FINITE HORIZON MARKOV DECISION PROBLEMS AND A CENTRAL LIMIT THEOREM FOR TOTAL REWARD

ALESSANDRO ARLOTTO AND J. MICHAEL STEELE

Abstract. We prove a central limit theorem for a class of additive processes that arise naturally in the theory of finite horizon Markov decision problems. The main theorem generalizes a classic result of Dobrushin (1956) for temporally non-homogeneous Markov chains, and the principal innovation is that here the summands are permitted to depend on both the current state and a bounded number of future states of the chain. We show through several examples that this added flexibility gives one a direct path to asymptotic normality of the optimal total reward of finite horizon Markov decision problems. The same examples also explain why such results are not easily obtained by alternative Markovian techniques such as enlargement of the state space.

Mathematics Subject Classification (2010): Primary: 60J05, 90C40; Secondary: 60C05, 60F05, 60G42, 90B05, 90C27, 90C39.

Key Words: sequential decision, non-homogeneous Markov chain, central limit theorem, Markov decision problem, dynamic inventory management, alternating subsequence.

1. Markov Decision Problems and Asymptotic Distributions

In a finite horizon Markov decision problem (MDP) with n periods, it is typical that the decision policy π_n^* that maximizes total expected reward will take actions that depend on both the current state of the system and on the number of periods that remain within the horizon. The total reward $R_n(\pi_n^*)$ that is obtained when one follows the mean-optimal policy π_n^* will have the expected value that optimality requires, but the actual reward $R_n(\pi_n^*)$ that is realized may — or may not — behave in a way that is well summarized by its expected value alone.

As a consequence, a well-founded judgement about the economic value of the policy π_n^* will typically require a deeper understanding of the random variable $R_n(\pi_n^*)$. One gets meaningful benefit from the knowledge of the variance of $R_n(\pi_n^*)$ or its higher moments (Arlotto, Gans and Steele, 2014), but, in the most favorable instance, one would hope to know the distribution of $R_n(\pi_n^*)$, or at least an asymptotic approximation to that distribution.

Limit theorems for the total reward (or the total cost) of an MDP have been studied extensively, but earlier work has focused almost exclusively on those problems where the optimal decision policy is stationary. The first steps were taken by Mandl (1973; 1974*a*; 1974*b*) in the context of finite state space MDPs. This work

Date: first version: May 4, 2015.

A. Arlotto: The Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC, 27708. Email address: alessandro.arlotto@duke.edu .

J. M. Steele: Department of Statistics, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA, 19104. Email address: steele@wharton.upenn.edu .

was subsequently refined and extended to more general MDPs by Mandl (1985), Mandl and Laušmanová (1991), Mendoza-Pérez (2008), and Mendoza-Pérez and Hernández-Lerma (2010). Through these investigations one now has a substantial limit theory for a rich class of MDPs that includes infinite-horizon MDPs with discounting and infinite horizon MDPs where one seeks to maximize the long-run average reward.

Here the focus is on finite horizon MDPs and, to deal with such problems, one needs to break from the framework of stationary decision policies. Moreover, for the purpose of the intended applications, it is useful to consider additive functionals that are more complex than those that have been considered earlier in the theory of temporally non-homogeneous Markov chains. These functionals are defined in the next subsection where we also give the statement of our main theorem.

A Class of MDP Linked Processes

In the theory of discrete-time finite horizon MDPs, one commonly studies a sequence of problems with increasing sizes. Here, it will be convenient to consider two parameters, m and n. The parameter m is fixed, and it will be determined by the nature of the actions and rewards of the MDP. The parameter n measures the size of the MDP; it is essentially the traditional horizon size, but it comes with a small twist.

Now, for a given m and n, we consider an arbitrary sequence of random variables $\{X_{n,i} : 1 \le i \le n+m\}$ with values in a Borel space $\mathcal{X}$, and we also consider an array of n real valued functions of $1+m$ variables,

$$f_{n,i} : \mathcal{X}^{1+m} \to \mathbb{R}, \quad 1 \le i \le n.$$

Further properties will soon be required for both the random variables and the array of functions, but, for the moment, we only note that the random variable of most importance to us here is the sum

$$S_n = \sum_{i=1}^{n} Z_{n,i} \quad \text{where} \quad Z_{n,i} = f_{n,i}(X_{n,i}, \dots, X_{n,i+m}). \tag{1}$$

In a typical MDP application, the random variable $Z_{n,i}$ has an interpretation as a reward for an action taken in period $i \in \{1, 2, \dots, n\}$. The size parameter n is then the number of periods in which decisions are made, and S_n is the total reward received over all periods $i \in \{1, 2, \dots, n\}$ when one follows the policy π_n. Here, of course, the actions chosen by π_n are allowed to depend on both the current *time* and the current *state*.

The parameter m is new to this formulation, and, as we will shortly explain, the flexibility provided by m is precisely what makes sums of the random variables $Z_{n,i} = f_{n,i}(X_{n,i}, \dots, X_{n,i+m})$ useful in the theory of MDPs. In the typical finite horizon setting, the index i corresponds to the decision period, and the realized reward that is associated with period i may depend on many things. In particular, it commonly depends on n, i, the decision period state $X_{n,i}$, and *one or more* values of the post-decision period realizations of the driving sequence $\{X_{n,i} : 1 \le i \le n+m\}$.

Requirements on the Driving Sequence

We always require the driving sequence $\{X_{n,i} : 1 \le i \le n+m\}$ to be a Markov process, but here the Markov kernel for the transition between time i and $i+1$ is allowed to change as i changes. More precisely, we take $\mathcal{B}(\mathcal{X})$ to be the set of Borel subsets of the Borel space $\mathcal{X}$, and we define $\{X_{n,i} : 1 \le i \le n+m\}$ to be the *time non-homogeneous Markov chain* that is determined by specifying a distribution for the initial value $X_{n,1}$ and by making the transition from time i to time $i+1$ in accordance with the Markov transition kernel

$$K^{(n)}_{i,i+1}(x,B) = \mathbb{P}(X_{n,i+1} \in B \mid X_{n,i} = x), \quad \text{where } x \in \mathcal{X} \text{ and } B \in \mathcal{B}(\mathcal{X}).$$

The transition kernels can be quite general, but we do require a condition on their minimal ergodic coefficient. Here we first recall that for any Markov transition kernel $K = K(x, dy)$ on $\mathcal{X}$, the *Dobrushin contraction coefficient* is defined by

$$\delta(K) = \sup_{\substack{x_1,x_2\in\mathcal{X}\\ B\in\mathcal{B}(\mathcal{X})}} |\, K(x_1,B) - K(x_2,B)\,|, \tag{2}$$

and the corresponding *ergodic coefficient* is given by

$$\alpha(K) = 1 - \delta(K).$$

Further, for an array $\{K^{(n)}_{i,i+1} : 1 \le i < n\}$ of Markov transition kernels on $\mathcal{X}$, the *minimal ergodic coefficient* of the n'th row is defined by setting

$$\alpha_n = \min_{1\le i<n} \alpha(K^{(n)}_{i,i+1}). \tag{3}$$

There is also a minor technical point worth noting here. Although we study additive functionals that can depend on the full row $\{X_{n,i} : 1 \le i \le n+m\}$ with $n+m$ elements, the last $1+m$ elements of the row are used in a way that does not require any constraint on the associated ergodic coefficients. Specifically, the last $1+m$ elements of the row are used only to determine value of the time n reward that one receives as a consequence of the last decision. It is for this reason that in expressions like (3) we need only to consider i in the range from 1 to $n-1$.

Main Theorem: A Central Limit Theorem for Reward Processes

When the sums $\{S_n : n \ge 1\}$ defined by (1) are centered and scaled, it is natural to expect that, in favorable circumstances, they will converge in distribution to the standard Gaussian. The next theorem confirms that this is the case provided that one has some modest compatibility between the size of the minimal ergodic coefficient α_n, the size of the functions $f_{n,i}$, $1 \le i \le n$, and the variance of S_n.

Theorem 1 (CLT for Reward Processes)**.** *If there are constants $C_1, C_2, \dots$ such that*

$$\max_{1\le i\le n} \|\, f_{n,i}\,\|_\infty \le C_n \quad \textit{and} \quad C_n^2\alpha_n^{-2} = o(\mathrm{Var}[S_n]), \tag{4}$$

then one has the convergence in distribution

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \Longrightarrow N(0,1), \qquad \textit{as } n \to \infty. \tag{5}$$

Corollary 2. *If there are constants $c > 0$ and $C < \infty$ such that*

$$\alpha_n \ge c \quad \textit{and} \quad C_n \le C \ \textit{ for all } n \ge 1,$$

then one has the asymptotic normality (5) *whenever* $\mathrm{Var}[S_n] \to \infty$ *as* $n \to \infty$.

Organization of the Analysis

Before proving this theorem, it is useful to note how it compares with the classic CLT of Dobrushin (1956) for non-homogeneous Markov chains. If we set $m = 0$ in Theorem 1 then we recover the Dobrushin theorem, so the main issue is to understand how one benefits from the possibility of taking $m \geq 1$. This is addressed in detail in Section 2 and in the examples of Sections 8 and 9.

After recalling some basic facts about the minimal ergodic coefficient in Section 3, the proof begins in earnest in Section 4 where we note that there is a martingale that one can expect to be a good approximation for S_n. The confirmation of the approximation is carried out in Sections 5 and 6. In Section 7 we complete the proof by showing that the assumptions of our theorem also imply that the approximating martingale satisfies the conditions of a basic martingale central limit theorem.

We then take up applications and examples. In particular, we show in Section 8 that Theorem 1 leads to an asymptotic normal law for the optimal total cost of a classic dynamic inventory management problem, and in Section 9 we see how the theorem can be applied to a well-studied problem in combinatorial optimization.

2. On $m = 0$ vs $m > 0$ and Dobrushin's CLT

Dobrushin (1956) introduced many of the concepts that are central to the theory of additive functionals of a non-homogenous Markov chain. In addition to introducing the contraction coefficient (2), Dobrushin also provided one of the earliest — yet most refined — of the CLTs for non-homogenous chains.

Theorem 3 (Dobrushin, 1956)**.** *If there are constants $C_1, C_2, \dots$ such that*

$$\max_{1\leq i\leq n} \| f_{n,i} \|_\infty \leq C_n \quad \text{and} \quad C_n^2 \alpha_n^{-3} = o\biggl(\sum_{i=1}^n \mathrm{Var}[f_{n,i}(X_{n,i})] \biggr), \tag{6}$$

then for $S_n = \sum_{i=1}^n f_{n,i}(X_{n,i})$ one has the asymptotic Gaussian law

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \Longrightarrow N(0,1), \qquad \text{as } n \to \infty.$$

After Dobrushin's work there were refinements and extensions by Sarymsakov (1961), Hanen (1963), and Statuljavičus (1969), but the work that is closest to the approach taken here is that of Sethuraman and Varadhan (2005). They used a martingale approximation to give a streamlined proof of Dobrushin's theorem, and they also used spectral theory to prove the variance lower bound

$$\frac{1}{4}\alpha_n \biggl(\sum_{i=1}^n \mathrm{Var}[f_{n,i}(X_{n,i})] \biggr) \leq \mathrm{Var}[S_n]. \tag{7}$$

This improves a lower bound of Iosifescu and Theodorescu (1969, Theorem 1.2.7) by a factor of two, and Peligrad (2012, Corollary 15) gives some further refinements.

