Invariant Measures for Markov Chains with no Irreducibility Assumptions

R. L. Tweedie

Dedicated to Joe Gani in gratitude for his advice, encouragement and support

Abstract

Foster's criterion for positive recurrence of irreducible countable space Markov chains is one of the oldest tools in applied probability theory. In various papers in JAP and AAP it has been shown that, under extensions of irreducibility such as $\phi$-irreducibility, analogues of and generalizations of Foster's criterion give conditions for the existence of an invariant measure $\pi$ for general space chains, and for $\pi$ to have a finite $f$-moment $\int \pi(dy) f(y)$, where $f$ is a general function. In the case $f = 1$ these cover the question of finiteness of $\pi$ itself.

In this paper we show that the same conditions imply the same conclusions without any irreducibility assumptions; Foster's criterion forces sufficient and appropriate regularity on the space automatically. The proofs involve detailed consideration of the structure of the minimal subinvariant measures of the chain.

The results are applied to random coefficient autoregressive processes in order to illustrate the need to remove irreducibility conditions if possible.

Foster's CRITERIA; RECURRENCE; ERGODICITY; TIME SERIES

1. Introduction

It is a privilege to be asked to help celebrate the 25th anniversary of the Applied Probability Trust. Largely as a consequence of the existence of the Journal of Applied Probability (JAP) and Advances in Applied Probability (AAP), applied probability has flourished over the past 25 years. Many authors have been encouraged, as I was, by the publication in JAP of their first paper, and I regularly discover the bulk of my reference lists are from JAP and AAP: this is not the exact equivalent of a Science Citation Index search, but it indicates at least to me the continuing importance of these journals in my professional existence.

I wish the Applied Probability Trust, and Joe Gani especially, a long and continuing fruitful existence. This paper is dedicated to Joe Gani in acknowledgments.
ledgement of his own influence on my development, both personally in CSIRO and SIROMATH, and professionally through the example of his work in applied probability and the Trust itself.

The work in this paper is, I hope, appropriate to the occasion, as it rounds off in a sense a series of papers stretching back almost to the beginning of modern applied probability. In 1951, in the discussion of Kendall [4], and in more detail in Foster [3], F. G. Foster developed a simple ‘drift condition’ criterion for irreducible Markov chains on countable spaces to be ergodic, or positive recurrent. These results have been widely used, especially in queueing and storage models. They have also been extended, generalised and rediscovered on several occasions: Section 8 of [12] contains a history as at 1975. In [13], which substantially improves [12], the role the ‘drift condition’ can play in proving existence of a finite invariant measure for Markov chains on more general spaces was explored. This was set in a context where certain irreducibility conditions hold which provide the same framework as ordinary irreducibility on countable spaces: namely, chains can be classified a priori as positive or null recurrent, or as transient. In [14], [10] the same type of result was extended to provide conditions for the invariant measure to have finite moments, again under suitable irreducibility assumptions.

In many cases these irreducibility conditions are not a problem, even on fairly general spaces. In [11] for example, it is shown that a certain amount of continuity in the transition functions of the chain imply at least a controlled form of irreducibility.

However, in practice it is often difficult to find satisfactory minimal (i.e. necessary and sufficient) conditions for irreducibility. This question is raised in [5], for example, in many modelling contexts. The time series models in [2], which we explore in Section 5 below, are also of this kind. In simple terms these models were shown in [2] to have a stationary version if the appropriate ‘drift conditions’ hold and if the distribution of the random ‘errors’ have some amount of positive density near zero. The latter condition is purely a convenience to force irreducibility and hence place the models in a context where general results hold: in principle one feels the result should hold in wider generality and the density assumption is in no sense necessary. We show this is indeed the case in Section 5.

The aim of this paper is to prove that the drift conditions (Conditions F and M below) have surprisingly wide applicability without any irreducibility assumptions.

