EXTENSION OF WALD'S FIRST LEMMA TO MARKOV PROCESSES

GEORGE V. MOUSTAKIDES,∗
University of Patras

Abstract

Let \( \xi_0, \xi_1, \xi_2, \ldots \) be a homogeneous Markov process and let \( S_n \) denote the partial sum
\[ S_n = \theta(\xi_1) + \ldots + \theta(\xi_n), \]
where \( \theta(\xi) \) is a scalar nonlinearity. If \( N \) is a stopping time
with \( \mathbb{E}N < \infty \) and the Markov process \( \{\xi_n\}_0^\infty \) satisfies certain ergodicity properties,
we then show that
\[ \mathbb{E}S_N = \left[ \lim_{n \to \infty} \mathbb{E}\theta(\xi_n) \right] \mathbb{E}N + \mathbb{E}\omega(\xi_0) - \mathbb{E}\omega(\xi_N). \]
The function \( \omega(\xi) \) is a well defined scalar nonlinearity directly related to
\( \theta(\xi) \) through a Poisson integral equation, with the characteristic that
\( \omega(\xi) \) becomes zero in the i.i.d. case. Consequently our result constitutes an extension to Wald's first lemma for the case of
Markov processes. We also show that, when \( \mathbb{E}N \to \infty \), the correction term is negligible
as compared to \( \mathbb{E}N \) in the sense that
\[ \mathbb{E}\omega(\xi_0) - \mathbb{E}\omega(\xi_N) = o(\mathbb{E}N). \]

Keywords: Wald's lemma; stopping times

AMS 1991 Subject Classification: Primary 60G40

1. Introduction

Wald's lemmas constitute a very powerful tool in sequential analysis for evaluating the
performance of sequential schemes. In particular they can be used in schemes where the
test statistic is a random sum of i.i.d. random variables. This includes both popular tests of
sequential analysis, namely the sequential probability ratio test (SPRT) used for the sequential
hypotheses problem, and the CUSUM test used for the disruption problem. Both problems
are known to have a large number of applications in the areas of signal, image and speech
processing, communications, systems monitoring, economics and so on. Since, in such applica-
tions, the data encountered are rarely i.i.d., the extension of Wald's lemmas to more general
data types is of primary interest. In this work we focus on one of Wald's lemmas and more
specifically on the one known as Wald's first lemma.

Let \( X_1, X_2, \ldots \) be a sequence of random variables with partial sum
\[ \sum_{i=1}^{n} X_i. \]
Let \( N \) denote any stopping time with respect to the filtration generated by
\( \{X_n\}_0^\infty \). Wald's first lemma [18], extended by Blackwell [1], states that if
\( \{X_n\}_0^\infty \) is an i.i.d. sequence with
\[ \mathbb{E}|X_1| < \infty \]
and the stopping time \( N \) has
\[ \mathbb{E}N < \infty, \]
then
\[ \mathbb{E}\sum_{i=1}^{n} X_i = \mathbb{E}X_1 \mathbb{E}N. \]
(1)

Generalizations of this lemma consider primarily the case
\[ \mathbb{E}X_1 = 0 \]
and conditions on the
stopping time \( N \), and the sequence
\( \{X_n\}_0^\infty \) that ensures
\[ \mathbb{E}\sum_{i=1}^{n} X_i = 0. \]
Burkholder and Gandy [2]
show that if
\[ \mathbb{E}|X_1|^{\alpha} < \infty \]
and
\[ \mathbb{E}N^{1/\alpha} < \infty \]
then
\[ \mathbb{E}\sum_{i=1}^{n} X_i = 0 \]
for the case \( \alpha = 2 \), for which a
generalization to any \( 1 < \alpha \leq 2 \) was made in [3]. Chow et al. [4] extended this result further,
to a class of denormalized U-statistics.

Received 9 June 1997; revision received 4 February 1998.

