyond the discrete models. In Chapter III we meet some of them; the first concerns coupling of continuous-time renewal processes. It turns out that virtually all asymptotic results for such processes may be obtained by suitable couplings, as shown in Part 1. Coupling and related methods have had a strong impact on the asymptotics theory of Markov chains with a general state space. Part 2 is an account of that.

At this stage it is natural to answer some questions that have emerged. Is there a best possible coupling of two random sequences? Can we generalize our results from renewal to regenerative processes? Are there particular observations to make concerning coupling of Markov processes? Affirmative answers are provided in Parts 3 and 4. In Part 5 of Chapter III we sum up many of our findings about Markov chain ergodicity and coupling in one major theorem.

Relation (1.8) and its consequences is a simple example of how to use coupling to establish inequalities. That use is the theme of Chapter IV, which has Strassen's theorem as a basic result. However, many interesting inequalities may be proved by simpler means, in terms of independent, uniformly distributed variables, for example; the last section is a gallery of such inequalities.

The coupling method is a fine tool in the study of Markov processes which are defined in terms of transition intensities. The first three parts of Chapter V show its use in a range of examples, from birth and death processes, through models for queueing networks and epidemics, to interacting particle systems. The possibilities of embedding in a bivariate Poisson process are demonstrated in Parts 4 and 5; among other things, we find that device useful for urn models and renewal theory.

Chapter VI is devoted to diffusions. In several cases, coupling of one-dimensional diffusions is easy, due to the path continuity. The multidimensional case is in general difficult, but there are interesting exceptions; it turns out, for example, that Brownian motions may efficiently and easily be coupled.

We finish with a bit of history of the method in the Appendix, where a tribute is paid to its inventor, Wolfgang Doeblin.

3. Notes. The reader is supposed to be familiar with probability theory based on measures and Lebesgue integrals learned from, for example, Ash [9], Billingsley [27], Chung [39], Dudley [55], Durrett [58] or Williams [160]. And there are few readers with this background who do not have the slight acquaintance with Markov chains, random walk, Brownian motion, and so on, required, learned either from one of these references or from an undergraduate textbook such as Grimmett and Stirzaker [69]. A few standard results are used without comment; they are easily found in some of the references noted above.

The enumeration is a standard one: By "(3.2)" we refer to the second enumerated item in § 3 in the same chapter; "(IV.3.2)" indicates a reference to another chapter; § IV.3 designates § 3 in Chapter IV. The sign " = " is used not only to mean "equal to" but also occasionally "be defined by" and even "which we abbreviate to".

The rule is that notation is defined at most once. A comprehensive list is provided following the Appendix. For measurable spaces $(E, \mathscr{E})$ and $(E^*, \mathscr{E}^*)$, we shall let $\mathscr{E}/\mathscr{E}^*$ denote the class of measurable mappings from E to E^*. If $(E^*, \mathscr{E}^*)$ equals $(\mathbb{R}, \mathscr{R})$, the real numbers with the Borel sets, we simply write $f \in \mathscr{E}$ in order to say that f is a measurable real-valued function.

For a class of functions, the prefixes b, c, and i are used to mean "bounded", "continuous", and "increasing", respectively. Hence $f \in ib\mathscr{E}$ means that f is an increasing (= nondecreasing) and bounded measurable function. The prefix i requires that E be ordered in some way, while c can be used when E has a topology. The prefix b is also used for measures.

CHAPTER I

Preliminaries

1. What is a coupling? The coupling method deals, formally, with comparisons of probability measures on a measurable space. Our first task is to define what is meant by a coupling. Throughout, we let a state space candidate be denoted by $(E, \mathscr{E})$.

(1.1) *A coupling of the probability measures P and P' on a measurable space $(E, \mathscr{E})$ is a probability measure $\hat{P}$ on $(E^2, \mathscr{E}^2)$ such that*

$$P = \hat{P}\pi^{-1} \quad \text{and} \quad P' = \hat{P}\pi'^{-1},$$

where $\pi(x, x') = x$, $\pi'(x, x') = x'$ for $(x, x') \in E^2$.

Thus P and P' are the marginal distributions (or just marginals) of $\hat{P}$.

Now typical uses of the coupling method are those demonstrated in § Int.1, and we find that our definition (1.1) is not well fit for immediate use. Once again we understand that probability theory is not animated measure theory, but rather, a mathematical discipline in its own right, concerned with random elements.

For a more fruitful definition of coupling, it is ecónomical to define a random element in the space $(E, \mathscr{E})$ to be a quadruple

$$(\Omega, \mathscr{F}, \mathbf{P}, X),$$

where $(\Omega, \mathscr{F}, \mathbf{P})$ is the underlying probability space (the sample space) and $X \in \mathscr{F}/\mathscr{E}$, the class of measurable mappings from Ω to E.

(1.2) *By a coupling of the random elements* $(\Omega, \mathscr{F}, \mathbf{P}, X)$ *and* $(\Omega', \mathscr{F}', \mathbf{P}', X')$ *in* $(E, \mathscr{E})$ *we mean a random element* $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{\mathbf{P}}, (\hat{X}, \hat{X}'))$ *in* $(E^2, \mathscr{E}^2)$ *such that*

$$X \stackrel{\mathscr{D}}{=} \hat{X} \quad \textit{and} \quad X' \stackrel{\mathscr{D}}{=} \hat{X}' .$$

Hence $\hat{\mathbf{P}}(\hat{X}, \hat{X}')^{-1}$ is a coupling of $\mathbf{P}X^{-1}$ and $\mathbf{P}'X'^{-1}$ in the sense of (1.1).

Once this definition is understood, we may safely be content with a reduced terminology and refer to $(\hat{X}, \hat{X}')$ as a coupling of the random elements X and X'. When no clarity is lost, we omit superscripts ($\hat{}$, etc.) on Ω, $\mathscr{F}$, and $\mathbf{P}$.

For the random elements involved, the superscripts will vary; you will meet X'', $\tilde{Y}$, $\hat{Z}$, and so on. In the first example of § Int.1 we produced a coupling (X'', X') of the random elements X and X' in $(\mathbb{Z}_+^\infty, \mathscr{Z}_+^\infty)$, and the second example was similar in nature. The third, concerning Poisson approximation, is a case where for a comparison of two probability measures P and P' on a certain space $(E, \mathscr{E})$, a random element $(\Omega, \mathscr{F}, \mathbf{P}, (X, X'))$ in $(E^2, \mathscr{E}^2)$ is constructed. We shall refer to such a construction as a "coupling solution of" or a "coupling approach to" a certain problem. Notice that the definitions (1.1) and (1.2) may be extended directly to be valid for any class of probability measures or random elements.

When should the term "coupling" be used? There is no general agreement about that; some prefer a usage restricted to topics of the type presented in § Int.1; others find it suitable in any situation where for the comparison of probability measures, a common probability space is constructed. From the latter point of view, embeddings in Brownian motion, for examples, are actually couplings; notice that the definition (1.2) does not reject the use of the term in such connections.

If you prefer a restrictive usage, you soon find yourself in trouble when trying to state a definition. But there is no urgent need for any.

2. The coupling inequality. The terms "probability" and "distribution" will occasionally appear as alternatives to "probability measure". We shall use the total variation norm to measure distances between probabilities. For $\nu \in b\mathscr{M}_s(E, \mathscr{E})$, the space of signed and bounded measures on $(E, \mathscr{E})$, that norm is defined by

$$\|\nu\| = \sup_{\substack{|f| \le 1 \\ f \in \mathscr{E}}} \left| \int f \, d\nu \right| . \tag{2.1}$$

Due to Jordan–Hahn decomposition, there is a set $D \in \mathscr{E}$ such that $\nu(\cdot \cap D)$ and $-\nu(\cdot \cap D^c)$ are positive measures, ν^+ and ν^-, say. We have $\nu = \nu^+ - \nu^-$,

$$\sup_{A \in \mathscr{E}} \nu(A) = \nu(D) = \nu^+(E) \tag{2.2}$$

and find that

$$\|\nu\| = \int f_0 \, d\nu = \nu^+(E) + \nu^-(E) \tag{2.3}$$

with $f_0 = I_D - I_{D^c}$. If ν has total mass 0, then $\nu^+(E) = \nu^-(E)$, so if P and P' are probabilities on $(E, \mathscr{E})$, then (2.2–2.3) renders

$$\|P - P'\| = 2 \cdot \sup_{A \in \mathscr{E}} (P(A) - P'(A)) = 2 \cdot \sup_{A \in \mathscr{E}} (P'(A) - P(A)) . \tag{2.4}$$

There are further details about $b\mathscr{M}_s$ and $\| \cdot \|$ in App.2.

Now suppose that the random element $(\Omega, \mathscr{F}, \mathbf{P}, (Z, Z'))$ is a coupling in order to estimate $\|P - P'\|$. Using (2.4), we obtain

$$\|P - P'\| = 2 \cdot \sup_{A \in \mathscr{E}} (\mathbf{P}(Z \in A) - \mathbf{P}(Z' \in A)) .$$

But

$$\begin{aligned} \mathbf{P}(Z \in A) - \mathbf{P}(Z' \in A) &= \mathbf{P}(Z \in A, Z = Z') + \mathbf{P}(Z \in A, Z \neq Z') \\ &\quad - \mathbf{P}(Z' \in A, Z = Z') \\ &\quad - \mathbf{P}(Z' \in A, Z \neq Z') \\ &= \mathbf{P}(Z \in A, Z \neq Z') \\ &\quad - \mathbf{P}(Z' \in A, Z \neq Z') \\ &\le \mathbf{P}(Z \neq Z') , \end{aligned} \tag{2.5}$$

(2.6) $$\|P - P'\| \leq 2 \cdot \mathbf{P}(Z \neq Z') .$$

This is the (basic) coupling inequality. Before proceeding, notice that we must have $\{Z = Z'\} \in \mathcal{F}$. But throughout we shall restrict ourselves to state spaces that are Polish, and for such we indeed have that $\Delta = \{(x, x'); x = x'\} \in \mathcal{E}^2$, so $\{Z = Z'\} = \{(Z, Z') \in \Delta\} \in \mathcal{F}$.

Now let $(\hat{X}, \hat{X}')$ be a coupling of $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$, where X_n, X'_n are random elements in $(E, \mathcal{E})$. If there is a random time $T \in \mathbb{Z}_+$ such that

(2.7) $$\hat{X}_n = \hat{X}'_n \quad \text{for} \quad n \geq T ,$$

then we call T a coupling time and obtain from (2.6) that

(2.8) $$\|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| \leq 2 \cdot \mathbf{P}(T > n)$$

since $\{\hat{X}_n \neq \hat{X}'_n\} \subset \{T > n\}$.

But we can do better than (2.8). For $n \in \bar{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$, define the shift $\theta_n \colon E^\infty \to E^\infty$ by

(2.9) $$\begin{aligned} &\theta_n x = (x_n, x_{n+1}, \ldots) \qquad \text{when} \quad n < \infty, \quad \text{and} \\ &\theta_\infty x = (z, z, z, \ldots) \end{aligned}$$

for $x = (x_0, x_1, x_2, \ldots) \in E^\infty$, where z is a fixed element in E. Notice that E^∞ is Polish, with a natural metric generating the standard product σ-field $\mathcal{E}^\infty$. Then $(\theta_n \hat{X})_0^\infty, (\theta_n \hat{X}')_0^\infty$ is a coupling of $(\theta_n X)_0^\infty$ and $(\theta_n X')_0^\infty$ with the same coupling times T as for $(\hat{X}, \hat{X}')$, and we obtain that (2.8) actually improves itself, to the inequality

(2.10) $$\begin{aligned} &\|\mathbf{P}(\theta_n X \in \cdot\,) - \mathbf{P}(\theta_n X' \in \cdot\,)\| \\ &= \|\mathbf{P}((X_n, X_{n+1}, \ldots) \in \cdot\,) - \mathbf{P}((X'_n, X'_{n+1}, \ldots) \in \cdot\,)\| \\ &\leq 2 \cdot \mathbf{P}(T > n) . \end{aligned}$$

Consider the space of functions $D_E[0, \infty)$, with values in a Polish space E and defined on $[0, \infty)$, which are right-continuous and have left-hand limits at all arguments t. Endowed with the Skorohod topology, that space, which we abbreviate as D_E, becomes a Polish space. This is the most general type of path space we shall use; hence we shall not violate the Polish assumption even in our investigations of continuous-time processes. The σ-field on D_E generated by the Skorohod topology will be denoted by $\mathscr{D}_E$; see also § App.1.

Let $(\hat{X}, \hat{X}')$ be a coupling of $X = (X_t)_0^\infty$ and $X' = (X_t')_0^\infty$, where X and X' are D_E valued. We call the random time $T \in \bar{\mathbb{R}}_+$ a coupling time if

$$(2.11) \qquad \hat{X}_t = \hat{X}_t' \qquad \text{for} \quad t \geq T .$$

The analogs of (2.8) and (2.10) are

$$(2.12) \qquad \|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X_t' \in \cdot\,)\| \leq 2 \cdot \mathbf{P}(T > t)$$

and

$$(2.13) \qquad \|\mathbf{P}(\theta_t X \in \cdot\,) - \mathbf{P}(\theta_t X' \in \cdot\,)\| \leq 2 \cdot \mathbf{P}(T > t) ,$$

where the shift θ_t is defined in the obvious way:

$$(2.14) \qquad \begin{aligned} &\theta_t x = x(t + \cdot\,) \qquad \text{for} \quad t \text{ finite}, \quad \text{and} \\ &\theta_\infty x = \text{constant function} \equiv z \end{aligned}$$

for $x \in D_E$, where z is a fixed element $\in E$. For (2.12)–(2.13), notice that we must know that $\theta_t X$ and $\theta_t X'$ are random elements in $(D_E, \mathscr{D}_E)$, in order to use the argument (2.5) again. But they are; in order that a mapping into D_E, h say, shall be measurable, it suffices that $h(t)$ is a measurable mapping into E for each $t > 0$, as is well known to a reader familiar with the Skorohod topology.

Let Z and Z' be random elements in $(E, \mathscr{E})$, and for another space $(E^*, \mathscr{E}^*)$ let $\psi \in \mathscr{E}/\mathscr{E}^*$. From (2.4) we obtain

$$(2.15) \qquad \begin{aligned} &\|\mathbf{P}(\psi(Z) \in \cdot\,) - \mathbf{P}(\psi(Z') \in \cdot\,)\| \\ &= 2 \cdot \sup_{B \in \mathscr{E}^*} (\mathbf{P}(\psi(Z) \in B) - \mathbf{P}(\psi(Z') \in B)) \\ &\leq 2 \cdot \sup_{A \in \mathscr{E}} (\mathbf{P}(Z \in A) - \mathbf{P}(Z' \in A)) \\ &= \|\mathbf{P}(Z \in \cdot\,) - \mathbf{P}(Z' \in \cdot\,)\| , \end{aligned}$$

so the total variation distance between distributions never increases with a mapping. If $(\hat{Z}, \hat{Z}')$ is a coupling of Z and Z', a combination of (2.6) and (2.15) yields

$$\text{(2.16)} \quad \sup_{\psi} \|\mathbf{P}(\psi(Z) \in \cdot) - \mathbf{P}(\psi(Z') \in \cdot)\| \leq 2 \cdot \mathbf{P}(\hat{Z} \neq \hat{Z}') ,$$

where ψ ranges over the class of measurable mappings with E as domain. Because of its importance, (2.16) deserves a name; let us call it the coupling–mapping inequality.

3. Rates of convergence. If a coupling of two random sequences $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ is successful, (i.e., it has a coupling time T that is finite a.s.), then $\mathbf{P}(T > n) \to 0$ as $n \to \infty$ and (2.8) and (2.10) give the results

$$\text{(3.1)} \qquad \|\mathbf{P}(X_n \in \cdot) - \mathbf{P}(X'_n \in \cdot)\| \to 0$$

$$\text{(3.2)} \qquad \|\mathbf{P}(\theta_n X \in \cdot) - \mathbf{P}(\theta_n X' \in \cdot)\| \to 0 \qquad \text{as} \quad n \to \infty .$$

Actually, (3.1) is a consequence of (3.2); use the coupling–mapping inequality with

$$\psi(x) = x_0 \qquad \text{for} \quad x = (x_k)_0^\infty$$

to obtain

$$\text{(3.3)} \quad \|\mathbf{P}(X_n \in \cdot) - \mathbf{P}(X'_n \in \cdot)\| \leq \|\mathbf{P}(\theta_n X \in \cdot) - \mathbf{P}(\theta_n X \in \cdot)\| .$$

Below, we display rate results only for the convergence (3.1) and its continuous-time analog, but at each occasion we may replace $\|\mathbf{P}(X_n \in \cdot) - \mathbf{P}(X'_n \in \cdot)\|$ by $\|\mathbf{P}(\theta_n X \in \cdot) - \mathbf{P}(\theta_n X' \in \cdot)\|$ and $\|\mathbf{P}(X_t \in \cdot) - \mathbf{P}(X'_t \in \cdot)\|$ by $\|\mathbf{P}(\theta_t X \in \cdot) - \mathbf{P}(\theta_t X' \in \cdot)\|$, as soon as our estimates are based on (2.8) and (2.12), respectively.

Now (3.1)–(3.2) may be improved if we can bound $\mathbf{P}(T > n)$. Indeed, if φ is a nonnegative and increasing function on $\mathbb{Z}_+$ and we know that $\mathbf{E}[\varphi(T)] < \infty$, then

$$\text{(3.4)} \qquad \varphi(n) \cdot \mathbf{P}(T > n) \leq \mathbf{E}[\varphi(T) \cdot I(T > n)] \to 0$$

as $n \to \infty$ due to dominated convergence, and (2.10) implies that

(3.5) $\|\mathbf{P}(\theta_n X \in \cdot\,) - \mathbf{P}(\theta_n X' \in \cdot\,)\| = o(1/\varphi(n)) \quad \text{as} \quad n \to \infty,$

and (3.3) gives

(3.6) $\|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| = o(1/\varphi(n)) \quad \text{as} \quad n \to \infty.$

We also have

$$\sum_{n=0}^{\infty} (\Delta\varphi)(n) \cdot \mathbf{P}(T > n) = \sum_{j=1}^{\infty} \mathbf{P}(T = j) \cdot (\varphi(j) - \varphi(0))$$
$$\leq \mathbf{E}[\varphi(T)] < \infty\,;$$

hence

(3.7) $$\sum_{n=0}^{\infty} (\Delta\varphi)(n) \|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| < \infty.$$

With $\varphi(n) = n^{\alpha}$, $\alpha > 0$, we get from (3.6)–(3.7) that

(3.8) $\|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| = o(n^{-\alpha}) \quad \text{as} \quad n \to \infty$

and

(3.9) $$\sum_{n=1}^{\infty} n^{\alpha-1} \cdot \|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| < \infty$$

if $\mathbf{E}[T^{\alpha}] < \infty$. Use $n^{\alpha-1} \leq C \cdot (\Delta\varphi)(n)$ for (3.9).

If $E[\rho^T] < \infty$ for $\rho > 1$, we apply (3.6)–(3.7) with $\varphi(n) = \rho^n$, $n \geq 0$, to get

(3.10) $\|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| = o(\rho^{-n}) \quad \text{as} \quad n \to \infty$

and

(3.11) $$\sum_{1}^{\infty} \rho^n \cdot \|\mathbf{P}(X_n \in \cdot\,) - \mathbf{P}(X'_n \in \cdot\,)\| < \infty.$$

This time (3.7) was applied with the observation that $\Delta\varphi(n) \leq C \cdot \varphi(n)$ for $n > 0$.

Now let φ be a nonnegative increasing function on $[0, \infty)$ so smooth that $\varphi(t) = \int_0^t \varphi'(s)\, ds$ for all $t \geq 0$, where φ' is the derivative. Fubini's theorem applies to the effect that

$$\begin{aligned}\int_0^\infty \varphi'(t) \cdot \mathbf{P}(T > t)\, dt &= \int_0^\infty \varphi'(t)[\mathbf{P}T^{-1}](t, \infty)\, dt \\ &= \int_0^\infty [\varphi(y) - \varphi(0)][\mathbf{P}T^{-1}](dy) \\ &\leq \int_0^\infty \varphi(y)[\mathbf{P}T^{-1}](dy) = \mathbf{E}[\varphi(T)]\,.\end{aligned}$$

Hence if $\mathbf{E}[\varphi(T)] < \infty$, the inequality (2.12) renders

$$\|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\| = o(\varphi(t)^{-1}) \quad \text{as} \quad t \to \infty \tag{3.12}$$

and

$$\int_0^\infty \varphi'(t) \cdot \|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\|\, dt < \infty \tag{3.13}$$

for a coupling of two continuous-time processes $X = (X_t)_0^\infty$ and $X' = (X'_t)_0^\infty$ with coupling time T. This implies

$$\|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\| = o(t^{-\alpha}) \quad \text{as} \quad t \to \infty \tag{3.14}$$

and

$$\int_0^\infty t^{(\alpha-1)} \cdot \|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\| < \infty \tag{3.15}$$

if $\mathbf{E}[T^\alpha] < \infty$ for an $\alpha > 0$. If even $\mathbf{E}[e^{\alpha T}] < \infty$ for an $\alpha > 0$, then

$$\|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\| = o(e^{-\alpha t}) \quad \text{as} \quad t \to \infty \tag{3.16}$$

and

$$\int_0^\infty e^{\alpha t} \cdot \|\mathbf{P}(X_t \in \cdot\,) - \mathbf{P}(X'_t \in \cdot\,)\|\, dt < \infty\,. \tag{3.17}$$

4. Weak coupling. We return to the coupling inequality (2.6) and the proof of that. It turns out that if $(\Omega, \mathscr{F}, \mathbf{P}, (Z, Z'))$ is a coupling to estimate $\|P - P'\|$ and $B, B' \in \mathscr{F}$ are such that

$$\mathbf{P}((Z \in \cdot\,) \cap B) = \mathbf{P}((Z' \in \cdot\,) \cap B') \tag{4.1}$$

[which implies that $\mathbf{P}(B) = \mathbf{P}(B')$], then modifying (2.5) in the obvious way, we obtain

$$\begin{aligned} \mathbf{P}(Z \in A) - \mathbf{P}(Z' \in A) &= \mathbf{P}(Z \in A, B^c) - \mathbf{P}(Z' \in A, B'^c) \\ &\leq \mathbf{P}(B^c)\,. \end{aligned} \tag{4.2}$$

Hence

$$\|P - P'\| \leq 2 \cdot \mathbf{P}(B^c)\,. \tag{4.3}$$

Of course, this is of interest only if $\mathbf{P}(B) > 0$. Assuming this, you may prefer to express (4.1) as

$$\mathbf{P}(B) = \mathbf{P}(B') \quad \text{and} \quad \mathbf{P}(Z \in \cdot \,\|\, B) = \mathbf{P}(Z' \in \cdot \,\|\, B')\,. \tag{4.4}$$

We call (4.3) the weak coupling inequality and B, B' the similarity sets. To find suitable similarity sets and then use (4.1) and (4.3) to estimate $\|P - P'\|$ characterizes a weak coupling.

Now let $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ be random sequences in $(E, \mathscr{E})$. For a weak coupling of X and X', let $\hat{X} = (\hat{X}_n)_0^\infty$ and $\hat{X}' = (\hat{X}'_n)_0^\infty$ be sequences, T and T' random times $\in \bar{\mathbb{Z}}_+$ such that

$$X \stackrel{\mathscr{D}}{=} \hat{X}, \qquad X' \stackrel{\mathscr{D}}{=} \hat{X}' \tag{4.5}$$

and

$$(\theta_T \hat{X}, T) \stackrel{\mathscr{D}}{=} (\theta_{T'} \hat{X}', \hat{T}') \tag{4.6}$$

are at hand; recall the definition (2.9) of θ_n for $n \in \bar{\mathbb{Z}}_+$. Then (4.1) is satisfied with $Z = \theta_n \hat{X}$, $Z' = \theta_n \hat{X}'$ and $B = \{T \leq n\}$, $B' = \{T' \leq n\}$:

$$\begin{aligned}\mathbf{P}(\theta_n\hat{X}\in A,\,T\le n) &= \sum_{k=0}^{n}\mathbf{P}(\theta_{n-k}(\theta_k\hat{X})\in A, T=k)\\ &= \sum_{k=0}^{n}\mathbf{P}(\theta_{n-k}(\theta_k\hat{X}')\in A, T'=k)\\ &= \mathbf{P}(\theta_n\hat{X}'\in A,\,T'\le n)\end{aligned}$$

for any $A\in\mathscr{E}^\infty$ due to (4.6), and (4.3) renders

$$(4.7)\qquad \|\mathbf{P}(\theta_nX\in\cdot\,)-\mathbf{P}(\theta_nX'\in\cdot\,)\|\le 2\cdot\mathbf{P}(T>n)\,.$$

Hence a weak coupling of sequences X and X', defined by (4.5)–(4.6), gives us a familiar inequality. Notice that such a coupling may involve two coupling times, T and T'.

The definition (4.5)–(4.6) seems peculiar, and you may wonder if it is useful. Indeed it is; often our construction work is complicated, and it turns out that the introduction of a weak coupling may reduce technical difficulties considerably. You will soon see an example of how well a weak coupling works in a simple connection: for discrete renewal processes. Also, measurability considerations are sometimes simplified: Notice that no assumption of the type "$\{Z=Z'\}\in\mathscr{F}$" is necessary for (4.2)!

For any random sequence $\hat{X}$ and time $T\in\bar{\mathbf{Z}}_+$ defined on the same probability space, $\theta_T\hat{X}$ is measurable, as you easily verify. But that is also true for our D_E-valued processes, where $T\in\bar{\mathbf{R}}_+$. This implies that for a definition of weak coupling of random processes $X=(X_t)_0^\infty$ and $X'=(X'_t)_0^\infty$, (4.5)–(4.6) may be used exactly as they stand.

When the need arises, we shall call a coupling in the sense of § 1 "strong".

5. The γ coupling. The purpose of this paragraph is to prove the existence of a coupling that is maximal in the sense that we get equality in the coupling inequality (2.6). On the way, we learn more about the total variation norm.

For the probability measures P and P' on $(E,\mathscr{E})$, let $\lambda=P+P'$ and

$$g=dP/d\lambda,\qquad g'=dP'/d\lambda\,.$$

Going back to the definition (2.1) of $\|P-P'\|$, we obtain

$$\|P-P'\| = \sup_{\substack{|f|\le 1\\ f\in\mathscr{E}}} \left|\int f\cdot(g-g')\,d\lambda\right| \tag{5.1}$$

$$= \int_{g\ge g'} 1\cdot(g-g')\,d\lambda + \int_{g<g'} (-1)\cdot(g-g')\,d\lambda$$

$$= \int |g-g'|\,d\lambda$$

$$= \int (g - g\wedge g')\,d\lambda + \int (g' - g\wedge g')\,d\lambda$$

$$= 2\cdot\left(1-\int g\wedge g'\,d\lambda\right).$$

Hence if we can produce a coupling (Z, Z') such that $\mathbf{P}(Z=Z') = \int g\wedge g'\,d\lambda$, it is best possible. To that end, let Q denote the subprobability defined by $dQ = (g\wedge g')\,d\lambda$ and γ its total mass. We search for a coupling $\hat{P}$ of P and P' such that $\hat{P}(\Delta)=\gamma$.

(5.2) Theorem. *For any two probability measures P and P' on a measurable space $(E, \mathscr{E})$ there exists a coupling (Z, Z') such that*

(i) $\|P-P'\| = 2\cdot\mathbf{P}(Z\ne Z')$

and

(ii) *Z and Z' are independent conditioned on $\{Z\ne Z'\}$, that is, we have $\mathbf{P}(Z\in A,\ Z'\in A' \| Z\ne Z') = \mathbf{P}(Z\in A\|Z\ne Z')\times \mathbf{P}(Z'\in A'\|Z\ne Z')$ for all $A, A'\in\mathscr{E}$ if $\gamma<1$.*

Proof. We may restrict ourselves to the case $\gamma<1$. The mapping $\psi\colon E\to E^2$ given by $x\to(x,x)$ belongs to $\mathscr{E}/\mathscr{E}^2$ due to the Polish assumption (see § 6). Let $\hat{Q} = Q\psi^{-1}$; of course, $\hat{Q}$ concentrates all its mass $(=\gamma)$ on Δ. Now let

$$\nu = P - Q \quad\text{and}\quad \nu' = P' - Q$$

and

$$\hat{P} = \nu\times\nu'/(1-\gamma) + \hat{Q}\,. \tag{5.3}$$

We have

$$\begin{aligned}\hat{P}(A \times E) &= \nu(A)\cdot(1-\gamma)/(1-\gamma) + \hat{Q}(A \times E)\\ &= \nu(A) + Q(A) = P(A)\end{aligned}$$

and $\hat{P}(E \times A') = P'(A)$ for all $A, A' \in \mathscr{E}$, so $\hat{P}$ is indeed a coupling of P and P'. For (i), let (Z, Z') be any pair with distribution $\hat{P}$. Now it is immediate from (5.3) that

$$\hat{P}(\cdot \,\|\Delta^c) = (\nu/(1-\gamma)) \times (\nu'/(1-\gamma))$$

and (ii) follows readily. $\square$

We shall refer to $\hat{P}$ of (5.3) as the γ coupling of P and P'; the term is used in the obvious way for random elements. The γ coupling is what can be called a maximal coupling; with condition (5.2)(ii) it is unique. Notice that the Poisson approximation in the last example of § Int.1 uses a γ coupling of Y_t and Y'_t.

6. The Polish assumption. A topological space E is called Polish if the topology is metrizable by a metric that is complete and makes it separable. There are examples in probability theory where non-Polish state spaces are required, but in general the natural topology on a state space E is Polish, and the Borel σ-field $\mathscr{E}$ (= the σ-field generated by the open sets in E) is the one we expect and want. This holds for state spaces ranging to D_E and even further. That has the consequence that many probabilists feel reluctant to enter spaces beyond the Polish.

On the other hand, there is a powerful and virtually complete theory for probabilities on Polish spaces, and several fine accounts of it. For a review and references, see §§ App.1 and App.3; whenever you meet a phrase such as "Due to the Polish assumption . . ." and hesitate, turn to that summary.

The purpose of this book is to be a reasonably sized presentation of the many possibilities of the coupling method. To concentrate on that, we restrict our attention to Polish state spaces.

7. Notes. Stimulated by an idea in Ney [122], weak coupling of random sequences and processes was introduced by Thorisson [148]; he calls it "distributional coupling". The γ coupling has appeared at several places in different shapes. I have not been able to trace its roots; that is the reason for choosing a neutral term.

CHAPTER II

Discrete Theory

1. RENEWAL THEORY

1. Basics. Let $Y_0, Y_1, Y_2, \ldots$ be independent random variables, taking nonnegative integer values and defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The variables $Y_1, Y_2, \ldots$ are identically distributed and strictly positive, playing the role of recurrence times in the discrete renewal process $S = (S_n)_0^\infty$, where $S_n = \Sigma_0^n Y_i$, $S_0 = Y_0$.

Throughout, we assume that a discrete state space E is equipped with the σ-field $\mathcal{P}(E)$ of all subsets of that space. For the variables Y_0, Y_1, the state space is $\mathbb{Z}_+$; we denote the σ-field by $\mathcal{Z}_+$.

We say that a renewal takes place at each of the times S_n, $n \geq 0$; the first renewal appears at Y_0, the delay of the process, which may be 0. If $Y_0 \equiv 0$, the process is zero delayed. It is often convenient to indicate the renewal times by a random sequence $V = (V_n)_0^\infty$, where

$$V_n = \begin{cases} 1 & \text{if } n \text{ is a renewal time} \\ 0 & \text{otherwise}. \end{cases}$$

Whenever convenient, we may denote the renewal process by V rather than S. The renewal sequence $v = (v_k)_0^\infty$ associated with it is defined through

$$v_k = \mathbf{P}(V_k = 1).$$

In the zero-delayed case, it is customary to let $u = (u_k)_0^\infty$ denote that sequence. Notice that $u_0 = 1$.

We avoid trite work by assuming that the distribution $p = (p_k)_1^\infty$

of the recurrence times $[\mathbf{P}(Y_j = k) = p_k$ for $j \geq 1]$ is aperiodic, that is,

$$\text{g.c.d. } \{k;\ p_k > 0\} = 1\,.$$

We denote the distribution of the delay Y_0 by $a = (a_k)_0^\infty$ and the mean recurrence time $\mathbf{E}[Y_1]$ by μ. The renewal intensity is defined by $\lambda = 1/\mu$.

What about $(\Omega, \mathcal{F}, \mathbf{P})$? When S is embedded in some process, then of course $(\Omega, \mathcal{F}, \mathbf{P})$ is the probability space on which that "host" process is defined. But when our aim is to establish some properties of v in terms of a and p, we choose let $\Omega = \mathbb{Z}_+^\infty$, $\mathcal{B} = \mathcal{X}_+^\infty$, $\mathbf{P} = \Pi_0^\infty \nu_i$, where $\nu_0 = a$ and $\nu_i = p$ for $i \geq 1$, and define $(Y_n)_0^\infty$ as the coordinate process in the usual way. The reader is assumed to be familiar with this sort of standard product space constructions; in this book, they will not be carried out without special reasons.

When the delay distribution needs to be emphasized, the notation $\mathbf{P}_a$ is used; $\mathbf{P}_i$ is short for $\mathbf{P}_{\delta_i}$, and $\mathbf{E}_a$, $\mathbf{E}_i$ are used for expectations w.r.t. $\mathbf{P}_a$, $\mathbf{P}_i$. Notice that we have assumed that p is nondefective (i.e., $\Sigma_1^\infty p_k = 1$). However, in a proof below, we need to consider renewal processes with possibly a defective recurrence distribution. With such, there are only a finite number of renewals. It is virtually a geometrically distributed variable, hence has a finite expectation.

We shall also have use for the fact that due to the aperiodicity

(1.1) there exists an n_0 such that $u_n > 0$ for $n \geq n_0$.

You probably remember (1.1) from your first course on Markov chains.

2. Stationarity. The coupling. Throughout this section assume that $\mu < \infty$. We say that a random sequence $X = (X_n)_0^\infty$ is stationary if

$$\theta_n X \stackrel{\mathcal{D}}{=} X \qquad \text{for all} \quad n \geq 0\,.$$

It is well known that if the delay Y_0 has distribution $c = (c_k)_0^\infty$ where

$$c_k = \lambda \cdot \sum_{k+1}^{\infty} p_i, \qquad k \geq 0, \tag{2.1}$$

then V is stationary. For a proof it suffices to show that the distributions of D_n, where

$$D_n = \min\{S_k - n; S_k - n \geq 0\},$$

are the same for all $n \geq 0$. But $D = (D_n)_0^\infty$ is a Markov chain, and easy calculations show that $c = cP$ where P is its transition matrix, rendering D stationary.

To study how $V = (V_n)_0^\infty$ forgets initial conditions as $n \to \infty$, introduce another renewal process $V' = (V'_n)_0^\infty$, independent of V, and with recurrence and delay distributions p and a', respectively. Let

$$T = \min\{k; V_k = V'_k = 1\}. \tag{2.2}$$

We shall prove $\mathbf{P}_{aa'}(T < \infty) = 1$ (the notation $\mathbf{P}_{aa'}$ is supposed to be self-explanatory). To that end, define $V^* = (V_n^*)_0^\infty$ by

$$V_n^* = V_n \cdot V'_n, \qquad n \geq 0,$$

and assume momentarily that $a = a' = c$. Then V^* is another aperiodic and stationary renewal process, possibly with defective delay or recurrence distribution. But its recurrence distribution must be nondefective ($\Sigma_1^\infty p_i^* = 1$) because otherwise $\mathbf{P}_{cc}(V_n^* = 1) \to 0$; $\mathbf{P}_{cc}(V_n^* = 1) \equiv \lambda^2$, however. Hence $\mathbf{P}_{00}(V_n^* = 1 \text{ i.o.}) = 1$. Now, use the n_0 of (1.1) in the following argument: Denoting $\{V_n^* = 1 \text{ i.o.}\}$ by A we obtain for $m \geq 0$,

$$\begin{aligned}
1 &= \mathbf{P}_{00}(A) \\
&= \mathbf{P}_{00}(A \| (V_{n_0}, V'_{n_0+m}) = (1,1)) \cdot \mathbf{P}_{00}((V_{n_0}, V'_{n_0+m}) = (1,1)), \\
&\quad + \mathbf{P}_{00}(A \| (V_{n_0}, V'_{n_0+m}) \neq (1,1)) \cdot \mathbf{P}_{00}((V_{n_0}, V'_{n_0+m}) \neq (1,1)),
\end{aligned}$$

which forces

$$\mathbf{P}_{00}(A \| (V_{n_0}, V'_{n_0+m}) = (1,1)) = \mathbf{P}_{0m}(V_n^* = 1 \text{ i.o.})$$

to equal 1. We used the fact that if x, y, z are numbers satisfying $0 \le x, y, z \le 1$ and $z > 0$, then $xz + y(1-z) = 1$ implies that $x = 1$. For $i, j \ge 0$ we have $\mathbf{P}_{ij}(V_n^* = 1 \text{ i.o.}) = \mathbf{P}_{0,|j-i|}(V_n^* = 1 \text{ i.o.})$ and may conclude that

$$(2.3)\quad \mathbf{P}_{aa'}(T < \infty) \ge \mathbf{P}_{aa'}(V_n^* = 1 \text{ i.o.}) = \sum_{i,j=0}^{\infty} \mathbf{P}_{ij}(V_n^* = 1 \text{ i.o.})\, a_i a_j = 1$$

for any delay distributions a, a'. With V'' defined by $V_n'' = V_n$ for $n < T$ and V_n' for $n \ge T$, we have completed the construction of the coupling (V'', V') of (V, V').

Throughout, we shall call a coupling where we let two independent processes run until they meet a Doeblin coupling. Actually, we shall find good reasons to let two processes, $(X_n)_0^\infty$ and $(X_n')_0^\infty$, say, be dependent at time 0, but thereafter be independent conditioned on X_0 and X_0'. The term "Doeblin coupling" is used even under these circumstances.

3. The discrete renewal theorem. We omitted the detailed argumentation that (V'', V') is indeed a coupling of (V, V'). The reason for this is that we have come to a point where a simple weak coupling is suitable to use. Actually, before the terms "distributional coupling" and "weak coupling" were coined, it had been observed concerning Markov chains and related processes that it is the use of the strong Markov property, with T of (2.2) as the stopping time involved, that really matters rather than a complete construction of a strong coupling.

Lengthy formulations of lemmas, theorems, and so on, will be avoided in this book.

(3.1) Theorem. *If p is aperiodic and $\mu < \infty$, then*

(i) $\|\mathbf{P}_a(\theta_n V \in \cdot\,) - \mathbf{P}_{a'}(\theta_n V' \in \cdot\,)\| \to 0$ *as* $n \to \infty$ *for all* a, a'.
In particular,

(ii) $\|\mathbf{P}_a(\theta_n V \in \cdot\,) - \mathbf{P}_c(V' \in \cdot\,)\| \to 0$ *as* $n \to \infty$ *for all* a.

Proof. Recalling the definition of weak coupling of sequences, we find, with $T = T'$, that it suffices to prove $(\theta_T V, T) \overset{\mathscr{D}}{=} (\theta_T V', T)$; cf. (I.4.5)–(I.4.6). We have for any set $A \in \mathscr{X}_+^\infty$ and $k \ge 0$,

$$\begin{aligned}\mathbf{P}_{aa'}((\theta_T V, T) \in (A, k)) &= \mathbf{P}_{aa'}(\theta_k V \in A, T = k) \\ &= \mathbf{P}_0(V \in A) \cdot \mathbf{P}_{aa'}(T = k)\end{aligned}$$

and we get exactly the same thing from elaborating $\mathbf{P}_{aa'}((\theta_T V', T) \in (A, k))$. The statements of the theorem now follows from the coupling inequality (I.4.7) and (2.3). The result (ii) is case (i), with $a' = c$. □

The classical result $v_n \to \lambda$ as $n \to \infty$ is a consequence of (ii) and the coupling–mapping inequality: Let $\psi((x_i)_0^\infty) = x_0$ for $(x_i)_0^\infty \in \{0, 1\}^\infty$. The reader is invited to try some other mappings ψ, in order to find how easily the asymptotics of $\psi(\theta_n V)$ follows from (ii); no renewal equations needed! We return to the convergence $v_n \to \lambda$ in § 5. As a consequence of successful dependent couplings in Part 2, (3.1)(i) also holds when $\mu = \infty$.

A final remark: Notice that it is formally quite correct to say that we are estimating $\|\mathbf{P}_a(\theta_n V \in \cdot) - \mathbf{P}_{a'}(\theta_n V \in \cdot)\|$. However, the prime must be introduced in a coupling proof, and it supports intuition to include it in the statements of theorems.

4. Finite moments of T. Under what conditions is $\mathbf{E}_{aa'}[T^\alpha] < \infty$ for an $\alpha > 0$? Once we know that,

$$\|\mathbf{P}_a(\theta_n V \in \cdot) - \mathbf{P}_{a'}(\theta_n V' \in \cdot)\| = o(n^{-\alpha})$$

follows from (I.3.8). But far more important than that first application is the fact that we obtain this sort of rate (= speed of convergence) results for the entire class of processes, including Markov chains, where properties of embedded discrete renewal processes are instrumental for the couplings. Hence this is a key section. Also, the methods of proof are of value for the continuous-time renewal theory; see § III.6.

For a probability distribution $b = (b_i)_0^\infty$ on $\mathbb{Z}_+$ and an $\alpha > 0$, let

$$m_\alpha(b) = \sum_0^\infty i^\alpha \cdot b_i \,.$$

For $b = p$, abbreviate $m_\alpha(p)$ to μ_α (= the moment of order α of the recurrence distribution). Of course, $\mu_1 = \mu$.

Recall the sequence $D=(D_n)_0^\infty$ from § 2: Think of D_n as the overshoot at the level n of the renewal process S. It is convenient to have certain moment estimates for D_n at our disposal.

(4.1) Lemma. *Suppose that $\mu_\alpha<\infty$ for an $\alpha>1$. Then*

(i) $\mathbf{E}_0[D_n^\alpha]\le\mu_\alpha\cdot n$ *for all* $n\ge 0$, *and*
(ii) *for each $\rho>0$ there exists a constant $C_1=C_1(\rho)$ such that* $\mathbf{E}_0[D_n]\le C_1+\rho\cdot n$ *for all* $n\ge 0$.

Proof. We pay no attention to the case $n=0$. For $n\ge 1$ we have

$$\begin{aligned}\mathbf{P}_0(D_n=k)&=\sum_0^\infty \mathbf{P}_0(D_n=k,S_j<n,S_{j+1}\ge n)\\&=\sum_{j=0}^\infty\sum_{i=0}^{n-1}\mathbf{P}_0(S_j=i,Y_{j+1}=n+k-i)\\&=\sum_0^{n-1}u_i\cdot p_{n+k-i}\le\sum_0^{n-1}p_{n+k-i}\,.\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{E}_0[D_n^\alpha]&\le\sum_{k=0}^\infty k^\alpha\sum_{i=0}^{n-1}p_{n+k-i}\\&=\sum_{i=0}^{n-1}\sum_{k=0}^\infty k^\alpha p_{n+k-i}\le\sum_0^{n-1}\mu_\alpha=n\cdot\mu_\alpha\,.\end{aligned}$$

Of course $(\Sigma_0^{n-1}u_i)/n\to\lambda$ since $u_n\to\lambda$. Now $\Sigma_0^{n-1}u_i=\mathbf{E}_0[\eta_n]$, where $\eta_n=$ number of renewals in $\{0,1,\ldots,n-1\}$, and $n+D_n=S_{\eta_n}$, so Wald's lemma applied to S_{η_n} implies that

$$\mathbf{E}_0[D_n]=\mu\cdot\mathbf{E}_0[\eta_n]-n$$

and (ii) follows. □

Wald's lemma states that if $Z_1,Z_2,\ldots$ are i.i.d. random variables with finite expectation μ and τ is a stopping time w.r.t. $(Z_i)_0^\infty$, then

$$\mathbf{E}\left[\sum_1^\tau Z_i\right]=\mu\cdot\mathbf{E}[\tau]$$

if τ has a finite expectation.

(4.2) Theorem. *For an $\alpha > 0$, suppose that μ_α, $m_\alpha(a)$, and $m_\alpha(a')$ are finite. Then $\mathbf{E}_{aa'}[T^\alpha] < \infty$.*

Proof. Recall the running assumption that $\mu < \infty$. First let $Y_0 \equiv 0$. We will carry out a series of trials to make S and S' hit each other. To obtain a uniform lower bound for the success probability in each trial, note that there exists an n_0 and a $\gamma > 0$ such that $u_n \geq \gamma$ for $n \geq n_0$ since $u_n \to \lambda > 0$, and define random variables A_i, B_i, ν_i for $i > 0$ by $A_0 \equiv 0$ and, by induction,

$$\begin{aligned}
B_{2n} &= \min\{S'_j - A_{2n};\, S'_j - A_{2n} \geq 0\} = S'_{\nu_{2n}} - A_{2n}\,,\\
A_{2n+1} &= S'_{\nu_{2n}+n_0}\,,\\
B_{2n+1} &= \min\{S_j - A_{2n+1};\, S_j - A_{2n+1} \geq 0\} = S_{\nu_{2n+1}} - A_{2n+1}\,,\\
A_{2n+2} &= S_{\nu_{2n+1}+n_0}\,.
\end{aligned}$$

Of course, $B_j = 0$ stands for a successful coupling; the reason for adding n_0 steps to each of the variables S_{ν_1}, $S'_{\nu_2}, \ldots$ is that the probability for such a success becomes $\geq \gamma$.

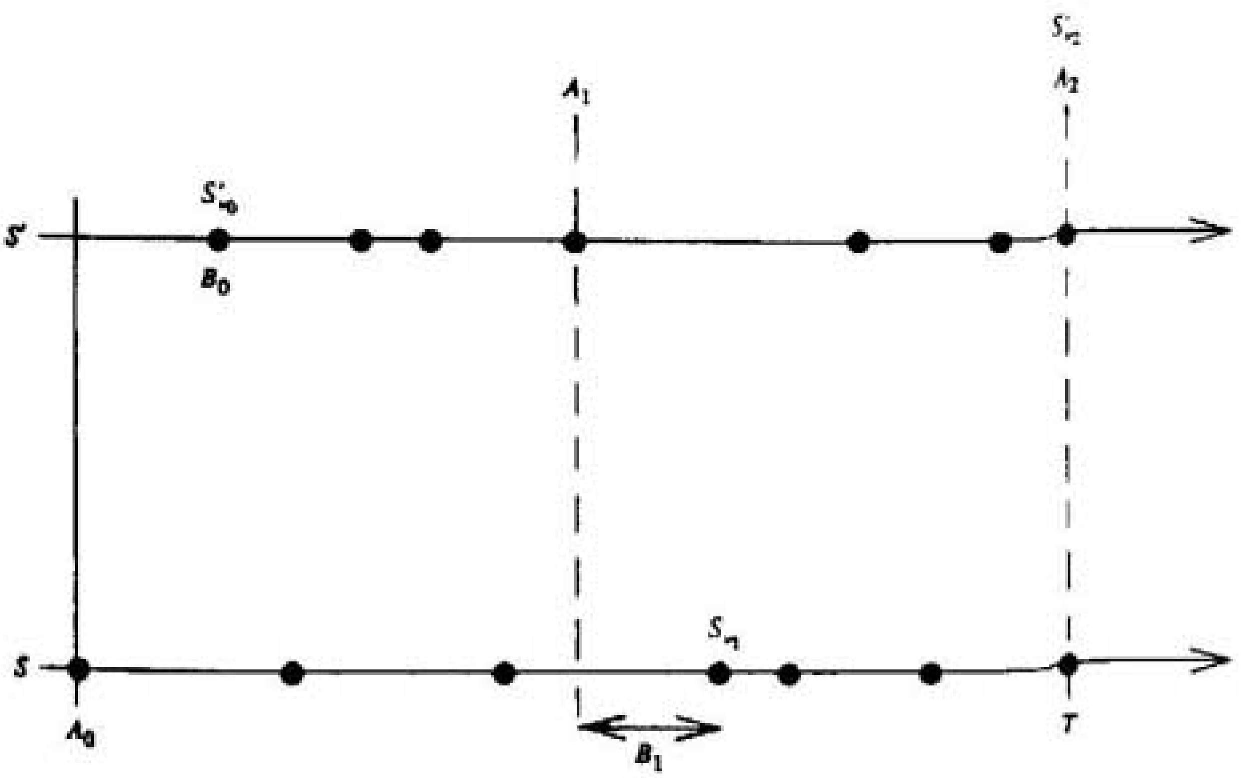

Figure 1. If $B_j = 0$, we have a successful coupling. In this example we have B_0 and $B_1 > 0$ and $B_2 = 0$, hence $\tau = 2$. Notice that $n_0 = 3$.

Define

$$U_1 = S_{\nu_1} - S'_{\nu_0}\,, \qquad U_2 = S'_{\nu_2} - S_{\nu_1}\,,$$

and so on, and

$$\tau = \min\{k;\, B_k = 0\}\,.$$

Obviously,

$$T \le Y'_0 + \sum_1^{\tau} U_i = Y'_0 + \sum_1^{\infty} U_i \cdot I(\tau \ge i)\,. \tag{4.3}$$

For $\alpha \ge 1$, let $\|Z\| = \mathbf{E}[|Z|^\alpha]^{1/\alpha}$ for random variables Z. Minkowski's inequality renders

$$\|T\| \le \|Y'_0\| + \sum_1^{\infty} \|U_i \cdot I(\tau \ge i)\|\,, \tag{4.4}$$

so our essential task is to bound $\|U_i \cdot I(\tau \ge i)\|$ for a proof of $\mathbf{E}[T^\alpha] < \infty$; that $\|Y'_0\| < \infty$ is an assumption of the theorem. To that end, let

$$\mathcal{B}_i = \sigma\{Y_j, Y'_k;\, j \le \nu_i, k \le \nu_{i-1} + n_0\}$$

for i odd, and

$$\mathcal{B}_i = \sigma\{Y_j, Y'_k;\, k \le \nu_i, j \le \nu_{i-1} + n_0\}$$

for i even. The σ-field $\mathcal{B}_i$ then comprises all random variables involved in the first i coupling trials (there are $i + 1$ of them if you count the possibility that $Y_0 = Y'_0$). We have

$$\|U_i \cdot I(\tau \ge i)\|^\alpha = \mathbf{E}[U_i^\alpha \cdot I(\tau \ge i)] = \mathbf{E}[\mathbf{E}[U_i^\alpha \| \mathcal{B}_{i-1}] \cdot I(\tau \ge i)]\,.$$

Now

$$\begin{aligned}\mathbf{E}[U_i^\alpha \| \mathcal{B}_{i-1}] &= \mathbf{E}[U_i^\alpha \| B_{i-1}] = \mathbf{E}[(S_{\nu_{i-1}+n_0} - S_{\nu_{i-1}} + B_i)^\alpha \| B_{i-1}] \\ &\le C + C \cdot \mathbf{E}[B_i^\alpha \| B_{i-1}];\end{aligned}$$

a prime may be missing on the S variables. Here, and often below, C and c are used to denote generic constants >0. A moment's thought and use of (4.1)(i) renders

$$\mathbf{E}[B_i^\alpha \| B_{i-1}] \leq C + C \cdot B_{i-1}\,.$$

Hence

$$\begin{aligned}\mathbf{E}[U_i^\alpha \cdot I(\tau \geq i)] &\leq \mathbf{E}[(C + C \cdot B_{i-1}) \cdot I(\tau \geq i)] \\ &\leq C \cdot (1-\gamma)^i + C \cdot \mathbf{E}[B_{i-1} \cdot I(\tau \geq i)]\,,\end{aligned}$$

and $\mathbf{E}[T^\alpha] < \infty$ now follows from (4.4) if we are able to prove that $\Sigma_1^\infty \mathbf{E}[B_{i-1} \cdot I(\tau \geq i)]^{1/\alpha} < \infty$. But for $i \geq 2$,

$$\mathbf{E}[B_{i-1} \cdot I(\tau \geq i)] \leq \mathbf{E}[\mathbf{E}[B_{i-1} \| \mathcal{B}_{i-2}] \cdot I(\tau \geq i-1)]\,,$$

and if $\mathbf{E}[B_{i-1} \| \mathcal{B}_{i-2}]$ is estimated as $\mathbf{E}[U_i^\alpha \| \mathcal{B}_{i-1}]$ above by using (4.1)(ii) with $\rho = \frac{1}{2}$, say, we obtain

$$\mathbf{E}[B_{i-1} \cdot I(\tau \geq i)] \leq C \cdot (1-\gamma)^{i-1} + \tfrac{1}{2} \cdot \mathbf{E}[B_{i-2} \cdot I(\tau \geq i-1)]$$

and it follows that

$$\mathbf{E}[B_{i-1} \cdot I(\tau \geq i)] \leq C \cdot i \cdot [\max(1-\gamma, \tfrac{1}{2})]^i\,,$$

so

$$\sum_1^\infty \mathbf{E}[B_{i-1} \cdot I(\tau \geq i)]^{1/\alpha} < \infty\,.$$

Now recall our restriction that $Y_0 \equiv 0$. To get rid of that, put a new zero point at $\min(Y_0, Y_0')$, and we find that

$$\mathbf{E}_{aa'}[T^\alpha] \leq C \cdot (\mathbf{E}_{aa'}[\min(Y_0, Y_0')^\alpha] + \mathbf{E}_{0,b}[T^\alpha])\,,$$

where b is the distribution of $\max(Y_0, Y_0') - \min(Y_0, Y_0')$.

The case $\alpha < 1$ remains to be handled. Without further elaboration this time, we assume that $Y_0 \equiv 0$. Since $(\Sigma_1^\infty x_i)^\alpha \leq \Sigma_1^\infty x_i^\alpha$ when $x_i \geq 0$ and $\alpha < 1$, the inequality (4.3) implies that

$$\mathbf{E}[T^\alpha] \le \mathbf{E}[Y_0'^\alpha] + \sum_1^\infty \mathbf{E}[U_i^\alpha \cdot I(\tau \ge i)]\,.$$

For a typical term of the sum we obtain

$$\begin{aligned}\mathbf{E}[U_i^\alpha \cdot I(\tau \ge i)] &= \mathbf{E}[\mathbf{E}[U_i^\alpha \cdot I(\tau \ge i) \| Y_0']] \\ &\le \mathbf{E}[\mathbf{E}[U_i \| Y_0']^\alpha \cdot \mathbf{E}[I(\tau \ge i) \| Y_0']^{\alpha_1}]\,,\end{aligned}$$

where $\alpha + \alpha_1 = 1$, due to Hölder's inequality. By using Lemma (4.1)(ii) as above, we find that $\mathbf{E}[U_i \| Y_0'] \le C \cdot i + C \cdot Y_0'$, implying that $\mathbf{E}[U_i \| Y_0']^\alpha \le C \cdot i^\alpha + C \cdot Y_0'^\alpha$. We also have $\mathbf{E}[I(\tau \ge i) \| Y_0']^{\alpha_1} \le (1-\gamma)^{\alpha_1}$, and $\mathbf{E}[T^\alpha] < \infty$ may now be concluded. □

Yes, that was a long proof. But now we have a powerful result at our disposal; the rewards will come soon.

Being traditional, we say that a sequence $(x_n)_1^\infty$ tends to 0 geometrically fast if there exists a $\rho > 1$ such that $|x_n| \le C \cdot \rho^{-n}$ for all $n \ge 1$, and use the term "exponentially fast" for functions defined on $\mathbb{R}_+$. If $\mathbf{E}[\rho^T] < \infty$ for a $\rho > 1$, then $\|\mathbf{P}(\theta_n V \in \cdot\,) - \mathbf{P}(\theta_n V' \in \cdot\,)\|$ tends to 0 geometrically fast. That finiteness is rather easily established, under the natural conditions that for a $\rho_1 > 1$

$$\sum_0^\infty \rho_1^i \cdot a_i, \quad \sum_0^\infty \rho_1^i \cdot a_i', \quad \text{and} \quad \sum_0^\infty \rho_1^i \cdot p_i$$

are finite. To prove that, we again make the innocent assumption that $Y_0 \equiv 0$ and notice that

$$\mathbf{E}[\rho^T] \le \lim_{n\to\infty} \mathbf{E}[\rho^{Y_0' + \Sigma_1^n U_i \cdot I(\tau \ge i)}]\,. \tag{4.5}$$

Now the right-hand expectation of (4.5) equals

$$1 + \sum_0^{n-1} \mathbf{E}[\rho^{Y_0'} \cdot (\rho^{\Sigma_1^{j+1} U_i \cdot I(\tau \ge i)} - \rho^{\Sigma_1^j U_i \cdot I(\tau \ge i)})]\,, \tag{4.6}$$

and the jth term of that sum,

$$\begin{aligned}&\le \mathbf{E}[\rho^{Y_0'} \cdot \rho^{\Sigma_1^{j+1} U_i} \cdot I(\tau \ge j+1)] \\ &\le \mathbf{E}[\rho^{4 \cdot Y_0'}]^{1/4} \cdot \mathbf{E}[\rho^{4 \cdot \Sigma_1^{j+1} U_i}]^{1/4} \cdot \mathbf{P}(\tau \ge j+1)^{1/2}\,.\end{aligned}$$

due to the Schwarz inequality. By conditioning on $\mathcal{B}_j$ and then on $\mathcal{B}_{j-1}$, we realize as above that

$$\mathbf{E}[\rho^{4\cdot\Sigma_1^{j+1}U_i}] \le C_\rho^{j+1},$$

where $C_\rho \downarrow 1$ as $\rho \downarrow 1$. We can now choose a $\rho > 1$ so small that $\mathbf{E}[\rho^{4\cdot Y_0'}]^{1/4} < \infty$ and $C_\rho \cdot (1-\gamma) < 1$, rendering the sum in (4.6) finite. That $\mathbf{E}[\rho^T] < \infty$ now follows from (4.5), and we use (I.3.10) to get a geometric rate of convergence result.

With methods similar to those above, finiteness of $\mathbf{E}_{aa'}[\varphi(T)]$ may be established for a wide class of nondecreasing functions φ: $\mathbf{Z}_+ \to \mathbf{R}_+$ under the natural conditions on a and a'. Finite moment results of the coupling time for random walks with drift are also available; cf. § 6.

5. Renewal sequences. What we need to compare the renewal sequences v, v' of S, S' are the inequality

$$|v_n - v_n'| \le 2 \cdot \mathbf{P}_{aa'}(T > n)$$

and Theorem (4.2). Actually, that "2" is superfluous: recall the discussion after (Int.1.2).

(5.1) Theorem. *For an $\alpha > 0$, suppose that μ_α, $m_\alpha(a)$, and $m_\alpha(a')$ are finite. Then*

(i) $|v_n - v_n'| = o(n^{-\alpha})$, *and*
(ii) $\Sigma_1^\infty n^{\alpha-1} \cdot |v_n - v_n'| < \infty$.
(iii) *If $\alpha \ge 1$, then the series $\Sigma_0^\infty (v_n - v_n')$ is convergent with the sum $s_{aa'} = \lambda \cdot \mathbf{E}_{aa'}[Y_0' - Y_0] = \lambda \cdot (m_1(a') - m_1(a))$, and*
(iv) $|\Sigma_0^n (v_i - v_i') - s_{aa'}| = o(n^{-(\alpha-1)})$.

Proof. The first lines of § 4 give the arguments for (i). Further, for $\alpha > 0$,

$$\sum_1^\infty n^{\alpha-1} \cdot |v_n - v_n'| \le 2 \cdot \sum_1^\infty n^{\alpha-1} \cdot \mathbf{P}_{aa'}(T > n)$$

$$\le C \cdot \sum_1^\infty n^\alpha \cdot \mathbf{P}_{aa'}(T = n) = C \cdot \mathbf{E}_{aa'}[T^\alpha]$$

[cf. (I.3.9)], which is finite due to Theorem (4.2), and (ii) is proved. For $\alpha \geq 1$, (ii) renders $\Sigma_0^\infty (v_n - v_n')$ absolutely convergent, hence convergent. There is a beautiful probabilistic idea for the calculation of that sum! Let ξ, ξ' be the random variables defined by

$$T = Y_0 + \sum_1^{\xi} Y_i = Y_0' + \sum_1^{\xi'} Y_i' .$$

An application of Wald's lemma yields

$$m_1(a) + \mu \cdot \mathbf{E}_{aa'}[\xi] = m_1(a') + \mu \cdot \mathbf{E}_{aa'}[\xi'] ,$$

so

$$\mathbf{E}_{aa'}[\xi - \xi'] = \lambda \cdot (m_1(a') - m_1(a)) .$$

But

$$\sum_0^T V_n = 1 + \xi \quad \text{and} \quad \sum_0^T V_n' = 1 + \xi' ;$$

thus

$$\sum_0^\infty (v_n - v_n') = \mathbf{E}_{aa'}\left[\sum_0^T (V_n - V_n')\right] = \mathbf{E}_{aa'}[\xi - \xi']$$

$$= \lambda \cdot (m_1(a') - m_1(a)) .$$

For $\alpha \geq 1$, (ii) implies that

$$\sum_n^\infty |v_i - v_i'| = o(n^{-(\alpha - 1)}) ;$$

hence, by (iii),

$$\left| \sum_0^n (v_i - v_i') - s_{aa'} \right| = o(n^{-(\alpha - 1)}) ,$$

and we have proved (iv), too. □

If $\mu_\alpha < \infty$ for an $\alpha \geq 1$, we have $m_{\alpha-1}(c) < \infty$ for the choice of a' that makes V' stationary: namely, the distribution c of (2.1). The moment $m_1(c)$ is easily calculated, it equals $\lambda \cdot (\mu_2 - \mu)/2$. Theorem (5.1) yields

(5.2) (i) $|v_n - \lambda| = o(n^{-(\alpha-1)})$ for $\alpha > 1$ if μ_α, $m_{\alpha-1}(a) < \infty$,
(ii) $\Sigma_1^\infty n^{\alpha-2} \cdot |v_n - \lambda| < \infty$ for $\alpha > 1$ if μ_α and $m_{\alpha-1}(a) < \infty$, and
(iii) $|\Sigma_0^n v_i - (n+1) \cdot \lambda - (\lambda^2(\mu_2 - \mu)/2 - \lambda \cdot m_1(a))| = o(n^{-(\alpha-2)})$ for $\alpha \geq 2$ if μ_α and $m_{\alpha-1}(a) < \infty$.

When using analytical methods to prove (iii), it is not easy at the outset to have an idea of what constant should be added to $(n+1) \cdot \lambda$ to get an improved approximation of $\Sigma_0^n v_i$, or, for that matter, to have the hope that such a constant exists! For us, an application of Wald's lemma was decisive.

Theorem (5.1) also renders fluctuation results for renewal sequences. Indeed, let $Y_0' \overset{\mathcal{D}}{=} Y_0 + 1$, which means that $a_i' = a_{i-1}$ for $i \geq 1$. Then $v_i' = v_{i-1}$ for $i \geq 1$, and (5.1) gives

(5.3) (i) $|v_n - v_{n-1}| = o(n^{-\alpha})$, and
(ii) $\Sigma_1^\infty n^{\alpha-1} \cdot |v_n - v_{n-1}| < \infty$ for $\alpha > 0$ if μ_α and $m_\alpha(a) < \infty$.

For the sequence $(u_n)_0^\infty$ we obtain from (5.2)–(5.3) that if $\mu_\alpha < \infty$ for an $\alpha > 1$, then

(5.4) (i) $|u_n - \lambda| = o(n^{-(\alpha-1)})$,
(ii) $\Sigma_1^\infty n^{(\alpha-2)} \cdot |u_n - \lambda| < \infty$,
(iii) $|\Sigma_0^\infty u_i - (n+1) \cdot \lambda - \lambda^2(\mu_2 - \mu)/2| = o(n^{-(\alpha-2)})$ for $\alpha \geq 2$,
(iv) $|u_n - u_{n-1}| = o(n^{-\alpha})$, and
(v) $\Sigma_1^\infty n^{(\alpha-1)} \cdot |u_n - u_{n-1}| < \infty$.

For the case $\mu = \infty$, the convergence result $v_n \to 0$ is a consequence of the ergodic results for null-recurrent Markov chains, to be established in § 9. Recall the Markov chain D from § 2; we have

$$v_n = \mathbf{P}(V_n = 1) = \mathbf{P}(D_n = 0)$$

[= the n-step transition probability $p_{i0}^{(n)}$ for D if $a = \delta_i$], and that probability does indeed tend to 0 as $n \to \infty$.

6. Notes. This section is based on Pitman [129] and Lindvall [106]; the generalized moment result mentioned at the end of § 4 can be found in the latter reference. Berbee [24] proves that the rate results in § 5 are the best possible. See also Davies and Grübel [47]. A survey of Russian coupling is given by Anichkin [7]. For a fine account of analytical discrete renewal theory, see Freedman [62], which also gives a proof of (1.1). An entire part on renewal theory without any renewal equation—that is purely intentional!

2. MARKOV CHAINS

7. Notation. Let $P = (p_{ij})$ be a transition matrix indexed by a countable set, meant to be a state space; it is no restriction to let it equal $\mathbb{Z}_+$. We assume that P governs an aperiodic and irreducible Markov chain $X = (X_n)_0^\infty$.

Faithful to the relevant literature (but unfortunately, giving new meaning to certain symbols used in Part 1), we denote initial distributions by λ or μ, and reserve π for stationary distributions. With $\mathbf{P}_\lambda$ for the probability measure on the sample space of X when $X_0 \overset{\mathcal{D}}{=} \lambda$, we have

$$\mathbf{P}_\lambda(X_n = j) = p_{\lambda j}^{(n)} = \sum \lambda_i p_{ij}^{(n)} = (\lambda P^n)_j$$

$$= \textit{the } j\text{th element of the vector } \lambda P^n \;;$$

we adopt the usage of row vectors for distributions on $\mathbb{Z}$ or a subset of that space. We let $\mathbf{P}_\lambda$ be $\mathbf{P}_i$ when $\lambda = \delta_i$, and $\mathbf{E}_i$ is the associated expectation.

8. Positive recurrent chains. For $i \geq 0$, let

$$T_i = \min\{n; n \geq 1, X_n = i\} .$$

If one $m_i = \mathbf{E}_i[T_i]$ is finite, so are all; this is the positive recurrent case. In this, it is well known that $\pi = (\pi_i)_0^\infty = (1/m_i)_0^\infty$ is a solution to $\pi = \pi P$, hence a stationary distribution (it is unique). We also have $\pi_i = \pi_i^{\circ}/m_0$, where

$$\pi_i^{\circ} = \mathbf{E}_0[\#\{j; 0 \le j < T_0, X_j = i\}]. \tag{8.1}$$

We shall prove a general version of that in Chapter III.

For the comparison of λP^n and μP^n, Doeblin coupling will be used. So let $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ be independent, with $X_0 \stackrel{\mathscr{D}}{=} \lambda$, $X'_0 \stackrel{\mathscr{D}}{=} \mu$, and defined on a standard space with probability measure $\mathbf{P}_{\lambda\mu}$. For applications of (I.2.8), let

$$T^* = \min\{n; X_n = X'_n\}.$$

It is very natural to hope T^* should serve us best. However, there seems to be no application in the positive recurrent case where T^* is essentially superior to

$$T = \min\{n; X_n = X'_n = 0\}, \tag{8.2}$$

the use of which makes our tools from discrete renewal theory immediately applicable. Indeed,

$$\{n; X_n = 0\} \quad \text{and} \quad \{n; X'_n = 0\}$$

constitute two independent, aperiodic renewal processes, S and S', say, with recurrence times distribution $p =$ that of T_0 under $\mathbf{P}_0$, with the expectation $m_0 = m_1(p)$, assumed throughout this section to be finite. As for the discrete renewal theorem, it is convenient to use a weak coupling in order to settle the basic inequality. To obtain

$$\|\lambda P^n - \mu P^n\| \le 2 \cdot \mathbf{P}_{\lambda\mu}(T > n) \tag{8.3}$$

it suffices to prove that $(\theta_T X, T) \stackrel{\mathscr{D}}{=} (\theta_T X', T)$. But the arguments for that are very close to those used in § 3 to prove that $(\theta_T V, T) \stackrel{\mathscr{D}}{=} (\theta_T V', T)$! Hence (8.3) holds, and since $T < \infty$ a.s. due to the result of § 2 and the observation above that the visits to 0 constitute renewal processes, we obtain

$$\|\lambda P^n - \mu P^n\| \to 0 \quad \text{as} \quad n \to \infty \tag{8.4}$$

for any initial distributions λ and μ. With $\mu = \pi$ we have in particular

$$(8.5) \qquad \|\lambda P^n - \pi\| \to 0 \quad \text{as} \quad n \to \infty$$

for any initial distribution λ.

The distributions a, a' of the delays Y_0, Y_0' in the embedded renewal processes S, S' are those of $T^{(0)}$ under $\mathbf{P}_\lambda$, $\mathbf{P}_\mu$, respectively, where

$$T^{(i)} = \min\{n; n \geq 0, X_n = i\}$$

for $i \in \mathbb{Z}_+$. We shall use the fact that $(b\mathcal{M}_s(\mathbb{Z}_+), \|\cdot\|)$ is a Banach space where $\|\nu\| = \Sigma\,|\nu_i|$ for signed bounded measures; of course that space is nothing but $l^1(\mathbb{Z}_+)$. In particular, every absolutely convergent series in that space has a sum. As for Theorem (5.1), we apply Theorem (4.2), now to (8.3); the proof will be brief.