We shall show that, essentially, the drift conditions in Condition F suffice for existence of finite invariant measures, and those in Condition M for the existence of invariant measures with finite moments. Extra structural assumptions, such as the $\phi$-irreducibility in [7], are only needed for ensuring the
Invariant measures for Markov chains

convergence of the $n$-step transition probabilities to the finite invariant measure. Our results do not address this deeper level of complexity.

Our context is as follows. Let $(S, \mathcal{F})$ be a measurable space, and let $\{X_n\}$ be a Markov chain with temporally homogeneous transition probabilities

$$P^n(x, A) = \mathbb{P}(X_n \in A \mid X_0 = x), \quad x \in S, \quad A \in \mathcal{F}.$$ 

We shall in Section 3 consider cases where $S$ admits a topology in which case $\mathcal{F}$ will be taken as the Borel $\sigma$-field. We do not however need such structure in general.

Our intention is to explore the consequences of assuming the following general form of Foster’s condition.

**Condition F.** There exists, for some set $A \in \mathcal{F}$, a non-negative function $g$ and an $\varepsilon > 0$ such that

\[
(F1) \quad \int_{A^c} P(x, dy)g(y) \leq g(x) - \varepsilon, \quad x \in A^c;
\]

and there further exists a (non-trivial) measure $\mu$ such that

\[
(F2) \quad \int_{A} \mu(dx) \left[ \int_{A^c} P(x, dy)g(y) \right] < \infty.
\]

When $S$ is countable and $A$ is a finite set, these conditions are known to imply positive recurrence for irreducible chains [3], [13]. Note that the form (F2) holds for any $\mu$ with $\mu(A) < \infty$ provided

\[
(F3) \quad \sup_{x \in A} \int_{A^c} P(x, dy)g(y) < \infty;
\]

this is the more common form usually checked.

Following [14], [10], we shall consider the following related conditions for finite moments.

**Condition M.** For a given non-negative function $f$ on $S$ there exists a set $A$ and a non-negative function $g$ such that

\[
(M1) \quad \int_{A^c} P(x, dy)g(y) \leq g(x) - f(x), \quad x \in A^c.
\]

Note that (F1) is (M1) in the particular case $f(x) = \varepsilon$ on $A^c$.

In the $\phi$-irreducible chain context, it is shown in [13] that when (F1) and (F3) hold, then under suitable conditions on $A$ the chain is positive recurrent; in [14] it is shown that in this positive recurrent context, if (M1) holds with
suitable conditions on \( A \), then

\[
\int_S \pi(dy)f(y) < \infty
\]

where \( \pi \) is the invariant probability measure for \( \{X_n\} \).

Conditions F and M have similar consequences in general, provided \( A \) is a suitable set. To define what we mean by 'suitable', we need the idea of subinvariant measures. We shall call a measure \( \mu \) on \( \mathcal{F} \) subinvariant for \( P \) if, for all sets \( B \in \mathcal{F} \)

\[
(1.1) \quad \mu(B) \geq \int_S \mu(dy)P(y, B).
\]

Such a measure will be called invariant if it satisfies (1.1) with equality for all \( B \in \mathcal{F} \). Conditions for the existence of subinvariant and invariant measures have been widely studied. A careful analysis of the structure of such measures enables us to prove the following result in Section 2.

**Theorem 1.** Suppose that \( \mu \) is a subinvariant measure and \( A \in \mathcal{F} \) is such that \( 0 < \mu(A) < \infty \). If (F1) holds for this \( A \) and some \( g \geq 0, \varepsilon > 0 \), then

(i) \( \mu \) is a \( \sigma \)-finite invariant measure for \( \{X_n\} \);

(ii) \( \mu \) is finite provided (F2) also holds for this \( A, \mu, g \);

(iii) \( \mu \) admits a finite \( f \)-moment, i.e.

\[
(1.2) \quad \int_S \mu(dy)f(y) < \infty,
\]

provided (M1) and (F2) hold and also \( \int_A \mu(dy)f(y) < \infty \).

Note that none of these results use any irreducibility assumptions whatsoever.

