International Statistical Review (2002), **70**, 2, 215–240, Printed in The Netherlands

MC's for MCMC'ists

Esa Nummelin

University of Helsinki, Finland

Summary

We develop a minimum amount of theory of Markov chains at as low a level of abstraction as possible in order to prove two fundamental probability laws for standard *Markov chain Monte Carlo algorithms:*

1. The *law of large numbers* **explains why the algorithm works: it states that the empirical means calculated from the samples converge towards their "true" expected values, viz. expectations with respect to the invariant distribution of the associated Markov chain (= the target distribution of the simulation).**

2. The *central limit theorem* **expresses the deviations of the empirical means from their expected values in terms of asymptotically normally distributed random variables. We also present a formula and an estimator for the associated variance.**

Key words: Markov chain Monte Carlo; Convergence; Central limit theorem.

1 Introduction

The theory of Markov chains (MC's) on countable state spaces comprises basic material for traditional courses on stochastic processes. Due to the advent of the *Markov chain Monte Carlo (MCMC) method*, the more abstract theory of MC's on "continuous" (general measurable) state spaces has become relevant for statisticians and applied mathematicians.

After the pioneering studies by W. Doeblin, J.L. Doob, T. Harris and S. Orey, several monographs have been written on general MC's (e.g. Revuz (1975), Nummelin (1984), Meyn & Tweedie (1993)) covering a broad area of MC theory as well as its applications. However, none of these focuses directly on the needs arising from the MCMC method.

The purpose of the MCMC method is to provide a tool for the simulation of p.d.f.'s whose direct simulation as sample means of i.i.d. sequences is impossible in practice (notably due to high dimensionality). The fundamental idea underlying the MCMC method involves the observation, according to which an apriori impossible simulation task may become possible if suitable statistical dependence (of Markovian type) is allowed within the simulating sequence. Thus the idea is to sample from a Markov chain whose invariant distribution is precisely the target distribution of the simulation. The time-averages will then produce estimates for the target distribution.

We shall here deal with the following fundamental questions:

1. "Why does an MCMC algorithm work, i.e., why does the *law of large numbers* hold true for the algorithm?"
2. "How accurate is the algorithm, i.e., does the *central limit theorem* hold true, and what is the associated variance?"

In exact mathematical terms, an MCMC algorithm is defined as a Markov chain $X_0, X_1, \ldots$ on a Euclidean space R^k and having the unknown *target distribution* $\pi(x), x \in R^k$, as its *invariant distribution*.

Let $\hat{\pi}_n(A)$ denote the mean number of visits by the MC to (measurable) subsets $A \subset R^k$ in n steps:

$$\hat{\pi}_n(A) := n^{-1}\#\{0 \leq m \leq n-1 : X_m \in A\}.$$

The collection $\hat{\pi}_n(A)$, $A \subset R^k$, comprises the *empirical distribution* (with *sampling time* n).

The *(strong) law of large numbers* asserts that the empirical distribution converges towards the target distribution:

$$\lim_{n\to\infty} \hat{\pi}_n(A) = \int_A \pi(x)dx, \quad \text{with probability 1.} \tag{1.1}$$

According to the *central limit theorem* the deviations of the empirical probabilities from their true values are asymptotically normally distributed:

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\hat{\pi}_n(A) - \int_A \pi(x)dx) = \mathcal{N}(0, \sigma_A^2), \quad \text{in distribution.} \tag{1.2}$$

(We denote by $\mathcal{N}(0, \sigma^2)$ a generic standard normally distributed random variable with mean 0 and variance σ^2. The subscript A refers to dependence of the variance on the subset A.)

In the MC literature the LLN and the CLT are formulated and proved within the general abstract MC framework, and consequently the learning of the underlying theory requires a substantial amount of work. We argue here that these fundamental results concerning the statistical behaviour of MCMC algorithms can be proved without relying on the (abstract) general theory of MC's.

This is because, for an MCMC-algorithm the basic set-up includes already the existence of an invariant distribution for the associated Markov chain, and hence it is natural to take this fact as the very basic starting point for the analysis (see Section 2.3). In fact, it turns out that, once the mere existence of an invariant p.d.f. $\pi(x)$ is known, the LLN and CLT can be proved with minor extra hypotheses. The proof of the LLN will require a weak "irreducibility hypothesis" which in practice is trivially true (see Section 2.2). For the CLT it will be necessary to introduce a somewhat more complicated condition of stochastic stability expressing ergodicity of degree 2 of the MC (see Sections 4.1–2). It is worth emphasizing that abstract concepts such as e.g. the (measure theoretic) irreducibility, existence of so-called C-sets, Harris recurrence, or convergence of the individual transition probabilitites, are not of primary relevance for the theory of MCMC algorithms.

The basic mathematical tool here will be the use of the *regeneration method*, cf. [MT, Section 5.1], [N, Section 4.3]. (We shall refer to the monographs by Meyn & Tweedie (1993) and by Nummelin (1984) briefly as [MT] and [N], respectively.) The principal idea of the regeneration method involves the construction of "artificial atoms" which will play the same role as individual states of Markov chains in countable state spaces. Since the target distributions of MCMC simulations are usually non-singular (w.r.t. the Lebesgue measure), i.e., they can be described in terms of probability density functions, it is desirable to modify the regeneration technique accordingly. An introduction to the central ideas of the regeneration method is given in Sections 3.2–6 and 4.2–3. After these preliminaries the actual proofs of the LLN and CLT become rather easy (Sections 3.7 and 4.4). The multivariate CLT is briefly discussed in Section 4.5*. In Section 4.6* we elaborate on the variance and its consistent estimation. (The asterisk indicates the supplementary character of these sections.)

We shall not dwell here on the importance of the MCMC method for statistical estimation but assume simply that a reader of this article is apriori motivated to study the underlying theory. Nor do we go into lengthy descriptions of the variety of different MCMC algorithms and their practical applicability in different concrete situations.

There exists today an abundance of scientific articles dealing with various theoretical aspects of the MCMC method. This poses the problem of a just choice of the references to be included. Since the principal purpose of the present article is pedagogical rather than expository we have chosen a

"minimalistic" solution. (The other available choice would involve the inclusion of "all relevant" references.) Thus, as regards the MCMC method itself, we refer only to the early, influential review article by Tierney (1994) and to the recent comprehensive monograph by Robert & Casella (1999).

In accordance with the chosen line we shall not refer to original articles on MC theory either but lean on the monographs [MT] and [N]. The bibliographies of these sources contain further references.

The article attempts to be self-contained. In particular, no previous knowledge on the general theory of MC's is required. As prerequisites the reader is supposed to know the ordinary LLN and CLT for sums of i.i.d. random variables. In the proof of the CLT Kolmogorov's inequality for sums of i.i.d. random variables will be utilized.

We point out the connections to the general MC theory in separate remarks (indicated with an asterix). In one remark we refer to the relevance of the present results outside the MCMC context (to the *theory of dynamical systems*).

The book [N] has been critized as being somewhat hard to read. At the time of its appearence in the mid 80's the appreciation of the practical importance of the general MC theory which resulted from the wide spread use of MCMC methods for carrying out numerical computations particularly in Bayesian statistics, laid still ahead in the future. The present article could also be seen as an attempt by the author to respond to this changed atmosphere around the subject.

The article grew out from lecture notes for a course on Markov chains intended for students and researchers of the MCMC method. I am grateful for the valuable criticism on the text which I obtained from Elja Arjas and Pekka Nieminen.

With the passing away of Richard Tweedie and Vladimir Kalashnikov the probability community recently lost two of its eminent members. Since having had Richard as my thesis examiner 25 years ago I got to know Richard also as an inspiring coworker. Riku (which was his Finnish nickname) was a long-time friend of the small group of Finnish probabilists. Vladimir Kalashnikov I learnt to know first as the skillful and devoted translator of my book [N] into Russian. Later I had the pleasure to get to know Volodja also as a friend. I dedicate this article to the memory of Richard Tweedie and Vladimir Kalashnikov.

2 Description of the Model

2.1 The Markov Chain

In this paper we study the following class of Markov chains:

Let $E \subset R^k$ be a given subset of the Euclidean space R^k, referred to as the *state space*. Let $X_0, X_1, \ldots$ be a sequence of E-valued random variables. We assume that the sequence (X_n) forms a *time-homogeneous Markov chain*. At each time $n \geq 0$ the MC will either *stay put* , or it will *make a jump*. In exact mathematical terms, conditional on knowing the *pre-n-path* of the MC (X_n), viz. conditionally on $X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x$, for states $x_0, \ldots, x_{n-1}, x \in E$, the next state X_{n+1}

(i) is equal to the previous state X_n : $X_{n+1} = x$ with a probability $0 \leq r(x) < 1$ depending on x; or

(ii) moves to a new state y according to a subprobability density function $p(x, y)$ (depending on the state x).

(A function $\lambda(y) \geq 0$ is called a *subprobability density function* if $\int \lambda(y)dy \leq 1$. If $\int \lambda(y)dy = 1$, then λ is called simply a *probability density function, or p.d.f.*) It follows that the the *total probability for a jump to occur* is

$$\int p(x, y)dy = 1 - r(x).$$

We can characterize the MC in terms of its *transition probabilities*

$$P(x, A) := P(X_{n+1} \in A | X_n = x)$$
$$= \int_A p(x, y)dy + r(x)\delta_x(A), \quad x \in E, A \subset E. \tag{2.1.1}$$

(The symbol δ_x refers to "unit mass at x": $\delta_x(A) := 1$ if $x \in A$, $:= 0$ if $x \in A^c$.)