There are also upper bounds for the variance of S_n in terms of the sum of the individual variances and the reciprocal α_n^{-1} of the minimal ergodic coefficient. The most recent of these are given by Szewczak (2012) where they are used in the analysis of continued fraction expansions among other things.

Comparison of Conditions

Theorem 1 requires that $C_n^2 \alpha_n^{-2} = o(\mathrm{Var}[S_n])$ as $n \to \infty$ — a condition that is directly imposed on the variance of the total sum S_n. On the other hand, Dobrushin's theorem imposes the condition (6) on the sum of the variances of the *individual* summands. This difference is not accidental; it actually underscores a notable distinction between the traditional setting where $m = 0$ and the present situation where $m \geq 1$.

When one has $m = 0$, the variance lower bound (7) tells us that condition (6) of Theorem 3 implies condition (4) of Theorem 1, but, when $m \geq 1$, there is not any analog to the lower bound (7). This is the nuance that forces us to impose an explicit condition on the variance of the sum S_n in Theorem 1.

A simple example can be used to illustrate the point. We take $m = 1$ and for each $n \geq 1$ we consider a sequence $X_{n,1}, X_{n,2}, \ldots, X_{n,n+1}$ of independent identically distributed random variables with $0 < \mathrm{Var}[X_{n,1}] < \infty$. The minimal ergodic coefficient in this case is just $\alpha_n = 1$. Next, for $1 \leq i \leq n$ we consider the function

$$f_{n,i}(x, y) = \begin{cases} x & \text{if } i \text{ is even} \\ -y & \text{if } i \text{ is odd;} \end{cases}$$

we then set $S_0 = 0$, and, more generally, we let

$$S_n = \sum_{i=1}^{n} f_{n,i}(X_{n,i}, X_{n,i+1}).$$

Now, for each $n \geq 0$ we see that cancellations in the sum give us $S_{2n} = 0$ and $S_{2n+1} = -X_{2n+1,2(n+1)}$, so, according to parity we find

$$\mathrm{Var}[S_{2n}] = 0 \qquad \text{and} \qquad \mathrm{Var}[S_{2n+1}] = \mathrm{Var}[X_{n,1}].$$

In particular, we have $\mathrm{Var}[S_n] = O(1)$ for all $n \geq 1$, while, on the other hand, for the sum of the individual variances we have that

$$\sum_{i=1}^{n} \mathrm{Var}[f_{n,i}(X_{n,i}, X_{n,i+1})] = n \,\mathrm{Var}[X_{n,1}] = \Omega(n).$$

The bottom line is that when $m \geq 1$, there is no analog of the lower bound (7), and, as a consequence, a result like Theorem 1 needs to impose an explicit condition on $\mathrm{Var}[S_n]$ rather than a condition on the sum of the variances of the individual summands.

Two Related Alternatives

One might hope to prove Theorem 1 by considering an enlarged state space where one could first apply Dobrushin's CLT (Theorem 3) and then extract Theorem 1 as a consequence. For example, given the conditions of Theorem 1 with $m = 1$, one might introduce the bivariate chain $\{\widehat{X}_{n,i} = (X_{n,i}, X_{n,i+1}) : 1 \leq i \leq n\}$ with the hope of extracting the conclusion of Theorem 1 by applying Dobrushin's theorem to $\{\widehat{X}_{n,i} : 1 \leq i \leq n\}$.

The fly in the ointment is that the resulting bivariate chain can be degenerate in the sense that the minimal ergodic coefficient of the chain $\{\widehat{X}_{n,i} : 1 \leq i \leq n\}$ can equal zero. In such a situation, Dobrushin's theorem does not apply to the process $\{\widehat{X}_{n,i} : 1 \leq i \leq n\}$, even though Theorem 1 may still provide a useful central limit theorem. We give two concrete examples of this phenomenon in Sections 8 and 9.

A further way to try to rehabilitate the possibility of using the bivariate chain $\{\widehat{X}_{n,i} : 1 \le i \le n\}$ is to appeal to theorems where the minimal ergodic coefficient α_n is replaced with some less fragile quantity. For example, Peligrad (2012) has proved that one can replace α_n in Dobrushin's theorem with the maximal coefficient of correlation ρ_n. Since one always has $\rho_n \le \sqrt{1-\alpha_n}$, Peligrad's CLT is guaranteed to apply at least as widely as Dobrushin's CLT. Nevertheless, the examples of Sections 8 and 9 both show that this refinement still does not help.

3. On Contractions and Oscillations

To prove Theorem 1, we need to assemble a few properties of the Dobrushin contraction coefficient. Much more can be found in Seneta (2006, Section 4.3), Winkler (2003, Section 4.2), or Del Moral (2004, Chapter 4).

If μ and ν are two probability measures, we write $\| \mu - \nu \|_{\mathrm{TV}}$ for the total variation distance between μ and ν. Dobrushin's coefficient (2) can then be written as

$$\delta(K) = \sup_{x_1, x_2 \in \mathcal{X}} \| K(x_1, \cdot) - K(x_2, \cdot) \|_{\mathrm{TV}},$$

and one always has $0 \le \delta(K) \le 1$. For any two Markov kernels K_1 and K_2 on $\mathcal{X}$, we also set

$$(K_1 K_2)(x, B) = \int K_1(x, dz) K_2(z, B),$$

so $(K_1K_2)(x, B)$ represents the probability that one ends up in B given that one starts at x and takes two steps: the first governed by the transition kernel K_1 and the second governed by the kernel K_2. A crucial property of the Dobrushin coefficient δ is that one has the product inequality

$$\delta(K_1 K_2) \le \delta(K_1)\delta(K_2). \tag{8}$$

Now, given any array $\{K^{(n)}_{i,i+1} : 1 \le i < n\}$ of Markov kernels and any pair of times $1 \le i < j \le n$, one can form the multi-step transition kernel

$$K^{(n)}_{i,j}(x, B) = (K^{(n)}_{i,i+1} K^{(n)}_{i+1,i+2} \cdots K^{(n)}_{j-1,j})(x, B),$$

and, as the notation suggests, the kernel $K^{(n)}_{i,i+1}$ can change as i changes. The product inequality (8) and the definition of the minimal ergodic coefficient (3) then tell us

$$\delta(K^{(n)}_{i,j}) \le (1 - \alpha_n)^{j-i} \qquad \text{for all } 1 \le i < j \le n. \tag{9}$$

Dobrushin's coefficient can also be characterized by the action of the Markov kernel on a natural function class. First, for any bounded measurable function $h : \mathcal{X} \to \mathbb{R}$ we note that the operator

$$(Kh)(x) = \int K(x, dz) h(z),$$

is well defined, and one also has that the oscillation of h

$$\mathrm{Osc}(h) = \sup_{z_1, z_2 \in \mathcal{X}} |\, h(z_1) - h(z_2) \,| < \infty.$$

Now, if one sets $\mathcal{H} = \{h : \mathrm{Osc}(h) \le 1\}$, then the Dobrushin contraction coefficient (2) has a second characterization,

$$\delta(K) = \sup_{\substack{x_1, x_2 \in \mathcal{X} \\ h \in \mathcal{H}}} |\, (Kh)(x_1) - (Kh)(x_2) \,|. \tag{10}$$

This tells us in turn that for any Markov transition kernel K on $\mathcal{X}$ and for any bounded measurable function $h : \mathcal{X} \to \mathbb{R}$, one has the oscillation inequality

$$\operatorname{Osc}(Kh) \leq \delta(K) \operatorname{Osc}(h). \tag{11}$$

This bound is especially useful when it is applied to the multi-step kernel given by $K^{(n)}_{i,j} = K^{(n)}_{i,i+1} K^{(n)}_{i+1,i+2} \cdots K^{(n)}_{j-1,j}$. In this case, the oscillation inequality (11) and the upper bound (9) combine to give us

$$\operatorname{Osc}(K^{(n)}_{i,j} h) \leq \delta(K^{(n)}_{i,j}) \operatorname{Osc}(h) \leq (1-\alpha_n)^{j-i} \operatorname{Osc}(h). \tag{12}$$

This basic bound will be used many times in the analysis of Section 5.

4. Connecting a Martingale to S_n

Our proof of Theorem 1 exploits a martingale approximation like the one used by Sethuraman and Varadhan (2005) in their proof of the Dobrushin central limit theorem. Similar plans have been used in many investigations including Gordin (1969), Kipnis and Varadhan (1986), Kifer (1998), Wu and Woodroofe (2004), Gordin and Peligrad (2011), and Peligrad (2012), but prior to Sethuraman and Varadhan (2005) the martingale approximation method seems to have been used only for stationary processes.

Here we only need a basic version of the CLT for an array of martingale difference sequences (MDS) that we frame as a proposition. This version is easily covered by any of the martingale central limit theorems of Brown (1971), McLeish (1974), or Hall and Heyde (1980, Corollary 3.1).

Proposition 4 (Basic CLT for MDS Arrays)**.** *If for each $n \geq 1$, one has a martingale difference sequence $\{\xi_{n,i} : 1 \leq i \leq n\}$ with respect to the filtration $\{\mathcal{G}_{n,i} : 0 \leq i \leq n\}$, and if one also has the negligibility condition*

$$\max_{1\leq i\leq n} \| \, \xi_{n,i} \, \|_\infty \longrightarrow 0 \qquad \text{as } n \to \infty, \tag{13}$$

then the "weak law of large numbers" for the conditional variances

$$\sum_{i=1}^{n} \mathbb{E}[\xi^2_{n,i} \,|\, \mathcal{G}_{n,i-1}] \xrightarrow{p} 1 \qquad \text{as } n \to \infty, \tag{14}$$

implies that one has convergence in distribution to a standard normal,

$$\sum_{i=1}^{n} \xi_{n,i} \Longrightarrow N(0,1) \qquad \text{as } n \to \infty.$$

A Martingale for a Non-Homogenous Chain

We let $\mathcal{F}_{n,0}$ be the trivial σ-field, and we set $\mathcal{F}_{n,i} = \sigma\{X_{n,1}, X_{n,2}, \ldots, X_{n,i}\}$ for $1 \leq i \leq n+m$. Further, we define the *value to-go process* $\{V_{n,i} : m \leq i \leq n+m\}$ by setting $V_{n,n+m} = 0$ and by letting

$$V_{n,i} = \sum_{j=i+1-m}^{n} \mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}], \qquad \text{for } m \leq i < n+m. \tag{15}$$

If we view the random variable $Z_{n,j}$ as a reward that we receive at time j, then the value to-go $V_{n,i}$ at time i is the conditional expectation at time i of the total of the rewards that stand to be collected during the time interval $\{i+1-m,\dots,n\}$. For $1+m \le i \le n+m$ we then let

$$d_{n,i} = V_{n,i} - V_{n,i-1} + Z_{n,i-m}, \tag{16}$$

and one can check directly from the definition that $\{d_{n,i} : 1+m \le i \le n+m\}$ is a martingale difference sequence (MDS) with respect to its natural filtration $\{\mathcal{F}_{n,i} : 1+m \le i \le n+m\}$.

