Standard deviation of recurrence times for piecewise linear transformations

Mimoon Ismael  
*University of Anbar, mii998@uowmail.edu.au*

Rodney Nillsen  
*University of Wollongong, nillsen@uow.edu.au*

Graham H. Williams  
*University of Wollongong, ghw@uow.edu.au*

Follow this and additional works at: [https://ro.uow.edu.au/eispapers1](https://ro.uow.edu.au/eispapers1)
Standard deviation of recurrence times for piecewise linear transformations

Abstract
This paper is concerned with dynamical systems of the form \((X, f)\) where \(X\) is a bounded interval and \(f\) comes from a class of measure-preserving, piecewise linear transformations on \(X\). If \(A \subseteq X\) is a Borel set and \(x \in A\), the Poincaré recurrence time of \(x\) relative to \(A\) is defined to be the minimum of \(\{n : n \in \mathbb{N} \text{ and } f^n(x) \in A\}\), if the minimum exists, and \(\infty\) otherwise. The mean of the recurrence time is finite and is given by Kac's recurrence formula. In general, the standard deviation of the recurrence times need not be finite but, for the systems considered here, a bound for the standard deviation is derived.

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers1/320
STANDARD DEVIATION OF RECURRENCE TIMES FOR
PIECEWISE LINEAR TRANSFORMATIONS

MIMoon ISMAEL
PURE MATHEMATICS DEPARTMENT
COLLEGE OF EDUCATION AND THE PURE SCIENCES
UNIVERSITY OF ANBAR, AL-ANBAR
REPUBLIC OF IRAQ

AND

RODNEY NILLSEN AND GRAHAM WILLIAMS
SCHOOL OF MATHEMATICS AND APPLIED STATISTICS
UNIVERSITY OF WOLLONGONG
NORTHFIELDS AVENUE, WOLLONGONG
NEW SOUTH WALES, 2522
AUSTRALIA

Abstract. This paper is concerned with dynamical systems of
the form \((X, f)\), where \(X\) is a bounded interval and \(f\) comes from a class of
measure-preserving, piecewise linear transformations on \(X\). If \(A \subseteq X\) is a
Borel set and \(x \in A\), the Poincaré recurrence time of \(x\) relative to \(A\) is defined
to be the minimum of \(\{n : n \in \mathbb{N} \text{ and } f^n(x) \in A\}\), if the minimum exists,
and \(\infty\) otherwise. The mean of the recurrence time is finite and is given by
Kac’s recurrence formula. In general, the standard deviation of the recurrence
times need not be finite but, for the systems considered here, a bound for the
standard deviation is derived.

Keywords: Dynamical systems; piecewise linear functions; recurrence times;
standard deviation, Kac’s formula

AMS Subject Classification: Primary 37C45, 37E05; Secondary 26A18, 37A05,
37B20

1. Introduction

A dynamical system is a set \(S\) together with a transformation \(f : S \rightarrow S\).
Letting \(\circ\) denote the composition of functions, we use the notation that \(f^1 = f\),
\(f^2 = f \circ f\), \(f^3 = f \circ f \circ f\), and so on. We denote the set \(\{1, 2, \ldots\}\) by \(\mathbb{N}\). Now if we
have a dynamical system \((S, f)\), let \(A \subseteq S\), let \(x \in A\) and assume that \(f^n(x) \in A\)
for some \(n \in \mathbb{N}\). Then, we put
\[\Theta_A(x) = \min\{n : n \in \mathbb{N} \text{ and } f^n(x) \in A\}.
\]
If \(x \in A\) it may happen that \(f^n(x) \notin A\) for all \(n \in \mathbb{N}\), in which case we put
\(\Theta_A(x) = \infty\). Thus, \(\Theta_A : A \rightarrow \mathbb{N} \cup \{\infty\}\) and we call \(\Theta_A(x)\) the recurrence time
of \(x\) relative to \(A\) and \((S, f)\), or simply the recurrence time if \((S, f)\) and \(A\) are
understood. See Brown [3] and Petersen [11], for example, for a discussion of this
notion.
We consider the case when \( S \) is a bounded interval \( X \) of length 1. Let \( B \) denote the \( \sigma \)-algebra of Borel subsets of \( X \), and let \( \mu \) denote the usual Lebesgue measure on \( B \). A formulation of the Recurrence Theorem of Henri Poincaré says that if \( \mu \) preserves the measure of the Borel sets under the action of a transformation \( f \) on \( X \), then \( \Theta_A(x) < \infty \) for \( \mu \)-almost all points of \( X \). In ergodic systems, Kac [7] showed that the average value of the recurrence time is \( 1/\mu(A) \). The question arises as to the variation, predictability and uncertainty of the recurrence times. A formula for the moments of recurrence times in the context of stationary stochastic processes was derived by Blum and Rosenblatt [2], and a formula for the standard deviation of the recurrence times was derived by Kasteleyn [8]. A discussion of some of these matters is also in Ismael [5] and Chapter 4 of [10].