2. The structure of subinvariant measures

To prove Theorem 1 we shall study in some detail the structure of subinvariant measures, and in particular we shall show that a variety of results hold under weaker conditions than they have previously been given. We define the usual ‘taboo’ probabilities \( _AP^n(x, B) \) by \( _AP^1(x, B) = P(x, B) \) and

\[
_AP^n(x, B) = \int_{_A} A P^{n-1}(x, dy)P(y, B);
\]

these have the well-known interpretation

\[
_AP^n(x, B) = P(X_n \in B, \tau_A \geq n \mid X_0 = x),
\]
where \( \tau_A = \inf(n \geq 1 : X_n \in A) \). If we write
\[
_A G(x, B) = \sum_{n=1}^{\infty} A P^n(x, B),
\]
then
\[
_A G(x, A) = P(\tau_A < \infty \mid X_0 = x).
\]
Assume now that \( \mu \) is subinvariant, and let \( A \) be such that \( 0 < \mu(A) < \infty \). Define the measure \( \mu^* \) by
\[
(2.1) \quad \mu^*(B) = \int_A \mu(dy) A G(y, B), \quad B \in \mathcal{F}.
\]

**Lemma 1.** For all \( B \in \mathcal{F} \), \( \mu(B) \geq \mu^*(B) \).

**Proof.** By induction, we have that for any \( n \)
\[
\mu(B) \geq \sum_{m=1}^{n} \int_A \mu(dy) A P^n(y, B) + \int_{A^c} \mu(dy) A P^n(y, B)
\]
and the result follows by letting \( n \to \infty \); see [7], p. 33 and note that the proof follows for subinvariant as well as invariant measures.

**Lemma 2.** Suppose \( _A G(x, A) = 1 \) for \( \mu \)-almost all \( x \in A \). Then \( \mu = \mu^* \) on \( A \).

**Proof.** Let \( B \subseteq A, B \in \mathcal{F} \). Then from Lemma 1,
\[
\mu(A) = \mu(B) + \mu(A \setminus B) \\
\geq \mu^*(B) + \mu^*(A \setminus B) \\
= \int_A \mu(dy) A G(y, A) \\
= \mu(A).
\]
Hence the inequality \( \mu(B) \geq \mu^*(B) \) must be an inequality for all \( B \subseteq A \).

**Lemma 3.** Suppose that \( _A G(x, A) = 1 \) for \( \mu \)-almost all \( x \in A \). Then \( \mu^* \) is an invariant \( \sigma \)-finite measure for \( P \) on \( S \).

**Proof.** By definition and Lemma 2, for \( B \in \mathcal{F} \)
\[
\mu^*(B) = \int_A \mu(dy) A G(y, B)
\]
\[
(2.2) \quad = \int_A \mu^*(dy) A G(y, B).
\]
In particular, in the terminology of [7], the measure \( \mu^* \) is invariant for the
process on $A$ with transition law $\mathcal{A}G(y, B)$, and the proof below is then similar to that in [7], p. 32.

First, for any $B \in \mathcal{F}$, from (2.2),

$$
\int_S \mu^*(dy)P(y, B) = \int_A \mu^*(dy)P(y, B)
$$

$$
+ \int_{A^c} \left[ \int_A \mu^*(dw)\mathcal{A}G(w, dy) \right]P(y, B)
$$

$$
= \int_A \mu^*(dy) \left[ P(y, B) + \sum_2^{\infty} \mathcal{A}P^n(y, B) \right]
$$

$$
= \mu^*(B)
$$

and so $\mu^*$ is invariant for $P$ on $S$.

Now write $\tilde{A}(n, j) = \{ y \in S : P^n(y, A) \in ((j + 1)^{-1}, j^{-1}] \}$ and let $\tilde{A} = \bigcup_{n,j} \tilde{A}(n, j)$. By invariance,

$$
\mu^*(A) \geq \int_{\tilde{A}(n,j)} \mu^*(dy)P^n(y, A)
$$

$$
\geq (j + 1)^{-1} \mu^*(\tilde{A}(n,j))
$$

and so $\mu^*$ is $\sigma$-finite on $\tilde{A}$. However, if $A^0 = \tilde{A}^c$, then for all $y \in A^0$, $\mathbb{P}(\tau_A < \infty \mid X_0 = y) = 0$. Since $\mathcal{A}G(x, A) = 1$ for $\mu$-almost all $x \in A$, we must have $\mathcal{A}G(x, A^0) = 0$ for $\mu$-almost all $x \in A$, and by construction $\mu^*(A^0) = 0$. Thus $\mu^*$ is $\sigma$-finite on $S$ as required.