When we sum the terms of (16), the summands $V_{n,i} - V_{n,i-1}$ telescope, and we are left with the basic decomposition

$$S_n = \sum_{i=1}^{n} Z_{n,i} = V_{n,m} + \sum_{i=1+m}^{n+m} d_{n,i}. \tag{17}$$

For the proof of Theorem 1, we assume without loss of generality that $\mathbb{E}[Z_{n,i}] = 0$ for all $1 \le i \le n$. Naturally, in this case we also have $\mathbb{E}[S_n] = \mathbb{E}[V_{n,m}] = 0$ since the sum of the martingale differences in (17) will always have total expectation zero. We now just need to analyze the components of the representation (17).

5. Oscillation Estimates

The first step in the proof of Theorem 1 is to argue that the summand $V_{n,m}$ in (17) makes a contribution to S_n that is asymptotically negligible when compared to the standard deviation of S_n. Once this is done, one can use the martingale CLT to deal with the last sum in (17). Both of these steps depend on oscillation estimates that exploit the multiplicative bound (12) on the Dobrushin contraction coefficient.

For any random variable X one has the trivial bound

$$\mathrm{Osc}(X) = \mathrm{esssup}(X) - \mathrm{essinf}(X) \le 2\| X \|_\infty, \tag{18}$$

together with its partial converse,

$$\| X - \mathbb{E}[X] \|_\infty \le \mathrm{Osc}(X). \tag{19}$$

Moreover for any two σ-fields $\mathcal{I} \subseteq \mathcal{I}'$ of the Borel sets $\mathcal{B}(\mathcal{X})$, the conditional expectation is a contraction for the oscillation semi-norm; that is, one has

$$\mathrm{Osc}(\mathbb{E}[X \,|\, \mathcal{I}]) \le \mathrm{Osc}(\mathbb{E}[X \,|\, \mathcal{I}']) \le \mathrm{Osc}(X). \tag{20}$$

Also, by comparison of $X(\omega)Y(\omega)$ and $X(\omega')Y(\omega')$, one has the product rule

$$\mathrm{Osc}(XY) \le \| X \|_\infty \mathrm{Osc}(Y) + \| Y \|_\infty \mathrm{Osc}(X). \tag{21}$$

In the next two lemmas we assume that there is a constant $C_n < \infty$ such that

$$\| f_{n,i} \|_\infty \le C_n \qquad \text{for all } 1 \le i \le n.$$

Since $Z_{n,i} = f_{n,i}(X_{n,i},\dots,X_{n,i+m})$ and $\mathbb{E}[Z_{n,i}] = 0$, this assumption gives us

$$\| Z_{n,i} \|_\infty \le C_n, \quad \text{and} \quad \mathrm{Osc}(\mathbb{E}[Z_{n,i} \,|\, \mathcal{I}]) \le 2C_n \tag{22}$$

for any σ-field $\mathcal{I} \subseteq \mathcal{B}(\mathcal{X})$.

Oscillation Bounds on Conditional Moments

Lemma 5 (Conditional Moments)**.** *For all $1 \le i < j \le n$ one has*

$$\| \mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}] \|_\infty \le \mathrm{Osc}(\mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}]) \le 2C_n(1-\alpha_n)^{j-i}, \tag{23}$$

and

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j}^2 \,|\, \mathcal{F}_{n,i}]) \le 2C_n^2(1-\alpha_n)^{j-i}. \tag{24}$$

Proof. Since $\mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}]$ has mean zero, the first inequality of (23) is immediate from (19). To get the second inequality, we note by the Markov property that we can define a function h_j on the support of $X_{n,j}$ by setting

$$h_j(X_{n,j}) = \mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,j}],$$

and by (22) we have the bound $\mathrm{Osc}(h_j) \le 2C_n$. For $i < j$ a second use of the Markov property gives us the pullback identity

$$\mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}] = (K_{i,j}^{(n)} h_j)(X_{n,i}),$$

so the bound (12) gives us

$$\mathrm{Osc}(K_{i,j}^{(n)} h_j) \le 2C_n(1-\alpha_n)^{j-i},$$

and this is all we need to complete the proof of (23).

One can prove (24) by essentially the same method, but now we define a map $x \mapsto s_j(x)$ by setting

$$s_j(X_{n,j}) = \mathbb{E}[Z_{n,j}^2 \,|\, \mathcal{F}_{n,j}],$$

so for $i < j$ the pullback identity becomes

$$\mathbb{E}[Z_{n,j}^2 \,|\, \mathcal{F}_{n,i}] = (K_{i,j}^{(n)} s_j)(X_{n,i}).$$

By (20) we have $\mathrm{Osc}(s_j) \le \mathrm{Osc}(Z_{n,j}^2)$, so (22) implies $\mathrm{Osc}(s_j) \le 2C_n^2$, and the inequality (12) then gives us (24). □

Oscillation Bounds on Conditional Cross Moments

The minimal ergodic coefficient α_n can also be used to control the oscillation of the conditional expectations of the products $Z_{n,j}Z_{n,k}$ given $\mathcal{F}_{n,i}$. All of the inequalities that we need tell a similar story, but the specific bounds have an inescapable dependence on the relative values of i, j, k, n, and m. Figure 1 gives a graphical representation of the constraints on the indices that feature in the next lemma.

Lemma 6 (Conditional Cross Moments)**.** *For each $i \in \{m, \dots, n+m\}$ we consider $i - m < j < n$ and $j < k \le n$. We then have the following oscillation bounds that depend on the range of the indices (see also Figure 1):*

Range 1. *If $j \le i$ and $k \le j + m$ then*

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \le 4C_n^2. \tag{25}$$

Range 2. *If $j \le i < j + m < k$ then*

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \le 6C_n^2(1-\alpha_n)^{k-j-m}. \tag{26}$$

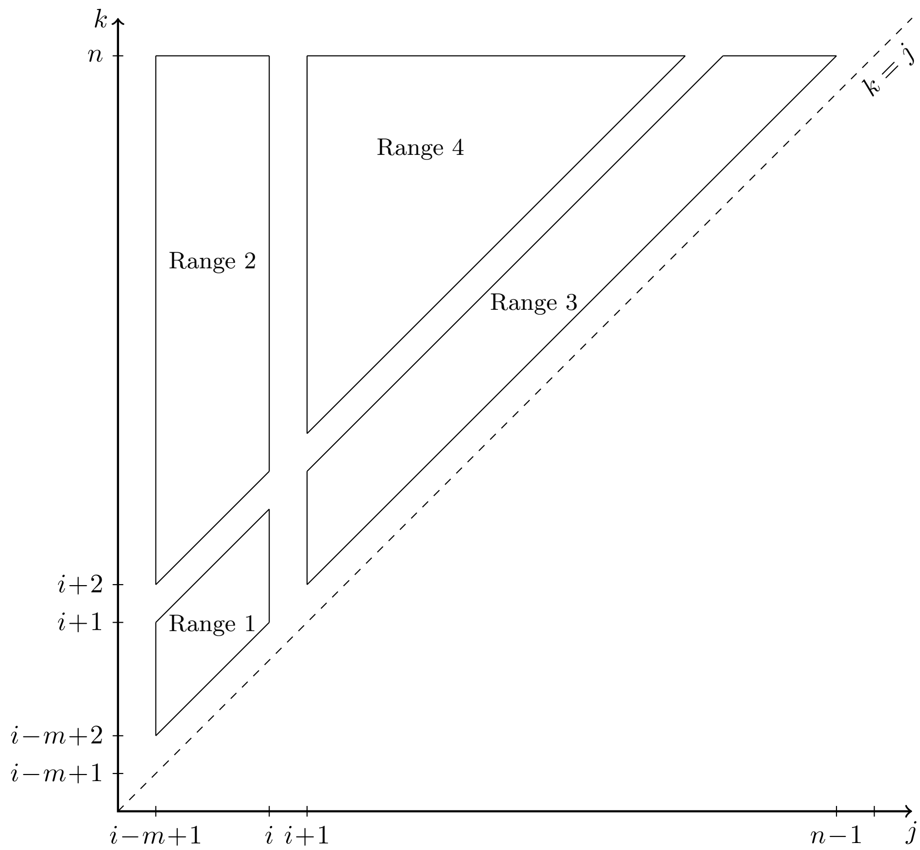
Range 3. *If $i < j < k \le j + m$ then*

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \le 2C_n^2(1-\alpha_n)^{j-i}. \tag{27}$$

Range 4. *If $i < j \le j + m < k$, then*

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \le 6C_n^2(1-\alpha_n)^{k-i-m}. \tag{28}$$

FIGURE 1. Cross Moments Index Relations



The estimates in Lemma 6 require attention to certain ranges of indices. In turn, these amount to a decomposition of the lattice triangle defined by the upper-left half of $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$.

Proof. Inequality (25) follows immediately from the product rule (21) and the bounds (22). To prove (26), we note that for $i < j + m$ we have $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n,j+m}$ so from the monotonicity (20) and the fact that $Z_{n,j}$ is $\mathcal{F}_{n,j+m}$-measurable, we obtain that

$$\operatorname{Osc}(\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \leq \operatorname{Osc}(\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,j+m}]) = \operatorname{Osc}(Z_{n,j} \mathbb{E}[Z_{n,k} \,|\, \mathcal{F}_{n,j+m}]).$$

The product rule (21) applied to the quantity on the right-hand side above gives us the inequality

$$\begin{aligned} &\operatorname{Osc}(\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \\ &\qquad \leq \| Z_{n,j} \|_\infty \operatorname{Osc}(\mathbb{E}[Z_{n,k} \,|\, \mathcal{F}_{n,j+m}]) + \operatorname{Osc}(Z_{n,j}) \| \mathbb{E}[Z_{n,k} \,|\, \mathcal{F}_{n,j+m}] \|_\infty, \end{aligned}$$

so if we recall that $\| Z_{n,i} \|_\infty \leq C_n$ and that $\operatorname{Osc}(Z_{n,j}) \leq 2C_n$ and use the conditional moment bounds in (23) we have

$$\operatorname{Osc}(\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \leq 2C_n^2 (1 - \alpha_n)^{k-j-m} + 4C_n^2 (1 - \alpha_n)^{k-j-m},$$

completing the proof of (26).