It follows that if the state X_n has probability density function $\lambda(x)$, i.e.,

$$P(X_n \in A) = \int_A \lambda(x)dx,$$

then the next state X_{n+1} will have p.d.f.

$$\lambda P(y) := \int \lambda(x)p(x, y)dx + \lambda(y)r(y). \tag{2.1.2}$$

Namely,

$$P(X_{n+1} \in A) = \int \lambda(x)P(x, A)dx$$
$$= \int_E \int_A \lambda(x)p(x, y)dxdy + \int_A \lambda(x)r(x)dx$$
$$= \int_A \lambda P(y)dy.$$

The transition probability kernel P induces in this way a *Markov operator* which maps a given p.d.f. λ to the p.d.f. λP. (Positive operators which preserve integrals, i.e., map p.d.f.'s to p.d.f.'s, are commonly referred to as Markov operators.) We denote by P^n the *n'th iterate*, viz. we define recursively:

$$\lambda P^n := (\lambda P^{n-1})P, \text{ for } n = 2, 3, \ldots.$$

By convention P^0 is interpreted as the identity operator: $\lambda P^0 := \lambda$.

Example 1. The *Metropolis–Hastings algorithm* is defined by specifying a p.d.f. $\pi(x)$, $x \in E$, and a transition p.d.f. $q(x, y)$:

$$\int q(x, y)dy \equiv 1.$$

It is assumed that the shape of the p.d.f. π is known, i.e., π is known up to a multiplicative constant. The algorithm evolves as follows:

If the Markov chain is currently in state $X_n = x$, then it generates a "candidate value" y for the next state from the p.d.f. $q(x, y)$. This candidate is "accepted with probability"

$$\alpha(x, y) := \min\left\{\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right\} \text{ if } \pi(x)q(x, y) > 0,$$
$$\alpha(x, y) := 1 \text{ if } \pi(x)q(x, y) = 0.$$

In the case of "rejection" the MC will stay put where it is, viz. at x. Thus the final (sub-)probability d.f. for a jump from state x becomes

$$p(x, y) = \alpha(x, y)q(x, y).$$

The probability for staying at x is

$$r(x) = 1 - \int p(x, y)dy = 1 - \int \alpha(x, y)q(x, y)dy.$$

In fact, most of the commonly used (time-homogeneous) MCMC algorithms are special cases of the Metropolis–Hastings algorithm, see e.g. Tierney (1994).

Example 2. The (standard) *Gibbs sampler* is defined in terms of (known) conditional p.d.f.'s $\pi(x_j|x_i, i \neq j)$, $j = 1, \ldots, k$, of an (unknown) target p.d.f. $\pi(x) = \pi(x_1, \ldots, x_k)$. The transition p.d.f. is:

$$p(x_1, \ldots, x_k; y_1, \ldots, y_k)$$

$$= \pi(y_1|x_2, \ldots, x_k)\pi(y_2|y_1, x_3, \ldots, x_k) \cdots \pi(y_{k-1}|y_1, \ldots, y_{k-2}, x_k)\pi(y_k|y_1, \ldots, y_{k-1}).$$

(It follows that $r(x) \equiv 0$.)

*Example 3**. It is worthwile to point that the theory presented here has some relevance to the *theory of dynamical systems*, too. Namely, the singular component of the transition probability density function (alternative (i)) could be "more complicated" and the continuous component (alternative (ii)) "negligible". To this end, let $g_i(x)$, $i = 1, \ldots, d$, denote d continuously differentiable maps defined on open connected subsets $E_i \subset E$, $i = 1, \ldots, d$, respectively. We assume that each of the maps $g_i : E_i \to E$ is injective, i.e., has an inverse map $g_i^{-1} : g_i(E_i) \to E_i$. Let $0 \leq r_i(x) < 1$, $i = 1, \ldots, d$, denote *probabilities for the d alternative singular moves* $x \to g_i(x)$. In precise terms, the Markov chain is supposed to evolve as follows:

Conditional on $X_n = x$ the next state X_{n+1} is

(i) equal to the state $g_i(x)$ with probability $r_i(x)$, $i = 1, \ldots, d$; or

(ii) distributed according to a subprobability density function $p(x, y)$ (depending on the state x).

It follows that the total probability of a "statistically continuous" move from x is

$$\int p(x, y)dy = 1 - r(x),$$

where $r(x) := \sum_{i=1}^{d} r_i(x)$. Now the corresponding Markov operator P mapping a given p.d.f. λ to the p.d.f. λP has the form

$$\lambda P(y) := \int \lambda(x)p(x, y)dx + \sum_{i:\ y \in g_i(E_i)} r_i(g_i^{-1}(y))|g_i'(g_i^{-1}(y))|^{-1}\lambda(g_i^{-1}(y)).$$

If the total probability for the "continuous jumps" is "small" (but still non-zero), i.e., if $r(x) \approx 1$, then we are dealing with a "small perturbation" of a *dynamical system induced by the random maps* $g_i(x)$.

An important special case is the case where the domains E_i are disjoint. Then we are dealing with (a small perturbation of) a *dynamical system induced by the deterministic map*:

$$g(x) := g_i(x), \text{ for } x \in E_i.$$

2.2 *Irreducibility*

We shall now formulate an *irreducibility hypothesis*. A subset $I \subset E$ having volume $|I| > 0$ will be called *small*, if there exists a subset $J \subset E$ with volume $|J| > 0$ and a positive constant $\beta > 0$ such that

$$p(x, y) \geq \beta \text{ whenever } x \in I,\ y \in J. \tag{2.2.1}$$

Hypothesis 1. There exists a small set I such that for each initial state $x \in E$,

$$P^{n(x)}(x, I) := P(X_{n(x)} \in I | X_0 = x) > 0,$$

for some integer $n(x) \geq 1$ depending on x.

Remarks. (i) In practice the p.d.f. $p(x, y)$ is often (at least piecewise) continuous as a function of its variables x and y. In fact, for the Hypothesis 1 it is sufficient to assume that there exists a pair of points $(x_0, y_0) \in E \times E$ such that $p(x_0, y_0) > 0$, $p(x, y)$ is continuous at (x_0, y_0), and the MC visits arbitrarily small x_0-centered balls with positive probability (eventually from an arbitrary initial state).

(ii) As a special case, suppose that $p(x, y)$ has the form $p(x, y) = f(y - x)$ where $f(x)$, $x \in E$, is a given continuous p.d.f. In this case the algorithm performs a *random walk* with *increment p.d.f.* f. Now it is well known that a random walk, which has a continuous increment distribution, visits all open balls with a positive probability, see e.g. Revuz (1975). Thus, in view of the previous remark, Hypothesis 1 is satisfied.

(iii)* Hypothesis 1 implies that the MC is *ϕ-irreducible*, where ϕ is the restriction of the Lebesgue measure to the subset J:

$$\phi(A) := |A \cap J| := \int_{A \cap J} dx, \quad A \subset E. \tag{2.2.2}$$

Namely,

$$P^{n(x)+1}(x, A) \geq \beta P^{n(x)}(x, I)\phi(A) > 0 \text{ for all } x \in E, \ A \subset E \text{ with } \phi(A) > 0. \tag{2.2.3}$$

(iv)* The present definition of a small set, as well as the definition of a *weakly small set* given later in Section 4.3, are special cases of a more abstract definition of a small set, cf. [MT, Section 5.2], [N, Section 2.3]. Small sets are also called *C-sets* in the literature. According to a deep theorem (originally due to W. Doeblin, J.L. Doob and S. Orey) *ϕ-irreducibility* implies the existence of small sets. (However, as noted above, the direct verification of the existence of small sets is usually easy. Thus the abstract concept of ϕ-irreducibility does not play an important role in practice.)

2.3 *The Invariant Distribution*

Definition 1. A p.d.f. π is called *invariant* if

$$\pi P = \pi. \tag{2.3.1}$$

In view of (2.1.2) this is the same as

$$\int \pi(x)p(x, y)dx \equiv (1 - r(y))\pi(y). \tag{2.3.2}$$

We now postulate the existence of (at least one) invariant p.d.f.:

Hypothesis 2. (i) The MC (X_n) has an invariant p.d.f. $\pi = (\pi(x), x \in E)$.
(ii) Moreover, we assume that the *support*

$$S := \{x \in E : \pi(x) > 0\}$$

of the invariant p.d.f. is (statistically) closed in the sense that

$$P(x, S) = 1 \text{ for all } x \in S.$$

Remarks. (i) We do not require uniqueness of the invariant p.d.f. (However, this turns out later to follow as a corollary from the Hypotheses 1 and 2, see Section 3.5.)

(ii) *Reversible MC's*, i.e., MC's for which there exists a p.d.f. λ such that

$$\lambda(x)p(x, y) \equiv \lambda(y)p(y, x),$$

possess an invariant p.d.f. This is because the p.d.f. λ itself is invariant:

$$\int \lambda(x)p(x, y)dx = \lambda(y) \int p(y, x)dx = (1 - r(y))\lambda(y).$$

Due to its reversibility, the Metropolis–Hastings algorithm has the target p.d.f. $\pi(x)$ as its invariant p.d.f.

(iii) The target distribution $\pi(x_1, \ldots, x_k)$ is invariant for the Gibbs sampler.

(iv)* The closedness hypothesis is not very restrictive. In fact, in practice usually $\pi(x) > 0$ for all $x \in E$ so that $S = E$. Even in the case where the inclusion $S \subset E$ is strict, closedness will automatically follow after a modification of the invariant p.d.f. π. Namely, in view of the identities

$$\int \pi(x)P(x, S)dx = \int_S \pi(y)dy = 1$$

we have $P(x, S) = 1$ for *almost every* initial state $X_0 = x \in S$ (i.e., for all $x \in S$ outside a set of volume zero). An invariant p.d.f. π can be modified on sets of volume zero and therefore it can be set equal to zero for $x \in S$ for which $P(x, S) < 1$. After this modification the support S becomes statistically closed.

3 The Law of Large Numbers

3.1 Formulation of the LLN

The principal goal of this paper is to prove that the strong law of large numbers holds true for MC's satisfying the Hypotheses 1 and 2. We say that a function $f(x)$, $x \in E$, is *π-integrable* if

$$\pi(f) := \int f(x)\pi(x)dx$$

is finite. We denote the arithmetic means by

$$\hat{\pi}_n(f) := n^{-1}(f(X_0) + \cdots + f(X_{n-1})).$$

THEOREM 1. *Assume the Hypotheses 1 and 2. Then*

$$\lim_{n\to\infty} \hat{\pi}_n(f) = \pi(f), \text{ with probability 1,}$$

for all π-integrable functions f and all initial states $X_0 = x$ belonging to the support S of the invariant p.d.f. π.