∗ Postal address: Department of Computer Engineering and Informatics, University of Patras, 26 500 Patras, Greece.
Email address: moustaki@cti.gr.
In this paper we intend to consider a similar data case as in [16], consequently we will make use of the results of Sadowsky [16] who considered the case where the random variable $\theta(\xi)$ is bounded. Before going into any detail, we need to call upon certain definitions and results regarding Dynkin's identity as we will see in Section 2.1.

Extension of Wald's first lemma to Markov processes

We will identify in a more precise way the two parameters of (3) that guarantee $\lim_{n \to \infty} E[Y_n - Y] = 0$. La Pena [5] introduced a tail probability version of Wald's lemma for expectations of randomly stopped sums and showed that the work of Gundy [7] could not be extended to this case. Recent publications aim to obtain conditions that are necessary and sufficient for the validity of Wald's first lemma. Gundy [7] considered the case of a scalar nonlinearity and showed that the maximum eigenvalue identified. It turns out that this is not easy most of the time, even for simple Markov processes (as for example finite state chains). The difficulty is mainly due to the parametrization of (3) with respect to the parameter $\xi$. In this paper we will present a solution in the form of a series (one of the examples being finite state chains). One last form solutions for the function $\omega(\xi)$ is related to $\theta(\xi)$ which enters in (2). Specifically for stopping times with approach we are going to follow, we will identify in a more precise way the two parameters of (3) that we do not simultaneously have $\mu(\xi)$. The case of a scalar nonlinearity and $\mu(\xi)$ is not considered in this paper as we will see in Section 2.1.

From Dynkin's identity as we will see in Section 2.1.

We will present a solution in the form of a series because it is not parametrized with respect to any parameter. Furthermore for other identity, involving products of random variables, he showed that $\mu(\xi)$ can be explicitly solved and the maximum eigenvalue identified. It turns out that this is not easy most of the time, even for simple Markov processes (as for example finite state chains). The difficulty is mainly due to the parametrization of (3) with respect to the parameter $\xi$. In this paper we will present a solution in the form of a series (one of the examples being finite state chains). One last form solutions for the function $\omega(\xi)$ is related to $\theta(\xi)$ which enters in (2). Specifically for stopping times with approach we are going to follow, we will identify in a more precise way the two parameters of (3) that we do not simultaneously have $\mu(\xi)$. The case of a scalar nonlinearity and $\mu(\xi)$ is not considered in this paper as we will see in Section 2.1.
notions. Several examples are presented in [12].

We must note that in many interesting cases it is relatively easy to verify these

the g.c.d. of the $n$ and $\phi(n)$

It is known (see [12, Chapter 5]), that for $\phi(n)$

existence of the

the main advantages of

maximal irreducibility measure. Of course this property is seldom used in practice since one of

necessary definitions. A

It turns out (see [12]) that if an invariant probability measure

is satisfied

for some constants

$P$

Let us first proceed with the

with criteria that can guarantee the proper form of convergence required by our analysis, and

exponential rate. The theory of

$\phi$

also with various background results that will facilitate our proofs. Let us first proceed with the

finite we have

$\pi$

there exists a function $V$

$\sigma$

Let also

$\sum_{i \in \mathbb{N}} \mathbb{I}_i$

Theorem 1.1.

and a function $V$

The process is geometrically ergodic in the sense that there exists a probability measure

$\pi$

and $\lambda$

Let $A$ be

$\mathbb{I}_i$

$\mathbb{I}_i$

There exists a function $V$

$\mathbb{E}$

$\mathbb{E}_n$

$\pi$

$\pi$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$

$\lambda$
Although for finite state chains (and processes with compact state space $X$) the exponential convergence toward the invariant probability measure is uniform, this is not the case for general processes where there is dependence on the initial state. It is this dependence that Theorem 1.1 reveals and controls efficiently for $\phi$-irreducible aperiodic processes.