To verify inequality (27), we consider the map $X_{n,j} \mapsto p_j(X_{n,j})$ given by

$$p_j(X_{n,j}) = \mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,j}],$$

and we note that for $i < j$ we have the pullback identity

$$\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}] = (K^{(n)}_{i,j} p_j)(X_{n,i}).$$

Since $\| Z_{n,j} \|_\infty$ and $\| Z_{n,k} \|_\infty$ are bounded by C_n, we have $\| p_j \|_\infty \leq C_n^2$ and $\mathrm{Osc}(p_j) \leq 2C_n^2$. We also have $i < j < k$ so (12) tells us that

$$\mathrm{Osc}(K^{(n)}_{i,j} p_j) \leq \delta(K^{(n)}_{i,j}) \,\mathrm{Osc}(p_j) \leq 2C_n^2 (1-\alpha_n)^{j-i},$$

completing the proof of (27).

Finally, for the last inequality (28) we have $j \leq j+m < k$, we consider the map $X_{n,j} \mapsto q_j(X_{n,j})$ defined by setting

$$q_j(X_{n,j}) = \mathbb{E}[Z_{n,j}(\mathbb{E}[Z_{n,k} \,|\, \mathcal{F}_{n,j+m}]) \,|\, \mathcal{F}_{n,j}],$$

and we obtain the identity

$$\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}] = (K^{(n)}_{i,j} q_j)(X_{n,i}).$$

By the multiplicative bound (12), this gives us

$$\mathrm{Osc}(\mathbb{E}[Z_{n,j} Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \leq (1-\alpha_n)^{j-i} \,\mathrm{Osc}(q_j),$$

and we also have $\mathrm{Osc}(q_j) \leq 6C_n^2 (1-\alpha_n)^{k-j-m}$ by (26), so the proof of (28) is also complete. $\square$

6. The Value To-Go Process and MDS L^∞-Bounds

We have everything we need to argue that the variance condition (4) implies the negligibility condition (13). The first step is to get simple L^∞-estimates of the value to-go $V_{n,i}$ that was defined in (15). We then need estimates of the martingale difference$d_{n,i}$ defined in (16). Here, and subsequently, we use $M = M(m)$ to denote a Hardy-style constant which depends only on m and which may change from one line to the next.

Lemma 7 (L^∞-Bounds for the Value To-Go and for the MDS)**.** *There is a constant $M < \infty$ such that for all $n \geq 1$ we have*

$$\| V_{n,i} \|_\infty \leq M C_n \alpha_n^{-1}, \qquad \text{for } m \leq i \leq n+m, \quad \text{and} \tag{29}$$

$$\| d_{n,i} \|_\infty \leq M C_n \alpha_n^{-1}, \qquad \text{for } 1+m \leq i \leq n+m. \tag{30}$$

Proof. We have $\| Z_{n,j} \|_\infty \leq C_n$, and when we use this estimate on the first m summands in the definition (15) of the value to-go $V_{n,i}$ we get the bound

$$\| V_{n,i} \|_\infty \leq m\, C_n + \sum_{j=i+1}^{n} \| \mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}] \|_\infty.$$

From (23) we know that $\| \mathbb{E}[Z_{n,j} \,|\, \mathcal{F}_{n,i}] \|_\infty \leq 2C_n(1-\alpha_n)^{j-i}$ for all $1 \leq i < j \leq n$ so, after completing the geometric series, we have

$$\| V_{n,i} \|_\infty \leq m\, C_n + 2C_n \alpha_n^{-1} \leq M C_n \alpha_n^{-1},$$

where one can take $M = 2m$ as a generous choice for M. This bound, the representation (16), and the triangle inequality then give us (30). $\square$

Conditional Variances L^2-Bounds

Everything is also in place to show that the variance condition (4) gives one the weak law of large numbers for the conditional variances (14). We begin by deriving some basic inequalities for the variance of S_n.

Lemma 8 (Variance Bounds)**.** *For all $n \geq 1$ we have*

$$\mathbb{E}[S_n^2] = \mathbb{E}[V_{n,m}^2] + \sum_{j=1+m}^{n+m} \mathbb{E}[d_{n,j}^2], \quad \textit{and} \tag{31}$$

$$\mathrm{Var}[S_n] - MC_n^2\alpha_n^{-2} \leq \sum_{j=1+m}^{n+m} \mathbb{E}[d_{n,j}^2] \leq \mathrm{Var}[S_n]. \tag{32}$$

Proof. When we square both sides of (17) we have

$$S_n^2 = V_{n,m}^2 + 2V_{n,m}\Big\{ \sum_{j=1+m}^{n+m} d_{n,j} \Big\} + \Big\{ \sum_{j=1+m}^{n+m} d_{n,j} \Big\}^2.$$

Since $V_{n,m}$ is $\mathcal{F}_{n,m}$-measurable, we obtain from the conditional orthogonality of the martingale differences that

$$\mathbb{E}[S_n^2 \,|\, \mathcal{F}_{n,m}] = V_{n,m}^2 + \sum_{j=1+m}^{n+m} \mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,m}],$$

and, when we take the total expectation, we then get (31). Finally, since $\mathbb{E}[S_n] = 0$, the representation (31) and the bound (29) for $\| V_{n,m} \|_\infty$ give us the two inequalities of (32). $\square$

Lemma 9 (Oscillation Bound)**.** *There is a constant $M < \infty$ such that*

$$\mathrm{Osc}\Big(\sum_{j=1+i}^{n+m} \mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,i}]\Big) \leq MC_n^2\alpha_n^{-2} \qquad \textit{for } m \leq i \leq n+m. \tag{33}$$

Proof. If we sum the identity (16) we have

$$\sum_{j=1+i}^{n+m} Z_{n,j-m} = V_{n,i} + \sum_{j=1+i}^{n+m} d_{n,j},$$

so, when we square both sides and use the fact that $V_{n,i}$ is $\mathcal{F}_{n,i}$-measurable, the orthogonality of the martingale differences gives us

$$\mathbb{E}\big[\big\{ \sum_{j=1+i}^{n+m} Z_{n,j-m} \big\}^2 \,|\, \mathcal{F}_{n,i}\big] = V_{n,i}^2 + \sum_{j=i+1}^{n} \mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,i}].$$

The triangle inequality then implies

$$\mathrm{Osc}\,\big(\sum_{j=i+1}^{n+m} \mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,i}]\big) \leq \mathrm{Osc}(V_{n,i}^2) + \mathrm{Osc}\,\Big(\mathbb{E}\big[\big\{ \sum_{j=1+i}^{n+m} Z_{n,j-m} \big\}^2 \,|\, \mathcal{F}_{n,i}\big]\Big). \tag{34}$$

By (29) we have $\| V_{n,i} \|_\infty \leq MC_n\alpha_n^{-1}$ so, by (18), we obtain

$$\mathrm{Osc}(V_{n,i}^2) \leq 2\| V_{n,i}^2 \|_\infty \leq MC_n^2\alpha_n^{-2}. \tag{35}$$

It only remains to estimate the second summand of (34), but this takes some work. Specifically, we will check that one can write

$$\operatorname{Osc}(\mathbb{E}[\{\sum_{j=1+i-m}^{n} Z_{n,j}\}^2|\mathcal{F}_{n,i}]) \leq \mathcal{S}_0 + \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3 + \mathcal{S}_4. \tag{36}$$

where $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$, and $\mathcal{S}_4$ are non-negative sums that one can estimate individually with help from our oscillation bounds. Here the first term $\mathcal{S}_0$ accounts for the oscillation of the conditional squared moments. It is given by

$$\mathcal{S}_0 = \sum_{j=1+i-m}^{i} \operatorname{Osc}(\mathbb{E}[Z_{n,j}^2|\mathcal{F}_{n,i}]) + \sum_{j=1+i}^{n} \operatorname{Osc}(\mathbb{E}[Z_{n,j}^2|\mathcal{F}_{n,i}]),$$

and by (22) and (24) we have the estimate

$$\mathcal{S}_0 \leq 2mC_n^2 + 2C_n^2 \sum_{j=1+i}^{n} (1-\alpha_n)^{j-i} \leq 2(1+m)C_n^2\alpha_n^{-1}.$$

The remaining sums $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and $\mathcal{S}_4$ are given by the oscillation of the conditional cross moments $Z_{n,j}Z_{n,k}$ given $\mathcal{F}_{n,i}$ where the ranges of the indices j and k are given by the corresponding four regions in Figure 1. Specifically, we have

$$\mathcal{S}_1 = 2 \sum_{j=1+i-m}^{i} \sum_{k=1+j}^{j+m} \operatorname{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]),$$

and (25) gives us $\mathcal{S}_1 \leq 8m^2C_n^2$ since $\mathcal{S}_1$ has m^2 summands. Next, if we set

$$\mathcal{S}_2 = 2 \sum_{j=1+i-m}^{i} \sum_{k=1+j+m}^{n} \operatorname{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}])$$

then the oscillation inequality (26) gives us

$$\mathcal{S}_2 \leq 12C_n^2 \sum_{j=1+i-m}^{i} \sum_{k=1+j+m}^{n} (1-\alpha_n)^{k-j-m} \leq 12mC_n^2\alpha_n^{-1}.$$

Similarly, for the third region, the bound (27) gives us

$$\begin{aligned} \mathcal{S}_3 &= 2 \sum_{j=1+i}^{n} \sum_{k=1+j}^{j+m} \operatorname{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \\ &\leq 4C_n^2 \sum_{j=1+i}^{n} \sum_{k=1+j}^{j+m} (1-\alpha_n)^{j-i} \leq 4mC_n^2\alpha_n^{-1}, \end{aligned}$$

and, for the fourth region, the bound (28) implies

$$\begin{aligned} \mathcal{S}_4 &= 2 \sum_{j=1+i}^{n} \sum_{k=1+j+m}^{n} \operatorname{Osc}(\mathbb{E}[Z_{n,j}Z_{n,k} \,|\, \mathcal{F}_{n,i}]) \\ &\leq 12C_n^2 \sum_{j=1+i}^{n} \sum_{k=1+j+m}^{n} (1-\alpha_n)^{k-i-m} \leq 12C_n^2\alpha_n^{-2}. \end{aligned}$$

Finally, by our decomposition (36), the upper bounds for $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$, and $\mathcal{S}_4$ tell us that there is a constant M for which we have

$$\operatorname{Osc}(\mathbb{E}[\{ \sum_{j=1+i-m}^{n} Z_{n,j} \}^2 | \mathcal{F}_{n,i}]) \leq M C_n^2 \alpha_n^{-2},$$

so, given (34) and (35), the proof of the lemma is complete. □

7. Completion of the Proof of Theorem 1

It only remains to argue that if we set

$$\eta_i = \mathbb{E}[d_{n,i}^2 \,|\, \mathcal{F}_{n,i-1}] \quad \text{and} \quad \Delta_n = \sum_{i=1+m}^{n+m} (\eta_i - \mathbb{E}[\eta_i]),$$

then the variance condition (4) implies that $\Delta_n = o(\operatorname{Var}[S_n])$ in probability as $n \to \infty$. We can get this as an easy consequence of the next lemma.