Remarks. (i) By setting $f(x) := \delta_x(A)$ for subsets $A \subset E$ we obtain the formulation (1.1) of the LLN appearing in the Introduction.

(ii) The expression $\hat{\pi}_n(f)$ can be interpreted as an integral of the "test function" f with respect to the empirical distribution defined as the (random) measure

$$\hat{\pi}_n := n^{-1}(\delta_{X_0} + \cdots + \delta_{X_{n-1}}).$$

Namely, since

$$\int f(y)\delta_x(dy) = f(x),$$

it follows that

$$\hat{\pi}_n(f) := n^{-1}(f(X_0) + \cdots + f(X_{n-1})) = \int f(y)\hat{\pi}_n(dy).$$

For subsets $A \subset E$ we have

$$\hat{\pi}_n(A) := n^{-1}\#\{0 \le m \le n-1 : X_m \in A\} = \int_A \hat{\pi}_n(dy).$$

The LLN can be interpreted as expressing that the (sequence of) *empirical distributions* $\hat{\pi}_n$ *is a consistent estimator for the invariant p.d.f.* π.

3.2 *Regeneration*

We shall now describe the regeneration scheme, i.e., show how the path of the MC can be decomposed into *independent and identically distributed blocks.*

We denote by ν the *uniform p.d.f.* on the set J:

$$\nu(y) \equiv |J|^{-1} \text{ for } y \in J, \equiv 0 \text{ elsewhere.}$$

(As before $|J| := \int_J dy$ denotes the volume of the set J.) We define the function

$$s(x) \equiv \beta|J| \text{ for } x \in I, \equiv 0 \text{ elsewhere.}$$

Note that, due to the Hypothesis 1,

$$p(x, y) \ge s(x)\nu(y) \text{ for all } x, y \in E.$$

Let

$$\begin{aligned} Q(x, A) :&= P(x, A) - s(x)\int_A \nu(y)dy \\ &= P(x, A) - \beta|A \cap J| \text{ for } x \in I, A \subset E, \\ &= P(x, A) \text{ for } x \in I^c, A \subset E. \end{aligned} \tag{3.2.1}$$

Note that $Q(x, A) \ge 0$ always. We shall interpret Q as a *sub-Markov operator* which maps a given subprobability density function λ to the subprobability d.f.

$$\lambda Q(y) := \lambda P(y) - (\int \lambda(x)s(x)dx)\nu(y), \quad y \in E.$$

We define a *bivariate Markov chain* (X_n, Y_n) as follows. Let $X_0, X_1, \ldots$ be a sequence of E-valued random variables and let $Y_0, Y_1, \ldots$ be a sequence of binary ($\{0, 1\}$-valued) random variables. The transition probabilities of the bivariate MC are defined by:

$$P(X_n \in A; Y_n = 1 | X_{n-1} = x; Y_{n-1}) := s(x)\int_A \nu(y)dy, \tag{3.2.2}$$

$$P(X_n \in A; Y_n = 0 | X_{n-1} = x; Y_{n-1}) := Q(x, A), \tag{3.2.3}$$

for all $n \ge 1$, $A \subset E$, independently of the r.v. Y_{n-1}.

Notice first that *the marginal probability law of the first component* X_n *is the original probability*

law of the MC (X_n). Namely, by (3.2.2) and (3.2.3) and by the Markov property, we have

$$\begin{aligned} &P(X_n \in A | X_0, \dots, X_{n-2}; X_{n-1} = x) \\ &= P(X_n \in A, Y_n = 0 | X_{n-1} = x) + P(X_n \in A, Y_n = 1 | X_{n-1} = x) \\ &= Q(x, A) + s(x) \int_A \nu(y) dy = P(x, A) \end{aligned}$$

as required.

Also note that

$$P(Y_n = 1 | X_{n-1} = x; Y_{n-1}) = s(x), \text{ and}$$

$$P(Y_n = 0 | X_{n-1} = x; Y_{n-1}) = 1 - s(x).$$

It may be helpful to imagine that the r.v. Y_n would be the result of a toss of a (biased) coin: Conditionally on that the state of the MC at time $n-1$ is $X_{n-1} = x$, the result is "head" or "tail" with respective probabilities $s(x)$ and $1 - s(x)$ (independently of the paths leading to x at time $n-1$).

It turns out that the *bivariate MC* (X_n, Y_n) *regenerates at the epochs at which* $Y_n = 1$. Namely, conditionally on the pre-$(n-1)$-path $(X_0, Y_0), \dots, (X_{n-1}, Y_{n-1})$ and on the event $Y_n = 1$, the r.v. X_n is distributed according to the p.d.f. ν:

$$\begin{aligned} &P(X_n \in A | X_{n-1} = x; Y_{n-1}; Y_n = 1) \\ &\quad = (P(Y_n = 1 | X_{n-1} = x; Y_{n-1}))^{-1} P(X_n \in A; Y_n = 1 | X_{n-1} = x; Y_{n-1}) \\ &\quad = s(x)^{-1} s(x) \int_A \nu(y) dy = \int_A \nu(y) dy. \end{aligned}$$

More generally, due to the Markov property,

$$\begin{aligned} &P(X_n \in A_0, X_{n+1} \in A_1, \dots; Y_{n+1} = y_1, Y_{n+2} = y_2, \dots | X_0, \dots, X_{n-1}; Y_0, \dots, Y_{n-1}; Y_n = 1) \\ &\quad = P(X_0 \in A_0, X_1 \in A_1, \dots; Y_1 = y_1, Y_2 = y_2, \dots | Y_0 = 1) \\ &\quad = P_\nu(X_0 \in A_0, X_1 \in A_1, \dots; Y_1 = y_1, Y_2 = y_2, \dots) \end{aligned} \tag{3.2.4}$$

for all $A_0, A_1, \dots \subset E$, $y_1, y_2, \dots \in \{0, 1\}$. (The subscript "$\nu$" refers to the initial p.d.f. ν.)

For a later purpose (see Section 4.3) we record the following inequality which is valid for all $n \geq 0, x \in I$:

$$\begin{aligned} &P(X_n \in A | X_0, \dots, X_{n-2}; Y_0, \dots, Y_{n-1}; X_{n-1} = x; Y_n = 0) \\ &\quad = (1 - s(x))^{-1} Q(x, A) \leq (1 - \beta |J|)^{-1} P(x, A) \\ &\quad = (1 - \beta |J|)^{-1} P(X_n \in A | X_0, \dots, X_{n-2}; Y_0, \dots, Y_{n-1}; X_{n-1} = x). \end{aligned} \tag{3.2.5}$$

Let $T \geq 1$ denote the *first regeneration epoch:*

$$T := \min\{n \geq 1 : \ Y_n = 1\}.$$

More generally, let $1 \leq T_1 \leq T_2 \leq \dots$ denote the *successive regeneration epochs:*

$$\begin{aligned} T_1 &:= T, \\ T_i &:= \min\{n > T_{i-1} : \ Y_n = 1\} \text{ for } i = 2, 3, \dots . \end{aligned}$$

We shall later prove that, under the Hypotheses 1 and 2, all the regeneration epochs T_i, $i = 1, 2, \dots$, are finite (for all initial states $X_0 = x \in S$, see Section 3.5).

It follows from (3.2.4) that

$$P(X_{T_i} \in A_0, X_{T_i+1} \in A_1, \dots, X_{T_i+m-1} \in A_{m-1}; T_{i+1} - T_i = m | X_0, \dots, X_{T_i-1}; T_1, \dots, T_{i-1}; T_i = n)$$

$$= P(X_0 \in A_0, \ldots, X_{m-1} \in A_{m-1}; T = m | Y_0 = 1)$$
$$= P_\nu(X_0 \in A_0, \ldots, X_{m-1} \in A_{m-1}; T = m).$$

Thus the *random blocks*

$$\xi_0 := (X_0, \ldots, X_{T-1}; T),$$
$$\xi_i := (X_{T_i}, \ldots, X_{T_{i+1}-1}; T_{i+1} - T_i), \text{ for } i = 1, 2, \ldots,$$

are independent. Moreover the blocks ξ_i, $i = 1, 2, \ldots$, *are identically distributed, obeying the same probability law as the block* ξ_0 *under the initial p.d.f.* ν. In particular, it follows that

$$P(T_{i+1} - T_i = m | X_0, \ldots, X_{n-1}; T_1, \ldots, T_{i-1}; T_i = n) = P_\nu(T = m). \tag{3.2.6}$$

For any function $f(x), x \in E$, we define the following *random sums over the blocks* ξ_i:

$$\zeta_0(f) := \sum_{m=0}^{T-1} f(X_m), \text{ and} \tag{3.2.7}$$

$$\zeta_i(f) := \sum_{m=T_i}^{T_{i+1}-1} f(X_m) \text{ for } i = 1, 2, \ldots . \tag{3.2.8}$$

These sums are independent, too. Moreover, the r.v.'s $\zeta_i(f)$, $i = 1, 2, \ldots$, are identically distributed, obeying the same probability law as the r.v. $\zeta_0(f)$ under the initial p.d.f. ν.

3.3 The Potential Function

The finiteness of the expectations of the regeneration times can be conveniently examined by using the so-called *potential function*, which is defined below in (3.3.4).