(8.6) Theorem. *For an $\alpha > 0$, suppose that $m_\alpha(p)$, $m_\alpha(a)$, and $m_\alpha(a')$ are finite. Then*

(i) $\|\lambda P^n - \mu P^n\| = o(n^{-\alpha})$,
(ii) $\Sigma_1^\infty n^{\alpha-1} \cdot \|\lambda P^n - \mu P^n\| < \infty$,
(iii) *If $\alpha \geq 1$, then the series $\Sigma_0^\infty (\lambda P^n - \mu P^n)$ is convergent with the sum $\nu = (\nu_i)_0^\infty$, where $\nu_i = (\mathbf{E}_\mu[T^{(i)}] - \mathbf{E}_\lambda[T^{(i)}])/m_i$, and*
(iv) $\|\Sigma_0^n (\lambda P^i - \mu P^i) - \nu\| = o(n^{-(\alpha-1)})$.

Proof. For $\alpha \geq 1$, (ii) makes $\Sigma_0^k (\lambda P^n - \mu P^n)$ to a Cauchy sequence in $b\mathcal{M}_s(\mathbb{Z}_+)$; it is hence convergent. After that observation you should be curious only about the evaluation of ν_i in (iii). For $i = 0$, ν_i is nothing but the sum $s_{aa'}$ of (5.1)(iii):

$$\nu_0 = \sum_0^\infty (p_{\lambda 0}^{(n)} - p_{\mu 0}^{(n)}) = \sum_0^\infty (v_n - v_n')$$

where (v_n), (v_n') are the renewal sequences associated with S and S'. For $i \neq 0$, change perspective and imagine that state i was our choice of reference state from the outset. $\square$

...oosing $\mu = \pi$, we have $\pi P^n = \pi$ for all $n \geq 0$, and obtain from ...6) that if $m_\alpha(p)$ and $m_{\alpha-1}(a)$ are finite, then

...7) (i) $\|\lambda P^n - \pi\| = o(n^{-(\alpha-1)})$,
(ii) $\Sigma_1^\infty n^{\alpha-2} \cdot \|\lambda P^n - \pi\| < \infty$ for $\alpha > 1$, and
(iii) $\|\Sigma_0^n \lambda P^i - (n+1)\cdot \pi - \nu\| = o(n^{-(\alpha-2)})$, where ν is displayed in (8.6)(iii), for $\alpha \geq 2$.

Considering the amount of attention paid to Markov chains, it is ...emarkable that these results are so new; it is one indication of the ...ower of the coupling method.

To obtain fluctuation results for the transition probabilities, put $\mu = \lambda P$. Theorem (8.6) then gives

(8.8) (i) $\|\lambda P^n - \lambda P^{n+1}\| = o(n^{-\alpha})$, and
(ii) $\Sigma_1^\infty n^{\alpha-1} \cdot \|\lambda P^n - \lambda P^{n+1}\| < \infty$ for $\alpha > 0$ if $m_\alpha(p)$ and $m_\alpha(a) < \infty$.

With $\lambda = \delta_i$ for some i, (8.8) says that

$$\sum_j |p_{ij}^{(n)} - p_{ij}^{(n+1)}| = o(n^{-\alpha}), \quad \text{and}$$

$$\sum_{n=0}^\infty n^{\alpha-1} \cdot \sum_j |p_{ij}^{(n)} - p_{ij}^{(n+1)}| = \sum_j \sum_{n=0}^\infty n^{\alpha-1} \cdot |p_{ij}^{(n)} - p_{ij}^{(n+1)}| < \infty$$

for $\alpha > 0$, if $m_\alpha(p) < \infty$.

If the state space is finite, equal to $\{0, \ldots, L\}$, say, it is easy to prove that the rate of convergence toward stationarity is geometric. Indeed, let $k > 0$ be so large that

$$\rho = \min_{i,j} p_{ij}^{(k)} > 0 .$$

Then, for X, X' independent, we have

$$\mathbf{P}_{\lambda\mu}(X_k \neq X'_k) \leq 1 - \rho$$

for any λ, μ. Hence

$$\mathbf{P}_{\lambda\mu}(X_i \neq X'_i \text{ for } i \leq n) \leq (1-\rho)^{[n/k]} ,$$

where $[\cdot]$ denotes "integer part of"; certainly, the right-hand side tends to 0 geometrically fast as $n\to\infty$, and the left equals $\mathbf{P}_{\lambda\mu}(T^* > n)$, where T^* is from p. 35.

So far, we have constructed our couplings in terms of the independent processes X and X'. We may also build them from the transition probabilities of a bivariate chain. For example, let

$$(8.9) \qquad \begin{aligned} p_{(ij)(km)} &= p_{ik} \cdot p_{jm} \quad \text{if} \quad i \neq j \quad \text{and} \\ p_{(ii)(kk)} &= p_{ik} \end{aligned}$$

for i, j, k and $m \in \mathbb{Z}_+$. Then (8.9) indicates Doeblin coupling, as is easily verified. But if we intend to strive toward the diagonal as quickly as possible, we should in each step of the bivariate chain make a γ coupling. This idea leads us to the transitions

$$(8.10) \qquad p_{(ij)(\cdot)} = \nu_{ij},$$

where ν_{ij} is the γ coupling of $p_{i\cdot}$ and $p_{j\cdot}$. Recalling the proof Theorem (I.5.2), it turns out that (8.10) means a coupling with

(8.11)

$$\begin{aligned} &p_{(ij)(kk)} = p_{ik} \wedge p_{jk}, \quad \text{hence} \quad p_{(ii)(kk)} = p_{ik}, \quad \text{and} \\ &p_{(ij)(km)} = (p_{ik} - p_{jk})^+ \cdot (p_{jm} - p_{im})^+/(1-\gamma_{ij}) \qquad \text{for} \quad i \neq j, k \neq m, \\ &\quad \text{where} \quad \gamma_{ij} = \sum_k p_{ik} \wedge p_{jk}. \end{aligned}$$

This is the Markov chain version of the so-called Vasershtein coupling; cf. § V.12. We shall have no use for it, despite its appeal.

Notice that for any $A \in \mathscr{X}_+^\infty$,

$$\begin{aligned} |\mathbf{P}_\lambda(\theta_n X \in A) - \mathbf{P}_\mu(\theta_n X' \in A)| &= \Big|\sum_i \mathbf{P}_i(A) \cdot (p_{\lambda i}^{(n)} - p_{\mu i}^{(n)})\Big| \\ &\leq \sum_i |p_{\lambda i}^{(n)} - p_{\mu i}^{(n)}| = \|\lambda P^n - \mu P^n\|. \end{aligned}$$

Hence

(8.12)

$$\|\mathbf{P}_\lambda(\theta_n X \in \cdot\,) - \mathbf{P}_\mu(\theta_n X' \in \cdot\,)\| \le 2 \cdot \|\lambda P^n - \mu P^n\|$$
$$= 2 \cdot \|\mathbf{P}_\lambda(X_n \in \cdot\,) - \mathbf{P}_\mu(X'_n \in \cdot\,)\| .$$

That inequality has nothing to do with coupling. The estimate to establish (8.12) was crude; actually, equality holds without that "2," as we shall learn in Chapter III.

9. Null-recurrent chains. For Markov chains that are not positive recurrent, consider the asymptotic properties

(9.1) $$p_{ij}^{(n)} \to 0 \quad \text{as} \quad n \to \infty \quad \text{for all} \quad i, j \ge 0 ,$$

and the usual

(9.2) $$\|\lambda P^n - \mu P^n\| \to 0 \quad \text{as} \quad n \to \infty \quad \text{for all} \quad \lambda, \mu .$$

We know that a transition matrix P governs transient chains if and only if

(9.3) $$\sum_n p_{ij}^{(n)} < \infty$$

for some states i, j, and if so, the sum is finite for all pairs i, j. Hence, in the transient case (9.1) is immediately seen to hold true. To the challenge (9.2) there is, however, no interesting general response for the transient chains; a number of special cases (random walks) are examined in Part 3.

What about the null-recurrent case? When the conditions on aperiodicity and irreducibility are in force, result (9.2) is the celebrated Orey's theorem. Ornstein's coupling, being the subject matter of § 12, seems to offer the best probabilistic proof of (9.2), hence we wait to get that settled. But the Doeblin coupling also throws light on (9.1)–(9.2) in the null-recurrent case. Indeed, let $\tilde{X}_n = (X_n, X'_n)$, $n \ge 0$, be the Markov chain on $\mathbb{Z}_+^2$ we get when X, X' are independent; it is aperiodic and irreducible. If it is transient, then due to (9.3) we have

$$\sum \tilde{p}_{(ii)(jj)}^{(n)} = \sum p_{ij}^{(n)2} < \infty ,$$

and hence $p_{ij}^{(n)} \to 0$ for all pairs i, j. When recurrent, we have of course

$$(9.4) \qquad \mathbf{P}_{(ij)}(T < \infty) = 1$$

for all pairs (i, j), where

$$T = \min\{n; \tilde{X}_n = (0,0)\} .$$

But T is a familiar coupling time, and (9.2) follows from (9.4) in the usual way.

It always holds in the null-recurrent case that (9.2) implies (9.1). To prove that, suppose that (9.1) is false; then for some i, a standard application of the diagonal method yields a subprobability $\pi^\circ = (\pi_j^\circ)_0^\infty \neq (0)_0^\infty$ such that

$$p_{ij}^{(k_n)} \to \pi_j^\circ \qquad \text{as} \quad n \to \infty$$

for some sequence $(k_n)_0^\infty$ ($\Sigma_0^\infty \pi_j^\circ \leq 1$ follows from Fatou's lemma). Letting $\lambda = \delta_i$ and $\mu = \delta_i P$ in (9.2), we obtain that π° must be an invariant measure for P. Hence π given by $\pi_j = \pi_j^\circ / \Sigma \pi_k^\circ$ is a stationary distribution. But there is no such distribution if X is not positive recurrent.

Actually, the result (9.1) is now set.

10. An observation. For sequences $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$, a coupling proof of $\|\mathbf{P}(X_n \in \cdot) - \mathbf{P}(X'_n \in \cdot)\| \to 0$ usually follows the line of § 9. But taking a close look at that, we realize that a coupling of X and X' is more than we need if only the distance between the distributions of X_n and X'_n are to be compared. Indeed, if $\hat{X} = (\hat{X}_n)_0^\infty$ and $\hat{X}' = (\hat{X}'_n)_0^\infty$ are random sequences such that

$$(10.1) \qquad X_n \stackrel{\mathcal{D}}{=} \hat{X}_n \quad \text{and} \quad X'_n \stackrel{\mathcal{D}}{=} \hat{X}'_n \qquad \text{for every} \quad n \geq 0$$

and

$$\hat{X}_n = \hat{X}'_n \qquad \text{for} \quad n \geq \text{a random time } T ,$$

$(\hat{X}_n, \hat{X}'_n)$ is a coupling of X_n and X'_n for every n, $\{\hat{X}_n \neq \hat{X}'_n\} \subset$ $\{T > n\}$, and $\|\mathbf{P}(X_n \in \cdot) - \mathbf{P}(X'_n \in \cdot)\| \to 0$ if $T < \infty$ a.s. Notice that this does not imply that $\|\mathbf{P}(\theta_n X \in \cdot) - \mathbf{P}(\theta_n X' \in \cdot)\| \to 0$ as $n \to \infty$. In the Markov chain case, however, (8.12) may be applied.

11. **Notes.** The account of positive recurrent chains is based on Pitman [129], a pioneering work. The Markov chain version of the Vasershtein coupling is from Griffeath [66]. For coupling of Markov chains in statistical theory, see Lauritzen [101].

3. RANDOM WALK

12. The Ornstein coupling. Let $Y_1, Y_2, \ldots,$ be i.i.d. integer-valued random variables, with common distribution $p = (p_i)_{-\infty}^{\infty}$ on $\mathbf{Z}$; $\mathbf{P}(Y_j = i) = p_i$ for all $j \geq 1$. The Y_i variables will play the role of the step sizes of a random walk $\tilde{S} = (\tilde{S}_n)_0^{\infty}$:

$$\tilde{S}_n = \sum_1^n Y_i \qquad \text{for} \quad n \geq 0$$

(hence $\tilde{S}_0 = 0$). As will be seen below, there are several good reasons to learn how to couple random walks. Our first undertaking is to prove that under minimal conditions

$$(12.1) \quad \|\mathbf{P}(k + \tilde{S}_n \in \cdot) - \mathbf{P}(j + \tilde{S}_n \in \cdot)\| \to 0 \qquad \text{as} \quad n \to \infty$$

for all k, j. Of course, with $m = k - j$,

$$\|\mathbf{P}(k + \tilde{S}_n \in \cdot) - \mathbf{P}(j + \tilde{S}_n \in \cdot)\| = \|\mathbf{P}(m + \tilde{S}_n \in \cdot) - \mathbf{P}(\tilde{S}_n \in \cdot)\|,$$

so we may let $j = 0$. Notice that

$$(12.2) \quad \|\mathbf{P}(k + \tilde{S}_n \in \cdot) - \mathbf{P}(\tilde{S}_n \in \cdot)\| = \|\delta_k * p^{*n} - p^{*n}\|.$$

For the convergence (12.1), it is immediate that the distribution p must be aperiodic. But that is not a strong enough condition, as one realizes by letting $\tilde{S}$ be a simple random walk (Y_i equals 1 or

-1). The natural condition to impose is that p is strongly aperiodic, by which we mean that

$$\text{(12.3)} \qquad \text{g.c.d.}\,\{k+i;\, p_i > 0\} = 1 \qquad \text{for all} \quad k.$$

This condition suffices to prove that $\tilde{S}$ is ergodic. To begin with, assume that p has a finite first moment ($\mathbf{E}[|Y_i|] < \infty$), and let $\tilde{S}'$ be a random walk independent of $\tilde{S}$ and with the same step size distribution p. Fix k, and let

$$T = \min\{n;\, k + \tilde{S}_n = \tilde{S}'_n\}.$$

Certainly,

$$\begin{aligned}\|\mathbf{P}(k + \tilde{S}_n \in \cdot\,) - \mathbf{P}(\tilde{S}_n \in \cdot\,)\| &= \|\mathbf{P}(k + \tilde{S}_n \in \cdot\,) - \mathbf{P}(\tilde{S}'_n \in \cdot\,)\| \\ &\le 2 \cdot \mathbf{P}(T > n).\end{aligned}$$

Now, with $V_i = Y_i - Y'_i$ and $Z_n = \Sigma_1^n V_i$, we have

$$T = \min\{n;\, Z_n = -k\}.$$

But $Z = (Z_n)_0^\infty$ is a random walk with step sizes V_i having expectation 0, which means that the condition of the Chung–Fuchs recurrence theorem is satisfied. Hence Z is recurrent, and from (12.3) we get that the distribution of the V_i variables is aperiodic, hence Z hits every state a.s., and $T < \infty$ a.s. is established.

An idea due to Ornstein is the key to the general case. Let Y''_1, $Y''_2, \ldots,$ be i.i.d. with distribution p and independent of $\tilde{S}$. Fix a large constant A (more about it later) and let

$$Y'_i = \begin{cases} Y_i & \text{if} \quad |Y_i - Y''_i| > A \\ Y''_i & \text{if not}. \end{cases}$$

An argument may be needed to convince you that $Y'_i \stackrel{\mathcal{D}}{=} p$. We have, for any k,

$$\begin{aligned}\mathbf{P}(Y'_i = k) &= \mathbf{P}(Y'_i = k, |Y_i - Y''_i| > A) + \mathbf{P}(Y'_i = k, |Y_i - Y''_i| \le A) \\ &= \mathbf{P}(Y_i = k, |Y_i - Y''_i| > A) + \mathbf{P}(Y''_i = k, |Y_i - Y''_i| \le A).\end{aligned}$$

But $P(Y_i'' = k,\ |Y_i - Y_i''| \le A) = P(Y_i = k,\ |Y_i - Y_i''| \le A)$ due to symmetry, and we are done.

The variables $V_i^* = Y_i - Y_i'$ are bounded and symmetrically distributed; hence $E[V_i^*] = 0$. The analysis above, now applied to the random walk $Z^* = (Z_n^*)_0^\infty$, where $Z_n^* = \Sigma_1^n V_i^*$, yields $T < \infty$ a.s. if Z^* is aperiodic. But it is: Choose A so large that the steps V_i^* get an aperiodic distribution.

If T_0 is a recurrence time of a zero-mean random walk, an analytical argument (Karamata's Tauberian theorem) may be used to prove that $P(T_0 > n) = O(n^{-1/2})$. Leaning on this, we can establish that $P(T > n) = O(n^{-1/2})$ for our coupling time T, and hence for any k,

$$\|\delta_k * p^{*n} - p^{*n}\| = O(n^{-1/2}) \tag{12.4}$$

when p is a strongly aperiodic distribution on $\mathbb{Z}$. In § 14, however, a simple argument for this turns up, so we do not go into further details here.

With no moment assumptions about p, (12.4) is a strong result. A well-respected analyst asked to prove it got into trouble. After fruitless work his response was: "Perhaps there are methods of real analysis that I am not aware of."

13. Null-recurrent Markov chains. The notation and assumptions of § 7 are now in force. In particular, let $Y_0, Y_1, \ldots$ denote the delay and the recurrence times of the renewal process $\{n; X_n = 0\}$ associated with a null-recurrent and aperiodic Markov chain $X = (X_n)_0^\infty$ with initial distribution λ; we denote its renewal times by S_n, $n \ge 0$, and let $\tilde{S}_0 = 0$, $\tilde{S}_n = \Sigma_1^n Y_i$ for $n \ge 1$. Hence $S_n = Y_0 + \tilde{S}_n$ for $n \ge 0$.

Let X'' be independent of X, with initial distribution μ. Define the blocks B_i, $i \ge 0$, of X as the sequences

$$B_0 = (X_0, \ldots, X_{S_0}),$$
$$B_i = (X_{S_{i-1}}, \ldots, X_{S_i}) \qquad \text{for} \quad i \ge 1$$

(cf. § 15), and let $B_0'', B_1'', \ldots$, have the obvious meaning. To begin with, assume that the recurrence times Y_i and Y_i'', $i \ge 1$, have a strongly aperiodic distribution p; it is then easy to extend the method of § 12 to a proof of $\|\lambda P^n - \mu P^n\| \to 0$. Indeed, define $X' = (X'_n)_0^\infty$ in terms of its blocks B'_i as follows:

$$B'_0 = B''_0 ,$$
$$B'_i = \begin{cases} B_i & \text{if } \ |Y_i - Y''_i| > A \\ B''_i & \text{if not.} \end{cases}$$

The constant A is from § 12. In the same way as there, one shows that $B'_i \overset{\mathscr{D}}{=} B_i$, so $X' \overset{\mathscr{D}}{=} X''$. We make a coupling of X and X' at the first common renewal of S and S'. And there is such a renewal, because $\Sigma_1^n (Y_i - Y'_i) = (Y'_0 - Y_0)$ for some n, as we have learned.

Fortunately, it is not difficult to get rid of that "strongly"; it is not natural in this connection. Assume that p is aperiodic and take K so large that

$$\{i;\ p_i^{*j} > 0 \text{ for some } j,\ 1 \le j \le K\}$$

contains two successive integers. Let τ_i, τ''_i, $i \ge 1$, be independent random variables, also independent of X, X'', such that

$$\tau_1,\ \tau_2 - \tau_1,\ \tau''_1,\ \tau''_2 - \tau''_1, \ldots$$

are i.i.d. and uniformly distributed on $\{1, 2, \ldots, K\}$. A moment's thought gives that the distribution of $\tilde{S}_{\tau_1}$ is strongly aperiodic. Redefine the blocks B_i, B''_i, $i \ge 1$, by replacing the indices S_{i-1}, S_i with $S_{\tau_{i-1}}$, S_{τ_i}, respectively, for the X process, and analogously for X'' with use of the τ''_i variables; we have let $\tau_0 = \tau''_0 = 0$. We may now proceed, as above, to obtain a common renewal at one of the S_{τ_i} times.

We conclude the proof of $\|\lambda P^n - \mu P^n\| \to 0$ by referring to (I.2.8); perhaps that will be forgotten more often than not in what follows, but the reader certainly knows that step well by now.

14. The Mineka coupling. The coupling to be introduced now was constructed for the analysis of sums of independent but not necessarily identically distributed random variables. We first consider the i.i.d. case; the notation of § 12 is in force. For $i \in \mathbb{Z}$, let

$$\alpha_i = (p_i \wedge p_{i+1})/2 ,$$

where $p_i = \mathbf{P}(Y_j = i)$. Define the distribution of the steps (Y_j, Y'_j) of a random walk $(\tilde{S}_n, \tilde{S}'_n)$, $n \ge 0$, in $\mathbb{Z}^2$ by

$$(14.1)\qquad \begin{aligned} \mathbf{P}((Y_j, Y'_j) = (i-1, i)) &= \alpha_{i-1}\,, \\ \mathbf{P}((Y_j, Y'_j) = (i, i-1)) &= \alpha_{i-1}\,, \\ \mathbf{P}((Y_j, Y'_j) = (i, i)) &= p_i - \alpha_{i-1} - \alpha_i \end{aligned}$$

for $i \in \mathbb{Z}$. Then certainly $\tilde{S}_n$, $n \ge 0$, is a random walk with step size distribution $p = (p_i)_{i\in\mathbb{Z}}$. Further, the variables $Y_j - Y'_j$ are distributed symmetrically around 0,

$$(14.2)\qquad |Y_j - Y'_j| \le 1 \quad \text{and} \quad \mathbf{P}(Y_j - Y'_j = 1) = \sum \alpha_i\,.$$

Denote that sum by α. To begin with, assume that $\alpha > 0$, which holds if and only if $p_i \wedge p_{i+1} > 0$ for some i. Due to (14.2) it is now trivial to prove that $T < \infty$ a.s. for $T = \min\{n; Z_n = -k\}$, where $Z = (Z_n)_0^\infty$ is the random walk with steps $V_i = Y_i - Y'_i$. Indeed, Z is a symmetric simple random walk (with the possibility that $V_i = 0$, however) so we just have to wait for Z to hit the state $-k$, something that eventually happens.

Hence $\|\delta_k * p^{*n} - p^{*n}\| \to 0$ for all k is proved again. For the estimate $O(n^{-1/2})$ we must show that $n^{1/2} \cdot \mathbf{P}(T > n)$ is bounded in n. Let

$$\xi = \#\{i; i \le n \text{ and } V_i \ne 0\}$$

and $\tilde{Z}_n = \Sigma_{i=0}^n \tilde{V}_i$, where $\tilde{V}_i =$ the ith V_j such that $V_j \ne 0$. We have that ξ is $\mathrm{Bin}(n, 2\alpha)$ distributed, and $\tilde{Z}_n$, $n \ge 0$, is a symmetric simple random walk. Put

$$\tilde{T} = \min\{j; \tilde{Z}_j = -k\}\,.$$

Elementary calculations, using the probability generating function of $\tilde{T}$, show that $P(\tilde{T} > n) = O(n^{-1/2})$. Indeed, for $k = 1$ that generating function equals $(1 - (1 - s^2)^{1/2})/s$, an expansion gives $\mathbf{P}(\tilde{T} = j) \le C \cdot j^{-3/2}$ for $j \ge 1$, and $\mathbf{P}(\tilde{T} > n) = O(n^{-1/2})$ follows from that, for all k. We get

$$\begin{aligned}\mathbf{P}(T>n) &= \mathbf{P}(T>n,\ \xi \geq \alpha n)+\mathbf{P}(T>n,\ \xi<\alpha n)\\ &\leq \mathbf{P}(T>n,\ \xi \geq \alpha n)+\mathbf{P}(|\xi-2\cdot \alpha n| \geq \alpha n)\,.\end{aligned}$$

Now the last probability is not greater than $1/\alpha n$ due to Chebyshev's inequality, and

$$\mathbf{P}(T>n,\ \xi \geq \alpha n) \leq \mathbf{P}(\tilde{T}>[\alpha n])\,;$$

hence $n^{1/2}\cdot \mathbf{P}(T>n)$ is indeed bounded in n.

What should be done if $\alpha=0$? If there exists an n_0 such that $\mathbf{P}(\tilde{S}_{n_0}=i) \wedge \mathbf{P}(\tilde{S}_{n_0}=i+1)>0$ for some i, then we achieve everything above, with new Y_i variables $\tilde{S}_{n_0}$, $\tilde{S}_{2n_0}-\tilde{S}_{n_0}, \ldots$ and use of

$$\|\delta_k * p^{*n}-p^{*n}\|=\|(\delta_k-\delta_0)*p^{*n}\| \leq \|(\delta_k-\delta_0)*(p^{*n_0})^{*[n/n_0]}\|\,.$$

Our new α is, of course, $\sum_i p_i^{*n_0} \wedge p_{i+1}^{*n_0}$. There is such an n_0 if p is strongly aperiodic. To prove that, let

$$G=\{j;\ p_i^{*n} \wedge p_{i+j}^{*n}>0 \text{ for some } n \text{ and } i\}\,.$$

This set is closed under addition. Indeed, for $j_1, j_2 \in G$ there exist i_1, i_2, n_1, n_2 such that $p^{*n_1}(i_1) \wedge p^{*n_1}(i_1+j_1)>0$ and $p^{*n_2}(i_1) \wedge p^{*n_2}(i_2+j_2)>0$, which implies that $p^{*(n_1+n_2)}(i_1+i_2) \wedge p^{*(n_1+n_2)}(i_1+i_2+j_1+j_2)>0$ [we write $p^{*n}(i)$ instead of p_i^{*n} for the readability].

We also have that if $j \in G$, so does $-j$. Hence G is a subgroup of $\mathbb{Z}$, so it equals L_m for some $m \geq 1$, where

$$L_m=\{m\cdot i;\ i \in \mathbb{Z}\}\,.$$

Let

$$D=\{i-j;\ p_i>0,\ p_j>0\}\,.$$

We have $D \subseteq G$, and, due to the strong aperiodicity of p, g.c.d. $(D)=1$. If $\bar{D}$ denotes the smallest additive group containing D, then $\bar{D} \subset G$. But g.c.d. $(\bar{D})=1$ forces $\bar{D}$ to equal $\mathbb{Z}$, and we have proved that $G=\mathbb{Z}$.

We finish this section with an exercise. Let Y_j, $j \geq 1$, be independent variables such that

$$\mathbf{P}(Y_j = 0) = \mathbf{P}(Y_j = 1) = \alpha_j, \quad \text{and}$$
$$\mathbf{P}(Y_j = k_j) = 1 - 2 \cdot \alpha_j ,$$

where $0 < \alpha_j \leq \frac{1}{2}$ and $(k_j)_1^\infty$ is an increasing sequence of positive integers. We search for a good coupling in order to establish ergodicity of $S_n = \Sigma_0^n Y_j$, $n \geq 0$. To that end, let S'_n, $n \geq 0$, be the parallel process, and put $S_0 = 0$, $S'_0 = 1$, for example. Show that

(i) the Vaserschtein coupling is useless if $(k_j)_1^\infty$ increases rapidly and the α_j's all equal $\frac{1}{4}$, say,

(ii) the Ornstein coupling is successful if and only if $\Sigma_j \alpha_j^2 = \infty$, and

(iii) the Mineka coupling is superior to Ornstein's: It is successful if and only if $\Sigma_j \alpha_j = \infty$.

A hint for (i): Simplify by observing that $|S_n - S'_n|$, $n \geq 0$, is actually a Markov chain, and pick the sequence $(k_j)_1^\infty$ such that $\mathbf{P}(|S_n - S'_n| \to \infty \text{ as } n \to \infty) > 0$.

15. Blocks. There are good reasons to be more precise than in § 13 about what to mean by blocks; they will also play a crucial role in what follows next. For a state space $(E, \mathscr{E})$, a block is an element in

$$E_b = E^\infty \cup \bigcup_{i=0}^\infty E_0^i ,$$

where E_0^i is the space $\{(x_0, \ldots, x_i);\ x_i \in E\}$, and $\mathscr{E}_0^i$ is the standard σ-field on that space; we include E^∞ in order to allow blocks of infinite length. The space E_b is endowed with the σ-field

$$\mathscr{E}_b = \sigma\{A;\ A \in \mathscr{E}^\infty \text{ or } A \in \mathscr{E}_0^k \text{ for some } k \geq 0\} .$$

In fact,

$$\mathscr{E}_b = \left\{ \bigcup_0^\infty A_i;\ A_i \in \mathscr{E}^\infty \text{ or } A_i \in \mathscr{E}_0^i \text{ for some } j \right\} .$$

If $X_0, X_1, \ldots$, are random elements in $(E, \mathscr{E})$ defined on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and $\tau \in \bar{\mathbb{Z}}_+$ is a random time on that same space, then

$$B = (X_0, X_1, \ldots, X_\tau)$$

$[= (X_0, X_1, \ldots)$ on $\{\tau = \infty\}]$ is a random element in $(E_b, \mathscr{E}_b)$. Indeed, let $A \in \mathscr{E}_b$. Then we have

$$\begin{aligned}\{(X_0, \ldots, X_\tau) \in A\} &= \Big(\bigcup_{k=0}^{\infty} \{(X_0, \ldots, X_k) \in A\} \cap \{\tau = k\}\Big) \\ &\quad \cup \{(X_0, X_1, \ldots) \in A\} \\ &= \Big(\bigcup_{k=0}^{\infty} \{(X_0, \ldots, X_k) \in A \cap E_0^k\} \cap \{\tau = k\}\Big) \\ &\quad \cup \{(X_0, X_1, \ldots) \in A\} \in \mathscr{E}_b\end{aligned}$$

since $\{(X_0, \ldots, X_k) \in A \cap E_0^k\} \in \mathscr{E}_0^k$ and $\{(X_0, X_1, \ldots) \in A\} \in \mathscr{E}^\infty$.

Let φ be the mapping $E_b \to \bar{\mathbb{Z}}_+$ given by $\varphi(x) = i$ for $x = (x_0, \ldots, x_i)$, an element of E_b; $i = \infty$ if x is infinite. Elements $B_0 = (Y_{00}, Y_{01}, \ldots)$, $B_1 = (Y_{10}, Y_{11}, \ldots)$, $B_2 = \cdots \in E_b$ such that $Y_{k,\varphi(B_k)} = Y_{k+1,0}$ for all k may be glued together to a sequence $X = (X_n)_0^\infty \in E^\infty$ in the obvious way: With $\tau_i = \varphi(B_i)$, let

$$X_n = \begin{cases} Y_{0n} & \text{for } n \le \tau_0, \\ Y_{1,n-\tau_0} & \text{for } \tau_0 < n \le \tau_0 + \tau_1, \end{cases} \tag{15.1}$$

and so on.

16. The Harris random walk. A Markov chain $X = (X_n)_0^\infty$ on $\mathbb{Z}$ with transition probabilities

$$\begin{aligned} p_{i,i+1} &= p_i, \\ p_{ii} &= r_i, \\ p_{i,i-1} &= q_i, \end{aligned}$$

where $p_i + r_i + q_i = 1$ for all i, is called a Harris random walk; let

$\mathbf{P}_j$ denote the probability measure governing X when $X_0 = j$. The criterion of ergodicity in the transient case that will be determined here is due to Rösler. We restrict ourselves to the chains where all the r_i values equal 0, and avoid trite thinking by assuming that all $p_i, q_i > 0$.

Without virtual restrictions, we may assume that

$$\mathbf{P}_i(X_n = i + 1 \text{ for some } n \geq 1) = 1 . \tag{16.1}$$

Indeed, with $A_+ = \{X_n \to \infty \text{ as } n \to \infty\}$, $A_- = \{X_n \to -\infty \text{ as } n \to \infty\}$, the transience and skip-freeness of paths render $\mathbf{P}_i(A_+ \cup A_-) = 1$ for all i. Now A_+ is a tail event, and if $0 < \mathbf{P}_i(A_+) < 1$ for some i, then X is not ergodic, as we shall learn in § III.21; hence we may exclude that possibility from our analysis. We notice further that if $\mathbf{P}_i(A_+) = 1$ for some i, then that holds for all i, which is equivalent to (16.1). For symmetry reasons we need not pay attention to the case $\mathbf{P}_i(A_-) = 1$ for all i.

We shall prove that X is ergodic (after attention has been paid to the periodicity) if and only if $\Sigma_{i=0}^{\infty} q_i = \infty$ (the Rösler criterion), and begin with the sufficiency of that. Define T_i, $i \geq 1$, through

$$T_i = \min\{n; X_n = i\} ,$$

and the probability measure Q_{ij} on $(\mathbb{Z}_b, \mathscr{Z}_b)$ through

$$Q_{ij} = \mathbf{P}_{i-1}((X_0, X_1, \ldots, X_j) \in \cdot \,\|\, T_i = j) .$$

For convenience let 0 and $2k$, $k > 0$, be the initial states of the processes to be coupled, and denote the distributions of T_{2k} and $T_{2k+j} - T_{2k+j-1}$, $j \geq 1$, under $\mathbf{P}_0$ by $\lambda_{0,2k}$ and $\lambda_{2k+j-1,2k+j}$, respectively. For the construction of $\hat{X}$ and $\hat{X}'$ we need independent variables Y_0, $(Y_1, Y_1'), \ldots$, such that for $j \geq 1$,

(i) $Y_0 \overset{\mathscr{D}}{=} \lambda_{0,2k}$, $Y_0' \equiv 0$,
(ii) $Y_j \overset{\mathscr{D}}{=} \lambda_{2k+j-1,2k+j}$, and
(iii) (Y_j, Y_j') is distributed according to (14.1), but with all the $(i-1)$'s there replaced by $i-2$, and $\alpha_i = (\mathbf{P}(Y_j = i) \wedge \mathbf{P}(Y_j = i+2))/2$.

We use a standard product space construction to produce these variables; let them be defined on $(\Omega_1, \mathcal{F}_1, P_1)$, where $\Omega_1 = (\mathbb{Z}_+^2)^\infty$, $\mathcal{F}_1 = $ etc. On $(\Omega, \mathcal{F}) = ((\mathbb{Z}_b^2, \mathcal{Z}_b^2)^\infty)$, let P be defined by

$$\mathbf{P} = \int dP_1((y_j, y'_j)_0^\infty)\, K((y_j, y'_j)_0^\infty, \cdot)\,,$$

where K is the transition kernel in $\Omega_1 \times \Omega$ given by

$$K((y_j, y'_j)_0^\infty, \cdot) = \prod_{j=0}^{\infty} Q_{j,y_j} \times Q_{2k+j,y'_j}\,.$$

Now let $(B_j, B'_j)_0^\infty$ have distribution $\mathbf{P}$, and use $(B_j)_0^\infty$, $(B'_j)_0^\infty$ to obtain $(\hat{X}_j)_0^\infty$, $(\hat{X}'_j)_0^\infty$ by (15.1). Our coupling time is

$$\begin{aligned} T &= \min\{n; \hat{X}_n = \hat{X}'_n\} \\ &\le \min\Big\{ \sum_0^k Y_j; \sum_0^k Y_j = \sum_0^k Y'_j \Big\} = \min\Big\{ \sum_0^k Y_j; \sum_1^k Y_j = \sum_1^k Y'_j - Y_0 \Big\}. \end{aligned}$$

Due to arguments in § 14 we know that coupling is successful if

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{P}(Y_i = 2j+1) \wedge \mathbf{P}(Y_i = 2j+3) = \infty.$$

But it is immediate that $\mathbf{P}(Y_i = 2j+1)$ is decreasing in j for each i, so that the double sum equals

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathrm{P}(Y_i = 2j+3) = \sum_0^{\infty} (1 - \mathbf{P}(Y_i = 1)) = \sum_0^{\infty} (1 - p_i) = \sum_0^{\infty} q_i\,,$$

and we have proved the sufficiency.

On the other hand, if $\Sigma_i\,(1 - p_i) < \infty$, then $\Pi_k^\infty\, p_i$ tends to 1 as $k \to \infty$. Let $\hat{X}$, $\hat{X}'$ now be independent and with $\hat{X}_0 = k$, $\hat{X}'_0 = k+2$. We obtain

$$\begin{aligned} &\|\mathbf{P}(\hat{X}_n \in \cdot) - \mathbf{P}(\hat{X}'_n \in \cdot)\| \\ &\quad \ge |\mathbf{P}(\hat{X}_n = k+2+n) - \mathbf{P}(\hat{X}'_n = k+2+n)| \\ &\quad = \mathbf{P}(\hat{X}'_n = k+2+n) = \prod_{k+2}^{k+1+n} p_i \ge \prod_{k+2}^{\infty} p_i\,. \end{aligned}$$

Hence for k large enough, we have

$$\limsup_n \|\mathbf{P}(\hat{X}_n \in \cdot) - \mathbf{P}(\hat{X}'_n \in \cdot)\| \geq \tfrac{1}{2},$$

and the proof of Rösler's criterion is complete.

17. A multidimensional random walk. We shall give no systematic account of the possibilities of producing good couplings of random walks in several dimensions. Rather, we content ourselves with a brief examination of the simple random walk; that will suffice to demonstrate a useful method and to understand that the dimension number is, in a sense, irrelevant.

Consider the simple random walk $S = (S_n)_0^\infty$ in $\mathbb{Z}^d$, $d \geq 2$; that is, we have

$$S_n = Y_0 + \sum_1^n Y_i,$$

where $Y_0, Y_1, \ldots$ are independent and the variables $Y_1, Y_2, \ldots$ all take values $(\pm 1, 0, \ldots, 0)$, $(0, \pm 1, \ldots, 0), \ldots, (0, 0, \ldots, \pm 1)$ with probability $1/(2d)$ for each. To achieve successful coupling with another such random walk $S' = (S'_n)_0^\infty$, with $S'_n = Y'_0 + \Sigma_1^n Y'_i$, the idea is to let S and S' coincide in an increasing number of coordinates. To that end, let $\xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots, \eta'_1, \eta'_2, \ldots$ be independent variables where the ξ_i variables are uniformly distributed on $\{1, 2, \ldots, d\}$ and η_i, η'_i, $i \geq 1$, all take the values 1 and -1 with probability $\frac{1}{2}$ each. Further, let ξ_i^* be the vector with 1 in coordinate ξ_i and zeros elsewhere. Define Y_i, Y'_i, $i \geq 1$, successively by

$$(17.1) \quad \begin{aligned} Y_i &= \xi_i^* \cdot \eta_i, \\ Y'_i &= \xi_i^* \cdot [\eta_i \cdot I(S_{\xi_i, i-1} = S'_{\xi_i, i-1}) + \eta'_i \cdot I(S_{\xi_i, i-1} \neq S_{\xi_i, i-1})], \end{aligned}$$

where the notation $S_n = (S_{1n}, S_{2n}, \ldots, S_{dn})$ is used.

No doubt, random walks S and S' taking steps according to (17.1), eventually coincide under the obviously necessary condition that $Y_{i0} - Y'_{i0}$ is even for $1 \leq i \leq d$.

18. Notes. Ornstein introduced his coupling in [128]. For a fine account of recurrence, see Chung [39]. The coupling of § 14 was discovered independently by Mineka [119] and Rösler [137, 138]. The exercise at the end of that section is from Rösler [140]. For accounts on how to use the idea of extracting a symmetric simple random walk in order to prove, for example, local limit theorems, see McDonald [117] and McDonald and Rösler [118]. The criterion of § 16 was presented in Rösler [137]. See Feller [61, (III.7.6)] for the estimate $\mathbf{P}(\tilde{T} = n) \leq C \cdot n^{-3/2}$, and Breiman [30] for the use of Karamata's theorem.

4. CARD SHUFFLING

19. Basics. Card games offer the probabilist an abundance of interesting and difficult problems to solve. They are also important for the person who takes applications seriously (i.e., is involved in real games, with money or glory at stake).

Consider a deck with N cards, labeled by the numbers 1, 2, ..., N. An arrangement of the deck is an element in

$$G = \{\mathbf{i} = (i_1, \ldots, i_N); \mathbf{i} \text{ is a permutation of } 1, \ldots, N\}.$$

It will be convenient to refer to i_1 as the top card and i_N as the bottom card.

The deck is now mixed by repeated single shuffles carried out independently of each other and according to the same principle each time. This renders a time-homogeneous Markov chain $X = (X_n)_0^\infty$ with G as the state space, where X_n is the card arrangement after n single shuffles.

Usually, X is aperiodic and irreducible and has a uniform distribution on G (to be called π) as the stationary distribution. Since G is finite, we know that X is asymptotically stationary with a geometric rate of convergence, so the deck is well shuffled for n sufficiently large. However, that general observation pales relative to what is true of X under certain shuffling principles. Here we demonstrate that fact in terms of one example, "top to random" shuffling.

For simplicity, assume that the deck at the outset is in the order 1, 2, ..., N, from top to bottom. We shall consider large N and say that X has a threshold at $t_N > 0$ if

$$\text{(19.1)} \quad \text{(i)} \ \|\mathbf{P}(X_{(1-\epsilon)\cdot t_N} \in \cdot\,) - \pi\| \to 2, \text{ and}$$
$$\text{(ii)} \ \|\mathbf{P}(X_{(1+\epsilon)\cdot t_N} \in \cdot\,) - \pi\| \to 0 \text{ as } N \to \infty$$

for all $\epsilon > 0$. Here $(t_N)_{N=1}^\infty$ is a sequence increasing in N.

The topic of thresholds is new to us. A rate result is a natural answer to the question: How well mixed is the deck after n single shuffles? If a threshold value t_N can be found, we have a precise answer to how many single shuffles suffice for a large N. We shall prove below that there is a threshold at $N \cdot \log N$ if "top to random" shuffling is used.

To study the asymptotics of $X = (X_n)_0^\infty$, it is natural to introduce a second deck of cards that is uniformly mixed to begin with, and then make a suitable coupling of X and the chain, $X' = (X'_n)_0^\infty$ say, of that second deck. Such couplings are the most efficient way to understand X for some shuffling principles, but not all. Actually, in many cases a parallel deck is not only unnecessary but should even be avoided, at least to begin with. There is another notion whose possibilities should be investigated first: that of strong uniform times.

Consider a stopping time T w.r.t. X with the properties

(19.2) (i) $X_T \overset{\mathcal{D}}{=} \pi$, and

(ii) X_T and T are independent.

These are the defining properties for a strong uniform time. Applying (ii) first and then (i), we obtain for any $A \subset G$ and $n \geq 0$ that

$$
\begin{aligned}
&\mathbf{P}(X_n \in A, T \leq n) \\
&\quad = \sum_{j=0}^{n} \mathbf{P}(X_n \in A, T = j) \\
&\quad = \sum_{j=0}^{n} \sum_{\mathbf{i} \in G} \mathbf{P}(X_n \in A \| X_j = \mathbf{i}, T = j) \cdot \mathbf{P}(X_j = \mathbf{i}, T = j) \\
&\quad = \sum_{j=0}^{n} \sum_{\mathbf{i} \in G} P^{n-j}(\mathbf{i}, A)\mathbf{P}(X_T = \mathbf{i}) \cdot \mathbf{P}(T = j) \\
&\quad = \sum_{j=0}^{n} \pi P^{n-j}(A) \cdot \mathbf{P}(T = j) \\
&\quad = \sum_{j=0}^{n} \pi(A) \cdot \mathbf{P}(T = j) = \pi(A) \cdot \mathbf{P}(T \leq n) .
\end{aligned}
$$

Hence

$$\begin{aligned}
&|\mathbf{P}(X_n \in A) - \pi(A)| \\
&\quad = |\mathbf{P}(X_n \in A, T \le n) + \mathbf{P}(X_n \in A, T > n) - \pi(A)| \\
&\quad = |\pi(A) \cdot \mathbf{P}(T \le n) + \mathbf{P}(X_n \in A, T > n) \\
&\qquad - \pi(A) \cdot \mathbf{P}(T \le n) - \pi(A) \cdot \mathbf{P}(T > n)| \\
&\quad \le \mathbf{P}(T > n)
\end{aligned}$$

and we obtain

$$\|\mathbf{P}(X_n \in \cdot\,) - \pi\| \le 2 \cdot \mathbf{P}(T > n)\,, \tag{19.3}$$

a familiar inequality. Actually, it may be seen as a coupling inequality: You may construct a strong coupling of X and a stationary X' such that T becomes a coupling time. In fact, let X' be identical to X after T.

Before the "top to random" shuffling, let us make two remarks. The first is that T need not be a stopping time w.r.t. X, the inequality (19.3) holds also if T is a so-called randomized stopping time w.r.t. X. The definition of that useful concept will be given in § III.10. An important special case is when T is a function not only of X but also of some random element independent of X.

The second remark is that a strong uniform time is a special case of what we shall call a conforming time. That term will also be defined in § III.10.

20. "Top to random" shuffling. A single shuffle now consists of inserting the top card at a random place, with probability $1/N$ for each of the N possibilities. Notice that a replacement at the top is one of them. Now let

$T =$ the first time the original bottom card comes to the top $+ 1$.

The "1" represents an insertion of that card at a random place. We omit the details of proof of the rather obvious fact that T is a strong uniform time. To prove that $N \cdot \log N$ is a threshold value, it is clarifying to note that

$$T \overset{\mathcal{D}}{=} V\,, \tag{20.1}$$

where V is the number of required draws from an urn with N balls, with replacements and equal probabilities, until each ball has been drawn (if the Coupon Collector is a familiar character, you recognize his random variable). To prove (20.1), rewrite T as

$$T = T_1 + (T_2 - T_1) + \cdots + (T_{N-1} - T_{N-2}) + (T - T_{N-1}), \tag{20.2}$$

where T_i is the first time there are i cards under the original bottom card (hence $T - T_{N-1} \equiv 1$). Now

$$V = (V - V_{N-1}) + (V_{N-1} - V_{N-2}) + \cdots + V_1, \tag{20.3}$$

where V_i is the number of draws required for i distinct balls to appear. The representations (20.2)–(20.3) express T and V in terms of sums of independent, geometrically distributed variables. With $T = T_N$, $V = V_N$, and $V_0 = T_0 \equiv 0$, we find that

$$T_{i+1} - T_i \stackrel{\mathcal{D}}{=} V_{N-i} - V_{N-i-1} \qquad \text{for all} \quad 0 \le i \le N-1 .$$

Hence $T \stackrel{\mathcal{D}}{=} V$.

Now put labels $1, \ldots, N$ on the balls and let A_i be the event that ball number i is not in the first m draws. We obtain

$$\mathbf{P}(V > m) = \mathbf{P}\Big(\bigcup_1^N A_i\Big) \le \sum_1^N \mathbf{P}(A_i) = N \cdot (1 - 1/N)^m .$$

With $m =$ the integer part of $(1+\epsilon) \cdot N \cdot \log N$, that expression tends to 0 as $N \to \infty$. Using (20.1) and (19.3) we have proved that

$$\|\mathbf{P}(X_{(1+\epsilon)\cdot N \cdot \log N} \in \cdot\,) - \pi\| \to 0 \qquad \text{as} \quad N \to \infty \quad \text{for all} \quad \epsilon > 0 .$$

To establish (19.1)(i), we shall find sets A_N such that given any $\epsilon > 0$,

$$\mathbf{P}(X_{(1-\epsilon)\cdot N \cdot \log N} \in A_N) - \pi(A_N) \to 1 \qquad \text{as} \quad N \to \infty . \tag{20.4}$$

As you have noticed, we have omitted indexing with N at the obvious places and using the "integer part of" function on $(1+\epsilon) \cdot N \cdot \log N$, for example. Now $N \cdot \log N$ is a good candidate for a threshold since $\mathbf{E}[T] \approx N \cdot \log N$. Indeed,

$$\mathbf{E}[T] = \sum_{0}^{N-1} \mathbf{E}[T_{i+1} - T_i] = \sum_{0}^{N-1} N/(i+1) \approx N \cdot \log N .$$

Now let $\epsilon_1 > 0$, and take a $j \geq 1$ so large that $1/j! < \epsilon_1/2$ and let

$$A_N = \{\mathbf{i} \in G;\, i'_{N-j+1} < \cdots < i'_k < \cdots < i'_N \text{ for } N-j+1 < k < N\} ,$$

where i'_m is the position of the card labeled m. Notice that $\pi\,(A_N) = 1/j!$ and that $X_n \in A_N$ is the event that the order of the original j bottom cards is retained at time n. A crucial observation now is that the first time the card labeled $(N - j + 1)$ comes to the top is distributed like V_{N-j+1}. This implies that

$$\mathbf{P}(X_{(1-\epsilon)\cdot N\cdot\log N} \in A_N) \geq \mathbf{P}(V_{N-j+1} > (1-\epsilon)\cdot N \cdot \log N) .$$

So if $\mathbf{P}(V_{N-j+1} \leq (1-\epsilon)\cdot N\cdot \log N) < \epsilon_1/2$ for N large enough, (20.4) is established since ϵ_1 is arbitrary. But

$$\begin{aligned}(20.5) \qquad \mathbf{P}(V_{N-j+1} &\leq (1-\epsilon)\cdot N\cdot\log N) \\ &\leq \mathbf{P}(|V_{N-j+1} - \mathbf{E}[V_{N-j+1}]| \\ &\geq \mathbf{E}[V_{N-j+1}] - (1-\epsilon)\cdot N\cdot\log N) .\end{aligned}$$

Now the latter probability tends to 0 as $N \to \infty$ due to Chebyshev's inequality. Indeed,

$$\begin{aligned}\mathrm{Var}[V_{N-j+1}] &= \sum_{i=j-1}^{N-1} \mathrm{Var}[V_{N-i} - V_{N-i-1}] \\ &= \sum_{i=j-1}^{N-1} (N/(i+1))^2 \cdot (1-(i+1)/N) = \mathrm{O}(N^2)\end{aligned}$$

and

$$\mathbf{E}[V_{N-j+1}] - (1-\epsilon)\cdot N\cdot\log N \geq c\cdot N\cdot\log N$$

for some constant $c > 0$ for N large enough; hence the Chebyshev inequality shows that the probability of (20.5) is bounded by $\mathrm{O}(N^2/N^2\cdot\log^2 N)$.

21. Notes. We have considered only one out of many interesting shuffling principles. For accounts, see Aldous [2], Diaconis [50], and Aldous and Diaconis [4, 5]. A random walk on multidimensional cubes is a closely related topic also treated in these references.

5. POISSON APPROXIMATION

22. Basics. Recall the last example of § Int.1 and its notation. The inequality

$$\|\mathbf{P}(X \in \cdot\,) - p_\lambda\| \le 2 \cdot \sum_1^n p_i^2 \tag{22.1}$$

proved in that section may be sharpened qualitatively to

$$\|\mathbf{P}(X \in \cdot\,) - p_\lambda\| \le C_1 \cdot \Big(\sum_1^n p_i^2\Big)/\lambda\,, \tag{22.2}$$

where C_1 is a universal constant not depending on $p_1, \ldots, p_n$ or n. Notice that the right-hand side of (22.2) is not larger than $C_1 \cdot \max_{1\le i\le n}(p_i)$. The value of the constant C_1 has been improved gradually since (22.2) was proved; the best known value equals 2. We refer to (22.2) as Le Cam's theorem. The main purpose of this section is to show how a coupling comes to use a certain generalization of the theorem to dependent variables Y_1, $Y_2, \ldots, Y_n$ and applications of that generalization. No entirely probabilistic proof of Le Cam's theorem has been found; some analytical thinking seems inevitable. The original proof used Fourier analysis, while today the proof that uses the Stein–Chen method is better known. The latter may be said to be simpler and can also be combined with coupling ideas to achieve good Poisson approximations for dependent summands. We dedicate §§ 24 and 25 to the Stein–Chen method and its possibilities. But before that, a coupling even more direct than that of § Int.1 merits our attention.

23. Another simple coupling. For the sequence $p_1, \ldots, p_n$, $0 \le p_i < 1$, let $Y_1', \ldots, Y_n'$ be independent with $Y_i' \stackrel{\mathcal{D}}{=} \mathrm{Poi}(\lambda_i')$, where $\lambda_i' = -\log(1 - p_i)$. Further, let

(23.1) $$Y_i = \min(Y_i', 1)\,.$$

Then $Y_i \overset{\mathcal{D}}{=} \mathrm{Ber}(p_i)$ since $\mathbf{P}(Y_i = 0) = \exp(-\lambda_i') = 1 - p_i$. Notice that $X' = \Sigma_1^n Y_i'$ has a $\mathrm{Poi}(\lambda')$ distribution, where $\lambda' = \Sigma_1^n \lambda_i'$. If $X = \Sigma_1^n X_i$ equals X' with high probability, it is again meaningful to use the coupling inequality for a Poisson approximation. Notice that the approximating distribution has parameter λ', not $\lambda = \Sigma_1^n p_i$.

We have

$$\mathbf{P}(X \neq X') \leq \sum_1^n \mathbf{P}(Y_i' \geq 2) \leq \tfrac{1}{2} \cdot \sum_1^n \lambda_i'^2$$

because for any $\alpha > 0$,

$$\sum_2^\infty e^{-\alpha} \cdot \alpha^i/i! = \alpha^2 \cdot \sum_0^\infty e^{-\alpha} \cdot \alpha^i/(i+2)!$$

$$\leq \alpha^2 \cdot \tfrac{1}{2} \cdot \sum_0^\infty e^{-\alpha} \cdot \alpha^i/i! = \tfrac{1}{2}\,\alpha^2\,.$$

Hence

(23.2) $$\|\mathbf{P}(X \in \cdot\,) - p_{\lambda'}\| \leq 2 \cdot \tfrac{1}{2} \cdot \sum_1^n \lambda_i'^2 = \sum_1^n \lambda_i'^2\,.$$

That bound is $\leq 2 \cdot \Sigma_1^n p_i^2$ [cf. (22.1)] if $p_i \leq \frac{1}{2}$ for all i, because then $\lambda_i'^2 \leq 2p_i^2$.

The paradox that we have different approximating Poisson distributions in (22.1) and (23.2) is resolved by the observation that p_λ and $p_{\lambda'}$ are close for values of $p_1, \ldots, p_n$ small enough to make approximation interesting. Indeed, let $0 \leq \alpha \leq \alpha'$ and take two independent variables Z and Z'', where $Z \overset{\mathcal{D}}{=} \mathrm{Poi}(\alpha)$ and $Z'' \overset{\mathcal{D}}{=} \mathrm{Poi}(\alpha' - \alpha)$. Then $Z' = Z + Z'' \overset{\mathcal{D}}{=} \mathrm{Poi}(\alpha')$ and

$$\|p_\alpha - p_{\alpha'}\| \leq 2 \cdot \mathbf{P}(Z \neq Z') = 2 \cdot \mathbf{P}(Z'' \geq 1) = 2 \cdot (1 - e^{-(\alpha' - \alpha)})$$

$$\leq 2 \cdot (\alpha' - \alpha)\,.$$

So

(23.3)

$$\|p_\lambda - p_{\lambda'}\| \leq 2 \cdot (\lambda' - \lambda) = 2 \cdot \sum_1^n [-\log(1 - p_i) - p_i] \leq 2 \cdot \sum_1^n p_i^2$$

if all $p_i \le \frac{1}{2}$, since $-\log(1-\alpha) - \alpha \le \alpha^2$ for $0 \le \alpha \le \frac{1}{2}$. Hence if we insist on approximating $\mathbb{P}(X \in \cdot\,)$ by p_λ with use of the simple idea (23.1), the inequalities (23.2) and (23.3) render

(23.4)

$$\|\mathbb{P}(X \in \cdot\,) - p_\lambda\| \le \|\mathbb{P}(X \in \cdot\,) - p_{\lambda'}\| + \|p_{\lambda'} - p_\lambda\| \le 4 \cdot \sum_1^n p_i^2$$

if the p_i values are small enough.

Notice that $X \le X'$. This makes it possible to bound the tail probabilities of X. Indeed, for any $k \ge 0$,

$$\mathbb{P}(X \ge k) \le \mathbb{P}(X' \ge k) = p_{\lambda'}[k, \infty)\,. \tag{23.5}$$

We conclude with a result that usually is proved with generating functions. Let

$$Y_{ni},\ 1 \le i \le k_n, \qquad n = 1, 2, \ldots$$

be an array of independent variables with $Y_{ni} \overset{\mathcal{D}}{=} \mathrm{Ber}(p_{ni})$. Suppose that

$$\max_{1 \le i \le k_n} p_{ni} \to 0 \qquad \text{as} \quad n \to \infty\,. \tag{23.6}$$

Then there exists a $\lambda \ge 0$ such that

$$X_n = \sum_{i=1}^{k_n} Y_{ni} \overset{\mathcal{D}}{\to} \mathrm{Poi}(\lambda)$$

if and only if

$$\sum_{i=1}^{k_n} p_{ni} \to \lambda \qquad \text{as} \quad n \to \infty\,.$$

For the sufficiency, use (23.6) to get

$$\sum_{i=1}^{k_n} p_{ni}^2 \le \big(\max_{1 \le i \le k_n} p_{ni}\big) \cdot \sum_{j=1}^{k_n} p_{nj} \to 0 \qquad \text{as} \quad n \to \infty\,. \tag{23.7}$$

Due to (23.6), we may assume that all p_{ni} are small enough for (23.4) to be valid. That inequality implies, with $\mu_n = \Sigma_{i=1}^{k_n} p_{ni}$, that

$$\begin{aligned}\|\mathbf{P}(X_n \in \cdot) - p_\alpha\| &\le \|\mathbf{P}(X_n \in \cdot) - p_{\mu_n}\| + \|p_{\mu_n} - p_\alpha\| \\ &\le 4 \cdot \sum_{i=1}^{k_n} p_{ni}^2 + \|p_{\mu_n} - p_\alpha\|\end{aligned}$$

for n large enough. But the right-hand side tends to 0 as $n \to \infty$, and convergence in total variation norm implies convergence in distribution.

For the necessity, the thing is to prove that $\limsup_{n\to\infty} \mu_n < \infty$. We have for $\mu_n \ge 1$, due to Chebyshev's inequality,

$$\begin{aligned}\mathbf{P}(X_n = 0) &\le \mathbf{P}(|X_n - \mu_n| \ge \tfrac{1}{2} \cdot \mu_n) \\ &\le \operatorname{Var}[X_n]/(\tfrac{1}{2} \cdot \mu_n)^2 = \sum_{i=1}^{k_n} p_{ni}(1 - p_{ni})/(\tfrac{1}{2}\mu_n)^2 \le 1/\mu_n .\end{aligned}$$

Now $\mathbf{P}(X_n = 0) \to p_\lambda(0) > 0$ as $n \to \infty$, implying $\liminf_{n\to\infty} (1/\mu_n) > 0$, and $\limsup_{n\to\infty} \mu_n < \infty$ follows.

If $\mu_n \to \lambda$ as $n \to \infty$ were not true, we might pick a subsequence of $(\mu_n)_1^\infty$ convergent to a $\lambda' \ne \lambda$. But this would imply that a subsequence of $(X_n)_1^\infty$ has $\mathrm{Poi}(\lambda')$ as a limiting distribution, a contradiction.

24. The Stein–Chen method. Consider variables Y_i, $1 \le i \le n$, such that $Y_i \overset{\mathcal{D}}{=} \mathrm{Ber}(p_i)$ but without an independence assumption. To indicate that we are now facing a wider problem, but also to adjust ourselves to notation that is standard in applications of the Stein–Chen method, we let the sum $\Sigma_1^n Y_i$ be denoted by W. The notation λ for $\Sigma_1^n p_i$ is retained.

For $1 \le k \le n$, let U_k and V_k be random variables on the same probability space, satisfying

$$U_k \overset{\mathcal{D}}{=} W, \qquad 1 + V_k \overset{\mathcal{D}}{=} \mathbf{P}(W \in \cdot \,\|\, Y_k = 1), \tag{24.1}$$

with the convention that $V_k \equiv 0$ if $\mathbf{P}(Y_k = 1) = 0$. If $U_k = V_k$ with high probability, indicating that if the dependence between the Y_i

variables is weak and the p_k values are small, a good Poisson approximation of W may be obtained.

For the understanding of the proof below, note that if $Z \stackrel{\mathscr{D}}{=} \text{Poi}(\lambda)$, then

$$\mathbf{E}[\lambda \cdot h(Z+1) - Z \cdot h(Z)] = 0 \tag{24.2}$$

for all bounded functions on $\mathbb{Z}_+$. Indeed,

$$\begin{aligned}\mathbf{E}[\lambda \cdot h(Z+1)] &= \lambda \cdot \sum_0^\infty h(i+1) \cdot e^{-\lambda} \cdot \lambda^i/i! \\ &= \sum_0^\infty (i+1)h(i+1) \cdot e^{-\lambda} \cdot \lambda^{i+1}/(i+1)! \\ &= \sum_1^\infty i \cdot h(i) \cdot e^{-\lambda} \cdot \lambda^i/i! = \mathbf{E}[Z \cdot h(Z)] .\end{aligned}$$

(24.3) Theorem. *It holds that*

$$\|\mathbf{P}(W \in \cdot\,) - p_\lambda\| \le 2 \cdot (1 \wedge \lambda^{-1}) \cdot \sum_1^n p_i \cdot \mathbf{E}[|U_i - V_i|] .$$

Proof. For any bounded function h on $\mathbb{Z}_+$ we get

$$\begin{aligned}&\mathbf{E}[\lambda \cdot h(W+1) - W \cdot h(W)] \\ &\quad = \sum_1^n \{p_i \cdot \mathbf{E}[h(W+1)] - \mathbf{E}[Y_i \cdot h(W)]\} \\ &\quad = \sum_1^n p_i \cdot \{\mathbf{E}[h(W+1)] - \mathbf{E}[h(W) \| Y_i = 1]\} \\ &\quad = \sum_1^n p_i \cdot \mathbf{E}[h(U_i+1) - h(V_i+1)] .\end{aligned}$$

Now for any $j, k \ge 0$ we have $|h(i) - h(j)| \le |i-j| \cdot \|\Delta h|$, where $\|\cdot\|$ denotes supremum norm. Using this, we obtain

$$|\mathbf{E}[\lambda \cdot h(W+1) - W \cdot h(W)]| \le \|\Delta h\| \cdot \sum_1^n p_i \cdot \mathbf{E}[|U_i - V_i|] . \tag{24.4}$$

Fix a set $A \subset \mathbb{Z}_+$ and let g be the solution to the equation

$$(24.5)\qquad g(0)=0\,,$$
$$\lambda\cdot g(i+1)-i\cdot g(i)=I_A(i)-p_\lambda(A),\qquad i\geq 0\,.$$

We omit the rather simple, but careful steps needed to prove that g is bounded and

$$(24.6)\qquad \|\Delta g\|\leq 1\wedge\lambda^{-1}\,.$$

Combine (24.4)–(24.6) to get

$$\begin{aligned}|\mathbf{P}(W\in A)-p_\lambda(A)| &= |\mathbf{E}[I_A(W)-p_\lambda(A)]|\\ &= |E[\lambda\cdot g(W+1)-W\cdot g(W)]|\\ &\leq (1\wedge\lambda^{-1})\cdot\sum_1^n p_i\cdot\mathbf{E}[|U_i-V_i|]\,.\end{aligned}$$

The assertion of the theorem now follows since

$$\|\mathbf{P}(W\in\cdot\,)-p_\lambda\|=2\cdot\sup_{A\subset\mathbf{Z}_+}|\mathbf{P}(W\in A)-p_\lambda(A)|\,.\qquad\square$$

Hence the Stein–Chen method for Poisson approximation may be summarized as follows:

(24.7) (i) Notice the equality (24.2), and surmise that if $\mathbf{E}[\lambda\cdot h(Z+1)-Z\cdot h(Z)]$ is small for functions h with $\|h\|\leq 1$, then Z has approximately a Poisson distribution.

(ii) Use (24.5): write $\mathbf{P}(W\in S)-p_\lambda(A)$ as $\mathbf{E}[\lambda\cdot g(W+1)-W\cdot g(W)]$ for a certain function g.

(iii) Bound g and Δg, to apply (24.4).

(iv) Produce a clever coupling satisfying (24.1) to make $\mathbf{E}[|U_k-V_k|]$ small for $1\leq k\leq n$.

A similar program works for Gaussian approximation of sums of weakly dependent random variables. For step (iii), see the primary reference given in § 26.

To deduce Le Cam's theorem from (24.3), no skill concerning step (iv) is needed. Indeed, let

$$U_k=W\quad\text{and}\quad V_k=\sum_{i\neq k}Y_i\,.$$

(24.8) (Le Cam's Theorem). *If $Y_1, \ldots, Y_n$ are independent, then*

$$\|\mathbf{P}(W \in \cdot\,) - p_\lambda\| \le 2 \cdot \sum_1^n p_i^2/\lambda \le 2 \cdot \max_{1 \le i \le n} (p_i)\,.$$

Proof. This follows from Theorem (24.3), with the observation that $\mathbf{E}[|U_i - V_i|] = \mathbf{E}[Y_i] = p_i$ and $1 \wedge \lambda^{-1} \le \lambda^{-1}$. □

The original proof had a universal constant larger than 2 on the right-hand side.

Theorem (24.3) becomes particularly valuable in cases where we can find a coupling satisfying (24.1) such that

$$U_k \ge V_k \qquad \text{for all} \quad k\,. \tag{24.9}$$

For the bound in the assertion of Theorem (24.3), we then have

$$\sum_1^n p_i \cdot \mathbf{E}[|U_i - V_i|] = \lambda - \mathrm{Var}[W] \tag{24.10}$$

and $\mathrm{Var}[W]$ can often be calculated explicitly. The equality (24.10) is easily proved:

$$\begin{aligned}
\sum_1^n p_i \cdot \mathbf{E}[|U_i - V_i|] &= \sum_1^n p_i \cdot \mathbf{E}[U_i - V_i] \\
&= \sum_1^n p_i \cdot \mathbf{E}[U_i] - \sum_1^n p_i \cdot \mathbf{E}[V_i] \\
&= \lambda^2 - \sum_1^n p_i \cdot \mathbf{E}[V_i] \\
&= \lambda^2 - \sum_1^n p_i \cdot \mathbf{E}[1 + V_i] + \sum_1^n p_i \\
&= \lambda^2 - \sum_1^n p_i \cdot \mathbf{E}[W \,\|\, Y_i = 1] + \lambda \\
&= \lambda^2 - \mathbf{E}\left[\sum_1^n Y_i \cdot W\right] + \lambda = \lambda - \mathrm{Var}[W]\,.
\end{aligned}$$

To have a practical criterion for (24.9), consider the vectors $(Y_{k1}, \ldots, Y_{kn})$, $(Y'_{k1}, \ldots, Y'_{kn})$, $1 \le k \le n$, defined on the same probability space such that for each k

$$(24.11)\quad \text{(i)}\qquad (Y_{k1}, \ldots, Y_{kn}) \stackrel{\mathscr{D}}{=} (Y_1, \ldots, Y_n),$$

$$\text{(ii)}\qquad (Y'_{k1}, \ldots, Y'_{kn}) \stackrel{\mathscr{D}}{=} \mathbf{P}((Y_1, \ldots, Y_n) \in \cdot \,\|\, Y_k = 1)$$

with the convention that $(Y'_{k1}, \ldots, Y'_{kn}) = (0, \ldots, 1, \ldots, 0)$ with the "1" at place k, if $\mathbf{P}(Y_k = 1) = 0$. With

$$U_k = \sum_{i=1}^{n} Y_{ki}, \qquad 1 + V_k = \sum_{i=1}^{n} Y'_{ki},$$

the variables U_k and V_k, $1 \le k \le n$, satisfy (24.1).

The variables $Y_1, \ldots, Y_n$ are said to be negatively related if there exist vectors $(Y_{k1}, \ldots, Y_{kn})$, $(Y'_{k1}, \ldots, Y'_{kn})$, $1 \le k \le n$, satisfying (24.11) and such that

$$(24.12)\qquad Y'_{ki} \le Y_{ki} \qquad \text{for all} \quad i \ne k \quad \text{and} \quad 1 \le k \le n\,.$$

The inequality of (24.9) holds then:

$$\begin{aligned} U_k - V_k &= \sum_{i=1}^{n} (Y_{ki} - Y'_{ki}) + 1 \\ &\ge \sum_{\substack{i=1 \\ i \ne k}}^{n} (Y_{ki} - Y'_{ki}) + (1 - Y'_{kk}) + Y_{kk} \ge 0 \end{aligned}$$

since $Y_{kk} \ge 0$.

Using Theorem (24.3) and (24.10), we are now ready to state the final theorem on Poisson approximation.

(24.13) Theorem. *If $Y_1, \ldots, Y_n$ are negatively related, then*

$$\|\mathbf{P}(W \in \cdot) - p_\lambda\| \le 2 \cdot (1 \wedge \lambda^{-1}) \cdot (\lambda - \operatorname{Var}[W])\,.$$

25. An example. Out of many possibilities, we choose to demonstrate the strength of Theorem (24.13) by a Poisson approximation of the hypergeometrical distribution. To visualize that distribution, sider N boxes with places for one ball only in each, and ribute m balls among the boxes uniformly. Here $m < N$ and by iformly" we mean that each of the $\binom{N}{m}$ configurations has the me probability. Let $Y_i = 1$ if there is a ball in box number i, $=0$ if not. Then the sum $W = \Sigma_1^n Y_i$, where $n \leq N$, has a ypergeometrical distribution, and it is a reasonable conjecture at W is approximately Poisson distributed if n and m are small lative to N.