Lemma 10 (L^2-Bound for Δ_n)**.** *There is a constant $M < \infty$ depending only on m such that for all $n \geq 1$ one has the inequality*

$$\mathbb{E}[\Delta_n^2] = \operatorname{Var}\big[\{ \sum_{i=1+m}^{n+m} \mathbb{E}[d_{n,i}^2 \,|\, \mathcal{F}_{n,i-1}]\}\big] \leq M C_n^2 \alpha_n^{-2} \operatorname{Var}[S_n].$$

Proof. By direct expansion we have

$$\mathbb{E}[\Delta_n^2] = \sum_{i=1+m}^{n+m} \operatorname{Var}[\eta_i] + 2 \sum_{i=1+m}^{n+m} \mathbb{E}\big[(\eta_i - \mathbb{E}[\eta_i])\{ \sum_{j=i+1}^{n+m} \big(\eta_j - \mathbb{E}[\eta_j]\big)\}\big], \tag{37}$$

and we estimate the two sums separately. First, by crude bounds and (30) we have

$$\mathbb{E}[\eta_i^2] \leq \| \eta_i \|_\infty \mathbb{E}[\eta_i] \leq \| d_{n,i} \|_\infty^2 \mathbb{E}[\eta_i] \leq M C_n^2 \alpha_n^{-2} \mathbb{E}[\eta_i],$$

so we obtain that the first sum of (37) satisfies the inequality

$$\sum_{i=1+m}^{n+m} \operatorname{Var}[\eta_i] \leq M C_n^2 \alpha_n^{-2} \sum_{i=1+m}^{n+m} \mathbb{E}[\eta_i].$$

The twin bounds of (32) and the definition $\eta_i = \mathbb{E}[d_{n,i}^2 \,|\, \mathcal{F}_{n,i-1}]$ then tell us that

$$\operatorname{Var}[S_n] - M C_n^2 \alpha_n^{-2} \leq \sum_{i=1+m}^{n+m} \mathbb{E}[\eta_i] \leq \operatorname{Var}[S_n], \tag{38}$$

so we also have the upper bound

$$\sum_{i=1+m}^{n+m} \operatorname{Var}[\eta_i] \leq M C_n^2 \alpha_n^{-2} \operatorname{Var}[S_n]. \tag{39}$$

To estimate the second sum of (37), we first note that η_i is $\mathcal{F}_{n,i-1}$-measurable and $\mathcal{F}_{n,i-1} \subseteq \mathcal{F}_{n,i}$, so, if we condition on $\mathcal{F}_{n,i}$ we have

$$\mathbb{E}\Big[(\eta_i - \mathbb{E}[\eta_i])\{ \sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j])\}\Big] = \mathbb{E}\Big[(\eta_i - \mathbb{E}[\eta_i])\, \mathbb{E}[\sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j]) \,|\, \mathcal{F}_{n,i}]\Big]. \tag{40}$$

The definition of η_j tells us that $\eta_j - \mathbb{E}[\eta_j] = \mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,j-1}] - \mathbb{E}[d_{n,j}^2]$ so, because $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n,j-1}$ for all $i < j$, one then has

$$\mathbb{E}[\sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j]) \,|\, \mathcal{F}_{n,i}] = \sum_{j=i+1}^{n+m} \{\mathbb{E}[d_{n,j}^2 \,|\, \mathcal{F}_{n,i}] - \mathbb{E}[d_{n,j}^2]\}.$$

These summands have mean zero, so the bound (19) and the oscillation inequality (33) give us

$$\| \, \mathbb{E}[\sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j]) | \mathcal{F}_{n,i}] \, \|_\infty \le M C_n^2 \alpha_n^{-2}.$$

When we use this estimate in (40), we see from the non-negativity of η_j and the triangle inequality that

$$\| \, \mathbb{E}\Big[(\eta_i - \mathbb{E}[\eta_i])\{\sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j])\}\Big] \, \|_\infty \le M C_n^2 \alpha_n^{-2} \mathbb{E}[\eta_i],$$

so, after summing over $i \in \{1+m, \ldots, n+m\}$ and recalling the second inequality of (38) we obtain

$$\| \sum_{i=1+m}^{n+m} \mathbb{E}\Big[(\eta_i - \mathbb{E}[\eta_i])\{\sum_{j=i+1}^{n+m} (\eta_j - \mathbb{E}[\eta_j])\}\Big] \, \|_\infty \le M C_n^2 \alpha_n^{-2} \operatorname{Var}[S_n]. \tag{41}$$

By (37), the bounds (39) and (41) complete the proof of the lemma. □

Now, at last, we can use the basic decomposition (17) to write

$$\frac{S_n}{\sqrt{\operatorname{Var}[S_n]}} = \sum_{i=1}^{n} \frac{d_{n,i+m}}{\sqrt{\operatorname{Var}[S_n]}} + O\big(\frac{\| \, V_{n,m} \, \|_\infty}{\sqrt{\operatorname{Var}[S_n]}}\big), \tag{42}$$

and it only remains to apply our lemmas. First, from our hypothesis (4) that $C_n^2 \alpha_n^{-2} = o(\operatorname{Var}[S_n])$ as $n \to \infty$, we see that the L^∞-bound $\| \, d_{n,i} \, \|_\infty \le M C_n \alpha_n^{-1}$ in Lemma 7 implies the asymptotic negligibility (13) of the scaled differences $d_{n,i+m}/\sqrt{\operatorname{Var}[S_n]}$, $1 \le i \le n$. Second, our hypothesis (4) and the variance bounds (32) imply the asymptotic equivalence

$$\operatorname{Var}[S_n] \sim \sum_{i=1}^{n} \mathbb{E}[d_{n,i+m}^2] \qquad \text{as } n \to \infty,$$

so the L^2-inequality in Lemma 10 tells us that the weak law (14) also holds for the scaled martingale differences.

Taken together, these two observations imply that the first sum on the right-hand side of (42) converges in distribution to a standard normal. Moreover, because of the L^∞-bound $\| \, V_{n,m} \, \|_\infty \le M C_n \alpha_n^{-1}$ given by (29), the last term in (42) is asymptotically negligible. In turn, these observations tell us that

$$\frac{S_n}{\sqrt{\operatorname{Var}[S_n]}} \Longrightarrow N(0,1) \qquad \text{as } n \to \infty,$$

and the proof of Theorem 1 is complete.

8. Dynamic Inventory Management: A Leading Example

We now consider a classic dynamic inventory management problem where one has n periods and n independent demands $D_1, D_2, \ldots, D_n$. For the purpose of illustration, we also assume that demand D_i in period i has the uniform distribution on $[0, a]$ for some $a > 0$, but this assumption is far from necessary.

In each period $1 \le i \le n$ one knows the current level of inventory x, and the task is to decide the level of inventory $y \ge x$ that one wants to hold after an order is placed and fulfilled. Here it is also useful to allow for x to be negative, and, in that case, $|x|$ would represent the level of backlogged demand. To stay mindful of this possibility, we sometimes call x the generalized inventory level.

We further assume that orders are fulfilled instantaneously at a cost that is proportional to the ordered quantity; so, for example, to move the inventory level from x to $y \ge x$, one places an order of size $y - x$ and incurs a purchase cost equal to $c(y - x)$ where the multiplicative constant c is a parameter of the model.

The model also takes into account the cost of either holding physical inventory or of managing a backlog. Specifically, if the current generalized inventory is equal to x, then the firm incurs additional carrying costs that are given by

$$L(x) = \begin{cases} c_h x & \text{if } x \ge 0 \\ -c_p x & \text{if } x < 0. \end{cases}$$

In other words, if $x \ge 0$, then $L(x)$ represents the cost for holding a quantity x of inventory from one period to the next, and, if $x < 0$, then $L(x)$ represents the penalty cost for managing a quantity $-x \ge 0$ of unmet demand.

Here we also assume that all unmet demand can be successfully backlogged, so customers in one period whose demand is incompletely met will return in successive periods until either their demand has been met or until the decision period n is completed. If there is still unmet demand at time n, then that demand is lost. Finally, we assume that the purchase cost rate c is strictly smaller than the penalty rate c_p, so it is never optimal to accrue penalty costs when one can place an order. Naturally, the manager's objective is to minimize the total expected inventory costs over the decision periods $1, 2, \ldots, n$.

This problem has been widely studied, and, at this point, its formulation as a dynamic program is well understood — cf. Bellman, Glicksberg and Gross (1955), Bulinskaya (1964), or Porteus (2002, Section 4.2). Specifically, if we let $v_k(x)$ denote the minimal expected inventory cost when there are k time periods remaining and when x is the current generalized inventory level, then dynamic programming gives us the backwards recursion

$$v_k(x) = \min_{y \ge x} \{c(y - x) + \mathbb{E}[L(y - D_{n-k+1})] + \mathbb{E}[v_{k-1}(y - D_{n-k+1})]\}, \tag{43}$$

for $1 \le k \le n$, and one computes $v_k(x)$ by iteration beginning with $v_0(x) \equiv 0$.

It is also known that for this model there is a *base-stock policy* that is optimal; that is, there are non-decreasing values

$$s_1 \le s_2 \le \cdots \le s_n \tag{44}$$

such that if the current time is i and the current inventory is x, then the optimal level $\gamma_{n,i}(x)$ at time i for the inventory after restocking is given by

$$\gamma_{n,i}(x) = \begin{cases} s_{n-i+1} & \text{if } x \le s_{n-i+1} \\ x & \text{if } x > s_{n-i+1}. \end{cases} \tag{45}$$

In other words, if at time i the inventory level is below s_{n-i+1} then the optimal action is to place an order of size $s_{n-i+1} - x$, but if the inventory level is s_{n-i+1} or higher, then the optimal action is to order nothing. Moreover, Bulinskaya (1964, Theorem 1) also showed that for demands with the uniform distribution on $[0, a]$ one has the two relations:

$$s_1 = a\left(\frac{c_p}{c_h + c_p} - \frac{c}{c_h + c_p}\right) \quad \text{and} \quad s_n \le a\left(\frac{c_p}{c_h + c_p}\right) \text{ for } n \ge 2.$$

Remark 11 (Accommodation of Lead Times for Deliveries)**.** Here, to keep the description of the inventory problem as brief as possible, we have assumed that order fulfillment is instantaneous. Nevertheless, in a more realistic model, one might want to accommodate the possibility of lead times for delivery fulfillments. One practical benefit of our "look-ahead" parameter m is that one can allow for lead times and still stay within the scope of Theorem 1, but we do not need to pursue this extension here.