2. Generalization of Wald's first lemma

Let us now assume that the process $\{\xi_n\}$ is $\phi$-irreducible and aperiodic satisfying either of the two conditions of Theorem 1.1 for some given function $V(\xi) \geq 1$. Furthermore let us assume that any stopping time we use is adapted to the filtration $\{F_n\}$ generated by the process $\{\xi_n\}$.

To this end consider the set $L_\infty V$ of all functions $g(\xi)$ that satisfy $\sup_{\xi} |g(\xi)| / V(\xi) < \infty$. The space $L_\infty V$ equipped with the norm $\|g\|_V = \sup_{\xi} |g(\xi)| / V(\xi)$ is a Banach space.

In a sense $L_\infty V$ contains all the functions for which we can a priori guarantee existence of their expectation under the invariant measure and also geometric ergodicity in the form of Theorem 1.1. With the help of Theorem 1.1, we can now show the following lemma.

Lemma 2.1. Let $g(\xi) \in L_\infty V$ then

$$\|P_\pi g - \pi g\|_V \leq \rho^n \|g\|_V R,$$

where $\rho$ and $R$ are the constants defined in Theorem 1.1.

Proof. The statement of this lemma is equivalent to (6), since we can observe that for any function $g(\xi) \in L_\infty V$ we have from the definition of the norm that $|g(\xi)| / \|g\|_V \leq V(\xi)$.

Let us now present an important property for functions in $L_\infty V$, involving randomly stopped sums, that will be used in the following.

Lemma 2.2. Let $g(\xi) \in L_\infty V$ then for any stopping time $N$ we have that

$$E\left[\left| \sum_{n=0}^{N-1} g(\xi_n) \right| \right] \leq \|g\|_V 1 - \lambda \left( E V(\xi_0) + b E N \right),$$

where $\lambda$ and $b$ are the constants defined in the drift condition (5).

Proof. The proof is a special case of Proposition 11.3.2 of [12] with $Z_k = \|g\|_V V(\xi_k)/(1 - \lambda)$, $f_k(\xi) = |g(\xi)|$, and $s_k(\xi) = \|g\|_V b/(1 - \lambda)$.

We call upon our last lemma before the presentation of our main theorem. This lemma introduces the function $\omega(\xi)$ that enters in the correction term of the generalized form of Wald's lemma.

Lemma 2.3. Let $\theta(\xi) \in L_\infty V$, and consider the following Poisson integral equation with respect to the unknown function $\omega(\xi)$, i.e.

$$P \omega = \omega - (P \theta - \pi \theta),$$

(7)

with the constraint $\pi \omega = 0$.

(8)
then the constrained Poisson equation has a unique solution in $L^\infty$ given by the series

$$\omega = \sum_{n=1}^{\infty} P_n \theta - \pi \theta,$$

(9)

which is convergent in norm in $L^\infty$.

Proof. Notice that if $\omega(\xi)$ is a solution to (7) so is $c + \omega(\xi)$ for any constant $c$. Setting the constraint $\pi \omega = 0$ results in a unique solution as we will soon see. That the series defined in (9) converges is an immediate consequence of Lemma 2.1. That $\omega(\xi)$ is a solution to the Poisson equation (7) satisfying the constraint in (8) can be easily verified using the Markov property of the process. Finally, to show that $\omega(\xi)$ is unique assume that there exist two functions $\omega_1(\xi)$ and $\omega_2(\xi)$, both in $L^\infty$, satisfying (7) and (8). Define $\delta(\xi) = \omega_1(\xi) - \omega_2(\xi)$, then $\delta(\xi) \in L^\infty$, $\pi \delta = 0$ and $P \delta = \delta$. Applying the last relation repeatedly, we conclude that for any $n \in \mathbb{Z}^+$ we have $P_n \delta = \delta$. This last equality combined with Lemma 2.1 yields $\|\delta\|_V \leq \rho^n \|\delta\|_V R$. Since $0 \leq \rho < 1$ we have $\|\delta\|_V = 0$ and this concludes the proof.

We are now in a position to prove our main result.