We bring these results together to prove the following result.

**Proposition 1.** Suppose $\mu$ is a subinvariant measure for $P$, and that $0 < \mu(A) < \infty$ for some $A$ such that $\mathcal{A}G(x, A) \equiv 1$ for $\mu$-almost all $x \in A$. Then $\mu$ restricted to $\tilde{A}$ is a $\sigma$-finite measure, invariant for $P$ and satisfying

(2.3) $$
\mu(B) = \int_A \mu(dy)\mathcal{A}G(y, B), \quad B \in \mathcal{F}, \quad B \subseteq \tilde{A}.
$$

**Proof.** By the minimality result of Lemma 1, the invariance of $\mu^*$ in Lemma 3 and the equality of $\mu$ and $\mu^*$ on $A$ in Lemma 2, for any $n$

$$
\mu(A) \geq \int_S \mu(dy)P^n(y, A) \geq \int_S \mu^*(dy)P^n(y, A) = \mu^*(A) = \mu(A).
$$

Hence $\mu = \mu^*$ on all of $\tilde{A}$ and the result follows.

Since $A^0$ is stochastically closed under the assumptions of Proposition 1, it is clear that we cannot say anything further about the behaviour of $\mu$ on $A^0$.

There are several results similar to the invariance of $\mu$ proved in Proposition
Invariant measures for Markov chains

1 in the literature (see e.g. [8], [9]) but none of them exploit as fully the identification of \( \mu \) with the minimal \( \mu^* \) given by (2.3) as we now do in the following proof.

**Proof of Theorem 1.** From Theorem 6.1 of [13], (F1) ensures

\[
AG(x, A) = 1, \quad x \in A^c,
\]

and in fact it ensures the much stronger result

\[
1 + AG(x, A^c) \leq g(x)/\epsilon, \quad x \in A^c.
\]

From (2.4) we have that also for \( x \in A, AG(x, A) = 1 \), so that \( A^0 \) is empty when (F1) holds. Hence \( \mu \) is \( \sigma \)-finite and invariant from Proposition 1, which proves Theorem 1(i).

Now from (2.3) and (2.5) we have

\[
\mu(A^c) = \int_A \mu(dy)AG(y, A^c)
\]

\[
= \int_A \mu(dy) \int_{A^c} P(y, dw)[1 + AG(w, A^c)]
\]

\[
\leq \int_A \mu(dy) \left[ \int_{A^c} P(y, dw)g(w)/\epsilon \right]
\]

which is finite if and only if (F2) holds. Since \( \mu(A) < \infty \) by hypothesis, \( \mu \) is finite as claimed in Theorem 1(ii).

Finally, to prove Theorem 1(iii), we note that the proof of Theorem 1 of [14] holds once we have the representation (2.3) for \( \mu \).

Note that the extra condition (F2) is needed for the finiteness result in Theorem 1(ii): it is easy to construct irreducible examples on \( S = \{0, 1, 2, \cdots \} \) with \( \mathbb{E}(\tau_0 \mid X_0 = j) < \infty \) for all \( j > 0 \) but with \( \mathbb{E}(\tau_0 \mid X_0 = 0) = \infty \).

### 3. Results for weak Feller chains

Many of the applications of these results in, say, operations research or time series examples are in the context of weak Feller chains. We are able to use known results for such chains to prove the following stronger result from Theorem 1.

**Theorem 2.** Suppose that \( S \) is a locally compact separable metric space with \( \mathcal{F} \) the Borel \( \sigma \)-field on \( S \). Suppose also that \( \{X_n\} \) is weak Feller; that is, the operator \( P(x, \cdot) \) maps bounded continuous functions into bounded continuous functions (see e.g. [9]).