A CLT for Optimally Managed Inventory Costs

We take the generalized inventory at the beginning of period $i = 1$ (before any order is placed) to be $X_{n,1} = x$, where x can be any element of the state space $[-a, a]$. Subsequently we take $X_{n,i}$ to be the generalized inventory at the beginning of period $i \in \{2, 3, \dots, n\}$; so, in view of the base-stock policy (45), we have the recursion

$$X_{n,i+1} = \gamma_{n,i}(X_{n,i}) - D_i \qquad \text{for all } 1 \le i \le n. \tag{46}$$

This key point here is that $\{X_{n,i} : 1 \le i \le n+1\}$ is a time non-homogenous Markov chain with state space $\mathcal{X} = [-a, a]$.

Now, if π_n^* is the policy that minimizes the total expected inventory cost that is incurred over n decision periods, then the total cost that is realized when one follows the policy π_n^* is given by

$$\mathcal{C}_n(\pi_n^*) = \sum_{i=1}^{n} \{c(\gamma_{n,i}(X_{n,i}) - X_{n,i}) + L(X_{n,i+1})\}, \tag{47}$$

and we see that the total inventory cost $\mathcal{C}_n(\pi_n^*)$ is a special case of the sum (1). To spell out the correspondence, we first take $m = 1$, then we take

$$f_{n,i}(x, y) = c(\gamma_{n,i}(x) - x) + L(y), \qquad \text{for } 1 \le i \le n,$$

and finally (46) gives us the driving Markov chain. Theorem 1 and Corollary 2 now give us a natural path to a central limit theorem for the realized optimal inventory cost.

Theorem 12 (CLT for Mean-Optimal Inventory Cost)**.** *For the inventory cost $\mathcal{C}_n(\pi_n^*)$ realized under the mean-optimal policy π_n^*, one has*

$$\frac{\mathcal{C}_n(\pi_n^*) - \mathbb{E}[\mathcal{C}_n(\pi_n^*)]}{\sqrt{\mathrm{Var}[\mathcal{C}_n(\pi_n^*)]}} \Longrightarrow N(0, 1), \qquad \text{as } n \to \infty.$$

Two steps are used to extract this result from Theorem 1. First we show that the minimal ergodic coefficient of the Markov chain (46) is bounded away from zero, and, second, we show that the variance of $\mathcal{C}_n(\pi_n^*)$ grows to infinity as $n \to \infty$.

Before we take these steps, we should explain why Theorem 12 does not follow from Dobrushin's theorem and the traditional device of state space extension. The trouble comes from the fact that when one extends the state space the coefficient of ergodicity can become degenerate.

State Space Extension: Degeneracy of a Bivariate Chain

One can write the realized cost (47) as an additive functional of a Markov chain if one moves from the basic chain $\{X_{n,i} : 1 \leq i \leq n\}$ on $\mathcal{X}$ to the Markov chain

$$\{\, \widehat{X}_{n,i} = (X_{n,i}, X_{n,i+1}) : 1 \leq i \leq n \,\} \tag{48}$$

on the enlarged state space $\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}$. The realized cost (47) then becomes

$$\mathcal{C}_n(\pi_n^*) = \sum_{i=1}^{n} f_{n,i}(\widehat{X}_{n,i}), \tag{49}$$

and one might hope to apply Dobrushin's CLT (Theorem 3) to get the asymptotic distribution of $\mathcal{C}_n(\pi_n^*)$. To see why this plan does not succeed, one just needs to calculate the minimal ergodic coefficient for the extended chain (48).

For any $x, y \in \mathcal{X}$ and any $B \times B' \in \mathcal{B}(\mathcal{X}^2)$, the transition kernel of the Markov chain (48) is given by

$$\begin{aligned} K^{(n)}_{i,i+1}((x,y), B \times B') &= \mathbb{P}(X_{n,i+1} \in B, X_{n,i+2} \in B' \,|\, X_{n,i} = x, X_{n,i+1} = y) \\ &= \mathbb{1}(y \in B)\mathbb{P}(\{\gamma_{n,i+1}(y) - D_{i+1}\} \in B' \,|\, X_{n,i+1} = y), \end{aligned}$$

where $\gamma_{n,i}(x)$ is the function defined in (45). If we now set $B' = \mathcal{X}$, we have

$$K^{(n)}_{i,i+1}((x,y), B \times \mathcal{X}) = \begin{cases} 1 & \text{if } y \in B, \\ 0 & \text{if } y \in B^c, \end{cases}$$

so for $y \in B$ and $y' \in B^c$ we have

$$K^{(n)}_{i,i+1}((x,y), B \times \mathcal{X}) - K^{(n)}_{i,i+1}((x,y'), B \times \mathcal{X}) = 1.$$

This tells us that the minimal ergodic coefficient of the chain (48) is given by

$$\alpha_n = 1 - \max_{1 \leq i < n} \big\{ \sup_{(x,y),(x',y')} \| \, K^{(n)}_{i,i+1}((x,y), \,\cdot\,) - K^{(n)}_{i,i+1}((x',y'), \,\cdot\,) \, \|_{\mathrm{TV}} \big\} = 0,$$

and, as a consequence, we see that Dobrushin's classic CLT simply does not apply to the sum (49).

Remark 13 (Related Degeneracies)**.** In Section 2 we noted the possibility of replacing the minimal ergodic coefficient α_n of the Markov chain (48) with a potentially less fragile measure of dependence such as the maximal coefficient of correlation ρ_n used by Peligrad (2012). For the bivariate chain (48), the maximal coefficient of correlation is given by

$$\rho_n = \max_{2 \leq i \leq n} \sup_{g} \left\{ \frac{\| \, \mathbb{E}[g(\widehat{X}_{n,i}) \,|\, \widehat{X}_{n,i-1}] \, \|_2}{\| \, g(\widehat{X}_{n,i}) \, \|_2} : \| \, g(\widehat{X}_{n,i}) \, \|_2 < \infty \text{ and } \mathbb{E}[g(\widehat{X}_{n,i})] = 0 \right\},$$

so for the functional

$$g(\widehat{X}_{n,i}) = g(X_{n,i}, X_{n,i+1}) = X_{n,i} - \mathbb{E}[X_{n,i}],$$

one has $\rho_n = 1$, and we see that the CLT of Peligrad (2012) does not help us here.

Ergodic Coefficients of the Univariate Chain

The situation is more pleasant when one seeks to apply Theorem 1. The critical point is that one can show that the minimal ergodic coefficient of the *univariate* Markov chain (46) is bounded away from zero.

If we write $|\,C\,|$ for the Lebesgue measure of the Borel set C, then, for any $x \in \mathcal{X}$ and any Borel set $B \subseteq \mathcal{X}$, the transition kernel of the Markov chain (46) is given by

$$K^{(n)}_{i,i+1}(x, B) = \mathbb{P}(X_{n,i+1} \in B \,|\, X_{n,i} = x) = a^{-1}|\,B \cap [\gamma_{n,i}(x) - a, \gamma_{n,i}(x)]\,|.$$

The formula (45) for $\gamma_{n,i}(x)$ tells us that

$$\gamma_{n,i}(x) - a \le 0 \le s_{n-i+1} \le \gamma_{n,i}(x) \qquad \text{for all } x \in \mathcal{X},$$

so one always has

$$\begin{aligned}&|\,B \cap [\gamma_{n,i}(x) - a, \gamma_{n,i}(x)]\,|\\ &\quad = |\,B \cap [\gamma_{n,i}(x) - a, 0]\,| + |\,B \cap [0, s_{n-i+1}]\,| + |\,B \cap [s_{n-i+1}, \gamma_{n,i}(x)]\,|.\end{aligned}$$

When we take the difference between the transition kernel $K^{(n)}_{i,i+1}$ started at x and at x', we see from the last decomposition that the two middle summands cancel and we are left with

$$\begin{aligned}K^{(n)}_{i,i+1}(x, B) - K^{(n)}_{i,i+1}(x', B) =& \frac{1}{a}\big\{|\,B \cap [\gamma_{n,i}(x) - a, 0]\,| + |\,B \cap [s_{n-i+1}, \gamma_{n,i}(x)]\,|\\ &- |\,B \cap [\gamma_{n,i}(x') - a, 0]\,| - |\,B \cap [s_{n-i+1}, \gamma_{n,i}(x')]\,|\big\}.\end{aligned}$$

Now, for any $b_1 \le b_2$ and any $B \in \mathcal{B}(\mathcal{X})$ one has $|\,B \cap [b_1, b_2]\,| \le b_2 - b_1$, so, after some simplification, we have

$$|\,B \cap [\gamma_{n,i}(x) - a, 0]\,| + |\,B \cap [s_{n-i+1}, \gamma_{n,i}(x)]\,| \le a - s_{n-i+1},$$

and

$$-|\,B \cap [\gamma_{n,i}(x') - a, 0]\,| - |\,B \cap [s_{n-i+1}, \gamma_{n,i}(x')]\,| \ge -(a - s_{n-i+1}).$$

These bounds tell us that for all x, x', and B, we have

$$|\,K^{(n)}_{i,i+1}(x, B) - K^{(n)}_{i,i+1}(x', B)\,| \le \frac{a - s_{n-i+1}}{a},$$

so the monotonicity (44) of the base-stock levels gives us

$$\delta(K^{(n)}_{i,i+1}) = \sup_{\substack{x,x' \in \mathcal{X}\\ B \in \mathcal{B}(\mathcal{X})}} |\,K^{(n)}_{i,i+1}(x, B) - K^{(n)}_{i,i+1}(x', B)\,| \le \frac{a - s_{n-i+1}}{a} \le 1 - \frac{s_1}{a}.$$

This then implies that for all $n \ge 1$ we have

$$\alpha_n = \min_{1 \le i < n} \{1 - \delta(K^{(n)}_{i,i+1})\} \ge \frac{s_1}{a},$$

and this uniform lower bound on the minimal ergodic coefficient completes the first step in the proof of Theorem 12.