Theorem 2.1. Let $E_V(\xi_0) < \infty$ and $\theta(\xi) \in L^\infty$. Define $S_n = \theta(\xi_1) + \cdots + \theta(\xi_n)$, then for any stopping time $N$ with $E_N < \infty$ we have

$$E_S = \lim_{n \to \infty} E_{\theta(\xi_n)} = E_{N} + E_{\omega(\xi_0) - \omega(\xi_N)},$$

(10)

with the function $\omega(\xi)$ defined in Lemma 2.3.

Proof. As in the proof for the i.i.d. case (see [17]) we need to introduce a proper martingale. Thus let us consider $U_n = S_n - \pi \theta + \omega(\xi_n)$; using (7) it is straightforward to show that $U_n$ is indeed a martingale. To complete our proof we also need to show that $\lim_{n \to \infty} E[|U_n| 1_{\{N > n\}}] = 0$.

We have $|U_n| \leq \sum_{k=1}^{n} |\theta(\xi_k)| + |\pi \theta| n + |\omega(\xi_n)|$.

Consequently on the set $\{N > n\}$ we can write

$$|U_n| \leq N - 1 \sum_{k=0}^{N-1} |\theta(\xi_k)| + |\pi \theta| N + \sum_{k=0}^{N-1} |\omega(\xi_k)|.$$

Since both functions $\theta(\xi)$ and $\omega(\xi)$ belong to $L^\infty$, with the help of Lemma 2.2 and the fact that we assumed that $E_V(\xi_0) < \infty$, we conclude that all three terms on the right-hand side of (12) have finite expectation. Thus (11) is satisfied.
It is worth noting that the expression \( E(\omega(\xi_0)) - E(\omega(\xi_N)) \) also appears in another popular identity that has been in the literature for several years—we refer to Dynkin's well-known identity. As we will see, it is possible to recover our proposed extension through this identity as well. We must point out, however, that, although not explicitly stated, Dynkin's identity was indirectly used in our previous proof since it is part of the proof of [12, Proposition 11.3.2], which we employed in showing Lemma 2.2.

Although the extension of Wald's first lemma was introduced for Markov processes \( \{\xi_n\} \) evolving on general state spaces, the processes in question, due to Theorem 1.1, were characterized by geometric ergodicity. This of course suggests that our analysis excludes Markov processes that are ergodic but not geometrically ergodic. As our referee pointed out, for the special case of countable state spaces and bounded \( \theta(\xi) \) functions it is possible to show the extension of Wald's lemma, for general ergodic Markov chains, through Dynkin's identity. Let us present this fact in the following theorem

**Theorem 2.2.** Let the Markov chain \( \{\xi_n\} \) be aperiodic, ergodic and evolving on a countable state space \( X \). For any bounded function \( \theta(\xi) \) define \( \omega(\xi) \) to satisfy (9), then the corresponding series is absolutely convergent. If \( \omega(\xi) \) is bounded then relation (10) is also valid.

**Proof.** It is possible to modify the proof of Theorem 2.1 by using alternative drift conditions guaranteeing more general forms of ergodicity (see [12]). In fact we could even go without the requirement of boundedness for the \( \omega(\xi) \) function. We prefer, however, to proceed along a different line in order to emphasize the association of the proposed extension with Dynkin's identity.

If \( \omega(\xi) \) satisfies (9) then by the standard coupling proof of the ergodic theorem for Markov chains [8] we have that the series converges absolutely, moreover we can easily show that (7) is also true.

We must note that although \( \theta(\xi) \) is (by assumption) bounded this is not necessarily the case for \( \omega(\xi) \), which can be unbounded even for countable state spaces. If \( \omega(\xi) \) turns out to be bounded (as in the case of finite state spaces) then we can apply Dynkin's identity and we obtain

\[
E(\xi_0 \omega(\xi_N)) - \omega(\xi_0) = E(\xi_0 \sum_{n=1}^{N} P(\omega(\xi_{n-1}) - \omega(\xi_n)))
\]

(13)

Now using Doob's optional sampling theorem we can write

\[
E(\xi_0 \sum_{n=1}^{N} \theta(\xi_n) - \sum_{n=1}^{N} P(\theta(\xi_{n-1}) - \theta(\xi_n))) = 0
\]

(14)

Substituting (7) and (14) into (13) one recovers the proposed extension.