We first define a *terminating MC* $(\tilde{X}_n)$ (with *lifetime* $T - 1$) as follows. Let the symbol ∂ denote an *auxiliary state* ("cemetery") not belonging to the state space E. We set

$$\tilde{X}_n = X_n \text{ for } 0 \leq n \leq T - 1,$$
$$\tilde{X}_n = \partial \text{ for } n \geq T.$$

The MC $(\tilde{X}_n)$ has transition (sub)probabilities

$$P(\tilde{X}_n \in A | \tilde{X}_{n-1} = x) = P(X_n \in A, T > n | X_{n-1} = x, T > n - 1)$$
$$= P(X_n \in A, Y_n = 0 | X_{n-1} = x, Y_{n-1} = 0)$$
$$= Q(x, A).$$

Let the common (initial) distribution of X_0 and $\tilde{X}_0$ be arbitrary, denoting it by λ:

$$p.d.f.\ (\tilde{X}_0) = p.d.f.\ (X_0) = \lambda.$$

It follows that the subprobability density function of the state $\tilde{X}_n$ is λQ^n, viz.

$$P(\tilde{X}_n \in A) = P(X_n \in A, T > n) = \int_A \lambda Q^n(x) dx. \tag{3.3.1}$$

Thus e.g.

$$P(T \geq n) = P(X_{n-1} \in E)$$
$$= \int \lambda Q^{n-1}(x) dx. \tag{3.3.2}$$

We also see that

$$P(T = n) = \int \lambda Q^{n-1}(x) P(Y_n = 1 | X_{n-1} = x, Y_{n-1} = 0) dx$$
$$= \int \lambda Q^{n-1}(x) s(x) dx. \tag{3.3.3}$$

Let

$$\mu(x) := \sum_{n=0}^{\infty} \nu Q^n(x) \tag{3.3.4}$$

denote the *potential function*. Conditionally on the initial p.d.f. ν, the expected number of visits by the MC (X_n) during the first block ζ_0 to any subset $A \subset E$ is obtained as the integral of the potential function over A:

$$E_\nu \sum_{n=0}^{T-1} \delta_{X_n}(A) = \sum_{n=0}^{\infty} P_\nu(X_n \in A, T > n)$$
$$= \sum_{n=0}^{\infty} \int_A \nu Q^n(x) dx \text{ by (3.3.1)}, \tag{3.3.5}$$
$$= \int_A \mu(x) dx.$$

In particular, by setting $A = E$ we see that the *expected regeneration time* is equal to the integral of the potential function over the whole state space:

$$M := E_\nu T = \int \mu(x) dx. \tag{3.3.6}$$

More generally, for any non-negative measurable function $f(x)$, $x \in E$, we have:

$$E_\nu \zeta_0(f) = E_\nu \sum_{n=0}^{T-1} f(X_n) = \int \mu(x) f(x) dx. \tag{3.3.7}$$

Finally, let us note that, by setting $\lambda = \nu$ and summing over n in (3.3.3) we obtain the formula

$$P_\nu(T < \infty) = \int \mu(x) s(x) dx. \tag{3.3.8}$$

3.4 Some Decomposition Identities

Let $n \geq 1$ be a fixed integer. We may decompose the event $\{T \leq n\}$ according to the *time elapsed since the last regeneration before n*:

$$\{T \leq n\} = \bigcup_{k=0}^{n-1} \{L_n = k\},$$

where

$$L_n := \min\{0 \leq k \leq n-1 : \ Y_{n-k} = 1\}.$$

Thus we obtain the following identity, which is valid for any initial p.d.f. λ, and for all $n \geq 1$, $A \subset E$:

$$\begin{aligned} P(X_n \in A) &= P(X_n \in A, T > n) + \sum_{k=0}^{n-1} P(L_n = k, X_n \in A) \\ &= P(X_n \in A, T > n) + \sum_{k=0}^{n-1} P(Y_{n-k} = 1, Y_{n-k+1} = 0, \ldots, Y_n = 0; X_n \in A) \\ &= P(X_n \in A, T > n) + \sum_{k=0}^{n-1} P(Y_{n-k} = 1) P_\nu(Y_1 = 0, \ldots, Y_k = 0; X_k \in A) \\ &= P(X_n \in A, T > n) + \sum_{k=0}^{n-1} \int \lambda P^{n-k-1}(y) s(y) dy \int_A \nu Q^k(x) dx. \end{aligned} \tag{3.4.1}$$

In particular, if the initial p.d.f. is invariant, i.e., $\lambda = \pi$, then this identity becomes

$$\int_A \pi(y) dy = P_\pi(X_n \in A, T > n) + \int \pi(x) s(x) dx \sum_{k=0}^{n-1} \int_A \nu Q^k(y) dy. \tag{3.4.2}$$

(The subscript π refers to the particular initial distribution $\lambda = \pi$.) Further, by setting $A = E$ we obtain

$$1 = \int \pi(y) dy = P_\pi(T > n) + \int \pi(x) s(x) dx \sum_{k=0}^{n-1} \int \nu Q^k(y) dy. \tag{3.4.3}$$

Letting $n \to \infty$ leads to the identity

$$1 = P_\pi(T = \infty) + M \int \pi(x) s(x) dx. \tag{3.4.4}$$

Since

$$\int \pi(x) s(x) dx = \beta |J| \int \pi(x) \sum_{n=1}^{\infty} 2^{-n} P^n(x, I) dx > 0 \tag{3.4.5}$$

by the basic Hypotheses 1 and 2, it follows that the expectation M of the first inter-regeneration time is finite:

$$M := E_\nu T = \int \mu(x) dx < \infty, \text{ see (3.3.6).} \tag{3.4.6}$$

Therefore also

$$P_\nu(T < \infty) = \int \mu(x) s(x) dx = 1. \tag{3.4.7}$$

By using induction, it is easy to see that, due to (3.2.6) and (3.4.7), we have

$$P(T_i < \infty | X_0 = x) = P(T < \infty | X_0 = x) \tag{3.4.8}$$

for an arbitrary initial state $x \in E$ and arbitrary $i \geq 1$.

3.5 The Invariant Distribution

By (3.4.6) the *function $M^{-1}\mu(x)$ is a probability distribution function. We claim that it is invariant.* But this is almost trivial; namely,

$$\begin{aligned}\mu(y) &\equiv \nu(y) + \sum_{n=1}^{\infty} \nu Q^n(y) \\ &\equiv \nu(y) + \sum_{n=0}^{\infty} (\nu Q^n) Q(y) \\ &\equiv (\int \mu(x)s(x)dx)\nu(y) + \mu Q(y) \text{ by (3.4.7)} \\ &\equiv \mu P(y) \text{ by (3.2.1).}\end{aligned}$$

Next we assert that the *invariant p.d.f. is unique, viz.*

$$\pi = M^{-1}\mu. \tag{3.5.1}$$

To this end, let $n \to \infty$ in the identity (3.4.2). The result is the inequality

$$\int_A \pi(y)dy \geq \int \pi(x)s(x)dx \int_A \mu(y)dy$$

which is valid for all $A \subset E$. It follows that the function

$$\lambda(y) := \pi(y) - (\int \pi(x)s(x)dx)\mu(y)$$

is non-negative (a.e.). If $\lambda(y) \equiv 0$ (a.e.), then $\pi(y) \equiv (\int \pi(x)s(x)dx)\mu(y)$ (a.e.) implying the uniqueness. By (3.5.1) we have

$$\int \pi(x)s(x)dx = M^{-1}. \tag{3.5.2}$$

If $\int \lambda(y)dy$ were > 0, then by (3.4.7) $(\int \lambda(y)dy)^{-1}\lambda(x)$ would be an invariant p.d.f. satisfying $\int \lambda(x)s(x)dx = 0$. But this is impossible by (3.4.5). Therefore $\pi = M^{-1}\mu$ is indeed unique.

3.6 Recurrence

Finally we show that the MC is *recurrent* in the sense that regenerations occur infinitely often with probability 1 (from an arbitrary initial state $X_0 = x$ belonging to the support $S = \{\pi > 0\}$ of the invariant p.d.f. π). Thus we claim that

$$P(T_i < \infty | X_0 = x) = 1 \text{ for all } i = 1, 2, \ldots, \text{ all } x \in S. \tag{3.6.1}$$

In view of (3.4.8) it is sufficient to prove the finiteness of only the first regeneration time $T = T_1$.

We claim that the first regeneration time T is finite from *almost every* initial state $X_0 = x \in S$. To this end, notice that by (3.4.4) and (3.5.2)

$$1 = P_\pi(T = \infty) + M^{-1}M,$$

whence $\int P(T < \infty | X_0 = x)\pi(x)dx = 1$, i.e.,

$$P(T < \infty | X_0 = x) = 1 \text{ for a.e. } x \in S. \tag{3.6.2}$$

In order to prove that T is finite from *every* initial state $X_0 = x \in S$, let us define

$$h_\infty(x) := P(T = \infty | X_0 = x).$$

In view of (3.3.1), $h_\infty(x)$ is obtained as the (monotone) limit

$$h_\infty(x) = \lim_{n\to\infty} P(T > n | X_0 = x) = \lim_{n\to\infty} Q^n(x, E).$$

Due to monotonicity,

$$\int Q(x, dy)h_\infty(y) \equiv h_\infty(x).$$

Since

$$\int \nu(y)h_\infty(y)dy = 0$$

by (3.4.7), it follows from the definition of the kernel Q in (3.2.1) that also

$$\int P(x, dy)h_\infty(y) \equiv h_\infty(x).$$

(In the MC terminology, $h_\infty(x)$ is a *harmonic function*.) In view of (2.1.1) this can be written as

$$\int p(x, y)h_\infty(y)dy \equiv (1 - r(x))h_\infty(x).$$

Suppose that $h_\infty(x_0)$ were > 0 for some $x_0 \in S$. Since $r(x) < 1$ for all $x \in E$ (see Section 2.1), it would then follow that

$$\int p(x_0, y)h_\infty(y)dy > 0.$$

But, due to the basic Hypothesis 2(ii),

$$\int_S p(x_0, y)h_\infty(y)dy = \int p(x_0, y)h_\infty(y)dy$$

which was > 0. This implies that

$$\int_S h_\infty(y)dy > 0.$$

However, this is impossible by (3.6.2).