Denote m/N by p. We have $\mathbf{E}[Y_i] = p$ for $1 \leq i \leq n$, and it is ell known that

$$\lambda = \mathbf{E}[W] = n \cdot p \tag{25.1}$$

and

$$\mathrm{Var}[W] = n \cdot p \cdot (1 - p) \cdot (N - n)/(N - 1) .$$

If $Y_1, \ldots, Y_n$ are negatively related, Theorem (24.13) implies that

$$\begin{aligned} \|\mathbf{P}(W \in \cdot\,) - p_\lambda\| &\leq 2 \cdot (1 \wedge \lambda) \cdot (1 - \mathrm{Var}[W]/\lambda) \\ &= 2 \cdot (1 \wedge \lambda) \cdot (1 - (1 - p)(N - n)/(N - 1)) \\ &\leq 2 \cdot (1 \wedge (np)) \cdot (n/N + p) \cdot N/(N - 1) \\ &\leq 2 \cdot (n + m)/(N - 1) , \end{aligned} \tag{25.2}$$

which is a good bound if n and m are small relative to N.

But are $Y_1, \ldots, Y_n$ negatively related? Yes. To prove that, fix a k, $1 \leq k \leq n$, and let $(Y'_{k1}, \ldots, Y'_{kn})$ be produced as follows: We first place a ball in box number k, then distribute the remaining $(m - 1)$ balls among the other $(N - 1)$ boxes. Then $Y'_{ki} = 1$ if there is a ball in box number i, and $Y'_{ki} = 0$ otherwise. Certainly, $(Y'_{k1}, \ldots, Y'_{kn})$ satisfies (24.11)(ii).

Now toss a coin with probability p for heads to turn up. If heads, then let

$$(Y_{k1}, \ldots, Y_{kn}) = (Y'_{k1}, \ldots, Y'_{kn}) .$$

If tails, pick up the ball assigned to box number k by

$(Y'_{k1}, \ldots, Y'_{kn})$ and let it drop into one of the $(N-1)$ other boxes that is empty, with equal probabilities. Then

$$(Y_{k1}, \ldots, Y_{kn}) \stackrel{\mathcal{D}}{=} (Y_1, \ldots, Y_n)$$

and (24.11) and (24.12) are satisfied.

26. Notes. Chen [35] (an elaborated version of his thesis from 1971) applied ideas of Stein [146] to establish Poisson approximations. A thorough account of the Stein–Chen method for Poisson approximation, with many interesting couplings and an abundance of examples, is that of Barbour, Holst, and Janson [21]. The coupling in § Int.1 is close to that in Hodges and Le Cam [73]. Le Cam presented his theorem (24.8) in [102]. The coupling based on (23.1) and several of the ideas in § 23 are from Serfling [141].

CHAPTER III

Continuous Theory

1. RENEWAL THEORY

1. Basics. The reader is assumed to be familiar with the role played by renewal theory; it is an important one because of the relevance to applications, but also because the renewal process is such an exciting object of study in its own right. The need for a motivating introduction is further reduced by several fine accounts in the literature.

We retain as much as possible of the notation in § II.1. The i.i.d. variables $Y_1, Y_2, \ldots$ are supposed to be strictly positive; hence $F(0) = 0$ for their common distribution function F. In the literature the Y_i variables are called (besides recurrence times) life lengths, interarrival times, or waiting times, depending on the circumstances; in this chapter we use "life lengths."

We assume that F is of nonlattice type; that is, there is no $\delta > 0$ such that $\operatorname{supp}(F) \subset L_\delta$, where

$$L_\delta = \{k \cdot \delta; k \in \mathbb{Z}_+\}.$$

The distribution of the delay Y_0 will be denoted by G. We shall not hesitate to let the same symbol, F say, denote both a probability measure and its corresponding distribution function: $F(x) = F([0, x])$ for $x \geq 0$. For any distribution H, let

$$\bar{H}(x) = 1 - H(x). \tag{1.1}$$

It is often appropriate to consider a renewal process as a point process on $\mathbb{R}_+ = [0, \infty)$. Let $\mathcal{N}_+$ be the space of integer-valued measures on $\mathbb{R}_+$ with finite mass on each bounded interval, endowed with the vague topology: $\nu_n \to \nu$ vaguely in $\mathcal{N}_+$ if $\int f\, d\nu_n \to \int f\, d\nu$ for all continuous function f with compact support. Let $\mathcal{B}_+$ be the σ-field generated by the vague topology; a point process on $\mathbb{R}_+$ is a random element in $(\mathcal{N}_+, \mathcal{B}_+)$. In particular, the point process N associated with a renewal process $S = (S_n)_0^\infty$ ($S_n = \Sigma_0^n Y_i$) gives mass 1 to each renewal:

$$(1.2) \qquad N(A) = \#\{n; S_n \in A\} \qquad \text{for} \quad A \in \mathcal{R}_+ .$$

In more compact form, $N = \Sigma_0^\infty \delta_{S_n}$.

The renewal measure M of S (or N) is just the expectation measure (sometimes called intensity measure) of N:

$$(1.3) \qquad M(A) = \mathbf{E}[N(A)] \qquad \text{for} \quad A \in \mathcal{R}_+ .$$

You should recall that M is finite for bounded sets A. As is standard, we give M a special name, U, in the zero-delayed case ($Y_0 \equiv 0$). Certainly,

(1.4) (i) $U = \sum_0^\infty F^{*n}$, and

(ii) $M = G * U$ if $Y_0 \stackrel{\mathcal{D}}{=} G$.

The counting process $N = (N_t)_0^\infty = \{N_t; t \geq 0\}$ associated with S is defined by

$$(1.5) \qquad N_t = N([0, t]) ,$$

and the renewal function M by

$$M(t) = M([0, t]) = \mathbf{E}[N_t] \qquad \text{for} \quad t \geq 0 .$$

Again, we use U in the zero-delayed case: $U(t) = U([0, t])$ for $t \geq 0$.

No harm should be caused by giving two meanings to each of N, M, and U; we save ourselves from trifling distinctions through that.

For $t \geq Y_0$, let

$$A_t = \min\{t - S_n; t - S_n \geq 0\} \tag{1.6}$$

and for $t \geq 0$,

$$D_t = \min\{S_n - t; S_n - t > 0\}\,. \tag{1.7}$$

The process $D = (D_t)_0^\infty$ is a continuous-time analog to the Markov chain $(D_n)_0^\infty$ in (I.1.2); it is a (even strong) Markov process. However, we have adopted the convention to let the paths of D be right-continuous; that renders $D_t > 0$. The right-continuity is used to prove the strong Markov property of D.

The process $A = (A_t)_0^\infty$ has been left undefined for $t < Y_0$. If we also wish A to be a Markov process, Y_0 must be given a particular meaning. To do that, think of A_t as the age at time t of the component at work at that time point, and take an $a \geq 0$ to be the age at time 0 of a component, real or imaginary. For a distribution G on $[0, \infty)$, define G_a, the residual life-length distribution at a, by

$$G_a(x) = \begin{cases} 1, & \text{if} \quad G(a) = 1, \\ (G(a + x) - G(a))/\bar{G}(a), & \text{if} \quad G(a) < 1. \end{cases} \tag{1.8}$$

[Hence $G_a = \delta_0$ if $G(a) = 1$.] Let Y_0 have distribution F_a; then with $A_t = a + t$ for $t < Y_0$, $(A_t)_0^\infty$ becomes a Markov process. Of course, that value a of A_0 can be randomized.

There is no general agreement on names for the processes A and D, defined by (1.6) and (1.7). Let us call A the age process and D the delay process; these names are at least as suitable as many others suggested.

2. Stationarity. For $t \geq 0$, define the shift operator θ_t: $\mathcal{N}_+ \to \mathcal{N}_+$ by

$$(\theta_t \nu)(\cdot) = \nu(t + \cdot) \qquad \text{for} \quad \nu \in \mathcal{N}_+\,.$$

The shift θ_t puts a new zero point at t.

For a renewal process N, $\theta_t N$ is another such process, with delay $= D_t$. It is stationary if $\theta_t N \stackrel{\mathcal{D}}{=} N$ for all $t \geq 0$, and this is the case if and only if $D_t \stackrel{\mathcal{D}}{=} Y_0$ for all $t \geq 0$. Let $\mu = \mathbf{E}[Y_1]$, the expected life length. Throughout this section we assume that $\mu < \infty$ if no special attention for the infinite case is called upon. Recall that $\lambda = 1/\mu$ = the renewal intensity.

We first notice that if $\mu = \infty$, then N cannot be stationary, because if it were, the renewal measure M would satisfy $M = M(t + \cdot)$ for all $t \geq 0$. This forces M to equal $\alpha \cdot l_+$, where α is a strictly positive constant and l_+ is the Lebesgue measure restricted to $[0, \infty)$. We obtain $M(t) = \alpha \cdot t$. However, the result known as the elementary renewal theorem says that $\lim_{t \to \infty} M(t)/t = \lambda$, which equals 0 if $\mu = \infty$, and a contradiction may be deduced.

But if $\mu < \infty$, then $Y_0 \overset{\mathcal{D}}{=} G_s$, where

$$dG_s(x) = g_s(x)\, dx = \lambda \cdot \bar{F}(x)\, dx \tag{2.1}$$

renders N stationary, as will now be proved. First we show that $M(t) = \lambda \cdot t$ for $t \geq 0$, where M is the renewal function with $Y_0 \overset{\mathcal{D}}{=} G_s$. The standard renewal argument (conditioning with respect to Y_0) gives $M = G_s * U$, so

$$M = G_s * U = G_s + G_s * U * F = G_s + M * F .$$

Hence M is a solution to the renewal equation

$$A = G_s + A * F \tag{2.2}$$

(A unknown). But it has only one solution that is bounded on finite intervals, so $M(t) = \lambda \cdot t$ if $\lambda \cdot t$ solves (2.2). And it does:

$$\begin{aligned} G_s(t) &+ \int_{[0,\,t]} \lambda \cdot (t - x)\, dF(x) \\ &= \lambda \cdot \left\{ \int_{[0,\,t]} \bar{F}(x)\, dx + \int_{[0,\,t]} (t - x)\, dF(x) \right\} \\ &= \lambda \cdot t \cdot \left\{ t - \int_{[0,\,t]} F(x)\, dx - \int_{[0,\,t]} (t - x)\, dF(x) \right\} = \lambda \cdot t , \end{aligned}$$

as you find after an integration by parts.

Now fix an x and let $\mathbf{P}(D_t > x)$ be denoted by $V(t)$ and $V_0(t)$ for the cases $Y_0 \overset{\mathcal{D}}{=} G_s$ and $Y_0 \equiv 0$, respectively. To prove that $V(t)$ does not vary with t, we apply the renewal argument and find that

$$V = 1 - G_s(x + \cdot) + V_0 * G_s .$$

We have $V_0 = \bar{F}(x + \cdot) * U$, and since $U * G_s = \lambda \cdot l_+$ we get

$$\begin{aligned} V(t) &= 1 - G_s(x+t) + (\bar{F}(x + \cdot) * (\lambda \cdot l_+))(t) \\ &= 1 - G_s(x+t) + \int_{[0,t]} g_s(x+t-u)\, du \\ &= 1 - G_s(x+t) + \int_{[x,x+t]} g_s(u)\, du = \bar{G}_s(x) . \end{aligned}$$

That was a well-known proof, demonstrating the strength of the renewal argument. But it is not very illuminating: what made G_s a candidate from the outset, for example? To throw light on that, assume that F has a density f and let r denote the failure rate function of F:

$$r(x) = f(x)/\bar{F}(x) . \tag{2.3}$$

We make the convention that $r(x) = \infty$ if $F(x) = 1$. The integrated failure rate function R is defined by

$$R(x) = \int_0^x r(y)\, dy . \tag{2.4}$$

Notice that the residual lifelength distribution F_a of (1.3) has failure rate function $r_a = r(a + \cdot)$, with integral $R_a = R(a + \cdot) - R(a)$.

Let H be the distribution of A_0, the value of the age process at time 0. If N is stationary, we have $A_0 \overset{\mathcal{D}}{=} A_t$ for all $t > 0$. Hence for any $x > 0$ and a small $\delta > 0$,

$$\begin{aligned} \bar{H}(x) &= \mathbf{P}(A_\delta > x) = \mathbf{P}(A_0 > x - \delta,\, Y_0 > \delta) \\ &\approx \int_{x-\delta}^{\infty} (1 - r(a)\cdot \delta)\, dH(a) , \end{aligned}$$

implying that

$$(\bar{H}(x) - \bar{H}(x-\delta))/\delta \approx -\int_{x-\delta}^{\infty} r(a)\, dH(a) ,$$

so if H has a differentiable density h, we are led to believe that

$$h'(x) = -r(x) \cdot h(x) \quad \text{for} \quad x > 0.$$

That yields $h(x) = C \cdot \exp(-R(x))$ for some constant C, and we must have $C = \lambda$ since H is a probability distribution. But $\exp(-R(x)) = \bar{F}(x)$, so $H = G_s$, and for Y_0 we get

$$\begin{aligned}
\mathbf{P}(Y_0 > x) &= \int \mathbf{P}(Y_0 > x \| A_0 = a)\, dH(a) = \int \exp(-R_a(x))\, dH(a) \\
&= \int \exp(-(R(a+x) - R(a)) \cdot \lambda \cdot \exp(-R(a))\, da \\
&= \int \lambda \cdot \exp(-R(a+x))\, da = \int_x^\infty \lambda \cdot \exp(-R(a))\, da \\
&= \bar{G}_s(x).
\end{aligned}$$

Hopefully, we now understand better why G_s turns up. We will learn more about stationarity in the account of regenerative processes in § 17.

That reasoning had the advantage in teaching us something about failure rates. For one that is shorter, let H denote the distribution of D_t (the same for all t, since N is stationary). We have

$$\begin{aligned}
\bar{H}(x) &= \mathbf{P}(D_0 > x) = \mathbf{P}(D_\delta > x) \\
&\approx \mathbf{P}(D_0 > x + \delta) + \mathbf{P}(D_0 < \delta) \cdot \mathbf{P}(Y_1 > x) \\
&= \bar{H}(x + \delta) + \mathbf{P}(D_0 < \delta) \cdot \bar{F}(x),
\end{aligned}$$

so if H has density h, we obtain from

$$(\bar{H}(x) - \bar{H}(x + \delta))/\delta \approx (H(\delta)/\delta) \cdot \bar{F}(x)$$

that

$$h(x) = h(0) \cdot \bar{F}(x).$$

But $h(0)$ is forced to equal λ.

We conclude this investigation with the remark that we should not be surprised by our finding that $A_t \stackrel{\mathscr{D}}{=} D_t$ for all $t > 0$ under stationarity; reverse the time to make that plausible.

3. Blackwell's renewal theorem. This celebrated result states that

$$U[x, x + A] \to A/\mu = A \cdot \lambda \qquad \text{as} \quad x \to \infty \tag{3.1}$$

for all $A > 0$, under the sole condition that F is of nonlattice type; that is, of course, also necessary. Using (1.4)(ii), we find without effort that (3.1) holds for any delay distribution.

There is a long row of different proofs of (3.1). Since Blackwell's renewal theorem is a result of some depth, there will never be a short and entirely elementary one. Proofs may be more or less appealing, however.

The following technical lemma, possibly recognized by the reader, is needed for our coupling proof of (3.1). For a set A, define $d(x, A)$ by $d(x, A) = \inf\{|x - y|; y \in A\}$.

(3.2) Lemma. *The support of U is asymptotically dense, i.e., $d(x, \operatorname{supp}(U)) \to 0$ as $x \to \infty$.*

Proof. For convenience, let $\operatorname{supp}(U)$ be denoted by Σ. For a, $b \in \Sigma$ such that $a < b$, let

$$B_m(a, b) = \{ma, (m-1)a + b, \ldots, a + (m-1)b, mb\}$$

for integers $m \geq 1$. We have $B_m(a, b) \subset \Sigma$, and the distance between successive points in $B_m(a, b)$ equals $\Delta = b - a$. Let

$$m_0 = \min\{m; (m+1) \cdot a \leq m \cdot b\}\,.$$

Now, any interval I of length $\geq \Delta$ such that $I \subset [(m_0 + 1) \cdot a, \infty)$ contains a point from

$$B(a, b) = \bigcup_{m=1}^{\infty} B_m(a, b)\,,$$

which is a subset of Σ. Hence if Σ is not asymptotically dense, then $\Sigma = \{x_1, x_2, \ldots\}$ for an increasing sequence $(x_n)_1^\infty$ such that $\delta =$

$\liminf_n(x_{n+1}-x_n)>0$. But $(x_{n+1}-x_n)\geq\limsup_m(x_{m+1}-x_m)$ for all n (you agree on that after a moment's thought), so $\delta=\lim_n(x_{n+1}-x_n)$. Now take an $x\in\Sigma$. Since $x+\Sigma\subset\Sigma$ we must have $x=k\cdot\delta$ for some $k\geq 0$ [because for each n there is a $n'>n$ such that $x+x_n=x_{n'}$, and we know about $(x_{n'}-x_n)$ that ... .]. Hence $\Sigma\subset L_\delta$, which contradicts the nonlattice assumption about F. $\square$

For the proof of (3.1) in the case $\mu<\infty$, let S and S' be independent, with S zero-delayed and S' stationary. We cannot hope for a successful coupling of S and S'; in fact, the probability that S, S' have a renewal in common is 0 if F is continuous. But it suffices that S comes close to S' with probability 1, in other words, that we have what shall be called a successful ϵ-coupling.

The notation $Y_0', Y_1', \ldots, N', M'$ has its obvious meaning; since S' is stationary, we have $M'=\lambda\cdot l_+$.

For $i\geq 0$, let

$$Z_i=\min\{S_j'-S_i;\, S_j'-S_i\geq 0\}\,,$$

and for a fixed $\epsilon>0$, let

$$A_i=\{Z_j<\epsilon \text{ for some } j\geq i\}\,.$$

We have

$$A_0\supset\ldots\supset\bigcap_{i=0}^{\infty}A_i=A_\infty=\{Z_i<\epsilon \text{ i.o.}\}\,.$$

Due to the independence of S and S', the stationarity of S' and the fact that $(S_{i+n}-S_i)_{n=0}^{\infty}$ is again a zero-delayed process for each i, we get that $\theta_i Z\overset{\mathcal{D}}{=}Z$, where $Z=(Z_n)_0^{\infty}$. Hence all the sets A_i have the same probability, and in particular, $\mathbf{P}(A_0)=\mathbf{P}(A_\infty)$. Now $\mathbf{P}(A_\infty\,\|\,Y_0'=t)$ equals 0 or 1 for every t by virtue of the Hewitt–Savage 0–1 law applied to the i.i.d. sequence (Y_1, Y_1'), $(Y_2, Y_2'), \ldots$ (or to the sequence $Y_1', Y_2', \ldots$, after conditioning on $Y_1, Y_2, \ldots$), Lemma (3.2) yields that $\mathbf{P}(A_0\,\|\,Y_0'=t)>0$ for every t. These observations and the equality

$$\int\mathbf{P}(A_0\,\|\,Y_0'=t)g_s(t)\,dt=\mathbf{P}(A_0)=\mathbf{P}(A_\infty)=\int P(A_\infty\,\|\,Y_0'=t)g_s(t)\,dt$$

force $\mathbf{P}(A_\infty \| Y_0' = t)$ to equal 1 a.e. with respect to the distribution of Y_0'. Hence $\mathbf{P}(A_\infty) = 1$, a fortiori

$$\mathbf{P}(A_0) = \mathbf{P}(Z_i < \epsilon \text{ for some } i) = 1 .$$

Let

$$\tau = \min\{i; Z_i < \epsilon\},\ \tau' = \min\{j; S_j' \geq S_\tau\} .$$

With

$$\begin{aligned} N''[x, x + A] = {} & N([x, x + A] \cap [0, S_\tau]) \\ & + N'([x + Z_\tau, x + A + Z_\tau] \cap (S'_{\tau'}, \infty)) \end{aligned}$$

we certainly have $N''[x, x + A] \stackrel{\mathcal{D}}{=} N[x, x + A]$. Hence

$$\begin{aligned} U[x, x + a] &= \mathbf{E}[N''[x, x + A]] \\ &= \mathbf{E}[N([x, x + A] \cap [0, S_\tau])] \\ &\quad + \mathbf{E}[N'[x + Z_\tau, x + A + Z_\tau]] \\ &\quad - \mathbf{E}[N'([x + Z_\tau, x + A + Z_\tau] \cap [0, S'_{\tau'}])] \\ &= V_1(x) + V_2(x) - V_3(x), \quad \text{say.} \end{aligned}$$

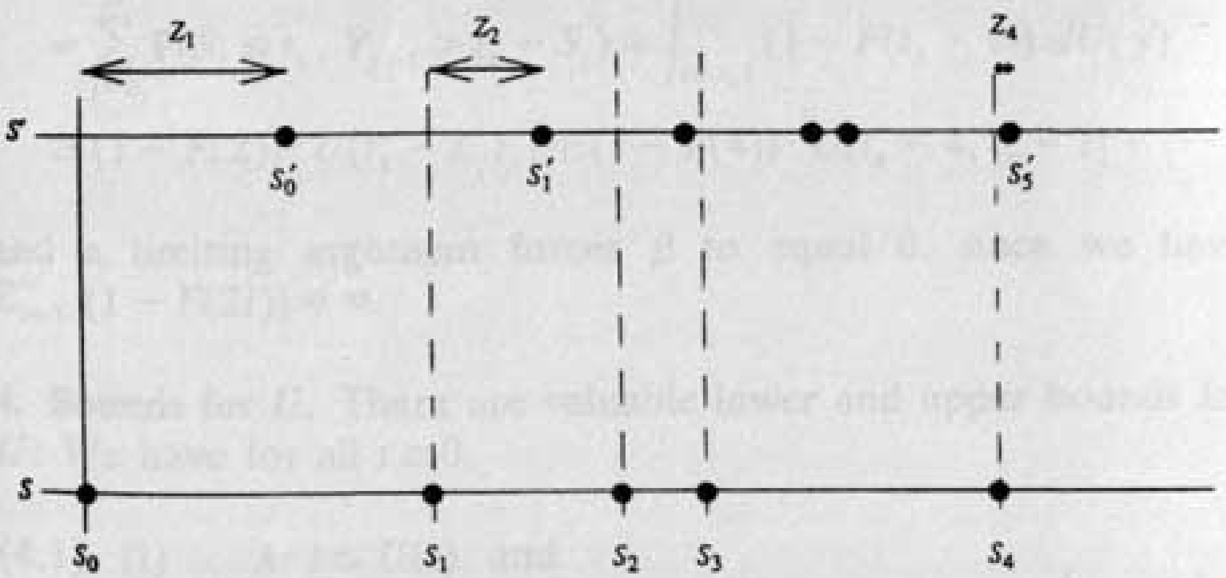

Figure 2. We wait until $Z_i < \epsilon$, which happens for $i = 4$.

Since we may choose $\epsilon > 0$ arbitrarily small, we can make $V_2(x)$ abitrarily close to $A \cdot \lambda$ uniformly in x, due to the fact that $Z_\tau < \epsilon$ implies

$$\begin{aligned}(A - \epsilon) \cdot \lambda = U'[x + \epsilon, x + A] \le V_2(x) \le U'[x, x + A + \epsilon] \\ = (A + \epsilon) \cdot \lambda .\end{aligned}$$

Next, since $N[x, x + A] \overset{\mathscr{D}}{\le} N[0, A]$ we have for any $a > 0$,

$$\begin{aligned}V_1(x) &\le \mathbf{E}[N[x, x + A] \cdot I(x \le S_\tau)] \\ &\le a \cdot \mathbf{P}(x \le S_\tau) + \mathbf{E}[N[0, A]; N[0, A] \ge a] .\end{aligned}$$

Now the last term tends to 0 as $a \to \infty$ due to dominated convergence, and then $a \cdot \mathbf{P}(x \le S_\tau) \to 0$ as $x \to \infty$ since $S_\tau < \infty$ a.s. We show similarly that $V_3(x) \to 0$ as $x \to \infty$, and the proof of Blackwell's renewal theorem is complete.

The crucial point was to show that an ϵ-coupling takes place a.s. If we try to do that relying on results of recurrence, we meet the same type of obstacle as in the discrete case (cf. § II.13): The step sizes $(Y_n - Y'_n)$ of the random walk $(S_n - S'_n)_0^\infty$ can be of lattice type even if F is not. For example, let $F = \frac{1}{2} \cdot \delta_1 + \frac{1}{2} \cdot \delta_{2^{1/2}}$. This difficulty may be sorted out by use of the randomization method in § II.13.

We now turn our attention to $\mu = \infty$. In the discrete case, the result $v_n \to 0$ came out as a by-product of our results on null-recurrent Markov chains, so there was no reason to present the rather simple analytical proof. Analogous theory that yields $U[x, x + A] \to 0$ as $x \to \infty$ when $\mu = \infty$ in the nonlattice case will not be developed, so this is a place for an analytical argument; it is just an extension of that often used for the discrete case.

Let $(t_k)_0^\infty$ be an increasing sequence tending to ∞ such that

$$\beta = \limsup_{t \to \infty} U(t, t + 1] = \lim_{k \to \infty} U(t_k, t_k + 1] . \tag{3.3}$$

We shall prove that $\beta = 0$. Conditioning with respect to S_i yields

$$\beta = \lim_{k \to \infty} \int U(t_k - y, t_k + 1 - y] \, dF^{*i}(y) \tag{3.4}$$

for every i. Now the asymptotic denseness of $\operatorname{supp}(U)$ implies that there exists a j_0 such that $U(j, j+1] > 0$ for $j \geq j_0$. Fix such a j, and choose an $i(j) = i$ such that $F^{*i}(j, j+1] > 0$. Using (3.3), we easily get from (3.4) that

$$\begin{aligned}
\beta = \liminf_{k\to\infty} \Big\{ \int_{y \notin (j,j+1]} & U(t_k - y, t_k + 1 - y]\, dF^{*i}(y) \\
& + \int_{(j,j+1]} U(t_k - y, t_k + 1 - y]\, dF^{*i}(y) \Big\} \\
\leq \beta \cdot (1 - F^{*i}&(j, j+1]) \\
& + \liminf_{k\to\infty} \int_{(j,j+1]} U(t_k - y, t_k + 1 - y]\, dF^{*i}(y) \\
\leq \beta \cdot (1 - F^{*i}&(j, j+1]) \\
& + (\liminf_{k\to\infty} U(t_k - j - 1, t_k - j + 1]) F^{*i}(j, j+1]
\end{aligned}$$

and we may conclude that

$$\liminf_{k\to\infty} U(t_k - j, t_k - j + 2] \geq \beta \qquad \text{for} \quad j > j_0 .$$

Now

$$\begin{aligned}
1 = \mathbf{P}(S_n > t_k \text{ for some } n) &= \sum_{j=0}^{\infty} \mathbf{P}(S_j \leq t_k, S_{j+1} > t_k) \\
= \sum_{j=0}^{\infty} \mathbf{P}(S_j \leq t_k, Y_{j+1} > t_k - S_j) &= \int_{(0,t_k]} (1 - F(t_k - y))\, dU(y) \\
\geq (1 - F(2)) \cdot U(t_k - 2, t_k] &+ (1 - F(4)) \cdot U(t_k - 4, t_k - 2] + \cdots
\end{aligned}$$

and a limiting argument forces β to equal 0, since we have $\Sigma_{i=1}^{\infty} (1 - F(2i)) = \infty$.

4. Bounds for U. There are valuable lower and upper bounds for U: We have for all $t \geq 0$,

$$\begin{aligned}
(4.1)\quad &\text{(i)} \qquad \lambda \cdot t \leq U(t), \text{ and} \\
&\text{(ii)} \qquad U(t) \leq \lambda \cdot t + \lambda^2 \cdot \mu_2 ,
\end{aligned}$$

where $\mu_2 = \int x^2\, dF(x)$. The bound (i) is well known (it is needed in the proof of the elementary renewal theorem) and usually established with use of Wald's lemma. Perhaps you are less familiar with the upper bound (ii); it is known as Lorden's inequality.

For (i), let $Y_0', Y_1, Y_2, \ldots$ be independent variables, $Y_0' \stackrel{\mathcal{D}}{=} G_s$ and $Y_i \stackrel{\mathcal{D}}{=} F$ for $i \geq 1$, and let N, N' be the renewal processes with renewals at $S_n = \Sigma_0^n\, Y_i$, $S_n' = Y_0' + S_n$ for $n \geq 0$; hence N is zero-delayed and N' stationary. Of course, $N'(t) \leq N(t)$ for all $t \geq 0$, so

$$\lambda \cdot t = \mathbf{E}[N_t'] \leq \mathbf{E}[N_t] = U(t)\,.$$

For the proof of (ii), let Y_0, Y_0' be independent and have distribution G_s. We shall use the fact (easily proved) that U is subadditive, that is,

$$U(s+v) \leq U(s) + U(v) \qquad \text{for all} \quad s, v;$$

notice that $U(s) = 0$ for $s < 0$. We have

$$(4.2)\quad U(t) = \mathbf{E}[U(t)] \leq \mathbf{E}[U(t + Y_0 - Y_0')] + \mathbf{E}[U(Y_0' - Y_0)]\,.$$

Now

$$\mathbf{E}[U(t + Y_0 - Y_0')] = \mathbf{E}[\mathbf{E}[U(t + Y_0 - Y_0') \| Y_0]] = \mathbf{E}[\lambda \cdot (t + Y_0)]$$

since $\mathbf{E}[U(s - Y_0')] = \lambda \cdot s$ for all $s \geq 0$. With that device used again, we get $\mathbf{E}[U(Y_0' - Y_0)] = \lambda \cdot \mathbf{E}[Y_0']$. Hence from (4.2),

$$U(t) \leq \lambda \cdot t + \lambda \cdot (\mathbf{E}[Y_0 + Y_0']) = \lambda \cdot t + 2 \cdot \lambda \cdot \mathbf{E}[Y_0]\,.$$

But $\mathbf{E}[Y_0] = \lambda \cdot \mu_2/2$, and (ii) is established.

Notice that (ii) is sharp. Indeed, if the life lengths have expected life length 1 and a very small variance, then $U(t) \approx 1 + [t]$ for t not too large; and $\lambda \cdot t + \lambda^2 \cdot \mu_2 \approx 1 + t$.

It would be interesting with a more direct proof of

$$\mathbf{E}[N[0,t] - N[0, t - Y_0']] \leq \lambda^2 \cdot \mu_2\,,$$

which implies (4.1)(ii), but one may doubt that there is any.

5. An exact coupling. To prove Blackwell's renewal theorem, an ϵ-coupling was sufficient. But for its improvements, close is not enough; an exact (hence necessarily dependent) coupling must be produced. Its construction requires a not negligible amount of work, but it is worthwhile; the yield will be rich, and several of the ideas will be valuable later.

We shall assume that F is nonsingular; that is, there exists a subprobability measure $F_0 \neq 0$ with density f_0 such that

$$F(A) \geq F_0(A) = \int_A f_0(x)\,dx$$

for all $A \in \mathcal{R}_+$. One also says that F has an absolutely continuous component (w.r.t. the Lebesgue measure) when this holds. Let $F_1 = F - F_0$. We may assume that f_0 is bounded and has a bounded support; there is no need for F_1 to be singular. Using only the boundedness condition, we find that f_0^{*2} is continuous. Indeed, for any function g that is bounded and integrable, there exists a sequence g_n, $n = 1, 2, \ldots$, of continuous such functions such that $\|g_n - g\|_1 \to 0$ as $n \to \infty$. Now $g_n * g$ is continuous due to dominated convergence, and since $\|g_n * g - g^{*2}\|_\infty \leq \|g\|_\infty \cdot \|g_n - g\|_1$ we may deduce that g^{*2} is that, too, as the uniform limit of continuous functions (we used $\|\cdot\|_1$ and $\|\cdot\|_\infty$ to denote L^1 and L^∞ norms). The coupling will be constructed with an easy proof of the rate results to follow in mind. Two lemmas will be requisite.

(5.1) Lemma. *If S is zero-delayed, there exist constants A, c_1, $c_2 > 0$ such that the distribution of A_t has an absolutely continuous component with a density $\geq c_1$ on $[0, c_2]$ for all $t \geq A$.*

Proof. Recall that we use C and c to denote strictly positive generic constants. We have

$$\sum_0^\infty F^{*n} \geq F^{*2} * U = U_0\,,$$

where U_0 has a component with continuous density $U * f_0^{*2} = u_0$. Picking an interval I where $f_0^{*2} > c$, we find $u_0(x) \geq c \cdot U(x - I) > c$ for $x \geq$ some A' due to Blackwell's renewal theorem. For the zero-delayed process, standard arguments render

$$\mathbf{P}(A_t \in [x, x+dx]) = \sum_0^\infty \mathbf{P}(S_n \in [t-x, t-x+dx], S_{n+1} > t)$$
$$= U([t-x, t-x+dx]) \cdot \bar{F}(x) \geq c \cdot \bar{F}(x)\,dx$$

when $t - x \geq A'$. Hence if we take c_2 so small that $F((c_2, \infty)) > 0$ and let $A = A' + c_2$, the lemma follows. □

(5.2) Lemma. *Let H_1, H_2 by probability measures on $\mathbb{R}_+$ with absolutely continuous component densities h_1, h_2, respectively, and put $\gamma_1 = \int h_1(y) \wedge h_2(y)\,dy$. Then there exists a coupling H of H_1 and H_2 satisfying $H(\{(x, y); x = y\}) = H(\Delta) \geq \gamma_1$.*

Proof. The result is a consequence of the γ coupling theorem (I.5.2). However, to settle the present lemma, you need just a part of the proof of that result: Let the Q there have density $h_1 \wedge h_2$ with respect to the Lebesgue measure. The rest is easy. □

We shall now construct an efficient coupling of two renewal processes S and S', with delay distributions G and G', respectively. Let $\tilde{Y} = (\tilde{Y}_i)_0^\infty$ and $\tilde{Y}' = (\tilde{Y}'_i)_0^\infty$ be independent sequences of independent variables, $\tilde{Y}_0$, $\tilde{Y}'_0$ with distributions G and G', respectively, all other $\tilde{Y}_i$, $\tilde{Y}'_i$ with distribution F. Let $\tilde{S}$, $\tilde{S}'$ be the renewal processes associated with $\tilde{Y}$, $\tilde{Y}'$. Define V_0, V'_0, and Z_0 through

$$V_0 = \tilde{Y}_0, \qquad V'_0 = \tilde{Y}'_0 \qquad Z_0 = \max(V_0, V'_0)$$

and let $W_1 = Z_0 + A$, where A is from Lemma (5.1). On $[0, W_1]$, let S, S' coincide with $\tilde{S}$, $\tilde{S}'$.

At our halt at W_1, we make the first coupling attempt: If $A_{W_1} = a$, put

$$H_1 = F_a$$

[recall the definition (1.8) of F_a], define H_2 analogously for S', and let the overshoots of S, S' at W_1 be denoted by V_1, V'_1, respectively, where (V_1, V'_1) has the distribution H of Lemma (5.2). Hence S and S' have their first renewals in (W_1, ∞) at $W_1 + V_1$, $W_1 + V'_1$.

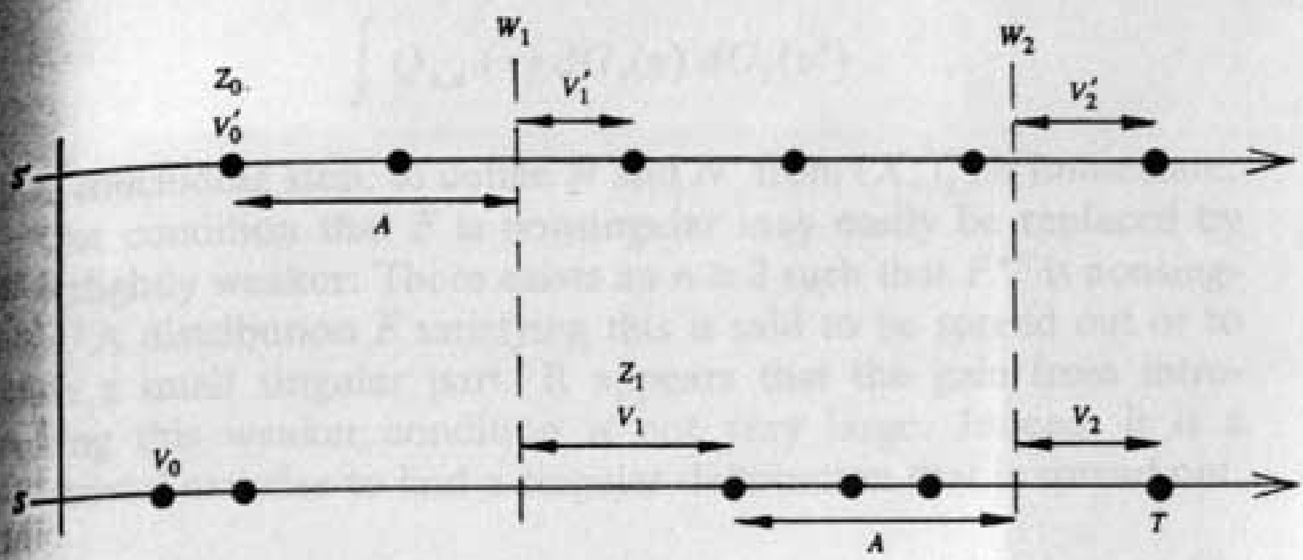

Figure 3. Two trials needed for a successful coupling. Notice that $Z_0 = V_0'$, $Z_1 = V_1$, and $Z_2 = V_2 = V_2'$.

With W_1 as a new starting point, put $Z_1 = \max(V_1, V_1')$ and take fresh segments of $\tilde{S}$, $\tilde{S}'$ to define S, S' on $(W_1 + V_1, W_1 + Z_1 + A]$ and $(W_1 + V_1', W_1 + Z_1 + A]$; $W_1 + Z_1 + A = W_2$ is our second halt time, and a new coupling attempt is carried out there like the one at W_1. This procedure is repeated until $V_i = V_i'$ at some halt W_i.

At that first successful trial, $W_i + V_i = T$ is a renewal common to S and S', hence a coupling time. The meaning of the notation W_i, V_i, V_i', Z_i for $i \geq 2$, some of it already used, is obvious.

But is $T < \infty$ a.s.? Yes, due to the crucial fact that $\mathbf{P}(V_i = V_i') \geq$ some $\gamma > 0$ uniformly in i. To show that, let $\delta_1 > 0$ be so small that $\int_{\delta_1}^{\infty} f_0(x)\,dx > 0$, and put $\delta = \delta_1 \wedge c_2$, where c_2 is from Lemma (5.1). Due to Lemmas (5.1)–(5.2), we have

$$\text{(5.3)}\quad \mathbf{P}(V_i = V_i') \geq c \cdot \iint_{s,t<\delta} \Big[\int_0^{\infty} f_0(x+s) \wedge f_0(x+t)\,dx \Big]\,ds\,dt$$

$$\geq c \cdot \iint_{s,t<\delta} \Big(\int_0^{\infty} I_B(x+s) \cdot I_B(x+t)\,dx \Big]\,ds\,dt$$

for all, i where B is a Borel set $\in \mathcal{R}_+$ such that $0 < l(B) < \infty$ and $f_0 \geq c \cdot I_B$. But the latter integral is strictly positive: indeed, it equals

$$\int_0^\infty \left[\int_{s<\delta} I_B(x+s)\,ds\right]^2 dx > 0\,.$$

And the number of coupling trials needed is dominated by a geometric distribution.

The construction of S and S' was hopefully instructive, but it was not complete. We have agreed to omit standard arguments for the existence of relevant probability spaces, but this time some elucidation might be in order. To provide that, let us build a Markov chain $(X_n)_0^\infty$ with state space $\mathcal{N}_+ \times \mathcal{N}_+$, where X_n is meant to represent $(N_{[w_n, w_{n+1}]}, N'_{[w_n, w_{n+1}]})$. Here N, N' are the point processes associated with S, S', respectively, and $\nu_I = \nu(\cdot \cap I)$ is the restriction to the interval I of $\nu \in \mathcal{N}_+$.

Notice that the vague topology on $\mathcal{N}_+$ is Polish and so is the product topology on $\mathcal{N}_+ \times \mathcal{N}_+$. Hence the space condition of Kolmogorov's consistency theorem is satisfied, so for the existence of $(X_n)_0^\infty$ we just have to produce the proper Markov kernel (cf. § 8).

To that end, let v, $v' \geq 0$, $z = \max(v, v')$ and $w = z + A$. We need the independent and zero-delayed processes $\tilde{S}$, $\tilde{S}'$ at our disposal; their existence follows indeed from standard arguments! Now let $\hat{N}$ and $\hat{N}'$ have points of occurrence at

$$v + \tilde{S}_n, \quad n \geq 0\,,$$

and

$$v' + \tilde{S}'_n, \quad n \geq 0\,,$$

respectively, if $v \neq v'$, and at $v + \tilde{S}_n$, $n \geq 0$ (common points of occurrence), if $v = v'$. Let $Q_{vv'}$ be the distribution of $(\hat{N}_{[0,w]}, \hat{N}'_{[0,w]})$. For ν, $\nu' \in \mathcal{N}_+$ such that ν, $\nu' \neq \mathbf{0}$, let z_0 be the largest of the first points of occurrence of ν and ν', $w_0 = z_0 + A$, and let a, a' be the "ages" of ν, ν' at w_0. Further, let H_1, H_2 be the residual life-length distributions w.r.t. to a and a', and H the coupling of H_1 and H_2 given by Lemma (5.2). Now define

$$P((\nu, \nu'), \cdot) = \int Q_{vv'}(\cdot)\, dH(v, v')\,. \tag{5.4}$$

We may let X_0 have the distribution

$$\int Q_{v,v'}(\cdot)\, dG_1(v)\, dG_2(v')\,.$$

The concluding step, to define N and N' from $(X_n)_0^\infty$, is immediate.

The condition that F is nonsingular may easily be replaced by one slightly weaker: There exists an $n \geq 2$ such that F^{*n} is nonsingular. A distribution F satisfying this is said to be spread out or to have a small singular part. It appears that the gain from introducing this weaker condition is not very large. Indeed, it is a nontrivial exercise to find a singular distribution that is spread out.

6. Finite moments of T. Rate results. It is not very hard to modify Theorem (II.4.2) to the present continuous-time setting, with F nonsingular. We now let

$$U_i = A + Z_i \qquad \text{for} \quad i \geq 1\,.$$

Obviously, with the notation of § 5,

$$T = Z_0 + \sum_1^\tau U_i\,,$$

where $\tau = \min\{i \geq 1;\ V_i = V_i'\}$. Also, let

$$\mathcal{B}_0 = \sigma\{V_0, V_0'\}$$

and

$$\mathcal{B}_i = \sigma\{Y_j, Y_j';\ S_j \leq W_i + V_i,\ S_j' \leq W_i + V_i'\}\,.$$

Notice that we proved in § 5 that $\mathbf{P}(V_i = V_i') \geq$ some $\gamma > 0$, uniformly in i. When transferring the proof of Theorem (II.4.2), we need the bound, for $\alpha \geq 1$,

$$\mathbf{E}[Z_i^\alpha \,\|\, \mathcal{B}_{i-1}] \leq C + C \cdot Z_{i-1}\,; \tag{6.1}$$

the rest is straightforward. But one of the variables V_i and V_i' is the overshoot at the level A for a zero-delayed process starting at $W_{i-1} + Z_{i-1}$, the other at the level $A + Z_{i-1} - \min(V_{i-1}, V_{i-1}')$, so

(6.1) follows if we can extend Lemma (II.4.1)(i). But if S is zero-delayed, then for $t>0$,

$$\mathbf{P}(D_t > x) = \int_{[0,t]} \bar{F}(t-s+x)\,dU(s)\,;$$

hence

$$\begin{aligned}\mathbf{E}[D_t^\alpha] &= \alpha\cdot\int x^{\alpha-1}\cdot\mathbf{P}(D_t>x)\,dx\\ &= \int_{[0,t]} dU(s)\left[\int \alpha\cdot x^{\alpha-1}\cdot\bar{F}(t-s+x)\,dx\right]\\ &\le \int_{[0,t]} dU(s)\left[\int \alpha\cdot x^{\alpha-1}\bar{F}(x)\,dx\right] \le \mu_\alpha\cdot U(t) \le C\cdot(1+t)\,,\end{aligned}$$

which suffices. The word "straightforward" was used above since the extension of Lemma (II.4.1)(ii) needs no comment.

Letting $m_\alpha(H) = \int x^\alpha\,dH(x)$ for a distribution H on $[0,\infty)$, and $\mu_\alpha = m_\alpha(F)$, we may now state:

(6.2) $$\mathbf{E}[T^\alpha] < \infty \qquad \text{for all} \quad \alpha > 0$$

if μ_α, $m_\alpha(G)$ and $m_\alpha(G') < \infty$.

We shall now use (6.2) to estimate the distance between the renewal measures of N and N' (i.e., $G*U$ and $G'*U$). The findings have uniform versions of the key renewal theorem as consequences. Letting $Q(t)$, $Q'(t)$ denote the distributions on $\mathcal{N}_+$ of $\theta_t N$, $\theta_t N'$, respectively, we obtain

(6.3) $$\|Q(t) - Q'(t)\| = \mathrm{o}(t^{-\alpha})$$

under the conditions of (6.2), and if N' is stationary ($G' = G_s$), then

(6.4) $$\|Q(t) - Q_s\| = \mathrm{o}(t^{-(\alpha-1)})$$

if $m_{\alpha-1}(G)$, $\mu_\alpha < \infty$ for an $\alpha > 1$ (the meaning of Q_s is obvious).

For $\nu \in \mathcal{N}_+$, let

$$\psi(\nu) = \inf\{s;\ \nu[0,s] \ge 1\}\,;$$

clearly, $\psi(\theta_t N)$ is the overshoot at t of S. Let H_t denote its distribution $[= Q(t)\psi^{-1}]$ and H'_t that of $\psi(\theta_t N')$. Due to the coupling–mapping inequality, we have

$$\|H_t - H'_t\| \le \|Q(t) - Q'(t)\| \tag{6.5}$$

and a moment's thought gives

$$\begin{aligned} \|M_{[t,t+1]} - M'_{[t,t+1]}\| &= \|(G * U)_{[t,t+1]} - (G' * U)_{t,t+1]}\| \\ &= \|(H_t * U)_{[0,1]} - (H'_t * U)_{[0,1]}\| \\ &\le U(1) \cdot \|H_t - H'_t\| \end{aligned} \tag{6.6}$$

[we used the inequality $\|\nu_1 * \nu_2\| \le \|\nu_1\| \cdot \|\nu_2\|$ (cf. § App.2) for signed bounded measures ν_1, ν_2].

From (6.3)–(6.6) we now obtain that if $\mu < \infty$, then for every $B > 0$,

(6.7) (i) $\|(G * U)_{[t,t+B]} - (G' * U)_{[t,t+B]}\| = o(t^{-\alpha})$ for $\alpha > 0$, if $m_\alpha(G)$, $m_\alpha(G')$ and $\mu_\alpha < \infty$, and

(ii) $\|(G * U)_{[t,t+B]} - \lambda \cdot l_{[0,B]}\| = o(t^{-(\alpha-1)})$ for $\alpha > 1$, if $m_{\alpha-1}(G)$ and $\mu_\alpha < \infty$.

For an extension of (6.7) to infinite intervals, we use (6.6) to find that

$$\|(G * U)_{[t,t+\infty)} - (G' * U)_{[t,t+\infty)}\| \le U(1) \cdot \sum_{k \ge [t]} \|H_k - H'_k\| .$$

But for that sum we have

$$\begin{aligned} t^{\alpha-1} \cdot \sum_{k \ge [t]} \|H_k - H'_k\| &\le 2 \cdot t^{\alpha-1} \cdot \sum_{k \ge [t]} \mathbf{P}(T > k) \\ &\le C \cdot \sum_{k \ge [t]} k^\alpha \cdot \mathbf{P}([T] = k) \\ &\le C \cdot \mathbf{E}[[T]^\alpha \cdot I(T \ge [t])] \to 0 \quad \text{as} \quad t \to \infty \end{aligned}$$

if $\mathbf{E}[T^\alpha] < \infty$. Hence:

(6.8) (i) $\|G * U(t+\cdot) - G' * U(t+\cdot)\| = o(t^{-(\alpha-1)})$ for $\alpha > 1$, if $m_\alpha(G)$, $m_\alpha(G')$, and $\mu_\alpha < \infty$,

(ii) $\|G * U(t+\cdot) - \lambda \cdot l_+\| = o(t^{-(\alpha-2)})$ for $\alpha > 2$; if $m_{\alpha-1}(G)$ and $\mu_\alpha < \infty$, and

(iii) $\|G * U - \lambda l_+\| < \infty$ if $m_1(G)$ and $\mu_2 < \infty$.

But more can be said about (ii) and (iii). For the sake of clarity, let us concentrate on the case where N is zero-delayed and N' is stationary. If $\mu_2 < \infty$, then (6.8)(iii), often referred to as Stone's decomposition, tells us that $U - \lambda l_+$ is a bounded signed measure, ν say. We know that

$$\nu(\mathbb{R}_+) = \mathbf{E}[N_T - N'_T].$$

But that expectation can be calculated with the method demonstrated after Theorem (II.5.1). Indeed,

$$T = S_{N_T - 1} = Y'_0 + \sum_{i=1}^{N'_T - 1} Y'_i.$$

An application of Wald's lemma yields

$$\mu \cdot \mathbf{E}[N_T - 1] = \lambda \cdot \mu_2/2 + \mu \cdot \mathbf{E}[N'_T - 1]$$

and we deduce that

$$\mathbf{E}[N_T - N'_T] = \lambda^2 \cdot \mu_2/2. \tag{6.9}$$

Using the observation $U(t) - \lambda \cdot t = \nu[0, t] = \nu(\mathbb{R}_+) - \nu(t, \infty) = \lambda^2\mu_2/2 - \nu(t, \infty)$, we may conclude: If $\mu_2 < \infty$, then

(6.10) (i) $\|U - \lambda \cdot l_+\| < \infty$,

(ii) $U(t) - \lambda \cdot t \to \lambda^2\mu_2/2$ as $t \to \infty$, and

(iii) $|U(t) - \lambda t - \lambda^2\mu_2/2| = o(t^{-(\alpha-2)})$ if $\mu_\alpha < \infty$, $\alpha > 2$.

The convergence (ii) holds without the assumption that F has an absolutely continuous component; it may be established by using the key renewal theorem.

It is natural to conjecture that if f is integrable, then

$$(6.11) \qquad U * f(t) \to \lambda \cdot \int_0^\infty f(x)\, dx \quad \text{as} \quad t \to \infty .$$

This is an important topic since $U * f$ is the unique solution to the renewal equation $A = f + A * F$ (f known, A unknown) under mild conditions. You know that (6.11) holds if f is directly Riemann integrable, as soon as F is nonlattice. But if F is nonsingular, that condition on f may be relaxed drastically.

For our result, we will need the fact that U can be decomposed into $U_0 + U_1$, where U_0 has a bounded density with respect to Lebesgue measure and U_1 is bounded. To understand that, recall the representation $F = F_0 + F_1$, where F_0 has a bounded density with a bounded support. We have

$$U = \sum_0^\infty F^{*n} = \sum_0^\infty (F_0 + F_1)^{*n} = F_0 * H + \sum_0^\infty F_1^{*n} = U_0 + U_1$$

for some measure H. Now, U_1 has finite mass $[= 1/(1 - F_1(\infty))]$. Also, U_0 has a bounded density; to see that, recall that $\sup_x U[x, x + A) < \infty$ for all $A > 0$ to prove that $\sup_x H[x, x + A) < \infty$ for all $A > 0$, then bound the density $H * f_0$ of U_0.

(6.12) Theorem. *If F is nonsingular and $\mu < \infty$, then*

$$\lim_{t\to\infty} \sup_{|f| \le g} \left| U * f(t) - \lambda \cdot \int_0^\infty f(s)\, ds \right| = 0$$

if

(i) *g is bounded,*
(ii) *g is integrable, and*
(iii) *$g(t) \to 0$ as $t \to \infty$.*

Proof. Let $g = g \cdot I_{[0,B]} + g \cdot I_{(B,\infty)} = g_1 + g_2$, where B is large and determined below. Split f in the same fashion. Due to (6.7)(ii), we get

$$\left| U * f_1(t) - \lambda \cdot \int_0^\infty f_1(s)\, ds \right| \le \| U_{[t-B,t]} - \lambda \cdot l_{[0,B]} \| \cdot \| g_1 \| ,$$

which tends to 0 uniformly in f as $t \to \infty$. Further,

$$(6.13)\quad \left| U * f_2(t) - \lambda \cdot \int_0^\infty f_2(s)\, ds \right| \le U * g_2(t) + \lambda \cdot \int_0^\infty g_2(s)\, ds\,.$$

Now

$$U * g_2(t) = U_0 * g_2(t) + U_1 * g_2(t) \le C \cdot \int_0^\infty g_2(s)\, ds$$
$$+ \|g_2\| / (1 - F_1(\infty))\,,$$

so, leaning on (ii) and (iii), we may choose B sufficiently large to make (6.13) arbitrarily small, and complete the proof. □

7. Notes. For introductions and applications of continuous-time renewal theory, see, for example, Karlin and Taylor [86] (gives no proof of Blackwell's renewal theorem, however) and Asmussen [10]. For the theory of point processes, see Daley and Vere-Jones [45]. The proof of Blackwell's renewal theorem presented here is from Lindvall [105]. A well-known analytical proof is that of [10, Chap. IV]. See Breiman [30] for one using Fourier transforms.

The elegant proof of Lorden's inequality is due to Carlsson and Nerman [34]. See also Asmussen [10] for that topic. Theorem (6.10) is from Arjas, Nummelin, and Tweedie [8], an impetus for Lindvall [108], which is the basis for § 6. For generalizations, alternative approaches, and special cases, see Athreya, McDonald, and Ney [15] [which ignores the possibility that $(Y_i - Y_i')$ may have a lattice distribution; cf. § 3], Berbee [22, 23], Lalley [96–98], Ney [122], Thorisson [153], and Parts 5 and 6 in Chapter V.

2. HARRIS CHAINS

8. Basics. This is an appropriate place to settle some notation, to be used not only in this section but also later, especially in Chapter IV. Let $(E_0, \mathscr{E}_0)$, $(E_1, \mathscr{E}_1)$ be two state spaces. A transition kernel in $E_0 \times E_1$ (w.r.t. $\mathscr{E}_0$ and $\mathscr{E}_1$) is a mapping $E_0 \times \mathscr{E}_1 \to [0, 1]$ such that

(8.1) (i) $K(x, \cdot\,)$ is a probability measure on $(E_1, \mathscr{E}_1)$ for each $x \in E_0$, and

(ii) $K(\cdot, A)$ is a measurable mapping $E_0 \to [0,1]$ for each $A \in \mathscr{E}_1$.

For a probability measure λ on $(E_0, \mathscr{E}_0)$, $\lambda \cdot K = \lambda K$ is the probability measure on $(E_1, \mathscr{E}_1)$ defined by

$$\int \lambda(dx) K(x, dy) . \tag{8.2}$$

With a third state space $(E_2, \mathscr{E}_2)$ and transition kernels K_1 in $E_0 \times E_1$ and K_2 in $E_1 \times E_2$ at hand, the product $K_1 \cdot K_2 = K_1 K_2$ is the transition kernel in $E_0 \times E_2$ given by

$$(K_1 K_2)(x, A) = \int K_1(x, dy) K_2(y, A) . \tag{8.3}$$

Let $\lambda \circ K$ denote the probability on $(E_0 \times E_1, \mathscr{E}_0 \times \mathscr{E}_1)$ defined by

$$\int_{A_0} \lambda(dx) K_1(x, A_1) \tag{8.4}$$

for rectangles $A_0 \times A_1$. For a sequence of state spaces $(E_i, \mathscr{E}_i)$, $i \geq 0$, and transition kernels K_i in $(\Pi_0^{i-1} E_j) \times E_i$, for $i \geq 1$, we may construct a probability $\lambda \circ K_1 \circ \cdots \circ K_n$ on $(\Pi_0^n E_i, \Pi_0^n \mathscr{E}_i)$, where $\Pi_0^n \mathscr{E}_i$ is the standard product σ-field, from (8.4) by induction. Due to the Kolmogorov consistency theorem, there exists a (unique) probability $\mathbf{P}$ on $(\Pi_0^\infty E_i, \Pi_0^\infty \mathscr{E}_i)$ such that

(8.5)

$$\mathbf{P}(A_0 \times \cdots \times A_n \times E_{n+1} \times \cdots) = (\lambda \circ K_1 \circ \cdots \circ K_n)(A_0 \times \cdots \times A_n)$$

for rectangles $A_0 \times \cdots \times A_n \times \cdots \in \Pi_0^\infty \mathscr{E}_i$. Notice that we do not define $K_1 \circ K_2$ per se, and that $\lambda \circ K_1 \circ \cdots \circ K_n$ should be read from left to right in the obvious way.

A transition kernel K as defined by (8.1) may be seen as a bounded linear operator with domain $b\mathscr{E}_1$ and range $b\mathscr{E}_0$ defined by

$$(Kf)(x) = \int f(y) K(x, dy) \qquad \text{for} \quad f \in b\mathscr{E}_1 . \tag{8.6}$$

If all the spaces $(E_i, \mathscr{E}_i)$ are equal, to $(E, \mathscr{E})$ say, and P is a transition kernel in $E \times E$, we refer to it as a Markov kernel in E. With $K_j((x_0, \ldots, x_{j-1}), \cdot) = P(x_{j-1}, \cdot)$ for $j \geq 1$, we use (8.5) to obtain a probability measure $\mathbf{P}_\lambda$ on $(E^\infty, \mathscr{E}^\infty)$ such that

$$\begin{aligned}&\mathbf{P}_\lambda(A_0 \times \cdots \times A_n \times E \times E \times \cdots)\\&\quad = \int_{A_0} \lambda(dx_0) \int_{A_1} P(x_0, dx_1) \cdots \int_{A_{n-1}} P(x_{n-2}, dx_{n-1}) P(x_{n-1}, A_n)\end{aligned}$$

for rectangles $A_0 \times \cdots \times A_n \times \cdots$. The probability measure λ on E will play the role of an initial distribution.

With $\Omega = E^\infty$ and $\mathscr{F} = \mathscr{E}^\infty$, $(\Omega, \mathscr{F})$ is the sample space and $(\Omega, \mathscr{F}, \mathbf{P}_\lambda)$ the underlying probability space for the canonical Markov chain $X = (X_n)_0^\infty$, that is, the coordinate process

$$X_n(\omega) = \omega_n$$

for $\omega = (\omega_n)_0^\infty \in \Omega$. The process X is indeed a Markov chain:

$$\mathbf{P}(X_{n+1} \in A \| \mathscr{F}_n) = P(X_n, A) \qquad \text{under } \mathbf{P}_\lambda\,, \tag{8.7}$$

where $\mathscr{F}_n = \sigma\{X_0, \ldots, X_n\}$, and it has initial distribution λ. With a cautious interpretation, the equality (8.7) implies that

$$\mathbf{P}_\lambda(X_{n+1} \in A \| X_n = x) = P(x, A)\,. \tag{8.8}$$

We call both $P(X_n, A)$ and $P(x, A)$ in (8.7) and (8.8) transition probabilities of the Markov chain. Notice that $(X_0, \ldots, X_n)$ has distribution $\lambda \circ P \circ \cdots \circ P$, and X_n distribution $\lambda \cdot P \cdot \ldots \cdot P = \lambda P^n$ for $n \geq 0$.

If $X = (X_n)_0^\infty$ is a sequence of E-valued random variables defined on some probability space with probability measure $\mathbf{P}$, we call it a Markov chain if there exists a Markov kernel P such that (8.4) holds for all $n \geq 0$. It is not unusual that a random sequence X fails to be a Markov chain, but there exists an extension

$$\tilde{X}_n = (X_n, \xi_n) \tag{8.9}$$

for some sequence $(\xi_n)_0^\infty$ such that $\tilde{X} = (\tilde{X}_n)_0^\infty$ is Markovian. If so, the asymptotics of X may be analyzed via that of $\tilde{X}$.

Since there are several thorough accounts of the steps taken so far in this section, we have been brief.

9. Harris chains. We call $X = (X_n)_0^\infty$ a Harris chain if there exist a set $A \in \mathscr{E}$, a probability measure φ on $(E, \mathscr{E})$, a number β, $0 < \beta < 1$, and an integer $n_0 \geq 1$ such that

(9.1) (i) $\mathbf{P}_x(X_n \in A \text{ for some } n \geq 1) = 1$ for all $x \in E$, and
(ii) $\mathbf{P}_x(X_{n_0} \in B) \geq \beta \cdot \varphi(B)$ for all $x \in A$.

Condition (i) says that A is a recurrent set of the chain. For simplicity we will assume that $n_0 = 1$. Below, we will make precise the following way to interpret (ii): When X hits A, at time n, say, then with probability β we may let X_{n+1} be distributed according to φ regardless of the value taken by X_n. The entire section leans on that uniformity.

The definitions in the literature of what is meant by a Harris chain differ; we have chosen the one that most directly serves our purposes. We need at our disposal a sequence of i.i.d. 0–1 variables $I_0, I_1, \ldots$, with $\mathbf{P}(I_n = 1) = \beta$, to indicate when the φ distribution shall be chosen for a next state of the chain X. To enliven things, let us call them bell variables; if $X_n = x \in A$ and the bell rings ($I_n = 1$), we shall let X_{n+1} have distribution φ, otherwise distribution $Q(x, \cdot)$, where

$$Q(x, \cdot) = (P(x, \cdot) - \beta \cdot \varphi)/(1 - \beta).$$

We achieve this by letting X_n be the first component of $\tilde{X}_n$, where $\tilde{X} = (\tilde{X}_n)_0^\infty$ is a Markov chain with state space $\tilde{E} = E \times \{0, 1\}$, equipped with the σ-field $\tilde{\mathscr{E}} = \mathscr{E} \times \mathscr{P}(\{0, 1\})$ and, rather naturally, governed by the Markov kernel $\tilde{P}$ defined through

$$\begin{aligned}(9.2)\quad &\tilde{P}((x,1), B \times \{\delta\}) = (\beta \cdot \delta + (1-\beta)\cdot(1-\delta)) \cdot \varphi(B),\\ &\tilde{P}((x,0), B \times \{\delta\}) = (\beta \cdot \delta + (1-\beta)\cdot(1-\delta)) \cdot Q(x, B)\end{aligned}$$

for $\delta = 0$ or 1 if $x \in A$, and

$$\tilde{P}((x,\delta'), B \times \{\delta\}) = (\beta \cdot \delta + (1-\beta)\cdot(1-\delta)) \cdot P(x, B)$$

for δ, $\delta' = 0$ or 1 if $x \notin A$. If $\mathbf{P}_\lambda$ is the probability measure on $(\tilde{E}^\infty, \tilde{\mathscr{E}}^\infty)$ for the canonical Markov chain $\tilde{X}_n = (X_n, I_n)$, $n \geq 0$, with initial distribution $\nu = \lambda \times ((1-\beta)\cdot\delta_0 + \beta\cdot\delta_1)$, then indeed

(9.3) (i) $(X_n)_0^\infty$ is a Markov chain with transition probabilities $P(x, B)$ and initial distribution λ,

(ii) $I_0, I_1, \ldots,$ are i.i.d. 0–1 variables with $\mathbf{P}_\lambda(I_n = 1) = \beta$, and

(iii) $(X_0, \ldots, X_n)$ and I_n are independent for each $n \geq 0$.

We let $\sigma(\tilde{X}_0, \ldots, \tilde{X}_n)$ be denoted by $\tilde{\mathscr{F}}_n$.

10. Regeneration and stationarity. Here is the key to what follows.

(10.1) Lemma. *There exists a random time $\tau \geq 1$ such that*

(i) *τ is a previsible stopping time w.r.t. the filtration $(\tilde{\mathscr{F}}_n)_0^\infty$, that is, $\{\tau = n\} \in \tilde{\mathscr{F}}_{n-1}$ for all $n \geq 1$,*

(ii) *τ is finite a.s., and*

(iii) *$\mathbf{P}_\lambda(X_n \in B, \tau = n) = \varphi(B)\cdot\mathbf{P}_\lambda(\tau = n)$ for all $n \geq 1$.*

Proof. Let $\eta = \min\{n \geq 0;\, X_n \in A, I_n = 1\}$, and $\tau = \eta + 1$. Certainly, $\{\eta = n\} \in \tilde{\mathscr{F}}_n$; hence (i) is proved. Taking the finiteness of τ for granted, we obtain

$$\begin{aligned}
\mathbf{P}_\lambda(X_n \in B, \tau = n) &= \mathbf{P}_\lambda(X_n \in B, \eta = n-1) \\
&= \int_A \tilde{P}((x,1), B \times \{0,1\})\cdot\mathbf{P}_\lambda(\eta = n-1, X_{n-1} \in dx) \\
&= \varphi(B)\cdot\mathbf{P}_\lambda(\eta = n-1) = \varphi(B)\cdot\mathbf{P}_\lambda(\tau = n)\,,
\end{aligned}$$

which follows from (9.2) and (9.3)(iii), and (iii) is proved. Notice that (iii) implies that X_τ has distribution φ, and that X_τ and τ are independent.

To prove (ii), put

$$\kappa_0 = \min\{j \geq 0;\, X_j \in A\}$$

and

$$\kappa_n = \min\{j > \kappa_{n-1}; X_j \in A\}.$$

We have

$$\{\eta = \infty\} = \{I_{\kappa_n} = 0 \text{ for all } n\}.$$

A moment's thought, using (9.3)(iii) renders $\mathbf{P}_\lambda(I_{\kappa_n} = 0 \text{ for } n \le k) \le (1-\beta)^k$, hence $\eta < \infty$ a.s. and so is τ. □

Notice that η and τ are stopping times for $(\tilde{X}_n)_0^\infty$ but not for $(X_n)_0^\infty$. However, τ is a typical example of what is called a randomized stopping time for $(X_n)_0^\infty$, that is, a random time $\xi \in \bar{\mathbb{Z}}_+$ such that

$$(10.2)\quad P(\xi > n \| \mathcal{F}) = P(\xi > n \| \mathcal{F}_n) \qquad \text{for all} \quad n \ge 0 \text{ under } \mathbf{P}_\lambda.$$

But that property is easily seen to be at hand when $\{\xi > n\} \in \sigma\{X_0, X_1, \ldots, X_n, Y\}$, where Y is a random element of the type $(I_n)_0^\infty$. Definition (10.2) was adjusted to the present setting.

The lemma says that we may stop a Harris chain appropriately so that X_τ gets a desired distribution; let us call such a randomized stopping time a conforming time, and say that X conforms to φ at time τ.

Regenerations of a Harris chain take place at times $(S_n)_0^\infty$, defined as follows: Let

$$\eta_0 = \min\{k \ge 0; X_k \in A, I_k = 1\},$$

$$S_0 = \begin{cases} 0, & \text{if } \lambda = \varphi \\ \eta_0 + 1, & \text{otherwise,} \end{cases}$$

$$\eta_n = \min\{k \ge S_{n-1}; X_k \in A, I_k = 1\},$$

$$S_n = \eta_n + 1$$

for $n \ge 1$. Due to the lemma, we have $X_{S_n} \overset{\mathcal{D}}{=} \varphi$ for all n, and the strong Markov property of $\tilde{X}$, used repeatedly with the stopping times $S_0, S_1, \ldots$, yields that $S = (S_n)_0^\infty$ is a renewal process. The rationale for the awkward definition of S_0 is that we prefer to be consistent with notation introduced earlier; if $\lambda = \varphi$, it is natural to let S be zero-delayed.

Let $\lambda = \varphi$, and recall τ from Lemma (10.1); τ equals the first regeneration time ≥ 1 of S. Put

$$(10.3) \qquad \nu(B) = \mathbf{E}_\varphi\left[\sum_0^{\tau-1} I(X_i \in B)\right]$$

for $B \in \mathscr{E}$; it is a sound conjecture that ν is an invariant measure (normalized, it becomes a stationary distribution when ν is finite), that is,

$$(10.4) \qquad \nu = \nu P = \int \nu(x) P(x, \cdot)$$

[we recalled (8.2)]. Notice that $Pf(x) = \mathbf{E}[f(X_{i+1}) \| X_i = x]$. To prove (10.4), assume that $\mathbf{E}_\varphi[\tau] < \infty$, which of course is necessary and sufficient for ν to be finite. We have

$$\mathbf{E}_\varphi\left[\sum_0^{\tau} I(X_i \in \cdot)\right] = \varphi + \mathbf{E}_\varphi\left[\sum_1^{\tau} I(X_i \in \cdot)\right].$$

But that expectation also equals $\nu + \mathbf{E}_\varphi[I(X_\tau \in \cdot)] = \nu + \varphi$, hence $\nu = \nu P$ follows if we can establish that

$$(10.5) \qquad \mathbf{E}_\varphi\left[\sum_1^{\tau} I(X_i \in \cdot)\right] = \nu P.$$

To prove that, we shall have use of the fact that for $i \geq 0$,

$$(10.6) \quad \mathbf{E}_\varphi[f(X_{i+1}) \cdot I(\tau \geq i+1)] = \mathbf{E}_\varphi[Pf(X_i) \cdot I(\tau \geq i+1)]$$

for $f \in b\mathscr{E}$. Indeed,

$$\begin{aligned} &\mathbf{E}_\varphi[f(X_{i+1}) \cdot I(\tau \geq i+1)] \\ &\quad = \mathbf{E}_\varphi[f(X_{i+1})] - \sum_{j=1}^{i} \mathbf{E}_\varphi[f(X_{i+1}) \cdot I(\tau = j)], \end{aligned}$$

and (10.6) follows after conditioning on τ. With $f = I_B$,

$$\begin{aligned} \mathbf{E}_\varphi\left[\sum_1^{\tau} I(X_i \in B)\right] &= \mathbf{E}_\varphi\left[\sum_1^{\infty} f(X_i) \cdot I(\tau \geq i)\right] \\ &= \mathbf{E}_\varphi\left[\sum_0^{\infty} f(X_{i+1}) \cdot I(\tau \geq i+1)\right] \\ &= \mathbf{E}_\varphi\left[\sum_0^{\infty} Pf(X_i) \cdot I(\tau \geq i+1)\right] = \nu(Pf). \end{aligned}$$

Hence we have proved that

$$\mathbf{E}_\varphi\Big[\sum_1^\tau I(X_i \in B)\Big] = \nu[Pf] = [\nu P](B)$$

for every $B \in \mathscr{E}$, and (10.5) is established.