Variance Lower Bound

Here, as in most stochastic dynamic programs, the value to-go process (15) can be expressed in terms of the value functions that solve the dynamic programming recursion (43). In particular, at time $1 \le i \le n$, when the current generalized inventory is $X_{n,i}$ and there are $n - i + 1$ demands yet to be realized, one has

$$V_{n,i} = v_{n-i+1}(X_{n,i}),$$

where the function $x \mapsto v_{n-i+1}(x)$ is calculated by (43). Moreover, since we start with $X_{n,1} = x$, the definition of $v_n(x)$ gives us

$$V_{n,1} = v_n(x) = \mathbb{E}[\mathcal{C}_n(\pi_n^*)],$$

and the martingale decomposition (17) can be written more explicitly as

$$\mathcal{C}_n(\pi_n^*) - \mathbb{E}[\mathcal{C}_n(\pi_n^*)] = \sum_{i=1}^{n} d_{n,i+1}.$$

To estimate $\mathrm{Var}[\mathcal{C}_n(\pi_n^*)]$ from below, we just need to find an appropriate lower bound on $\mathbb{E}[d_{n,i+1}^2]$ for $1 \le i \le n$. In our inventory problem we begin by writing the martingale differences (16) more explicitly as

$$d_{n,i+1} = c(\gamma_{n,i}(X_{n,i}) - X_{n,i}) + L(X_{n,i+1}) + v_{n-i}(X_{n,i+1}) - v_{n-i+1}(X_{n,i}). \tag{50}$$

Next, we introduce the shorthand $\widehat{v}_{n-i}(x) = L(x) + v_{n-i}(x)$, and we obtain from the recursion (43) and the policy characterization (45) that

$$\begin{aligned} v_{n-i+1}(x) &= c(\gamma_{n,i}(x) - x) + \mathbb{E}[L(\gamma_{n,i}(x) - D_i)] + \mathbb{E}[v_{n-i}(\gamma_{n,i}(x) - D_i)]\} \\ &= c(\gamma_{n,i}(x) - x) + \mathbb{E}[\widehat{v}_{n-i}(\gamma_{n,i}(x) - D_i)]. \end{aligned} \tag{51}$$

We now replace x with $X_{n,i}$ in (51) to get a new expression for $v_{n-i+1}(X_{n,i})$, and we replace the last summand of (50) with this expression. If we recall from (46) that $X_{n,i+1} = \gamma_{n,i}(X_{n,i}) - D_i$, then we find after simplification that

$$d_{n,i+1} = \widehat{v}_{n-i}(\gamma_{n,i}(X_{n,i}) - D_i) - \mathbb{E}[\widehat{v}_{n-i}(\gamma_{n,i}(X_{n,i}) - D_i) \,|\, \mathcal{F}_{n,i}],$$

where, just as before, one has $\mathcal{F}_{n,i} = \sigma\{X_{n,1}, X_{n,2}, \dots, X_{n,i}\}$. This representation gives us a key starting point for estimating the second moment of $d_{n,i+1}$.

Lemma 14. *For the inventory cost $\mathcal{C}_n(\pi_n^*)$ realized under the mean-optimal policy π_n^*, one has for all $n \ge 1$ that*

$$\mathrm{Var}[\mathcal{C}_n(\pi_n^*)] = \sum_{i=1}^{n} \mathbb{E}[d_{n,i+1}^2] \ge \left\{ \frac{(c + c_h)^2 s_1^4}{12a^2} \right\} n.$$

Proof. We now let $(D_1', D_2', \dots, D_n')$ be an independent copy of $(D_1, D_2, \dots, D_n)$. Since $X_{n,i}$ is $\mathcal{F}_{n,i}$-measurable, one then has the further representation

$$\mathbb{E}[d_{n,i+1}^2 \,|\, \mathcal{F}_{n,i}] = \frac{1}{2}\mathbb{E}[\{\widehat{v}_{n-i}(\gamma_{n,i}(X_{n,i}) - D_i) - \widehat{v}_{n-i}(\gamma_{n,i}(X_{n,i}) - D_i')\}^2 \,|\, \mathcal{F}_{n,i}].$$

Next, we consider the set $G(X_{n,i})$ of all ω such that

$$D_i(\omega) \in [\gamma_{n,i}(X_{n,i}) - s_1, \gamma_{n,i}(X_{n,i})] \quad \text{and} \quad D_i'(\omega) \in [\gamma_{n,i}(X_{n,i}) - s_1, \gamma_{n,i}(X_{n,i})].$$

In other words, at time i when the generalized inventory begins with $X_{n,i}$, if $\omega \in G(X_{n,i})$ then both the demand $D_i(\omega)$ and the demand $D_i'(\omega)$ would cause one to order up to the level s_{n-i} in period $i + 1$.

If we now replace i with $i+1$ in the recursion (51) we see that

$$\{\widehat{v}_{n-i}(x) - \widehat{v}_{n-i}(y)\}\mathbb{1}((x,y) \in [0,s_1]^2) = (c+c_h)(y-x)\mathbb{1}((x,y) \in [0,s_1]^2),$$

because the two new inventory levels for the next period $i+1$ are both given by $\gamma_{n,i+1}(x) = \gamma_{n,i+1}(y) = s_{n-i}$ and because one incurs holding costs that are proportional to the difference $y-x$. This last equivalence implies the lower bound

$$\mathbb{E}[d_{n,i+1}^2 \,|\, \mathcal{F}_{n,i}] \geq \frac{1}{2}(c+c_h)^2 \mathbb{E}[\{D_i' - D_i\}^2 \mathbb{1}(G(X_{n,i})) \,|\, \mathcal{F}_{n,i}].$$

We now recall that D_i and D_i' are independent and uniformly distributed on $[0,a]$ and integrate the right-hand side to obtain that

$$\mathbb{E}[d_{n,i+1}^2 \,|\, \mathcal{F}_{n,i}] \geq \frac{(c+c_h)^2 s_1^4}{12a^2} > 0.$$

After one takes the total expectation and sums over $1 \leq i \leq n$, the proof of the lemma is complete. □

9. An Application in Combinatorial Optimization: Online Alternating Subsequences

Given a sequence $y_1, y_2, \ldots, y_n$ of n distinct real numbers, we say that a subsequence $y_{i_1}, y_{i_2}, \ldots, y_{i_k}$, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, is *alternating* provided that the relative magnitudes alternate as in

$$y_{i_1} < y_{i_2} > y_{i_3} < y_{i_4} > \cdots \quad \text{or} \quad y_{i_1} > y_{i_2} < y_{i_3} > y_{i_4} < \cdots .$$

Combinatorial investigations of alternating subsequences go back to Euler (cf. Stanley, 2010), but probabilistic investigations are more recent; Widom (2006), Pemantle (cf. Stanley, 2007, p. 568), Stanley (2008) and Houdré and Restrepo (2010) all considered the distribution of the length of the longest alternating subsequence of a random permutation or of a sequence $\{Y_1, Y_2, \ldots, Y_n\}$ of independent random variables with the uniform distribution on $[0,1]$. There have also been recent applications of this work in computer science (e.g. Romik, 2011; Bannister and Eppstein, 2012) and in tests of independence (cf. Brockwell and Davis, 2006, p. 312).

Here we consider alternating subsequences in a *sequential*, or *online*, context where we are presented with the values $Y_1, Y_2, \ldots, Y_n$ one at the time, and the goal is to select an alternating subsequence

$$Y_{\tau_1} < Y_{\tau_2} > Y_{\tau_3} < Y_{\tau_4} > \cdots \lesseqgtr Y_{\tau_k} \tag{52}$$

that has maximal expected length.

A sequence of selection times $1 \leq \tau_1 < \tau_2 < \cdots < \tau_k \leq n$ that satisfy (52) is called a *feasible policy* if our decision to accept or reject Y_i as member of the alternating subsequence is based only on our knowledge of the observations $\{Y_1, Y_2, \ldots, Y_i\}$. In more formal terms, the feasibility of a policy is equivalent to requiring that the indices τ_k, $k = 1, 2, \ldots$, are all stopping times with respect to the increasing sequence of σ-fields $\mathcal{A}_i = \sigma\{Y_1, Y_2, \ldots, Y_i\}$, $1 \leq i \leq n$.

We now let Π denote the set of all feasible policies, and for $\pi \in \Pi$, we let $A_n^o(\pi)$ be the number of alternating selections made by π for the realization $\{Y_1, Y_2, \ldots, Y_n\}$, so

$$A_n^o(\pi) = \max\{k : Y_{\tau_1} < Y_{\tau_2} > \cdots \lesseqgtr Y_{\tau_k} \text{ and } 1 \leq \tau_1 < \tau_2 < \cdots < \tau_k \leq n\}.$$

We say that a policy $\pi_n^* \in \Pi$ is optimal (or, more precisely, *mean-optimal*) if

$$\mathbb{E}[A_n^o(\pi_n^*)] = \sup_{\pi \in \Pi} \mathbb{E}[A_n^o(\pi)].$$

Arlotto, Chen, Shepp and Steele (2011) found that for each n there is a unique mean-optimal policy π_n^* such that

$$\mathbb{E}[A_n^o(\pi_n^*)] = (2 - \sqrt{2})n + O(1),$$

and it was later found that there is a CLT for $A_n^o(\pi_n^*)$.

Theorem 15 (CLT for Optimal Number of Alternating Selections)**.** *For the mean-optimal number of alternating selections $A_n^o(\pi_n^*)$ one has*

$$\frac{A_n^o(\pi_n^*) - \mathbb{E}[A_n^o(\pi_n^*)]}{\sqrt{\mathrm{Var}[A_n^o(\pi_n^*)]}} \Longrightarrow N(0,1) \qquad \text{as } n \to \infty.$$

The main goal of this section is to show that Theorem 1 leads to a proof of this theorem that is quicker, more robust, and more principled than the original proof given in Arlotto and Steele (2014). In the process, we also get a second illustration of the ways in which Theorem 1 helps one sidestep the degeneracy that sometimes arises when one tries to use Dobrushin's theorem on a naturally associated bivariate chain. In fact, it is this feature of Dobrushin's theorem that initially motivated the development of Theorem 1.

Structure of the Additive Process

To formulate the alternating subsequence problem as an MDP, we first consider a new state space that consists of pairs (x, s) where x denotes the value of the last selected observation and where we set $s = 0$ if x is a local minimum and set $s = 1$ if x is a local maximum. The decision problem then has a notable reflection property: the optimal expected number of alternating selections that one makes when k observations are yet to be seen is the same if the system is in state $(x, 0)$ or if the system is in state $(1 - x, 1)$. Earlier analyses exploited this symmetry to show that there is a sequence $\{g_k : 1 \le k < \infty\}$ of optimal threshold functions such that if one sets $X_{n,1} = 0$ and lets

$$X_{n,i+1} = \begin{cases} X_{n,i} & \text{if } Y_i < g_{n-i+1}(X_{n,i}) \\ 1 - Y_i & \text{if } Y_i \ge g_{n-i+1}(X_{n,i}), \end{cases} \tag{53}$$

then the optimal number of alternating selections has the representation

$$A_n^o(\pi_n^*) = \sum_{i=1}^{n} \mathbb{1}\left(Y_i \ge g_{n-i+1}(X_{n,i})\right) = \sum_{i=1}^{n} \mathbb{1}(X_{n,i+1} \ne X_{n,i}). \tag{54}$$

The derivation of these relations requires a substantial amount of work, but for the purpose of illustrating Theorem 1 and Corollary 2, one does not need to go into the details of the construction of these optimal threshold functions. Here it is enough to note that this representation for $A_n^o(\pi_n^*)$ is exactly of the form (1) that is addressed by Theorem 1.