2. Estimating the Standard Deviation for the Baker’s Transformations

We now introduce some further notations. Let \( X \) be an interval in \( \mathbb{R} \) of length 1. Let \( B \) denote the \( \sigma \)-algebra of Borel subsets of \( X \), and let \( \mu \) denote the usual Lebesgue measure on \( B \). Let \( I = \mathbb{N} \) or let \( I \) denote an interval in \( \mathbb{N} \) of the form \( \{1, 2, \ldots, r\} \) with \( r \geq 2 \). For each \( j \in I \) let \( X_j \) be a subinterval of \([0, 1]\) of positive length such that the sets in the family \( (X_j)_{j \in I} \) form a partition of \( X \). That is, \( X_i \cap X_j = \emptyset \) when \( i \neq j \) and \( \bigcup_{j \in I} X_j = X \). Such a partition we call a standard partition and may be denoted by \( \mathcal{P} \). Note that a standard partition has at least two elements so, as \( \sum_{j \in I} \mu(X_j) = 1 \), for a standard partition the set \( \{\mu(X_j) : j \in I\} \) has a maximum value and

\[
\max\{\mu(X_j) : j \in I\} < 1. \tag{2.1}
\]

We call a transformation \( f : X \rightarrow X \) piecewise linear if there is a standard partition \( \mathcal{P} = (X_j)_{j \in I} \) such that \( f \) is linear on each interval \( X_j \in \mathcal{P} \), by which is meant that there are \( a_j, b_j \in \mathbb{R} \) such that \( f(x) = a_jx + b_j \) for all \( x \in X_j \). If, in addition, the range of \( f \) on each interval \( X_j \) is either \( X \) or \( X \) less one or two endpoints, then \( f \) is called a (generalised) baker’s transformation relative to \( \mathcal{P} \). A particular type of baker’s transformation is discussed in [12, page 77]. Note that a baker’s transformation is a piecewise linear function which is not one-to-one, but is one-to-one when restricted to each interval \( X_j \). A Baker’s transformation is Lebesgue measure preserving. Note that a piecewise linear function will be either increasing or decreasing on each interval \( X_j \) in the associated partition.

**Lemma 2.1.** Let \( f \) be a baker’s transformation on an interval \( X \) of length 1. Then, \( f \) is Lebesgue measure preserving on \( B \). That is, \( \mu(f^{-1}(A)) = \mu(A) \), for all \( A \in B \).

**Proof.** As \( f \) is linear on each interval in the standard partition associated with \( f \), it is easy to see that \( \mu(f^{-1}(A)) = \mu(A) \) for all subintervals \( A \) of \( X \). Let \( A = \{A : A \in B \text{ and } \mu(f^{-1}(A)) = \mu(A)\} \). Then \( A \) is a monotone class in the sense of [1, page 6] that contains all subintervals of \( X \). By the Monotone Class Lemma, see [1, pages 6-7] for example, \( A \) contains the Borel sets. \( \square \)