With

$$(10.7) \qquad \pi(B) = \nu(B)/\nu(E) \quad \text{for} \quad B \in \mathscr{E}\,,$$

π is a stationary distribution. The proof of the invariance of ν also turns out to be valid for ν infinite, so our precaution of assuming ν to be finite can be relaxed.

To show that ν is always σ-finite, let

$$E_{nm} = \{x; \mathbf{P}_x(\tau \le n) \ge 1/m\}$$

for n, $m \ge 1$. A "geometric number of trials" argument shows that $\nu(E_{nm}) < \infty$, and $E = \bigcup_{n,m} E_{nm}$.

Uniqueness of the stationary distribution π is a consequence of the ergodic result below. When ν is infinite, it is the unique invariant measure (up to a multiplicative constant). Since that is a topic of no relevance to us, we content ourselves with a reference in § 13.

11. Ergodicity. It is now assumed that P is such that the embedded renewal process S is aperiodic. In addition to the process X, introduce a parallel independent process $X' = (X'_n)_0^\infty$ governed by P and with initial distribution π [cf. (10.4)]. With S' denoting its renewal process, we get a weak coupling at T = the first common renewal time, and have

$$(11.1) \qquad \|\mathbf{P}_\lambda(X_n \in \cdot\,) - \pi\| \le 2 \cdot \mathbf{P}_{\lambda\pi}(T > n)\,.$$

There remains little to do to establish the following result.

(11.2) Theorem. *The stationarity distribution π is unique, and*

$$\|\mathbf{P}_\lambda(X_n \in \cdot\,) - \pi\| = \|\lambda P^n - \pi\| \to 0 \qquad \text{as} \quad n \to \infty$$

for every initial distribution λ.

Proof. The embedded renewal process S and S' are aperiodic and have a finite mean recurrence time. We know from (II.2.3) that T is finite, and the ergodic result follows from (11.1). For the uniqueness, suppose that π^* is another stationary distribution, and let $\lambda = \pi^*$. We get

$$0 \neq \|\pi^* - \pi\| = \|\mathbf{P}_{\pi^*}(X_n \in \cdot\,) - \pi\| \to 0\,,$$

a contradiction. □

The construction of a strong coupling is not difficult. We wait for X and X' to be in A at the same time; if both bells ring then, we let X and X' coincide (with distribution φ, of course) in the next step. If not, we .... But all that is unnecessary!

We shall avoid repeating the rate results for discrete Markov chains. As an example of what can be obtained for a Harris chain, suppose we know that $\mathbf{E}_\varphi[\tau^\alpha] < \infty$ for an $\alpha > 1$. Then

$$\|P^n(x, \cdot\,) - \pi\| = o(n^{-(\alpha-1)})\,. \tag{11.3}$$

To achieve that, use $\mathbf{E}_{x,\pi}[T^{\alpha-1}] < \infty$ and (11.1) in the obvious way. For the extension of (II.8.6) with an $\alpha \geq 1$, we need the result that $b\mathcal{M}_s(E, \mathscr{E})$ is a Banach space, which is proved in § App.2.

There does exist a coupling proof of Orey's theorem for the null-recurrent case, but it is rather a specimen of mastership than a demonstration of the strength of the method, because of the technical complications. Hence there is good reason to examine a competing proof, using the theory to be presented in § 21. Remember that recurrence is a part of the definition of a Harris chain.

(11.4) Theorem. *For any two initial distributions λ and μ it holds that*

$$\|\lambda P^n - \mu P^n\| \to 0 \qquad as \quad n \to \infty\,.$$

Proof. Let f be a bounded space-time harmonic function, with domain $E \times \mathbb{Z}_+$:

$$f(x, n) = \int f(y, n+1) P(x, dy)$$

for all $(x, n) \in E \times \mathbb{Z}_+$. Fix $k \in \mathbb{Z}_+$, and define Z_n, $n \geq 0$, through $Z_n = f(X_n, k+n)$. That sequence is a bounded martingale; hence the optional sampling theorem renders

$$\begin{aligned} (11.5) \qquad f(x, k) &= \mathbf{E}_x[Z_0] = \mathbf{E}_x[Z_\tau] = \mathbf{E}_x[f(X_\tau, k+\tau)] \\ &= \sum g(k+j) \cdot \mathbf{P}_x(\tau = j) = \mathbf{E}_x[g(k+\tau)], \end{aligned}$$

where $g(i) = \mathbf{E}_\varphi[f(X_0, i)]$. Integrating (11.5) with respect to φ gives

$$g(k) = \mathbf{E}_\varphi[g(k+\tau)]$$

for all $k \geq 0$. But since g is bounded and τ has an aperiodic distribution under $\mathbf{P}_\varphi$, g is forced to be constant, as is well known. That constancy is equivalent to $\|\lambda P^n - \mu P^n\| \to 0$, as we shall see in § 21. □

12. Random walk. Now let μ be a probability on $(\mathbb{R}, \mathcal{R})$ having an absolutely continuous component with respect to Lebesgue measure. We aim to extend (II.12.1) to a continuous result:

$$(12.1) \qquad \|\delta_z * \mu^{*n} - \mu^{*n}\| \to 0 \quad \textit{as} \quad n \to \infty$$

for all $z \in \mathbb{R}$. Since μ^{*2} has a continuous component [cf. the proof of Lemma (5.1)], it is innocent to assume that μ has one. Now recall the Ornstein coupling from § II.12. Let $S_n^* = (S_n, S_n')$, $n \geq 0$ with $S_0 = z$, $S_0' = 0$. The random walk $(S_n - S_n')_0^\infty$ is recurrent in the (weak) sense that $\mathbf{P}(S_n - S_n' \in I$ for some $n \geq 0) = 1$ for any open interval I. However, that is not good enough for a coupling proof of (12.1): "Exact" recurrence is needed. To that end, let P^* be the Markov kernel of S^*, considered as a Markov chain in $\mathbb{R}^2$, and let $\epsilon > 0$ be so small that

$$\gamma = \inf_{|x-x'| \leq \epsilon} \int f_0(s) \wedge f_0(s - (x - x'))\, ds > 0,$$

where f_0 is a continuous component of μ. For (x, x') satisfying $|x - x'| \leq \epsilon$, let P^{**} be a Markov kernel such that for $B \in \mathcal{R}$,

$$P^{**}((x, x'), B \times \mathbf{R}) = \mu(B - x),$$
$$P^{**}((x, x'), \mathbf{R} \times B) = \mu(B - x'), \quad \text{and}$$
$$P^{**}((x, x'), \Delta) \geq \gamma .$$

That may be constructed along the lines of the proof of Lemma (5.2). Now let $\tilde{P}$ equal P^* if $|x - x'| > \epsilon$, and P^{**} if $|x - x'| \leq \epsilon$. Certainly, $\tilde{P}$ governs a Markov chain $(\tilde{S}_n, \tilde{S}'_n)$, $n \geq 0$, where $\tilde{S}_n$ and $\tilde{S}'_n$, $n \geq 0$, are random walks with step size distribution μ. And

(12.2) $\mathbf{P}_{z0}((\tilde{S}_n, \tilde{S}'_n) \in \Delta \text{ for some } n) = 1 \qquad \text{for all} \quad z \in \mathbb{R}$;

a "geometric number of trials" argument is used to prove that. Of course, (12.2) implies (12.1).

13. Notes. For a thorough treatment of general state space Markov chains in discrete time, the reader is referred to Revuz [134]. The term "Harris chain" is so well established that we find it fruitless to try to introduce "Doeblin–Harris chains"; that would also give due credit to the pioneer. The study of Harris chains took a dramatic step forward in the mid 1970s due to work by Griffeath [67] and the Ph.D. thesis on which that is based, Athreya and Ney [13], and Nummelin [123]; see also Nummelin's book [124] and its bibliography. We follow [13] rather closely in §§ 2, 3 and 5. However, there is an important difference: In [13], no parallel process is introduced, and a discrete renewal equation turns out to play a key role for the proof of Theorem (11.2); we find it natural to apply our results from § II.2.

The proof that ν of (10.1) is an invariant measure is essentially that of Pitman [130]. Learn more about randomized stopping times from that paper and from Pitman and Speed [131]. For ν infinite, the uniqueness up to a multiplicative constant is not trivial to prove; see [10]. For applications of Harris chains to queueing theory, see Sigman [142]. The proof of (12.1) is close to that of Griffeath [67]. That paper also contains a coupling proof of Orey's theorem. For coupling of random sequences that remember more of the past than Markov chains, see Harris [71], Kaiser [80], and Berbee [25] (infinite memory, chains with complete connection), and Brandt and Künsch [28] (an example from time-series analysis).

3. MAXIMAL COUPLING

14. The coupling. Goldstein's theorem. Recall the basic coupling inequality [(I.2.6) and (I.4.7)]: For any coupling $\hat{X} = (\hat{X}_n)_0^\infty$, $\hat{X}' = (\hat{X}'_n)_0^\infty$ we have

$$\|\mathbf{P}(\theta_n X \in \cdot\,) - \mathbf{P}(\theta_n X' \in \cdot\,)\| \le 2 \cdot \mathbf{P}(T > n)\,.$$

The purpose of this section is to show that equality can be achieved: There exists a coupling such that

$$\|\mathbf{P}(\theta_n X \in \cdot\,) - \mathbf{P}(\theta_n X' \in \cdot\,)\| = 2 \cdot \mathbf{P}(T > n) \tag{14.1}$$

for all $n \ge 0$. This is interesting: The existence of a successful coupling is not only sufficient to get that X and X' are asymptotically equally distributed, but actually equivalent to that. We call a coupling satisfying (14.1) a maximal coupling.

We first obtain (14.1) for a weak coupling. But to each coupling there corresponds a strong one with the same coupling time distribution, as will be proven in the next section. Hence (14.1) holds for a strong coupling. To realize the weak coupling, we shall apply repeatedly the idea underlying the γ coupling of § I.5; notice from the construction there that for any subprobabilities μ and μ' there exists a greatest minorization of μ and μ', to be denoted by $\mu \wedge \mu'$, namely the subprobability with density $g \wedge g'$ with respect to $\lambda = \mu + \mu'$, where $g = d\mu/d\lambda$ and $g' = d\mu'/d\lambda$. If μ and μ' are probability measures, then

$$\|\mu - \mu'\| = 2 \cdot (1 - \|\mu \wedge \mu'\|)\,. \tag{14.2}$$

Of course, $\|\mu \wedge \mu'\|$ is the total mass of $\mu \wedge \mu'$.

Our sequences X and X' have $(E^\infty, \mathscr{E}^\infty)$ as state space. For $n \ge 0$, let $\mathscr{T}_n$ be the σ-field $\subset \mathscr{E}^\infty$ defined by

$$\mathscr{T}_n = \theta_n^{-1}(\mathscr{E}^\infty) \tag{14.3}$$

while

$$\mathscr{T} = \bigcap_{n=0}^{\infty} \mathscr{T}_n$$

is called the tail σ-field. Notice that $\mathcal{T}_0 = \mathcal{E}^\infty$. For a measure μ on $(E^\infty, \mathcal{E}^\infty)$, let $\mu_{(n)}$ be its restriction to $(E^\infty, \mathcal{T}_n)$. Let P and P' denote the distributions of X and X'. We have

$$\begin{aligned}\|\mathbf{P}(\theta_n \hat{X} \in \cdot\,) - \mathbf{P}(\theta_n \hat{X}' \in \cdot\,)\| &= \|P_{(n)} - P'_{(n)}\| \\ &= 2 \cdot (1 - \|P_{(n)} \wedge P'_{(n)}\|)\end{aligned}$$

for any coupling $\hat{X}, \hat{X}'$ of X and X' due to (14.2). Hence if we can produce a weak coupling such that

$$\|P_{(n)} \wedge P'_{(n)}\| = \mathbf{P}(T \le n) \tag{14.4}$$

for all $n \ge 0$, we are done. But rather a lot of subtlety is needed for that, so a few words about the proof might be due. We shall construct certain subprobability measures $\tilde{\mu}_n$ and $\tilde{\mu}'_n$ on $(E^\infty, \mathcal{E}^\infty)$ for each $n \ge 0$ to be such that

$$\mathbf{P}(X \in A, T = n) = \tilde{\mu}_n(A) \quad \text{and} \quad \mathbf{P}(X' \in A, T' = n) = \tilde{\mu}'_n(A)\,.$$

That is achieved when we have reached the crucial (14.7). Some technicalities are postponed to the end of the proof.

As the first steps, let $\mu_0, \mu_1, \ldots$ be the subprobabilities on $\mathcal{T}_0$, $\mathcal{T}_1, \ldots$, defined by

$$\mu_0 = P \wedge P'$$

and

$$\mu_n = P_{(n)} \wedge P'_{(n)} - (P_{(n-1)} \wedge P'_{(n-1)})_{(n)}$$

for $n \ge 1$. Let $\mu'_n = \mu_n$. Notice that

$$\sum_0^n \mu_i = \sum_0^n \mu'_i = P_{(n)} \wedge P'_{(n)} \quad \text{on } \mathcal{T}_n \quad \text{for each} \quad n \ge 0\,. \tag{14.5}$$

Now suppose that we can find subprobabilities $\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_\infty$ and $\tilde{\mu}'_0, \tilde{\mu}'_1, \ldots, \tilde{\mu}'_\infty$ on $(E^\infty, \mathcal{E}^\infty)$ such that

$$(14.6) \qquad \tilde{\mu}_0 + \cdots + \tilde{\mu}_\infty = P, \quad \tilde{\mu}'_0 + \cdots + \tilde{\mu}'_\infty = P'$$

and

$$\tilde{\mu}_n, \tilde{\mu}'_n \text{ are extensions of } \mu_n, \mu'_n \text{ from } \mathcal{T}_n \text{ to } \mathscr{E}^\infty \text{ for all } n < \infty .$$

We may then define probabilities Q, Q' on $(E^\infty \times \bar{\mathbb{Z}}_+, \mathscr{E}^\infty \times \mathscr{Z}^\infty_+)$ by

$$(14.7) \qquad Q(A \times \{n\}) = \tilde{\mu}_n(A), \qquad Q'(A \times \{n\}) = \tilde{\mu}'_n(A)$$

for $n \in \bar{\mathbb{Z}}_+$, and random elements $(\hat{X}, T)$, $(\hat{X}', T')$ with distributions Q and Q', respectively, constitute a maximal weak coupling of X and X'. Indeed, $X \stackrel{\mathcal{D}}{=} \hat{X}$ and $X' \stackrel{\mathcal{D}}{=} \hat{X}'$ follow from (14.6) and (14.7), and $(\theta_T \hat{X}, T) \stackrel{\mathcal{D}}{=} (\theta_{T'} \hat{X}', T')$ from (14.7); recall the convention that $\theta_\infty x =$ a fixed element $(z, z, \ldots)$ for all $x \in E^\infty$. And (14.4) is satisfied due to (14.5) since

$$\mathbf{P}(T \le n) = Q(E^\infty \times \{0, 1, \ldots, n\}) = \sum_0^n \tilde{\mu}_i(E^\infty) = \sum_0^n \mu_i(E^\infty)$$
$$= (P_{(n)} \wedge P'_{(n)})(E^\infty) = \|P_{(n)} \wedge P'_{(n)}\| .$$

An induction argument will be used to establish (14.6). Suppose that we have been able to find extensions $\tilde{\mu}_k$ of μ_k for all $k \le n - 1$ in such a way that $\Sigma_0^{n-1} \tilde{\mu}_k \le P$; obviously, the case $n = 1$ causes no problem. Let $\lambda_n = P - \Sigma_0^{n-1} \tilde{\mu}_k$. Now there exists a kernel K_n in $E^\infty \times E^\infty$ w.r.t. $\mathcal{T}_n$ and $\mathcal{T}_0$ such that

$$(14.8) \qquad \lambda_n(A \cap B) = \int_B \lambda_n(dx) K_n(x, A)$$

for all $A \in \mathcal{T}_0$ and $B \in \mathcal{T}_n$ (you may wish to normalize λ_n to a probability measure and think of K_n as a conditional probability to get the right feeling for that). Define $\tilde{\mu}_n$ by

$$(14.9) \qquad \tilde{\mu}_n(A) = \int \mu_n(dx) K_n(x, A) .$$

Certainly $\tilde{\mu}_n$ is an extension of μ_n from $\mathcal{T}_n$ to $\mathscr{E}^\infty$. Further, (14.5) implies that $\mu_n \le \lambda_n$ on $\mathcal{T}_n$. Using this it follows from (14.9) that

$$\tilde{\mu}_n(A) \le \int \lambda_n(dx) K_n(x, A) = \lambda_n(A),$$

hence

$$\sum_0^n \tilde{\mu}_i \le P.$$

With analogous arguments to get $\tilde{\mu}'_n$, and defining $\tilde{\mu}_\infty$ and $\tilde{\mu}'_\infty$ to be $P - \Sigma_0^\infty \tilde{\mu}_i$ and $P' - \Sigma_0^\infty \tilde{\mu}'_i$, respectively, we are done. Notice that $\tilde{\mu}_n = \tilde{\mu}'_n$ on $\mathcal{T}_n$ for $n \in \mathbb{Z}_+$. □

Possibly you are exhausted, but it is a beautiful and deep result we have proved! We are now ready to state and prove a main result, Goldstein's theorem.

(14.10) Theorem. *For random sequences $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ with distributions P and P' on $(E^\infty, \mathcal{E}^\infty)$, the following are equivalent*:

(i) *There exists a successful coupling of X and X'*,
(ii) *P and P' agree on $\mathcal{T}$, and*
(iii) $\|\mathbf{P}(\theta_n X \in \cdot) - \mathbf{P}(\theta_n X' \in \cdot)\| = \|P_{(n)} - P'_{(n)}\| \to 0$ *as* $n \to \infty$.

Proof. The implication (i)⇒(iii) is a consequence of the basic coupling inequality, and (iii)⇒(i) of the existence of a maximal coupling. That (iii)⇒(ii) follows from the inequality

$$\|P_{(\infty)} - P'_{(\infty)}\| \le \|P_{(n)} - P'_{(n)}\|,$$

where $P_{(\infty)}$, $P'_{(\infty)}$ are the restrictions of P, P' to $\mathcal{T}$. It remains to prove (ii)⇒(iii), which is not as easy as you may think. Let $2\alpha = \lim_{n\to\infty} \|P_{(n)} - P'_{(n)}\|$ and pick a sequence $A_n \in \mathcal{T}_n$ such that $\lim_{n\to\infty} (P(A_n) - P'(A_n)) = \alpha$. Now introduce the Hilbert space

$$\mathcal{H} = L^2(E^\infty, \mathcal{E}^\infty, P + P')$$

and let $(\cdot, \cdot)$ denote the associated inner product. Recall that in a Hilbert space

(14.11) every bounded sequence has a weakly convergent subsequence ($f_n \to f$ weakly if $(f_n, g) \to (f, g)$ for all $g \in \mathscr{H}$).

Let $f_n = I_{A_n}$, $g = dP/d(P+P')$, and $g' = dP'/d(P+P')$. The sequence $(f_n)_0^\infty$ is bounded; hence $f_{k_n} \to$ some f_∞ weakly for a subsequence $(f_{k_n})_0^\infty$. Now the linear functionals $f \to \int f\, dP = (f, g)$ and $f \to \int f\, dP' = (f, g')$ are certainly bounded; hence due to (14.11) we obtain

$$\begin{aligned}\int f_\infty\, dP - \int f_\infty\, dP' &= \lim_{n\to\infty} ((f_{k_n}, g) - (f_{k_n}, g')) \\ &= \lim_{n\to\infty} \left(\int f_{k_n}\, dP - \int f_{k_n}\, dP'\right) \\ &= \lim_{n\to\infty} (P(A_{k_n}) - P'(A_{k_n})) = \alpha, \quad \text{say.}\end{aligned}$$

But since each $L^2(\mathscr{E}^\infty, \mathscr{T}_n, P + P')$ is weakly closed, we find that $f_\infty \in \mathscr{T}_n$ for all n, hence $f_\infty \in \mathscr{T}$. Now P and P' agree on, $\mathscr{T}$, so $\int f_\infty\, dP = \int f_\infty\, dP'$ and $\alpha = 0$ is established.

In fact, a few more arguments render that we always have $\|P_{(\infty)} - P'_{(\infty)}\| = \lim_{n\to\infty} \|P_{(n)} - P'_{(n)}\|$, even if P and P' do not agree on $\mathscr{F}$. □

Hence the basic result (14.11) from Hilbert space theory was instrumental in the proof of (ii)$\Rightarrow$(iii).

This is an appropriate place for some definitions. For a probability space $(S, \mathscr{S}, Q)$, we say that $\mathscr{S}$ is trivial (Q) if

$$Q(A) = 0 \text{ or } 1 \quad \text{for all} \quad A \in \mathscr{S}. \tag{14.12}$$

When clear which Q is referred to, we say that $\mathscr{S}$ is trivial. Notice that (14.12) is equivalent to

(14.13) all functions $f \in b\mathscr{S}$ are constant a.s. (Q), that is, there exists a constant c such that $Q(\{y; f(y) = c\}) = 1$.

Of course, that constant equals $\int f(y)\, dQ(y)$.

Recall the tail σ-field $\mathcal{T}$ of $\mathscr{E}^\infty$ defined by (14.3). For a random sequence $X = (X_n)_0^\infty$ defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and with values in $(E^\infty, \mathscr{E}^\infty)$,

$$\mathcal{T}(X) = X^{-1}(\mathcal{T}) \tag{14.14}$$

is the tail σ-field of X; it is a sub-σ-field of $\mathcal{F}$ (yes, we are using $\mathcal{T}$ here with different meanings, but that is harmless). If $\mathcal{T}(X)$ is trivial $(\mathbf{P})$, we say that X has a trivial tail σ-field $(\mathbf{P})$, or shorter, that X has a trivial tail.

We say that a random sequence $X = (X_n)_0^\infty$ defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with distribution $P = \mathbf{P}X^{-1}$ on $(E^\infty, \mathscr{E}^\infty)$, is mixing if for all $B \in \mathscr{E}^\infty$ we have

$$\lim_{n\to\infty} \sup_{A \in \mathcal{T}_n} |P(A \cap B) - P(A)\cdot P(B)| = 0\,. \tag{14.15}$$

Mixing and tail triviality turn out to be the same thing.

(14.16) Theorem. *A random sequence is mixing if and only if its tail σ-field is trivial.*

Proof. For (14.15)$\Rightarrow$(14.12), let $B = A \in \mathcal{T}$. We get

$$\begin{aligned} |P(A) - P(A)^2| &= |P(A \cap B) - P(A)\cdot P(B)| \\ &\le \sup_{A' \in \mathcal{T}_n} |P(A' \cap B) - P(A')\cdot P(B)| \to 0 \end{aligned}$$

as $n \to \infty$, hence $P(A) = P(A)^2$, and (14.12) follows.

To prove the converse, take a B such that $P(B) > 0$ and let $P' = P(\cdot \cap B)/P(B)$. Certainly, P and P' agree on $\mathcal{T}$, hence from (14.10) [(ii)$\Rightarrow$(iii)] we get

$$\|P_{(n)} - P'_{(n)}\| = 2\cdot \sup_{A \in \mathcal{T}_n} |P(A) - P(A \cap B)/P(B)| \to 0\,,$$

and (14.15) is set. $\square$

Theorems (14.10) and (14.16) hold also for continuous-time processes. To prove that, it suffices to consider (14.10). For the σ-field $\mathcal{D}_E$ on D_E, let

(14.17) $$\mathcal{T}_t = \theta_t^{-1}\mathcal{D}_E \quad \text{for} \quad t \geq 0$$

and

$$\mathcal{T} = \bigcap_{t\geq 0} \mathcal{T}_t .$$

For random functions X and X' in $(D_E, \mathcal{D}_E)$ with distributions $P = \mathbf{P}X^{-1}$ and $P' = \mathbf{P}X'^{-1}$, let $P_{(t)}$, $P'_{(t)}$ be the restrictions of P, P' to $\mathcal{T}_t$. For a proof of Theorem (14.10) in this new setting, we find that only (iii) $\Rightarrow$ (i) need attention. Here is the trick to do that: Introduce the sequences $Y = (Y_n)_0^\infty$ and $Y' = (Y'_n)_0^\infty$ in $(D_E^\infty, \mathcal{D}_E^\infty)$, where

(14.18) $$Y_n = \theta_n X, \qquad Y'_n = \theta_n X' .$$

Now suppose that $\|P_{(n)} - P'_{(n)}\| \to 0$. Use the implication (iii) $\Rightarrow$ (i) of Theorem (14.10) as it stands to produce a successful strong coupling $((\hat{Y}_n)_0^\infty, (\hat{Y}'_n)_0^\infty)$ of $Y = (Y_n)_0^\infty$ and $Y' = (Y'_n)_0^\infty$. This yields that $\hat{Y}_n = \hat{Y}'_n$ for n sufficiently large. For a coupling of X and X', simply let $\hat{X} = \hat{Y}_0$ and $\hat{X}' = \hat{Y}'_0$. Since $\hat{Y}'_n = \theta_n \hat{Y}_0$ and $\hat{Y}'_n = \theta_n \hat{Y}'_0$ and $\hat{Y}'_n = \theta_n \hat{Y}'_0$ we certainly have that $\hat{X}_t = \hat{X}'_t$ for t large enough.

15. From weak to strong coupling. Let $(\tilde{X}, T)$, $(\tilde{X}', T')$, defined on spaces with probability measures $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{P}}'$, be a weak coupling of $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$. To produce a strong coupling at least as efficient as X and X', we shall, virtually, let $\tilde{X}$ be as it is and adjust $\tilde{X}'$ to $\tilde{X}$ in a suitable way. To that end, notice that due to the Polish assumption there exists a regular version of the conditional probabilities $\mathbf{P}(\tilde{X}' \in \cdot \,\|(\theta_{T'}\tilde{X}', T') = (y, j))$, that is, a particular transition kernel in $(E^\infty \times \bar{\mathbb{Z}}_+) \times E^\infty$, which we denote by K'. We have

(15.1) $$K'(((x_i)_0^\infty, j), \{(y_i)_0^\infty; (y_{j+i})_0^\infty = (x_i)_0^\infty\}) = 1$$

for all $(x_i)_0^\infty$ and j.

Now let $\hat{\mathbf{P}}$ be the probability on $E^\infty \times E^\infty \times \bar{\mathbb{Z}}_+$ defined by

(15.2) $$\hat{\mathbf{P}}(A \times B \times \{i\}) = \int I_i(j) \cdot I_A(x) \cdot K'((\theta_j x, j), B)\, dQ((x, j))$$

for rectangles $A \times B \times \{i\}$, where Q is the distribution of (X, T).

A triple $(\hat{X}, \hat{X}', \hat{T})$ with distribution $\hat{\mathbf{P}}$ is a strong coupling of X and X', with a coupling time $\hat{T}$ distributed as T and T'. Indeed, we have $\hat{X} \stackrel{\mathcal{D}}{=} X$ because

$$\hat{\mathbf{P}}(A \times E^{\infty} \times \bar{\mathbb{Z}}_{+}) = \int I_A(x)\, dQ((x, j)) = \mathbf{P}(X \in A).$$

Further,

$$\begin{aligned}\hat{\mathbf{P}}(E^{\infty} \times B \times \bar{\mathbb{Z}}_{+}) &= \int K'((\theta_j x, j), B)\, dQ((x, j)) \\ &= \int K'((\theta_j x, j), B) \cdot \tilde{\mathbf{P}}((X, T) \in d(x, j)) \\ &= \int K'((y, j), B) \cdot \tilde{\mathbf{P}}((\theta_T X, T) \in d(y, j)) \\ &= \int K'((y, j), B) \tilde{\mathbf{P}}'((\theta_T X', T') \in d(y, j)) \\ &= \tilde{\mathbf{P}}'(X' \in B).\end{aligned}$$

We allowed ourselves the notation $\mathbf{P}(W \in dw)$ for the distribution of a random element W. Notice how we made use of $(\theta_T \hat{X}, T) \stackrel{\mathcal{D}}{=} (\theta_{T'} \tilde{X}', T')$. Also, for any $i \in \bar{\mathbb{Z}}_{+}$,

$$\begin{aligned}&\hat{\mathbf{P}}(\{(x, y, i); x_n = y_n \quad \text{for} \quad n \geq i\}) \\ &= \hat{\mathbf{P}}(E^{\infty} \times E^{\infty} \times \{i\}) = \tilde{\mathbf{P}}(T = i)\end{aligned}$$

due to (15.1), and we are done.

For a shorter proof, let H be the common distribution of $(\theta_T \tilde{X}, T)$ and $(\theta_{T'} \tilde{X}', T')$ on $(E^{\infty} \times \bar{\mathbb{Z}}_{+}, \mathscr{E}^{\infty} \times \bar{\mathscr{Z}}_{+})$ and let K be a kernel for a regular version of $\mathbf{P}(X \in \cdot \,\|(\theta_T X, T) = (y, j))$. Then $\hat{P}$ on $(E^{\infty} \times E^{\infty}, \mathscr{E}^{\infty} \times \mathscr{E}^{\infty})$ defined by

$$\hat{P}(A \times B) = \int K((y, j), A) K'((y, j), B)\, dH((y, j))$$

for rectangles $A \times B$ provide a coupling of X and X', and for a pair $(\hat{X}, \hat{X}')$ with distribution $\hat{P}$ we have $T^* \stackrel{\mathcal{D}}{\leq} T$, where

$$T^* = \min\{k; \hat{X}_n = \hat{X}'_n \quad \text{for all } n \geq k\} .$$

We meet no obstacles when using the ideas above for processes in D_E rather than in discrete time; again we may lean on the Polish assumption for the existence of regular conditional probabilities, since the spaces D_E are Polish.

16. Notes. The existence of a maximal coupling was first established for Markov chains by Griffeath [66]. The coupling idea presented here and Theorem (14.10) are due to Goldstein [64]. Thorisson [152] simplified the construction by using a weak coupling. Our Theorem (14.16) is Theorem 4.1 in Orey [127]. Orey uses a backward martingale to prove that a trivial tail implies mixing. The strong coupling construction of § 15 is based on Thorisson [148, Construction 1.1]. Kallenberg [82] shows an example of a minimal extension of an underlying probability space in order to get a strong coupling. For an impetus to simulation, see Devroye [49].

4. REGENERATIVE PROCESSES

17. Basics. Stationarity. Loosely speaking, a regenerative process $Z = (Z_t)_0^\infty$ is a process that "starts from scratch" at renewal times S_n, $n \geq 0$. With a state space $(E, \mathscr{E})$ for the variables Z_t, one possible formulation of that is the following: Z is a regenerative process with respect to the renewal process $S = (S_n)_0^\infty$ if

(17.1) (i) the distributions of $\theta_{S_n} Z$ are equal for all $n \geq 0$, and
(ii) $\theta_{S_n} Z$ is independent of $\{Z_t; t < S_n\}$ and $\{S_0, \ldots, S_n\}$ for each $n \geq 0$.

Probably you have seen a definition like this, and some examples. One such is the number of customers in an GI/GI/1 queueing system at time t; then Z is regenerative w.r.t. $S = (S_n)_0^\infty$, where S_n is the nth time a customer arrives to an idle service station. Furthermore, a renewal equation may be set for $A(t) = \mathbf{P}(Z_t \in B)$, $B \in \mathbb{Z}_+$, in order to establish an asymptotic distribution of Z_t as $t \to \infty$.

But we like to replace that renewal equation technique by a coupling argument, of course, involving a stationary version of Z. The coupling is the topic of the paragraph to follow; here we shall make it clear to ourselves how stationarity can be realized. To that end, a new approach to stationarity of a renewal process will give crucial insight. We assume that the life-length distribution F of such a process S is of nonlattice type, and use the notation in § 1 without further notice. Throughout this section we assume that F, which is suitable to call the cycle-length distribution here, has a finite expectation μ. Since $\theta_t N$ is approximately stationary for t large, and since it is a reasonable guess that t is uniformly distributed in the interval $[S_k, S_{k+1})$ containing t, one should obtain a stationary process by first letting $Y_0 \stackrel{\mathscr{D}}{=}$ the limiting distribution of the length of that interval, then choosing a zero-point uniformly in $[0, Y_0]$. To see that this works, let S be zero-delayed to begin with, and put

$$I_t = S_{k+1} - S_k \qquad \text{where} \quad t \in [S_k, S_{k+1}) .$$

Then for $x \geq 0$, $A(t) = \mathbf{P}(I_t > x)$ equals

$$\sum_0^\infty \mathbf{P}(I_t > x, S_k \leq t, S_{k+1} > t) = U * \bar{F}(t \vee x)$$

and we find that the limiting distribution G_0 of I_t satisfies $dG_0(x) = \lambda \cdot x \cdot dF(x)$, a well-known fact from renewal theory.

Now if $Y_0 \stackrel{\mathscr{D}}{=} G_0$ and V is uniformly distributed on $[0, 1]$ (and independent of S), the delay of $\theta_{V \cdot Y_0} N$ is $(1 - V) \cdot Y_0$, which is immediately seen to have distribution $G_s[dG_s(x) = \lambda \bar{F}(x)\, dx$. Remember?]. Hence $\theta_{V \cdot Y_0} N$ is stationary.

We now return to the regenerative processes, with an idea how to create stationarity. The segments $\{Z_t; S_n \leq t < S_{n+1}\}$, $n \geq 0$, are called the cycles of Z, while $\{Z_t; t < S_0\}$ is the delay part of it. It is void if $S_0 \equiv 0$; this is the case when $Z \stackrel{\mathscr{D}}{=} \theta_{S_n} Z$ for all n. Then S is zero-delayed, and it is natural to use the same term for Z. The variables $Y_n = S_n - S_{n-1}$, $n \geq 1$ (with distribution F), are the cycle lengths.

It is not immediately obvious how to proceed in our program in order to create the delay part, but fortunately there is a device from the theory of stationary point processes that turns out to be useful. Indeed, let Z be zero-delayed and defined on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Since $S_0 \equiv 0$, the first regeneration in positive time takes place at $S_1 = Y_1$. Define a new probability $\hat{\mathbf{P}}$ on $(\Omega, \mathscr{F})$ by

$$d\tilde{\mathbf{P}} = \lambda \cdot Y_1 \, d\mathbf{P} . \tag{17.2}$$

We may do the innocent assumption that the variable V, uniformly distributed on $[0, 1]$ and independent of Z under $\tilde{\mathbf{P}}$, is defined on $(\Omega, \mathscr{F}, \tilde{\mathbf{P}})$. Then

(17.3) (i) $Z^* = \theta_{V \cdot Y_1} Z$ is a stationary process under $\tilde{\mathbf{P}}$, and
(ii) the distribution of $\theta_{Y_1} Z$ under $\tilde{\mathbf{P}}$ equals that of Z under $\mathbf{P}$, and
(iii) $\theta_{Y_1} Z$ and $\{Z_t; t < Y_1\}$ are independent under $\tilde{\mathbf{P}}$,

from which it follows that Z^* is a stationary version of Z. To prove (i), notice that for an $f \in b\mathscr{D}_E$ we have

(17.4)

$$\begin{aligned}
\tilde{\mathbf{E}}[f(Z^*)] &= \tilde{\mathbf{E}}[f(\theta_{V \cdot Y_1} Z)] \\
&= \tilde{\mathbf{E}}\Big[\int_0^1 f(\theta_{uY_1} Z)\, du\Big] = \tilde{\mathbf{E}}\Big[\int_0^{Y_1} f(\theta_u Z) \cdot (1/Y_1)\, du\Big] \\
&= \int \Big[\int_0^{Y_1} f(\theta_u Z)(1/Y_1)\, du\Big] \lambda \cdot Y_1 d\mathbf{P} \\
&= \lambda \cdot \mathbf{E}\Big[\int_0^{Y_1} f(\theta_u Z)\, du\Big] .
\end{aligned}$$

For an $A \in \mathscr{D}_E$ and a $t \geq 0$, we use (17.4) with $f = I_{\theta_t^{-1} A}$ to get

(17.5)

$$\begin{aligned}
\tilde{\mathbf{P}}(\theta_t Z^* \in A) = \tilde{\mathbf{E}}[f(Z^*)] &= \lambda \cdot \mathbf{E}\Big[\int_0^{Y_1} I_{\theta_t^{-1} A}(\theta_u Z)\, du\Big] \\
&= \lambda \cdot \mathbf{E}\Big[\int_0^{Y_1} I_A(\theta_{u+t} Z)\, du\Big] = \lambda \cdot \mathbf{E}\Big[\int_t^{t+Y_1} I_A(\theta_u Z)\, du\Big] \\
&= \lambda \cdot \Big(\mathbf{E}\Big[\int_t^{Y_1} I_A(\theta_t Z)\, du\Big]\Big) + \mathbf{E}\Big[\int_{Y_1}^{t+Y_1} I_A(\theta_u Z)\, du\Big]\Big) .
\end{aligned}$$

Now
$$\mathbf{E}\Big[\int_{Y_1}^{t+Y_1} I_A(\theta_u Z)\,du\Big] = \mathbf{E}\Big[\int_0^t I_A(\theta_u Z)\,du\Big]$$

since $Z \overset{\mathscr{D}}{=} \theta_{Y_1} Z$. Hence the last expression in (17.5) equals

$$\lambda \cdot \mathbf{E}\Big[\int_0^{Y_1} I_A(\theta_u Z)\,du\Big],$$

which is independent of t.

To understand (17.3)(ii), retain the meaning of A. We get

$$\begin{aligned}\tilde{\mathbf{P}}(\theta_{Y_1} Z \in A) &= \tilde{\mathbf{E}}[I_A(\theta_{Y_1} Z)]\\ &= \mathbf{E}[I_A(\theta_{Y_1} Z)\cdot\lambda\cdot Y_1] = \mathbf{E}[\mathbf{E}[\theta_{Y_1} Z)\| Y_1]\cdot\lambda\cdot Y_1]\\ &= \mathbf{E}[I_A(Z)]\cdot\mathbf{E}[\lambda\cdot Y_1] = \mathbf{P}(Z\in A)\cdot 1\end{aligned}$$

due to (17.1). To establish (17.3)(iii), take $B \in \sigma\{Z_t; t < Y_1\}$ to obtain

$$\begin{aligned}&\tilde{\mathbf{E}}[I_A(\theta_{Y_1} Z)\cdot I_B]\\ &\quad= \mathbf{E}[I_A(\theta_{Y_1} Z)\cdot I_B\cdot\lambda\cdot Y_1]\\ &\quad= \mathbf{E}[I_A(\theta_{Y_1} Z)]\cdot\mathbf{E}[I_B\cdot\lambda\cdot Y_1] \qquad \text{[due to (17.1)]}\\ &\quad= \tilde{\mathbf{E}}[I_A(\theta_{Y_1} Z)]\cdot\tilde{\mathbf{E}}[I_B] \qquad \text{[due to (17.1) and (17.3)(ii)]}.\end{aligned}$$

Thus $\theta_{Y_1} Z$ and $\{Z_t; t < Y_1\}$ are independent under $\tilde{\mathbf{P}}$, and since V is independent of Z, we get (17.3)(iii).

18. Coupling of regenerative processes. To make a coupling of two regenerative processes Z and Z', they must be of the same type (i.e., have the same cycle distribution). When so, their embedded renewal processes $S = (S_n)_0^\infty$ and $S' = (S'_n)_0^\infty$ have the same cycle-length distribution, F say, now assumed to be nonsingular.

Only little work is needed to produce an efficient coupling of Z and Z' with § 5 as a base. Let K be the transition kernel in $\mathbb{R}_+^\infty \times D_E$ such that $K(s, A)$ means a regular transition probability $\mathbf{P}(Z \in A \| S = s)$, and define K' analogously for S' and Z'. Further, let $(\hat{S}, \hat{S}')$ be the coupling of S and S' constructed in § 5, and Q the distribution of $(\hat{S}, \hat{S}')$ on $((\mathbb{R}_+^\infty)\ (\mathcal{R}_+^\infty)^2) = (\Omega_1, \mathcal{F}_1)$; we may let $(\hat{S}, \hat{S}')$ be defined on that space. Now define the probability measure $\hat{P}$ on $(\hat{\Omega}, \hat{\mathcal{F}}) = (\Omega_1 \times D_E^2, \mathcal{F}_1 \times \mathcal{D}_E^2)$ by

$$(18.1)\quad \hat{\mathbf{P}}(A \times (B \times B')) = \int_A K(s, B) \cdot K'(s', B')\, d\hat{Q}((s, s')) .$$

[Hence $\hat{\mathbf{P}} = Q \circ (K(\cdot\,, \cdot) \times K'(\cdot\,, \cdot))$, in the notation of § 8.] Then $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{P}})$ serves as the underlying probability space for $((\hat{S}, \hat{S}'), (\hat{Z}, \hat{Z}'))$; $(\hat{Z}, \hat{Z}')$ is the projection $\Omega \to D_E^2$.

The idea behind (18.1) is simple: we have constructed $((\hat{S}, \hat{S}'), (\hat{Z}, \hat{Z}'))$ such that

(18.2) (i) $\hat{Z} \stackrel{\mathcal{D}}{=} Z$ and $\hat{Z}' \stackrel{\mathcal{D}}{=} Z'$,
(ii) conditional on $(\hat{S}, \hat{S}')$ the processes $\hat{Z}$ are $\hat{Z}'$ are independent, and
(iii) $\hat{Z}$ depends on $(\hat{S}, \hat{S}')$ only through $\hat{S}$, and $\hat{Z}'$ only through $\hat{S}'$.

Now let

$$T = \min\{\hat{S}_i;\ \hat{S}_i = \hat{S}_j' \text{ for some } i \text{ and } j\} .$$

Then $(\hat{Z}, T)$ and $(\hat{Z}', T)$ constitute a weak coupling of Z and Z' in the sense of (I.4.5) and (I.4.6) (time is continuous). If you care for an argument, here is one. For $A \in \mathcal{D}_E$ and $B \in \mathcal{R}_+$, we shall settle

$$\hat{\mathbf{P}}(\theta_T \hat{Z} \in A,\ T \in B) = \hat{\mathbf{P}}(\theta_T \hat{Z}' \in A,\ T \in B) .$$

But with ξ and ξ' defined by

$$T = \hat{S}_\xi = \hat{S}_{\xi'}'$$

and

$$\mathcal{G} = \sigma\{\hat{S}_k, \hat{S}_m';\ k \le \xi,\ m \le \xi'\}$$

we have

$$\hat{\mathbf{P}}(\theta_T\hat{Z} \in A,\ T \in B) = \hat{\mathbf{E}}[I_A(\theta_T\hat{Z}) \cdot I_B(T)]$$
$$= \hat{\mathbf{E}}[I_B(T) \cdot \mathbf{E}[I_A(\theta_T\hat{Z}) \| \mathscr{G}]] .$$

But due to (17.1)(ii) and (18.2)(iii),

$$\mathbf{E}[I_A(\theta_T Z) \| \mathscr{G}] = \mathbf{E}[I_A(Z^\circ)] ,$$

where Z° is zero-delayed and of the same type as Z, defined on some probability space with governing probability measure **P** and expectation **E**. Hence

$$\hat{\mathbf{P}}(\theta_T\hat{Z} \in A,\ T \in B) = \hat{\mathbf{P}}(T \in B) \cdot \mathbf{P}(Z^\circ \in \mathbf{A}) ,$$

and an elaboration of $\hat{\mathbf{P}}(\theta_T\hat{Z}' \in A,\ T \in B)$ gives the same thing.

The continuous-time version of (I.4.7) yields

$$\|\mathbf{P}(\theta_t Z \in \cdot\,) - \mathbf{P}(\theta_t Z' \in \cdot\,)\| \le 2 \cdot \hat{\mathbf{P}}(T > t) . \tag{18.3}$$

But T is defined in terms of $\hat{S}$ and $\hat{S}'$, and all we know about T from § 6 now has consequences for our regenerative processes. For example, if Z is zero-delayed and Z' is stationary, then for $\alpha \ge 1$

$$\|\mathbf{P}(\theta_t Z \in \cdot\,) - \mathbf{P}(Z' \in \cdot\,)\| = o(t^{-(\alpha-1)}) \tag{18.4}$$

if $\int x^\alpha \, dF(x) < \infty$. Hence to establish results much as (18.4), the crucial thing is to establish finite moments of cycle lengths. A considerable amount of work on that task has been carried out concerning the regenerative processes in queueing theory.

19. Notes. We have been somewhat brief, partly because there exists thorough accounts of coupling of regenerative processes: Thorisson [148, 150, 156]. In these papers, regeneration is treated under weaker conditions than (17.1). See also Asmussen [10] for an account on regenerative processes (without emphasis on coupling).

Under a definition that allows regenerative processes to be non-time-homogeneous, we cannot hope for ergodicity. However, if we let the initial time point tend to $-\infty$, we may obtain a limit result for the part of the process that has a nonnegative time argument (a backward limit); see Thorisson [154].

Kallenberg [81] uses a coupling of embedded renewal processes in his study of regenerative sets. We collect here our references to applications in service systems and dam theory, even them without emphasis on regeneration. See Thorisson [149, 151], Asmussen and Thorisson [11, 12], Hordijk and Ridder [77, 78], Kendall [88], Smith and Whitt [144], and Sonderman [145].

5. ON MARKOV PROCESSES

20. Some remarks. For a homogeneous Markov process $X = (X_t)_0^\infty$ in continuous time, we will let P_t, $t \geq 0$, denote the governing Markov kernels. Hence, if X_t has state space $(E, \mathscr{E})$, $(P_t)_0^\infty$ is a family of Markov kernels in $E \times E$ satisfying

(20.1) (i) $P_0(x, \cdot) = \delta_x$ for all $x \in E$, and
(ii) $P_{t+s} = P_t \cdot P_s$ for all $s, t \geq 0$;

notation from § 8 will be used. Due to (ii), $(P_t)_0^\infty$ is called the transition semigroup governing X. For X with initial distribution λ we have

(20.2) (i) $\mathbf{P}_\lambda(X_t \in \cdot) = \lambda P_t = \int \lambda(dx) P_t(x, \cdot)$, and
(ii) $\mathbf{P}_\lambda(X_{t+s} \in \cdot \,\|\, \mathscr{F}_t) = P_s(X_t, \cdot)$ for all $s, t \geq 0$,

where $\mathscr{F}_t = \sigma\{X_s; s \leq t\}$.

A transition always means a reduction of the norm of a signed bounded measure. Indeed, for spaces $(E_1, \mathscr{E}_1)$, $(E_2, \mathscr{E}_2)$ and a transition kernel K in $E_1 \times E_2$, we have for $(Kf)(x) = \int f(y) K(x, dy)$ that if $f \in b\mathscr{E}_2$, then

$$\left| \int f \, d(\nu K) \right| = \left| \int (Kf) \, d\nu \right| \leq \|Kf\| \cdot \|\nu\|$$

and

$$\|\nu K\| \leq \|\nu\| \tag{20.3}$$

follows, since obviously $|Kf| \le 1$ if $|f| \le 1$. In particular, if λ and μ are initial distributions of Markov processes X and X' governed by $(P_t)_0^\infty$, (20.1) and (20.2) imply that

$$\|(\lambda - \mu)P_{t+s}\| = \|(\lambda - \mu)P_t P_s\| \le \|(\lambda - \mu)P_t\| . \tag{20.4}$$

Hence the distance

$$\|\mathbf{P}_\lambda(X_t \in \cdot\,) - \mathbf{P}_\mu(X'_t \in \cdot\,)\| = \|(\lambda - \mu)P_t\| \tag{20.5}$$

is nonincreasing in $t \ge 0$.

The discrete-time version is that

$$\|\mathbf{P}_\lambda(X_n \in \cdot\,) - \mathbf{P}_\mu(X_n \in \cdot\,)\| = \|(\lambda - \mu)P^n\| \tag{20.6}$$

is nonincreasing in n.

We gave a promise in § II.8 to improve the inequality (II.8.12). Consider an arbitrary state space. We have

$$\mathbf{P}_\lambda(\theta_n X \in \cdot\,) = \mathbf{P}_{\lambda P^n}(X \in \cdot\,)$$

due to the Markov property, so to prove

$$\begin{aligned} &\|\mathbf{P}_\lambda(\theta_n X \in \cdot\,) - \mathbf{P}_\mu(\theta_n X' \in \cdot\,)\| \\ &\qquad = \|\lambda P^n - \mu P^n\| = \|(\lambda - \mu)P^n\| \end{aligned} \tag{20.7}$$

it suffices to establish $\|\mathbf{P}_\lambda(X \in \cdot\,) - \mathbf{P}_\mu(X' \in \cdot\,)\| = \|\lambda - \mu\|$. Let g_λ and g_μ be the Radon–Nikodym derivatives

$$g_\lambda = d\lambda/d\nu \quad \text{and} \quad g_\mu = d\mu/d\nu ,$$

where $\nu = \lambda + \mu$. For an arbitrary $A \in \mathscr{E}^\infty$ we have

$$\begin{aligned} \mathbf{P}_\lambda(A) - \mathbf{P}_\mu(A) &= \int \mathbf{P}_x(A) g_\lambda(x)\, d\nu(x) - \int \mathbf{P}_x(A) g_\mu(x)\, d\nu(x) \\ &\le \int_{g_\lambda \ge g_\mu} (g_\lambda - g_\mu)\, d\nu = \tfrac{1}{2} \cdot \|\lambda - \mu\| , \end{aligned}$$

and $\|\mathbf{P}_\lambda - \mathbf{P}_\mu\| = \|\lambda - \mu\|$ follows from the definition of $\|\cdot\|$; we know already that $\|\lambda - \mu\| \le \|\mathbf{P}_\lambda - \mathbf{P}_\mu\|$. In the same way we obtain

(20.8)
$$\|\mathbf{P}_\lambda(\theta_t X \in \cdot) - \mathbf{P}_\mu(\theta_t X' \in \cdot)\| = \|\lambda P_t - \mu P_t\| = \|(\lambda - \mu)P_t\|$$

for the continuous-time case.

Our next remark concerns a simple consequence of (20.5) and (20.6). For any continuous-time Markov process $X = (X_t)_0^\infty$ we have that $\tilde{X} = (X_{n\delta})_0^\infty$ is a Markov chain for every $\delta > 0$, with Markov kernel $\tilde{P} = P_\delta$. To estimate $\|\mathbf{P}_\lambda(\theta_t X \in \cdot) - \mathbf{P}_\mu(\theta_t X' \in \cdot)\|$ we may exploit (20.8) and (20.5) and find that, with $k = k_t = [t/\delta] + 1$,

$$\begin{aligned}(20.9)\qquad \|\mathbf{P}_\lambda(\theta_k \tilde{X} \in \cdot) - \mathbf{P}_\mu(\theta_k \tilde{X}' \in \cdot)\| &= \|(\lambda - \mu)\tilde{P}_k\| \\ &\le \|(\lambda - \mu)P_t\| = \|\mathbf{P}_\lambda(\theta_t X \in \cdot) - \mathbf{P}_\mu(\theta_t X' \in \cdot)\| \\ &\le \|\mathbf{P}_\lambda(\theta_{k-1}\tilde{X} \in \cdot) - \mathbf{P}_\mu(\theta_{k-1}\tilde{X}' \in \cdot)\|\end{aligned}$$

since $(k-1)\cdot\delta \le t \le k\cdot\delta$. Hence since t and k_t are of the same order of magnitude, a successful coupling of $\tilde{X}$ and $\tilde{X}'$ yields an estimate for $\|\mathbf{P}_\lambda(\theta_t X \in \cdot) - \mathbf{P}_\mu(\theta_t X' \in \cdot)\|$ of the same quality as for $\|\mathbf{P}_\lambda(\theta_n \tilde{X} \in \cdot) - \mathbf{P}_\mu(\theta_n \tilde{X}' \in \cdot)\|$. This is worthwhile to remember, because

(20.10) (i) we may avoid some of the technical difficulties that appear as soon as we measure time on a continuous scale, and

(ii) there are opportunities of using results already at hand. For example $\tilde{X}$ may be a Harris chain for some $\delta > 0$; we then exploit § 11.

The last remark is concerned with the idea to produce a strong coupling by exchanging the future of one of the paths when they meet. Let $(\tilde{P}_t)_0^\infty$ be a transition semigroup governing a Markov process $\tilde{X} = (\tilde{X}_t)_0^\infty = ((X_t, X'_t))_0^\infty$ in $(E^2, \mathscr{E}^2)$ such that

(20.11)

$$\tilde{P}_t((x, x'), A \times E) = P_t(x, A) \text{ and } \tilde{P}_t((x, x'), E \times A') = P_t(x', A')$$

for all $x, x' \in E$ and $A, A' \in \mathscr{E}$. Then $X = (X_t)_0^\infty$ and $X' = (X'_t)_0^\infty$ are both governed by $(P_t)_0^\infty$, and if ν is an initial distribution for $\tilde{X}$ with marginals λ and μ, respectively, then the initial distributions of X and X' are λ and μ.

Now let

$$(20.12) \qquad X''_t = \begin{cases} X_t & \text{if } t < T \\ X'_t & \text{if } t \geq T, \end{cases}$$

where $T = \inf\{s; X_s = X'_s\}$.

The important question is: When do we have $X'' \stackrel{\mathscr{D}}{=} X$? It suffices to prove

(20.13)

$$\tilde{\mathbf{P}}_\nu(X''_{t_1} \in A_1, \ldots, X''_{t_k} \in A_k) = \tilde{\mathbf{P}}_\nu(X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k)$$

for all setups $t_1, t_2, \ldots, t_k$ and $A_1, \ldots, A_k$, where $0 < t_1 < \cdots < t_k$ and $A_i \in \mathscr{E}$, since the finite-dimensional rectangles are measure determining. With $t_0 = 0$, we have for each j, $1 \leq j \leq k$, that

$$\begin{aligned}
&\tilde{\mathbf{P}}_\nu(X''_{t_1} \in A_1, \ldots, X''_{t_k} \in A_k, t_{j-1} \leq T < t_j) \\
&\quad = \tilde{\mathbf{P}}_\nu(X_{t_1} \in A_1, \ldots, X_{t_{j-1}}, X'_{t_j} \in A_j, \ldots, X'_{t_k} \in A_k, t_{j-1} \leq T < t_j) \\
&\quad = \tilde{\mathbf{E}}_\nu[I_{A_1}(X_{t_1}) \cdot \ldots \cdot I_{A_{j-1}}(X_{t_{j-1}}) \\
&\qquad \times \mathbf{E}[I_{[t_{j-1}, t_j)}(T) \cdot I_{A_j}(X'_{t_j}) \cdot \ldots \cdot I_{A_k}(X'_{t_k}) \| \mathscr{F}_T]]
\end{aligned}$$

and we understand that (20.13) holds if

(20.14) $\tilde{X}$ has the strong Markov property, and $X_T = X'_T$

(after that attention has been paid to the event $\{T \geq t_k\}$), because if this is the case, then

$$\mathbf{E}[I_{[t_{j-1},t_j)}(T)I_{A_j}(X'_{t_j})\cdot \ldots \cdot I_{A_k}(X'_{t_k})\|\tilde{\mathcal{F}}_T]$$
$$= I_{[t_{j-1},t_j)}(T)\cdot e(X_T, X'_T, T),$$

where

$$e(x, x', u) = \tilde{\mathbf{E}}_{x,x'}[I_{A_j}(X'_{(t_j-u)^+})\cdot \ldots \cdot I_{A_k}(X'_{(t_k-u)^+})].$$

But due to (20.11),

$$e(x, \cdot, u) = e(\cdot, x, u) = \mathbf{E}_x[I_{A_j}(X^\circ_{(t_j-u)^+})\cdot \ldots \cdot I_{A_k}(X^\circ_{(t_k-u)^+})],$$

where X° is a process governed by $(P_t)_0^\infty$. Hence

$$e(x, x, u) = \tilde{\mathbf{E}}_{x,x}[I_{A_j}(X_{(t_j-u)^+})\cdot \ldots \cdot I_{A_k}(X_{(t_k-u)^+})].$$

We may now go backwards and realize that we have been able to replace all the X'_{t_i} variables in (20.13) by X_{t_i} variables.

It is well known that for Markov processes with paths in D_E, the Feller property is sufficient for the process to be (sufficiently) strongly Markovian. Notice that $\tilde{X}$ is D_{E^2}-valued, and that we certainly have $X_T = X'_T$. The Feller property means that for all $f \in bc\mathscr{E}$ (the space of continuous and bounded functions on E) we have $P_t f \in bc\mathscr{E}$ for all $t > 0$. If

$$\tilde{P}_t((x, x'), A \times A') = P_t(x, A)\cdot P_t(x', A'), \tag{20.15}$$

meaning that X and X' are independent, then (20.11) is immediate if $(P_t)_0^\infty$ has the Feller property. For cases where it suffices to produce a weak coupling, we shall prove that

$$(\theta_T X, T) \overset{\mathscr{D}}{=} (\theta_T X', T).$$

That is easier than settling $X \overset{\mathscr{D}}{=} X''$. Indeed, for $A \in \mathscr{D}_E$, $B \in \mathscr{R}_+$ we have under the stated conditions that

$$\begin{aligned}\tilde{\mathbf{E}}_\nu[I_A(\theta_T X)\cdot I_B(T)] &= \tilde{\mathbf{E}}_\nu[\tilde{\mathbf{E}}_{X_T, X'_T}[I_A(X)]\cdot I_B(T)]\\ &= \tilde{\mathbf{E}}_\nu[\mathbf{E}_{X'_T}[I_A(X')]\cdot I_B(T)]\\ &= \tilde{\mathbf{E}}_\nu[I_A(\theta_T X')\cdot I_B(T)]\end{aligned}$$

for a successful coupling. If $\tilde{\mathbf{P}}_\nu(T=\infty)>0$, recall the convention $\theta_\infty x =$ a fixed path for all x, and modify.

21. Ergodicity. For Markov chains, we use the term "ergodic" when

$$(21.1)\qquad \|\lambda P^n - \mu P^n\| = \|\mathbf{P}_\lambda(X_n \in \cdot\,) - \mathbf{P}_\mu(X_n \in \cdot\,)\| \to 0$$

as $n\to\infty$, for all initial distributions λ and μ. Hence ergodicity concerns forgetfulness of initial conditions. As mentioned before, the term is used in a number of ways. In the Markov theory context, some reserve it for chains (or processes in general) that possesses a unique stationary distribution π, and $\|\lambda P^n - \pi\| \to 0$ as $n\to\infty$ for all λ. If adopting that, some other term must be found for property (21.1); "weak ergodicity" has been suggested.

Due to the basic coupling inequality, successful coupling implies ergodicity. Theorem (14.10) deepened our understanding; It asserts that ergodicity is equivalent to the existence of successful coupling and that the restrictions of $\mathbf{P}_\lambda$ to tail events are the same for all λ (i.e., forgetfulness of initial conditions = identical distributions in the infinitely remote future).

But that theorem holds for arbitrary random sequences. For Markov chains we shall prove that ergodicity is equivalent to triviality of tail events. That enables us to combine Theorems (14.10) and (14.16); you now have the presentiment that we are approaching the statement of a major theorem on what coupling of Markov chains means.

But we are not ready for that yet: First, we must understand the role of harmonic functions and invariant events. A function $f\in b\mathscr{E}$ is said to be harmonic (w.r.t. P) if $Pf=f$. Of course, all constant functions are harmonic; the case when there are no other is most interesting, as we shall see. It is well known that if f is harmonic, then $f(X_n)$, $n=0,1,\ldots$, is a martingale w.r.t. the filtration $(\mathscr{F}_n)_0^\infty$, where $\mathscr{F}_n=\sigma(X_0,\ldots,X_n)$. Indeed, for all $n\ge 0$,

$$(21.2)\qquad \mathbf{E}[f(X_{n+1})\,\|\,\mathscr{F}_n] = \mathbf{E}[f(X_{n+1})\,\|\,X_n] = (Pf)(X_n) = f(X_n)\,.$$

Let $\mathscr{I}$ denote the σ-field of sets $A\in\mathscr{E}^\infty$ satisfying

$$A=\theta^{-1}A\,.$$

These are the invariant events, and we call $\mathscr{I}$ the invariant σ-field; we have $\mathscr{I} \subset \mathscr{T}$. Note that $\mathscr{I}$ also may be defined by

$$(21.3) \qquad f \in \mathscr{I} \Leftrightarrow f = f \circ \theta .$$

We shall at some place use θf for $f \circ \theta$. An $f \in \mathscr{I}$ is called an invariant function. For a sequence X in $(E^\infty, \mathscr{E}^\infty)$, we define $\mathscr{I}(X)$, the invariant σ-field of X, in analogy with $\mathscr{T}(X)$ [recall (14.14)]; hence $\mathscr{I}(X) = X^{-1}(\mathscr{I})$.

To illuminate that $\mathscr{I}$ is smaller than $\mathscr{T}$, notice that

$$(21.4) \qquad f \in \mathscr{T} \Leftrightarrow \text{there exist functions } f_n \in \mathscr{E}^\infty \text{ such that } f = f_n \circ \theta_n \text{ for all } n \geq 0.$$

Until the formulation of Theorem (21.12), we shall assume that $X = (X_n)_0^\infty$ is a canonical process. With no loss, that facilitates and even clarifies some work.

We are on our way to the result that a Markov chain has only trivial invariant events if and only if all its harmonic functions are constant. For that, associate to $\psi \in b\mathscr{E}^\infty$ the function $h_\psi \in b\mathscr{E}$ defined by

$$h_\psi(x) = \mathbf{E}_x[\psi(X)]$$

($= \mathbf{E}_x[\psi]$, since X is a canonical process).

(21.5) Lemma. *If $\psi \in b\mathscr{E}^\infty$ is invariant, then h_ψ is harmonic and $h_\psi(X_n) \to \psi$ a.s. $(\mathbf{P}_\lambda)$ as $n \to \infty$ for all λ.*

Proof. Fix a governing $\mathbf{P}_\lambda$. We have for any $n \geq 1$,

$$(21.6) \qquad h_\psi(X_n) = E_{X_n}[\psi] = \mathbf{E}[\theta^n \psi \,\|\, \mathscr{F}_n] = \mathbf{E}[\psi \,\|\, \mathscr{F}_n] .$$

In particular for $n = 1$,

$$h_\psi(x) = \mathbf{E}_x[\psi] = \mathbf{E}_x[\mathbf{E}[\psi \,\|\, \mathscr{F}_1]] = E_x[h_\psi(X_1)] = (Ph_\psi)(x) ,$$

hence h_ψ is harmonic. Further,

$$\mathbf{E}[\psi \,\|\, \mathscr{F}_n] \to \mathbf{E}[\psi \,\|\, \mathscr{F}] \quad \text{as} \quad n \to \infty ,$$

a well-known application of martingale theory. Since $\mathbf{E}[\psi \| \mathcal{F}] = \psi$ a.s. and $h_\psi(X_n) = \mathbf{E}[\psi \| \mathcal{F}_n]$ due to (21.6), the convergence assertion is proved. □

(21.7) Lemma. *If $h \in b\mathscr{E}$ is harmonic, there exists an invariant $\psi \in b\mathscr{E}^\infty$ such that $h = h_\psi$.*

Proof. Let $\psi = \liminf_{n\to\infty} h(X_n)$, which is finite since h is bounded. We have

$$\psi = \liminf_{n\to\infty} h(X_{n+1}) = \liminf_{n\to\infty} (\theta h)(X_n) = \theta\psi\ ,$$

hence $\psi \in \mathcal{I}$. Since h is harmonic, we have $h(X_n) \to \psi$ a.s. $(\mathbf{P}_x)$ for all $x \in E$. But from Lemma (21.5) we know that $h_\psi(X_n) \to \psi$ a.s. $(\mathbf{P}_x)$. We may conclude that

$$h(x) = \mathbf{E}_x[h(X_0)] = \mathbf{E}_x[\lim_{n\to\infty} h(X_n)] = \mathbf{E}_x[\psi] = h_\psi(x)\ .$$

Hence $h = h_\psi$. □

We are now prepared for the following result.

(21.8) Theorem. *All bounded harmonic functions are constant if and only if the invariant σ-field $\mathcal{I}$ is trivial $(\mathbf{P}_\lambda)$ for all λ.*

Proof. Recall (14.12)–(14.13). For the "only if" part, take $A \in \mathcal{I}$ and let $\psi = I_A$. Due to Lemma (21.5), h_ψ is harmonic, hence constant. Since $\psi = \lim_{n\to\infty} h_\psi(X_n)$ a.s., ψ is constant a.s. $(\mathbf{P}_\lambda)$.

Now suppose that h is a nonconstant bounded harmonic function, take x, y such that $h(x) \neq h(y)$. Find the $\psi \in b\mathcal{I}$ of Lemma (21.7) satisfying $h = h_\psi$, and let $\lambda = \frac{1}{2}\cdot(\delta_x + \delta_y)$. That ψ is not constant a.s. $(\mathbf{P}_\lambda)$ because it takes different mean values on the sets $\{X_0 = x\}$ and $\{X_0 = y\}$, which have $\mathbf{P}_\lambda$-probabilities $\frac{1}{2}$ each. Indeed,

$$2\cdot \int_{X_0 = x} \psi\, d\mathbf{P}_\lambda = h(x) \neq h(y) = 2\cdot \int_{X_0 = y} \psi\, d\mathbf{P}_\lambda\ . \qquad \square$$

But this theorem is not a sharp enough tool to decide which

Markov chains have trivial tails: Recall that $\mathcal{I} \subset \mathcal{T}$. Fortunately, the extra effort needed is not large: It is only a matter of knowing what time it is.

To introduce a clock, extend $X = (X_n)_0^\infty$ to a Markov chain $\tilde{X}$ in $\tilde{E} = E \times \mathbb{Z}_+$ with $\tilde{X}_n = (X_n, K_n)$ such that $K_{n+1} = K_n + 1$ for $n = 0, 1, \ldots$. The extended chain, to be called the space-time chain associated with X, is defined by the transition probabilities

(21.9)

$$\tilde{P}((x, j), A \times \{j+1\}) = P(x, A) \qquad \text{for} \quad x \in E, A \in \mathcal{E}, j \geq 0 .$$

For $A \in \mathcal{E}^\infty$ and $k \geq 0$, let

$$A_{(k)} = \{((x_i, k+i))_0^\infty;\ (x_i)_0^\infty \in A\} .$$

Let the product σ-field on $\tilde{E}$ be denoted by $\tilde{\mathcal{E}}$, and the invariant σ-field on $\tilde{\mathcal{E}}^\infty$ by $\tilde{\mathcal{I}}$. For an initial distribution ν of $\tilde{X}$, let $\tilde{P}_\nu$ have the obvious meaning.

(21.10) Lemma. *Fix a $k \geq 0$. Then we have $A \in \mathcal{T}$ if and only if $A_{(k)} \in \tilde{\mathcal{I}}$.*

Proof. Suppose that $A \in \mathcal{T}$, and let f_n be a sequence of functions such that $I_A = f_n \circ \theta_n$. Define $\tilde{f} \in \tilde{\mathcal{E}}^\infty$ by

$$\tilde{f}((x_i, k+n+i)_{i=0}^\infty) = f_n(x_0, x_1, \ldots) \tag{21.11}$$

for sequences of the type $(x_i, j+i)_{i=0}^\infty$ with $j \geq k$, and let $\tilde{f}$ be 0 for all other arguments. We have

$$\tilde{f}(\theta_n((x_i, k+i))_{i=0}^\infty) = f_n(\theta_n(x_i)_{i=0}^\infty) \qquad \text{for all} \quad n \geq 0 .$$

Now $f_n \circ \theta_n = I_A$ does not depend on n, hence $\tilde{f} \circ \theta = \tilde{f}$. But the invariant $\tilde{f}$ equals $I_{A_{(k)}}$.

For the converse, reverse the order of arguments. Let $A_{(k)} \in \tilde{\mathcal{I}}$ for an $A \in \mathcal{E}^\infty$, and put $\tilde{f} = I_{A_{(k)}}$. With

$$f_n(x_0, x_1, \ldots) = \tilde{f}((x_0, k+n), (x_1, k+n+1), \ldots)$$

for $n \geq 0$, we get $I_A = f_n \circ \theta_n$, hence $A \in \mathcal{T}$. $\square$

We are ready for the main theorem on ergodicity and coupling of Markov chains. The assumption that the chains are canonical is not in force any more.

(21.12) Theorem. *For Markov chains $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$, governed by the Markov kernel P, the following assertions are equivalent*:

For all initial distributions λ

(i) *X has a trivial tail σ-field $\mathcal{T}(X)$*,

(ii) *X is mixing under $\mathbf{P}_\lambda$*;

for all initial distributions λ and μ

(iii) *there exists a successful coupling of X* and *X'*,

(iv) $\|\lambda P^n - \mu P^n\| \to 0$ *as* $n \to \infty$,

(v) $\|\mathbf{P}_\lambda(\theta_n X \in \cdot\,) - \mathbf{P}_\mu(\theta_n X' \in \cdot\,)\| \to 0$ *as* $n \to \infty$;

and

(vi) *all the bounded space-time harmonic functions are constant*;

and for all initial distributions λ

(vii) *the invariant σ-field of $\tilde{X} = ((X_n, k+n))_0^\infty$ is trivial* $(\mathbf{P}_{\lambda \times \delta_k})$ *for at least one k.*

Proof. We have already proved that

(i)$\Leftrightarrow$(ii) [Theorem (14.16)],
(i)$\Rightarrow$(iii) [Theorem (14.10): (i) implies that $\mathbf{P}_\lambda$ and $\mathbf{P}_\mu$ agree on $\mathcal{T}$]),
(iii)$\Rightarrow$(iv) [basics],
(iv)$\Leftrightarrow$(v) [equality (20.7)],
(vi)$\Rightarrow$(vii) [Theorem (21.8)], and
(vii)$\Leftrightarrow$(i) [Lemma (21.10)].