The proof of Theorem 15 then takes two steps. First, one needs an appropriate lower bound for the minimal ergodic coefficients of the chain (53), and second one needs to check that the variance of $A_n^o(\pi_n^*)$ goes to infinity as $n \to \infty$.

The second property is almost baked into the cake, and it is even proved in Arlotto and Steele (2014) that $\mathrm{Var}[A_n^o(\pi_n^*)]$ grows linearly with n. Still, to keep our discussion brief, we will not repeat that proof. Instead we focus on the new — and more strategic — fact that minimal ergodic coefficients of the Markov chains (53) are uniformly bounded away from zero for all $1 \le i \le n-2$ and all $n \ge 3$.

A Lower Bound for the Minimal Ergodic Coefficient

For any $x \in [0,1]$ and any Borel set $B \subseteq [0,1]$, the Markov chain (53) has the transition kernel

$$\begin{aligned} K^{(n)}_{i,i+1}(x,B) &= \mathbb{1}(x \in B) g_{n-i+1}(x) + \int_{g_{n-i+1}(x)}^{1} \mathbb{1}(1-u \in B)\, du \\ &= \mathbb{1}(x \in B) g_{n-i+1}(x) + |\, B \cap [0, 1-g_{n-i+1}(x)]\,|, \end{aligned}$$

where the first summand of the top equation accounts for the rejection of the newly presented value $Y_i = u$, and the second summand accounts for its acceptance.

To obtain a meaningful estimate for the contraction coefficient of $K^{(n)}_{i,i+1}$ we recall from the earlier analyses that the optimal threshold functions $\{g_k : 1 \le k < \infty\}$ have the two basic properties: (i) $g_k(x) = x$ for all $x \in [1/3, 1]$ and all $k \ge 1$, and (ii) $g_k(x) \ge 1/6$ for all $x \in [0,1]$ and all $k \ge 3$. Property (ii) and the recursion (53) give us $X_{n,i} \le 5/6$ for all $1 \le i \le n-2$, and we see from property (i) that

$$\delta(K^{(n)}_{i,i+1}) = \sup_{x,x'} \| \, K^{(n)}_{i,i+1}(x,\cdot) - K^{(n)}_{i,i+1}(x',\cdot) \, \|_{\mathrm{TV}} \le \frac{5}{6} \quad \text{for all } 1 \le i \le n-2.$$

This estimate gives us in turn that

$$\alpha_{n-2} = \min_{1 \le i < n-2} \{1 - \delta(K^{(n)}_{i,i+1})\} \ge \frac{1}{6},$$

so by Corollary 2 we have the CLT for $A^o_{n-2}(\pi^*_n)$. Since $A^o_n(\pi^*_n)$ and $A^o_{n-2}(\pi^*_n)$ differ by at most 2, this also completes the proof of Theorem 15.

10. A Final Observation

Theorem 1 generalizes the classical CLT of Dobrushin (1956), and it offers a prepackaged approach to the CLT for the kinds of additive functionals that one meets in the theory of finite horizon Markov decision processes. The technology of MDPs is wedded to the pursuit of policies that maximize total expected rewards, but such policies may not make good economic sense unless the realized reward is "well behaved." While there are several ways to characterize good behavior, asymptotic normality of the realized reward is likely to be high on almost anyone's list. The orientation of Theorem 1 addresses this issue in a direct and practical way.

The examples of Sections 8 and 9 illustrate more concretely what one needs to do to apply Theorem 1. In a nutshell, one needs to show that the variance of the total reward goes to infinity and one needs an *a priori* lower bound on the minimal coefficient of ergodicity. These conditions are not trivial, but, as the examples show, they are not intractable. Now, whenever one faces the question of a CLT for the total reward of a finite horizon MDP, there is an explicit agenda that lays out what one needs to do.

References

Arlotto, A., Chen, R. W., Shepp, L. A. and Steele, J. M. (2011), 'Online selection of alternating subsequences from a random sample', *J. Appl. Probab.* **48**(4), 1114–1132.

Arlotto, A., Gans, N. and Steele, J. M. (2014), 'Markov decision problems where means bound variances', *Oper. Res.* **62**(4), 864–875.

Arlotto, A. and Steele, J. M. (2014), 'Optimal online selection of an alternating subsequence: a central limit theorem', *Adv. in Appl. Probab.* **46**(2), 536–559.

Bannister, M. J. and Eppstein, D. (2012), Randomized speedup of the Bellman–Ford algorithm, *in* 'Proceedings of the Meeting on Analytic Algorithmics & Combinatorics, Kyoto, Japan, January 16, 2012', Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 41–47.

Bellman, R., Glicksberg, I. and Gross, O. (1955), 'On the optimal inventory equation', *Management Sci.* **2**, 83–104.

Brockwell, P. J. and Davis, R. A. (2006), *Time series: theory and methods*, Springer Series in Statistics, Springer, New York. Reprint of the second (1991) edition.

Brown, B. M. (1971), 'Martingale central limit theorems', *Ann. Math. Statist.* **42**, 59–66.

Bulinskaya, E. V. (1964), 'Some results concerning optimum inventory policies', *Theory of Probability & Its Applications* **9**(3), 389–403.

Del Moral, P. (2004), *Feynman-Kac formulae*, Probability and its Applications (New York), Springer-Verlag, New York.

Dobrushin, R. L. (1956), 'Central limit theorem for nonstationary Markov chains. I, II', *Theory of Probability & Its Applications* **1**(1), 65–80; ibid. **1**(4), 329–383.

Gordin, M. I. (1969), 'The central limit theorem for stationary processes', *Dokl. Akad. Nauk SSSR* **188**, 739–741.

Gordin, M. and Peligrad, M. (2011), 'On the functional central limit theorem via martingale approximation', *Bernoulli* **17**(1), 424–440.

Hall, P. and Heyde, C. C. (1980), *Martingale limit theory and its application*, Academic Press, Inc., New York-London.

Hanen, A. (1963), 'Théorèmes limites pour une suite de chaînes de Markov', *Ann. Inst. H. Poincaré* **18**, 197–301.

Houdré, C. and Restrepo, R. (2010), 'A probabilistic approach to the asymptotics of the length of the longest alternating subsequence', *Electron. J. Combin.* **17**(1), Research Paper 168, 1–19.

Iosifescu, M. and Theodorescu, R. (1969), *Random processes and learning*, Springer-Verlag, New York.

Kifer, Y. (1998), 'Limit theorems for random transformations and processes in random environments', *Trans. Amer. Math. Soc.* **350**(4), 1481–1518.

Kipnis, C. and Varadhan, S. R. S. (1986), 'Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions', *Comm. Math. Phys.* **104**(1), 1–19.

Mandl, P. (1973), 'A connection between controlled Markov chains and martingales', *Kybernetika (Prague)* **9**, 237–241.

Mandl, P. (1974*a*), 'Estimation and control in Markov chains', *Adv. in Appl. Probab.* **6**, 40–60.

Mandl, P. (1974*b*), On the asymptotic normality of the reward in a controlled Markov chain, *in* 'Progress in statistics (European Meeting Statisticians, Budapest, 1972), Vol. II', North-Holland, Amsterdam, pp. 499–505. Colloq. Math. Soc. János Bolyai, Vol. 9.

Mandl, P. (1985), 'Local asymptotic normality in controlled Markov chains', *Statist. Decisions* **2**, 123–127. Selected papers presented at the 16th European meeting of statisticians (Marburg, 1984).

Mandl, P. and Laušmanová, M. (1991), 'Two extensions of asymptotic methods in controlled Markov chains', *Ann. Oper. Res.* **28**(1-4), 67–79.

McLeish, D. L. (1974), 'Dependent central limit theorems and invariance principles', *Ann. Probab.* **2**, 620–628.

Mendoza-Pérez, A. F. (2008), 'Asymptotic normality of average cost Markov control processes', *Morfismos* **12**, 33–52.

Mendoza-Pérez, A. F. and Hernández-Lerma, O. (2010), 'Asymptotic normality of discrete-time Markov control processes', *J. Appl. Probab.* **47**(3), 778–795.

Peligrad, M. (2012), 'Central limit theorem for triangular arrays of non-homogeneous Markov chains', *Probab. Theory Related Fields* **154**(3-4), 409–428.

Porteus, E. (2002), *Foundations of Stochastic Inventory Theory*, Business/Marketing, Stanford University Press.

Romik, D. (2011), Local extrema in random permutations and the structure of longest alternating subsequences, *in* '23rd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2011)', Discrete Math. Theor. Comput. Sci. Proc., AO, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, pp. 825–834.

Sarymsakov, T. A. (1961), 'Inhomogeneous Markov chains', *Theory of Probability & Its Applications* **6**(2), 178–185.

Seneta, E. (2006), *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York. Revised reprint of the second (1981) edition.

Sethuraman, S. and Varadhan, S. R. S. (2005), 'A martingale proof of Dobrushin's theorem for non-homogeneous Markov chains', *Electron. J. Probab.* **10**, no. 36, 1221–1235 (electronic).

Stanley, R. P. (2007), Increasing and decreasing subsequences and their variants, *in* 'International Congress of Mathematicians. Vol. I', Eur. Math. Soc., Zürich, pp. 545–579.

Stanley, R. P. (2008), 'Longest alternating subsequences of permutations', *Michigan Math. J.* **57**, 675–687. Special volume in honor of Melvin Hochster.

Stanley, R. P. (2010), 'A survey of alternating permutations', *Contemp. Math.* **531**, 165–196.

Statuljavičus, V. A. (1969), 'Limit theorems for sums of random variables that are connected in a Markov chain. I, II, III', *Litovsk. Mat. Sb.* **9**, 345–362; ibid. **9**, 635–672; ibid. **10**, 161–169.

Szewczak, Z. S. (2012), 'On Dobrushin's inequality', *Statist. Probab. Lett.* **82**(6), 1202–1207.

Widom, H. (2006), 'On the limiting distribution for the length of the longest alternating sequence in a random permutation', *Electron. J. Combin.* **13**(1), Research Paper 25, 1–7.

Winkler, G. (2003), *Image analysis, random fields and Markov chain Monte Carlo methods*, Vol. 27 of *Applications of Mathematics (New York)*, second edn, Springer-Verlag, Berlin.

Wu, W. B. and Woodroofe, M. (2004), 'Martingale approximations for sums of stationary processes', *Ann. Probab.* **32**(2), 1674–1690.