**Lemma 2.2.** Let \( X \) be an interval of length 1 and let \( f \) be a baker’s transformation relative to a standard partition \( (X_j)_{j \in I} \) of \( X \), as described above. Then, if \( n_1, n_2, \ldots, n_k \in I \), the set \( X_{n_1n_2\ldots n_k} \) is defined by

\[
X_{n_1n_2\ldots n_k} = \left\{ x : x \in X_{n_1}, f(x) \in X_{n_2}, f^2(x) \in X_{n_3}, \ldots, f^{k-1}(x) \in X_{n_k} \right\}.
\]
Then, $X_{n_1n_2...n_k}$ is a subinterval of $X$ having positive length, and the restriction of $f^k$ to $X_{n_1n_2...n_k}$ is a linear function whose range is either $X$, or $X$ less one or two endpoints.

Proof. When $k = 1$, the result is true from the definitions. So, we proceed by induction. Let $\text{int}(J)$ denote the interior of any interval $J$. Let $k \in \mathbb{N}$ and assume that the result holds for any choice of $n_1, n_2, \ldots, n_k \in I$. Now, let $n_1, \ldots, n_k, n_{k+1} \in I$ and observe that

$$X_{n_1n_2...n_kn_{k+1}} = \{x : x \in X_{n_1n_2...n_k} \text{ and } f^k(x) \in X_{n_k+1}\}.$$ 

It follows that $X_{n_1n_2...n_kn_{k+1}}$ is an interval of positive length because it is assumed that $f^k$ is linear on the interval $X_{n_1n_2...n_k}$ of positive length and that the range of $f^k$ on $X_{n_1n_2...n_k}$ is $X$ or $X$ less one or two endpoints. It is immediate from the definition that $f : X_{n_1n_2...n_kn_{k+1}} \to X_{n_2...n_kn_{k+1}}$, and $f^k$ is linear on $X_{n_2...n_kn_{k+1}}$ by the inductive assumption. Also, as $f$ is linear on $X_{n_1}$, $f$ is linear on the smaller set $X_{n_1n_2...n_kn_{k+1}}$. So we have

$$f : X_{n_1n_2...n_kn_{k+1}} \to X_{n_2...n_kn_{k+1}} \text{ and } f^k : X_{n_2...n_kn_{k+1}} \to X,$$

where both $f$ and $f^k$ are linear. Since the composition of linear functions is linear, it follows that $f^{k+1} = f^k \circ f$ is linear on $X_{n_1n_2...n_kn_{k+1}}$.

Now, as earlier, let $\text{int}(J)$ denote the interior of any sub-interval $J$ of $X$, and let $z \in \text{int}(X)$. By the inductive assumption, there is $y$ belonging to $\text{int}(X_{n_2n_3...n_{k+1}})$ such that $f^k(y) = z$. Then, as $f$ maps $X_{n_1}$ onto either $X$ or $X$ less one or two endpoints, there is $x \in \text{int}(X_{n_1})$ such that $f(x) = y$ and it follows that $x \in X_{n_1n_2...n_k}$. Also, $f^{k+1}(x) = f^k(f(x)) = f^k(y) = z$. We deduce that $f^{k+1}$ maps $\text{int}(X_{n_1n_2...n_k})$ onto $\text{int}(X)$. As $f^{k+1}$ is linear on $\text{int}(X_{n_1n_2...n_k})$, $f^{k+1}$ must map $\text{int}(X_{n_1n_2...n_k})$ onto $X$, or onto $X$ less one or two endpoints. Thus, assuming the Lemma holds for $k$, we have seen it also holds for $k + 1$, and the result follows by induction. \hfill \square