Hence we are done if we can prove (iv)$\Rightarrow$(vi) (draw a graph!). Let h be a bounded harmonic function on $E \times \mathbb{Z}_+$ w.r.t. $\tilde{P}$, that is,

$$(21.13) \qquad h(x, k) = \int h(z, k+n) P^n(x, dz)$$

for all $x \in E$ and $k, n \geq 0$. Hence for $x, y \in E$ and $k \geq 0$,

$$|h(x, k) - h(y, k)| \leq$$
$$\|h\| \cdot \|P^n(x, \cdot) - P^n(y, \cdot)\| \to 0 \quad \text{as} \quad n \to \infty;$$

thus h is constant in its first argument—but in its second too, due to (21.13). $\square$

If you feel uneasy about the fact that we are leaning on Theorem (14.10), a result with a complicated proof, you may withdraw (iii) from the list and content yourself with the result that (iii) implies each of the equivalent (i), (ii) (iv)–(vii); for that you need to prove (ii)$\Rightarrow$(iv), which is rather easy. Also recall that Theorem (14.16) may be proved without reliance on Theorem (14.10); see § 16.

For an extension of the theorem to continuous-time processes, let us postpone the discussion about harmonic functions and space-time processes for a while, and concentrate on (i)–(v), the analogs of which will be denoted by c(i)–c(v). Recall (14.17)–(14.18). Since X is a Markov process, $Y = (\theta_n X)_0^\infty$ is a Markov chain in a Polish space, hence Theorem (21.12) applies to that. What remains is to transfer the knowledge thus obtained to X. The Markov chain Y' is defined from a second process X'.

(21.14) Theorem. *For D_E-valued Markov processes X and X', governed by the Markov kernels $(P_t)_0^\infty$, the analogs* c(i)–c(v) *of assertions* (i)–(v) *of Theorem* (21.12) *are equivalent.*

Proof. We let X' and X' be canonical processes in $(D_E, \mathscr{D}_E)$. Then $(D_E, \mathscr{D}_E, \mathbf{P}_\lambda)$ serves as the underlying probability space for Y. We abbreviate "c(ii) holds for X if and only if (ii) holds for Y", e.g., to "c(ii)$\Leftrightarrow$(ii)". We have c(i)$\Leftrightarrow$(i) since

$$\mathscr{T}_{n+1} \subset \mathscr{T}_t \subset \mathscr{T}_n(Y) \qquad \text{for} \quad n \leq t \leq n+1$$

which implies that $\mathscr{T} = \mathscr{T}(Y)$. This relation yields that for all $B \in \mathscr{D}_E$

$$\sup_{A\in\mathcal{T}_t} |\mathbf{P}_\lambda(A\cap B)-\mathbf{P}_\lambda(A)\cdot\mathbf{P}_\lambda(B)|$$
$$\leq \sup_{A\in\mathcal{T}_{[t]}} |\mathbf{P}_\lambda(\{Y_{[t]}\in A\}\cap B)-\mathbf{P}_\lambda(\{Y_{[t]}\in A\})\cdot\mathbf{P}_\lambda(B)|\,,$$

with the inequality reversed if we replace $[t]$ by $[t]+1$. Hence c(ii)$\Leftrightarrow$(ii).

Also, (iii)$\Rightarrow$c(iii), because if $\hat{Y}$ and $\hat{Y}'$ is a successful strong coupling of Y and Y', we get a successful coupling of X and X', too, namely $\hat{X}=\hat{Y}_0$ and $\hat{X}'=\hat{Y}'_0$ (cf. the last lines of § 14).

Since obviously c(iii)$\Rightarrow$c(iv)$\Leftrightarrow$c(v), a moment's thought gives that if c(v)$\Rightarrow$(iv), the equivalence of c(i)–c(v) is proved. But that implication is immediate. $\square$

As was shown in the proof of Orey's theorem for Harris chains, the use of space-time harmonic functions is an efficient tool to prove ergodicity. Also, it was a link in the chain or arguments to settle Theorem (21.12). However, it was not needed for the latter purpose in the proof of Theorem (21.14). Since it is not very interesting to establish that all space-time harmonic functions are constant for its own sake, we content ourselves with a sufficiency result.

A function h is harmonic w.r.t. $(P_t)_0^\infty$ if $P_t h = h$ for all $t\geq 0$; that is, $h(x)=\mathbf{E}_x[h(X_t)]$ if $X=(X_t)_0^\infty$ is governed by $(P_t)_0^\infty$. As for the discrete case, we extend X to a space-time Markov process, with paths

$$\tilde{X}_t=(X_t, a+t), \qquad t\geq 0\,;$$

the clock shows a at time 0. The process $\tilde{X}$ has state space $\tilde{E}=E\times\mathbb{R}_+$, and Markov kernel defined by

$$\tilde{P}_t((x,a), A\times B)=P_t(x,A)\cdot I_B(a+t)$$

for $t\geq 0$, $(x,a)\in\tilde{E}$, $A\in\mathscr{E}$ and $B\in\mathscr{R}_+$.

(21.15) Theorem. *If all the bounded space-time harmonic functions associated with a Markov process X are constant, then X is ergodic; hence* c(i)–c(v) *of Theorem* (21.14) *hold.*

Proof. Simple, because the condition implies that all the space-time harmonic functions of the Markov chain $(Y_n)_0^\infty$ are constant. Due to the implication (vi) $\Rightarrow$ (v) of Theorem (21.12), this renders $(Y_n)_0^\infty$ ergodic. But that is equivalent to each of c(i)–c(iv) of Theorem (21.14), as was shown in the proof of that theorem. $\square$

22. Notes. For a serious study of continuous-time Markov processes, consult Chung [40], Williams [159], and/or Ethier and Kurtz [60]. The material of § 20 uses standard theory. The account of ergodicity owes much to Orey [127, Sec. I.4]. Consult also Aldous and Thorisson [6], Greven [65], and Thorisson [155].

CHAPTER IV

Inequalities

1. STRASSEN'S THEOREM

1. Basics. Let P and P' be probability measures on $(\mathbb{R}, \mathcal{R})$. We say that P' dominates P stochastically if

$$(1.1) \qquad P([x, \infty)) \le P'([x, \infty)) \qquad \text{for all} \quad x \in \mathbb{R},$$

and write $P \overset{\mathcal{D}}{\le} P'$ when this holds. In terms of the distribution functions F and F' corresponding to P and P', this equals $F \ge F'$, as is easily seen.

If X and X' are random variables such that $X \overset{\mathcal{D}}{=} P$ and $X' \overset{\mathcal{D}}{=} P'$, where $P \overset{\mathcal{D}}{\le} P'$, we say that X is stochastically smaller than X' and write $X \overset{\mathcal{D}}{\le} X'$. It is well known that (1.1) is equivalent to

$$(1.2) \qquad \int f\, dP \le \int f\, dP' \qquad \text{for} \quad f \text{ bounded and nondecreasing},$$

hence for such a function we have $\mathbf{E}[f(X)] \le \mathbf{E}[f(X')]$ if $X \overset{\mathcal{D}}{\le} X'$.

Probably you know that if $X \overset{\mathcal{D}}{\le} X'$, there exists a coupling $(\hat{X}, \hat{X}')$ of X and X' such that $\hat{X} \le \hat{X}'$. The standard proof of this is the following: Let G be a distribution function and define a generalized inverse G^* of G by

$$(1.3) \qquad G^*(u) = \inf\{x; G(x) \ge u\} \qquad \text{for} \quad u \in (0, 1).$$

Then $G^*(U) \overset{\mathcal{D}}{=} G$ if U is uniformly distributed on $(0, 1)$. Letting F, F' denote the distribution functions of X, X' respectively, we have

$F \geq F'$, which implies that $F^* \leq F'^*$. With $\hat{X} = F^*(U)$ and $\hat{X}' = F'^*(U)$ we have completed the desired coupling.

The inequality (1.2) becomes entirely trivial with this coupling at hand; indeed,

$$\int f\,dP = \mathbf{E}[f(X)] = \mathbf{E}[f(\hat{X})] \leq \mathbf{E}[f(\hat{X}')] = \mathbf{E}[f(X')] = \int f\,dP' . \tag{1.4}$$

Expressed in terms of probability measures, we have proved that if $P \overset{\mathcal{D}}{\leq} P'$, there exists a coupling $\hat{P}$ of P and P' such that

$$\hat{P}(\{(x, x'); x \leq x'\}) = 1 . \tag{1.5}$$

A coupling satisfying (1.5) exists under general conditions on the state space and its ordering; that is the point of Strassen's theorem.

2. The theorem. Recall the Polish assumption. A relation $\preceq$ on E is called a partial ordering if

(2.1) (i) $x \preceq x$ for all $x \in E$,
(ii) $x \preceq y$, $y \preceq z \Rightarrow x \preceq z$, and
(iii) $x \preceq y$, $y \preceq x \Rightarrow x = y$.

As agreed in § Int.3, we let $i\mathscr{E}$ the class of functions $f \in \mathscr{E}$ such that

$$x \preceq y \Rightarrow f(x) \leq f(y) . \tag{2.2}$$

For probability measures P and P' on $(E, \mathscr{E})$ we write $P \overset{\mathcal{D}}{\preceq} P'$ if

$$\int f\,dP \leq \int f\,dP' \qquad \text{for all} \quad f \in ib\mathscr{E} , \tag{2.3}$$

and give $X \overset{\mathcal{D}}{\preceq} X'$ the obvious meaning. Notice that $X \overset{\mathcal{D}}{\preceq} X'$ if and only if $g(X) \overset{\mathcal{D}}{\preceq} g(X')$ for all $g \in ib\mathscr{E}$.

We need to assume that $\preceq$ is closed, that is, the set

$$M = \{(x, x') \in E^2; x \preceq x'\}$$

is closed in the product topology on E^2. Of course, $\hat{P}(M)=1$ for a coupling $\hat{P}$ of P and P' on $(E, \mathscr{E})$ implies that $P \overset{\mathscr{D}}{\leq} P'$. The remarkable thing is that the converse is true if $\leq$ is a closed partial ordering.

(2.4) Theorem. *If P and P' are probability measures on $(E, \mathscr{E})$ satisfying $P \overset{\mathscr{D}}{\leq} P'$, then there exists a coupling $\hat{P}$ of P and P' such that $\hat{P}(M)=\mathbf{1}$.*

Proof. There is no hope for a complete proof of reasonable length, because of the technical complications. An outline, however, is possible.

Due to the fact that a probability measure on a Polish space is tight, (cf. § App.1), it is justifiable to assume that E is compact (the approximation problems we save ourselves are not trivial, however). The continuous functions on E or E^2 are bounded, due to the compactness. This makes boundedness assumptions superfluous. If L is a linear functional on $c\mathscr{E}^2$ which

(i) is nonnegative, and
(ii) satisfies $L\mathbf{1}=1$ ($\mathbf{1}$ is the function $\equiv 1$),

then due to the Riesz representation theorem, there is a unique probability measure Q on $(E^2, \mathscr{E}^2)$ such that

$$Lf = \int f \, dQ\,. \tag{2.5}$$

If we can produce such a functional satisfying

(2.6) (i) $Lf \leq 0$ if $f(x,x')=0$ for $(x,x') \in M$, and

(ii) $Lf = \int g_1 \, dP + \int g_2 \, dP'$ if $f(x,x') = g_1(x) + g_2(x')$ for some $g_1, g_2 \in c\mathscr{E}$,

then the corresponding Q of (2.5) is the desired $\hat{P}$. Denote the subspace of functions f that has the representation in (2.6)(ii) by $\mathscr{L}$ (it is unique up to addition of constants), and define L for $f \in \mathscr{L}$ by (2.6)(ii). For the extension of the domain of L from $\mathscr{L}$ to $c\mathscr{E}^2$, the

Hahn-Banach theorem will be instrumental. To find a suitable seminorm to apply that theorem, define

$$\hat{f}(x') = \inf\{h(x'); h \in ic\mathscr{E} \text{ and } h(x) \geq f(x, x') \text{ for all } x\}$$

and

$$\rho(f) = \int \hat{f}\, dP'$$

for $f \in c\mathscr{E}^2$. It is easy to check that

$$\rho(f_1 + f_2) \leq \rho(f_1) + \rho(f_2), \quad \text{and} \qquad (2.7)$$
$$\rho(cf) = c \cdot \rho(f) \qquad \text{for} \quad c \geq 0 .$$

Further, let

$$\bar{g}(x) = \inf\{h(x); h \in ic\mathscr{E} \text{ and } h \geq g\} .$$

Then for $f \in \mathscr{L}$ with representing functions g_1, g_2 we have

$$\begin{aligned} Lf &= \int g_1\, dP + \int g_2\, dP' \qquad (2.8)\\ &\leq \int \bar{g}_1\, dP + \int g_2\, dP' \leq \int \bar{g}_1\, dP' + \int g_2\, dP' \\ &= \int \hat{f}\, dP' = \rho(f) \end{aligned}$$

since $\bar{g}_1 \in i\mathscr{E}$, $P \overset{\mathscr{D}}{\leq} P'$, and actually, $\bar{g}_1 + g_2 = \hat{f}$. Granted (2.7)–(2.8), the Hahn-Banach theorem provide an extension of L to $c\mathscr{E}^2$ such that $Lf \leq \rho(f)$ for all $f \in c\mathscr{E}^2$. Of course, $L\mathbf{1} = 1$, and since $\hat{f} \leq 0$ if $f \leq 0$, we get $Lf \leq 0$ from (2.8) for such functions f. Hence L is nonnegative.

It remains to settle (2.6)(i). But for every $f \in c\mathscr{E}^2$ with $f(x, x') = 0$ for $(x, x') \in M$, there exists an $\tilde{f}$: $E^2 \rightarrow R$ such that

$$f \leq \tilde{f},$$
$$\tilde{f}(x, x') = 0 \qquad \text{if} \quad (x, x') \in M, \quad \text{and}$$
$$\tilde{f}(\cdot\,, x') \in ic\mathscr{E} \qquad \text{for all} \quad x' .$$

This implies that

$$\tilde{f}(x') \le \tilde{f}(x, x') = 0.$$

Hence

$$Lf \le \rho(f) = \int \hat{f}\, dP' = 0$$

for $f \in c\mathscr{E}^2$, and we are done. $\square$

3. Alternative formulations. If K is a transition kernel in E satisfying

$$K(x, \{y; x \preceq y\}) = 1 \qquad \text{for all} \quad x\,,$$

we call it an upward kernel. For alternative (i) below, we use the probability space $((0,1), \mathscr{R}_{(0,1)}, l_{(0,1)})$ and the identity mapping to produce a uniformly distributed variable U.

For the probability measures P and P' on $(E, \mathscr{E})$, $P \overset{\mathscr{D}}{\preceq} P'$ is equivalent to any of the following statements: There exist

(3.1) (i) measurable mappings h, $h' \in \mathscr{R}_{(0,1)}/\mathscr{E}$ such that $h < h'$ and $h(U) \overset{\mathscr{D}}{=} P$, $h'(U) \overset{\mathscr{D}}{=} P'$,

(ii) random elements Z, Z' in $(E, \mathscr{E})$ such that $Z \overset{\mathscr{D}}{=} P$, $Z' \overset{\mathscr{D}}{=} P'$, and $Z \preceq Z'$, and

(iii) an upward kernel K in E such that $PK = P'$.

Of course (i) $\Rightarrow$ (ii), and (ii) $\Rightarrow$

$(*)$ there exists a coupling $\hat{P}$ of P and P' such that $\hat{P}(M) = 1$,

which holds if and only if $P \overset{\mathscr{D}}{\preceq} P'$. Further, $(*) \Rightarrow$ (i) follows from the Skorohod representation (cf. § App.1).

It remains to prove (iii) $\Leftrightarrow$ (i), (ii) or $(*)$; we prove (iii) $\Rightarrow (*)$ and (ii) $\Rightarrow$ (iii). For (iii) $\Rightarrow (*)$, define $\hat{P}$ on $(E^2, \mathscr{E}^2)$ by

$$\hat{P}(A \times A') = \int_A P(dx) K(x, A')$$

for rectangles $A \times A'$ [hence $\hat{P} = P \circ K$; recall (III.8.4)–(III.8.5)]. Certainly, $\hat{P}(M) = 1$, $\hat{P}(A \times E) = P(A)$, $\hat{P}(E \times A') = (PK)(A) = P'(A)$.

To establish (ii)$\Rightarrow$(iii), recall that the Polish assumption is sufficient for the existence of regular conditional probabilities

$$K^*(x, A') = \mathbf{P}(Z' \in A' \| Z = x) .$$

Letting $B = \{x; K^*(x, \{y; x \leq y\}) < 1\}$, we modify K^* to K by

$$K(x, A') = \begin{cases} K^*(x, A') & \text{if } x \notin B \\ \delta_x & \text{if } x \in B . \end{cases}$$

K is an upward kernel, and since $P(B) = 0$, we have $PK = P'$.

4. Notes. For partial orderings, Theorem (2.4) is a special case of Theorem 11 in Strassen [147]: take his ϵ and ω to be 0 and M, respectively. Our outline of proof follows that of Liggett [104, Sec. II.2] rather closely. That general theorem of Strassen may also be used for other purposes than domination; see Dudley [55]. The alternative formulations in § 3 are from Kamae, Krengel, and O'Brien [84].

2. DOMINATION

5. The general result. Let the state spaces E_i be equipped with the close partial orderings $\preceq_i$, respectively. For transition kernels K and K' in $E_1 \times E_2$, we say that K is dominated by K' if

$$(5.1) \qquad K(x, \cdot) \overset{\mathcal{D}}{\preceq}_2 K'(x', \cdot) \qquad \text{for} \quad x \preceq_1 x' .$$

If $K = K'$, we say that K is monotone.

The purpose of this section is to prove that if $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ are random sequences such that for each $n \geq 1$, the kernel given by $\mathbf{P}(X_n \in \cdot \| (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1}))$ is dominated by the analogous transition probability for X', there exists a coupling $(\hat{X}, \hat{X}')$ of X and X' such that $\hat{X}$ is dominated by $\hat{X}'$ a.s., in a sense to be made precise. To that end, some preparatory work is needed. For the spaces $(E_i, \mathscr{E}_i)$ endowed with the (closed) partial orderings $\leq_i$, $i \geq 0$, define for $0 \leq n < \infty$ partial orderings $\leq_n^*$ on $\Pi_0^n E_i$ by

$$(5.2) \quad x \leq_n^* x' \quad \text{for} \quad x = (x_0, \ldots, x_n),\, x' = (x_0', \ldots, x_n')$$
$$\text{if} \quad x_i \leq_i x_i' \quad \text{for each } i, \quad 0 \leq i \leq n\,.$$

Let λ, μ be probability measures on $(E_0, \mathscr{E}_0)$, and for $i \geq 1$, let K_i, K_i' be transition kernels in $\Pi_0^{i-1} E_j \times E_i$.

(5.3) Lemma. *If* $\lambda \overset{\mathscr{D}}{\leq}_0 \mu$ *and if* K_i *is dominated by* K_i' *for* $1 \leq i \leq n < \infty$ *w.r.t.* $\leq_{(i-1)}^*$ *and* $\leq_i$, *then*

$$(5.4) \qquad \lambda \circ K_1 \circ \cdots \circ K_n \overset{\mathscr{D}}{\leq}{}_n^* \mu \circ K_1' \circ \cdots \circ K_n'\,.$$

Proof. Suppose that (5.4) holds with n replaced by $j-1$, where $j \leq n$. Let f be bounded and increasing on $\Pi_0^j E_i$. We have

$$\int f(x_0, \ldots, x_j)(\lambda \circ K_1 \circ \cdots \circ K_j)(dx_0, \ldots, dx_j)$$
$$= \int (\lambda \circ K_1 \circ \cdots \circ K_{j-1})(dx_0, \ldots, dx_{j-1})$$
$$\times \int f(x_0, \ldots, x_j) K_j(x_0, \ldots, x_{j-1}, dx_j)\,.$$

Denote $\int f(x_0, \ldots, x_j) K_j(x_0, \ldots, x_{j-1}, dx_j)$ by $g(x_0, \ldots, x_{j-1})$. Certainly, g is increasing, and $g \leq g'$, where $g'(x_0, \ldots, x_{j-1}) = \int f(x_0, \ldots, x_j) K_j'(x_0, \ldots, x_{j-1}, dx_j)$. An induction argument completes the proof. □

We also need the result that the domination in (5.4) is retained as $n \to \infty$. Define the projection $\pi_n\colon \Pi_0^\infty E_i \to \Pi_0^n E_i$ by

$$\pi_n(x_0, x_1, \ldots) = (x_0, \ldots, x_n)\,.$$

(5.5) Lemma. *For the probability measures* P *and* P' *on* $(\Pi_0^\infty E_i, \Pi_0^\infty \mathscr{E}_i)$, *suppose that*

(5.6) $$P\pi_n^{-1} \overset{\mathcal{D}}{\leq}{}_n^* P'\pi_n^{-1} \quad \textit{for all} \quad n \geq 1 .$$

Then $P \overset{\mathcal{D}}{\leq}{}_\infty^* P'$.

Proof. We shall prove that

(5.7) $$\int f \, dP \leq \int f \, dP'$$

for bounded and increasing functions f on $\Pi_0^\infty E_i$. Fix a sequence $(z_k)_0^\infty$ in $\Pi_0^\infty E_i$, and define a pseudoinverse to π_n by

$$\pi_n^*(x_0, \ldots, x_n) = (x_0, x_1, \ldots, x_n, z_{n+1}, \ldots) .$$

Let

$$f_n = f \circ \pi_n^* .$$

We have $f_n \circ \pi_n \to f$ a.s. w.r.t. P and P' as $n \to \infty$ (why?), so dominated convergence renders $\int (f_n \pi_n) \, dP \to \int f \, dP$, $\int (f_n \circ \pi_n) \, dP' \to \int f \, dP'$ as $n \to \infty$. Certainly, f_n is increasing on $\Pi_0^n E_i$; hence (5.6) implies that

$$\int f_n \, d(P\pi_n^{-1}) \leq \int f_n \, d(P'\pi_n^{-1}) .$$

But $\int f_n \, d(P\pi_n^{-1}) = \int (f_n \circ \pi_n) \, dP$ and $\int f_n \, d(P\pi_n^{-1}) = \int (f_n \circ \pi_n) \, dP$; (5.7) may be concluded. □

We are now ready to state the discrete-time domination theorem. Recall the notation, and so on, preceding Lemma (5.3).

(5.8) Theorem. *If* $\lambda \overset{\mathcal{D}}{\leq}_0 \mu$ *and if* K_i *is dominated by* K_i' *for* $i \geq 1$, *then there exist random sequences* $Z = (Z_n)_0^\infty$ *and* $Z' = (Z_n')_0^\infty$ *in* $(\Pi_0^\infty E_i, \Pi_0^\infty \mathscr{E}_i)$ *such that*

$$(Z_0, \ldots, Z_n) \overset{\mathcal{D}}{=} \lambda \circ K_1 \circ \cdots \circ K_n \quad \text{and}$$

$$(Z_0', \ldots, Z_n') \overset{\mathcal{D}}{=} \mu \circ K_1' \circ \cdots \circ K_n'$$

for each $n \geq 1$, and

$$Z_0 \leq_0 Z_0', \; Z_1 \leq_1 Z_1', \ldots \text{ a.s.}$$

Proof. Use Kolmogorov's consistency theorem to provide probability measures P and P' on $(E_0^\infty, \mathscr{E}_0^\infty)$ such that $P\pi_n^{-1} = \lambda \circ K_1 \circ \cdots \circ K_n$ and $P'\pi_n^{-1} = \mu \circ K_1' \circ \cdots \circ K_n'$ for each $n \geq 1$. Due to Lemmas (5.3) and (5.5), $P \overset{\mathscr{D}}{\leq_\infty^*} P'$. Now, apply (3.1)(ii). □

The notation $\leq_n'$, and so on, is rather trying. In most applications, however, an obvious partial ordering is used; for example, when studying random sequences in $\mathbb{R}$, we normally use the partial ordering $\leq$ on $\mathbb{R}^\infty$ defined by

$$x \leq x' \quad \text{iff} \quad x_i \leq x_i' \qquad \text{for all} \quad i \geq 0$$

where $x = (x_i)_0^\infty$, $x' = (x_i')_0^\infty$.

To illuminate our theory, let $P(\cdot\,,\cdot)$ be a monotone Markov kernel in $\mathbb{R}$, that is,

$$P(x, [y, \infty)) \leq P(x', [y, \infty)) \tag{5.9}$$

$$\text{for all} \quad x, x', y \in \mathbb{R} \text{ with } x \leq x'\,.$$

Theorem (5.8) ensures the existence of Markov chains $X = (X_n)_0^\infty$ and $X' = (X_n')_0^\infty$ governed by P such that a.s.

$$X_n \leq X_n' \qquad \text{for all} \quad n \geq 0 \tag{5.10}$$

if the initial distributions λ, μ satisfy $\lambda \overset{\mathscr{D}}{\leq} \mu$. For such λ and μ, it is strongly appealing to intuition that, for example,

(5.11) $\tau_y' \overset{\mathscr{D}}{\leq} \tau_y$, for every y, where $\tau_y = \min\{n \geq 0; X_n \geq y\}$

and τ_y' is defined similarly for X',

and

$$\mathbf{P}(X_n \in \cdot\,) = \lambda P^n \overset{\mathscr{D}}{\leq} \mu P^n = \mathbf{P}(X_n' \in \cdot\,) \qquad \text{for all} \quad n \geq 0\,.$$

These properties follow immediately from (5.10).

If we rather deal with two Markov kernels $P(\cdot\,,\cdot)$ and $P'(\cdot\,,\cdot)$ on $\mathbb{R}$ such that P' dominates P, that is,

$$P(x, [y, \infty)) \leq P'(x', [y, \infty)) \quad \text{for all} \quad x, x', y \in \mathbb{R} \text{ with } x \leq x', \tag{5.12}$$

then we may construct Markov chains $X = (X_n)_0^\infty$ and $X' = (X'_n)_0^\infty$ governed by $P(\cdot, \cdot)$ and $P'(\cdot, \cdot)$, respectively, such that (5.10) holds if the initial distributions λ, μ satisfy $\lambda \overset{\mathcal{D}}{\leq} \mu$. In particular, if these kernels admit stationary distributions π and π' with $\pi \overset{\mathcal{D}}{\leq} \pi'$, we obtain stationary Markov chains X, X' such that X' dominates X if we have $X_0 \overset{\mathcal{D}}{=} \pi$, $X'_0 \overset{\mathcal{D}}{=} \pi'$.

The next result, Harris's inequality, is a key device in percolation theory. We shall make no particular use for it, but it is interesting per se, and its proof has a trick you should know. The proof is also related to that of Lemma (5.3), making this a natural place to present the inequality. Consider $\mathbb{R}^n$ endowed with the natural partial ordering:

$$x \leq x' \quad \text{iff} \quad x_i \leq x'_i \qquad \text{for all} \quad 1 \leq i \leq n\,,$$

where $x = (x_i)_1^n$, $x' = (x'_i)_1^n$, and let f and g be bounded and increasing functions on $\mathbb{R}^n$ w.r.t. that partial ordering. Further, let $X_1, \ldots, X_n$ be independent random variables and denote $(X_1, \ldots, X_n)$ by X.

(5.13) Theorem (Harris's Inequality). *The variables $f(X)$ and $g(X)$ are positively correlated, that is,*

$$\mathbf{E}[f(X) \cdot g(X)] \geq \mathbf{E}[f(X)] \cdot \mathbf{E}[g(X)]\,.$$

Proof. We use induction. The result is true for $n = 1$. Indeed, pick an X' independent of X such that $X' \overset{\mathcal{D}}{=} X$. Since

$$(f(x) - f(x')) \cdot (g(x) - g(x')) \geq 0 \qquad \text{for all} \quad x, x'$$

we get

$$\begin{aligned} 0 &\leq \mathbf{E}[(f(X) - f(X')) \cdot (g(X) - g(X')] \\ &= 2 \cdot (\mathbf{E}[f(X) \cdot g(X)] - \mathbf{E}[f(X)] \cdot \mathbf{E}[f(X)])\,, \end{aligned}$$

which is the assertion of the theorem for $n = 1$. For convenience, we refer to this as the one-dimensional case.

Now take any $n > 1$, and suppose that the inequality is established for $n - 1$. In particular, this means that

$$\begin{aligned}&\mathbf{E}[f_1(X_1, \ldots, X_{n-1}) \cdot g_1(X_1, \ldots, X_{n-1})]\\ &\quad \geq \mathbf{E}[f_1(X_1, \ldots, X_{n-1})] \cdot \mathbf{E}[g_1(X_1, \ldots, X_{n-1})],\end{aligned}$$

where $f_1(x_1, \ldots, x_{n-1}) = \mathbf{E}[f(x_1, \ldots, x_{n-1}, X_n)]$ and $g_1(x_1, \ldots, x_{n-1}) = \mathbf{E}[g(x_1, \ldots, x_{n-1}, X_n)]$, since f_1 and g_1 so defined are increasing. But $f_1(X_1, \ldots, X_{n-1}) = \mathbf{E}[f(X) \| X_1, \ldots, X_{n-1}]$ and $g_1(X_1, \ldots, X_{n-1}) = \mathbf{E}[g(X) \| X_1, \ldots, X_{n-1}]$; hence

$$\begin{aligned}\mathbf{E}[f(X) \cdot g(X)] &= \mathbf{E}[\mathbf{E}[f(X) \cdot g(X) \| X_1, \ldots, X_{n-1}]]\\ &\geq \mathbf{E}[\mathbf{E}[f(X) \| X_1, \ldots, X_{n-1}] \cdot \mathbf{E}[g(X) \| X_1, \ldots, X_{n-1}]]\\ &= \mathbf{E}[f_1(X_1, \ldots, X_{n-1}) \cdot g_1(X_1, \ldots, X_{n-1})]\\ &\geq \mathbf{E}[f_1(X_1, \ldots, X_{n-1})] \cdot \mathbf{E}[g_1(X_1, \ldots, X_{n-1})]\\ &= \mathbf{E}[\mathbf{E}[f(X) \| X_1, \ldots, X_{n-1}]] \cdot \mathbf{E}[\mathbf{E}[g(X) \| X_1, \ldots, X_{n-1}]]\\ &= \mathbf{E}[f(X)] \cdot \mathbf{E}[g(X)]\end{aligned}$$

if

$$\begin{aligned}&\mathbf{E}[f(X) \cdot g(X) \| X_1, \ldots, X_{n-1}]\\ &\quad \geq \mathbf{E}[f(X) \| X_1, \ldots, X_{n-1}] \cdot \mathbf{E}[g(X) \| X_1, \ldots, X_{n-1}].\end{aligned}$$

But that last inequality is a particular one-dimensional case. Indeed, freeze the values of $X_1, \ldots, X_{n-1}$ to realize that. □

6. Monotonicity and convergence. Our first result is a generalization of the observations in § 3. Let the state space $(E, \mathscr{E})$ be equipped with the partial ordering $\preceq$.

(6.1) Theorem. *Let $P_1, P_2, \ldots$ be a sequence of probability measures on $(E, \mathscr{E})$. The following are equivalent:*

(i) $P_1 \overset{\mathscr{D}}{\preceq} P_2 \overset{\mathscr{D}}{\preceq} \cdots$,

(ii) there exist random elements $X_1, X_2, \ldots$ in $(E, \mathscr{E})$ such that $X_i \overset{\mathscr{D}}{=} P_i$ and $X_1 \leq X_2 \leq \cdots$ a.s., and

(iii) there exist mappings $h_1, h_2, \ldots \in \mathscr{R}_{(0,1)}/\mathscr{E}$ such that $h_i(U) \overset{\mathscr{D}}{=} P_i$ and $h_1 \leq h_2 \leq \cdots$, where U is uniformly distributed on $((0,1), \mathscr{R}_{(0,1)})$.

Proof. Of course, (iii)$\Rightarrow$(ii)$\Rightarrow$(i). To prove (i)$\Rightarrow$(iii), use (3.1)(iii) to find upward kernels K_i, $i \geq 2$, such that $P_{i-1}K_i = P_i$. Let $Q_n = P_1 \circ K_2 \circ \cdots \circ K_n$. Kolmogorov's consistency theorem provides a probability measure P on $(E_1^\infty, \mathscr{E}_1^\infty)$ such that $P\pi_n^{-1} = Q_n$. (Here, $E_1^\infty = \{(x_i)_1^\infty;\ x_i \in E\}$ and π_n: $E_1^\infty \to E^n$ is given by $(x_1, x_2, \ldots) \to (x_1, \ldots, x_n)$). Certainly, $P\{(x_i)_1^\infty;\ x_1 \leq x_2 \leq \cdots\} = 1$. Use Skorohod's representation again, to find $h_i \in \mathscr{R}_{(0,1)}/\mathscr{E}$ such that $(h_1(U), h_2(U), \ldots) \overset{\mathscr{D}}{=} P$; that $h_1 \leq h_2 \leq \cdots$ follows from this (possibly after a null-set modification). □

For the next results, assume that $\leq$ is a partial ordering (closed, as usual). Let d be a complete and separable metric generating $\mathscr{E}$.

(6.2) Theorem. *Let $X_1, X_2, \ldots$ be random elements in $(E, \mathscr{E})$ defined on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ such that $X_1 \leq X_2 \leq \cdots$ a.s., and $(X_n)_1^\infty$ is tight. Then $(X_n)_1^\infty$ is convergent in probability.*

Proof. Fix $\epsilon > 0$. Due to the tightness, there exists a compact set $K \subset E$ such that $\mathbf{P}(X_n \in K) > 1 - \epsilon$ for all n. Let

$$A = \{\omega;\ X_n(\omega) \in K \text{ i.o.}\}\ .$$

Certainly, $\mathbf{P}(A) > 1 - \epsilon$. Let $S(\omega) = \{n;\ X_n \in K\}$. We shall prove that $(X_n(\omega))_{n \in S(\omega)}$ is convergent if $\omega \in A$, by showing that the sequence has exactly one limit point. The compactness of K yields one such point. Suppose that there is more than one, $X'(\omega)$ and $X''(\omega)$, say. But a moment's thought, using the monotonicity of $(X_n(\omega))_1^\infty$ and the closedness of $\leq$, gives that $X'(\omega) \leq X''(\omega)$ and $X''(\omega) \leq X'(\omega)$, hence $X'(\omega) = X''(\omega)$. Call that common value $X(\omega)$. We get

$$\mathbf{P}(d(X_n, X) > \epsilon) \leq \mathbf{P}(A^c) + \mathbf{P}(A, d(X_n, X) > \epsilon \text{ and } X_n \in K) + \mathbf{P}(X_n \not\in K) < 3\epsilon$$

for $n \geq$ some n_ϵ, since the second probability obviously tends to 0 as $n \to \infty$. The convergence in probability is proved. □

We were close to an a.s. convergence result. With the extra condition

$$(6.3) \qquad x_n \leq y_n \leq x_{n+1}, \qquad x_n \to x \Rightarrow y_n \to x$$
$$\text{for sequences } (x_n)_1^\infty, (y_n)_1^\infty \text{ in } E,$$

we have that stronger convergence.

(6.4) Corollary. *Let $P_1, P_2, \ldots$ be probability measures on $(E, \mathscr{E})$ such that $P_1 \overset{\mathscr{D}}{\leq} P_2 \overset{\mathscr{D}}{\leq} \cdots$ and $(P_n)_1^\infty$ is tight. Then $(P_n)_1^\infty$ is weakly convergent.*

Proof. Use (6.1)(ii) to find a sequence $(X_n)_0^\infty$ such that $X_n \overset{\mathscr{D}}{=} P_n$ and $X_1 \leq X_2 \leq \cdots$ a.s. Due to Theorem (6.2), $X_n \to$ some X in probability as $n \to \infty$. This implies that $P_n \Rightarrow$ the distribution of X as $n \to \infty$. □

For a continuous-time version of (6.4), note that tightness implies that each sequence (P_{t_k}) has a weakly convergent subsequence, and apply the corollary to understand that $(P_t)_0^\infty$ can have no more than one weak limit point.

7. Notes. This section is based on Kamae, Krengel, and O'Brien [84]. That reference, and Kamae and Krengel [83], also contain continuous-time theory analogous to the one given here. Harris's inequality was proved in its first version in Harris [72]. For all the details concerning the condition (6.3) for a.s. convergence in (6.2), see [84].

3. DOMINATION AND MONOTONICITY OF MARKOV PROCESSES

8. Basics. Recall the notation of § III.20. In the following sections we shall see several examples of pairs of Markov processes X, X' such that if the initial distributions λ and μ satisfy $\lambda \overset{\mathscr{D}}{\leq} \mu$, there exists a coupling $(\hat{X}, \hat{X}')$ with the property

$$(8.1)\qquad \hat{X}_t \le \hat{X}'_t \quad \text{for all} \quad t \ge 0\,.$$

Here $X = (X_t)_0^\infty$ and $X' = (X'_t)_0^\infty$ are governed by semigroups $(P_t)_0^\infty$ and $(P'_t)_0^\infty$, respectively, are D_E-valued as usual and have a state space $(E, \mathscr{E})$ equipped with a partial ordering $\le$.

In Part 2, no general theory was developed to produce a continuous-time analog of Theorem (5.8); rather, (8.1) will be established by using special properties of the processes under investigation. Recall the second example of § Int.1, where (8.1) is immediate from the skip-free property of birth and death processes. Notice that a necessary condition for the existence of a coupling satisfying (8.1) for any λ, μ such that $\lambda \overset{\mathscr{D}}{\le} \mu$ is that P_t is dominated by P'_t for all $t \ge 0$ [cf. (5.1)]. Indeed, if $x \le x'$, then $\delta_x \overset{\mathscr{D}}{\le} \delta_{x'}$, and if $\hat{X}$, $\hat{X}'$ satisfy (8.1) with $\hat{X}_0 = x$, $\hat{X}'_0 = x'$, we get $P_t(x, \cdot) \overset{\mathscr{D}}{\le} P'_t(x', \cdot)$ from $\hat{X}_t \le \hat{X}'_t$.

Most often (8.1) will be established and used when $(P_t)_0^\infty$ and (P'_t) agree; then X and X' are realizations of the same Markov processes with different initial conditions. We will prove domination and monotonicity properties as complements to our ergodicity results for Markov processes. For a measurable space $(E^\circ, \mathscr{E}^\circ)$, consider mappings $\psi \in \mathscr{D}_E / \mathscr{E}^\circ$ such that

(8.2) (i) $x \le^* x' \Rightarrow \psi(x) \le^\circ \psi(x')$, or

(ii) $x \le^* x' \Rightarrow \psi(x') \le^\circ \psi(x)$

for all $x, x' \in D_E$, where $\le^\circ$ is a partial ordering of E° and $\le^*$ is the natural partial ordering of D_E given by $x \le^* x'$ if and only if $x_t \le^\circ x'_t$ for all $t \ge 0$, where $x = (x_t)_0^\infty$, $x' = (x'_t)_0^\infty$. If (i) [(ii)] holds, then ψ is called increasing [decreasing]. If $(\hat{X}, \hat{X}')$ is a coupling satisfying (8.1), then of course

$$(8.3)\qquad \begin{aligned} &\psi(\hat{X}) \le^\circ \psi(\hat{X}') \quad \text{if } \psi \text{ is increasing, and} \\ &\psi(\hat{X}') \le^\circ \psi(\hat{X}) \quad \text{if } \psi \text{ is decreasing.} \end{aligned}$$

9. A monotonicity result. Now suppose that there exists an $a \in E$ such that

$$(9.1)\qquad a \le x \quad \text{for all} \quad x \in E\,.$$

Of course, this minorization implies that

(9.2) $$\delta_a \overset{\mathcal{D}}{\leq} \mu \qquad \text{for all} \quad \mu .$$

(9.3) Theorem. *Suppose that the semigroup $(P_t)_0^\infty$ is such that a coupling satisfying (8.1) exists for $\lambda = \delta_a$ and any μ. Let X be governed by $(P_t)_0^\infty$ and such that $X_0 \equiv a$. Then $\psi(\theta_t X)$ is stochastically increasing (decreasing) in t if $\psi \in \mathcal{D}_E/\mathscr{E}^\circ$ is increasing (decreasing).*

Proof. We pay attention only to ψ increasing. Let $f \in ib\mathscr{E}^\circ$. For s, t such that $0 \leq s \leq t$, we shall prove that

(9.4) $$\mathbf{E}_a[f(\psi(\theta_s X))] \leq \mathbf{E}_a[f(\psi(\theta_t X))] .$$

Now since $\theta_t X = \theta_s(\theta_{t-s} X)$ and $\psi(\theta_s \cdot)$ is increasing if ψ is, it suffices to establish (9.4) for $s = 0$. We have

$$\mathbf{E}_a[f(\psi(\theta_t X))] = \mathbf{E}_a[\mathbf{E}[f(\psi(\theta_t X)) \| X_t]] = \mathbf{E}_\mu[f(\psi(X))] ,$$

where $\mu = \mathbf{P}_a(X_t \in \cdot\,) = \delta_a P_t$. Now let $\hat{X}$ be the first component of a coupling $(\hat{X}, \hat{X}')$ of X and X', with $X_0' \overset{\mathcal{D}}{=} \mu$, satisfying (8.1). Since $f \circ \psi \in ib\mathcal{D}_E$ we certainly have

$$\mathbf{E}_\mu[f(\psi(X))] = \mathbf{E}_{a,\mu}[f(\psi(\hat{X}'))] \geq \mathbf{E}_{a,\mu}[f(\psi(\hat{X}))] = \mathbf{E}_a[f(\psi(X))]$$

and we are done. $\square$

The simplest increasing mapping ψ of interest is

$$\psi(x) = x_0 \qquad \text{for} \quad x = (x_t)_0^\infty \in D_E .$$

For that, $\psi(\theta_t x) = x_t$, and under the conditions of Theorem (9.3) we obtain that X_t is stochastically increasing [as a random element in $(E, \mathscr{E})$] in t. A consequence of this is that any birth and death process $X = (X_t)_0^\infty$ starting from 0 is stochastically increasing, with $E = \mathbb{Z}_+$ naturally ordered. However, we have not yet rigorously proved that (Int.1.4) means a coupling satisfying (8.1); that will be done in § V.2.

A close look at the proof of Theorem (9.3) reveals that

(9.5) if

(i) (8.1) holds,

(ii) we have a minimal element a according to (9.1), and $\psi \in \mathcal{D}_E/\mathscr{E}^\circ$ is such that $\mathbf{P}_x(\psi(X) \in \cdot)$ is stochastically increasing (decreasing) in $x \in E$,

then $\psi(\theta_t X)$ is stochastically increasing (decreasing) in t if $X_0 = a$.

For the proof, make use of the fact that X_t is stochastically increasing in t. As will be seen, there are cases where ψ is not monotone but where it is of interest to know that $\psi(\theta_t X)$ is stochastically increasing or decreasing. In these, it is sometimes possible to construct couplings sharp enough to prove condition (9.5).

4. EXAMPLES OF DOMINATION

10. Direct constructions. Several applications of Strassen's theorem and the results in Part 2 will appear in later sections. Here we point out a number of examples of domination that have no other natural place.

Strassen's theorem is a result with a complicated proof. Hence it is important to realize that in many situations a direct and often simple coupling may be brought about. Actually, we have already seen an example of that: the coupling of birth and death processes in § Int.1, which produced the inequalities (Int.1.7) and (Int.1.8).

We show first how couplings may be instrumental for establishing a.s. results. Let $Y_1, Y_2, \ldots$ be independent random variables which are majorized in the sense that there exists a distribution F such that $Y_i \overset{\mathcal{D}}{\leq} F$ for all i. If F has a finite expectation μ, then

$$(10.1) \qquad \limsup_{n\to\infty} S_n/n \leq \mu \quad \text{a.s.},$$

where $S_n = \Sigma_1^n Y_i$, and if F also has a finite second moment, hence a variance $\sigma^2 < \infty$, then

(10.2) $\limsup_{n\to\infty} (S_n - n\cdot\mu)/(2\sigma^2\cdot n\cdot \log\log n)^{1/2} \le 1$ a.s.

To prove that, just produce independent random pairs (Y_i, Y'_i), $i \ge 1$, such that $Y'_i \stackrel{\mathcal{D}}{=} F$ and $Y_i \le Y'_i$ a.s. for all i (cf. § 1). Then, with $S'_n = \Sigma_1^n Y'_i$, $\lim_{n\to\infty} S'_n/n = \mu$ a.s. due to the strong law of large numbers for i.i.d. variables, and (10.1) follows:

$$\mathbf{P}(\limsup_{n\to\infty} S_n/n \le \mu) \ge \mathbf{P}(\limsup_{n\to\infty} S'_n/n \le \mu) = 1 .$$

Using the law of the iterated logarithm instead of the law of large numbers, we obtain (10.2)

The case of 0–1 variables is particularly simple. Let U be uniformly distributed on $[0, 1]$, and

(10.3) $$Y = I_{[0,p]}(U) .$$

Then $Y \stackrel{\mathcal{D}}{=} \mathrm{Ber}(p)$. For a sequence $Y_1, Y_2, \ldots$ of independent such variables with p-values $p_1, p_2, \ldots$ we shall prove that

(10.4) $$\limsup_{n\to\infty} S_n/n \le \rho^* \quad \text{a.s.} ,$$

where $S_n = \Sigma_1^n Y_i$ and $\rho^* = \limsup_{i\to\infty} p_i$. For that, let $U_1, U_2, \ldots$ be independent and Uni$[0, 1]$-distributed. Take an $\epsilon > 0$, let $p = \rho^* + \epsilon$ and

(10.5) $$Y_i = I_{[0,p_i]}(U_i),\ Y'_i = I_{[0,p]}(U_i) .$$

With $S'_n = \Sigma_1^n Y'_i$, we have

$$S_n/n \le S'_n/n + \#\{i \le n;\ p_i > p\}/n .$$

But the set involved is finite, hence

$$\limsup_{n\to\infty} S_n/n \le \limsup_{n\to\infty} S'_n/n = p .$$

Since $\epsilon > 0$ is arbitrary, (10.4) follows. Of course,

$$\liminf_{n\to\infty} S_n/n \ge \rho_* \quad \text{a.s.} ,$$

where $\rho_* = \liminf_{i\to\infty} p_i$, so S_n/n is a.s. convergent if $\rho_* = \rho^*$ [i.e., when the sequence $(p_i)_1^\infty$ is convergent].

We also use (10.5) for a central limit result. For $0<p<1$, it has been known for some time now (a number of centuries) that $(S'_n - np)/(np(1-p))^{1/2} \xrightarrow{\mathcal{D}} N(0,1)$ as $n\to\infty$. But how much can we allow the p_i values to deviate from p and still have

$$(10.6) \qquad (S_n - np)/(np(1-p))^{1/2} \xrightarrow{\mathcal{D}} N(0,1)?$$

The last convergence holds if $(S_n - S'_n)/n^{1/2} \xrightarrow{P} 0$ as $n\to\infty$, which in its turn is true if $\mathbf{E}[|S_n - S'_n|/n^{1/2}] \to 0$ as $n\to\infty$. But

$$\mathbf{E}[|S_n - S'_n|] \le \sum_1^n \mathbf{E}[|Y_i - Y'_i|] = \sum_1^n \mathbf{P}(Y_i \ne Y'_i) = \sum_1^n |p_i - p|;$$

hence (10.6) holds if

$$(10.7) \qquad \sum_1^n |p_i - p| = o(n^{1/2}) \qquad \text{as} \quad n\to\infty.$$

11. Percolation. We continue in this and the following sections to use independent Uni[0, 1]-distributed variables, now to establish monotonicity results. For the first example, consider a random graph (not oriented) with nodes = the integer points $\mathbb{Z}^2$ in the plane, and edges possible only between neighboring nodes (so there are at most four edges with endpoints at a particular node).

We shall call such an edge a channel and suppose that there exists such a channel between any two neighboring nodes with constant probability p. Let $C(\{\mathbf{i},\mathbf{j}\})$ be 1 if there is a channel between $\mathbf{i}$ and $\mathbf{j}$, and 0 if not. The variables $C(\{\mathbf{i},\mathbf{j}\})$ are supposed to be independent as $\{\mathbf{i},\mathbf{j}\}$ ranges the class of neighboring nodes; let that class be denoted by M.

Now let water flow to node $\mathbf{0}$ from an exterior source. The classical problems of planar percolation theory are: Can we find a $p<1$ large enough so that the plane is not flooded with probability 1, that is, that the part of the channel system containing $\mathbf{0}$ is infinite with a strictly positive probability? If so, what is the lower bound for those p values?

Let

$$\theta(p) = \mathbf{P}_p(|C_0| = \infty)$$

where C_0 represents the class of nodes that will be reached from $\mathbf{0}$. No one doubts that θ is an increasing function, so the lower bound mentioned can be defined by

$$p_c = \inf\{p; \theta(p) > 0\} .$$

But how do we prove that θ is increasing? Let $U_{\{\mathbf{i},\mathbf{j}\}}$, $\{\mathbf{i},\mathbf{j}\} \in M$, be independent and Uni[0, 1]-distributed; a standard product space is used to construct these variables. Then let

$$C(\mathbf{i},\mathbf{j}) = I_{[0,p]}(U_{\{\mathbf{i},\mathbf{j}\}})$$

for $\{\mathbf{i},\mathbf{j}\} \in M$. Certainly, there becomes more and more channels available as p increases, hence θ is increasing.

To be honest: the purpose of this paragraph was to raise your interest for percolations. The sharp result that $p_c = \frac{1}{2}$ calls for your attention!

12. Bernstein polynomials. The next topic is hopefully unexpected. For a continuous function f on $[0, 1]$, let f_n (often denoted by $B_n f$ in the literature) for $n \geq 1$ be the associated Bernstein polynomial:

$$f_n(x) = \sum_{i=0}^{n} f(i/n) \cdot \binom{n}{i} x^i (1-x)^{n-i} . \tag{12.1}$$

Bernstein introduced these polynomials to give a new proof of the Weierstrass approximation theorem based on the weak law of large numbers. Indeed, we have

$$f_n(x) = \mathbf{E}[f(S_n/n)] ,$$

where $S_n \stackrel{\mathscr{D}}{=} \mathrm{Bin}(n, x)$, and the proof of $f_n \to f$ uniformly is easy. Several tools from probability theory may be used to illuminate that approximation. Here, we shall use the representation (10.3) to prove that if f is increasing, so is f_n. But that is immediate: Let $0 \leq x \leq y \leq 1$, and let $Y_1, Y_2, \ldots, Y_1', Y_2', \ldots$ be as above, with p-values x and y, respectively. Since $S_n/n \leq S_n'/n$ again, we have

$$f_n(x) = \mathbf{E}[f(S_n/n)] \leq \mathbf{E}[f(S_n'/n)] = f_n(y) .$$

13. Increasing power functions. Using a random variable X that has a $\mathrm{Bin}(n, p)$ distribution, we wish to test the hypothesis

$$H\colon p \le p_0 \quad \text{against} \quad K\colon p > p_0 \tag{13.1}$$

at the level of significance α. Test theory tells us that H shall be rejected if $X \ge k_\alpha$, where k_α is the smallest number k such that

$$\mathbf{P}_{p_0}(X \ge k) \le \alpha\,. \tag{13.2}$$

We omit the possibility of introducing randomized tests. It is desirable, and natural to hope, that the power function of this test

$$\beta(p) = \mathbf{P}_p(X \ge k_\alpha) \tag{13.3}$$

is increasing as p ranges from 0 to 1. And that follows as in the examples of the preceding section immediately from (10.3); let $U_1, \ldots, U_n$ be independent and uniformly distributed on $[0, 1]$, $Y_1, \ldots, Y_n$ and $Y'_1, \ldots, Y'_n$ given by (10.3) with parameter values p and p' respectively. Then $X = \Sigma_1^n Y_i$, $X' = \Sigma_1^n Y'_i$ have $\mathrm{Bin}(n, p)$, $\mathrm{Bin}(n, p')$ distributions, $X \le X'$, and we obtain

$$\beta(p) = \mathbf{P}(X \ge k_\alpha) \le \mathbf{P}(X' \ge k_\alpha) = \beta(p')\,. \tag{13.4}$$

Resources may be saved testing H of (13.1) by use of the statistic

$$Z = \min\{\, j;\ Y_i = 1 \text{ for } m \text{ out of the } j \text{ variables } Y_1, \ldots, Y_j\}\,, \tag{13.5}$$

where m is a fixed integer ≥ 1. We reject H if $Z \le k_\alpha$, where k_α is the largest k such that $\mathbf{P}_{p_0}(Z \le k) \le \alpha$. To prove that the power function of that test is increasing, extend the sequence $U_1, U_2, \ldots$ above to an infinite one, to get variables Y_i, Y'_i for each $i \ge 1$. Letting Z' be the analog of Z for the Y'_j sequence, we certainly have

$$\beta(p) = \mathbf{P}(Z \le k_\alpha) \le \mathbf{P}(Z' \le k_\alpha) = \beta(p')$$

since $Z' \le Z$.

As a third case, consider a lot containing N items, out of which d are defective. If a sample of size n is selected at random (each subset equally likely), the number X of defective items in that sample has a hypergeometric distribution, that is,

(13.6)

$$\mathbf{P}_d(X=i)=\binom{d}{i}\binom{N-d}{n-i}\Big/\binom{N}{n}, \qquad i=0,\ldots,\min(n,d),$$

as you know. We reject

(13.7) $$H: d \le d_0 \quad \text{in favor of} \quad K: d > d_0$$

at the α level if $X \ge$ an appropriate k_α. To prove that $\mathbf{P}_d(X \ge k_\alpha)$ is increasing in d, we should avoid using (13.6). Rather, imagine one of the nondefective items to be defective, and let X' be the number of defective items in the sample, possibly including that imaginary item. Of course, $X \le X'$ and we may conclude that $\mathbf{P}_d(X \ge k_\alpha)$ increases if d is changed to $d+1$.

For the last example, consider a Poisson process N on $[0,\infty)$, that is, a point process such that, for all sets $A, B \in \mathcal{R}_+$, $N(A)$ and $N(B)$ are independent whenever A and B are disjoint, and $N(A)$ has a Poisson distribution. If there exists a function λ such that the parameter of that Poisson distribution is

$$\Lambda(A)=\int_A \lambda(t)\,dt,$$

we call that the intensity function of the process. If we intend to test

(13.8) $$H: \lambda=\lambda_0 \quad \text{against} \quad K: \lambda \ge \lambda_0,\ \lambda \ne \lambda_0$$

by observing the process during the time interval $[0, t_0]$, we reject H if $N_{t_0} \ge k_\alpha$ for $k_\alpha =$ the smallest j such that

$$\sum_{i=j}^{\infty} p_\gamma(i) \le \alpha,$$

where $\gamma=\int_0^{t_0}\lambda_0(t)\,dt$, $p_\gamma(i)=\exp(-\gamma)\cdot\gamma^i/i!$.

To prove that the probability for rejection increases with λ, let N, N'' be independent Poisson processes with intensity functions λ_0, $\lambda - \lambda_0$, respectively, and $N' = N + N''$. Then N' has an intensity function λ, and since $N \le N'$ we get $\mathbf{P}(N_{t_0} \ge k_\alpha) \le \mathbf{P}(N'_{t_0} \ge k_\alpha)$.

With the exception of (13.5), we have considered fixed sample sizes only. However, the idea of constructing couplings with a.s. domination also has consequences for sequential methods. When using such, it is typical that a hypothesis is rejected if and when a sequence of statistics crosses the (a) border of a domain. In the four cases considered above, sequential tests in which rejection takes place if and when an upper border is crossed may be constructed. Using the couplings introduced, we obtain, when testing the H of (13.1), for example, that $\tau' \le \tau$ a.s. if $p \le p'$, where τ, τ' are the sample sizes and p, p' are the parameter values.

14. Cox processes. Now consider a point process ξ on $\mathbb{R}$ generated as follows. Let Λ be a random measure on $\mathbb{R}$, that is, a random element in $(\mathcal{M}, \mathcal{B})$, the space of measures on $(\mathbb{R}, \mathcal{R})$ giving finite mass to bounded sets. That Λ plays the role of a random parameter for ξ, to the effect that ξ conditioned on Λ is a Poisson process with expectation measure Λ_ω. This means that

$$\mathbf{P}(\xi(A) = k) = \int p_y(k)\, d[\mathbf{P}\Lambda^{-1}](y)$$

for $A \in \mathcal{R}$. This was a sketchy definition of a Cox process ξ.

Now endow $\mathcal{M}$ with the natural partial ordering: $\nu_1 \le \nu_2$ if and only if $\nu_1(A) \le \nu_2(A)$ for all $A \in \mathcal{R}$. If ξ and ξ' are two Cox processes governed by Λ and Λ', respectively, it should be true that

$$\Lambda \overset{\mathcal{D}}{\le} \Lambda' \quad \text{implies} \quad \xi \overset{\mathcal{D}}{\le} \xi' .$$

And it is. Indeed, use Strassen's theorem to obtain a coupling $(\hat{\Lambda}, \hat{\Lambda}')$ of (Λ, Λ') such that $\hat{\Lambda} \le \hat{\Lambda}'$. Notice that if Λ and Λ' happen to have random intensities λ and λ' such that $\lambda_s \le \lambda'_s$ for all s, Strassen's theorem is not needed: We then have $\Lambda \le \Lambda'$ directly.

Now the Cox processes $\hat{\xi}$ and $\hat{\xi}'$ represent a coupling of ξ and ξ': for all $A \in \mathcal{R}$ and $k \ge 0$,

$$\hat{\xi}'(A) = \hat{\xi}(A) + \eta(A) ,$$

where η is a Cox process governed by $\hat{\Lambda}' - \hat{\Lambda}$. Hence $\hat{\xi} \leq \hat{\xi}'$, and $\xi \overset{\mathcal{D}}{\leq} \xi'$ is proved.

We chose the space $(\mathbb{R}, \mathcal{R})$ for the sake of concreteness. However, everything said makes perfect sense for Cox processes on any Polish space. You will see more of point process domination in Chapter V.

15. Notes. O'Brien [125] and Guttorp, Kulperger, and Lockhart [70] are the bases for § 10. See Durrett [57], Grimmett [68], and Kesten [92, 93] for percolations. Probabilistic methods to establish properties of Bernstein polynomials are surveyed in Lindvall [109]. Monotonicity of power functions is usually proved by use of monotone likelihood ratios, cf. Lehmann [103]. In § 3.11 of [103] a coupling is used in a sequential test context. The details concerning Cox process, see Daley and Vere-Jones [45].

For other uses of Uni[0, 1] variables, see Hodges and Rosenblatt [74] (inequalities for a hitting time of a random walk), Harris [71] (representation of certain finite-state random sequences), and Brandt, Lisek, and Nerman [29] (representation of stationary discrete Markov chains). Couplings are also used in Cambanis and Simons [33] to obtain inequalities $\mathbf{E}[f(Z)] \leq \mathbf{E}[f'(Z')]$ for certain pairs of functions f, f' and normally or elliptically distributed variables Z, Z' in $\mathbb{R}^d$, $d \geq 2$. See also Dubins and Meilijson [54].

CHAPTER V

Intensity-Governed Processes

1. BIRTH AND DEATH PROCESSES

1. Basics. For a Markov process $X = (X_t)_0^\infty$ with a countable state space E, the governing semigroup P_t, $t \geq 0$, consists of matrices $(p_{ij}(t))_{i,j \in E}$. We content ourselves with a brief setup since there are thorough accounts available.

For $i, j \in E$, $i \neq j$, let $q_i \geq 0$, $q_{ij} \geq 0$ be such that

$$(1.1) \qquad q_i = \sum_{j \neq i} q_{ij}$$

for all i. The paths of X are constructed as follows: After entering (or starting in) a state i, X stays there for a time that is $\operatorname{Exp}(q_i)$-distributed. At the end of such a sojourn, X jumps to state $j \neq i$ with probability q_{ij}/q_i, independent of the history of X up to that jump time.

We shall denote these jump times by $\sigma_1, \sigma_2, \ldots$, let $\sigma_0 = 0$, and assume that sufficient conditions for nonexplosiveness ($\sigma_n \to \infty$ a.s.) are satisfied. Letting the paths be constant on the intervals $[\sigma_i, \sigma_{i+1})$, $i \geq 0$, X becomes D_E-valued and satisfies for $j \neq i$,

$$(1.2) \qquad \begin{aligned} p_{ij}(h) &= \mathbf{P}_\lambda(X_{t+h} = j \| X_t = i) = q_{ij} \cdot h + o(h), \quad \text{and} \\ 1 - p_{ii}(h) &= \mathbf{P}_\lambda(X_{t+h} \neq i \| X_t = i) = q_i \cdot h + o(h) \end{aligned}$$

as $h \to 0$; the notation $\mathbf{P}_\lambda$ means that $X_0 \stackrel{\mathscr{D}}{=} \lambda$. The relations in (1.2) justify what we call q_i and q_{ij} the transition intensities from i and from i to j, respectively. If $q_i = 0$ for a state i, that is an absorbing state. If X hits such a state, the subsequent σ_j times are all set to equal ∞ [notice the consistency with our definition of the Exp(0) distribution].

The process X is irreducible if for all $i, j \in E$ there exists a $t_{ij} > 0$ such that $p_{ij}(t_{ij}) > 0$. Actually, X is such if and only if $p_{ij}(t) > 0$ for all $i, j \in E$ and $t > 0$. If X is irreducible, it has a stationary distribution π if and only if it is positive recurrent; then π is unique and satisfies

$$\pi Q = 0, \tag{1.3}$$

where $Q = (q_{ij})_{i,j \in E}$ is the intensity matrix ($q_{ii} = -q_i$).

A birth and death process has a state space $E \subset \mathbb{Z}_+$ and takes jumps of size 1 or -1 only. For a state i, the possibly nonvanishing intensities are then $q_{i,i+1}$ and $q_{i,i-1}$; these are the birth and death intensities, denoted by β_i and δ_i, respectively. A simple and useful sufficient condition for X (always denoting a birth and death process in this section) to be nonexplosive is

$$\sum_{i \in E} 1/\beta_i = \infty. \tag{1.4}$$

To understand that, dominate X by a pure birth process (all $\delta_i = 0$) with intensities β_i and use the fact that $\Sigma\, Z_i = \infty$ a.s. if Z_i, $i \in E$, are independent and Exp(β_i) distributed, with the sum (1.4) infinite. The usefulness of the criterion (1.4) comes from the fact that in most birth and death models, we have $\beta_i \le C \cdot (1 + i)$.

For the state space, those β_i and δ_i that are strictly positive decide what the choice should be. We shall restrict our attention to

(1.5) (i) $E = \mathbb{Z}_+$, or
(ii) $E = \{0, 1, \ldots, m\}$ for some $m \ge 1$.

In case (i), X is irreducible if $\beta_i > 0$ for $i \ge 0$ and $\delta_i > 0$ for $i \ge 1$. As you know, we then have recurrence if and only if

$$\sum_{n=1}^{\infty} \delta_1 \cdot \ldots \cdot \delta_n / \beta_1 \cdot \ldots \cdot \beta_n = \infty$$

and positive recurrence (hence the existence of a unique stationary distribution π) if and only if

$$\sum_{n=1}^{\infty} \beta_0 \cdot \ldots \cdot \beta_{n-1}/\delta_1 \cdot \ldots \cdot \delta_n < \infty ,$$

and $\pi = (\pi_n)_0^\infty$ may, if this holds, be calculated through

$$(1.6) \qquad \pi_n = \pi_0 \cdot (\beta_0 \cdot \ldots \cdot \beta_{n-1}/\delta_1 \cdot \ldots \cdot \delta_n) \qquad \text{for} \quad n \geq 1 .$$

In the finite state space (1.5)(ii) (now $\beta_m = 0$), all irreducible X are positive recurrent, and π is obtained by (1.6), with $1 \leq n \leq m$.

The well developed theory of birth and death processes, in one or several dimensions, should be considered a major achievement in probability. Whatever your inclination may be (for basic, animated, applicable, or applied probability), it has much to offer.

2. Ergodicity. Recall the second example of § Int.1. The skip-freeness of paths was used there to convince us that in the recurrent case we always have

$$(2.1) \qquad \|\lambda P_t - \mu P_t\| \to 0 \qquad \text{as} \quad t \to \infty$$

for any initial distributions λ, μ, and if we have positive recurrence, then

$$(2.2) \qquad \|\lambda P_t - \pi\| \to 0 \qquad \text{as} \quad t \to \infty$$

for any initial λ; π is the unique stationary distribution.

To complete the proof of (2.1)–(2.2), we must show that X'' defined by (Int.1.4) satisfies $X'' \stackrel{\mathscr{D}}{=} X$. But that follows from the last remark of § III.20: see (III.20.12) and what follows. Notice that all Markov processes with a countable state space always have the Feller property; hence the strong Markov property consideration is easy.

Define τ_0 by

$$\tau_0 = \inf\{s \geq 0; X_s = 0\}$$

(i.e., the hitting time of the state 0), and τ_0' analogously in terms of X'. We have

(2.3) $$T \le \max(\tau_0, \tau_0') \quad \text{a.s.}$$

That follows if we establish that the sets of jump times $\{\sigma_1, \sigma_2, \dots\}$ and $\{\sigma_1', \sigma_2', \dots\}$ are disjoint a.s. But

(2.4) $$\begin{aligned} &\mathbf{P}_{\lambda\mu}(\{\sigma_1, \sigma_2, \dots\} \cap \{\sigma_1', \sigma_2', \dots\} \ne \emptyset) \\ &\qquad = \int \mathbf{P}_{\nu\mu}(\{t, t+\sigma_1, \dots\} \cap \{\sigma_1', \sigma_2', \dots\} \ne \emptyset) h(t)\, dt, \end{aligned}$$

where ν is the distribution of X_{σ_1} and h the density of the distribution of σ_1 under $\mathbf{P}_\lambda$ (a mixture of exponential distributions). But for any outcome of $\sigma_1, \sigma_2, \dots, \sigma_1', \sigma_2', \dots$, under $\mathbf{P}_{\lambda\mu}$, the set $\{t, t+\sigma_1, \dots\} \cap \{\sigma_1', \sigma_2', \dots\}$ is nonempty for at most countably many t values. Hence the probability in (2.4) equals 0. For (2.1)–(2.2) a weak coupling suffices. But to establish inequalities such as (Int.1.8), a strong coupling is required; we postpone examples to § 3.

A Doeblin coupling may also be obtained analogously to (II.8.9). Let $\tilde{X} = (\tilde{X}_t)_0^\infty = ((X_t, X_t'))_0^\infty$ be a jump process with state space $\mathbb{Z}_+^2$ governed by transition intensities

(2.5) $$\begin{aligned} q_{(i,j)(i+1,j)} &= \beta_i, \qquad & q_{(i,j)(i-1,j)} &= \delta_i, \\ q_{(i,j)(i,j+1)} &= \beta_j, \qquad & q_{(i,j)(i,j-1)} &= \delta_j, \end{aligned}$$

if $i \ne j$, and

$$q_{(i,i)(i+1,i+1)} = \beta_i \qquad q_{(i,i)(i-1,i-1)} = \delta_i$$

for $i, j \in \mathbb{Z}_+$. Then X and X' are indeed birth and death processes, with the right intensities.

In this book we pay no attention to ergodicity of transient birth and death processes (or diffusions), but refer the reader to the rather extensive literature. In fact, for both birth–death and diffusion processes, ergodicity is equivalent to the Doeblin coupling being successful.

3. Rates. We shall now exploit (2.3) to obtain two rate results. As usual, we denote the initial distributions of the independent processes X and X' by λ and μ. To prove $\mathbf{E}_{\lambda\mu}[T^\alpha]<\infty$ for an $\alpha \geq 1$, it is natural first to observe that

$$T \leq \tau_0 + \tau_0'$$

and

$$\mathbf{E}_{\lambda\mu}[T^\alpha] \leq C \cdot (\mathbf{E}_\lambda[\tau_0^\alpha] + \mathbf{E}_\mu[\tau_0'^\alpha]) ,$$

and then search for simple conditions in terms of $(\beta_i)_1^\infty$ and $(\delta_i)_1^\infty$ for $\mathbf{E}_\lambda[\tau_0^\alpha]<\infty$ to hold. Unfortunately, such conditions do not exist. We then simplify matters by introducing the condition

$$\inf_{i \geq 1} q_i = c_0 > 0 \tag{3.1}$$

($q_i = \beta_i + \delta_i$), and considering the embedded Markov chain $Z = (Z_n)_0^\infty$ where $Z_n = X_{\sigma_n}$, $n \geq 0$. That chain is a Harris random walk [cf. II.16] with $p_i = \beta_i/q_i$ and $1 - p_i = \delta_i/q_i$. Let

$$\eta = \min\{i; Z_i = 0\} .$$

Then $\mathbf{E}_\lambda[\tau_0^\alpha]<\infty$ if $\mathbf{E}_\lambda[\eta^\alpha]<\infty$ for an $\alpha \geq 1$. Indeed,

$$\tau_0 \overset{\mathscr{D}}{\leq} \sum_1^\eta U_i , \tag{3.2}$$

where $U_1, U_2, \ldots$ are i.i.d. Exp(c_0)-distributed variables. Condition on η and use Minkowski's inequality to get

$$\mathbf{E}_\lambda[\tau_0^\alpha] \leq \mathbf{E}_\lambda[U_1^\alpha] \cdot \mathbf{E}_\lambda[\eta^\alpha] < \infty .$$

But also to establish a condition for $\mathbf{E}_1[\eta^2]<\infty$, for example, is disappointingly complicated. However, when we have a uniform bound for p_i,

$$\sup_{i \geq 1} p_i = c_1 \leq \tfrac{1}{2} , \tag{3.3}$$

examining τ_0 via η is a fruitful way to obtain rate results for $\|\lambda P_t - \mu P_t\|$. We state them in terms of two theorems, for the cases $c_1 < \frac{1}{2}$ and $c_1 = \frac{1}{2}$, respectively.