*Remark.** It follows that the MC (X_n) is *Harris recurrent* (on the state space S) with *recurrence measure*

$$\phi(A) := |A \cap J| := \int_{A\cap J} dx, \quad A \subset E,$$

cf. [MT, Section 9.1], [N, Section 3.6]. Namely, using the inequality (2.2.3) and an elementary "argument of geometric trials", it follows that the MC (X_n) visits every ϕ-positive set $A \subset E$ infinitely often with probability 1, from an arbitrary initial state $X_0 = x \in S$. Thus *in the present set-up Harris recurrence follows as a corollary from the basic Hypotheses 1 and 2*. This is also noted in Tierney (1994, Corollary 2 of Theorem 2).

3.7 The Proof of the LLN

After these preliminaries the actual proof of the LLN will be short and simple. *The following statements hold true for an arbitrary fixed initial state $X_0 = x$ belonging to the support S and for an arbitrary fixed π-integrable function $f(x), x \in E$.*

We use the shorthand notation ζ_i for the random sums $\zeta_i(f)$ defined in (3.2.7–8). Since the r.v.'s $\zeta_0, \zeta_1, \ldots$ are independent and the r.v.'s $\zeta_1, \zeta_2, \ldots$ are identically distributed (obeying the same

probability law as the r.v. ζ_0 under the initial p.d.f. ν), it follows from the ordinary LLN for i.i.d. random variables and from the finiteness of the r.v. ζ_0 (see 3.6.1) that

$$\lim_{i\to\infty} i^{-1}\sum_{j=0}^{i}\zeta_j = E\zeta_1 = E_\nu\zeta_0 = \int f(x)\mu(x)dx \tag{3.7.1}$$
$$= M\pi(f) \text{ with probability 1, see (3.3.7) and (3.5.1).}$$

Similarly,

$$\lim_{i\to\infty} i^{-1}T_i = E(T_2 - T_1) = E_\nu T \tag{3.7.2}$$
$$= M \text{ with probability 1, see (3.4.6).}$$

Let us denote by $N(n)$, $n = 1, 2, \ldots$, the random number of regeneration epochs T_i up to time n, so that

$$T_{N(n)} \le n < T_{N(n)+1}.$$

Since $N(n) \to \infty$ with probability 1 by (3.6.1), it follows from (3.7.2) that

$$\lim_{n\to\infty} n^{-1}N(n) = \lim_{n\to\infty} (T_{N(n)})^{-1}N(n) \tag{3.7.3}$$
$$= M^{-1} \text{ with probability 1.}$$

We can decompose the sum $S_n := S_n(f) = \sum_{m=0}^{n-1} f(X_m)$ as

$$S_n = \sum_{j=0}^{N(n)-1}\zeta_j + \zeta'_{N(n)},$$

where

$$\zeta'_{N(n)} := \sum_{m=T_{N(n)}}^{n-1} f(X_m) \text{ if } T_{N(n)} \le n-1, \text{ and} \tag{3.7.4}$$
$$\zeta'_{N(n)} := 0 \text{ if } T_{N(n)} = n.$$

Since

$$|\zeta'_{N(n)}| \le \sum_{m=T_{N(n)}}^{T_{N(n)+1}-1} |f(X_m)|,$$

where the random variable on the right hand side has the same probability distribution as the (with probability 1) finite random variable $\sum_{m=0}^{T-1}|f(X_m)|$ (under the initial p.d.f. ν), we can deduce that

$$\lim_{n\to\infty} n^{-1}\zeta'_{N(n)} = 0 \text{ with probability 1.} \tag{3.7.5}$$

By combining (3.7.1–5), we see finally that

$$\lim_{n\to\infty} n^{-1}S_n = \pi(f)$$

as required.

4 The Central Limit Theorem

4.1 *Formulation of the CLT*

A third hypothesis will be needed for the CLT. It has the nature of a *stability condition*, expressing *ergodicity of degree* of the MC (X_n), see Sections 4.2 and 4.3.

We shall call a set $K \subset E$ having volume $|K| > 0$ *weakly small*, if it can be covered with a finite collection $\{I_k\}$ of small sets. Thus K is weakly small, provided that there exists a constant $\beta > 0$ and a finite collection of subsets $I_k, J_k \subset E$, having strictly positive volumes $|I_k|, |J_k| > 0$ such that the sets I_k cover K and the inequalitiy

$$p(x, y) \geq \beta, \text{ for } x \in I_k, \ y \in J_k, \tag{4.1.1}$$

holds true for all $k = 1, \ldots, d$, cf. (2.2.1).

Note that, trivially, a small set is weakly small.

Remark. The following mild topological conditions clearly imply the weak smallness of a set K:

(i) K is compact, and

(ii) there is a continuous function $q(x, y) \leq p(x, y)$ satisfying

$$\int_E q(x, y)dy > 0 \text{ for all } x \in K.$$

(In the literature $q(x, y)$ is referred to as a *continuous component*, cf. [MT, Chapter 6].)

HYPOTHESIS 3. There exists a weakly small set $K \subset E$, a finite non-negative π-integrable function $h(x) \geq 0$, and constants $C < \infty$ and $\delta > 0$ such that

$$E(h(X_1)|X_0 = x) \leq C \text{ for all } x \in K, \text{ and} \tag{4.1.2}$$

$$E(h(X_1) - h(X_0)|X_0 = x) \leq -\delta \text{ for all } x \in K^c. \tag{4.1.3}$$

The following "drift condition" represents the special case where $h(x) \equiv x$:

Example. Suppose that

(i) the invariant p.d.f. π has finite first moment:

$$\int |x|\pi(x)dx < \infty,$$

(ii) the MC has a "drift from infinity"; namely, there exist a weakly small set K and a positive constant $\delta > 0$ such that

$$E(X_1|X_0 = x) \leq x - \delta \text{ for all } x \in K^c, \text{ and}$$

(iii) the jumps out from K have bounded expectations, i.e., there exists a finite constant $C < \infty$ such that

$$E(X_1|X_0 = x) \leq C \text{ for all } x \in K.$$

Let $f(x), x \in E$, be an arbitrary bounded function. According to the CLT the normalized error between the empirical mean $\hat{\pi}_n(f)$ and the "true mean" $\pi(f)$ is asymptotically normally distributed

with variance

$$\sigma_f^2 := \int (f(x) - \pi(f))^2 \pi(x) dx + 2 \sum_{n=1}^{\infty} \int \int \pi(x)(f(x) - \pi(f)) Q^n(x, dy)(f(y) - \pi(f)) dx. \tag{4.1.4}$$

THEOREM 2. *Assume the Hypotheses 1, 2 and 3. Then*

$$\lim_{n \to \infty} n^{\frac{1}{2}} (\hat{\pi}_n(f) - \pi(f)) = \mathcal{N}(0, \sigma_f^2) \text{ in distribution,}$$

for all bounded functions f and all initial states $X_0 = x \in S$.

Remarks. (i) We obtain the CLT for the numbers of visits to subsets $A \subset E$ (formula 1.2 in the Introduction) as the special case where $f(x) := \delta_x(A)$. The associated variance is

$$\sigma_A^2 := \int (\delta_x(A) - \pi(A))^2 \pi(x) dx + 2 \sum_{n=1}^{\infty} \int \int \pi(x)(\delta_x(A) - \pi(A)) Q^n(x, dy)(\delta_y(A) - \pi(A)) dx.$$

(ii) The function h appearing in the stability criterion for ergodicity of degree 2 is referred to as the *Lyapunov–Foster function*. Note that, in order to check its integrability w.r.t. the invariant p.d.f. π, it is sufficient to know π only up to a multiplicative constant (as is the case in MCMC practice).

In the MCMC literature the hypothesis of ergodicity of degree 2 is sometimes replaced by stronger hypotheses, notably by *geometric ergodicity* (associated with remarks to the difficulty of the practical verification of ergodicity of degree 2, cf. e.g. Tierney (1994, Section 3.2). The present approach to the CLT, which is based on the use of the concept of the ergodicity of degree 2, is influenced by the more abstract treatment in [N, Chapter 7]. For general expositions of stability criteria for the CLT see [MT, Section 17.5], Robert & Casella (1999, Sections 4.7.2 and 4.10).

The assumption of *reversibility* leads to a simplification of the variance formula, see Robert & Casella (1999, Section 4.7.2).

4.2 *Ergodicity of Degree 2*

In the context of MC's the *finiteness of the variance in the CLT turns out to be equivalent to the MC being ergodic of degree 2*. The MC (X_n) is called ergodic of degree 2 if, conditionally on the initial p.d.f. ν, the second moment of the first regeneration epoch is finite:

$$E_\nu T^2 < \infty. \tag{4.2.1}$$

Ergodicity of degree 2 can be characterized with the aid of the *potential function of degree 2* defined by

$$\mu^{(2)}(x) := \sum_{n=0}^{\infty} \pi Q^n(x). \tag{4.2.2}$$

Namely, in view of (3.3.4) and (3.5.1), the second order potential function can be written as

$$\mu^{(2)}(x) = M^{-1} \sum_{n=0}^{\infty} (\sum_{m=0}^{\infty} \nu Q^m) Q^n(x) = M^{-1} \sum_{n=0}^{\infty} (n+1) \nu Q^n(x),$$

and therefore, using (3.3.1) and (3.4.6),

$$\begin{aligned} M \int \mu^{(2)}(x)dx &= \sum_{n=0}^{\infty}(n+1) \int \nu Q^n(x)dx \\ &= \sum_{n=0}^{\infty}(n+1)P_\nu(T > n) \\ &= \frac{1}{2}E_\nu(T^2 + T) \\ &= \frac{1}{2}(E_\nu T^2 + M). \end{aligned}$$

Thus we obtain the following formula for the second moment:

$$E_\nu T^2 = M(2 \int \mu^{(2)}(x)dx - 1). \tag{4.2.3}$$

So we see that *ergodicity of degree 2 is also equivalent to the potential function of degree 2 being integrable*. Note that, in view of (3.3.1), the *integral of the potential function is equal to the expectation of the first regeneration time T with the invariant p.d.f. π as the initial p.d.f.*:

$$\int \mu^{(2)}(x)dx = \sum_{n=0}^{\infty} P_\pi(T > n) = E_\pi T. \tag{4.2.4}$$

Thus *ergodicity of degree 2 is also equivalent to the finiteness of the expectation $E_\pi T$*.