**Lemma 2.3.** Let $X$ be an interval of length 1 and let $f$ be a baker’s transformation relative to a standard partition $(X_j)_{j \in I}$ of $X$. Let $k \in \mathbb{N}$ and let $n_1, n_2, \ldots, n_k \in I$. Then,

$$\mu(X_{n_1n_2...n_k}) = \prod_{j=1}^{k} \mu(X_{n_j}).$$

Proof. If $k = 1$, the result is obviously true. So, assume that for some $k$, the result holds for any choice of $n_1, n_2, \ldots, n_k \in I$. Then

$$\mu(X_{n_1n_2...n_kn_{k+1}}) = \mu \left( \{x : x \in X_{n_1n_2...n_k} \text{ and } f^k(x) \in X_{n_{k+1}}\} \right),$$

$$= \mu(X_{n_1n_2...n_k}) \cdot \mu(X_{n_{k+1}}), \text{ by using Lemma 2.2,}$$

$$= \mu(X_{n_{k+1}}) \cdot \prod_{j=1}^{k} \mu(X_{n_j}),$$

as the result is assumed true for $k$,

$$= \prod_{j=1}^{k+1} \mu(X_{n_j}).$$
Thus, if the result is true for $k$, it is true for $k+1$. The result follows by induction. □

Lemma 2.4. Let $X$ be an interval of length 1 and let $f$ be a baker’s transformation relative to a standard partition $(X_j)_{j \in I}$ of $X$. Let $k, s \in \mathbb{N}$ and for each $j \in \{1, 2, \ldots, s\}$, suppose there are given $n_{j_1}, n_{j_2}, \ldots, n_{j_k} \in I$ and put

$$Y_j = X_{n_{j_1}n_{j_2}\ldots n_{j_k}}.$$

Then,

$$\mu \left( Y_1 \cap f^{-k}(Y_2) \cap f^{-2k}(Y_3) \cdots \cap f^{-(s-1)k}(Y_s) \right) = \prod_{j=1}^{s} \mu(Y_j).$$

Proof. Observe from the definitions that

$$Y_1 \cap f^{-k}(Y_2) \cap f^{-2k}(Y_3) \cdots \cap f^{-(s-1)k}(Y_s) = X_{n_{11} \ldots n_{1k}n_{21} \ldots n_{2k}n_{31} \ldots n_{3k} \ldots n_{s-1} \ldots n_{sk}}.$$

Then, by Lemma 2.3,

$$\mu(Y_1 \cap f^{-k}(Y_2) \cap \cdots \cap f^{-(s-1)k}(Y_s)) = \mu(X_{n_{11}}) \cdots \mu(X_{n_{1k}}) \cdots \mu(X_{n_{s1}}) \cdots \mu(X_{n_{sk}}) = \prod_{j=1}^{s} \mu(Y_j).$$

Lemma 2.5. Let $X$ be an interval of length 1 and let $f$ be a baker’s transformation relative to a standard partition $P = (X_j)_{j \in I}$ of $X$. Let $k \in \mathbb{N}$ and let $Z$ be a set that is a union of some family of sets in the partition $P_k$ of $X$. Then, for $\ell = 1, 2, \ldots,$

$$\mu \left( Z \cap f^{-k}(Z) \cap f^{-2k}(Z) \cap \cdots \cap f^{-(\ell-1)k}(Z) \right) = \mu(Z)^\ell.$$

Let $P = (X_j)_{j \in I}$ be a standard partition of an interval $X$ of length 1. Given $k \in \mathbb{N}$, make the definition that

$$P_k = \left\{ X_{n_1n_2\ldots n_k} : n_1, n_2, \ldots, n_k \in I \right\}.$$

Then $P_k$ is a partition of $X$ in that distinct sets of $P_k$ are disjoint and the union of the sets in $P_k$ is $X$. Lemma 2.2 shows that the sets in $P_k$ are intervals of positive length. In the case when $I$ has $r$ elements, $P_k$ consists of $r^k$ disjoint sets. If $\ell > k$, each set in $P_k$ is the union of $2^{\ell-k}$ sets in $P_\ell$, as expressed by the identity

$$X_{n_1n_2\ldots n_k} = \bigcup \left\{ X_{n_1n_2\ldots n_kn_{k+1}\ldots n_\ell} : n_{k+1}, n_{k+2}, \ldots, n_\ell \in I \right\}.$$
Proof. First, assume that $Z$ is a finite union of sets in $\mathcal{P}_k$. Then, there are disjoint sets $Y_1, Y_2, \ldots, Y_s$ in $\mathcal{P}_k$ such that $Z = \bigcup_{j=1}^s Y_j$. We have