(3.4) Theorem. *Assume that* $c_0 > 0$ *and* $c_1 < \frac{1}{2}$. *If* $\Sigma_i\, a^i\lambda_i$ *and* $\Sigma_i\, a^i\mu_i$ *are finite, where* $a = ((1-c_1)/(c_1))^{1/2}$, *then*

(i) $\|\lambda P_t - \mu P_t\| = o(\exp(-\gamma t))$

with $\gamma = c_0 \cdot (1 - 2\cdot(c_1\cdot(1-c_1))^{1/2})$. *If* $\Sigma_i\, a^i\lambda_i$ *is finite, then*

(ii) $\|\lambda P_t - \pi\| = o(\exp(-\gamma t))$,

where π *is the stationary distribution.*

Proof. Due to a domination argument (which?), it is no restriction to assume that $q_i = c_0$ for all $i \geq 1$, $p_i = c_1$ for all $i \geq 0$. Then

(3.5) $$\tau_0 \stackrel{\mathscr{D}}{=} \sum_1^{\eta} U_i$$

and

(3.6) $$\mathbf{E}_1[\exp(\alpha\tau_0)] = g(c_0/(c_0-\alpha)),$$

where g is the probability generating function for η if $X_0 = Z_0 = 1$. Now

$$g(s) = (1 - (1 - 4c_1(1-c_1)s^2)^{1/2})/2c_1 s$$

(cf. § II.4), hence $g(c_0/(c_0-\alpha))$ is finite iff $4c_1(1-c_1)\cdot(c_0/(c_0-\alpha))^2 \leq 1$, which is the same as $\alpha \leq \gamma$. We have $g(c_0/(c_0-\gamma)) = a$, and if $X_0 = Z_0 \stackrel{\mathscr{D}}{=} \lambda$, then

$$\mathbf{E}_\lambda[\exp(\gamma\tau_0)] = \sum_i g(c_0/(c_0-\gamma))^i \cdot \lambda_i,$$

which is finite by assumption.

Now $T \leq \max(\tau_0, \tau_0') \leq \tau_0 + \tau_0'$ and since the assumptions on λ and μ are identical and X and X' are independent, we get

$$\begin{aligned}\mathbf{E}_{\lambda\mu}[\exp(\gamma T)] &\leq \mathbf{E}_{\lambda\mu}[\exp(\gamma(\tau_0 + \tau_0'))] \\ &= \mathbf{E}_\lambda[\exp(\gamma\tau_0)]\cdot\mathbf{E}_\mu[\exp(\gamma\tau_0')] < \infty,\end{aligned}$$

and (i) follows.

For (ii), we must prove that $\mathbf{E}_\pi[\exp(\gamma\tau_0')] < \infty$. To do so, observe that the times of entrance to state 0,

$$\{t > 0;\ X'_{t-} = 1,\ X'_t = 0\}\,,$$

constitute a stationary renewal process, S' say. Let H denote the recurrence distribution of S'. We have

$$\int e^{\gamma x}\, dH(x) = g(c_0/(c_0 - \gamma)) \cdot c_0/(c_0 - \gamma) < \infty \tag{3.7}$$

since $\gamma < c_0$. That means that the delay Y_0' of S' satisfies $\mathbf{E}_\pi[\exp(\gamma Y_0')] < \infty$, because (3.6) implies that

$$\int e^{\gamma x} \bar{H}(x)\, dx < \infty\,,$$

and Y_0' has a density proportional to $\bar{H}$. Since $\tau_0' \leq Y_0'$, we are done. □

Due to (I.3.17) we also have

$$\int e^{\gamma t} \cdot \| \lambda P_t - \mu P_t \| < \infty$$

under the assumptions of Theorem (3.4). If $p_i \leq c_2 < \frac{1}{2}$ for $i \geq$ some i_0, then $\| \lambda P_t - \mu P_t \|$ is geometrically decreasing if $\Sigma_i\, a^i \lambda_i$ and $\Sigma_i\, a^i \mu_i$ are finite for an $a > 1$.

We now turn to the case $c_1 = \frac{1}{2}$ and will see an example of estimating $\mathbf{P}_{\lambda\mu}(T > t)$ without use of Markov's inequality.

(3.8) Theorem. *Assume that $c_0 > 0$ and $c_1 = \frac{1}{2}$. If the expectations $\Sigma_i\, i\lambda_i$ and $\Sigma_i\, i\mu_i$ are finite, then*

$$\| \lambda P_t - \mu P_t \| = \mathrm{O}(t^{-1/2}) \qquad \textit{as} \quad t \to \infty\,.$$

Proof. Since $\mathbf{P}_{\lambda\mu}(T > t) \leq \mathbf{P}_\lambda(\tau_0 > t) + \mathbf{P}_\mu(\tau_0' > t)$ and since we put the same conditions on λ and μ, it is sufficient to establish $\mathbf{P}_\lambda(\tau_0 > t) = \mathrm{O}(t^{-1/2})$. Again we set all q_i to c_0, and now all p_i equals $\frac{1}{2}$. Recall (3.5), and choose $\beta > 0$ so small that

$\beta \cdot \mathbf{E}[U_1] = \beta / c_0 < 1$. Then by splitting the sum $\Sigma_1^\eta U_i$ in an obvious way and using Chebyshev's inequality, we obtain

$$\mathbf{P}_\lambda\Big(\sum_1^\eta U_i \geq t\Big) \leq \beta \cdot \mathrm{Var}[U_1]/\{t \cdot (1 - \beta \cdot \mathbf{E}[U_1])^2\} + \mathbf{P}_\lambda(\eta \geq \beta t)\,,$$

so it remains to prove that $\sup_{t \geq 0} t^{1/2} \cdot \mathbf{P}_\lambda(\eta \geq \beta t) < \infty$, or what is the same thing, $\sup_{k \geq 0} k^{1/2} \cdot \mathbf{P}_\lambda(\eta \geq k) < \infty$. Since

$$\mathbf{P}_\lambda(\eta \geq k) = \sum_{j \geq 1} [\mathbf{P}_j(\eta \geq k)/j] \cdot j\lambda_j\,,$$

it is sufficient to prove that $\sup_{k \geq 0, j \geq 1} [k^{1/2} \cdot \mathbf{P}_j(\eta \geq k)/j] < \infty$. Now η has p.g.f. $g(s) = (1 - (1 - s^2)^{1/2})/s$ under $\mathbf{P}_1$ and g^j under $\mathbf{P}_j$, and the sequence $\mathbf{P}_j(\eta > k)$, $k \geq 0$, has p.g.f. $(g^j(s) - 1)/(s - 1)$. Hence with

$$h_j(u) = (g^j(e^{iu}) - 1)/(e^{iu} - 1)\,,$$

we get

$$\mathbf{P}_j(\eta > k) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-iku} \cdot h_j(u)\, du\,.$$

Partial integration renders

$$\int_{-\pi}^{\pi} e^{-iku} h_j(u)\, du = \int_{-\pi}^{\pi} (e^{-iku} - 1) \cdot (ik)^{-1} \cdot h_j'(u)\, du\,.$$

Now $|g'(z)| \leq C \cdot |1 - z|^{-1/2}$ for $|z| \leq 1$, and it follows that $|h_j'(u)| \leq C \cdot j \cdot |e^{iu} - 1|^{-3/2}$. We get

$$\begin{aligned} k^{1/2} \cdot j^{-1} \cdot \mathbf{P}_j(\eta > k) &\leq C \cdot k^{-1/2} \cdot \int_{-\pi}^{\pi} |e^{-iku} - 1| \cdot |e^{iu} - 1|^{-3/2}\, du \\ &\leq C \cdot k^{-1/2} \cdot \int_{-\pi}^{\pi} |e^{-iku} - 1| \cdot |u|^{-3/2}\, du \\ &= C \cdot \int_{-\pi k}^{\pi k} |e^{-iu} - 1| \cdot |u|^{-3/2}\, du \\ &\leq C \cdot \int_{-\infty}^{\infty} |e^{-iu} - 1| \cdot |u|^{-3/2}\, du < \infty\,. \end{aligned}$$

Hence

$$\sup_{j,k\geq 1} k^{1/2}\cdot j\cdot \mathbf{P}_j(\eta>k)<\infty,$$

which completes the proof. $\square$

4. Domination and monotonicity. For any set up of birth and death intensities there exist processes X, X' governed by these satisfying

$$(4.1)\qquad X_t\leq X'_t \qquad \text{for all} \quad t\geq 0$$

if the initial distributions λ, μ satisfy $\lambda\overset{\mathcal{D}}{\leq}\mu$. For that, only the nonexplosiveness assumption is needed; irreducibility and recurrence are non-essential. To achieve (4.1), let $X_0\leq X'_0$ and establish the Doeblin coupling given by (2.5).

Theorem (IV.9.3) applies to the effect that if $X_0\equiv 0$, then $\psi(\theta_t X)$ is stochastically increasing (decreasing) as $t\to\infty$ if ψ is an increasing (decreasing) mapping. This implies that

(4.2) (i) X_t is stochastically increasing as $t\to\infty$,

(ii) $\mathbf{E}_0[X_t]$ is increasing in t (possibly, it equals ∞ for all $t>0$), and

(iii) if the birth and death intensities allow the existence of a stationary distribution π, then $\mathbf{E}_0[X_t]\leq\Sigma_0^\infty i\pi_i$ for all t.

Here (i)$\Rightarrow$(ii), and (iii) follows from (4.1) by letting X' have π as the initial distribution.

The functional $\psi_1\in\mathcal{D}/\mathcal{R}_+$ given by

$$(4.3)\qquad \psi_1(x)=\inf\{t;\, x_t=0\}$$

is certainly increasing; hence $\psi_1(\theta_t X)$ (= the elapse after time t before state 0 is hit) is stochastically increasing in t if $X_0\equiv 0$.

Now consider a birth and death process with state space $E=\{0,\dots,m\}$, $m\geq 1$. You may have an $M/M/1$ queue with limited waiting room in mind (that process has birth intensities $\beta_n=\beta$, $0\leq n\leq m-1$, and death intensities $\delta_n=\delta$, $1\leq n\leq m$). Consider the functionals $\in\mathcal{D}/\mathcal{R}_+$:

(4.4) (i) $\psi_2(x) = \inf\{t; x_t = m\}$,
(ii) $\psi_3(x) = I_{\{m\}}(x_0)$, and
(iii) $\psi_4(x) = l_+\{s; 0 \le s \le A \text{ and } x_s = m\}$,

where in (iii) A is a positive constant, and l_+ is the Lebesgue measure on $(R_+, \mathcal{R}_+)$. Certainly, ψ_2 is decreasing, while ψ_3 and ψ_4 are increasing. That $\psi_3(\theta_t X)$ is stochastically increasing in t if $X_0 \equiv 0$ may be expressed as

$$p_{0m}(t) \text{ is increasing in } t, \tag{4.5}$$

and we understand that

$$p_{00}(t) \text{ is decreasing in } t, \tag{4.6}$$

since $\psi(x) = I_{\{0\}}(x_0)$ defines a decreasing functional.

You have now got an idea about the possibilities of Theorem (IV.9.3). But the Doeblin coupling (2.5) and that theorem are not good enough to prove the following: If X is a pure birth process (all $\delta_n = 0$) such that $\beta_0 \le \beta_1 \le \cdots$, then

$$\mathbf{E}_i[X_t] \text{ is convex in } t \tag{4.7}$$

for all i, and hence for all initial distributions. This is a simple example of what an improved coupling and the remark (IV.9.5) render; we shall now investigate that.

We construct a coupling of two processes using transition intensities as follows:

$$(4.8)\quad q_{(i,j)(i+1,j+1)} = \beta_i, \quad q_{(i,j)(i,j+1)} = \beta_j - \beta_i \quad \text{for} \quad i<j,$$
$$q_{(i,j)(k,m)} = q_{(j,i)(m,k)} \quad \text{for} \quad j<i, \text{ and } q_{(i,i)(i+1,i+1)} = \beta_i.$$

A jump process $\tilde{X} = (\tilde{X}_t)_0^\infty$ in $\mathbb{Z}_+^2$ with these intensities means a coupling of birth and death processes X and X' $[\tilde{X}_t = (X_t, X'_t)]$ governed by $\beta_0, \beta_1, \ldots$. The crucial observation is that

$$X'_t - X_t \text{ is increasing in } t \tag{4.9}$$

if $i \le j$ and $X_0 = i$, $X'_0 = j$. For initial distributions λ, μ satisfying

$\lambda \overset{\mathscr{D}}{\leq} \mu$, we let $X_0 \leq X_0'$ as usual, and (4.9) turns out to hold under that condition. Now let

$$\psi(x) = x_h - x_0$$

for $x = (x_t)_0^\infty \in D$ and a fixed $h > 0$. If we are able to prove that

$$\psi(X) \overset{\mathscr{D}}{\leq} \psi(X') \tag{4.10}$$

if $X_0 \equiv i \leq j \equiv X_0'$, then condition (IV.9.5) is satisfied and we conclude that

$$\psi(\theta_t X) \text{ is stochastically increasing in } t \text{ if } X_0 \equiv 0.$$

This implies that $\mathbf{E}_0[\psi(\theta_{t-h}X)] \leq \mathbf{E}_0[\psi(\theta_t X)]$ for all $t \geq h$. We obtain

$$\mathbf{E}_0[\psi(\theta_{t-h}X)] = \mathbf{E}_0[X_t - X_{t-h}] \leq \mathbf{E}_0[X_{t+h} - X_t] = \mathbf{E}_0[\psi(\theta_t X)],$$

so $\mathbf{E}_0[X_t] \leq \frac{1}{2} \cdot (\mathbf{E}_0[X_{t+h}] + \mathbf{E}_0[X_{t-h}])$ for all $t \geq h$. Since h was arbitrary, we have proved that

$$\mathbf{E}_0[X_t] \text{ is convex in } t \tag{4.11}$$

for a pure birth process with increasing birth rates if (4.10) has been set. But (4.10) follows immediately from (4.9). For a pure birth process X starting at i, $(X_t - i)_0^\infty$ is again such a process, with rates $\beta_{i+\cdot}$, starting at 0. Hence (4.7) follows from (4.11).

We now consider processes governed by intensities satisfying

$$\beta_0 \geq \beta_1 \geq \cdots \quad \text{and} \quad \delta_1 \leq \delta_2 \leq \cdots. \tag{4.12}$$

A coupling with intensities analogous to (4.8) is obtained from the following table of transitions:

(4.13) from	to	with intensity
(i, j)	$(i+1, j+1)$	β_j
	$(i+1, j)$	$\beta_i - \beta_j$ for $i \leq j$, and
	$(i-1, j-1)$	δ_i
	$(i, j-1)$	$\delta_j - \delta_i$ for $1 \leq i \leq j$,

and the obvious extension by symmetry to the case $i > j$. Notice the transitions from states (i, i). This yields a bivariate process $\tilde{X}_t = (X_t, X'_t)$, $t \geq 0$, such that

(4.14) $$X'_t - X_t \quad \text{is decreasing in } t$$

if $\lambda \overset{\mathcal{D}}{\leq} \mu$. Arguing as above, we now obtain that

(4.15) $\mathbf{E}_0[X_t]$ is increasing and concave in t under condition (4.12).

Now suppose that all $\beta_0, \beta_1, \ldots, \delta_1, \delta_2, \ldots$ are strictly positive and govern recurrent processes. It is then rather easy to see that (X, X') provide a successful coupling, due to (4.14). For this, prove and use that

$$\mathbf{P}_{ij}(X'_t - X_t = j - i - 1 \text{ for some } t) = 1$$

when $i < j$. You should be worried about the case when all β_i and all δ_i values are equal, but the recurrence rules out the possibility that $X'_t - X_t$ remains constant. Hence $X'_t - X_t$ decreases to 0 as $t \to \infty$ if $\lambda \overset{\mathcal{D}}{\leq} \mu$, and due to dominated convergence we obtain that

(4.16) $\mathbf{E}_{\lambda\mu}[X'_t - X_t]$ decreases to 0 as $t \to \infty$ if μ has a finite first moment.

An important birth and death process satisfying (4.13) is the $M/M/k$ queue, where X_t represents the number of customers in the system at time t. It has intensities $\beta_n = \beta$ for $n \geq 0$, $\delta_n = n \cdot \delta$ for $1 \leq n \leq k$, and $\delta_n = k \cdot \delta$ for $n > k$. As you know, it is positive recurrent and hence possesses a stationary distribution π if and only if $\beta < k \cdot \delta$. Using our latest findings, we can now state the following about a positive recurrent $M/M/k$ queue, idle at time 0 $(X_0 \equiv 0)$:

(4.17) (i) X_t is stochastically increasing in t,
(ii) $\mathbf{E}_0[X_t]$ is concave in t, and
(iii) $\mathbf{E}_0[X_t]$ increases to $\Sigma_0^\infty i\pi_i$ as $t \to \infty$.

Here (ii) is just (4.15), and (iii) follows from (4.16) with X' stationary. Notice that $\Sigma_0^\infty i\pi_i$ is finite since π_i tends to 0 geometrically fast as $i \to \infty$.

Recall criterion (1.4) for nonexplosiveness. The domination argument to prove its sufficiency was rather vague. But now we are in position to be precise: Construct a coupling with a table like (4.13), to get a bivariate process (X_t, X'_t), $t \geq 0$, where X' is a pure birth process with rates β_i dominating X.

We conclude this section by showing how a variant of the idea in (4.13) may be used for comparisons of different birth and death processes. Indeed, let X be governed by $\beta_0, \beta_1, \ldots, \delta_1, \delta_2, \ldots$ and X' by $\beta'_0, \beta'_1, \ldots, \delta'_1, \delta'_2, \ldots$. Suppose that

$$(4.18) \qquad \beta_n \leq \beta'_n \quad \text{for all} \quad n \geq 0, \qquad \delta_n \geq \delta'_n \quad \text{for all} \quad n \geq 1\,.$$

We let X and X' be the coordinate processes of $\tilde{X}$ with the following intensity table:

(4.19)

from	to	with intensity
(i, j)	$(i+1, j)$	β_i
	$(i, j+1)$	β'_j for $i \neq j$,
	$(i-1, j)$	δ_i for $1 \leq i \neq j$,
	$(i, j-1)$	δ'_j for $1 \leq j \neq i$,
(i, i)	$(i+1, i+1)$	β_i
	$(i, i+1)$	$\beta'_i - \beta_i$ and
	$(i-1, i-1)$	δ'_i
	$(i-1, j)$	$\delta_i - \delta'_i$ for $i \geq 1$.

These intensities work to the effect that X and X' evolve independently until (if ever) they meet. If $X_t = X'_t$ for some t, (4.19) uses (4.18) in such a way that $X_s \leq X'_s$ for all $s \geq t$. Hence for initial distributions λ, μ satisfying $\lambda \leq \mu$, we obtain that

$$(4.20) \qquad X_t \leq X'_t \qquad \text{for all} \quad t \geq 0\,,$$

after initial values X_0, X'_0 satisfying $X_0 \leq X'_0$ have been put up in the standard way. For one illustration, use (4.18)–(4.20) to compare two $M/M/k$ queues with different parameters. For another, consider the Yule process (i.e., the birth and death process with $\beta_i = i \cdot \beta$ and $\delta_i = i \cdot \delta$, where β and δ are constants ≥ 0). Use (4.18)–(4.20) for a comparison of two Yule processes. Also, make the use of domination in the proof of Theorem (3.4) precise by suitable couplings. As will be seen, (4.19) is just a first example of fruitful technique to compare Markov processes.

5. Notes. The material of § 1 is standard (cf. Freedman [62], Asmussen [10], and Karlin and Taylor [86]). The succeeding §§ 2 and 3 are based on Lindvall [107], a paper prepared independently of Küchler and Lunze [94]. The latter consider birth–death and diffusion processes as particular cases of so-called quasi-diffusion processes, and prove that success of Doeblin coupling is necessary for ergodicity. For analytical methods to prove monotonicity properties, see van Doorn [157]. Much attention has been paid to the ergodicity problem for transient birth–death and diffusion processes with or without couplings in focus; see Aldous [1] and Chapter VI and its references.

2. GENERAL BIRTH AND DEATH PROCESSES

6. Basics. When it comes to generalize the results above, we face some terminology problems. It is not unnatural to find the term "birth and death process" suitable for a wide class of processes, even including many interacting particle systems (cf. Part 3). But the latter name is well established, and the common use of "birth and death processes" is restricted to the one-dimensional case above.

By a general birth and death process (a GBD process) we shall mean any Markov jump process used to model the development of a system, in which it is, more or less, natural to use the words "births" and "deaths" to describe the transitions (individuals, or whatever the objects in the system are called, are born and die).

There are certainly interesting examples of GBD processes with uncountable state spaces (cf. § 10), but we shall restrict our attention here to the countable case $E = \mathbb{Z}_+^L$, $L \geq 2$, or a subset thereof. Even if you are not a friend of abbreviations, hopefully you agree that repeating the phrase "general birth and death process" becomes tiresome.

States in E are denoted by $\mathbf{i} = (i_1, \ldots, i_L)$; for example, the minimal requirement for a Markov process in E to be called a GBD process is the following:

(6.1) for $\mathbf{i}, \mathbf{j} \in E$, $q_{\mathbf{ij}} > 0$ only if $\mathbf{j} = \mathbf{i} + I_B - I_D$ for some subsets B and D of $\{1, \ldots, L\}$ such that $B \cap D = \emptyset$.

Certainly, the standard models for queueing networks and epidemics (topics to be discussed below) satisfy (6.1). Usually, only single-point sets B and D appear; we then have the following transition possibilities:

(6.2) (i) $(i_1, \ldots, i_k, \ldots, i_L)$ to $(i_1, \ldots, i_k + 1, \ldots, i_L)$,
(ii) $(i_1, \ldots, i_k, \ldots, i_L)$ to $(i_1, \ldots, i_k - 1, \ldots, i_L)$, and
(iii) $(i_1, \ldots, i_k, \ldots, i_m, \ldots, i_L)$ to $(i_1, \ldots, i_k - 1, \ldots, i_m + 1, \ldots, i_L)$, or $(i_1, \ldots, i_k + 1, \ldots, i_m - 1, \ldots, i_L)$,

where $1 \le k < m \le L$. Often i_k denotes the number of individuals of type k, according to some classification with L types. A transition (i) or (ii) means that an individual of type k is born or dies, while (iii) indicates that an individual changes type.

Before the elaboration of some of our earlier ideas to certain GBD processes, we turn our attention to the ergodicity of Markov processes with arbitrary countable state spaces.

7. Ergodicity. For the semigroup $(P_t)_0^\infty$, $P_t = (p_{ij}(t))_{i,j \in E}$, governing an irreducible and recurrent Markov process X with the countable state space E, we shall prove that

$$\|\lambda P_t - \mu P_t\| \to 0 \quad \text{as} \quad t \to \infty \tag{7.1}$$

for any initial distributions λ, μ. But $\|\lambda P_t - \mu P_t\| \le \|\lambda P_{[t]} - \mu P_{[t]}\|$ according to (III.20.5); hence (7.1) follows if we can prove that

$$\|\lambda P_n - \mu P_n\| \to 0 \quad \text{as} \quad n \to \infty. \tag{7.2}$$

Now $P_n = \tilde{P}^n$, where $\tilde{P}$ is the transition matrix $(p_{ij}(1))_{i,j\in E}$, so if $\tilde{X} = (X_n)_0^\infty$ is a recurrent Markov chain, then (7.2) is a consequence of the ergodic theorem for null-recurrent chains; recall § II.13. But $\tilde{X}$ is recurrent: Take a state i, let $X_0 \equiv i$, and define $(\tau_j)_1^\infty$ to be the time points in the set

$$\{t; X_{t-} \neq i, X_t = i\}.$$

Once X remains in the state i for longer than one time unit, we have that $X_n = i$ for some $n \geq 1$. But to have such a duration we need observe X in the intervals after the τ_j's only a geometrically distributed number of times.

If X is positive recurrent, we can refer to (II.8.4)–(II.8.5) for (7.1) via (7.2) if $\tilde{X}$ is also positive recurrent. In particular,

$$\|\lambda P_t - \pi\| \to 0 \quad \text{as} \quad t \to \infty \tag{7.3}$$

for any initial distribution λ. The process $\tilde{X}$ has that desired property, as we understand from sharpening the argument above. Indeed, let $\tau_0 \equiv 0$ and

$$\eta = \min\{k \geq 1; X_t = i \text{ for } t \in [X_{\tau_k - 1}, X_{\tau_k - 1} + 1]\}.$$

Then $\tilde{X}_n = i$ for an $n \geq 1$ not greater than

$$\sum_1^\eta (\tau_i - \tau_{i-1}),$$

which has finite expectation due to Wald's lemma.

8. Networks. We consider a queueing system, with L service stations, forming a network. To one of the (single-server) stations, we call them nodes, a customer arrives either from an external source or from finished service at some other node, is served, and then either leaves the network or goes to another node.

If we have no external arrivals or drain, customers move only internally in the network and we call it closed while a network allowing customers both to enter and leave is called open. Let

$$X_t = (X_{1t}, \ldots, X_{Lt}), \qquad t \geq 0,$$

be the process describing the network in the sense that X_{kt} is the number of customers of node k at time t. For an open network, $\mathbb{Z}_+^L$ is the natural state space. In a closed network $\Sigma_1^L X_{kt}$ is constant, equal to N say, and then

$$E_N = \left\{ \mathbf{i} \in \mathbb{Z}_+^L ; \sum i_k = N \right\}$$

is chosen.

We assume that

(8.1) (i) customers arrive at node k according to a Poisson process with intensity β_k,

(ii) service times at node k are $\mathrm{Exp}(\delta_k)$ distributed,

(iii) after finished service at node k, a customer goes to node $m \neq k$ with probability γ_{km} or leaves the network with probability $\gamma_k = 1 - \Sigma_m \gamma_{km}$, and

(iv) all the random quantities in (i) to (iii) are independent.

These assumptions make $X = (X_t)_0^\infty$ a GBD process, with transitions of type (6.2). We have transitions, with $\mathbf{i} = (i_1, i_2, \ldots, i_L)$, as follows:

(8.2) from	to	with intensity
$\mathbf{i}$	$\mathbf{i} + I_{(k)}$	β_k
	$\mathbf{i} - I_{(k)}$	$\delta_k \cdot \gamma_k \cdot I(i_k > 0)$
	$\mathbf{i} + I_{(m)} - I_{(k)}$	$\delta_k \cdot \gamma_{km} \cdot I(i_k > 0)$

for $1 \leq k \neq m \leq L$, as the possibly nonvanishing transition intensities. In a closed network, only transitions of the last type occur, while in an open one β_k and $\gamma_m > 0$ for some k and m, and we have transitions of all three types. This model with single servers at each node is known as the Jackson network.

Above, we replaced the term "service station" by the general "node." Models similar to that given by (8.2) are certainly of interest not only for queueing systems; we may also leave the "customers" behind and let objects of some other sort find their paths through a network. Hence it is appropriate to choose general terms.

To get an overview of the flow in the network, it is useful to draw a graph consisting of the nodes, arrows from node k to node m for pairs (k, m) such that $\gamma_{km} > 0$ and arrows to and from the nodes k such that $\beta_k > 0$ and $\gamma_k > 0$, respectively. For ergodicity of X, we must check the irreducibility. If for all $\mathbf{i} = (i_1, \ldots, i_L) \in \mathbb{Z}_+^L$ and k

$$
\begin{aligned}
(8.3) \qquad & p(\mathbf{i}, \mathbf{i} + I_{\{k\}}, t) > 0 \quad \text{and} \\
& p(\mathbf{i}, \mathbf{i} - I_{\{k\}}, t) > 0 \quad \text{when} \quad i_k > 0
\end{aligned}
$$

for some $t > 0$, we have that property for an open network [we used notation $p(\mathbf{i}, \mathbf{j}, t)$ rather than $p_{\mathbf{ij}}(t)$ for the readability]. The conditions of (8.3) are easy to check. Once the irreducibility is set, we may refer to (7.1) and (7.3) for the asymptotics of X. Concerning recurrence criteria, we content ourselves by stating the fact that if for each node k the rate of arriving customers is strictly smaller than δ_k, then X is positive recurrent, the open network has a stationary distribution π, and (7.3) yields the fact that X is asymptotically stationary. A closed network has a finite number of states; hence (7.3) holds for a unique π as soon as X is irreducible. You are left to contemplate when that is the case.

To extend some of the results of § 4, we now demonstrate that coupling satisfying (IV.8.1) can be done for networks satisfying (8.2). We use the natural partial ordering $\leq$ on $\mathbb{Z}_+^L$: $\mathbf{i} \leq \mathbf{j}$ for $\mathbf{i} = (i_1, \ldots, i_L)$, $\mathbf{j} = (j_1, \ldots, j_L)$ iff $i_k \leq j_k$ for all k. With that ordering, (IV.8.1) is of interest only for open networks, because there are no pairs of distributions on E_N that can be compared.

For initial distributions λ, μ satisfying $\lambda \overset{\mathcal{D}}{\leq} \mu$, first use Strassen's theorem to obtain initial values $X_0 \overset{\mathcal{D}}{=} \lambda$, $X_0' \overset{\mathcal{D}}{=} \mu$ of $X = (X_t)_0^\infty$ and $X' = (X_t')_0^\infty$ satisfying $X_0 \leq X_0'$. Then let (X_t, X_t'), $t \geq 0$, be governed by the following intensities:

(8.4)

from	to	with intensity
$(\mathbf{i}, \mathbf{j})$	$(\mathbf{i} + I_{\{k\}}, \mathbf{j} + I_{\{k\}})$	β_k
	$(\mathbf{i} - I_{\{k\}} \cdot I(i_k > 0), \mathbf{j} - I_{\{k\}} \cdot I(j_k > 0))$	$\delta_k \cdot \gamma_k$
	$(\mathbf{i} + (I_{\{m\}} - I_{\{k\}}) \cdot I(i_k > 0),$ $\mathbf{j} + (I_{\{m\}} - I_{\{k\}}) \cdot I(j_k > 0))$	$\delta_k \cdot \gamma_{km}$

for $1 \leq k$, $m \leq L$. Certainly, $X_t \leq X_t'$ for all $t \geq 0$.

As in § 4, we can now apply Theorem (IV.9.3), and find that if $X_0 \equiv \mathbf{0}$ (the network is idle at time 0), then $\psi(\theta_t X)$ is stochastically increasing in t if ψ is an increasing mapping. Using the appropriate mappings ψ, we find that

(8.5) (i) $(X_{k_1 t}, \ldots, X_{k_n t})$ for any $\{k_1, \ldots, k_n\} \subset \{1, \ldots, L\}$,
(ii) $\max_{1 \le k \le L} X_{kt}$,
(iii) $\inf\{s - t; s \ge t, X_s = \mathbf{0}\}$, and
(iv) $l_+\{s;\ t \le s \le t + A$ and $\Sigma_{k=1}^{L} X_{ks} \ge m\}$ for constants $A > 0,\ m \in \mathbb{N}$,

for example, are stochastically increasing in t, with state spaces $\mathbb{Z}_+^n$, $\mathbb{Z}_+$, $\mathbb{R}_+$, and $\mathbb{R}_+$, respectively. An example of a process stochastically decreasing in t is

$$\text{(8.6)} \qquad \inf\left\{s - t; s \ge t, \sum_1^L X_{ks} \ge m\right\}$$

[cf. (4.4)].

For results of the type (8.5)–(8.6), no irreducibility or recurrence conditions are needed. If these and the positive recurrence are reerected so that a unique stationary distribution π exists, we may let X' be stationary and combine our achievements to the effect that if $X_0 \equiv \mathbf{0}$, then

(8.7) (i) X_t is stochastically increasing in t,
(ii) $X_t \overset{\mathcal{D}}{\le} \pi$ for all $t \ge 0$, and
(iii) the distribution of X_t tends in total variation to π as $t \to \infty$.

In § 4 it was shown how birth and death processes with different intensities satisfying (4.18) may be compared. In the present setting, we must be careful when modifying (8.4) in the spirit of (4.19) to obtain

(8.8) $X_t \le X'_t$ for all $t \ge 0$
when the initial distributions λ, μ satisfy $\lambda \overset{\mathcal{D}}{\le} \mu$.

However, if

(8.9) $$\beta_k \le \beta'_k \quad \text{for} \quad 1 \le k \le L, \quad \text{and}$$
$$\delta_k = \delta'_k,\ \gamma_{km} = \gamma'_{km} \quad \text{for} \quad 1 \le k,\ m \le L\,,$$

then we have the following transition table:

(8.10)

from	to	with intensity
$(\mathbf{i}, \mathbf{j})$	$(\mathbf{i} + I_{(k)}, \mathbf{j} + I_{(k)})$	β_k
	$(\mathbf{i}, \mathbf{j} + I_{(k)})$	$\beta'_k - \beta_k$
	etc.	

and (8.8) follows.

9. Propagations. Couplings like that based on (8.4) are also valuable to reveal the properties of different sorts of propagations, epidemics for example. We consider two simple cases.

Before a vote, a population is divided into three groups: In the first we have those in favor of a certain proposition, in the second those against, and in the third the undecided. Let X_{it} denote the number in the group i at the time t. We assume that the population size is N and that $X_t = (X_{1t}, X_{2t}, X_{3t})$, $t \ge 0$, is a GBD process in E_N with the following intensities:

(9.1)

from	to	with intensity
$\mathbf{i}$	$\mathbf{i} + (i_1 + 1, i_2, i_3 - 1) \cdot I(i_3 > 0)$	$\beta_1 \cdot i_1 \cdot i_3$
	$\mathbf{i} + (i_1, i_2 + 1, i_3 - 1) \cdot I(i_3 > 0)$	$\beta_2 \cdot i_2 \cdot i_3$

where β_1, β_2 are constants >0. The graph is simple enough: We have three nodes and arrows from the third to each of the first two. The idea behind the choice of intensities is obvious.

We shall construct a coupling (X, X') to illuminate the intuitively obvious fact that if we increase the number in group 1, that advantage is retained as time passes. To be precise: Our coupling is such that

(9.2) $X_{1t} \le X'_{1t}$ for all $t \ge 0$
if X_0, X'_0 take initial values (n_1, n_2, n_3) and (n'_1, n'_2, n'_3) such that $n'_1 \ge n_1$ and $n_3 = n'_3$.

To achieve that, we use intensities as follows:

(9.3)

from	to	with intensity
$(\mathbf{i}, \mathbf{j})$	$((i_1+1, i_2, i_3-1),$ $(j_1+1, j_2, j_3-1))$	$\beta_1 \cdot ((i_1 \cdot i_3) \wedge (j_1 \cdot j_3))$
	$(\mathbf{i}, (j_1+1, j_2, j_3-1))$	$\beta_1 \cdot (j_1 \cdot j_3 - i_1 \cdot i_3)^+$
	$((i_1+1, i_2, i_3-1), \mathbf{j})$	$\beta_1 \cdot (i_1 \cdot i_3 - j_1 \cdot j_3)^+$

and define similarly for transitions affecting i_2 and/or j_2. Due to the initial condition $n_3 = n_3'$, we find that (X_t, X_t'), $t \geq 0$, takes values only in the set

$$\{(\mathbf{i}, \mathbf{j}); i_1 \leq j_1 \text{ and } i_3 \leq j_3, \text{ or } i_3 = j_3 + 1 \text{ and } j_1 \geq i_1 + 1\}$$

and (9.2) follows after a moment's thought.

Our second example is the so-called general epidemic process $X_t = (X_{1t}, X_{2t})$, $t \geq 0$. Here X_{1t} and X_{2t} are the number of susceptibles and infectives, respectively, at time t, in a finite population. We let $X = (X_t)_0^\infty$ be a GBD process with intensities

$$\text{(9.4)} \quad \text{(i)}\ q((i_1, i_2), (i_1 - 1, i_2 + 1)) = \beta \cdot i_1 \cdot i_2, \text{ and}$$
$$\text{(ii)}\ q((i_1, i_2), (i_1, i_2 - 1)) = \delta \cdot i_2\,,$$

where β, δ are constants > 0. We assume that at the outset that $X_0 = (n_1, n_2)$, where $n_1 + n_2 = N =$ the size of the population. To understand how X_{2t} increases in t initially, observe that the intensity $\beta \cdot i_1 \cdot i_2$ is rather close to $\beta \cdot N \cdot i_2$ if i_2 is small. But the birth and death process with intensities $\beta_j = \beta \cdot N \cdot j$ and $\delta_j = \delta \cdot j$ is the familiar Yule process, (cf. the comment at the end of § 4), which is exhaustively analyzed; among the elementary things we know that it grows exponentially fast. We may hope for a domination result

$$\text{(9.5)} \qquad X_{2t} \leq X_{3t} \qquad \text{for all} \quad t \geq 0\,,$$

where $(X_{3t})_0^\infty$ is a Yule process with the parameters given above. And $X_{3t} - X_{2t}$ should be small for small $t \geq 0$, which supports the notion that "an epidemic grows exponentially fast."

To achieve (9.5), consider an extension of X_t, $t \geq 0$, to a three-dimensional process $X_t^* = (X_{1t}, X_{2t}, X_{3t})$, $t \geq 0$, with the following transitions:

(9.6)

from	to	with intensity
(i_1, i_2, i_3)	$(i_1 - 1, i_2 + 1, i_3 + 1)$	$\beta \cdot i_1 \cdot i_2$
	$(i_1, i_2, i_3 + 1)$	$\beta \cdot N \cdot i_3 - \beta \cdot i_1 \cdot i_2$
	$(i_1, i_2 - 1, i_3 - 1)$	$\delta \cdot i_2$
	$(i_1, i_2, i_3 - 1)$	$\delta \cdot (i_3 - i_2)$

for the relevant states (i_1, i_2, i_3) such that $i_1 + i_2 \leq N$ and $i_2 \leq i_3$. The inequality (9.5) is readily deduced. Notice, however, that $X_{3t} - X_{2t}$ is not increasing a.s. in t.

The approximation result that has been promised is the following:

$$\text{(9.7)} \qquad \sup_{t \leq A/N} (X_{3t} - X_{2t}) \to 0 \quad \text{a.s.} \quad \text{as} \quad N \to \infty$$

for all constants $A > 0$, if $X_{30} = X_{20} = n_2$. To prove that, consider the embedded Markov chain $Z = (Z_n)_0^\infty$, where $Z_n = (Z_{1n}, Z_{2n}, Z_{3n}) \in \mathbb{Z}_+^3$, with Z_n = the state of X^* at the nth jump of that process. Now if $Z_{20} = Z_{30} = n_2$, then for all $K \geq 1$ we have that

$$\mathbf{P}(Z_{2i} = Z_{3i} \text{ for } 0 \leq i \leq K) \to 1 \quad \text{as} \quad N \to \infty .$$

This follows with an induction argument and a close look at (9.6), which gives

$$\mathbf{P}(Z_{2i} < Z_{3i} \| Z_{2j} = Z_{3j} \text{ for } j < i) \leq C_K/N \qquad \text{for all} \quad i \leq K ,$$

where C_K is a constant depending on K and n_2 but not on N. Now the number of jumps of X^* in the interval A/N is stochastically bounded, so (9.7) follows.

Suppose that N is large and n_2 small; unless the infectives are all removed quickly according to (ii), the epidemic spreads (exponentially fast, as we have just seen) to begin with. But eventually, there are few left of the susceptibles, the effect of (ii) manifests itself, and X_{2t} decreases. It is natural to expect that this development goes more quickly if n_2 is increased. We clarify this by showing that an obvious coupling of two processes X and X' governed by (9.4) has the following property: There exists a random time τ such that

$$
\begin{aligned}
(9.8) \qquad & X_{2t} \le X'_{2t} \quad \text{for } t \le \tau , \\
& X_{2t} \ge X'_{2t} \quad \text{for } t > \tau
\end{aligned}
$$

if $n_2 \le n'_2$. For the proof, let (X_t, X'_t), $t \ge 0$, have intensities related to those of (9.3):

(9.9)

from	to	with intensity
$(\mathbf{i}, \mathbf{j})$	$((i_1 - 1, i_2 + 1), (j_1 - 1, j_2 + 1))$	$\beta \cdot (i_1 \cdot i_2) \wedge (j_1 \cdot j_2)$
	$(\mathbf{i}, (j_1 - 1, j_2 + 1))$	$\beta \cdot (j_1 \cdot j_2 - i_1 \cdot i_2)^+$
	$((i_1 - 1, i_2 + 1), \mathbf{j})$	$\beta \cdot (i_1 \cdot i_2 - j_1 \cdot j_2)^+$
	$((i_1, i_2 - 1), (j_1, j_2 - 1))$	$\gamma \cdot (i_2 \wedge j_2)$

etc.

[the two transitions omitted have intensities $\delta \cdot (i_2 - i_2 \wedge j_2)$ and $\delta \cdot (j_2 - i_2 \wedge j_2)$, respectively]. Now with $X_0 = (n_1, n_2)$ and $X'_0 = (n'_1, n'_2)$ such that $n_1 + n_2 = n'_1 + n'_2 = N$ and $n_2 \le n'_2$, let

$$\tau = \inf\{t; X_{2t} = X'_{2t}\} .$$

With that τ, (9.8) follows from a close look at (9.9).

10. Notes. References to queueing networks are Kelly [87] and Walrand [158]. For a survey, see [10], where explicit calculations of the stationary distributions omitted here can be found. The first example of § 9 was inspired by Kendall and Saunders [91]. For other couplings relevant to epidemics, see Ball [17–19], and Ball and Donnelly [20]. The paper by Mollison [120] has been an inspiration to the UK couplings. Couplings of GBD processes with state spaces beyond the countable are used in Preston [133] and Lotwick and Silverman [115]. Chen [36, 37] couples general Markov jump processes. For a special case, see § 19.

3. INTERACTING PARTICLE SYSTEMS

11. A signpost. Basics and examples. An interacting particle system is a Markov process $\xi = (\xi_t)_0^\infty$ of a certain type, usually with state space $E = \{0, 1\}^S$, where S is countable, often equal to $\mathbb{Z}^d$ for some $d \geq 1$. The word "particle" suggests that we shall consider each $x \in S$ as a site, and let $\xi_t(x)$ equal 1 if that site is occupied at time t, and 0 otherwise. However, this type of process is useful for many purposes, and the possibilities 0 or 1 can also have the meaning that a person at x is against or for a proposal, or that he or she is healthy or infected by an epidemic. We may also represent the spin of an iron atom at x by 0 or 1. The word "interacting" is used to stress that the intensity for a flip at x depends on the whole configuration of particles; the dependence is often restricted to the neighboring sites of x.

Since about 1970, the theory of interacting particle systems has constituted a major area of probability. From the start, the coupling method has been an indispensable tool; in fact, several couplings have been developed within that theory and have then turned out to be of value for the analysis of other, often simpler processes. Actually, we have seen several examples of that above, in this chapter and earlier, without mentioning the histories of the particular couplings.

By now, there are excellent accounts of interacting particle systems, including presentations of the relevant couplings. As a consequence, we can be brief and sketchy. Consider this part as the text on a signpost, whose purpose it is to raise your interest and send you in the right direction. The notation used is a compromise between the one established here and in the particle system literature.

Fix E to be $\{0, 1\}^S$, and equip E with the standard product σ-field, denoted by $\mathscr{E}$. Let η be a typical (nonrandom) element in E. An interacting particle system with that E (or, which is virtually the same thing, $E = \{-1, 1\}^S$) as the state space is called a spin system; we shall restrict our attention to this type of systems. The basic instrument to describe and construct such a system is the flip rate function c: $S \times E \to \mathbb{R}_+$, governing a particle system to the effect that

$$\mathbf{P}_\eta(\xi_t(x) \neq \eta(x)) = c(x, \eta) \cdot t + o(t) \tag{11.1}$$

and

$$\mathbf{P}_\eta(\xi_t(x) \neq \eta(x), \xi_t(y) \neq \eta(y)) = o(t) \tag{11.2}$$

as $t \to 0$ for all $x, y \in S$ satisfying $x \neq y$ and all $\eta \in E$. The relation (11.2) means that $c(x, \eta)$ is the intensity for a flip at site x when the configuration equals η, while due to (11.2) the possibility for more than one flip in any finite set of sites during a short time interval can be ignored.

But $\xi = (\xi_t)_0^\infty$ appeared in (11.1)–(11.2) without any arguments for its existence. The construction of a unique D_E-valued Markov process ξ from a flip rate function (on which certain conditions must be imposed) is carried out in several steps, first by setting a generator; from that, a semigroup; and finally, the desired process. To penetrate all the aspects involved is an excellent course in the basics of Markov process theory; we shall not enter that here, however.

Our first example is the contact process. Here $S = \mathbb{Z}^d$ for some $d \geq 1$, and $\eta(x) = 0$ or 1 represents the possibilities healthy or infected concerning the individual $x \in S$. Hence the contact process models an epidemic in S. A natural choice of flip rates is the following:

$$c(x, \eta) = \begin{cases} \beta \cdot \sum\limits_{|y-x|=1} \eta(y) & \text{if } \eta(x) = 0 \\ 1 & \text{if } \eta(x) = 1\,, \end{cases} \tag{11.3}$$

where β is a constant > 0.

In connection with the examples here, a number of conjectures and questions are made and posed; answers to them are given in the following sections. Concerning the contact process, we first notice that $\mathbf{0}$ is an absorbing state. Is that state reached eventually regardless of the initial distribution, or do we have, as in branching processes and epidemics, a critical value, β_c say, such that "extinction" is certain if $\beta < \beta_c$ but not if $\beta > \beta_c$?

Endow E with a partial ordering defined by $\eta \leq \eta'$ if and only if $\eta(x) \leq \eta'(x)$ for all $x \in S$. Of course, $\eta \leq \mathbf{1}$ for all $\eta \in E$ (the meaning of **1** and **0** above, is the obvious one). Let $\xi_0 = \mathbf{1}$. Encouraged by earlier achievements, we assume here that ξ_t is then stochastically decreasing in t. Since E is compact, any family of probability measures on E is trivially tight. Hence ξ_t is weakly convergent as $t \to \infty$, due to the comment after Corollary (IV.6.4).

Let us denote the limit distribution by ν_β. Certainly, ν_β is a stationary distribution; we have

$$\nu_\beta = \lim_{t\to\infty} \delta_{\mathbf{1}} P_t = \lim_{t\to\infty} \delta_{\mathbf{1}} P_{s+t} = \lim_{t\to\infty} (\delta_{\mathbf{1}} P_t) P_s = \nu_\beta P_s \tag{11.4}$$

for all $s \geq 0$. Here $(P_t)_0^\infty$ is the semigroup governing ξ.

It is customary to call a particle system ergodic if and only if there is a unique stationary distribution and λP_t converges weakly to that as $t \to \infty$ for any initial distribution λ; we adopt that convention in this section. $\delta_{\mathbf{0}}$ is a stationary distribution, of course. The probability for nonextinction is maximized if we let $\xi_0 = \mathbf{1}$; hence if a critical value β_c exists, then $\nu_\beta \neq \delta_{\mathbf{0}}$ for $\beta > \beta_c$, and we have two distinct stationary distributions and no ergodicity.

For $0 \leq \alpha \leq 1$, $\alpha \cdot \delta_{\mathbf{0}} + (1 - \alpha) \cdot \nu_\beta$ is a stationary distribution. Are the possibilities exhausted by these linear combinations, or is the class of such distributions richer? With that difficult question we leave the contact process behind for the moment.

The second example of interacting particle systems is the voter model. Again $S = \mathbb{Z}^d$. In the simplest version of this model

$$c(x, \eta) = (2d)^{-1} \cdot \sum_{\substack{y \\ |y-x|=1}} I_{\{\eta(y) \neq \eta(x)\}} . \tag{11.5}$$

We interpret $\eta(x) = 1$ or 0 to mean that the person at x is for or against a certain proposal. He or she sticks to this opinion for intervals that are Exp(1) distributed, then picks one of the $2d$ (nearest) neighbors at random and adopts the opinion of that person (possibly the same one that he or she already has). If you are not attracted by this as a model for voting, you may use it for spatial conflicts. Indeed, let $\{x; \eta(x) = 1\}$ and $\{x; \eta(x) = 0\}$ represent the territories held by each of two competing populations, or fighting armies.

Since **0** and **1** are absorbing, the voter model is not ergodic. But are there more stationary distributions than the linear combinations of δ_0 and δ_1, and if we restrict the initial distribution in some suitable way, do we then have asymptotic stationarity? Another interesting question: Is stochastic domination retained as time passes?

Our third and last example of a spin system is the most celebrated, the (stochastic) Ising model. For convenience, let now $E = \{-1, 1\}^S$; S is again $= \mathbb{Z}^d$. Think of the possible values ± 1 of $\eta(x)$ to represent the spin of an iron atom at x. The flip rates are, for the simplest version of the Ising model,

$$c(x, \eta) = \exp\Big[-\beta \cdot \eta(x) \cdot \sum_{|y-x|=1} \eta(y)\Big], \tag{11.6}$$

where β is a constant ≥ 0. If $\beta = 0$, we have no interaction between the atoms, whereas if $\beta > 0$, the flip rates force $\eta(x)$ to have the same sign as the majority of the neighboring η values, roughly speaking.

You should now expect which questions are interesting concerning the Ising model. Is there a unique stationary distribution, and if so, do we have ergodicity? If not so, how many extremal stationary distributions are there? Do the answers to these questions depend on d and β?

12. The Vasershtein coupling. For $\eta \in E$ and $x \in S$, let

$$\eta_x(y) = \begin{cases} \eta(y) & \text{if } y \neq x \\ 1 - \eta(y) & \text{if } y = x, \end{cases} \tag{12.1}$$

and for two flip rate functions c and c', let

$$c_0(x, \eta, \eta') = \min(c(x, \eta), c'(x, \eta')). \tag{12.2}$$

The function c_0 comes into use only for arguments $x \in S$ and η, $\eta' \in E$ such that $\eta(x) = \eta'(x)$.

In Vasershtein coupling, also called the basic coupling, of two particle systems with flip rate functions c and c', respectively, we have transitions as follows:

(12.3)

$$\begin{array}{lll}
\text{from} & \text{to} & \text{with intensity} \\[6pt]
(\eta, \eta') & (\eta_x, \eta') & c(x, \eta) \\
 & (\eta, \eta'_x) & c'(x, \eta')
\end{array}$$

when $\eta(x) \neq \eta(x)$, and

$$\begin{array}{lll}
(\eta, \eta') & (\eta_x, \eta'_x) & c_0(x, \eta, \eta') \\
 & (\eta_x, \eta') & c(x, \eta) - c_0(x, \eta, \eta') \\
 & (\eta, \eta'_x) & c'(x, \eta') - c_0(x, \eta, \eta')
\end{array}$$

when $\eta(x) = \eta'(x)$,

where x ranges over S. If $\tilde{\xi}$ is the Markov process with state space E^2 governed by the intensities of (12.3), then $\tilde{\xi}$ has components ξ and $\xi'(\tilde{\xi}_t = (\xi_t, \xi'_t))$ with the flip rate functions c and c'.

For applications to follow, it is important to notice that if c and c' satisfy

(12.4) $$\begin{array}{lll}
c(x, \eta) \leq c'(x, \eta') & \text{if} \quad \eta(x) = \eta'(x) = 0, & \text{and} \\
c(x, \eta) \geq c'(x, \eta') & \text{if} \quad \eta(x) = \eta'(x) = 1 &
\end{array}$$

whenever $\eta \leq \eta'$, then $\xi_t \leq \xi'_t$ for all $t \geq 0$ if $\xi_0 \leq \xi'_0$. A well-known consequence of this is that if λ and μ are initial distributions satisfying $\lambda \overset{\mathcal{D}}{\leq} \mu$, then

(12.5) $$\lambda P_t \overset{\mathcal{D}}{\leq} \mu P_t \qquad \text{for all} \quad t \geq 0\,,$$

where $(P_t)_0^\infty$ and $(P'_t)_0^\infty$ are the semigroups given by c and c'. To obtain (12.5), use Strassen's theorem again to get initial values ξ_0, ξ'_0 satisfying $\xi_0 \leq \xi'_0$. The pathwise domination yields

$$\lambda P_t \overset{\mathcal{D}}{=} \xi_t \leq \xi'_t \overset{\mathcal{D}}{=} \mu P'_t\,.$$

13. Attractiveness and monotonicity. If (12.4) holds when $c = c'$, the spin system governed by c is called attractive. The voter model is a typical example of an attractive spin system; check that and you will understand why the term "attractive" has been coined. The contact process and our version of the Ising model are also attractive spin systems; hence stochastic domination is retained as time passes for all three of our example processes, due to (12.4).

We now find (cf. § 4) that

(13.1) (i) $\delta_0 P_t$ is stochastically increasing in t.

(ii) $\delta_1 P_t$ is stochastically decreasing in t, and

(iii) $\delta_0 P_t \overset{\mathcal{D}}{\leq} \lambda P_t \overset{\mathcal{D}}{\leq} \delta_1 P_t$ for all initial distributions λ and all $t \geq 0$.

When applied to (13.1), argument (11.4) and the discussion preceding it yield that

(13.2) (i) $\underline{\nu} = \lim_{t\to\infty} \delta_0 P_t$ and $\bar{\nu} = \lim_{t\to\infty} \delta_1 P_t$ exist and are stationary distributions,

(ii) if λ and $(t_n)_0^\infty$ are such that $\nu = \lim_{t_n\to\infty} \lambda P_{t_n}$ exists, then $\underline{\nu} \overset{\mathcal{D}}{\leq} \nu \overset{\mathcal{D}}{\leq} \bar{\nu}$, and

(iii) if π is a stationary distribution, then $\underline{\nu} \overset{\mathcal{D}}{\leq} \pi \overset{\mathcal{D}}{\leq} \bar{\nu}$.

We call $\underline{\nu}$ and $\bar{\nu}$ the lower and the upper stationary distribution, respectively. Now notice that the class of stationary distributions constitutes a convex set in the space of signed bounded measures on E. A distribution in that set is called extremal if it is not a proper linear combination of any two distinct distributions in the same set. It is convenient and illuminating to describe the class of stationary distributions of a particle systems in terms of its extremal elements. Actually, it agrees with the closed convex hull of the extremals.

As you may have conjectured, $\underline{\nu}$ and $\bar{\nu}$ are extreme. To prove that for $\bar{\nu}$, for example, let $\bar{\nu} = \alpha \cdot \mu_1 + (1-\alpha) \cdot \mu_2$, where $0 < \alpha < 1$ and μ_1, μ_2 are stationary distributions. Due to (13.2)(iii), we have

$$\int f \, d\mu_1 \leq \int f \, d\bar{\nu} \quad \text{and} \quad \int f \, d\mu_2 \leq \int f \, d\bar{\nu} \tag{13.3}$$

for $f \in i\mathscr{C}$. But

$$\int f \, d\bar{\nu} = \alpha \cdot \int f \, d\mu_1 + (1-\alpha) \cdot \int f \, d\mu_2 \, ,$$

which forces the inequalities in (13.3) to equalities. Actually, the integrals of the functions $\in i\mathscr{E}$ determine the measure with respect to which we integrate, hence $\bar{\nu} = \mu_1 = \mu_2$ follows.

With the current definition of ergodicity, that property of a particle system implies that there is a unique stationary distribution. For an attractive spin system, the reverse is true. In fact, the following properties are equivalent:

(13.4) (i) we have ergodicity,
(ii) there is a unique stationary distribution, and
(iii) $\underline{\nu} = \bar{\nu}$.

For a proof, recall that $\underline{\nu}$ and $\bar{\nu}$ are stationary, hence (ii)$\Rightarrow$(iii). The implications (iii)$\Rightarrow$(ii) and (iii)$\Rightarrow$(i) follow from (13.2)(iii) and (13.2)(ii), respectively.

We are now, at least to some extent, able to illuminate the answers to the questions raised concerning the example processes in § 11.

14. On the example processes. For the contact process with parameter β, let ν_β be the upper stationary distribution: $\nu_\beta = \lim_{t\to\infty} \delta_1 P_t$. It is a consequence of (12.4) that

(14.1) $$\nu_{\beta_1} \overset{\mathscr{D}}{\leq} \nu_{\beta_2} \quad \text{if} \quad \beta_1 \leq \beta_2 .$$

Hence the function ρ defined by

(14.2) $$\rho(\beta) = \nu_\beta\{\eta; \eta(x) = 1\}$$

(which does not depend on x) is nondecreasing in β.

The critical value of the contact process is

(14.3) $$\begin{aligned}\beta_c &= \sup\{\beta \geq 0; \nu_\beta = \delta_{\mathbf{0}} \text{ (the process dies out)}\} \\ &= \inf\{\beta \geq 0; \rho(\beta) > 0\} .\end{aligned}$$

It is hard to estimate β_c. Using ingenious methods, one has found inter alia that

(14.4) (i) $d \cdot \beta_{dc} \le \beta_{1c}$,
(ii) $1/(2d-1) \le \beta_{dc} \le 2/d$, and
(iii) $\lim_{d\to\infty} d \cdot \beta_{dc} = \frac{1}{2}$.

Here β_{dc} is β_c, with indication of dimension. Not even the fact that $\beta_c < \infty$, although intuitively obvious, seems to have a short proof. A cunning coupling may be used to prove (i).

In one dimension, there are no extremal stationary distributions except δ_0 and ν_β. A sharper result is the following: For $\beta > \beta_c$ we have

$$\lim_{t\to\infty} \lambda P_t = \gamma \cdot \delta_0 + (1-\gamma) \cdot \nu_\beta \tag{14.5}$$

for all initial distributions λ, where γ is the extinction probability under $\mathbf{P}_\lambda$. When restricted to translation invariant λ measures, (14.5) holds also for $d \ge 2$.

For $d \le 2$, the only extremal stationary distributions for the voter model are δ_0, and δ_1, while for $d \ge 3$ there is a continuum $\{\mu_\alpha; 0 \le \alpha \le 1\}$ of translation invariant such distributions. We omit the opportunity to demonstrate any of the extensive use of couplings in the study of the voter model; one possibility would have been a sketch of a proof of the result that for $d \le 2$,

$$\lim_{t\to\infty} \lambda P_t = \alpha \cdot \delta_1 + (1-\alpha) \cdot \delta_0 \tag{14.6}$$

for initial distributions λ that are translation invariant. Here $\alpha = \lambda\{\eta; \eta(x) = 1\}$. Hence, roughly speaking, a consensus among the voters is approached if $d \le 2$, but not if $d \ge 3$.

Recall the Ising model. Is the initial spin configuration always forgotten, so that $\underline{\nu} = \bar{\nu}$? The distribution $\underline{\nu}$ is now $\lim_{t\to\infty} \delta_{-1} P_t$. The answer is yes if $d = 1$, but for $d \ge 2$ there is a $\beta_d > 0$ such that $\underline{\nu} = \bar{\nu}$ for $\beta < \beta_d$ and $\underline{\nu} \ne \bar{\nu}$ for $\beta > \beta_d$. Hence if the interaction is strong enough, we do not have ergodicity for $d \ge 2$. Another interesting result: There are at most two extremal stationary distributions if $d \le 2$, while if $d \ge 3$, there are infinitely many such distributions if β is large enough.

15. Notes. It is due time to finish the text of the signpost by telling you where to go. The major account of interacting particle systems is Liggett [104]; you will find all the details of topics mentioned above in that masterly written book. See also Durrett [57].

4. EMBEDDING IN POISSON PROCESSES

16. A multivariate exponential distribution. Our main application of embedding in a Poisson process is to renewal theory, presented in Part 5. However, the device has several other interesting uses you should be aware of; it deserves a section of its own.

We say that a variable $X = (X_1, X_2)$ has a bivariate exponential distribution if each of its coordinates X_1, X_2 is exponentially distributed in one dimension. What is the simplest interesting example of such a variable (we do not count the case when X_1, X_2 are independent)? Think of a system suffering from shocks according to a Poisson process, but in which not all components are stricken. Indeed, let us have two particular components in focus, and let N_1, N_2, and N_{12} be three independent Poisson processes, with intensities λ_1, λ_2, and λ_{12}, respectively. The points of occurrence of these processes indicate the times of shocks to the first, second, and both components, respectively. All the shocks then appear according to the Poisson process N^*, where

$$N^* = N_2 + N_2 + N_{12} \,. \tag{16.1}$$

With Z_1, Z_2, and Z_{12} denoting the times of the first points of occurrence in N_1, N_2, and N_{12}, respectively, let

$$\begin{aligned} X_1 &= \min(Z_1, Z_{12}) \,, \\ X_2 &= \min(Z_2, Z_{12}) \,. \end{aligned} \tag{16.2}$$

Clearly, $X_i \stackrel{\mathcal{D}}{=} \mathrm{Exp}(\alpha_i)$, where $\alpha_i = \lambda_i + \lambda_{12}$ for $i = 1, 2$ since X_i is the first point in $N_i + N_{12}$, a Poisson process with intensity α_i.

Notice that

$$\min(X_1, X_2) \stackrel{\mathcal{D}}{=} \mathrm{Exp}(\alpha) \,, \tag{16.3}$$

where $\alpha = \lambda_1 + \lambda_2 + \lambda_{12}$. In the next section you will find the solution to the following problem: Give an example of a bivariate exponential variable such that $\min(X_1, X_2)$ is not exponentially distributed.

For an extension of the construction (16.2), let

(16.4) $S = \{\mathbf{i} = (i_1, \ldots, i_k);\ i_j = 0 \text{ or } 1 \text{ and } i_j = 1 \text{ for at least one } j,\ 1 \le j \le k\}$,

and let $N_\mathbf{i}$, $\mathbf{i} \in S$, be independent Poisson processes with intensities $\lambda_\mathbf{i}$. The meaning of $N_\mathbf{i}$ for a system with k components is obvious. Let $Z_\mathbf{i}$ be the time of the first point of $N_\mathbf{i}$, and define for $1 \le j \le k$

(16.5) $$X_j = \min\{Z_\mathbf{i};\ i_j = 1\}\ .$$

Then $X_j \overset{\mathcal{D}}{=} \mathrm{Exp}(\alpha_j)$, where $\alpha_j = \Sigma_{\mathbf{i}, i_j = 1}\ \lambda_\mathbf{i}$.

The properties of $X = (X_1, \ldots, X_k)$ are left to the reader to explore. One exercise: Find a measure on $(\mathbb{R}_+^k, \mathcal{R}_+^k)$ with respect to which the distribution of X is absolutely continuous; we need that in order to do a maximum likelihood estimation.

17. Embedding in a bivariate Poisson process. For the rest of this chapter, let l denote the Lebesgue measure on $(\mathbb{R}_+^2, \mathcal{R}_+^2)$ and ξ a Poisson process on $(\mathbb{R}_+^2, \mathcal{R}_+^2)$ with expectation measure l, that is, a point process such that

(17.1) (i) $\xi(B_1), \ldots, \xi(B_n)$ are independent for any class $B_1, \ldots, B_n$ of disjoint sets $\in \mathcal{R}_+^2$, and
(ii) $\xi(B) \overset{\mathcal{D}}{=} \mathrm{Poi}(l(B))$ for $B \in \mathcal{R}_+^2$.

If A is a measurable set such that $l(A \cap (B \times \mathbb{R}_+)) < \infty$ for all bounded $B \in \mathcal{R}_+$, then

(17.2) $$N = \xi(A \cap (\cdot \times \mathbb{R}_+))$$

is a one-dimensional Poisson process with expectation measure $l(A \cap (\cdot \times \mathbb{R}_+))$. In particular, with $A = \mathbb{R}_+ \times [a, a + \lambda]$, N is a Poisson process with intensity λ.

We may use (17.1)(i) and (17.2) to construct a class of independent Poisson processes N_i, $1 \le i < \infty$, with prescribed expectation measures m_i. Indeed, let $A_i \in \mathcal{R}_+^2$ be disjoint sets such that

(17.3) $$m_i = l(A_i \cap (\cdot \times \mathbb{R}_+))\ .$$

Then the processes given by (17.2), $N_i = \xi(A_i \cap (\cdot \times \mathbb{R}_+))$, have the desired properties. In particular, with prescribed intensities $\lambda_1, \lambda_2, \ldots$, we may use

$$A_i = \mathbb{R}_+ \times \Big[\sum_1^{i-1} \lambda_j, \sum_1^{i} \lambda_j \Big) \tag{17.4}$$

to construct independent Poisson processes with intensities λ_i, $i \geq 1$. Use this device to get N_1, N_2, and N_{12} of the previous section. This is not only convenient, but also gives us a means to visualize the variables X_1 and X_2, and their dependence: In the first quadrant, use the strips $\mathbb{R}_+ \times [0, \lambda_1)$, $\mathbb{R}_+ \times [\lambda_1, \lambda_1 + \lambda_2)$, and $\mathbb{R}_+ \times [\lambda_1 + \lambda_2, \lambda_1 + \lambda_2 + \lambda_{12})$ for the generation of N_1, N_2, and N_{12}, mark the first few points of occurrence of ξ in the strip $\mathbb{R}_+ \times [0, \lambda_1 + \lambda_2 + \lambda_{12})$, and indicate X_1 and X_2 to obtain a diagram.

To develop the embedding method further, for any set $B \in \mathcal{R}_+^2$ and $t \geq 0$, let

$$B_t = \{(x, y) \in B; x \leq t\} \tag{17.5}$$

and

$$\tau_B = \inf\{t \geq 0; \xi(B_t) \geq 1\} .$$

We have

$$\mathbf{P}(\tau_B > x) = \mathbf{P}(\xi(B_x) = 0) = \exp(-l(B_x)) ; \tag{17.6}$$

hence $\tau_B \stackrel{\mathcal{D}}{=} \mathrm{Exp}(\alpha)$ for an $\alpha \geq 0$ if and only if $l(B_x) = \alpha \cdot x$ for all $x \geq 0$. But this is true not only for the strips used above, but also for any set of the type

$$\{(x, y) \in \mathbb{R}_+^2; f(x) \leq y \leq f(x) + \alpha\} , \tag{17.7}$$

where f is a nonnegative function $\in \mathcal{R}_+$.

We are now ready to solve the exercise concerning $\min(X_1, X_2)$, where (X_1, X_2) has a bivariate exponential distribution. Let $A_1 = \{(x, y) \in \mathbb{R}_+^2;\ 0 \leq y \leq 1\}$ and $A_2 =$ the set of (17.7) with $f(x) = 1 + \sin(x)$, for example. Then $X_1 = \tau_{A_1}$ and $X_2 = \tau_{A_2}$ are both Exp(1) distributed, which is not the case for $\min(X_1, X_2) = \tau_{A_1 \cup A_2}$.

The next topic is crucial for Part 5. Let F be a distribution on $(\mathbb{R}_+, \mathcal{R}_+)$. Does there exist a $B \in \mathcal{R}_+^2$ such that $\tau_B \overset{\mathcal{D}}{=} F$? For this to hold, we understand that

$$\exp(-l(B_x)) = \bar{F}(x) \quad \text{for} \quad x \geq 0 \tag{17.8}$$

is required. Now recall the definition (III.2.3) of the failure rate function r of F, and the crucial fact that

$$\bar{F}(x) = \exp(-R(x)) \quad \text{for} \quad x \geq 0\,, \tag{17.9}$$

where $R(x) = \int_0^x r(s)\,ds$. Hence (17.8) holds for any set B such that

$$l(B_x) = R(x) \quad \text{for} \quad x \geq 0\,. \tag{17.10}$$

A natural choice of such a set is

$$B = \{(x, y);\, y \leq r(x)\}\,. \tag{17.11}$$

We notice that a restriction to finite random variables is not necessary; with a $B \in \mathcal{R}_+^2$ such that $l(B) < \infty$, we have $\mathbf{P}(\tau_B = \infty) = \exp(-l(B)) > 0$.

We may now generalize the multivariate exponential distribution and its graphical representation. Recall (16.4), and let $F_{\mathbf{i}}$, $\mathbf{i} \in S$, be distributions on $(\bar{\mathbb{R}}_+, \bar{\mathcal{R}}_+)$ with failure rate functions $r_{\mathbf{i}}$, respectively. Choose sets $B_{\mathbf{i}} \in \mathcal{R}_+^2$ such that

$$l((B_{\mathbf{i}})_x) = R_{\mathbf{i}}(x) \quad \text{for} \quad x \geq 0\,, \tag{17.12}$$

where $R_{\mathbf{i}}(x) = \int_0^x r_{\mathbf{i}}(s)\,ds$. The variable $X = (X_1, \ldots, X_k)$ is now defined by

$$X_j = \min_{\substack{\mathbf{i} \\ i_j = 1}} (\tau_{B_{\mathbf{i}}}), \qquad 1 \leq j \leq k\,, \tag{17.13}$$

where $\mathbf{i} = (i_1, \ldots, i_j, \ldots, i_k)$; X_j is the time of failure of the jth component. Notice that X_j has failure rate function r_j given by

$$r_j(x) = \sum_{\substack{\mathbf{i} \\ i_j = 1}} r_{\mathbf{i}}(x), \qquad x \geq 0\,,$$

and that X_j and X_m are independent if and only if B_j and B_m are disjoint, where

$$B_n = \bigcup_{i_n=1} B_{\mathrm{i}}, \qquad 1 \le n \le k\,.$$

When drawing a first-quadrant diagram with the sets B_{i}, the restrictions (17.12) allow considerable freedom for your graphical taste.