4.3 The Proof of Ergodicity of Degree 2

We shall show here that the stability condition expressed by Hypothesis 3 implies the ergodicity of degree 2 of the Markov chain X_n. To this end, we prove first a lemma concerning the finiteness of the expectation of the hitting times of the set K.

Let $\tau \geq 1$ denote the *time of the first visit by the MC (X_n) to the set K*:

$$\tau := \min\{n \geq 1 : X_n \in K\}.$$

More generally, let $0 \leq \tau_0 \leq \tau_1 \leq \dots$ denote the *successive epochs of the visits by the MC (X_n) to the set K*:

$$\begin{aligned} \tau_0 &:= \min\{n \geq 0 : X_n \in K\}, \\ \tau_i &:= \min\{n > \tau_{i-1} : X_n \in K\}, \text{ for } i = 1, 2, \dots . \end{aligned}$$

LEMMA 1. *Assume the Hypotheses 2 and 3. Then*

$$M := \max_{x \in K} E(\tau | X_0 = x) < \infty; \text{ and} \tag{4.3.1}$$

$$E_\pi \tau_0 < \infty. \tag{4.3.2}$$

Proof Lemma 1. We define the terminating MC $(X_{K;n})$ which is "killed at the set K" (i.e., has lifetime τ_0) as follows. Let again ∂ denote a state not belonging to E (the "cemetery"). We set

$$\begin{aligned} X_{K;n} &= X_n \text{ for } 0 \leq n \leq \tau_0, \\ X_{K;n} &= \partial \text{ for } n \geq \tau_0 + 1. \end{aligned}$$

The transition (sub-)probabilities of the killed MC are

$$P_K(x, A) := P(x, A), \text{ for } x \in K^c, A \subset E,$$
$$P_K(x, A) := 0 \text{ for } x \in K, A \subset E.$$

The hypothesis (4.1.3) can be expressed also in the form

$$h(x) \geq \delta P_K(x, E) + \int P_K(x, dy)h(y).$$

By using induction it is easy to show that in fact

$$h(x) \geq \delta \sum_{k=1}^{n} P_K^k(x, E) + \int P_K^n(x, dy)h(y) \text{ for all } n = 1, 2, \ldots .$$

Letting $n \to \infty$ here leads to the inequality

$$h(x) \geq \delta \sum_{k=1}^{\infty} P_K^k(x, E) = \delta \sum_{k=0}^{\infty} P(\tau_0 \geq k | X_0 = x) = \delta(1 + E(\tau_0 | X_0 = x)). \tag{4.3.3}$$

Further, by using the Markov property we obtain

$$\begin{aligned} E(\tau | X_0 = x) &= \int P(x, dy)(1 + E(\tau_0 | X_0 = y)) \\ &\leq \delta^{-1} \int P(x, dy)h(y) \\ &= \delta^{-1} E(h(X_1) | X_0 = x) \\ &\leq \delta^{-1} C \text{ for all } x \in K, \end{aligned} \tag{4.3.4}$$

by the hypothesis (4.1.2).

Integration of the inequality (4.3.3) with respect to the invariant p.d.f. π gives the inequality

$$E_\pi \tau_0 \leq \delta^{-1} \pi(h) < \infty. \tag{4.3.5}$$

This completes the proof of the lemma.

LEMMA 2. *Suppose that there exists a small set $K = I$ for which the conditions (4.3.1–2) hold true. Then the MC is ergodic of degree 2.*

Proof of Lemma 2. As noted at the end of the previous section, we need to show that the expectation $E_\pi T$ of the first regeneration time T is finite.

To this end, recall that regeneration, viz. the event $Y_n = 1$, is possible only at time epochs n such that $X_{n-1} \in I$, i.e., $n - 1 = \tau_k$ for some $k \geq 0$. Hence in particular, $T = \tau_\kappa + 1$, where κ denotes the (random) index

$$\kappa := \min\{k \geq 0 : Y_{\tau_{k+1}} = 1\}.$$

Since by (3.2.2) we have

$$P(T = n | X_0, \ldots, X_{n-2}; X_{n-1} = x, T > n - 1) \equiv \beta > 0 \text{ whenever } x \in I,$$

independently of $n \geq 1$ and of $X_0, \ldots, X_{n-2}$, it follows that also

$$P(\kappa = k + 1 | \kappa > k) \equiv P(T = n | X_{n-1} = x, \tau_{k+1} = n - 1, T > n - 1) \equiv \beta,$$

independently of $x \in I$ and $n \geq 1$. Thus κ *is Geom(β)-distributed.* Now, in view of the inequalities

(3.2.5) and (4.3.4), we have for all $k \geq 0, n \geq 1, x \in I$:

$$\begin{aligned} E(\tau_{k+1} - \tau_k | X_0, \ldots, X_{n-2}; X_{n-1} = x; \tau_k = n-1; Y_0, \ldots, Y_{n-1}; Y_n = 0) \\ \leq (1 - \beta|J|)^{-1} E(\tau_{k+1} - \tau_k | X_{n-1} = x; \tau_k = n-1) \\ = (1 - \beta|J|)^{-1} E(\tau | X_0 = x) \\ \leq (1 - \beta|J|)^{-1} M. \end{aligned}$$

Since the event $\kappa > k$ implies that $X_{n-1} = x, \tau_k = n-1, Y_n = 0$ for some $n \geq 1, x \in I$, it follows from the previous inequality that

$$E(\tau_{k+1} - \tau_k | \kappa > k) \leq (1 - \beta|J|)^{-1} M$$

as well.

The first regeneration epoch can be written as

$$\begin{aligned} T &= 1 + \sum_{k=0}^{\infty} \tau_k 1_{\{\kappa = k\}} \\ &= 1 + \tau_0 + \sum_{k=0}^{\infty} (\tau_{k+1} - \tau_k) 1_{\{\kappa > k\}}. \end{aligned}$$

(The symbol $1_{\{\kappa=k\}}$ denotes the *indicator of the event* $\kappa = k$, i.e., it equals 1 if $\kappa = k$ and 0 if $\kappa \neq k$.) Consequently, with $\lambda = \pi$ as the initial p.d.f.,

$$\begin{aligned} E_\pi T &= 1 + E_\pi \tau_0 + \sum_{k=0}^{\infty} E_\pi(\tau_{k+1} - \tau_k | \kappa > k) P(\kappa > k) \\ &= 1 + E_\pi \tau_0 + (1 - \beta|J|)^{-1} M \, E\kappa \end{aligned}$$

which, due to the assumptions (4.3.1–2), is finite as required.

After these two lemmas we are finally able to prove that Hypothesis 3 implies the ergodicity of degree 2.

Since the set K is weakly small, it can be covered with a finite number of small sets I_k satisfying 4.1.1. Now an *embedded regeneration structure* can be constructed *simultaneously* for each of the indices k. Namely, due to the inequalities (4.1.1), at each time when the MC visits the set K, a "β-coin can be tossed", cf. Section 3.2. If "head" occurs, then, depending on to which subset I_k the current state of the MC belongs, the next state is distributed according to the corresponding k'th regeneration density

$$\nu_k := \text{ the uniform p.d.f. on the set } J_k.$$

Now, by Lemma 1, the expectation $E_\pi \tau_K$ as well as $\max_{x \in K} E(\tau_K | X_0 = x)$ are finite, where τ_K denotes the epoch of the first visit to the set K. It follows further, by using again in an obvious way an "argument of geometric trials", that there exists an index k such that the expectation $E_\pi \tau_{I_k}$ as well as $\sup_{x \in I_k} E(\tau_{I_k} | X_0 = x)$ are finite, where τ_{I_k} denotes the epoch of the first visit to the subset I_k. In view of Lemma 2, this implies the required ergodicity of degree 2.

4.4 *The Proof of the CLT*

The proof of the CLT imitates the proof of the CLT for countable MC's (see e.g. Chung (1960)). Due to a simpler minorization structure the following proof will be considerably simpler than the proof for general Harris recurrent MC's (as presented e.g. in [N, Chapter 7]).

The following calculation holds true for an arbitrarily fixed initial state $X_0 = x \in S$. We shall also fix a bounded function $f(x)$, $x \in E$, satisfying $\pi(f) = 0$. (Otherwise the function $f(x)$ ought to be centered at $\pi(f)$, considering then the function $f(x) - \pi(f)$.) We adopt again the short hand notation ζ_i for the random sums $\zeta_i = \zeta_i(f)$ defined in (3.2.7–8). We denote $M := E_\nu T$.