$$
\mu\left(Z \cap f^{-k}(Z) \cap f^{-2k}(Z) \cap \cdots \cap f^{-(\ell-1)k}(Z)\right)
= \mu\left(\bigcup_{j=1}^s Y_j \cap f^{-k}(\bigcup_{j=1}^s Y_j) \cap f^{-2k}(\bigcup_{j=1}^s Y_j) \cap \cdots \cap f^{-(\ell-1)k}(\bigcup_{j=1}^s Y_j)\right)
= \mu\left(\bigcup_{j=1}^s Y_j \cap (\bigcup_{j=1}^s f^{-k}(Y_j)) \cap \cdots \cap (\bigcup_{j=1}^s f^{-(\ell-1)k}(Y_j))\right)
= \mu\left(\bigcup_{j_0,j_1,\ldots,j_{\ell-1} \in \{1,\ldots,s\}} Y_{j_0} \cap f^{-k}(Y_{j_1}) \cap \cdots \cap f^{-(\ell-1)k}(Y_{j_{\ell-1}})\right)
= \sum_{j_0,j_1,\ldots,j_{\ell-1} \in \{1,\ldots,s\}} \prod_{t=0}^{\ell-1} \mu(Y_{j_t}), \text{ by Lemma 2.4},
= \left(\sum_{j=1}^s \mu(Y_j)\right)^\ell
= \mu(Z)^\ell.
$$

(2.2)

Now, when $Z$ is an infinite union of sets in $\mathcal{P}_k$, there is an increasing sequence $(Z_n)$ of sets, each set $Z_n$ is a finite union of sets in $\mathcal{P}_k$, and $Z = \bigcup_{n=1}^\infty Z_n$. Using (2.2) for the sets $Z_n$, and the fact that $\mu$ is $\sigma$-additive on the Borel sets, if $k, \ell \in \mathbb{N}$ are given we now have

$$
\mu\left(Z \cap f^{-k}(Z) \cap f^{-2k}(Z) \cap \cdots \cap f^{-(\ell-1)k}(Z)\right)
= \lim_{n \to \infty} \mu\left(Z_n \cap f^{-k}(Z_n) \cap f^{-2k}(Z_n) \cap \cdots \cap f^{-(\ell-1)k}(Z_n)\right)
= \lim_{n \to \infty} \mu(Z_n)^\ell
= \mu(Z)^\ell.
$$

Lemma 2.6. Let $X$ be an interval of length 1, let $\mathcal{P} = (X_j)_{j \in I}$ be a standard partition of $X$ and put $\rho = \max\{\mu(X_j) : j \in I\}$. Then, for each $k \in \mathbb{N}$, $\mathcal{P}_k$ is a standard partition of $X$, and we have

$$
\max\{\mu(Y) : Y \in \mathcal{P}_k\} \leq \rho^k \text{ and } \lim_{k \to \infty} (\max\{\mu(Y) : Y \in \mathcal{P}_k\}) = 0.
$$

Proof. This is immediate from (2.1), and Lemmas 2.2 and 2.3. \hfill \Box

The following argument extends an idea in [9, pages 150-151] to prove ergodicity for the dynamical systems considered here. We use $A^c$ to denote the complement of a set $A$.