If (F1) holds with \( A \) compact, then there exists a \( \sigma \)-finite invariant measure \( \mu \) for \( P \) with \( 0 < \mu(A) < \infty \), and if (F2) also holds then \( \mu \) is finite.
Proof. Lin has shown (see [9], p. 182) that under the topological conditions above there exists a non-trivial subinvariant measure $\mu$, which is finite on compact sets, provided that there exists a non-negative continuous function $h$ on $S$ with compact support satisfying

$$\sum_{n=1}^{\infty} \int_S P^n(x, dy)h(y) > 0, \quad x \in S.$$  

Let $h$ be any continuous function with compact support such that $h(y) > 0, y \in A$. Since (F1) holds then as in the proof of Theorem 1, $A = S$ and so (3.1) holds. Hence we have a subinvariant measure $\mu$ with $\mu(A) < \infty$, and since $\mu(A) > (j + 1)^{-1}\mu(A(n, j))$ for each $n, j$ and $\mu$ is non-trivial, we also have $\mu(A) > 0$. The invariance and finiteness claims then follow from Theorem 1.

Corollary. For weak Feller chains, if (F1) and (F2) hold then there is a measure $\mu > 0$ such that for $\mu$-almost all $x \in S$, whenever $M$ is a neighbourhood of $x$

$$\mathbb{P}(X_n \in M \text{ infinitely often} \mid X_0 = x) = 1.$$  

Proof. This follows from Lemma 8 of [9].

This corollary shows that, for weak Feller chains, the consequences of Foster's criteria (F1) and (F2) holding for $A$ are not only that $A$ is then a recurrent set in the sense that $\mathbb{P}(\tau_A < \infty \mid X_0 = x) \equiv 1$ for all $x$, but that $\mu$-almost all points in the space have the same property for all their neighbourhoods.

We can apply Theorem 2 to the bulk of the examples of [5], or to the time series results in [2], and see that the assumption of $\phi$-irreducibility employed there for existence of moments and of finite measures is superfluous.

4. Chains with strongly continuous components

Continuity of a different type is considered in [11]. Suppose that $S$ has a topology which is $T_1$ and countably generated, and that $\mathcal{F}$ contains all the open sets. We say that $\{X_n\}$ has an everywhere non-trivial continuous component if, for every $x$ and $A$, $\lambda G(x, A) > T(x, A)$ where $T$ is a (substochastic) transition operator such that, for each $A \in \mathcal{F}$, $T(\cdot, A)$ is a lower semi-continuous function on $S$; and $T(x, S) > 0$ for every $x \in S$. If $T$ can be chosen as $P$ itself the chain is called strong Feller. In [5] some of the more practical consequences of assuming $\{X_n\}$ to have a continuous component are explored, and we now look at the consequences of assuming (F1) and the related conditions in this context.
3. If \( \{X_n\} \) has an everywhere non-trivial continuous component and (F1) holds for some compact \( A \) and some \( g \) then there exists a \( \sigma \)-finite invariant measure \( \mu \). If also (F2) holds for this \( \mu \), then \( \mu \) is finite and the general \( f \)-moment (1.2) is finite if (M1) also holds.

Proof. From Theorem 6.1 of [11], there is a decomposition of \( S \) as

\[
S = \sum_i H_i + E
\]

where each of the \( H_i \) is a Harris set and \( E \) is not properly essential. We first note that we cannot have \( A \subseteq E \) when (F1) holds for \( A \); for (F1), as in (2.4), implies \( P(\tau_A < \infty \mid X_0 = x) = 1 \), but \( P(y, E) = 0 \) for any \( y \in \sum H_i \). Hence \( \sum H_i \) is non-empty. Moreover, for any \( H_i, H_j \) such that \( i \neq j \), \( P(x, H_j) = 0 \) for \( x \in H_i \), so \( A \cap H_i \) is non-empty for each of the Harris sets \( H_i \), and in fact \( A G(x, A \cap H_i) = 1, x \in H_i \).