First notice that, due to the fact that $n^{-1}\zeta_0 \to 0$, $n^{-1}\zeta'_{N(n)} \to 0$ and $n^{-1}N(n) \to M^{-1}$ with probability 1 (see Section 3.7), we have only to prove that

$$n^{-\frac{1}{2}} \sum_{j=1}^{N(n)} \zeta_j \to \mathcal{N}(0, M\sigma_f^2) \text{ in distribution.} \tag{4.4.1}$$

To this end, let $\varepsilon, \eta > 0$ be arbitrary fixed numbers. Let us write $a := M^{-1}$ for short. By *Egorov's theorem* there exists an event Λ having probability $P(\Lambda^c) < \eta$ and an index n_η such that

$$|n^{-1}N(n) - a| < \eta \text{ for all } n \geq n_\eta, \text{ conditionally on the event } \Lambda.$$

Thus (conditionally on Λ and for $n \geq n_\eta$) we can approximate as follows:

$$|\sum_{j=1}^{N(n)} \zeta_j - \sum_{j=1}^{[n(a-\eta)]} \zeta_j| \leq \max_{n(a-\eta)<j<n(a-\eta)} |\sum_{n(a-\eta)<i\leq j} \zeta_i|. \tag{4.4.2}$$

By using the CLT for i.i.d. r.v.'s and recalling that each ζ_i, $i = 1, 2, \ldots$, has the same distribution as ζ_0 under the initial p.d.f. ν, we obtain

$$n^{-\frac{1}{2}} \sum_{j=1}^{[n(a-\eta)]} \zeta_j \to \mathcal{N}(0, E_\nu \zeta_0^2) \text{ in distribution.}$$

Notice that, due to ergodicity of degree 2, the variance $E_\nu \zeta_0^2$ is finite. Namely,

$$E_\nu \zeta_0^2 \leq E_\nu T^2 \sup_{x\in E} |f(x)| < \infty.$$

By using *Kolmogorov's inequality* for sums of i.i.d. r.v.'s the right hand side of (4.4.2) can be bounded as follows:

$$P(\max_{n(a-\eta)<j<n(a-\eta)} |\sum_{n(a-\eta)<i\leq j} \zeta_i| > n^{\frac{1}{2}}\varepsilon) \leq 2\eta\varepsilon^{-2} E_\nu \zeta_0^2,$$

and therefore

$$\begin{aligned} P(n^{-\frac{1}{2}}|\sum_{j=1}^{N(n)} \zeta_j - \sum_{j=1}^{[n(a-\eta)]} \zeta_j| > \varepsilon) &\leq P(\Lambda^c) + P(n^{-\frac{1}{2}}|\sum_{j=1}^{N(n)} \zeta_j - \sum_{j=1}^{[n(a-\eta)]} \zeta_j| > \varepsilon;\ \Lambda) \\ &\leq P(\Lambda^c) + P(\max_{n(a-\eta)<j<n(a-\eta)} |\sum_{n(a-\eta)<i\leq j} \zeta_i| > n^{\frac{1}{2}}\varepsilon). \\ &\leq \eta + 2\eta\varepsilon^{-2} E_\nu \zeta_0^2. \end{aligned}$$

The last expression can be made $< \varepsilon$ by choosing η small enough. Thus

$$\lim_{n\to\infty} n^{-\frac{1}{2}}|\sum_{j=1}^{N(n)} \zeta_j - \sum_{j=1}^{[n(a-\eta)]} \zeta_j| = 0 \text{ in probability,}$$

and therefore

$$\lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{j=1}^{N(n)} \zeta_j = \lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{j=1}^{[n(a-\eta)]} \zeta_j = \mathcal{N}(0, E_\nu \zeta_0^2) \text{ in distribution.}$$

It remains to identify the variance, for which we recall the formulas (3.2.2), (3.2.3), (3.3.1), (3.3.7) and (3.5.1). We obtain

$$\begin{aligned}
E_\nu \zeta_0^2 &= E_\nu \Big(\sum_{m=0}^{T-1} f(X_m)\Big)^2 \\
&= E_\nu \sum_{m=0}^{T-1} f^2(X_m) + 2E_\nu \sum_{k=0}^{T-1} f(X_k) \sum_{m=1}^{T-1-k} f(X_{k+m}) \\
&= E_\nu \sum_{m=0}^{T-1} f^2(X_m) \\
&\quad + 2 \int \sum_{k=0}^{\infty} E_\nu[f(X_k); X_k \in dx, T > k] \sum_{m=1}^{\infty} E[f(X_m); T > m | X_0 = x, Y_0 = 0] \\
&= M \int f^2(x)\pi(x)dx + 2M \sum_{m=1}^{\infty} \int \int \pi(x) f(x) Q^m(x, dy) f(y) dx.
\end{aligned}$$

This completes the proof of the CLT.

4.5 *. *The Multivariate CLT*

We shall here describe the multivariate extension of the CLT. To this end we define first the following covariance:

For any two bounded functions $f(x)$ and $g(x)$ we set

$$\begin{aligned}
\Sigma(f, g) := &\int (f(x) - \pi(f))(g(x) - \pi(g))\pi(x)dx \\
&+ \sum_{n=1}^{\infty} \int \int \pi(x)(f(x) - \pi(f)) Q^n(x, dy)(g(y) - \pi(g))dx \\
&+ \sum_{n=1}^{\infty} \int \int \pi(x)(g(x) - \pi(g)) Q^n(x, dy)(f(y) - \pi(f))dx.
\end{aligned} \tag{4.5.1}$$

Note that the variance σ_f^2 is obtained as

$$\sigma_f^2 = \Sigma(f, f). \tag{4.5.2}$$

For any R^d-valued function $\mathbf{f}(x) = (f_1(x), \ldots, f_d(x)), x \in E$, we denote

$$\begin{aligned}
\pi(\mathbf{f}) &:= (\pi(f_1), \ldots, \pi(f_d)), \text{ and} \\
\hat{\pi}_n(\mathbf{f}) &:= (\hat{\pi}_n(f_1), \ldots, \hat{\pi}_n(f_d)).
\end{aligned}$$

The $d \times d$-matrix

$$\sigma_{\mathbf{f}}^2 := (\Sigma(f_i, f_j);\ i, j = 1, \ldots, d) \tag{4.5.3}$$

will be the covariance matrix in the multivariate CLT. The proof of Theorem 3 is an exact analog of the proof of the scalar case (Theorem 2) and is therefore omitted.

THEOREM 3. *Assume the Hypotheses 1, 2 and 3. Then*

$$\lim_{n \to \infty} n^{\frac{1}{2}} (\hat{\pi}_n(\mathbf{f}) - \pi(\mathbf{f}) = \mathcal{N}(0, \sigma_{\mathbf{f}}^2) \text{ in distribution,}$$

for all bounded functions $\mathbf{f} : E \to R^d$ *and all initial states* $X_0 = x \in S$.

Remark. Let $\mathbf{A} = (A_1, \ldots, A_d)$ denote a *finite partition* of the state space E and let

$$\pi(\mathbf{A}) := (\pi(A_1), \ldots, \pi(A_d)),$$
$$\hat{\pi}_n(\mathbf{A}) := (\hat{\pi}_n(A_1), \ldots, \hat{\pi}_n(A_d)), \text{ and}$$
$$\sigma_{\mathbf{A}}^2 := (\Sigma(A_i, A_j);\ i, j = 1, \ldots, d),$$

where

$$\begin{aligned}\Sigma(A_i, A_j) := &\int (\delta_x(A_i) - \pi(A_i))(\delta_x(A_j) - \pi(A_j))\pi(x)dx \\ &+ \sum_{n=1}^{\infty} \int\int \pi(x)(\delta_x(A_i) - \pi(A_i))Q^n(x, dy)(\delta_y(A_j) - \pi(A_j))dx \\ &+ \sum_{n=1}^{\infty} \int\int \pi(x)(\delta_x(A_j) - \pi(A_j))Q^n(x, dy)(\delta_y(A_i) - \pi(A_i))dx.\end{aligned} \tag{4.5.4}$$

Then

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\hat{\pi}_n(\mathbf{A}) - \pi(\mathbf{A}) = \mathcal{N}(0, \sigma_{\mathbf{A}}^2) \text{ in distribution.}$$

4.6 *. *Consistent Estimation of the Variance*

As an auxiliary study we shall investigate the estimation of the variance σ_f^2, and more generally, of the covariances $\Sigma(f, g)$. To this end we define first the following symmetric measure $\pi^{(2)}$ on the product space $E \times E$. Let $\zeta_0(A)$ denote the random number of visits by the MC (X_n) during the first block to the subset $A \subset E$:

$$\zeta_0(A) := \sum_{n=0}^{T-1} \delta_{X_n}(A).$$

For any two subsets $A, B \subset E$, we set

$$\begin{aligned}\pi^{(2)}(A \times B) :=& M^{-1}E_\nu(\zeta_0(A)\zeta_0(B)) \\ =& M^{-1}E_\nu \sum_{m=0}^{T-1} \delta_{X_m}(A)\delta_{X_m}(B) + M^{-1}E_\nu \sum_{m=1}^{T-1}\sum_{k=1}^{m} \delta_{X_{m-k}}(A)\delta_{X_m}(B) \\ &+ M^{-1}E_\nu \sum_{m=1}^{T-1}\sum_{k=1}^{m} \delta_{X_{m-k}}(B)\delta_{X_m}(A) \\ =& \int_{A\cap B} \pi(x)dx + \sum_{m=1}^{\infty} \int_A \int_B \pi(x)Q^m(x, B)dx + \sum_{m=1}^{\infty} \int_B \int_A \pi(x)Q^m(x, A)dx,\end{aligned}$$

cf. the calculation of the variance $E_\nu\zeta_0^2$ in (4.4.3). Note that, in particular,

$$\pi^{(2)}(E \times E) = M^{-1}E_\nu T^2,$$

whence, due to the ergodicity of degree 2, $\pi^{(2)}$ *is a finite measure.* Also note that, in view of its definition, π is symmetric:

$$\pi(A \times B) = \pi(B \times A).$$

More generally, for any two bounded functions f and g we define

$$\begin{aligned}\pi^{(2)}(f,g) &:= M^{-1}E_\nu(\zeta_0(f)\zeta_0(g))\\ &= M^{-1}E_\nu\sum_{m=0}^{T-1} f(X_m)g(X_m) + M^{-1}E_\nu\sum_{m=1}^{T-1}\sum_{k=1}^{m} f(X_{m-k})g(X_m)\\ &\quad + M^{-1}E_\nu\sum_{m=1}^{T-1}\sum_{k=1}^{m} g(X_{m-k})f(X_m)\\ &= \int f^2(x)\pi(x)dx + \sum_{m=1}^{\infty}\int\int \pi(x)f(x)Q^m(x,dy)g(y)dx\\ &\quad + \sum_{m=1}^{\infty}\int\int \pi(x)g(x)Q^m(x,dy)f(y)dx.\end{aligned}$$

Clearly,

$$\pi^{(2)}(f,g) = \int\int f(x)g(y)\pi^{(2)}(dx\times dy).$$

The variance and covariance are obtained as

$$\begin{aligned}\sigma_f^2 &= \pi^{(2)}(f-\pi(f), f-\pi(f))\\ &= \pi^{(2)}(f,f) - 2\pi(f)\pi^{(2)}(1,f) + \pi(f)^2\pi^{(2)}(1,1)\end{aligned}$$

and

$$\begin{aligned}\Sigma(f,g) &= \pi^{(2)}(f-\pi(f), g-\pi(g))\\ &= \pi^{(2)}(f,g) - \pi(f)\pi^{(2)}(1,g) - \pi(g)\pi^{(2)}(1,f) + \pi(f)\pi(g)\pi^{(2)}(1,1),\end{aligned}$$

respectively.