Lemma 2.7. Let $X$ be an interval of length 1 and let $f$ be a baker’s transformation relative to a standard partition $(X_j)_{j \in I}$ of $X$. Then if $A$ is a Borel subset of $X$ such that $f^{-1}(A) = A$, then $\mu(A) = 0$ or $\mu(A) = 1$. That is, $(X, f)$ is an ergodic system.
Proof. Let $A$ be a Borel subset of $X$ with $f^{-1}(A) = A$ and $0 \leq \mu(A) < 1$. Then, $0 < \mu(A^c) \leq 1$. Let $\varepsilon > 0$. Then, there is a closed subinterval $U$ of $X$ such that $\mu(U) > 0$ and $\mu(U \cap A^c) > (1 - \varepsilon/2)\mu(U)$. Now by Lemma 2.6 there is $k \in \mathbb{N}$ and a sequence $V_1, V_2, \ldots$ of disjoint sets in $\mathcal{P}_k$ such that
\begin{equation}
\sum_{n=1}^{\infty} \mu(A^c \cap V_n) = \mu \left( A^c \cap \left( \bigcup_{n=1}^{\infty} V_n \right) \right) = \mu \left( U \cap A^c \cap \left( \bigcup_{n=1}^{\infty} V_n \right) \right) = \mu(U \cap A^c) - \mu \left( U \cap A^c \cap \left( \bigcup_{n=1}^{\infty} V_n \right)^c \right) \geq \mu(U \cap A^c) - \mu \left( U \cap \left( \bigcup_{n=1}^{\infty} V_n \right)^c \right) > \left( 1 - \frac{\varepsilon}{2} \right) \mu(U) - \frac{\varepsilon \mu(U)}{2} = (1 - \varepsilon) \mu(U).
\end{equation}

Hence, we have
\begin{equation}
\sum_{n=1}^{\infty} \mu(V_n \cap A^c) > (1 - \varepsilon) \mu \left( \bigcup_{n=1}^{\infty} V_n \right) = (1 - \varepsilon) \sum_{n=1}^{\infty} \mu(V_n).
\end{equation}

By choosing a suitable $V_n$, and putting $W = V_n$, we deduce that there is $W \in \mathcal{P}_k$ such that
\begin{equation}
\mu(W \cap A^c) > (1 - \varepsilon) \mu(W).
\end{equation}

Now, by Lemma 2.2, $f^k : W \rightarrow [0, 1]$ is onto except maybe for one or two endpoints. Let $B = A^c$ except that the endpoints of $X$ that are in $A^c$, if any, are omitted. Put $C = \{ x : x \in W$ and $f^k(x) \in B \}$. Note that
\begin{equation}
\mu(B) = \mu(A^c) \text{ and } f^k(C) = B.
\end{equation}

Also, as $f^{-1}(A) = A$, $f^{-1}(A^c) = A^c$ and so $f^{-k}(A^c) = A^c$. It follows that if $x \in C$, $f^k(x) \in B \subseteq A^c$ and $x \in f^{-k}(A^c) = A^c$. Hence, $C \subseteq A^c \cap W$.

Now as $f$ is one-to-one on $W$ and maps $W$ onto $X$ except perhaps for the endpoints, if it exists we let $w_0 \in W$ be the point in $W$ such that $f^k(w_0)$ is the left endpoint of $X$ and, if it exists, we let $w_1 \in W$ be the point in $W$ such that $f^k(w_1)$ is the right endpoint of $X$. Put $D = \emptyset, \{ w_0 \}, \{ w_1 \}$ or $\{ w_0, w_1 \}$, depending on the existence of $w_0, w_1$. Then,
\begin{equation}
A^c \cap W \cap D \subseteq C \subseteq A^c \cap W.
\end{equation}
We now have, using (2.4), (2.5) and (2.6),
\[ \mu(A^c) = \mu(B) = \mu(f^k(C)) = \frac{\mu(C)}{\mu(W)}, \]
as \( f^k \) is linear on \( W \),
\[ \geq \frac{\mu(A^c \cap W)}{\mu(W)} > 1 - \varepsilon. \]

As this holds for all \( \varepsilon > 0 \), we see that \( \mu(A^c) = 1 \), and so \( \mu(A) = 0 \). Thus, \((X, f)\) is an ergodic system, by definition. □

Definitions. Let \( f \) be a baker’s transformation on an interval \( X \) of length 1 relative to a standard partition \((X_j)_{j \in I} \) of \( X \). Let \( U \) be a subinterval of \( X \) of positive length and let \( \Theta_U(x) \) denote the corresponding recurrence time for \( x \in U \).