3. Study of the correction term

The generalized form of Wald's lemma reduces to the usual form (1) when \( \omega(\xi) = 0 \). From Lemma 2.3 we can see that this can happen if and only if \( P\theta = \pi\theta \) or equivalently \( P[\theta - \pi\theta] = 0 \).
with the last equality true because the event positive $M$ has finite expectation if Theorem 1.1 and such that 

$$\bigcup \{n \in \mathbb{N} : V_n \neq \emptyset \}$$

is a special combination of Markov processes and nonlinearities. Notice that if $V$ and using the Markov property of the process, we have for $n \geq 1$

$$V_n = \int_0^{\Pi_n} \omega(\xi) \, d\theta(\xi)$$

The stopping time $\Pi_n$ is i.i.d. Equation (15) may also hold, except in these two cases, for condition (15) is equivalent to saying that this operator has a linear operator from $L^m \rightarrow L^n$. Moreover, if $\Pi_n = \infty$ becomes negligible as $n \rightarrow \infty$, then for any function $\omega(\xi)$, we also have

$$\lim_{n \rightarrow \infty} \sum_{\xi \in F_n} \omega(\xi) = 0$$

where $F_n$ is the set of all functions $\omega(\xi)$ that belongs to this null space. Therefore, if $\Pi_n = \infty$ or the set $\bigcup \{n \in \mathbb{N} : V_n \neq \emptyset \}$ is a constant but also for any function $\omega(\xi)$, condition (15) is valid.

Proof. In Lemma 2.1 we have seen that for any $\rho \in \mathbb{R}$, $V_n = \int_0^{\Pi_n} \omega(\xi) \, d\theta(\xi)$, $\Pi_n \leq M$ is an integer greater than zero and let $\Pi_n = \infty$

where the stopping time $\Pi_n$ is infinite for the case where $\Pi_n < 1$. Applying Lemma 2.1 for $\Pi_n = \infty$, we have

$$\lim_{n \rightarrow \infty} \sum_{\xi \in F_n} \omega(\xi) = 0$$

In order to prove (16), it suffices to show that

$$\lim_{n \rightarrow \infty} \sum_{\xi \in F_n} \omega(\xi) = 0$$

...
Extension of Wald's first lemma to Markov processes

where

\[ R' = \max\{R, 1\} \]

Using Lemma 2.2 to bound the last sum, then dividing by \( E_N \) and taking the limit as \( E_N \to \infty \), we conclude that

\[
\lim \sup_{E_N \to \infty} \frac{E(V(\xi_N))}{E_N} \leq \pi \left[ V_1 l_{CM} + \rho m Rb \right] + \lambda.
\]

Since \( \pi V < \infty \) and \( 0 < \rho < 1 \), the right-hand side of the previous inequality can become arbitrarily small by tending \( M \) and \( m \) to infinity. This concludes the proof.

4. Examples

In this section we are going to present several examples and identify explicitly all the terms entering in the generalized form of Wald's lemma. For every example, in order to apply our results, we also need to show the existence of a function \( V(\xi) \) that satisfies the drift condition (5). At this point we must stress that the drift condition does not have a unique solution, and the different solutions are not necessarily equivalent, in the sense that they produce the same space of functions \( L_\infty \). Consequently if we are interested in applying our results to some given function \( \theta(\xi) \), it is clear that it suffices to find one solution to the drift condition such that \( \theta(\xi) \in L_\infty \).