The following symmetric *random measures* $\hat\pi_n^{(2)}$ turn out to provide an estimator for the measure $\pi^{(2)}$ introduced above. We define:

$$\begin{aligned}\hat\pi_n^{(2)}(A\times B) &:= n^{-1}\sum_{m=0}^{n-1}\delta_{X_m}(A\cap B) + n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m}\delta_{X_{m-k}}(A)\delta_{X_m}(B)\\ &\quad + n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m}\delta_{X_{m-k}}(B)\delta_{X_m}(A),\end{aligned}\tag{4.6.2}$$

where $L_m := \min\{k\ge 0 : Y_{m-k} = 1\}$ for $m \ge T$ (as before, see Section 3.4), and $L_m := 0$ for $m < T$. (By convention we interpret $\sum_{k=1}^{0} = 0$.) The random measure $\hat\pi_n^{(2)}$ can be written as a (random) linear combination of (random) δ-measures on the product space $E\times E$:

$$\begin{aligned}\hat\pi_n^{(2)} &:= n^{-1}\sum_{m=0}^{n-1}\delta_{(X_m,X_m)} + n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m}\delta_{(X_{m-k},X_m)}\\ &\quad + n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m}\delta_{(X_m,X_{m-k})}.\end{aligned}\tag{4.6.3}$$

For any two functions f and g we define

$$\hat\pi_n^{(2)}(f,g) := n^{-1}\sum_{m=0}^{n-1} f(X_m)g(X_m) + n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m} f(X_{m-k})g(X_m)$$

$$+n^{-1}\sum_{m=0}^{n-1}\sum_{k=1}^{L_m} g(X_{m-k})f(X_m).$$

Clearly

$$\hat{\pi}_n^{(2)}(f,g) = \int\int f(x)g(y)\hat{\pi}_n^{(2)}(dx \times dy).$$

We claim that *the random measures* $\hat{\pi}_n^{(2)}$ *provide a consistent estimator for the measure* $\pi^{(2)}$ in the sense that the integrals $\hat{\pi}_n(f,g)$ converge towards the integral $\pi^{(2)}(f,g)$.

THEOREM 4. *Suppose that the Hypotheses 1, 2 and 3 hold true. Then*

$$\lim_{n\to\infty} n^{-1}\hat{\pi}_n^{(2)}(f,g) = \pi^{(2)}(f,g) \text{ with probability 1,}$$

for all bounded functions f *and* g *and for all initial states* $X_0 = x \in S$.

Proof. As before (see 3.7.4) we define

$$\zeta'_{N(n)}(f) := \sum_{m=T_{N(n)}}^{n-1} f(X_m), \text{ if } T_{N(n)} \le n-1,$$

$$\zeta'_{N(n)}(f) := 0, \text{ if } T_{N(n)} = n.$$

Since

$$\zeta_0(f)\zeta_0(g) = \sum_{m=0}^{T-1} f(X_m)g(X_m) + \sum_{m=0}^{T-1}\sum_{k=1}^{m} f(X_{m-k})g(X_m) + \sum_{m=0}^{T-1}\sum_{k=1}^{m} g(X_{m-k})f(X_m),$$

$$\zeta_i(f)\zeta_i(g) = \sum_{m=T_i}^{T_{i+1}-1} f(X_m)g(X_m) + \sum_{m=T_i}^{T_{i+1}-1}\sum_{k=1}^{L_m} f(X_{m-k})g(X_m) + \sum_{m=T_i}^{T_{i+1}-1}\sum_{k=1}^{L_m} g(X_{m-k})f(X_m)$$

for $i = 1, 2, \ldots, N(n)-1$, and

$$\zeta'_{N(n)}(f)\zeta'_{N(n)}(g) = \sum_{m=T_{N(n)}}^{n-1} f(X_m)g(X_m) + \sum_{m=T_{N(n)}}^{n-1}\sum_{k=1}^{m-T_{N(n)}} f(X_{m-k})g(X_m)$$

$$+ \sum_{m=T_{N(n)}}^{n-1}\sum_{k=1}^{m-T_{N(n)}} g(X_{m-k})f(X_m),$$

and since $L_m = m$ for $0 \le m \le T_1 - 1$, $L_m = m - T_i$ for $T_i \le m \le T_{i+1} - 1$, $i = 1, \ldots, T_{N(n)-1}$ and $L_m = m - T_{N(n)}$ for $T_{N(n)} \le m \le n-1$, we can decompose the random measure $\hat{\pi}_n^{(2)}$ as follows:

$$\hat{\pi}_n^{(2)}(f,g) = \sum_{j=0}^{N(n)-1} \zeta_j(f)\zeta_j(g) + \zeta'_{N(n)}(f)\zeta'_{N(n)}(g).$$

Now, by imitating the proof of the LLN (see Section 3.7) we can deduce that

$$\lim_{n\to\infty} n^{-1}\hat{\pi}_n^{(2)}(f,g) = M^{-1}E_\nu(\zeta_0(f)\zeta_0(g)) = \pi^{(2)}(f,g)$$

as required. So the proof of Theorem 4 is completed.

Let us finally note that Theorem 4 yields an estimator for the variance σ_f^2 occurring in the CLT.

To this end, let us define the *empirical covariances* $\hat{\Sigma}_n(f, g)$ by the formula

$$\begin{aligned}\hat{\Sigma}_n(f, g) &:= \hat{\pi}_n^{(2)}(f - \hat{\pi}_n(f), g - \hat{\pi}_n(g)) \\ &= \hat{\pi}_n^{(2)}(f, g) - \hat{\pi}_n(f)\hat{\pi}_n^{(2)}(1, g) - \hat{\pi}_n(g)\hat{\pi}_n^{(2)}(1, f) + \hat{\pi}_n(f)\hat{\pi}_n(g)\hat{\pi}_n^{(2)}(1, 1).\end{aligned}$$

It follows as a corollary of Theorem 4 that the *empirical covariances* $\hat{\Sigma}_n(f, g)$ *form a consistent estimator for the covariance* $\Sigma(f, g)$, i.e.,

$$\lim_{n\to\infty} \hat{\Sigma}_n(f, g) = \Sigma(f, g) \text{ w.p.1.}$$

Namely,

$$\begin{aligned}\hat{\Sigma}_n(f, g) &= \hat{\pi}_n^{(2)}(f, g) - \hat{\pi}_n(f)\hat{\pi}_n^{(2)}(1, g) - \hat{\pi}_n(g)\hat{\pi}_n^{(2)}(1, f) + \hat{\pi}_n(f)\hat{\pi}_n(g)\hat{\pi}_n^{(2)}(1, 1) \\ &\to \pi^{(2)}(f, g) - \pi(f)\pi^{(2)}(1, g) - \pi(g)\pi^{(2)}(1, f) + \pi(f)\pi(g)\pi^{(2)}(1, 1) \\ &= \Sigma(f, g).\end{aligned}$$

Hence in particular, the *empirical variances*

$$\begin{aligned}\hat{\sigma}_{f;n}^2 &:= \hat{\Sigma}_n(f - \hat{\pi}_n(f), f - \hat{\pi}_n(f)) \\ &= \hat{\pi}_n^{(2)}(f, f) - 2\hat{\pi}_n(f)\hat{\pi}_n^{(2)}(1, f) + \hat{\pi}_n(f)^2\hat{\pi}_n^{(2)}(1, 1)\end{aligned}$$

form a consistent estimator for the variance σ_f^2:

$$\lim_{n\to\infty} \hat{\sigma}_{f;n}^2 = \sigma_f^2 \text{ w.p.1.}$$

Note that the computation of the covariance estimators $\hat{\Sigma}_n$ requires the implementation of the regenerative structure in the algorithm. At each time $m = 0, 1, \ldots, n - 1$ the current storage ought to contain the present state X_m as well as the previous states $X_{m-1}, \ldots, X_{m-k}$ till the time $m - k = m - L_m$ of the last regeneration epoch before m. (Before the occurrence of the first regeneration time the whole history $X_m, X_{m-1}, \ldots, X_0$ has to be retained.)

References

Chung, K.L. (1960). *Markov Chains with Stationary Transition Probabilities*. Berlin: Springer.
Meyn, S. & Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. London: Springer.
Nummelin, E. (1984). General Irreducible Markov Chains and Non-negative Operators. *Cambridge Tracts in Mathematics*, **83**. Cambridge: Cambridge University Press.
Revuz, D. (1975). *Markov Chains*. Amsterdam: North-Holland.
Robert, C.P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.
Tierney, L. (1994). Exploring posteriori distributions. *Annals of Statistics*, **22**, 1701–1734.

Résumé

Nous développons un minimum de la théorie des chaînes de Markov, à un niveau d'abstraction aussi bas que possible, afin de démontrer deux lois fondamentales probabilistes pour des *algorithmes de Monte Carlo à chaîne de Markov* standards:

1. La *loi des grands nombres* explique pourquoi l'algorithme fonctionne: elle dit que les moyennes empiriques calculées à partir déchantillons convergent vers leur "vraie" valeur d'ésperance, c'est-à-dire vers leur espérance par rapport à la distribution invariante de la chaîne de Markov associée (= la distribution cible de la simulation).

2. Le *théorème central limite* décrit les déviations des moyennes empiriques de leur espérance en termes de variables aléatoires avec distribution asymptotique normale. Nous présentons également une formule et un estimateur pour la variance associée.

[*Received February, 2001, accepted October, 2001*]