Then, the average or expectation \( E(\Theta_U) \) of \( \Theta_U \) over \( U \) is defined as
\[ E(\Theta_U) = \frac{1}{\mu(U)} \sum_{n=1}^{\infty} n\mu(\{x : \Theta_U(x) = n\}). \]

Also, the standard deviation of \( \Theta_U \) over \( U \) is defined as
\[ \sigma(\Theta_U) = \sqrt{\frac{1}{\mu(U)} \sum_{n=1}^{\infty} |n - E(\Theta_U)|^2 \mu(\{x : \Theta_U(x) = n\})}. \]

Theorem 2.8. Let \( f \) be a baker’s transformation on an interval \( X \) of length 1 relative to a standard partition \( \mathcal{P} = (X_j)_{j \in I} \) of \( X \). Let \( \rho = \max\{\mu(X_j) : j \in I\} \), as in Lemma 2.6. Let \( U \) be a subset of \( X \) that is a union of \( r \) disjoint subintervals \( U_1, U_2, \ldots, U_r \) of positive length, and put \( \tau = \min\{\mu(U_j) : 1 \leq j \leq r\} \). Then, the following hold.

(i) The average of the recurrence function \( \theta_U \) of \( U \) relative to \( f \) is \( \mu(U)^{-1} \).

(ii) The standard deviation \( \sigma(\theta_U) \) of \( \theta_U \) is finite.

(iii) Let \( \lfloor x \rfloor \) denote the integer part of \( x \in \mathbb{R} \), and put
\[ \nu(U) = 1 + \left\lfloor \log(\tau/2) \log \rho \right\rfloor. \]

Then, \( 2r\rho^{\nu(U)} < \mu(U) \) and
\[ \sigma(\theta_U) \leq \sqrt{-\frac{1}{\mu(U)^2} + \frac{3}{\mu(U)} - 2 + \frac{2\nu(U)}{\mu(U)} - \frac{1}{\mu(U)^2} - 2r\rho^{\nu(U)} - 1}. \]

Proof. (i) By Lemma 2.7, the system \((X, f)\) is ergodic, so that the fact that the average of \( \Theta_U \) over \( U \) is \( 1/\mu(U) \) is immediate from Kac’s formula [3], [7]. Alternatively, it can be deduced as a consequence of Theorem 4.15 in [10], by applying part of the proof of that Theorem to show the assumptions in Theorem 4.11 in [10] are satisfied.

(ii) and (iii). If \( U = X \) except possibly for end points, then \( \sigma(\Theta_U) = 0 \), so the result is true in this case. So, we may assume that \( 0 < \mu(U) < 1 \). By Lemma 2.6, there is \( k \in \mathbb{N} \) and a set \( Y \) that is a non-empty union of sets in \( \mathcal{P}_k \) such that \( Y \subseteq U \). Note that there may be many such \( k \) and sets \( Y \) such that this holds; we
simply choose any such pair \(k, Y\). Put \(Z = Y^c\). Then, \(Z\) is also a union of sets in \(\mathcal{P}_k, U^c \subseteq Z\) and \(0 < \mu(Z) < 1\). Now, using any set \(Z\) obtained in this way we have

\[
\sum_{n=1}^{\infty} \mu(U^c \cap f^{-n}(U^c)) = \sum_{\ell=1}^{\infty} \left( \sum_{n=(\ell-1)k+1}^{\ell k} \mu(U^c \cap f^{-n}(U^c)) \right) \\
\leq \sum_{\ell=1}^{\infty} \left( \sum_{n=(\ell-1)k+1}^{\ell k} \mu(Z \cap f^{-n}(Z)) \right) \\
\leq \sum_{\ell=1}^{\infty} \left( \sum_{n=(\ell-1)k+1}^{\ell k} \mu(Z \cap f^{-\ell k}(Z) \cap \cdots \cap f^{-1}(Z)) \right) \\
= \sum_{\ell=1}^{\infty} k \mu(Z)^{\ell}, \text{ by Lemma 2.5,} \\
= \frac{k \mu(Z)}{1 - \mu(Z)} \\
(2.8) \\
< \infty.
\]

By Theorem 4.15 of [10, page 275], (2.8) implies that the standard deviation \(\sigma(\theta_U)\) is finite and, what is more, it implies that formula (4.39) in [10] holds for the standard deviation of \(\Theta_U\) in the present case.