This ensures that an invariant measure \( \mu \) (the invariant measure \( \pi_i \) for a non-trivial \( H_i \)) exists, and \( \mu(A) = \pi_i(A \cap H_i) > 0 \). By Proposition 3.4 of [11] \( \pi_i(A) < \infty \). The finiteness of any such \( \pi_i \) under (F2), and the finite moment property under (M1), then follow directly from Theorem 1.

5. Random coefficient autoregressive models

Results of the kind above have been used increasingly in time series analysis: see not only [2] but also [1]. We apply our results here to extend [2]. We consider (3.1) of [2], namely a process of \( q \)-vectors \( \{X_t\} \) with

\[
X_t = (\theta + \Gamma_t)X_{t-1} + R_t, \quad t = 0, 1, \ldots
\]

where \( \theta \) is a \( q \times q \) non-random matrix, \( \Gamma_t \) is a sequence of random \( q \times q \) matrices, and \( R_t \) is a sequence of \( q \)-vectors. As defined in [2], this scheme also incorporates vector-valued and \( n \)th-order autoregressions.

We make the following assumptions.

Assumption 1. The sequences \( \{\Gamma_t\}, \{R_t\} \) are independent and identically distributed variables and are independent of each other.

Assumption 2. The following expectations exist with the prescribed values:

\[
\mathbb{E}(R_t) = 0, \quad \mathbb{E}(R_t R'_t) = G \quad (q \times q)
\]

\[
\mathbb{E}(\Gamma_t) = 0 \quad (q \times q) \quad \mathbb{E}(\Gamma_t \otimes \Gamma_t) = C \quad (q^2 \times q^2),
\]

where \( \otimes \) is the Kronecker product.

In [2], an assumption similar to requiring \( R_t \) to have a positive density over a ball around zero was shown to imply \( \phi \)-irreducibility. This is artificial in this
setting, as also are conditions in [6] designed to force other types of regularity similar to irreducibility.

We can avoid these by the use of Theorem 2.

**Theorem 4.** (a) If \( \{X_t\} \) is as in (5.1), and if
(i) Assumptions 1 and 2 hold
(ii) the eigenvalues of \( \theta \otimes \theta + C \) have moduli less than unity
then \( \{X_t\} \) has a finite invariant measure \( \pi \), and

\[
E_\pi(X_t X'_t) < \infty;
\]
that is, the process can be assumed second-order stationary.

(b) If further \( H_t = \theta + \Gamma_t \), and
(iii) the eigenvalues of \( D = \mathbb{E}(H_t^{\otimes 2m}) \) have modulus less than unity
(iv) \( E_\pi(\|R_t\|^{2m}) < \infty \)
then we also have

\[
E_\pi(\|X_t\|^{2m}) < \infty.
\]

**Proof.** Parts (a) and (b) of the proof of Theorem 4 of [2] show that \( \{X_t\} \) is weak Feller, and that (F1) holds for some compact \( A \), whilst the proof of Theorem 4 of [2] further shows that (F2) and also (M1) hold for \( g(x) = 1 + x'Vx \), where \( \text{vec}(V) = (1 - \theta' \otimes \theta' - C')^{-1} \text{vec}(W) \) for any \( q \times q \) positive definite \( W \) and for \( f(x) = \delta g(x) \), some \( \delta < 1 \).

From Theorem 2 above there then exists a finite invariant measure \( \pi \), and from Theorem 1(iii) we have (5.2).

We can prove (b) similarly by using the proof of Theorem 5 of [2] and Theorem 1(iii) above.

We note that the existence of these higher moments is of more than curiosity value. They are of importance in ensuring the validity of a number of strong laws (consistency results) and central limit theorems for estimators in these models, and it is to my mind pleasing to see strong results for such seemingly detailed models emerge within such a general applied probability framework.

**Acknowledgement**

I am grateful to Paul Feigin for asking why we really needed irreducibility for these results to hold.

**References**


Invariant measures for Markov chains