4.1. Finite state Markov chains

Let the chain \( \{\xi_n\} \) have \( K \) states and let \( P \) be the corresponding transition matrix with the \( i \)-th row of \( P \) denoting the transition probabilities of state \( i \). We know that any transition matrix \( P \) has a unit eigenvalue whereas all other eigenvalues have magnitude that cannot exceed unity. Let us assume that the unit eigenvalue is simple and that all other eigenvalues have magnitude strictly less than unity. Then the invariant probability vector \( \pi \) exists and it is the left eigenvector to the unit eigenvalue, \( J = [1 \cdots 1] \) being the corresponding right eigenvector.

We can now verify (by induction) that

\[
P^n - J \pi^T = (P - J \pi^T)^n.
\]

The matrix \( P - J \pi^T \) has the same eigenvalues as \( P \) except the unit eigenvalue that has become zero. This means that all eigenvalues of \( P - J \pi^T \) have a magnitude that is strictly smaller than unity. Since for any square matrix \( A \) we have \( \lim_{n \to \infty} n^{\frac{1}{2}} \|A^n\| = \max_i |a_i| \), where \( a_i \) represents the eigenvalues of \( A \) and \( \|\| \) represents any matrix norm (see for example [10, pp. 36–38]), we conclude that there exist \( R \) and \( 0 \leq \rho < 1 \) such that

\[
\|P^n - J \pi^T\| \leq \rho^n R.
\]

This in turn suggests [12, Theorem 16.0.2] that the chain is aperiodic and the drift condition is satisfied by an everywhere bounded function \( V(\xi) \). Finally, as we stated in Section 1, the chain is also \( \pi \)-irreducible.

For this example, any function \( \theta(\xi) \) can in fact be regarded as a vector \( \theta \) of length \( K \) and the space \( L_\infty \) coincides with the space \( R^K \). To find the two quantities entering in the generalized form of the lemma, notice first that the expectation of \( \theta \) under the invariant measure is simply \( \pi^T \theta \). To find the vector \( \omega \) we have the following linear system of equations with respect to \( \omega \) (which is the analogue of the Poisson equation (7) and the constraint (8)),

\[
(P - I)\omega = -(P - J \pi^T)\theta, \quad \pi^T \omega = 0.
\]

It should also be noted that a sufficient condition for the assumption regarding the eigenvalues of the matrix \( P \) to hold is that there exists \( m \in \mathbb{Z}^+ \) such that all the entries of the matrix \( P^m \) are positive (see [9, Chapter IV]).

To examine whether it is possible to have validity of Wald's lemma in its original form, notice that relation (15) takes the form

\[
P\left[\theta - (\pi^T \theta)J\right] = 0.
\]

This can be true if the matrix \( P \) has a zero eigenvalue.
we have that previous sections to any function \( \theta(\xi) \). Notice that we need to satisfy (6) only for \( \{\omega(\zeta) \} \), and consequently its corresponding function \( \pi \). Taking conditional expectation we then conclude that for small enough \( E \) with compact sets being small sets (see [12, Chapter 16]).

With function \( \lambda \), we can now show that \( V \) is irreducible and aperiodic with \( E \leq 1 \). Indeed notice that by using Hölder's inequality we have \( \pi \theta(\zeta) \). From this last relation we also note that \( C \) satisfies the drift condition provided that \( \lambda \), with function \( \sum \pi \).

For this example it seems more appropriate to use \( \pi \) and consider \( \{\omega(\zeta)\} \) has this form then it satisfies (6) is satisfied for \( \{\pi \theta(\zeta)\} \).
such that the pair \( \omega(\xi) \) and, more generally, to processes of the form \( \xi_\theta(\xi) \) is a polynomial then the correction term cannot be zero (when \( \omega(\xi) \) is a constant). This is because the expectation of a randomly stopped log-likelihood ratio function. Specifically, see whether a density, \( \pi_\theta \), is available, then we can define polynomials \( s_j \) and thus we have identified both quantities entering the generalized form of Wald's lemma. To avoid giving any further details, selected \( \lambda, \pi_\theta \) is true for example for any 1 \( \leq \) \( p \) such that Wald's lemma is not as straightforward as it was in the previous examples, although an interesting class of nonlinearities for which this is possible is the class of polynomials. Notice that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is Gaussian). If such a condition holds then any polynomial \( j \) \( \leq \) \( n \) can then be written as a linear combination of the polynomials \( s_j \). Due to (19) we have

\[
\sum_{j=0}^{\infty} a_j s_j \theta(\xi) = \theta(\xi)
\]