Our last topic will be crucial for the urn problems to be considered next. Now let ξ be a Poisson process in $\mathbb{R}_+ \times [0, 1)$, with expectation measure = the Lebesgue measure on that set. Let (T_i, η_i), $i \ge 1$, be the enumeration of the points of occurrence of ξ satisfying $T_1 < T_2 < \cdots$. Due to (17.2), the points $(T_i)_0^\infty$ constitute a one-dimensional Poisson process with intensity 1; we denote that by N. We shall prove that

(17.14) (i) the variables η_i, $i \ge 1$, are independent and Uni[0, 1)-distributed, and
(ii) N and $(\eta_i)_1^\infty$ are independent.

These properties are folklore, but it is common to leave results like these with an incomplete proof or no proof; we shall be a bit more serious. Recall that the distribution of a point process ζ on $(\mathbb{R}^d, \mathcal{R}^d)$, $d \ge 1$, is determined by its Laplace functional

$$L_\zeta(f) = \mathbf{E}[\exp(-\zeta f)]$$

$[\zeta f = \int f(x)\, d\zeta(x)]$, defined for nonnegative $f \in b\mathcal{R}^d$ with bounded support, and that a Poisson process with expectation measure ν has Laplace functional

$$L(f) = \exp\Big(\int (e^{-f} - 1)\, d\nu\Big)\,. \tag{17.15}$$

Let N' be a Poisson process with intensity 1, and η_i', $i \ge 1$, independent Uni[0, 1)-distributed variables, also independent of N'. Define the point process ξ' in $\mathbb{R}_+ \times [0, 1)$ by

$$\xi' = \sum_{i=1}^\infty \delta_{(T_i', \eta_i')}\,,$$

where $T_1', T_2', \ldots$ are the points of N' in increasing order. If we can establish that ξ and ξ' have the same Laplace transform, we have proved (17.14). Let f be a test function, defined on $\mathbb{R}_+ \times [0,1)$, of the type specified above. We have

$$(17.16) \qquad L_\xi(f) = \exp\left\{ \int_{u=0}^{\infty} \int_{s=0}^{1} (e^{-f(s,u)} - 1)\, ds\, du \right\}$$

$$= \exp\left\{ \int_0^\infty (g(u) - 1)\, du \right\},$$

where $g(u) = \int_0^1 e^{-f(s,u)}\, ds$. Also,

$$(17.17) \qquad L_{\xi'}(f) = \mathbf{E}[\exp(-\xi' f)] = \mathbf{E}[\mathbf{E}[\exp(-\xi' f) \| N]]$$

$$= \mathbf{E}\left[\prod_{i=1}^{\infty} \left\{ \int_0^1 e^{-f(T_i', u)}\, du \right\} \right]$$

$$= \mathbf{E}\left[\prod_{i=1}^{\infty} g(T_i') \right] = L_N(-\log g)$$

$$= \exp\left(\int_0^\infty (g(u) - 1)\, du \right),$$

where the last equality is due to (17.15), applied to N'. Hence (17.14) is proved.

18. Urns and boxes. Like coins and dice, urns with balls are basic requisites in probability and statistics. A typical problem: From an urn with balls of n different colors, independent random one-ball drawings with replacement are made until a particular configuration appears (e.g., each color is represented a specified number of times). An alternative formulation is in terms of balls falling independently and randomly into one of n boxes, or cells; we choose that one here.

A Uni[0, 1)-distributed variable η may be used to decide which box a ball falls into. Indeed, if box k has probability p_k, we let a ball fall into that if

$$(18.1) \qquad \sum_1^{k-1} p_j \le \eta < \sum_1^{k} p_j .$$

We shall consider only the case when all the p_k values are equal to $1/n$.

Recall the Poisson process ξ on $\mathbb{R}_+ \times [0, 1)$, and its associated $(T_i)_1^\infty$, N, and $(\eta_i)_1^\infty$. What follows leans heavily on (17.14). The variables η_i, $i \leq 1$, will be used to determine which box the ith ball falls into according to the principle (18.1).

Now let W denote the number of balls used in a certain box model. The analysis of such variables often leads to difficult or tedious combinatorial problems; embedding may save you from them.

We shall let the balls fall at the times T_i, $i \geq 1$. If we split N into the n process N_k, $1 \leq k \leq n$, where

$$N_k = \xi(\cdot \times [(k-1)/n, k/n)) ,$$

then

(18.2) (i) $N_1, N_2, \ldots$ are independent Poisson processes with intensity $1/n$, and
(ii) $N = \Sigma_1^n N_k$.

Let the points of occurrence of N_k in increasing order be denoted by T_{kj}, $j \geq 1$. Notice that

(18.3) T_{kj} = the time when the jth ball falls into box k.

We use Z_i, $i \geq 1$, to denote the interoccurrence times of N: $Z_i = T_i - T_{i-1}$, and let Z_{ki}, $i \geq 1$, have the analogous meaning for N_k. The variables Z_i are i.i.d. and Exp(1)-distributed, while the Z_{ki} variables are i.i.d. and Exp($1/n$)-distributed.

If S denotes the time when the last ball falls, then

$$S = \sum_{i=1}^{W} Z_i . \tag{18.4}$$

Since $(Z_i)_1^\infty$ and W are independent, we have

$$L_S(t) = g_W((1+t)^{-1}) , \tag{18.5}$$

where L_S is the Laplace transform of S ($= \mathbf{E}[\exp(-tS)]$, $t \geq 0$) and

g_W the p.g.f. of W; (18.5) is achieved by conditioning on W and use of the fact that the Laplace transform of Z_i equals $(1+t)^{-1}$. Due to (18.5), the distribution of S determines that of W, and relations between moments of S and W may be obtained through derivation. Concerning the expectations, we turn directly to (18.4) and get

$$\mathbf{E}[W] = \mathbf{E}[S]\,. \tag{18.6}$$

The relations (18.4)–(18.6) make it possible to find properties of W via S. Due to (18.3), that is a fruitful idea, as we now show in two examples. In the first, let W be the number of balls that have fallen when we observe that there are at least m balls in each box.

$$S = \max_{1 \le k \le n} T_{km}\,. \tag{18.7}$$

Denote T_{km}/n by Y_k. The Y_k variables are independent and $\Gamma(m, 1)$-distributed, and (18.7) brings us into extreme value theory. The $\Gamma(m, 1)$ distribution function equals

$$H_m(x) = 1 - \sum_{i=0}^{m-1} x^i e^{-x}/i!\,, \qquad x \ge 0\,. \tag{18.8}$$

Of course,

$$\mathbf{P}(S_n/n \le x) = H_m(x)^n\,.$$

For any y, let $a_n(y) = a_n$ be such that

$$e^{-y}/n = 1 - H_m(y + a_n)\,.$$

Then

$$\mathbf{P}(S_n/n \le y + a_n) \to \exp(-e^{-y}) \quad \text{as} \quad n \to \infty \tag{18.9}$$

for all y; S_n is S, with the number of boxes indicated. Hence

$$S_n/n - a_n \xrightarrow{\mathcal{D}} G \quad \text{as} \quad n \to \infty\,, \tag{18.10}$$

where $G(y) = \exp(-e^{-y})$, a distribution well known from extreme value theory.

But this is interesting only if a_n is asymptotically independent on y. Some calculations render

$$a_n = \log n + (m-1)\cdot \log\log n - \log(m-1)! + o(1)\,,$$

and that is actually the trickiest part in our proof of (18.10), a demanding result without the tools presented here.

Now recall (18.4). We have

$$S_n/n = \Big(\sum_1^{W_n} Z_i\Big)/W_n \cdot (W_n/n)\,, \tag{18.11}$$

with a number of boxes indicated on W. Since $W_n \to \infty$ as $n \to \infty$ and W_n and $(Z_i)_1^\infty$ are independent, we have $(\Sigma_1^{W_n} Z_i)/W_n \to E[Z_i] = 1$ in probability, and we conclude easily that

$$W_n/n - a_n \xrightarrow{\mathscr{D}} G \quad \text{as} \quad n \to \infty\,. \tag{18.12}$$

For the second example, let W now be the number of balls that fall until one box has m of them. This time we have

$$S = \min_{1\le k\le n} T_{km} = n \cdot \min_{1\le k\le n} Y_k\,. \tag{18.13}$$

With $\beta_n = (m!/n)^{1/m}$ we have

$$\begin{aligned}\mathbf{P}(\min_{1\le k\le n} Y_k/\beta_n > y) &= \bar{H}_m(\beta_n \cdot y)^n = \Big(1 - \sum_m^\infty (\beta_n y)^i \cdot e^{-\beta_n y}/i!\Big)^n \\ &= (1 - (y^m + o(1))/n)^n\,,\end{aligned}$$

which tends to $\exp(-y^m) = \bar{G}_1(y)$ as $n \to \infty$. Hence, with $b_n = n\cdot\beta_n$ and $S_n = S$, we get

$$S_n/b_n \xrightarrow{\mathscr{D}} G_1 \quad \text{as} \quad n \to \infty\,. \tag{18.14}$$

We recognize G_1 as a standardized Weibull distribution; contemplate why such a lifelength distribution, well known from reliability theory, appears here!

To come from (18.10) to (18.12), the crucial observation was that $\Sigma_1^{W_n} Z_i/W_n \xrightarrow{P} 1$ as $n \to \infty$. But this holds if $W_n \xrightarrow{P} \infty$, so the present $W_n = W$ has that property. Using arguments similar to those for (18.12), we obtain

$$W_n/b_n \xrightarrow{\mathscr{D}} G_1 \quad \text{as} \quad n \to \infty. \tag{18.15}$$

19. On free parking spaces. We consider a parking lot with spaces labeled 1, 2, ... according to their attractiveness. Assume that

(19.1) (i) cars arrive as a Poisson process with intensity λ to the lot,

(ii) an arriving car parks in the free space with lowest number, and

(iii) each car occupies its space for a time that is Exp(1)-distributed.

A natural choice of state space for this process is $E^\circ = \{0, 1\}^{\mathbb{N}}$, where $x_i = 1$ for $x = (x_i)_1^\infty \in E^\circ$ represents that space i is occupied. Let the process be denoted by $Y = (Y_t)_0^\infty$. The behavior of Y as $\lambda \to \infty$ (heavy traffic) is rewarding to study; we shall briefly consider the asymptotics at the parking places with numbers $1, 2, \ldots, [\rho\lambda]$, where $0 < \rho < 1$. It turns out that it is natural to formulate this in terms of the free places.

To construct what shall be the limit process, let ξ be a Poisson process on $\mathbb{R}_+ \times [0, \rho]$ with expectation measure = the Lebesgue measure restricted to that set. Also, let N^* be a Poisson process on $\mathbb{R}_+$ with intensity 1 and independent of ξ. As a state space, we shall use $\mathcal{N}_{[0,\rho]}$ = the integer-valued finite measures on $[0, \rho]$ (cf. § III.1). Abbreviate $\mathcal{N}_{[0,\rho]}$ as $\mathcal{N}_\rho$.

Now let $X = (X_t)_0^\infty$ be the GBD process with state space $\mathcal{N}_\rho$ defined as follows: Whenever a point of occurrence (T_i, η_i) of ξ appears, we add δ_{η_i} to X_{T_i-}, and for a point in N^* we delete the leftmost point of X at that time, if there is any. Recall (17.14) for the properties of $(T_i)_1^\infty$ and $(\eta_i)_1^\infty$; in particular, the T_i variables constitute a Poisson process, N say, with intensity ρ.

The process X is irreducible and positive recurrent; hence it has a unique stationary distribution. Indeed, in $X_t[0, \rho]$ = the total number of points of X_t we recognize the birth and death process counting customers in an $M/M/1$ queueing system, and with the present parameter values that process is positive recurrent. Hence so is X, with $\mathbf{0} \in \mathcal{N}_\rho$ as a positive recurrent state.

Let $\nu, \nu' \in \mathcal{N}_\rho$ be two initial states satisfying $\nu \leq \nu'$. To obtain a good coupling (X, X') with $X_0 = \nu$ and $X_0' = \nu'$, use ξ and N^* to add and delete points of occurrence in both X and X'. We then obviously get $X_t \leq X_t'$ for all $t \geq 0$. If we let $X_0 \equiv \mathbf{0}$ and X' be stationary, then $X_t = X_t'$ for times t after X' hits $\mathbf{0}$; hence we have a successful coupling and the asymptotic stationarity of X follows. Notice that Strassen's theorem may be used to get initial values X_0, X_0' satisfying $X_0 \leq X_0'$ as soon as their distributions λ, μ satisfy $\lambda \overset{\mathcal{D}}{\leq} \mu$.

Now recall the Y process, and write Y_t as $(Y_i(t))_1^\infty \in E^\circ$ whenever convenient. Letting

$$Y_t^* = \sum_{i=1}^{[\lambda\rho]} \delta_{i/\lambda} \cdot (1 - Y_i(t/\lambda)) \tag{19.2}$$

for $t \geq 0$, we transform Y to be a process $Y^* = (Y_t^*)_0^\infty$, which

- has a new time scale: cars now arrive with intensity 1,
- describes, on a new space scale, where the free parking places are, and
- satisfies $Y_t^* \in \mathcal{N}_\rho$.

Notice that for the definition of Y^*, no attention is paid to the parking spaces with numbers greater than $[\lambda\rho]$. You may think of the parking lot as having $[\lambda\rho]$ spaces only; when they are occupied, arriving cars just pass.

We are now ready for another embedding, to prove that Y^* is close to X in distribution for large λ values. Indeed, use ξ and N^* to obtain a parking process $\bar{Y} = (\bar{Y}_t)_0^\infty$ as follows:

(19.3) (i) cars arrive according to N^* and a new one parks in the lowest-numbered free space. If no space (there are $[\lambda\rho]$ of them) is free, the car passes. Further,

(ii) for every point of occurrence (T_i, η_i) of ξ, the car (if any) at space $([\eta_i \cdot [\rho\lambda]/\rho] + 1)$ leaves.

Notice that $[\eta_i \cdot [\rho\lambda]/\rho] + 1$ is uniformly distributed on $1, 2, \ldots, [\rho\lambda]$ because $\eta_i \overset{\mathscr{D}}{=} \mathrm{Uni}[0, \rho]$. With

$$(19.4) \qquad Y_t^{**} = \sum_{i=1}^{[\lambda\rho]} \delta_{i/\lambda} \cdot (1 - \bar{Y}_i(t)),$$

it is quite clear that $Y^{**} = (Y_t^{**})_0^\infty \overset{\mathscr{D}}{=} Y^*$.

From (19.3)(ii) we understand that Y^{**} is virtually a discretization of X. For any $t_0 > 0$, the probability of the set

$$\{\xi \text{ has at most one point of occurrence in each of the sets}$$
$$[0, t_0] \times [i \cdot \rho/[\rho\lambda], (i+1) \cdot \rho/[\rho\lambda]], 0 \le i \le [\rho\lambda] - 1\}$$

tends to 1 as $\lambda \to \infty$. For an outcome in that set, we have

$$(19.5) \qquad \sup_{0 \le x \le \rho} |X_t[0, x] - Y_t^{**}[0, x]| \le \rho/[\rho\lambda] + 1/\lambda$$

uniformly in $t \le t_0$, and that is more than is needed to prove that $Y_t^{**} \overset{\mathscr{D}}{=} Y_t^*$ converges in distribution to X_t as $\lambda \to \infty$, for each $t \ge 0$. In fact, (19.5) may be used to settle a result of weak convergence in $D_{\mathscr{N}_\rho}$ for Y^* as $\lambda \to \infty$, under the condition, say, that $Y_0 \equiv \mathbf{0}$ (no space free at the outset).

20. Notes. The multivariate exponential distribution in § 16 is from Marshall and Olkin [116]. Consult Daley and Vere-Jones [45] concerning Poisson processes on general spaces and Laplace functionals of point processes. For details and further results about urns and boxes, see Holst [76] and Donnelly and Whitt [52]. In § 19, we have only treated one aspect of the parking lot model; for asymptotic theory concerning the places beyond $[\lambda\rho]$, including interesting couplings, see Aldous [3].

5. MORE RENEWAL THEORY

21. Basics. In this concluding section on renewal theory, we assume throughout that the life-length distribution F on $(0, \infty)$ has a density f w.r.t. the Lebesgue measure. Our purpose is to sharpen the results in Chapter III, under conditions on the failure rate function r. The most interesting results are obtained when F has decreasing failure rates (we also say: is of DFR type); that is, when r is decreasing, but also in the increasing failure rate (IFR) case, we state some findings.

Recall § III.1; the notation there will be used. To avoid some trite comments, we shall assume that F does not have a bounded support. This always holds (why?) if F is of DFR type.

The choice B of (17.11) for a set $\in \mathscr{R}^2_+$ satisfying $\tau_B \overset{\mathscr{D}}{=} F$ is now crucial. Fix an $a > 0$ and let

$$B_0 = \{(x, y) \in R^2_+ ;\ y \leq r_a(x)\} \tag{21.1}$$

[remember: $r_a(x) = r(a + x)$, the failure rate function of F_a]; then let

$$Y_0 = \tau(B_0) \tag{21.2}$$

[cf. (17.5)]. Clearly, $Y_0 \overset{\mathscr{D}}{=} F_a$. If $a = 0$, we put $Y_0 = 0$. Now define recursively for $n \geq 1$ (remember: $S_0 = Y_0$)

$$B_n = \{(x, y) \in \mathbb{R}^2_+ ;\ x > S_{n-1},\ y \leq r(x - S_{n-1})\} \tag{21.3}$$

and

$$S_n = \tau(B_n)\,.$$

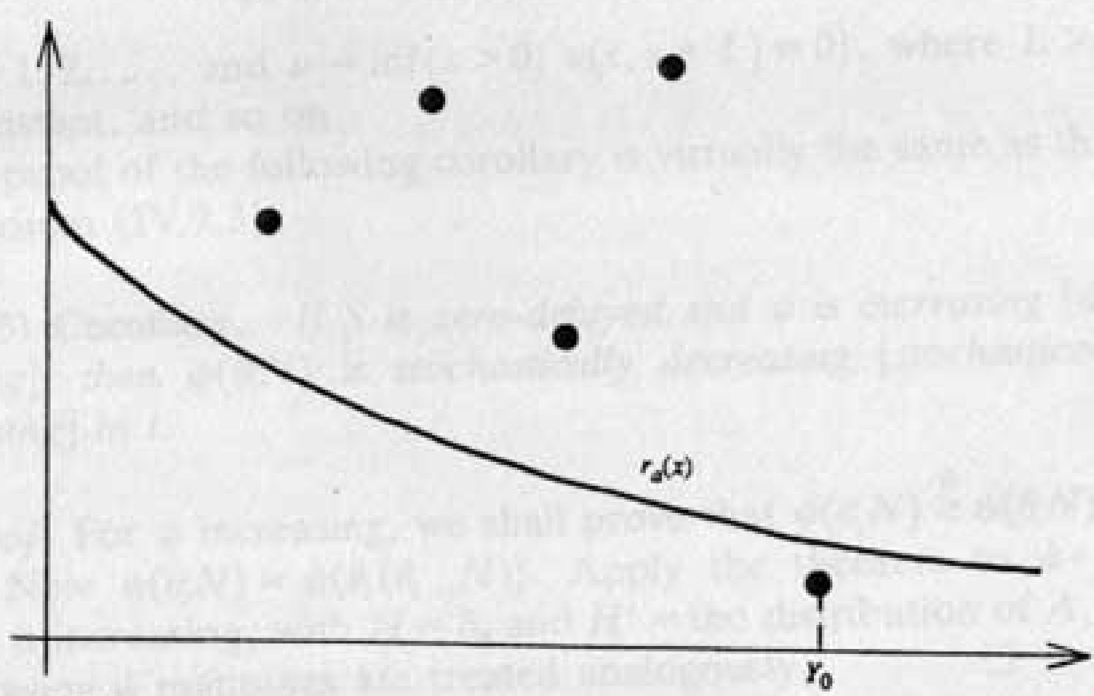

Figure 4. The value of Y_0 equals the x-coordinate of the first point under the curve $r_a(x)$, $x \geq 0$.

This renders $S=(S_n)_0^\infty$ a renewal process, with lifelength distribution F and initial age a. For the construction, only the Poisson process ξ on $(\mathbb{R}_+^2, \mathscr{R}_+^2)$ was used. To study S with a random initial age having distribution H, say, introduce a variable $\eta \overset{\mathscr{D}}{=} H$ independent of ξ, and redefine B_0 to

$$(21.4)\qquad B_0=\begin{cases}\{(x,y)\in R_+^2;\ y\le r_\eta(x)\} & \text{on}\quad \{\eta>0\}\\ \mathbb{R}_+^2 & \text{on}\quad \{\eta=0\}\,.\end{cases}$$

With that definition, no special attention to the zero-delayed case is needed. Use (21.2)–(21.4) to produce S. Of course, $S_0=0$ on $\{\eta=0\}$. Our construction renders the age process $A=(A_t)_0^\infty$ well defined for all $t\ge 0$, with $A_0=\eta \overset{\mathscr{D}}{=} H$.

You should simulate a few points of occurrence of ξ in $\mathbb{R}_+^2$ and indicate the sets B_0, B_1 of (21.2)–(21.4). Once that is done, you surmise the results that follow.

22. The DFR case. Let H, H' be initial age distributions satisfying $H \overset{\mathscr{D}}{\le} H'$. Pick variables η, η' independent of the Poisson process ξ such that

$$H \overset{\mathscr{D}}{=} \eta \le \eta' \overset{\mathscr{D}}{=} H'\,,$$

use (21.4) to obtain sets B_0, B_0' with η, η' inserted, and generate $B_1, B_2, \ldots,$ and $B_1', B_2', \ldots,$ in the obvious way. A quick look at these sets reveals that if F is of DFR type, then for the renewal processes $S=(S_n)_0^\infty$ and $S'=(S_n')_0^\infty$ we have

(22.1) $\qquad S$ and S' coincide from Y_0' and onward;

hence

(22.2) $\qquad$ (i) Y_0' is a coupling time, and
$\qquad\qquad\quad$ (ii) $N \ge N'$.

After recovering from the surprise of (22.2), we first explore (ii). Understand from § IV.8 the meaning that a mapping ψ is increasing or decreasing; we consider the partial ordering on $\mathscr{N}_+$ defined by

$$\nu \le \nu' \text{ iff } \nu(A) \le \nu'(A) \qquad \text{for all} \quad A \in \mathcal{R}_+ .$$

After the discussion above, little is needed to complete the proof of the following result.

(22.3) Theorem. *Suppose that F is of DFR type. If the initial age distributions H, H' of S, S' satisfy $H \overset{\mathcal{D}}{\le} H'$ and ψ is increasing [decreasing], then $\psi(\theta_t N) \overset{\mathcal{D}}{\ge} \psi(\theta_t N')$ [$\psi(\theta_t N) \overset{\mathcal{D}}{\le} \psi(\theta_t N')$] for $t \ge 0$. In particular, if S is zero-delayed and S' stationary, then $\psi(\theta_t N) \overset{\mathcal{D}}{\ge} \psi(N')$ [$\psi(\theta_t N) \overset{\mathcal{D}}{\le} \psi(N')$] for $t \ge 0$.*

Proof. Use (22.2)(ii) and the observation that if ψ is increasing (decreasing), so is $\psi \circ \theta_t$ for all $t \ge 0$. □

If $g \in \mathcal{R}_+$ is nonnegative, then

$$\psi(\nu) = \int g \, d\nu$$

means an increasing mapping. Examples of decreasing mappings are

$$\psi_k(\nu) = \inf\{s > 0;\ \nu(0, s] \ge k\} \tag{22.4}$$

for $k = 1, 2, \ldots$, and $\nu \to \inf\{s > 0;\ \nu(s, s + L] = 0\}$, where $L > 0$ is a constant, and so on.

The proof of the following corollary is virtually the same as that of Theorem (IV.9.3).

(22.5) Corollary. *If S is zero-delayed and ψ is increasing [decreasing], then $\psi(\theta_t N)$ is stochastically decreasing [stochastically increasing] in t.*

Proof. For ψ increasing, we shall prove that $\psi(\theta_s N) \overset{\mathcal{D}}{\ge} \psi(\theta_t N)$ if $s \le t$. Now $\psi(\theta_t N) = \psi(\theta_s(\theta_{t-s} N))$. Apply the theorem to $\psi \circ \theta_s$, which is increasing, with $H = \delta_0$ and H' = the distribution of A_{t-s}. Decreasing ψ mappings are treated analogously. □

As consequences of the corollary, we find that

(22.6) (i) $N(t+B)$ is stochastically decreasing in t for all $B \in \mathcal{R}_+$,
(ii) U is concave,
(iii) the delay D_t is stochastically increasing in t, and
(iv) the age A_t is stochastically increasing in t

in the zero-delayed case. For (i), notice that $\nu \to \nu(A)$ is an increasing mapping. In particular, $N(t, t+h]$ is stochastically decreasing for any $h > 0$, hence $\mathbf{E}[N(t, t+h]] = U(t+h) - U(t)$ is decreasing in t. Hence $U(t+h) - U(t) \geq U(t+2h) - U(t+h)$, so $U(t+h) \geq \frac{1}{2} \cdot (U(t+2h) + U(t))$ for all $h > 0$ and $t \geq h$, which implies (ii). For (iii), make an application to ψ_1 of (22.4).

The result (iv) needs an argument beyond (22.5), due to the problem to define the "age" before the first renewal. Here is one: Extend the state space to $\mathbb{R}_+ \times \mathcal{N}_+$ and find the right decreasing mapping on that space to elaborate!

For the rest of this section, assume that S is zero-delayed and S' stationary. Since (22.2)(ii) may be sharpened to $N \geq N' + 1$ under that assumption, we get

$$U(t) \geq \lambda \cdot t + 1 \quad \text{for} \quad t \geq 0 . \tag{22.7}$$

To obtain an upper bound for U, observe that

$$0 \leq D'_t - D_t \leq (Y'_0 - t)^+ \quad \text{for} \quad t \geq 0 ; \tag{22.8}$$

$D' = (D'_t)_t^\infty$ is the delay process of S'. We have

$$\mathbf{E}[D'_t] = \int x \, dG_s(x) = \lambda \mu_2 / 2 \quad \text{for all} \quad t \geq 0 .$$

Further, $\mathbf{E}[D_t]$ is increasing in t due to (22.6)(iii) and $\mathbf{E}[t + D_t] = \mu \cdot U(t)$ with use of Wald's lemma. Hence taking expectations in (22.8) yields

$$0 \leq \lambda^2 \mu_2 / 2 - (U(t) - \lambda \cdot t) \leq \lambda \cdot \mathbf{E}[(Y'_0 - t)^+] . \tag{22.9}$$

Recalling (22.6)(ii), we may summarize: The concave function

$U(t) - \lambda t$ increases to $\lambda^2\mu_2/2$. To understand how quickly, let us elaborate $\mathbf{E}[(Y'_0 - t)^+]$. Assume $\mu_\beta < \infty$, $\beta \geq 2$. We have

$$\mathbf{E}[(Y'_0 - t)^+] = \int_t^\infty (x-t)\, dG_s(x) = \tfrac{1}{2}\lambda \cdot \int_t^\infty (x-t)^2\, dF(x)\,,$$

which is of magnitude $o(t^{-(\beta-2)})$; the integral is dominated by $\mu_\beta \cdot c_\beta \cdot t^{-(\beta-2)}$, where $c_\beta = \sup_{y \geq 1} (y-1)^2/y^\beta$. Show that, and calculate c_β if the need arises.

From (22.2)(i) we obtain that

$$\text{(22.10)}\quad \|\mathbf{P}(\theta_t N \in \cdot\,) - \mathbf{P}(N' \in \cdot\,)\| \leq 2 \cdot \mathbf{P}(Y'_0 > t) = 2 \cdot \bar{G}_s(t) = 2\lambda \cdot \int_t^\infty \bar{F}(u)\, du$$

and rate results follow directly from what we know about the tail of F.

23. The IFR case. Repeating the arguments leading to (22.1) but now with r increasing, we obtain a coupling such that

$$\text{(23.1)}\quad \begin{aligned} &\text{(i)}\ S'_0 \leq S_0 \leq S'_1 \leq S_1 \leq \cdots \text{ on } \{\eta > 0\}, \text{ and} \\ &\text{(ii)}\ 0 = S_0 \leq S'_0 \leq S_1 \leq S'_1 \leq \cdots \text{ on } \{\eta = 0\}. \end{aligned}$$

When trying to exploit the alternating structure, one finds that the yield is not at all as rich as in the DFR case. However, there are a number of interesting observations to make. We restrict our attention to the case when S is zero-delayed and S' stationary; it is then (23.1)(ii) that is relevant. We obtain

$$N_t - N'_t = 0 \quad \text{or} \quad 1 \qquad \text{for} \quad t \geq 0\,;$$

hence

$$\text{(23.2)}\qquad 0 \leq U(t) - \lambda t \leq 1 \qquad \text{for} \quad t > 0\,,$$

a well-known relation. But (23.2) is coarse for small t: Notice that $U(0) = 1$. For a simple improvement, notice that

$$N'_t - (N_t - 1) \le I(Y'_0 \le t) .$$

Hence

(23.3) $$\bar{G}_s(t) \le U(t) - \lambda t \qquad \text{for} \quad t \ge 0 .$$

As a coupling time, we now have

(23.4) $$T = \min\{S_n;\, S_n = S'_n\} .$$

Due to the IFR property, there is an $\alpha > 0$ such that

$$\sup_{a \ge 0} \int e^{\alpha x}\, dF_a(x) < \infty .$$

Using this and the ideas of the proof of $\mathbf{E}[\rho^T] < \infty$ in (II.4.6), we find that the T of (23.4) has a finite exponential moment, and hence $\theta_t N$ tends to stationary exponentially fast as $t \to \infty$.

24. Notes. The ideas and results in this section are due to Brown [32] and Lindvall [111]. The paper [32] contains estimates of $U(t) - \lambda t$ not mentioned here. For use of embedding in Poisson processes, see El Karoui and Lepeltier [59] and Çinlar [41].

6. ON A CLASS OF POINT PROCESSES

25. Basics. We now proceed to consider a class of point processes where the past (the history) at each time point $t \ge 0$ is allowed to influence the future in a more general way than is the case for renewal processes. The analysis will be in terms of the class of intensity functions governing the first point of occurrences in the future when the history is known.

Our main concern will be to generalize the results of § 22. For that a new DFR concept is needed; it shall be based on the principle "if there have been many points of occurrence recently, we will soon experience another one." This is the topic of the next paragraph.

For a point process to be ergodic, the influences of its remote past must be limited. If for some constants $A > 0$ and $m \in \mathbb{N}$, only the points in the interval $[t - A, t]$, and the m points preceding t are allowed to influence the behavior after time t, we say that N is an (A, m) process. For them we have ergodicity under natural conditions; this is discussed briefly in § 27.

We use as much as possible of the notation from §§ III.1 and III.2. However, the points of occurrence of the process N to be constructed will be denoted by $T_1, T_2, \ldots$, in order not to violate our convention that S_n means the sum of independent variables. Let $(\mathcal{N}_-, \mathcal{B}_-)$ be the analog of $(\mathcal{N}_+, \mathcal{B}_+)$ for the interval $(-\infty, 0]$. For a $\nu \in \mathcal{N}_-$, let $z = (z_i)_1^K$ denote the increasing sequence of nonnegative numbers such that $\nu = \Sigma_1^K \delta_{-z_i}$; here $0 \leq K \leq \infty$, and we associate a z sequence of length 0 ($K = 0$) with the measure $\nu = \mathbf{0}$. At each occasion we represent an element $\in \mathcal{N}_-$ the way we find most convenient: by ν, by z, or by $(z_1, z_2, \ldots)$. Actually, only ν measures without multiple points will appear, making the corresponding z sequences strictly increasing.

Think of $\ldots, -z_2, -z_1$ as the points of occurrence of some phenomena you have observed during the time interval $(-\infty, 0]$; the intensity for the first point in the future $[= (0, \infty)]$ when the history at time 0 is $\nu = \Sigma_1^K \delta_{-z_i}$ will be denoted by

$$r(z, x) \quad \text{or} \quad r(\nu, x), \qquad x > 0$$

where $r \in (\mathcal{B}_- \times \mathcal{R}_+)/\mathcal{R}_+$. Hence for a fixed ν, $r(\nu, \cdot)$ is a failure rate function. We call r the intensity regime function or shorter: the regime. All that follows will be governed by it.

The meanings of $F(\nu, x)$ and $R(\nu, x)$ for $\nu \in \mathcal{N}_-$, $x \geq 0$, are the obvious ones. When the initial history is random, we denote it by ζ and its distribution by H. If $Z_1, Z_2, \ldots$ is the sequence representation of ζ, the history at time T_1 is

$$(0, T_1 + Z_1, T_2 + Z_2, \ldots).$$

With the outcome, ν_1 say, of that history at hand, T_2 can now be obtained: We let $T_2 - T_1$ have distribution $F(\nu_1, \cdot)$. The method for constructing the entire sequence $T_1, T_2, T_3, \ldots$ recursively should now be obvious. The point process in $\mathcal{N}_+$ to be investigated is $N = \Sigma_1^\infty \delta_{T_i}$. We use $\mathbf{P}_\nu$ and $\mathbf{P}_H$ to denote the underlying probability measure, for the cases ζ fixed $\equiv \nu$ and $\zeta \overset{\mathcal{D}}{=} H$, respectively.

Throughout, the time homogeneity condition

$$(25.1) \quad r(\theta_t \nu, \cdot) = r(\nu, t + \cdot) \qquad \text{for all} \quad \nu \in \mathcal{N}_-, \quad t \geq 0$$

is in force. Notice that the shift θ_t moves the ν mass t steps to the left on $\mathbb{R}$ for $t \geq 0$.

The conditions on the regime r will be given for $\nu \in \mathcal{N}_-^0$, where

$$\mathcal{N}_-^0 = \{\nu \in \mathcal{N}_-; \nu(\{0\}) = 1\}.$$

That is natural: At every point of occurrence T of N, its future distribution is determined by the history

$$(25.2) \qquad (0, T_k - T_{k-1}, \ldots, T_k - T_1, T_k + Z_1, \ldots)$$

at time T_k. The conditions are:

(25.3) There exist constants a_1, a_2 satisfying $0 < a_1 < a_2 < \infty$ such that

(i) $\sup_{\nu \in \mathcal{N}_-^0} R(\nu, a_2) < \infty$,
(ii) $\inf_{\nu \in \mathcal{N}_-^0, x \in [a_1, a_2]} r(\nu, x) > 0$, and
(iii) $\sup_{\nu \in \mathcal{N}_-^0} \mathbf{E}_\nu[T_1 | T_1 > a_2] < \infty$.

Notice that (i) implies the nonexplosiveness of N. That condition is assumed from now on, while (ii) and (iii) are needed for the ergodicity analysis only.

26. On the DFR concept. The standard definition of decreasing failure rates concerns one distribution, F. That suffices for renewal processes, but since that is a very special class of point processes, there are good reasons to ask what should be meant by decreasing failure rates more generally. For the class of point processes considered here, we shall suggest an answer in terms of closed partial orderings on $\mathcal{N}_-$, where $\nu \preceq \nu'$ stands for

(26.1) there have been more points of occurrence recently in the history ν' than in ν, considered from the time point 0.

We suppose that $\preceq$ satisfies the following:

(26.2) (i) $\mathbf{0} \preceq \nu$ for all $\nu \in \mathcal{N}_-$, and
(ii) if $\nu \preceq \nu'$, then $\theta_t \nu \preceq \theta_t \nu'$, $\theta_t \nu \preceq \theta_t \nu' + \delta_0$, and $\theta_t \nu + \delta_0 \preceq \theta_t \nu' + \delta_0$ for all $t > 0$.

These conditions are satisfied for the following two choices of $\preceq$:

(26.3) (i) $\nu \preceq \nu'$ if and only if $\nu \leq \nu'$, and
(ii) $\nu \preceq \nu$ if and only if $\nu[-y, 0] \leq \nu'[-y, 0]$ for all $y \geq 0$.

We say that a point process with regime r has decreasing failure rates (is of DFR type) with respect to the closed partial ordering $\preceq$ on $\mathcal{N}_-$ if (26.2) is satisfied and

$$\nu \preceq \nu' \text{ implies } r(\nu, \cdot) \leq r(\nu', \cdot)\,. \tag{26.4}$$

Recall (21.1)–(21.3). To copy the ideas there for a precise construction of N, let ζ be the history available at time 0 and define recursively for $n \geq 1$:

$$B_n = \{(x, y) \in \mathbb{R}^2_+ ; x > T_{n-1}, y \leq r(Y_{T_{k-1}}, x - T_{n-1})\} \tag{26.5}$$

and

$$T_n = \tau(B_n)\,.$$

Here $T_0 \equiv 0$ (used for this definition only) and $Y_{T_{n-1}}$ is the history (25.2) at time T_{n-1}. Hence $Y_0 = \zeta$.

In general, we shall let $Y = (Y_t)_0^\infty$ denote the process in $\mathcal{N}_-$ of histories:

$$Y_t = (t - T_{N_t}, t - T_{N_t - 1}, \ldots, t - T_1, t + Z_1, \ldots) \tag{26.6}$$

for $t \geq 0$; that explains the notation $Y_{T_{n-1}}$ above.

We have now shown how to construct an underlying probability space $(\Omega, \mathcal{F}, \mathbf{P}_H)$ for N when $\zeta \stackrel{\mathcal{D}}{=} H$, as a product of $(\mathcal{N}_-, \mathcal{B}_-, H)$ and another probability space carrying a Poisson process ξ in $(\mathbb{R}_+^2, \mathcal{R}_+^2)$. But that ξ may also be used to produce another process N' with initial history $\zeta' \stackrel{\mathcal{D}}{=} H'$ (cf. the beginning of § 22). Extend $(\mathcal{N}_-, \mathcal{B}_-, H)$ to $(\mathcal{N}_-^2, \mathcal{B}_-^2, \tilde{H})$, where $\tilde{H}$ has marginals H and H', and we have obtained a probability space on which both N and N' are defined.

Now if $\zeta \preceq \zeta'$ and r is of DFR type, then $N \leq N'$. This follows after introducing the sets $B_1', B_2', \ldots$ analogous to $B_1, B_2, \ldots$ of (26.5), and recalling (26.2)(ii). Indeed, with $\hat{T}_1 = \min(T_1, T_1')$ we have $\hat{T}_1 = T_1' \leq T_1$ due to (26.4). Hence the history of N' at time T_1' equals

$$(0, T_1' + Z_1', T_1' + Z_2', \ldots),$$

which dominates, w.r.t. $\preceq$, the history of N at that time, due to (26.2)(ii). Repeating this argument, we find that all points of occurrence of N are such also of N'. Hence $N \leq N'$.

There remains only little of the proof of the following. We use H, H' to denote the distributions of the initial histories of N, N', respectively.

(26.7) Theorem. *If the regime r is of DFR type and $H \stackrel{\mathcal{D}}{\preceq} H'$, then N and N' may be constructed such that*

(i) *$N \leq N'$, which implies that*
(ii) *if ψ is an increasing (decreasing) functional on $\mathcal{N}_+$, then*

$$\psi(\theta_t N) \stackrel{\mathcal{D}}{\leq} \psi(\theta_t N') \; (\psi(\theta_t N) \stackrel{\mathcal{D}}{\geq} \psi(\theta_t N')) \qquad \text{for} \quad t \geq 0 .$$

Proof. Use Strassen's theorem to obtain a probability measure $\tilde{H}$ on $(\mathcal{N}_-^2, \mathcal{B}_-^2)$ with marginals H and H' satisfying $\tilde{H}(\{(\nu, \nu'); \nu \preceq \nu'\}) = 1$, and from that initial histories $\zeta \stackrel{\mathcal{D}}{=} H$, $\zeta' \stackrel{\mathcal{D}}{=} H'$, and $\zeta \preceq \zeta'$. Then $N \leq N'$ from the discussion above. For (ii), use the proof of Theorem (22.3). □

Denoting the expectation measures of N and N' by M_H and $M_{H'}$, respectively, it is immediate from (i) that

(26.8)
$$M_H \leq M_{H'} .$$

Corollary (22.5) of Theorem (22.3) has the following analog.

(26.9) Corollary. *If $H = \delta_0$ and ψ is increasing (decreasing), then $\psi(\theta_t N)$ is stochastically increasing (decreasing) in t.*

Proof. Use (26.2)(i) and the proof of (22.5). □

Hence if N has $\mathbf{0}$ as initial history, we may argue as for (22.6) to obtain

(26.10) (i) $N(t + B)$ is stochastically increasing in t for all $B \in \mathcal{R}_+$,
(ii) M_0 is convex and $\leq M_{H'}$ for all H',
(iii) the delay D_t is stochastically decreasing in t, and
(iv) the "age" A_t is stochastically decreasing in t.

Here $D_t = \min\{T_n - t; T_n - t > 0\}$ and $A_t = \min\{t - T_n; t - t_n \geq 0\}$ for $t \geq T_1$ and $= \infty$ for $t < T_1$.

27. The (A, m) processes. Let $\nu \in \mathcal{N}_-$ have sequence representation $z = (z_i)_1^K$, define $n_1 = n_1(\nu)$ by

$$n_1 = \max\{i; z_i < A\}$$

and $n_2 = n_2(\nu)$ by

$$n_2 = \min(K, \max(n_1, m)) .$$

Further, put

$$\nu^0 = \sum_1^{n_2} \delta_{-z_i}$$

and

$$z^0 = (z_1, \ldots, z_{n_2}) .$$

We call N an (A, m) process if

$$r(\nu, \cdot) = r(\nu^0, \cdot) \qquad \text{for all} \quad \nu \in \mathcal{N}_- . \tag{27.1}$$

For an (A, m) process, ν^0 is called the memory of the history ν. Of course, in a $(0, m)$ process we recognize what is known as an m-dependent renewal process.

To prove ergodicity for an (A, m) process, one possibility is a coupling where we wait for two processes to have the same memory. However, another possibility has the advantage that we can lean on established theory; namely, to use the fact that $X = (X_t)_0^\infty$, where

$$X_t = Y_t^0 ,$$

is a positive recurrent Harris process under conditions (25.3)(ii)–(iii). We give a reference in the Notes to an account of Harris recurrence of continuous-time Markov processes. From that we learn that there exists a unique stationary distribution π for X, and that X is ergodic.

Now let $\theta_t^+ \mu$ be the restriction of $\theta_t \mu$ to $\mathcal{R}_+$ for a measure μ on $\mathcal{R}_+$. For any distribution G on the memories of N, we have

$$\begin{aligned} &|\mathbf{P}_G(\theta_t^+ N \in A) - \mathbf{P}_\pi(N \in A)| \\ &\quad = \left| \int \mathbf{P}_\nu(N \in A) \cdot \mathbf{P}_G(X_t \in d\nu) - \int \mathbf{P}_\nu(N \in A) \pi(d\nu) \right| \\ &\quad \le \|\mathbf{P}_G(X_t \in \cdot\,) - \pi\| \to 0 \end{aligned} \tag{27.2}$$

as $t \to \infty$, for every $A \in \mathcal{B}_+$. Hence

$$\|\mathbf{P}_G(\theta_t^+ N \in \cdot\,) - \mathbf{P}_\pi(N \in \cdot\,)\| \to 0 \quad \text{as} \quad t \to \infty . \tag{27.3}$$

Now N is stationary if $X_0 \overset{\mathcal{D}}{=} \pi$, and $M_\pi = \alpha \cdot l_+$ where $\alpha = E_\pi[N(0, 1]] = M_\pi(0, 1]$. Under condition (25.3)(i), it holds that

$$\sup_{\nu \in \mathcal{N}_-^0} \mathbf{E}_\nu[N(0, B)] < \infty$$

for all $B > 0$. This fact and an argument similar to (27.2) render

$$\|(M_G)_{[t,\, t+B]} - \lambda \cdot l_{[0,\, B]}\| \to 0 \quad \text{as} \quad t \to \infty \tag{27.4}$$

for all memory distributions G, and $B > 0$.

28. Notes. This section is based on Lindvall [112]. For Harris recurrence in continuous time, see Asmussen [10]. Athreya, Tweedie, and Vere-Jones [16] provide the asymptotics results of § 27 for two-dependent renewal processes.

CHAPTER VI

Diffusions

1. ONE-DIMENSIONAL PROCESSES

1. Basics. In Chapter V we learned that the coupling method is well fitted to be used in the study of birth and death processes, due to the fact that two paths of such processes cannot pass without hitting each other. Paths of one-dimensional diffusion processes share the property of being skip-free since they are continuous, hence making a related analysis of diffusions possible.

Not much of the relevant theory is given here: You are supposed to be either familiar with it or willing to learn (there are excellent sources, two of them mentioned in § 6). We restrict ourselves to settle notation and recall some concepts and properties.

Let I be an interval $\subset \mathbb{R}$. A diffusion process, or diffusion, $X = (X_t)_0^\infty$ on I is a Markov process with continuous paths having the strong Markov property. The notation of § III.20 will be used; in particular, $\mathbf{P}_x$ indicates that the diffusion starts at $x \in I$; $X_0 \equiv x$. It is standard to use the space of continuous functions with values in I, $C_I[0, \infty)$, as sample space Ω, and endow it with the σ-field $\mathscr{C}_I[0, \infty) =$ the Borel σ-field when $C_I[0, \infty)$ is given a "uniform convergence on compacts" metric. We abbreviate to $(C_I, \mathscr{C}_I)$, and drop the I when $I = \mathbb{R}$. This choice of probability space works not only when X is a coordinate process $[X_t(\omega) = \omega(t)$ for $\omega \in C_I]$ or such a process transformed by a scale change, but also when X is an Itô diffusion, solving the stochastic differential equation (SDE)

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t\,, \tag{1.1}$$

where $B = (B_t)_0^\infty$ is a Brownian motion, with drift and dispersion functions b and σ, respectively. In the latter case, we use $(C, \mathscr{C}, \mathbf{P})$ as underlying probability space, where the driving Brownian motion B is the coordinate process under $\mathbf{P}$.

For $y \in I$, let

$$\tau_y = \inf\{t; X_t = y\}\,.$$

and for a set A, let

$$\tau_A = \inf\{t; X_t \in A\}\,.$$

A diffusion is called regular if

$$\text{(1.2)} \qquad \mathbf{P}_x(\tau_y < \infty) > 0 \qquad \text{for all} \quad x \in \operatorname{int}(I), \quad y \in I\,.$$

We consider only processes satisfying this condition, which reminds you of irreducibility for discrete-state-space Markov processes.

A strictly increasing and continuous function s defined on I such that

$$\text{(1.3)} \qquad \mathbf{P}_x(\tau_b < \tau_a) = (s(x) - s(a))/(s(b) - s(a))$$

for $a < x < b \in I$ is called a scale function for X. To every regular diffusion X there is such a function.

A diffusion is recurrent if

$$\text{(1.4)} \qquad \mathbf{P}_x(\tau_y < \infty) = 1 \qquad \text{for all} \quad x, y \in I\,.$$

Recurrence is virtually a property of the scale function; consider, for example, a diffusion with state space $(0, \infty)$. It is recurrent if and only if $s(0+) = -\infty$ and $s(\infty) = \infty$.

A diffusion is in natural scale if

$$\mathbf{P}_x(\tau_b < \tau_a) = (x - a)/(b - a)$$

whenever $a < x < b$ and $[a, b] \subset I$. Any diffusion may be given a natural scale through a space transformation; we assume that our diffusion is in such a scale. Actually, $s(X_t)$, $t \geq 0$, is in a natural scale, where s is the scale function; that is seen immediately if I is a finite and closed interval and is rather easily proved in general. Notice that if X is a martingale, (1.3) is satisfied. But the solution of (1.1) is a local martingale if and only if b vanishes, so think of a diffusion in natural scale as one without drift.

There are virtually six cases to consider: $I=\mathbb{R}$, $[0,\infty)$, $(0,\infty)$, $[0,1]$, $[0,1)$, or $(0,1)$. It is rather easy to prove the following results, with use of the strong Markov property and the natural scale:

(1.5) (i) if $I=\mathbb{R}$, then $\limsup_{t\to\infty} X_t=\infty$ and $\liminf_{t\to\infty} X_t=-\infty$,
(ii) if $I=[0,\infty)$ and 0 not absorbing, then $\limsup_{t\to\infty} X_t=\infty$,
(iii) if $I=(0,\infty)$, or if $I=(0,1]$ and 1 is not absorbing, then $X_t\to 0$ a.s. as $t\to\infty$ for any initial distribution of X, and
(iv) if $I=(0,1)$, then $\mathbf{P}_x(X_t\to 1 \text{ as } t\to\infty)=x$ and $\mathbf{P}_x(X_t\to 0 \text{ as } t\to\infty)=1-x$.

From (iii) and (iv) we learn about the possibly surprising path behavior at a finite inaccessible endpoint of I.

To prove ergodicity for X recurrent turns out to be rather easy; that is a part of our first topic.

2. Ergodicity. I closed. Recall the discussion around (III.20.14)–(III.20.15). We assume now that $X=(X_t)_0^\infty$ and $X'=(X'_t)_0^\infty$ are independent, with initial distributions λ and μ, respectively, and governing probability $\mathbf{P}_{\lambda\mu}$. Ergodicity follows if

$$T=\inf\{t;\, X_t=X'_t\} \tag{2.1}$$

is finite a.s. ($\mathbf{P}_{xy}$) for all $x, y\in I$.

If $I=[0,\infty)$, we have $\mathbf{P}_x(\tau_0<\infty)=1$ for all $x\in I$. As for birth and death processes, we have again due to the skip-freeness that

$$T\le\max(\tau_0,\tau'_0) \tag{2.2}$$

[cf. (V.2.3)] and $T<\infty$ follows. We conclude that all diffusions in natural scale with state space $I=[0,\infty)$ are ergodic.

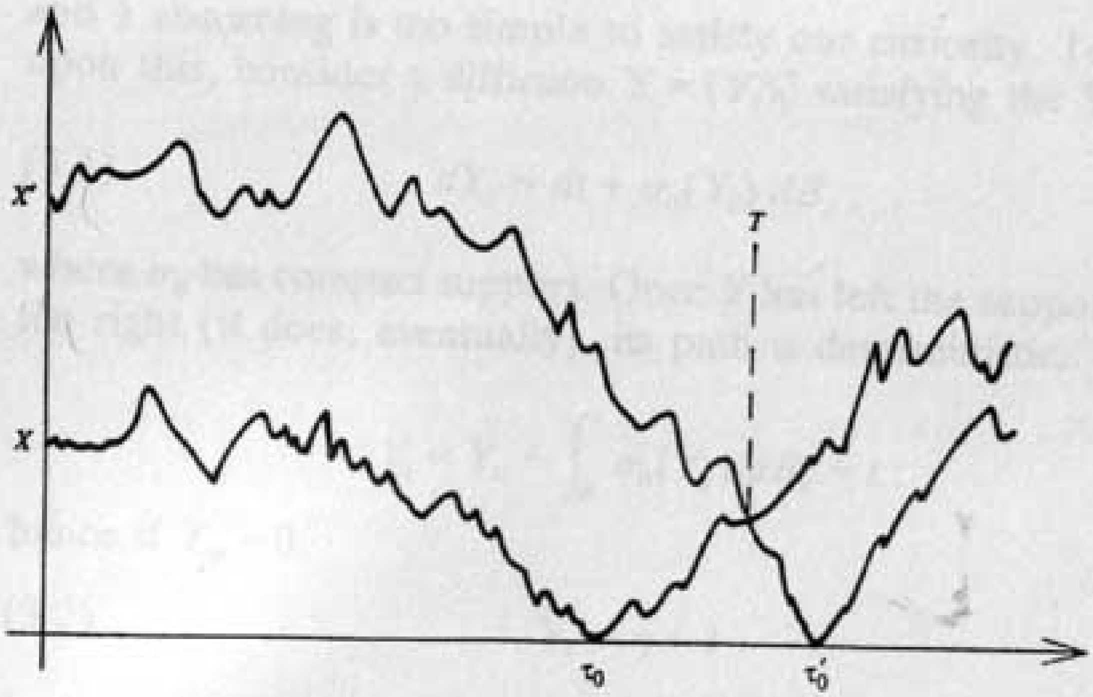

Figure 5. For $I = [0, \infty)$, coupling is easy due to (2.2).

Arguments close to these lead to the same result for the case $I = [0, 1]$ if not both 0 and 1 are absorbing. But if they are, then X is not ergodic. Indeed, due to the natural scale we have

$$\|\mathbf{P}_x(X_t \in \cdot) - \nu_x\| \to 0 \quad \text{as} \quad t \to \infty$$

where $\nu_x = x \cdot \delta_1 + (1 - x) \cdot \delta_0$. Since

$$\begin{aligned} 0 &< \|\nu_x - \nu_y\| \\ &\le \|\mathbf{P}_x(X_t \in \cdot) - \nu_x\| + \|\mathbf{P}_y(X_t' \in \cdot) - \nu_y\| \\ &\quad + \|\mathbf{P}_x(X_t \in \cdot) - \mathbf{P}_y(X_t' \in \cdot)\| \end{aligned}$$

for $x \neq y$, we conclude $\liminf_{t\to\infty} \|\mathbf{P}_x(X_t \in \cdot) - \mathbf{P}_y(X_t' \in \cdot)\| > 0$. This implies that T cannot be finite a.s. A quick inspection convinces you that it is not.

For the case $I = \mathbb{R}$, fix $x < y$ and pick $A_1, A_2, \ldots$ inductively such that $|x|, |y| < A_1$, $0 < A_1 < A_2 < \cdots$ and

$$\text{(2.3)} \qquad |\mathbf{P}_z(\kappa_n = A_n) - \tfrac{1}{2}| \le \tfrac{1}{4}$$

for $z = A_{n-1}$ and $-A_{n-1}$ (when $n = 1$: for $z = x$ and y), where

$\kappa_n = \tau_{(-A_n, A_n)}$ for $n \geq 1$. Such a sequence A_n, $n = 1, 2, \ldots$, exists due to the natural scale. With primes on the corresponding κ_n times for the X' process, we obtain for $X_0 = x$, $X'_0 = y$ from (2.3) that

$$\mathbf{P}_{xy}(X_{\kappa_i} \leq X'_{\kappa'_i} \text{ for all } i \leq n) \leq (1 - (\tfrac{1}{4})^2)^n,$$

implying that

$$\mathbf{P}_{xy}(T < \infty) = \mathbf{P}_{xy}(X_s = X'_s \text{ for some } s \geq 0) = 1$$

since $X_{\kappa_i} > X'_{\kappa'_i}$ for an $i \geq 1$ yields an $s \leq \min(\kappa_i, \kappa'_i)$ such that $X_s = X'_s$. Hence all diffusions on $\mathbb{R}$ in a natural scale are ergodic. That exhausts the recurrent cases, because diffusions in a natural scale with the remaining types of state space ((0, 1), (0, 1] and (0, ∞)) are all transient, which follows from (1.5).

Only little attention will be paid to rate results since that topic is rather technical. However, if $I = [0, 1]$ and at least one of the boundary points is reflecting, to prove

$$\text{(2.4)} \quad \sup_{\lambda, \mu} \|\mathbf{P}_\lambda(X_t \in \cdot\,) - \mathbf{P}_\mu(X'_t \in \cdot\,)\| = \sup_{\lambda, \mu} \|\lambda P_t - \mu P_t\| = \mathrm{o}(\rho^{-t})$$

as $t \to \infty$ for a $\rho > 1$ is simple and should not be omitted. For (2.4), let 1 be reflecting. We shall find a γ such that

$$\text{(2.5)} \qquad \sup_{\lambda, \mu} \mathbf{P}_{\lambda\mu}(T \geq n) \leq (1 - \gamma)^n$$

and (2.4) follows. Now $\mathbf{P}_{\lambda\mu}(T \geq n)$ is maximized if $\lambda = \delta_1$ and $\mu = \delta_0$ ($X_0 \equiv 1$, $X'_0 \equiv 0$); you are supposed to find that very plausible, and are asked to prove it with a coupling argument. But under $\mathbf{P}_{10}$, $T \leq \tau_0$ and

$$\begin{aligned} \mathbf{P}_1(\tau_0 \geq n) &= \mathbf{P}_1(\tau_0 \geq n \| \tau_0 \geq n-1) \cdot \mathbf{P}_1(\tau_0 \geq n-1) \\ &\leq (1 - \gamma) \cdot \mathbf{P}_1(\tau_0 \geq n-1) \end{aligned}$$

for $n \geq 1$ where $\gamma = \mathbf{P}_1(\tau_0 < 1)$, and (2.5) is set.

3. Ergodicity. *I* not closed. So far we have seen no example of an interesting nonergodic diffusion: The case $I = [0, 1]$ with both 0 and 1 absorbing is too simple to satisfy our curiosity. To improve upon this, consider a diffusion $Y = (Y_t)_0^\infty$ satisfying the SDE

$$dY_t = dt + \sigma_0(Y_t)\, dB_t\,, \tag{3.1}$$

where σ_0 has compact support. Once Y has left the support of σ_0 to the right (it does, eventually), its path is deterministic:

$$Y_t = Y_0 + \int_0^t \sigma_0(Y_s)\, dB_s + t\,;$$

hence if $Y_0 = 0$,

$$Y_t = \eta + t \tag{3.2}$$

for t large enough, where $\eta = \int_0^\infty \sigma_0(Y_s)\, dB_s$. Notice that η is tail measurable w.r.t. $Y = (Y_t)_0^\infty$; That tail variable is not constant, hence Y is not ergodic.

The Y process is not regular. But if σ_0 does not have compact support but $\sigma_0(y)$ tends to 0 sufficiently rapidly as $|y|$ becomes large, the fluctuation of Y slows down as time passes so that $Y_t - t$ is a.s. convergent to η as $t \to \infty$. And that new Y process is regular.

Let I_0 denote the state space of Y, and s the scale function so that $X_t = s(Y_t)$, $t \ge 0$, is in the natural space $I = s(I_0)$. The interval I must be open at a finite endpoint (why?) to which X_t converges a.s. as $t \to \infty$, and we have understood that there are interesting nonergodic diffusions in natural scale when I is not closed.

We lose no interesting features by concentrating our attention on $I = (0, 1]$. Forget about the special meaning given to $X = (X_t)_0^\infty$ above, and let it now be any diffusion in natural scale on I, with 1 not absorbing. We know that $X_t \to 0$ a.s. The important questions are:

(3.3) (i) How can nonergodicity of X be described? When is its fluctuation slow enough for that property?

(ii) How can the tail σ-field $\mathcal{T}(X)$ be described when nontrivial (= the nonergodic case)?

We continue to be sketchy, answering (i) and (ii) in terms of statements without proofs; see § 6 for thorough accounts.

To answer (i), let X' be independent of and with the same transition probabilities as X. The first answer to (i) is striking:

(3.4) X is nonergodic if and only if the coupling of X and X' is unsuccessful,

that is, if and only if $\mathbf{P}_{xy}(T<\infty)<1$ for some $x, y \in I$. The "only if" implication is known to us for a long time. That ergodicity implies the success of Doeblin coupling is a deep result.

For the special diffusion $X_t = s(Y_t)$, $t \geq 0$, in a natural scale, where Y is given by (3.1), we use (3.2) to find that

$$c(X_t) = \eta + t \tag{3.5}$$

for t large enough, where $c = s^{-1}$. How is that suitably generalized to any diffusion X in a natural scale? For $z \in I = (0, 1]$, let

$$c(z) = \mathbf{E}_1[\tau_z]\,.$$

Notice that c is strictly decreasing on I. Actually, $M = (M_t)_0^\infty$, where

$$M_t = t - c(X_t) \tag{3.6}$$

defines a continuous local martingale under each $\mathbf{P}_x$. The second question in (3.3)(i) will now be answered in terms of M; we may view M as X suitably normalized.

Consider the statements

(3.7) (i) $\mathbf{E}_x[\langle M\rangle_\infty]<\infty$ for all $x \in I$,

(ii) $(M_t)_0^\infty$ is L^2-bounded under all $\mathbf{P}_x$, $x \in I$, and

(iii) $(M_t)_0^\infty$ is convergent a.s. as $t \to \infty$ under all $\mathbf{P}_x$, $x \in I$.

Here $\langle M\rangle_t$, $t \geq 0$, is the quadratic variation process of M. From martingale theory we know that (i)$\Leftrightarrow$(ii) and (ii)$\Rightarrow$(iii). But if (iii) holds, then X is nonergodic, because $M_\infty = \lim_{t\to\infty} M_t$ is not constant, it is $\mathcal{T}(M)$-measurable, and $\mathcal{T}(M) \subset \mathcal{T}(X)$.

Nonergodicity implies (ii). Indeed, suppose that $\langle M\rangle_\infty = \infty$.

Take X' independent of X, and let X, X' start at x, y, respectively. Due to the independence, we have

$$\langle M - M' \rangle_t = \langle M \rangle_t + \langle M' \rangle_t ,$$

with the obvious meaning of M'. But on $\{\langle M - M' \rangle_\infty = \infty\}$ we have that

$$\limsup_{t\to\infty} (M_t - M'_t) = \infty \quad \text{and} \quad \liminf_{t\to\infty} (M_t - M'_t) = -\infty .$$

Hence due to continuity

$$c(X_t) - c(X'_t) = 0 \qquad \text{for some} \quad t \geq 0$$

and since c is strictly monotone, $X_t = X'_t$ for that t. The successful coupling is achieved.

We have now "proved" that nonergodicity of X is equivalent to each of (3.7)(i)–(iii), with use of (3.4). These properties of X are also equivalent to

$$\lim_{y\to 0} \mathrm{Var}_x[\tau_y] < \infty$$

for all $x \in I$, where Var_x denotes variance under $\mathbf{P}_x$.

We now consider the question (3.3)(ii). Since c is strictly monotone, $\mathcal{T}(M) = \mathcal{T}(X)$. It is a sound conjecture that

(3.8) $$\sigma(M_\infty) = \mathcal{T}(M)$$

(remember: $M_\infty = \lim_{t\to\infty} M_t$). Indeed, (3.8) holds, hence

(3.9) $$\mathcal{T}(X) = \sigma(M_\infty) .$$

But (3.8) is far from immediate.

4. The strong Feller property. Assumptions are retained: $X = (X_t)_0^\infty$ is a regular diffusion in a natural scale in an interval I. We shall demonstrate how to use Doeblin coupling to understand that X has the strong Feller property, that is, that for any $t > 0$ the function $P_t f$ defined by

$$P_t f(x) = \mathbf{E}_x[f(X_t)]$$

(a recall) is continuous in x for all $f \in b\mathcal{R}$. Fix a $t>0$. With X' independent of X, and X, X' starting at x, y, respectively, we have

$$(4.1)\qquad \begin{aligned}|P_t f(x) - P_t f(y)| &= |\mathbf{E}_{xy}[f(X_t)] - \mathbf{E}_{xy}[f(X'_t)]| \\ &\le 2 \cdot \mathbf{P}_{xy}(T>t) \cdot \|f\| ;\end{aligned}$$

hence we shall show that $\mathbf{P}_{xy}\ (T>t) \to 0$ as $y \to x$. Although intuitively rather obvious, this is not immediate.

To begin with, suppose that I is open [$= \mathbb{R}$, $(0, \infty)$ or $(0, 1)$]. Fix $x \in I$, $\epsilon > 0$, and choose $\delta_1 > 0$ so small that

$$\mathbf{P}_x(\langle X \rangle_t > \delta_1) > 1 - \epsilon .$$

Remember: It is the same $t>0$ appearing throughout. Next choose $\delta_2 > 0$ so small that $(x - \delta_2, x + \delta_2) \subset I$ and

$$\mathbf{P}(\sup_{s \le \delta_1} B_s \le \delta_2 \text{ or } \sup_{s \le \delta_1} - B_s \le \delta_2) \le \epsilon$$

for a Brownian motion $B = (B_s)_0^\infty$. For $y \in (x - \delta_2,\ x + \delta_2)$ we obtain

$$(4.2)\qquad \begin{aligned}\mathbf{P}_{xy}(T>t) &\le \mathbf{P}_{xy}(T>t, \langle X \rangle_t > \delta_1) + \epsilon \\ &= \mathbf{P}_{xy}(B(\langle X - X' \rangle_s) \\ &\qquad + (x-y) \ne 0 \text{ for } s \le t, \langle X \rangle_t > \delta_1) + \epsilon \\ &\le \mathbf{P}_{xy}(B_s + (x-y) \ne 0 \text{ for } s \le t) + \epsilon \le 2\epsilon .\end{aligned}$$

We made use of the existence of a Brownian motion B such that $X_t - X'_t = B(\langle X - X' \rangle_t) + (x-y)$ for $t \ge 0$; there is one since $X - X'$ is a continuous local martingale. Further, we also used the fact that $\langle X - X' \rangle_s = \langle X \rangle_s + \langle X' \rangle_s$ since X and X' are independent. This implies that $\langle X - X' \rangle_s \ge \langle X \rangle_s$.

Combining (4.1) and (4.2) now gives

$$|P_t f(x) - P_t f(y)| \le 2\epsilon \cdot \|f\| ;$$

hence $P_t f$ is continuous.

For I not open, some further remarks are needed, because X is then not necessarily a local martingale. We may restrict our attention to $I=[0,\infty)$. Recall that $T\le\max(\tau_0,\tau_0')$, implying that

$$\mathbf{P}_{xy}(T>t)\le\mathbf{P}_x(\tau_0>t)+\mathbf{P}_y(\tau_0'>t)\,. \tag{4.3}$$

Hence to prove that P_tf is continuous at 0, it suffices to know that

$$\mathbf{P}_y(\tau_0'>t)\to 0 \qquad \text{as}\quad y\to 0\,. \tag{4.4}$$

which is not that difficult to prove.

Now let $x>0$. To avoid paying attention to the behavior of X at state 0, simply kill the process there! That interference does not affect (4.3) because τ_0 and τ_0' are unchanged and (4.3) holds for the new $T=$ the coupling time of the killed X and X', and that T is not smaller than the original T.

So we may assume that the state 0 is absorbing. But in that case X is a continuous local martingale, and the proof above for I open works.

5. Domination. For a semigroup $(P_t)_0^\infty$ governing a diffusion we have

$$\lambda P_t\overset{\mathcal{D}}{\le}\mu P_t \qquad \text{for all}\quad t\ge 0 \qquad \text{if}\quad \lambda\overset{\mathcal{D}}{\le}\mu\,. \tag{5.1}$$

That property follows from the existence of a coupling (X,X') such that $X_0\overset{\mathcal{D}}{=}\lambda$, $X_0'\overset{\mathcal{D}}{=}\mu$, and

$$X_t\le X_t' \qquad \text{for all}\quad t\ge 0\,. \tag{5.2}$$

But as for birth and death processes [cf. (V.4)], Doeblin coupling provides (5.2), with X_0, X_0' constructed as usual to satisfy $X_0\le X_0'$.

No account of consequences of (5.2) similar to that around (V.4.2)–(V.4.6) will be given; find some possibilities! Rather, we shall try to find a diffusion analog of the result that (V.4.18) implies (V.4.20) [i.e., that (5.2) holds when $X_0\le X_0'$]; such results often appear in the literature called "comparison theorems."

Consider the SDEs

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t \quad \text{and} \tag{5.3}$$
$$dX'_t = b'(X'_t)\,dt + \sigma(X'_t)\,dB'_t$$

driven by Brownian motions $B = (B_t)_0^\infty$, $B' = (B'_t)_0^\infty$, and with the same dispersion function σ. We assume that σ and the drift functions b, b' satisfy the standard smoothness and growth conditions

$$\text{(i)}\ |b(x) - b(y)|,\quad |b'(x) - b'(y)|,\quad |\sigma(x) - \sigma(y)| \le C \cdot |x - y|\,, \tag{5.4}$$

which imply that

$$\text{(ii)}\ |b(x)|,\ |b'(x)|,\ |\sigma(x)| \le C \cdot (1 + |x|)$$

for all $x, y \in \mathbb{R}$. These assumptions are stronger than necessary but make you comfortable with the stochastic calculus that follows.

If $b \le b'$ and $x \le y$, it is a sound conjecture that the solutions X, X' to (5.3) with $X_0 = x$, $X'_0 = y$ satisfy

$$X_t \le X'_t \qquad \text{for all} \quad t \ge 0 \tag{5.5}$$

if $B = B'$. To prove this, let $(\varphi_n)_1^\infty$ be a sequence of increasing C^2 functions satisfying

(5.6) (i) $(x - 1/n)^+ \le \varphi_n(x) \le x^+$,
(ii) φ'_n is increasing and $0 \le \varphi'_n \le 1$, and
(iii) $0 \le \varphi''_n(x) \cdot x^2 \le 2/n$.