To get an upper bound for \(\sigma(\Theta_U)\), we choose \(k\) and \(Y\) in (2.8) in a particular way, noting that \(k\) and \(Y\) are inter-related. We have from the definition of \(\nu(U)\) that \(2^r \rho^{\nu(U)} < \tau\), so for each \(j = 1, 2, \ldots, r\) we have \(2^r \rho^{\nu(U)} < \mu(U_j)\). Thus,

\[
2^r \rho^{\nu(U)} < \sum_{j=1}^{r} \mu(U_j) = \mu(U). \tag{2.9}
\]

Also, using Lemma 2.6, we now see that if we put \(Y_j\) for the union of all the sets in \(\mathcal{P}_{\nu(U)}\) that are subsets of \(U_j\), then for each \(j = 1, 2, \ldots, r\) we have

\[
0 < \mu(Y_j) \leq \mu(U_j) < \mu(Y_j) + 2^r \rho^{\nu(U)}.
\]

So, if we put \(Y = \bigcup_{j=1}^{r} Y_j\), this union is disjoint and we have

\[
\mu(U) = \sum_{j=1}^{r} \mu(U_j) < \sum_{j=1}^{r} \mu(Y_j) + 2^r \rho^{\nu(U)} = \mu(Y) + 2^r \rho^{\nu(U)}. \tag{2.10}
\]
Using (2.9) and (2.10), and taking $\nu(U)$ for $k$ in (2.8), we now have
\[
\sum_{n=1}^{\infty} \mu(U^c \cap f^{-1}(U^c) \cap \cdots \cap f^{-n}(U^c)) \\
\leq \nu(U) \left( \frac{1}{\mu(Y)} - 1 \right) \\
\leq \nu(U) \left( \frac{1}{\mu(U)} - 2r \rho \nu(U) - 1 \right).
\]

It now follows from (2.8) and the formula for the standard deviation in Theorem 4.15 in [10] that the expression inside the square root sign in (2.7) is non-negative and that (2.7) holds.

The estimate in (2.7) can be simplified in the special case when the partition $\mathcal{P}$ is finite and there is $k \in \mathbb{N}$ such that the intervals $U_1, U_2, \ldots, U_r$ are in $\mathcal{P}_k$. This means that $U$ is a finite union of sets in $\mathcal{P}_k$. Using the notations in the above proof we have, by Lemma 2.6, that $\mu(U) = \sum_{j=1}^{r} \mu(U_j) \leq r \rho^k$. Consequently,
\[
k \leq \log \left( \frac{\mu(U)}{r} \right) / \log \rho.
\]

Now putting this value of $k$ and putting $U$ for $Y$ in (2.8), we have
\[
\sum_{n=1}^{\infty} \mu(U^c \cap f^{-1}(U^c) \cap \cdots \cap f^{-n}(U^c)) \leq \log \left( \frac{\mu(U)}{r} \right) / \log \rho \left( \frac{1}{\mu(U)} - 1 \right),
\]
and instead of (2.7) we have the simpler-looking estimate
\[
\sigma(\theta_U) \leq \sqrt{-\frac{1}{\mu(U)^2} + \frac{3}{\mu(U)} - 2 + \frac{2 \log \left( \frac{\mu(U)}{r} \right)}{\mu(U) \log \rho} \left( \frac{1}{\mu(U)} - 1 \right)}.
\]

Note that the estimate for $\sigma(\Theta_U)$ in (2.7) cannot be expected to be the best possible. One reason is the fact that the inequality (2.8), upon which (2.7) depends, was obtained by replacing smaller terms by larger ones. However, note that in a general dynamical system the standard deviation of the recurrence times need not be finite [6].

REFERENCES