Due to (20) we have

\[
E w_1 \log L(\xi) = \sum_{j=0}^{\infty} \frac{a_j}{j+1} s_j \theta(\xi) \theta(\xi)
\]

It must be noted that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is a vector of i.i.d. random variables and \( \theta(\xi) \) is a vector of constants). This is possible only when \( \omega(\xi) \) is Gaussian. If such a condition holds then we can find a constant \( \delta > 0 \) such that Wald's lemma is valid in its original form, notice that if \( \omega(\xi) \) is a polynomial this is possible is the class of polynomials. Notice that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is Gaussian). If such a condition holds then any polynomial \( j \) \( \leq \) \( n \) can then be written as a linear combination of the polynomials \( s_j \). Due to (19) we have

\[
\sum_{j=0}^{\infty} a_j s_j \theta(\xi) = \theta(\xi)
\]

Due to (20) we have

\[
E w_1 \log L(\xi) = \sum_{j=0}^{\infty} \frac{a_j}{j+1} s_j \theta(\xi) \theta(\xi)
\]

It must be noted that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is a vector of i.i.d. random variables and \( \theta(\xi) \) is a vector of constants). This is possible only when \( \omega(\xi) \) is Gaussian. If such a condition holds then we can find a constant \( \delta > 0 \) such that Wald's lemma is valid in its original form, notice that if \( \omega(\xi) \) is a polynomial this is possible is the class of polynomials. Notice that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is Gaussian). If such a condition holds then any polynomial \( j \) \( \leq \) \( n \) can then be written as a linear combination of the polynomials \( s_j \). Due to (19) we have

\[
\sum_{j=0}^{\infty} a_j s_j \theta(\xi) = \theta(\xi)
\]

Due to (20) we have

\[
E w_1 \log L(\xi) = \sum_{j=0}^{\infty} \frac{a_j}{j+1} s_j \theta(\xi) \theta(\xi)
\]

It must be noted that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is a vector of i.i.d. random variables and \( \theta(\xi) \) is a vector of constants). This is possible only when \( \omega(\xi) \) is Gaussian. If such a condition holds then we can find a constant \( \delta > 0 \) such that Wald's lemma is valid in its original form, notice that if \( \omega(\xi) \) is a polynomial this is possible is the class of polynomials. Notice that the computation of the coefficients of each polynomial involves the solution of a linear system of equations (when \( \omega(\xi) \) is Gaussian). If such a condition holds then any polynomial \( j \) \( \leq \) \( n \) can then be written as a linear combination of the polynomials \( s_j \). Due to (19) we have

\[
\sum_{j=0}^{\infty} a_j s_j \theta(\xi) = \theta(\xi)
\]

Due to (20) we have

\[
E w_1 \log L(\xi) = \sum_{j=0}^{\infty} \frac{a_j}{j+1} s_j \theta(\xi) \theta(\xi)
\]
Nikodym derivative has the form

\[ \tilde{\omega}(\xi) \]

To solve the Poisson integral equation for \( \tilde{\omega}(\xi) \)

\[ \theta(\xi) \]

The final correction term that enters in the generalized identity is, as we stated previously,

\[ \sigma | \begin{array}{c} \end{array} \]

\[ \lim_{n \to \infty} \left( \begin{array}{c} \end{array} \right) \]

\[ \beta \]

However, we wish to propose a slightly different approach that

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]

\[ \alpha \]
Acknowledgement
We would like to thank an anonymous reviewer for pointing out the relation of the proposed extension with Dynkin's identity.

References