Now fix the initial states x, y satisfying $x \le y$. Itô's formula applied to $Y_t = X_t - X'_t$ and φ_n gives

$$\begin{aligned} \varphi_n(Y_t) = &\int_0^t \varphi'_n(Y_s)[b(X_s) - b'(X'_s)]\,ds \\ &+ \int_0^t \varphi'_n(Y_s)[\sigma(X_s) - \sigma(X'_s)]\,dB_s \\ &+ \tfrac{1}{2} \cdot \int_0^t \varphi''_n(Y_s) \cdot [\sigma(X_s) - \sigma(X'_s)]^2\,ds\,. \end{aligned} \tag{5.7}$$

[Notice that $\varphi(x)=0$ for $x \le 0$.] The second integral has expectation 0; the finiteness of that follows from (5.6)(ii) and SDE theory. For the third integral of (5.7), we use (5.4)(i) and (5.6)(iii) to get

$$\text{(5.8)} \qquad \left|\mathbf{E}\left[\int_0^t \varphi_n''(Y_s)[\sigma(X_s)-\sigma(X_s')]^2\,ds\right]\right|$$
$$\le C\cdot\mathbf{E}\left[\int_0^t \varphi_n''(Y_s)\cdot Y_s^2\,ds\right] \le C\cdot t/n\,.$$

Due to 5.4(i), 5.6(ii), and the fact that $b \le b'$ we have

$$\text{(5.9)} \qquad \varphi_n'(Y_s)\cdot[b(X_s)-b'(X_s')] \le C\cdot Y_s^+\,.$$

We combine (5.7)–(5.9) to get

$$\mathbf{E}[\varphi_n(Y_t)] \le C\cdot\int_0^t \mathbf{E}[Y_s^+]\,ds + C\cdot t/n\,.$$

Now let $n\to\infty$ to find that

$$\mathbf{E}[Y_t^+] \le C\cdot\int_0^t E[Y_s^+]\,ds\,.$$

But any nonnegative locally integrable function f defined on $[0,\infty)$ satisfying $f(t) \le C\cdot\int_0^t f(s)\,ds$ must vanish. Hence $\mathbf{E}[Y_t^+]=0$ for all $t\ge 0$, and since $Y=(Y_t)_0^\infty$ has continuous paths, this implies that $X_t \le X_t'$ for all $t\ge 0$ a.s.

6. Notes. Two broad accounts of stochastic calculus and SDE theory are Rogers and Williams [136] and Karatzas and Shreve [85]. Ergodicity was studied by Fristedt and Orey [63], Rösler [139], Küchler and Lunze [94], and Lindvall [110]. For a beautiful summary to which we owe much, see Rogers [135]. You find the proof of (3.9) in that paper. The result (3.4) is from [94].

Almost all of § 4, on the strong Feller property, is due to Chris Rogers (private communication). Use the arguments of [136, Prop. V.50.1] to prove (4.4). The pioneering work on domination is due to Skorohod [143] and Yamada [161]; for detailed accounts, see [136] or [85]. See also O'Brien [126], Kulperger and Guttorp [95], Day [48], and Zhiyuan [162].

2. MULTIDIMENSIONAL PROCESSES

7. Basics. In the study of ergodicity of one-dimensional diffusions (§§ 2 and 3), only Doeblin coupling came to use; it turned out to be simple and efficient, and due to the sharp result that ergodicity implies the success of that coupling, we got no incitement to explore any other. But once we enter $\mathbb{R}^d$, $d \geq 2$, as state space, we lose the "no passing without meeting" property, and Doeblin coupling becomes useless. Indeed, let X and X' be independent Brownian motions in $\mathbb{R}^d$ starting at x and y respectively, where $x \neq y$. Then $X_t \neq X'_t$ for all $t \geq 0$ a.s. Hence we are challenged to produce efficient coupling even for the most basic multidimensional diffusion, and we surmise a wide field of interesting problems generally, for such diffusions. In this section a few topics from that field are selected to pique your interest. First we show how to couple Brownian motions by reflection and give some simple applications of that. The next topic is a certain coupling for diffusions with radials drift and constant dispersion.

8. Brownian motion. Equip $(C_{\mathbb{R}^d}, \mathscr{C}_{\mathbb{R}^d})$ with a probability measure $\mathbf{P}$ such that the coordinate process becomes a Brownian motion starting at 0 in $\mathbb{R}^d$; we denote that process by $B = (B_t)_0^\infty$. For an $x \in \mathbb{R}^d$, $X = (X_t)_0^\infty$ defined by $X_t = x + B_t$ is a Brownian motion starting at x; in this section, let X have that meaning.

A Brownian motion X' starting at y may be constructed in terms of X as follows. For $y \neq x$, let

$$L_{xy} = \{u; (u - (x+y)/2, x - y) = 0\} \tag{8.1}$$

be the hyperplane right between x and y such that $x - y$ is orthogonal to L_{xy}. No precision is needed to define L_{xx}; let it be any hyperplane containing x.

With $T_{xy}z$ denoting the mirror image of z with respect to L_{xy}, we let

$$X'_t = \begin{cases} T_{xy}X_t, & t < \kappa_{xy} \\ X_t, & t \geq \kappa_{xy}, \end{cases} \tag{8.2}$$

where

$$\kappa_{xy} = \inf\{s \geq 0; X_s \in L_{xy}\},$$

which is finite a.s. Certainly, $X' = (X'_t)_0^\infty$ is a Brownian motion starting at y.

We have $X_t \overset{\mathscr{D}}{=} N(x, tI)$ and $X'_t \overset{\mathscr{D}}{=} N(y, tI)$ for $t > 0$. Trite analytical calculations render

$$\text{(8.3)}\quad \begin{aligned} \|N(x, tI) - N(y, tI)\| &= 2 \cdot \Phi(-|x-y|/2t^{1/2}, |x-y|/2t^{1/2}) \\ &\le 2 \cdot |x-y|/(2\pi t)^{1/2}, \end{aligned}$$

where Φ is the standard normal distribution function on $(\mathbb{R}, \mathscr{R})$; you may replace the inequality by "$\sim$" as $t \to \infty$.

Now X and X' have the coupling time $T = \kappa_{xy}$. The basic inequality gives

$$\text{(8.4)}\qquad \|N(x, tI) - N(y, tI)\| \le 2 \cdot \mathbf{P}(\kappa_{xy} > t).$$

But κ_{xy} is distributed as the time it takes for a one dimensional Brownian motion to hit the level $|x-y|/2$. That time is well known to have the distribution function $2 - 2 \cdot \Phi(|x-y|/2t^{1/2})$, $t > 0$. This means, due to (8.3), that equality holds in (8.4), so our reflection coupling is actually maximal.

As a first illustration of how our coupling can be used, let ξ_t, $t \ge 0$, be the configuration of n gas molecules at time t, moving according to n independent Brownian motions; as state space for ξ_t we may use $((\mathbb{R}^d)^n, (\mathscr{R}^d)^n)$. To understand how quickly an initial configuration $\{x_1, \ldots, x_n\}$ is forgotten, we estimate $\|\mathbf{P}\xi_t^{-1} - \mathbf{P}\xi_t'^{-1}\|$, where ξ'_t, $t \ge 0$, is the analog of ξ_t, $t \ge 0$, with initial configuration $\{y_1, \ldots, y_n\}$. For the estimate, let X_i, X'_i be Brownian motions with $X_i(0) = x_i$, $X'_i(0) = y_i$ and coupled as above. We find that

$$\text{(8.5)}\quad \begin{aligned} \|\mathbf{P}\xi_t^{-1} - \mathbf{P}\xi_t'^{-1}\| &\le 2 \cdot \mathbf{P}(\xi_t \ne \xi'_t) \\ &\le 2 \cdot \mathbf{P}(\kappa_{x_i y_i} > t \text{ for some } i,\ 1 \le i \le n) \\ &\le 2 \cdot \sum_1^n \mathbf{P}(\kappa_{x_i y_i} > t), \end{aligned}$$

which is bounded by $2 \cdot (\Sigma_1^n |x_i - y_i|)/(2\pi t)^{1/2}$ due to (8.3) and that equality holds in (8.4). The last inequality in (8.5) is coarse.

Our second illustration concerns the use of hitting times of a Brownian motion and classical potential theory. The reader not familiar with the few concepts we use should consult the references in § 11.

To allow initial values of X and X' that are random, let $\mathbf{P}_{xy}$ be the governing measure on $((C_{\mathbb{R}^d})^2, (\mathscr{C}_{\mathbb{R}^d})^2)$ when $X_0 = x$ and $X_0' = y$, and define $\mathbf{P}_{\lambda\mu}$ by

$$\mathbf{P}_{\lambda\mu} = \int \mathbf{P}_{xy}(\cdot)[\lambda \times \mu](dx, dy) . \tag{8.6}$$

If we let (X_t, X_t'), $t \geq 0$ be governed by $\mathbf{P}_{\lambda\mu}$, then X and X' are Brownian motions with $X_0 \overset{\mathscr{D}}{=} \lambda$ and $X_0' \overset{\mathscr{D}}{=} \mu$. Of course, $\lambda \times \mu$ may be replaced by any probability measures on $((\mathbb{R}^d)^2, (\mathscr{R}^d)^2)$ with the right marginals. Definition (8.6) works to the effect that X_0 and X_0' are dropped independently, a mirror is set up between these points, and X, X' are constructed by a reflection as above. In our applications, only X_0' will be randomized; the meaning of $\kappa_{x\mu}$ is the obvious one.

Now, for a $\mathscr{F}_\sigma$ set $B \subset \mathbb{R}^d$, let

$$\tau_B = \inf\{x > 0; X_s \in B\} , \tag{8.7}$$

and let $h_B(x, \cdot)$ be the hitting distribution defined by

$$h_B(x, A) = \mathbf{P}_x(\tau_B < \infty, X_{\tau_B} \in A) . \tag{8.8}$$

Of course, $h_B(x, \cdot)$ is supported by $\overline{B}$.

To estimate the distance between $h_B(x, \cdot)$ and $h_B(y, \cdot)$, let τ_B' be the analog of (8.7) for the X' process and notice that

$$\tau_B = \tau_B' \quad \text{on} \quad \{\kappa_{xy} \leq \tau_B \wedge \tau_B'\} . \tag{8.9}$$

Using the fact that

$$\begin{aligned} h_B(x, A) = {} & \mathbf{P}_{xy}(\tau_B < \infty, \tau_B' < \infty, X_{\tau_B} \in A) \\ & + \mathbf{P}_{xy}(\tau_B < \infty, \tau_B' = \infty, X_{\tau_B} \in A) \end{aligned}$$

and the analogous split of $h_B(y, A) = \mathbf{P}_y(\tau_B' < \infty, X'_{\tau_B'} \in A)$, we obtain

$$(8.10)\quad \|h_B(x,\cdot)-h_B(y,\cdot)\| \le 2\cdot \mathbf{P}_{xy}(\kappa_{xy}>\tau_B\wedge\tau'_B) + \mathbf{P}_{xy}(\{\tau_B=\infty\}\Delta\{\tau'_B=\infty\})$$

for all $x, y\in\mathbb{R}^d$, due to (8.9).

We give two examples of how (8.10) can be applied. The first concerns the Dirichlet problem: For an open set D and a bounded and continuous function φ defined on ∂D, we search for a function f such that $f(x)=\varphi(x)$ for $x\in\partial D$ and $\Delta f(x)=0$ for $x\in D^\circ$, where $\Delta f=\Sigma_1^d\partial_i^2 f$ is the Laplacian of f. If the boundary of D is reasonably regular and D^c is recurrent, the unique solution is

$$f(x)=\mathbf{E}_x[\varphi(x_{\tau_{D^c}})]=\int\varphi(y)h_{D^c}(x,dy)\,.$$

For $x, y\in D$, (8.10) gives

$$(8.11)\qquad |f(x)-f(y)|\le 2\cdot\|\varphi\|\cdot\mathbf{P}_{xy}(\kappa_{xy}>\tau_{D^c}\wedge\tau'_{D^c})\,.$$

This is the simplest way to describe a method that has come to use in estimates of gradients of the solution to potential theory equations.

For the second application of (8.10), let μ_B be the equilibrium measure of a bounded nonpolar $\mathcal{F}_\sigma$ set B. We give a short proof of the result that

$$(8.12)\qquad \|h_B(x,\cdot)-\mu_B\|\to 0 \quad\text{as}\quad |x|\to\infty$$

when $d=2$. To that end, note that

$$\mu_B=\int h_B(z,\cdot)\sigma_r(dz)$$

if $B\subset B_r$, where σ_r is the uniform distribution on the surface of the ball B_r with radius r, centered at 0. Hence, if $|x|$ is large enough,

$$\|h_B(x,\cdot)-\mu_B\|=\|h_B(x,\cdot)-\int h_B(z,\cdot)\sigma_{|x|}(dz)\| \le 2\cdot\mathbf{P}_{x,\sigma_{|x|}}(\kappa_{x,\sigma_{|x|}}>\tau_B\wedge\tau'_B)\,.$$

Fix an r_0 such that $B\subset B_{r_0}$. For $|x|\ge r_0$ we have

$$\mathbf{P}_{x,\sigma_{|x|}}(\kappa_{x,\sigma_{|x|}} > \tau_B \wedge \tau'_B) \le \mathbf{P}_{x,\sigma_{|x|}}(\kappa_{x,\sigma_{|x|}} > \tau_{B_{r_0}}) .$$

But the latter probability tends to 0 as $|x| \to \infty$, which is seen after a scale change such that the two starting points lie on the unit circle, and use of $\mathbf{P}_z(\tau_{B_\alpha} \to \infty \text{ as } \alpha \to 0) = 1$ for all z with $|z| = 1$.

Now let $d \ge 3$. For the last example from potential theory of a property that can be proved easily by analytical means but is immediate with use of our reflection coupling, consider the equilibrium potential $e_B(x)$, $x \in \mathbb{R}^d$, of a $\mathcal{F}_\sigma$ set B. The probabilistic formulation of e_B, the unique solution to an equation we do not state, is the following:

$$e_B(x) = \mathbf{P}_x(\tau_B < \infty) .$$

Physical reasons make it plausible that e_B decreases as x remotes from B. To illuminate that, fix x, $y \in \mathbb{R}^d$. If B is a subset of the half-space bounded by L_{xy} and containing x, then

$$e_B(y) \le e_B(x)$$

as follows directly from $\tau_B \le \tau'_B$.

It is easy to produce a reflection coupling for Brownian motions on the unit sphere to prove that

$$\begin{aligned} &\|\mathbf{P}_\lambda(X_t \in \cdot\,) - \mathbf{P}_\mu(X'_t \in \cdot\,)\| \\ &\quad \le \|\mathbf{P}_{\delta_n}(X_t \in \cdot\,) - \mathbf{P}_{\delta_s}(X'_t \in \cdot\,)\| \to 0 \quad \text{as} \quad t \to \infty , \end{aligned}$$

for any initial distributions λ, μ; δ_n and δ_s denote unit masses at the north and south poles $(0, \ldots, 1)$ and $(0, \ldots, -1)$, respectively, and one to prove ergodicity of a Brownian motion on a torus.

A last remark: Any nontrivial diffusion on the unit circle is ergodic. Find the quick coupling proof!

9. Radial drift. Now consider the diffusion $X = (X_t)_0^\infty$ *in* $\mathbb{R}^d$, which solves the SDE

$$dX_t = dB_t - X_t \cdot \beta(|X_t|)\, dt , \tag{9.1}$$

where $\beta: \mathbb{R}_+ \to \mathbb{R}$ is assumed to be Lipschitz and bounded to make SDE theory work smoothly, and $B = (B_t)_0^\infty$ is a standard Brownian motion in $\mathbb{R}^d$. Notice that we do not assume β to be nonnegative, so a particle moving according to X may be pushed away from $\mathbf{0}$, along the ray from that point.

There are several ways to couple processes X and X' governed by (9.1). We shall employ one that is reasonably efficient and easy to explain. To begin with, let X and X' be independent and run them until

$$(9.2) \qquad \tau = \inf\{s; |X_s| = |X'_s|\}\,.$$

After that, let X'_t be the mirror image of X_t, with reflection in the subspace orthogonal to $X_\tau - X'_\tau$ until X hits that mirror surface; subsequently, the paths of X and X' coincide.

No one doubts that the distribution of X' so constructed is the right one. The number of details, however, in a thorough proof is considerable: trite work if you have gone through similar things before, useful if not.

We have two problems: to determine when

(9.3) (i) $\tau < \infty$, and
(ii) X hits the mirror in finite time

a.s., regardless of initial values for X and X'. Whether these stopping times are finite or not depends on properties of the diffusion $|X| = (|X_t|)_0^\infty$; that is obvious concerning (i), but also for (ii) as we shall see.

We know much about the finiteness of τ since $|X|$ is a diffusion; recall §§ 2 and 3! That stopping time is finite if and only if $|X|$ is ergodic, and this is the case if either $|X|$ is recurrent, or transient but not of the type given by (3.7), after the introduction of a natural scale. But these matters may be formulated explicitly in terms of the scale function, s say, of $|X|$. We have

$$(9.4) \qquad s(v) = \int_1^v x^{-(d-1)} \exp\Big[2\int_1^x u\beta(u)\, du\Big]\, dx$$

for $v > 0$, due to the fact that $|X|$ satisfies

(9.5) $$d|X_t| = d\tilde{B}_t + \{\tfrac{1}{2}\cdot(d-1)|X_t|^{-1} - |X_t|\beta(|X_t|)\}\,dt$$

for a certain Brownian motion $\tilde{B}$, as follows from Itô's formula applied to (9.1). Now $s(0+) = -\infty$ since β is bounded, so the comment after (1.4) tells us that $|X|$ is recurrent if and only if $s(\infty) = \infty$. For the transient case $s(\infty) < \infty$, define s_0 by $s_0(v) = s(v) - s(\infty)$. The property of being a scale function is not affected by addition of a constant, hence s_0 is again a scale function for $|X|$. To be ergodic, $|X|$ should not tend to be ∞ in a too determined way (cf. § 3). A (necessary and sufficient) condition may be given in terms of s_0 for that. Hence $\tau < \infty$ a.s. if $s(\infty) = \infty$, or if $s(\infty) < \infty$ but $|X|$ is still ergodic.

To take the second step of (9.3) we use a skew-product representation of X. Indeed, if $Z = (Z_t)_0^\infty$ is a Brownian motion on the unit sphere S^{d-1} in $\mathbb{R}^d$, and $R = (R_t)_0^\infty$ is a diffusions on $(0, \infty)$ independent of Z and satisfying (9.5) (replace $|X_t|$ by R_t there!), then

(9.6) $$X_t = R_t \cdot Z\Big(\int_0^t R_s^{-2}\,ds\Big), \qquad t \ge 0,$$

satisfies (9.1). Hence (9.3)(ii) holds if and only if the clock

$$C_t = \int_0^t |X_s|^{-2}\,ds, \qquad t \ge 0,$$

of the representations (9.6) diverges a.s. If $|X|$ is recurrent, this obviously happens. In the transient case there is again a necessary and sufficient condition, in terms of s, for that divergence. This completes the "solutions" to the problems (9.3)(i)–(ii). We found that both (i) and (ii) hold if $|X|$ is recurrent.

So far in this chapter, we have said nothing about stationary distributions of diffusions. In the one-dimensional case, the condition for the existence of a stationary distribution has a compact formulation. The speed measure m of a regular diffusion is, up to a multiplicative constant, the unique invariant measure; hence there is a stationary distribution π if and only if m is finite, and $\pi = m/m(I)$, of course, where I is the state space.

For our radial drift process X, note that once $|X|$ has a stationary distribution, π_0 say, then so has X. Indeed, use (9.6) with $R_0 \stackrel{\mathcal{D}}{=} \pi_0$ and Z_0 uniformly distributed on S^{d-1}. Then X is stationary, and we may conclude our account of coupling of radial drift diffusions with a result in the main vein of this book: Regardless of the initial state, the distribution of X_t tends to equilibrium as $t \to \infty$.

10. Another reflection coupling. We now turn to a generalization of the reflection idea used for coupling of Brownian motions in § 8. Consider the diffusions X_t and X'_t, $t \geq 0$, which solve the SDEs

$$\begin{aligned} dX_t &= \sigma(X_t)\, dB_t + b(X_t)\, dt\,, \\ dX'_t &= \sigma(X'_t)\, dB'_t + b(X'_t)\, dt\,. \end{aligned} \tag{10.1}$$

We display two equations to emphasize that the driving Brownian motions are not the same. Rather, we shall let the increments dB'_t of B' be a function (a suitable reflection) of those of B, with the aim of obtaining a coupling of X and X'. To that end, fix the Brownian motion B and notice that for B' we may use a distortion of B satisfying

$$dB'_t = H_t\, dB_t\,, \tag{10.2}$$

where H_t, $t \geq 0$, is some previsible process with values in the set of orthonormal $d \times d$ matrices. It is natural to let $H_t = H(X_t, X'_t)$ for a suitable mapping H: $(\mathbb{R}^d)^2 \to (\mathbb{R}^d)^2$.

Notice that if $|u| = 1$ for $u \in \mathbb{R}^d$, the $d \times d$ matrix

$$I - 2 \cdot uu^* \tag{10.3}$$

as linear operator is the reflection in the subspace orthogonal to u.

For $x, x' \in \mathbb{R}^d$, let $z = x - x'$. The mapping H is defined by $H(x, x') = I$ for $x = x'$, and by (10.3) with

$$u = \sigma^{-1}(x')z / |\sigma^{-1}(x')z| \tag{10.4}$$

for $x \neq x'$ (we assume that σ^{-1} exists). This coupling is easy to understand when σ is constant. Check also to find that for Brownian motions ($\sigma = I$, $b = 0$), it equals the one in § 8. A substantial amount of stochastic calculus is needed on the way to sufficient conditions for successful coupling.

11. Notes. This part is based on Lindvall and Rogers [114]. For further details about the radial drift case, see § 4 of that paper, which also contains an investigation of how far the reflection idea reaches. For couplings in terms of the generator of a diffusion, related to those we used for birth and death processes, see Chen and Li [38]. Davies [46] employs ideas related to those we have used for Harris chains to couple multidimensional diffusions, and establishes rate of convergence results. Kendall [89, 90] investigates the existence of successful couplings of Brownian motions on certain Riemannian manifolds and its consequences concerning harmonic functions.

For the use of Brownian motions in potential theory, see Port and Stone [132]; the notation and other details used here are adopted from that book. Check also Durrett [56]. For the use of Brownian motion coupling for gradient estimates of the solution to the Schrödinger and Poisson equations, for examples, see Cranston [42, 43] and Cranston and Zhao [44]. There are virtually couplings at work in the branching diffusions of Lalley and Sellke [99, 100].

Appendix

1. POLISH SPACES

1. A quick survey. Recall § I.6. A metric d on a set E is complete if every Cauchy sequence is convergent, and E is separable with that metric if there is a countable dense set in E. The Borel σ-field $\mathscr{E}$ is the smallest one containing the open sets in (E, d); due to the separability, $\mathscr{E}$ is generated already by a countable number of open balls in E. Notice that the continuous functions are measurable on $(E, \mathscr{E})$.

For topics (i) to (vi) to follow, let $E_0, E_1, \ldots$ be equipped with metrics $d_0, d_1, \ldots$ making all (E_i, d_i) Polish.

(i) Product spaces. With the metric

$$\rho_n(x, y) = \sum_0^n d_i(x_i, y_i)$$

for $x = (x_i)_1^n$ and $y = (y_i)_1^n$ the product space $\Pi_1^n E_i$ becomes Polish. This is true also for $\Pi_1^\infty E_i$ with the metric

$$\rho(x, y) = \sum_0^\infty 2^{-i} \cdot d_i(x_i, y_i)/(1 + d_i(x_i, y_i))$$

with x and y extended to infinite sequences. The standard product σ-fields $\Pi_0^n \mathscr{E}_i$ and $\Pi_0^\infty \mathscr{E}_i$ on $\Pi_0^n E_i$ and $\Pi_0^\infty E_i$ agree with the Borel σ-fields generated by ρ_n and ρ. For the construction of probability measures on $(\Pi_0^\infty E_i, \Pi_0^\infty \mathscr{E}_i)$, the state space condition of Kolmogorov's consistency theorem is satisfied.

(ii) Tightness and weak convergence. A family of probability measures $\{P_\alpha;\ \alpha \in \Lambda\}$ is called tight if for every $\epsilon > 0$ there is a compact set $K_\epsilon = K$ such that

$$\inf_{\alpha \in \Lambda} P_\alpha(K) > 1 - \epsilon\,.$$

Every single measure P is tight; that is, there is a K such that $P(K) > 1 - \epsilon$ for every $\epsilon > 0$.

A sequence $(P_n)_1^\infty$ is weakly convergent to P if $\int f\,dP_n \to \int f\,dP$ as $n \to \infty$ for all $f \in cb\mathscr{E}$; this is written $P_n \Rightarrow P$. Prohorov's theorem says that if $\{P_n;\ n \geq 1\}$ is tight, there is a subsequence of $(P_n)_1^\infty$ that is weakly convergent. For random elements X_n in E, we write $X_n \xrightarrow{\mathscr{D}} X$ as $n \to \infty$ if the distribution of X_n converges weakly to that of X.

(iii) Skorohod's representation. Suppose that $(P_n)_1^\infty$ and P are probability measures on $(E, \mathscr{E})$ such that $P_n \Rightarrow P$. Then there exists a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and random elements $(X_n)_1^\infty$, X in $(E, \mathscr{E})$ defined on $(\Omega, \mathscr{F}, \mathbf{P})$ such that $X_n \stackrel{\mathscr{D}}{=} P_n$, $X \stackrel{\mathscr{D}}{=} P$ and

$$X_n \to X \quad \text{a.s.} \qquad \text{as} \quad n \to \infty\,.$$

In fact, we may always use $([0,1], \mathscr{R}_{[0,1]}, l_{[0,1]})$ for $(\Omega, \mathscr{F}, \mathbf{P})$.

(iv) Function spaces. For the space (E, d), let $C_E[0,\infty) = C_E$ be the set of continuous function x: $[0,\infty) \to E$. With

$$\rho_k(x, y) = \sup_{0 \leq t \leq k} d(x(t), y(t)) \tag{1.1}$$

and

$$\rho(x, y) = \sum_k^\infty 2^{-k} \cdot \rho_k(x, y)/(1 + \rho_k(x, y))$$

for x and $y \in C_E$, ρ is a Polish metric on C_E. The Borel σ-field $\mathscr{C}_E[0,\infty) = \mathscr{C}_E$ equals that generated by the finite-dimensional sets in C_E, that is,

$$\mathscr{C}_E = \sigma\{\pi^{-1}_{t_1 \ldots t_k} B;\ B \in \mathscr{E}^k,\ t_i \geq 0,\ k = 1, \ldots\}\,,$$

where $\pi_{t_1 \ldots t_k}(x) = (x(t_1), \ldots, x(t_k))$ for $x \in C_E$.

Let $D_E[0,\infty) = D_E$ be the space of functions x: $[0,\infty) \to E$ that are right-continuous and have a left-hand limit at each $t \geq 0$. In the Skorohod topology on D_E, $x_n \to x$ as $n \to \infty$ if and only if there exists a sequence $(\lambda_n)_0^\infty$ of strictly increasing continuous functions $[0,\infty) \to [0,\infty)$ such that $\lambda_n(0) = 0$,

$$\sup_{0 \leq t < \infty} |\lambda_n(t) - t| \to 0 \qquad \text{as} \quad n \to \infty$$

and

$$\rho_k(x_n \circ \lambda_n, x) \to 0 \qquad \text{as} \quad n \to \infty \quad \text{for all} \quad k \geq 1\,,$$

where ρ_k is from (1.1). This topology renders D_E Polish. Again, the Borel σ-field $\mathcal{D}_E$ can be shown to equal that generated by the finite-dimensional sets.

(v) Measure spaces. Let $\mathcal{M}(E, \mathcal{E}) = \mathcal{M}$ be the class of measures that gives finite mass to compact sets, and let $\mathcal{N}(E, \mathcal{E}) = \mathcal{N}$ be those $\nu \in \mathcal{M}$ that are integer valued. We say that $\nu_n \to \nu$ vaguely in $\mathcal{M}$ if $\int f\, d\nu_n \to \int f\, d\nu$ for all $f \in c\mathcal{E}$ with compact support. That convergence criterion gives $\mathcal{M}$ and $\mathcal{N}$ a topology that is Polish. Actually, $\mathcal{N}$ is closed in $\mathcal{M}$.

This is the trickiest example of a Polish space; it is not trivial even to prove that this topology is metrizable. To get a feeling for the separability, let $\{x_i;\, i = 1, 2, \ldots\}$ be a dense set in E. Then the class of measures that give positive rational masses to a finite number of points x_i is dense in $\mathcal{M}$.

(vi) Regular conditional probabilities. Let X and Y be random elements in $(E_0, \mathcal{E}_0)$ and $(E_1, \mathcal{E}_1)$, respectively, defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then there exists a transition kernel K in $E_0 \times E_1$ such that $K(X, A)$ is a version of $\mathbf{P}(Y \in A \| X) = \mathbf{E}[I(Y \in A) \| X]$ for all $A \in \mathcal{E}_1$, that is,

$$\mathbf{P}(X \in B, Y \in A) = \int_{\{X \in B\}} K(X, A)\, d\mathbf{P}$$

for all $B \in \mathcal{E}_0$ and $A \in \mathcal{E}_1$. The integral equals

$$\int K(x, A)\, d[\mathbf{P}X^{-1}](x)\,.$$

A definition of $\mathbf{P}(Y \in A \| X = x)$ should be such that

$$\mathbf{P}(X \in B, Y \in A) = \int_B \mathbf{P}(Y \in A \| X = x)\, d[\mathbf{P}X^{-1}](x)$$

for all $B \in \mathscr{E}_0$ and $A \in \mathscr{E}_1$. This holds if we let $\mathbf{P}(Y \in A \| X = x) = K(x, A)$.

2. The Banach space $b\mathscr{M}_s$. It is natural that the little Banach space theory we need appear in a Polish section. Recall § I.2 and the decomposition there of a signed bounded measure ν into $\nu^+ - \nu^-$. We call $|\nu| = \nu^+ + \nu^-$ the total variation measure of ν. The result (I.2.3) says that $\|\nu\| = |\nu|(E)$.

It is easy to see that $\|\cdot\|$ makes $b\mathscr{M}_s(E, \mathscr{E}) = b\mathscr{M}_s$ a normed linear space. In particular,

$$\begin{aligned}\|\nu_1 + \nu_2\| &= \sup_{|f| \le 1} \left| \int f\, d(\nu_1 + \nu_2) \right| \\ &\le \sup_{|f| \le 1} \left| \int f\, d\nu_1 \right| + \sup_{|f| \le 1} \left| \int f\, d\nu_2 \right| = \|\nu_1\| + \|\nu_2\|\end{aligned}$$

for any ν_1 and $\nu_2 \in b\mathscr{M}_s$. To prove the completeness of the metric given by $\|\cdot\|$, let $(\nu_n)_1^\infty$ be a Cauchy sequence. Since $\|\nu_n - \nu_m\| \le 1$ for all n, $m \ge$ some index, we have $\sup_n \|\nu_n\| < \infty$. Hence

$$\mu = \sum_1^\infty 2^{-n} |\nu_n|$$

is finite. Certainly, ν_n is absolutely continuous w.r.t. μ; with $g_n = d\nu_n / d\mu$ we have for all n, $m \ge 1$,

$$\begin{aligned}\|\nu_n - \nu_m\| &= \sup_{|f| \le 1} \left| \int f\, d(\nu_n - \nu_m) \right| \\ &= \sup_{|f| \le 1} \left| \int f \cdot (g_n - g_m)\, d\mu \right| = \int |g_n - g_m|\, d\mu\end{aligned}$$

[let $f = \operatorname{sgn}(g_n - g_m)$]; hence $(g_n)_1^\infty$ is a Cauchy sequence in $L^1(E, \mathscr{E}, \mu)$. But the completeness of that space is well known.

The second result we have used is the well-known inequality

$$\|\nu_1 * \nu_2\| \le \|\nu_1\| \cdot \|\nu_2\| \tag{2.1}$$

for ν_1 and $\nu_2 \in b\mathcal{M}_s(R, \mathcal{R})$. Here $\nu_1 * \nu_2$ is the convolution of ν_1 and ν_2 defined by

$$[\nu_1 * \nu_2](B) = \int \nu_1(B - x)\, d\nu_2(x)$$

for $B \in \mathcal{R}$. Now

$$\begin{aligned}\|\nu_1 * \nu_2\| &= \sup_{|f|\le 1} \left| \int f(y)\, d[\nu_1 * \nu_2](y) \right| \\ &= \sup_{|f|\le 1} \left| \int\int f(y+x)\, d\nu_1(y)\, d\nu_2(x) \right| \\ &\le \int_x \left[\int_y d|\nu_1|(y) \right] d|\nu_2|(x) = \|\nu_1\| \cdot \|\nu_2\| .\end{aligned}$$

3. Notes. The classic about probability measures on Polish spaces and weak convergence is Billingsley [26]. Up-to-date accounts are Ethier and Kurtz [60] and Jacod and Shiryaev [79], with a wealth of material about weak convergence in C_E and D_E. For the vague topology of the random measure and point process state spaces $\mathcal{M}$ and $\mathcal{N}$, respectively, consult Daley and Vere-Jones [45]. See also Dudley [55] for the Skorohod representation and regular probabilities. The proof that $b\mathcal{M}_s$ is a Banach space is from Neveu [121].

2. EPILOGUE

4. Some history. It is generally agreed to consider Doeblin [51] as the inventor of the coupling method. That paper was published in 1938; Doeblin wrote it in the summer of 1936. A fascimile of the crucial passages in [51] completes the Appendix. For a survey of the life and work of Doeblin, see Lindvall [113]. Actually, there is a hint of the method in a note by Doeblin published before [51], but it does not give us reason to leave the accepted history.

Unfortunately, the paper [51] appeared in a journal that existed for only two years before World War II and never became known. The work of Doeblin was a major source of inspiration for Doob [53], but [51] is not in the reference list of that canonical book of postwar probability theory.

Harris [71] in 1955 mentioned Doeblin [51], but in 1991 (private communication) he has only vague memories of how he came across the paper; he mentions Kai Lai Chung, an ardent "student" of Doeblin, as a possible messenger. In that fine paper by Harris coupling is at work in the way we are used to; examples of that are not easy to find from the period 1945–1965. Actually, Harris coins the term "engagement" for coupling, but that did not last long; I have not seen it used since.

Elementary books by Breiman [31] and Hoel, Port, and Stone [75] prove the basic ergodic theorem for discrete state space Markov chains with coupling. That had influence, but it was the upsurge of interest in interacting particle systems starting about 1970 that established the method. The term "coupling" was coined in this period but it has not been possible to establish exactly when, although Frank Spitzer is agreed to be the person who did it.

Crucial for the awareness of the method were the papers by Griffeath [67] and Pitman [129], prepared independently. No doubt, Griffeath learned about couplings from his supervisor Spitzer, while Pitman, who does not use the term in [129], was inspired by Breiman [31].

My attention was caught by [129] in 1975. It is my conviction that we have most of the adventure ahead of us.

We finish with a line from the foremost contemporary Swedish poet, Tomas Tranströmer:

> Den som är framme har en lång väg att gå.

The word "framme" has two meanings, which explains why translations differ:

> He who has gone furthest has a long way to go. (Robin Fulton)

> The one who has arrived has a long way to go. (Robert Bly)

§ 2.—LE CAS RÉGULIER

4. Considérons deux systèmes matériels ne pouvant prendre que les états $E_1 \ldots E_\nu$, évoluant d'une façon réciproquement indépendante et suivant tous les deux la loi de la chaine, c.à d. chacun a une probabilité p_{ij} de passer en une épreuve de l'état E_i à l'état E_j. Nous les faisons partir à un même instant initial d'états E_i et E_k. Il pourra alors être commode de nommer le premier «système témoin» S_i et le second S_k. S_i à la probabilité $P_{ij}^{(n)}$ d'être après n épreuves en E_j, S_k la probabilité $P_{kj}^{(n)}$.

Soit A_{ik} la probabilité pour que le système S_i et le système S_k se rencontrent, c'est la somme de la probabilité $B_{ik}^{(1)}$ pour que S_i et S_k se rencontrent à la première épreuve plus la probabilité $B_{ik}^{(2)}$ qu'ils se rencontrent pour la première fois à la seconde épreuve etc.

$$A_{ik} = \sum_{n=1}^{\infty} B_{ik}^{(n)} = \sum_{n=1}^{\infty} B_{ki}^{(n)} = A_{ki}$$

On peut séparer $B_{ik}^{(n)}$ en r parties: c'est la somme des r probabilités pour que S_i et S_k se rencontrent pour la première fois à la n-ième épreuve et que cette rencontre ait lieu à l'état E_j $(j = 1, \ldots \nu)$

$$B_{ik}^{(n)} = \sum_{j=1}^{\nu} C_{ik}^{(n)}(j) = \sum_{j=1}^{\nu} C_{ki}^{(n)}(j)$$

On a le *théorème : Soit tous les A_{ik} sont $= 1$ $(i, k = 1 \ldots \nu)$, ou il y a un $A_{ik} = O$.* Suposons que tous les A_{ik} sont plus grands que $a > 0$, alors écrivons que la probabilité pour que S_i et S_k se reucontrent est la somme de la probabilité pour que S_i et S_k se rencontrent pendant les n premières épreuves et de la probabilité que S_i et S_k ne se rencontrent pas avant et à la n-ième épreuve, mais se rencontrent après. Or quels que soient les positions de S_i et S_k à la n-ième épreuve, la probabilité pour qu'ils se rencontrent dans le mouvement ultérieur est $> a$. D'autre part nous pouvous prendre n suffisamment grand pour que $\sum_{t=1}^{n} B_{ik}^{(t)} > A_{ik} - \varepsilon$, alors

$$A_{ik} > A_{ik} - \varepsilon + (1 - A_{ik})\,a$$

ε étant arbitrairement petit. Ceci entraine $A_{ik} = 1$. C.q.f.d. Si donc les deux systèmes témoins peuvent se rencontrer à partir de n'importe quelles positions, ils se rencontrent presque sûrement. Nous nommerons ce dernier cas le cas régulier (on verra que cette définition est équivalente à la définition de M. Fréchet).

Plaçons nous dans le *cas de Markoff* où tous les p_{ij} sont $> a > 0$, alors la probabilité pour que S_i et S_k se rencontrent en une épreuve est évidemment $> a$, quels que soient E_i et E_k. Il suit que $A_{ik} = 1$ pour tous i, k et même que la probabilité pour que S_i et S_k ne se rencontrent une seule fois pendant les n premières épreuves est $< (1 - a)^n$.

Facsimile of parts of pp. 78–80 in Doeblin [51], where the coupling method is presented and the first coupling inequality is proved.

Dans le cas plus général où il y a un ϱ tel que tous les $P_{ik}^{(\varrho)} > a > 1$, on voit de même que la probabilité pour que S_i et S_k ne se rencontrent pas pendant les $n \cdot \varrho$ premières épreuves est $< (1 - a)^n$. Dans tous les deux cas cette probabilité est $< k \lambda^n$ où $\lambda < 1$, nous disons quelle tend exponentiellement vers zéro.

5.— Théorème: Dans le cas régulier

$$P_{ij}^{(n)} \to P_j \qquad . \qquad \text{si } n \to \infty$$

quels que soient i *et* j.

Démonstration : Soit $T_{ik}^{(n)}$ la probabilité pour que S_i et S_k ne se sont pas encore rencontrés dans les n premières épreuves. Démontrons d'abord que

$$| P_{ij}^{(n)} - P_{kj}^{(n)} | < T_{ik}^{(n)}$$

Nous avons

$$P_{ij}^{(n)} = \sum_{t=1}^{n-1} \sum_{l=1}^{\nu} C_{ik}^{(t)}(l)\, P_{lj}^{(n-t)} + C_{ik}^{(n)}(j) + \Theta_{ikj}^{(n)}$$

$\Theta_{ikj}^{(n)}$ étant la probabilité pour que S_i se trouve à la n ième épreuve en E_j, sans s'être rencontré avec S_k avant et sans s'y rencontrer avec S_k.

Les $C_{ik}^{(t)}(l)$ étant symétriques en i et k et comme $\Theta_{ikj}^{(n)} \leq T_{ik}^{(n)}$:

$$| P_{ij}^{(n)} - P_{kj}^{(n)} | = | \Theta_{ikj}^{(n)} - \Theta_{kij}^{(n)} | \leq T_{ik}^{(n)}$$

Dans le cas régulier $T_{ik}^{(n)} \to 0$, et par suite

$$P_{ij}^{(n)} - P_{kj}^{(n)} \to 0$$

Or si $P_j^{(n)}$ et $p_j^{(n)}$ sont le maximum et minimum de $P_{kj}^{(n)}$ par rapport à k, il résulte de

$$P_{ij}^{(n)} = \sum_{\varepsilon=1}^{\nu} p_{i\varepsilon}\, P_{\varepsilon j}^{(n-1)} \qquad \text{que}$$

$$P_j^{(n-1)} > P_j^{(n)} > \ldots > p_j^{(n)} > p_j^{(n-1)}$$

et comme $P_j^{(n)} - p_j^{(n)}$ tend vers zéro, il suit

$$P_{ij}^{(n)} \to P_j$$

Et il résulte de ce que nous avons montré plus haut que dans le cas où tous les $P_{ik}^{(\varrho)}$ sont > 0, la différence $P_{ij}^{(n)} - P_j$ tend exponentiellement vers zéro; il est facile de le montrer aussi dans le cas régulier le plus général Des équations (1) et (2) résulte que

$$P_j = \sum_{i=1}^{\nu} P_i\, p_{ik} = \sum_{i=1}^{\nu} P_i\, P_{ik}^{(n)} \quad , \quad \sum_{j=1}^{\nu} P_j = 1 \tag{3}$$

Les P_j sont donc tous $> a$ si tous les $P_{ik}^{(\varrho)}$ sont $> a$.

Frequently Used Notation

I. Sets and certain functions

$\mathbb{N}$	that natural numbers $\{1, 2, \ldots\}$
$\mathbb{Z}, \mathbb{Z}_+$	the integers $\{\ldots, -1, 0, 1, \ldots\}$, the nonnegative integers $\{0, 1, \ldots\}$
$\mathbb{R}, \mathbb{R}_+$	the real numbers, the nonnegative real numbers $= [0, \infty)$
$\bar{\mathbb{Z}}_+$ etc.	$\mathbb{Z}_+$ etc. extended with a $+\infty$ point
E^n	the product space $\{(x_1, \ldots, x_n); x_i \in E\}$
E^∞	the product (sequence) space $\{(x_0, x_1, \ldots); x_i \in E\}$
A^c	the complement of the set A
$A^\circ, \partial A$	the interior, the boundary, of the set A
I_A	the indicator function of A; $I_A(x) = 1$ for $x \in A$, $= 0$ if not
$a \vee b, a \wedge b$	$\max(a, b)$, $\min(a, b)$
a^+, a^-	$a \vee 0, -a \wedge 0$
$[a]$	the largest integer $\leq a$
Δ	the diagonal $\{(x, x); x \in E\}$ in E^2
$\mathscr{F}_\sigma$	the class of countable unions of closed sets

II. Measurable spaces, measures, probabilities, and moments

$(\Omega, \mathscr{F}, \mathbf{P})$	sample space, underlying probability space
$(E, \mathscr{E})$	state space: a Polish space E with its Borel sets

$\mathscr{E}^n$, $\mathscr{E}^\infty$	the product σ-fields on E^n and E^∞
$\mathscr{E}_1/\mathscr{E}_2$	the class of measurable mappings $E_1 \to E_2$ for two measurable spaces, $(E_1, \mathscr{E}_1)$ and $(E_2, \mathscr{E}_2)$
$\mathscr{E}$	the class of measurable real-valued functions on $(E, \mathscr{E})$
$\mathbf{P}(A \| B)$	the conditional probability of A given B
$\mathbf{E}[Y \| X = x]$	the conditional expectation of Y given $X = x$
$\mathbf{P}(A \| \mathscr{F})$	the conditional probability of A given the σ-field $\mathscr{F}$
$\mathbf{E}[Y \| \mathscr{F}]$	the conditional expectation of Y given the σ-field $\mathscr{F}$
$\mathscr{P}(E)$	the σ-field of all subsets of E
$\mathscr{Z}$, $\mathscr{Z}_+$ etc.	$\mathscr{P}(\mathbb{Z})$, $\mathscr{P}(\mathbb{Z}_+)$ etc.
$\mathscr{R}$, $\mathscr{R}_+$, $\mathscr{R}^d$ etc.	the Borel σ-fields on $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}^d$ etc.
l	the Lebesgue measure on $(\mathbb{R}, \mathscr{R})$ or $(\mathbb{R}^d, \mathscr{R}^d)$ or specified subsets thereof
l_+, l_A	l restricted to $\mathbb{R}_+$, A
$m_\alpha(\cdot)$	the moment of order α of a distribution on $\mathbb{Z}_+$ or $\mathbb{R}_+$
μ_α	$m_\alpha(p)$ or $m_\alpha(F)$ for specified distributions p or F on $\mathbb{Z}_+$ or $\mathbb{R}_+$
μ	μ_α for $\alpha = 1$, often denoting an expected recurrence time
$\\|f\\|$	usually, $\sup_{x \in E} \|f(x)\|$, the sup norm of a function f defined on E
$\\|\cdot\\|_p$	the L^p norm
$\\|\nu\\|$	the total variation norm for a signed measure ν

III. Function and measure spaces

$(C_E, \mathscr{C}_E)$	short for $(C_E[0, \infty), \mathscr{C}_E[0, \infty))$; see (App.1)(iii)
$(D_E, \mathscr{D}_E)$	short for $(D_E[0, \infty), \mathscr{D}_E[0, \infty))$; see (App.1)(iii)
$(D, \mathscr{D})$	short for $(D_E, \mathscr{D}_E)$ when $E \subset \mathbb{R}$
$\mathscr{M}(E, \mathscr{E})$, $\mathscr{M}$	the measures on $(E, \mathscr{E})$ with finite mass to compact sets
$\mathscr{N}(E, \mathscr{E})$, $\mathscr{N}$	the measures in $\mathscr{M}$ that are $\mathbb{Z}_+$-valued
$\mathscr{N}_+$, $\mathscr{N}_-$	$\mathscr{N}(E, \mathscr{E})$ for $E = \mathbb{R}_+$, $(-\infty, 0]$
$b\mathscr{M}_s(E, \mathscr{E})$	the bounded signed measures on $(E, \mathscr{E})$

IV. Certain distributions

$\mathrm{Poi}(\lambda)$, p_λ	the Poisson distribution with parameter λ
$\mathrm{Bin}(n, p)$	the binomial distribution with parameters n and p
$\mathrm{Ber}(p)$	the Bernoulli distribution = that of a 0–1 variable with probability p for outcome 1
$\mathrm{Uni}[a, b]$	the uniform distribution on the interval $[a, b]$
$\mathrm{Exp}(\alpha)$	the exponential distribution with parameter α
$N(\mu, \sigma^2)$	the normal distribution with expectation μ and variance σ^2
Φ	the distribution function of $N(0, 1)$

V. Miscellaneous

$X \overset{\mathscr{D}}{=} Y$	the variables X and Y have the same distribution
$X \overset{\mathscr{D}}{=} P$	the variable X has distribution P
$P_n \Rightarrow P$	P_n converges weakly to P
$X_n \xrightarrow{\mathscr{D}} X$	X_n converges in distribution to X
$X_n \xrightarrow{\mathscr{D}} P$	the distribution of X_n converges weakly to P
$X_n \xrightarrow{\mathrm{P}} X$	X_n converges to X in probability
$a_n \sim b_n$	$a_n/b_n \to 1$ as $n \to \infty$
C, c	generic constants > 0
C_1, c_1	specific constants > 0
b, c, i	as prefixes: bounded, continuous, increasing
a.s.	almost surely
i.i.d.	independent and identically distributed
p.g.f.	probability generating function
w.r.t.	with respect to
i.o.	infinitely often

References

ALDOUS, D. J.

[1] Tail behaviour of birth and death and stochastically monotone processes, *Z. Wahrsch. Verw. Gebiete*, **62**, 375–394 (1983).

[2] Random walks on finite groups and rapidly mixing Markov chains, in *Seminaire de Probabilités XVII*, Springer Lecture Notes in Mathematics, Vol. 986, Springer-Verlag, Berlin, 1983.

[3] Some interesting processes arising as heavy traffic limits in an $M/M/\infty$ storage process, *Stochastic Process. Appl.*, **22**, 291–313 (1986).

ALDOUS, D. and DIACONIS, P.

[4] Shuffling cards and stopping times, *Amer. Math. Monthly*, **93**, 333–348 (1986).

[5] Strong uniform times and finite random walks, *Adv. Appl. Math.*, **8**, 69–97 (1987).

ALDOUS, D. and THORISSON, H.

[6] Shift-coupling, *Stochastic Process. Appl.*, (1993).

ANICHKIN, S. A.

[7] The coupling of renewal processes and its applications, in *Stability Problems for Stochastic Models* (ed. V. V. Kalashnikov, translated from the Russian), Plenum, New York, 1986.

ARJAS, E., NUMMELIN, E. and TWEEDIE, R. L.

[8] Uniform limit theorems for non-singular renewal and Markov renewal processes, *J. Appl. Probab.*, **15**, 112–125 (1978).

ASH, R. B.

[9] *Real Analysis and Probability*, Academic Press, New York, 1972.

Asmussen, S.

[10] *Applied Probability and Queues*, Wiley, New York, 1987.

Asmussen, S. and Thorisson, H.

[11] A Markov chain approach to periodic queues, *J. Appl. Probab.*, **24**, 215–225 (1987).

[12] Large deviation results for time-dependent queue length distributions, *Stochastic Models*, **4**, 99–116 (1988).

Athreya, K. B. and Ney, P.

[13] A new approach to the limit theory of recurrent Markov chains, *Trans. Amer. Math. Soc.*, **245**, 493–501 (1978).

Athreya, K. B., McDonald, D. and Ney, P.

[14] Limit theorems for semi-Markov processes and renewal theory for Markov chains, *Ann. Probab.*, **6**, 788–797 (1978).

[15] Coupling and the renewal theorem, *Amer. Math. Monthly*, **85**, 809–814 (1978).

Athreya, K. B., Tweedie, R. L. and Vere-Jones, D.

[16] Asymptotic behaviour of point processes with Markov-dependent intervals, *Math. Nachr.*, **99**, 301–313 (1980).

Ball, F.

[17] The threshold behaviour of epidemic models, *J. Appl. Probab.*, **20**, 227–241 (1983).

[18] Deterministic and stochastic epidemics with several kinds of susceptibles, *Adv. Appl. Probab.*, **17**, 1–22 (1985).

[19] A note on the total size distribution of epidemic models, *J. Appl. Probab.*, **23**, 832–836 (1986).

Ball, F. and Donnelly, P.

[20] Interparticle correlation in death processes with application to variability in compartmental models, *Adv. Appl. Probab.*, **19**, 755–766 (1987).

Barbour, A., Holst, L. and Janson, S.

[21] *Poisson Approximation*, Oxford University Press, Oxford, 1992.

Berbee, H.

[22] *Random Walk with Stationary Increments and Renewal Theory*, Mathematisch Centrum, Amsterdam, 1979.

[23] Periodicity and absolute regularity, *Israel J. Math.*, **55**, 289–304 (1986).

[24] Convergence rates in the strong law for bounded mixing sequences, *Probab. Theory Related Fields*, **74**, 255–270 (1987).

[25] Chains with infinite connections: uniqueness and Markov representation, *Probab. Theory Related Fields*, **76**, 243–253 (1987).

BILLINGSLEY, P.

[26] *Convergence of Probability Measures*, Wiley, New York, 1968.

[27] *Probability and Measures*, 2nd ed., Wiley, New York, 1986.

BRANDT, A. and KÜNSCH, H. R.

[28] On the stability of robust filter-cleaners, *Stochastic Process. Appl.*, **30**, 253–262 (1988).

BRANDT, A., LISEK, B. and NERMAN, O.

[29] On stationary Markov chains and independent random variables, *Stochastic Process. Appl.*, **34**, 19–24 (1990).

BREIMAN, L.

[30] *Probability*, Addison-Wesley, Reading, Mass., 1968.

[31] *Probability and Stochastic Processes*, Houghton Mifflin, Boston, 1969.

BROWN, M.

[32] Bounds, inequalities and monotonicity properties for some specialized renewal processes, *Ann. Probab.*, **8**, 227–240 (1980).

CAMBANIS, S. and SIMONS, G.

[33] Probability and expectation inequalities, *Z. Wahrsch. Verw. Gebiete*, **59**, 1–25 (1982).

CARLSSON, H. and NERMAN, O.

[34] An alternative proof of Lorden's renewal inequality, *Adv. Appl. Probab.*, **18**, 1015–1016 (1986).

CHEN, L.

[35] Poisson approximation for dependent trials, *Ann. Probab.*, **3**, 534–545 (1975).

CHEN, M. F.

[36] Coupling for jump processes, *Acta Math. Sinica*, **2**, 123–136 (1986).

[37] On coupling of jump processes, *Technical report* (1987).

CHEN, M. F. and LI, S. F.

[38] Coupling methods for multidimensional diffusion processes, *Ann. Probab.*, **17**, 151–177 (1989).

CHUNG, K. L.

[39] *A Course in Probability Theory*, 2nd ed., Academic Press, New York, 1974.

[40] *Lectures from Markov Processes to Brownian Motion*, Springer-Verlag, New York, 1982.

ÇINLAR, E.

[41] Markov and semi-Markov models of deterioration, in *Reliability Theory and Models* (ed. M. S. Abdel-Hameed et al.), Academic Press, New York, 1984.

CRANSTON, M.

[42] Gradient estimates on manifolds using coupling, *J. Funct. Anal.*, **99**, (1991).

[43] A probabilistic approach to gradient estimates, *Can. Math. Bull.*, **34** (1991).

CRANSTON, M. and ZHAO, Z.

[44] Some regularity results and eigenfunction estimates for the Schrödinger operator, in *Diffusion Processes and Related Problems in Analysis*, Vol. I (ed. M. Pinsky), Birkhäuser, Boston, 1991.

DALEY, D. J. and VERE-JONES, D.

[45] *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, 1988.

DAVIES, P. L.

[46] Rates of convergence to the stationary distribution for k-dimensional diffusion processes, *J. Appl. Probab.*, **23**, 370–384 (1986).

DAVIES, P. L. and GRÜBEL, R.

[47] Spaces of summable sequences in renewal theory and the theory of Markov chains, *Math. Nachr.*, **104**, 119–128 (1981).

DAY, M.V.

[48] Comparison results for diffusions conditioned on positivity, *J. Appl. Probab.*, **20**, 766–777 (1983).

DEVROYE, L.

[49] Coupled samples in simulation, *Oper. Res.*, **38** (1990).

DIACONIS, P.

[50] *Group Representation in Probability and Statistics*, IMS Lecture Notes – Monograph Series, Vol. 11, IMS, Hayward, 1988.

DOEBLIN, W.

[51] Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états, *Rev. Math. Union Interbalkanique*, **2**, 77–105 (1938).

DONNELLY, P. and WHITT, W.

[52] Stochastic comparisons in reinforcement-depletion compartmental urn models, *J. Appl. Probab.*, **26**, 477–489 (1989).

DOOB, J. L.

[53] *Stochastic Processes*, Wiley, New York, 1953.

DUBINS, L. E. and MEILIJSON, I.

[54] On stability for optimization problems, *Ann. Probab.*, **2**, 243–255 (1974).

DUDLEY, R. M.

[55] *Real Analysis and Probability*, Wadsworth and Brooks/Cole, Monterey, Calif., 1989.

DURRETT, R.

[56] *Brownian Motion and Martingales in Analysis*, Wadsworth, Belmont, Calif., 1984.

[57] *Lecture Notes on Particle Systems and Percolation*, Wadsworth and Brooks/Cole, Monterey, Calif., 1988.

[58] *Probability: Theory and Examples*, Brooks/Cole, Pacific Grove, Calif., 1991.

EL KAROUI, N. and LEPELTIER, J.-P.

[59] Représentation des processus ponctuel multivariés à l'aide d'un processus de Poisson, *Z. Wahrsch. Verw. Gebiete*, **39**, 111–133 (1977).

ETHIER, S. N. and KURTZ, T. G.

[60] *Markov Processes*, Wiley, New York, 1986.

FELLER, W.

[61] *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., Wiley, New York, 1968.

FREEDMAN, D.

[62] *Markov Chains*, Springer-Verlag, New York, 1983.

FRISTEDT, B. and OREY, S.

[63] The tail σ-field of one dimensional diffusions, in *Stochastic Analysis* (ed. A. Friedman and M. Pinsky), Academic Press, New York, 1978.

GOLDSTEIN, S.

[64] Maximal coupling, *Z. Wahrsch. Verw. Gebiete*, **46**, 193–204 (1979).

GREVEN, A.

[65] Coupling of Markov chains by randomized stopping times, *Probab. Theory Related Fields*, **75**, I: 195–212 and II: 431–458 (1987).

GRIFFEATH, D.

[66] A maximal coupling for Markov chains, *Z. Wahrsch. Verw. Gebiete*, **31**, 95–106 (1975).

[67] Coupling methods for Markov processes, in: *Studies in Probability and Ergodic Theory*, Advances in Mathematics Supplementary Studies, Vol. 2 (ed. G.-C. Rota), Academic Press, New York, 1978.

GRIMMETT, G. R.

[68] *Percolation*, Springer-Verlag, New York, 1989.

GRIMMETT, G. R. and STIRZAKER, D. R.

[69] *Probability and Random Processes*, 2nd ed., Oxford University Press, Oxford, 1992.

GUTTORP, P., KULPERGER, R. and LOCKHART, R.

[70] A coupling proof of weak convergence, J. Appl. Probab., **22**, 447–453 (1985).

HARRIS, T. E.

[71] On chains of infinite order, *Pacific J. Math.*, **5**, 713–724 (1955).

[72] Lower bound for the critical probability in a certain percolation process, *Proc. Cambridge Philos. Soc.*, **56**, 13–20 (1960).

HODGES, J. L., JR. and LE CAM, L.

[73] The Poisson approximation to the Poisson binomial distribution, *Ann. Math. Statist.*, **31**, 737–740 (1960).

HODGES, J. L., JR. and ROSENBLATT, M.

[74] Recurrence-time moments in random walks, *Pacific J. Math.*, **3**, 127–136 (1953).

HOEL, P. G., PORT, S. C. and STONE, C. J.

[75] *Introduction to Stochastic Processes*, Houghton Mifflin, Boston, 1972.

HOLST, L.

[76] On birthday, collector's, occupancy and other classical urn problems, *Internat. Statist. Rev.*, **54**, 15–27 (1986).

HORDIJK, A. and RIDDER, A.

[77] Stochastic inequalities for queueing networks, in *Teletraffic Analysis and Computer Performance Evaluation* (ed. O. J. Boxma et al.), Elsevier Science Publishers, New York, 1986.

[78] Stochastic inequalities for an overflow model, *J. Appl. Probab.*, **24**, 696–708 (1987).

JACOD, J. and SHIRYAEV, A. N.

[79] *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin, 1987.

KAISER, T.

[80] On couplings, Markov chains and random systems with complete connection, in *Proceedings of the Sixth Conference on Probability Theory* (ed. B. Bereanu et al.), Editura Academiei Republicii Socialiste România, Bucuresti, 1981.

KALLENBERG, O.

[81] Splitting at backward times in regenerative sets, *Ann. Probab.*, **9**, 781–799 (1981).

[82] Spreading and predictable sampling in exchangeable sequences and processes, *Ann. Probab.*, **16**, 508–534 (1988).

KAMAE, T. and KRENGEL, U.

[83] Stochastic partial ordering, *Ann. Probab.*, **6**, 1044–1049 (1978).

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L.

[84] Stochastic inequalities on partially ordered spaces, *Ann. Probab.*, **5**, 899–912 (1977).

KARATZAS, I. and SHREVE, S. E.

[85] *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.

KARLIN, S. and TAYLOR, H. M.

[86] *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975.

KELLY, F.

[87] *Reversibility and Stochastic Networks*, Wiley, New York, 1979.

KENDALL, W. S.

[88] Coupling methods and the storage equation, *J. Appl. Probab.*, **20**, 436–441 (1983).

[89] Stochastic differential geometry, a coupling property, and harmonic maps, *J. London Math. Soc. (2)*, **33**, 554–566 (1986).

[90] Nonnegative Ricci curvature and the Brownian coupling property, *Stochastics*, **19**, 111–129 (1986).

KENDALL, W. S. and SAUNDERS I. W.

[91] Epidemic in competition II: the general epidemic, *J. Roy. Statist. Soc. Ser. B*, **45**, 238–244 (1983).

KESTEN, H.

[92] *Percolation Theory for Mathematicians*, Birkhäuser, Boston, 1982.

[93] Percolation theory and first passage percolation, *Ann. Probab.*, **15**, 1231–1271 (1987).

KÜCHLER, U. and LUNZE, U.

[94] On the tail σ-field and the minimal parabolic functions for one-dimensional quasi-diffusions, *Z. Wahrsch. Verw. Gebiete*, **51**, 303–322 (1980).

KULPERGER, R. and GUTTORP, P.

[95] Criticality conditions for some random environment population processes, *Stochastic Process. Appl.*, **11**, 207–212 (1981).

LALLEY, S. P.

[96] Conditional Markov renewal theory I: finite and denumerable state space, *Ann. Probab.*, **12**, 1113–1148 (1984).

[97] Renewal theorem for a class of stationary sequences, *Probab. Theory Related Fields*, **72**, 195–213 (1986).

[98] An extension of Kesten's renewal theorem for random walk in a random environment, *Adv. Appl. Math.*, **7**, 80–100 (1986).

LALLEY, S. and SELLKE, T.

[99] Travelling waves and inhomogeneous branching diffusions, *Ann. Probab.*, I: **16**, 1051–1062 (1988); II: **17**, 116–127 (1989).

[100] Limit theorems for the frontier of a one-dimensional branching diffusion, *Ann. Probab.*, **19** (1992).

LAURITZEN, S.

[101] *Extremal Families and Systems of Sufficient Statistics*, Springer Lecture Notes in Statistics, Vol. 49, Springer-Verlag, New York, 1988.

LE CAM, L.

[102] An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.*, **10**, 1181–1197 (1960).

LEHMANN, E. L.

[103] *Testing Statistical Hypotheses*, Wiley, New York, 1959.

LIGGETT, T. M.

[104] *Interacting Particle Systems*, Springer-Verlag, New York, 1985.

LINDVALL, T.

[105] A probabilistic proof of Blackwell's renewal theorem, *Ann. Probab.*, **5**, 482–485 (1977).

[106] On coupling of discrete renewal processes, *Z. Wahrsch. Verw. Gebiete*, **48**, 57–70 (1979).

[107] A note on coupling of birth and death processes, *J. Appl. Probab.*, **16**, 505–512 (1979).

[108] On coupling of continuous-time renewal processes, *J. Appl. Probab.*, **19**, 82–89 (1982).

[109] Bernstein polynomials and the law of large numbers, *Math. Sci.*, **7**, 127–139 (1982).

[110] On coupling of diffusion processes, *J. Appl. Probab.*, **20**, 82–93 (1983).

[111] On coupling of renewal processes with use of failure rates, *Stochastic Process. Appl.*, **22**, 1–15 (1986).

[112] Ergodicity and inequalities in a class of point processes, *Stochastic Process. Appl.*, **30**, 121–131 (1988).

[113] W. Doeblin 1915–1940, *Ann. Probab.*, **19**, 929–934 (1991).

LINDVALL, T. and ROGERS, L. C. G.

[114] Coupling of multidimensional diffusions by reflection, *Ann. Probab.*, **14**, 860–872 (1986).

LOTWICK, H. W. and SILVERMAN, B. W.

[115] Convergence of spatial birth-and-death processes, *Math. Proc. Cambridge Philos. Soc.*, **90**, 155–165 (1981).

MARSHALL, A. W. and OLKIN, I.

[116] A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, **62**, 30–44, (1983).

MCDONALD, D. R.

[117] On local limit theorem for integer-valued random variables, *Theory Probab. Appl.*, **24**, 613–619 (1979).

MCDONALD, D. R. and RÖSLER, U.

[118] An application of the Bernoulli part to renewal theory and local limit theorems, *Technical report* (1987).

MINEKA, J.

[119] A criterion for tail events for sums of independent random variables, *Z. Wahrsch. Verw. Gebiete*, **25**, 163–170 (1973).

MOLLISON, D.

[120] Spatial contact models for ecological and epidemic spread, *J. Roy. Statist. Soc. Ser. B*, **39**, 283–326 (1977).

NEVEU, J.

[121] *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.

NEY, P.

[122] A refinement of the coupling method in renewal theory, *Stochastic Process. Appl.*, **11**, 11–26 (1981).

NUMMELIN, E.

[123] A splitting technique for Harris recurrent Markov chains, *Z. Wahrsch. Verw. Gebiete*, **43**, 309–318 (1978).

[124] *General Irreducible Markov Chains and Non-negative Operators*, Cambridge University Press, Cambridge, 1984.

O'BRIEN, G. L.

[125] The comparison method for stochastic processes, *Ann. Probab.*, **3**, 80–88 (1975).

[126] A new comparison theorem for solutions of stochastic differential equations, *Stochastics.*, **3**, 245–249 (1980).

OREY, S.

[127] *Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand, London, 1971.

ORNSTEIN, D. S.

[128] Random walks I, *Trans. Amer. Math. Soc.*, **138**, 1–43 (1969).

PITMAN, J. W.

[129] Uniform rates of convergence for Markov chain transition probabilities, *Z. Wahrsch. Verw. Gebiete*, **29**, 193–227 (1974).

[130] An identity for stopping times of a Markov process, in *Studies in Probability and Statistics* (ed. E. J. Williams), North-Holland, Amsterdam, 1975.

PITMAN, J. W. and SPEED, T. P.

[131] A note on random times, *Stochastic Process. Appl.*, **1**, 369–374 (1973).

Port, S. C. and Stone, C. J.

[132] *Brownian Motion and Classical Potential Theory*, Academic Press, New York, 1978.

Preston, C.

[133] Spatial birth-and-death processes, *Bull. Inst. Internat. Statist.*, **46**, 371–391 (1976).

Revuz, D.

[134] *Markov Chains*, North-Holland, Amsterdam, 1984.

Rogers, L. C. G.

[135] Coupling and the tail σ-field of a one-dimensional diffusion, in *Stochastic Calculus in Application* (ed. J. Norris), Longman, Harlow, Essex, England, 1988.

Rogers, L. C. G. and Williams, D.

[136] *Diffusions, Markov Processes, and Martingales*, Vol. 2: *Itô Calculus*, Wiley, New York, 1987.

Rösler, U.

[137] Das 0–1–Gesetz der terminalen σ-Algebra bei Harrisirrfahrten, *Z. Wahrsch. Verw. Gebiete*, **37**, 227–242 (1977).

[138] Eine Mischungseigenschaft von Zufallsvariablen, *Z. Wahrsch. Verw. Gebiete*, **37**, 329–340 (1977).

[139] The tail σ-field of time-homogeneous one-dimensional diffusion processes, *Ann. Probab.*, **7**, 847–857 (1979).

[140] A new Markovian coupling and some applications, *Technical report* (1982).

Serfling, R. J.

[141] Some elementary results on Poisson approximation in a sequence of Bernoulli trials, *SIAM Rev.*, **20**, 567–579 (1977).

Sigman, K.

[142] Queues as Harris recurrent Markov chains, *Queueing Systems*, **3**, 179–198 (1988).

Skorohod, A. V.

[143] *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, Mass., 1965.

Smith, D. R. and Whitt, W.

[144] Resource sharing for efficiency in traffic systems, *Bell System Tech. J.*, **60**, 39–55 (1981).

SONDERMAN, D.

[145] Comparing multi-server queues with finite waiting rooms, *Adv. Appl. Probab.*, **11**, I: 439–447; II: 448–455 (1979).

STEIN, C.

[146] A bound for the error in the normal approximation to the distribution of a sum of independent random variables, in *Proc. 6th Berkeley Symp. Math. Stat.*, Vol. 2. University of California Press, Berkeley, 1970.

STRASSEN, V.

[147] The existence of probability measures with given marginals, *Ann. Math. Statist.*, **36**, 423–439 (1965).

THORISSON, H.

[148] The coupling of regenerative processes, *Adv. Appl. Probab.*, **15**, 531–561 (1983).

[149] The queue $GI/G/k$: finite moments of the cycle variables and uniform rates of convergence, *Stochastic Models*, **1**, 221–238 (1985).

[150] Periodic regeneration, *Stochastic Proc. Appl.*, **20**, 85–104 (1985).

[151] On regenerative and ergodic properties of the k-server queue with non-stationary Poisson arrivals, *J. Appl. Probab.*, **22**, 893–902 (1985).

[152] On maximal and distributional coupling, *Ann. Probab.*, **14**, 873–876 (1986).

[153] A complete coupling proof of Blackwell's renewal theorem, *Stochastic Process. Appl.*, **26**, 87–97 (1987).

[154] Backward limits, *Ann. Probab.*, **16**, 914–924 (1988).

[155] Future independent times and Markov chains, *Probab. Theory Related Fields*, **78**, 143–148 (1988).

[156] Construction of a stationary regenerative process, *Stochastic Process. Appl.*, (1992).

VAN DOORN, E.

[157] *Stochastic Monotonicity and Queueing Applications of Birth–Death Processes*, Springer Lecture Notes in Statistics, Vol. 4, Springer-Verlag, New York, 1981.

WALRAND, J.

[158] *An introduction to Queueing Networks*, Prentice Hall, Englewood Cliffs, N.J., 1988.

WILLIAMS, D.

[159] *Diffusions, Markov Processes and Martingales*, Vol. 1: *Foundations*, Wiley, New York, 1979.

[160] *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

YAMADA, T.

[161] On a comparison theorem for solutions of stochastic differential equations and its applications, *J. Math. Kyoto Univ.*, **13**, 497–512 (1973).

ZHIYUAN, H.

[162] A comparison theorem for solutions of stochastic differential equations and its applications, *Proc. Amer. Math. Soc.*, **91**, 611–617 (1984).

Index