J. Appl. Prob. **45,** 640–649 (2008)
Printed in England

HITTING TIME AND INVERSE PROBLEMS FOR MARKOV CHAINS

VICTOR DE LA PEÑA,* *Columbia University*
HENRYK GZYL,** *IESA Caracus*
PATRICK MCDONALD,*** *New College of Florida*

Abstract

Let W_n be a simple Markov chain on the integers. Suppose that X_n is a simple Markov chain on the integers whose transition probabilities coincide with those of W_n off a finite set. We prove that there is an $M > 0$ such that the Markov chain W_n and the joint distributions of the first hitting time and first hitting place of X_n started at the origin for the sets $\{-M, M\}$ and $\{-(M+1), (M+1)\}$ algorithmically determine the transition probabilities of X_n.

Keywords: Inverse problem; Markov chain; first hitting time

2000 Mathematics Subject Classification: Primary 60G40; 35R30; 39A12

1. Introduction

Many problems arising in the natural sciences involve situations in which first hitting times for an unknown diffusion process driving particles in an inaccessible region are given, and from this data one seeks to determine properties of the underlying region and/or the particle dynamics (see [4] for a general reference, see [2] for applications in neuroscience). As an illustrative model problem, consider a long thin tube containing a liquid whose diffusivity is known outside a given interval, say $[-1, 1]$. Suppose that particles are injected at the point 0 and that the first hitting time probabilities are given at a number of locations outside the inaccessible interval $[-1, 1]$. We ask: What properties of the diffusivity can be determined from the given data?

In this paper we study a discrete analog of the model diffusion problem sketched above. We formalize the problem as follows.

Suppose that W_n is a simple Markov chain on the integers, $\mathbb{Z}$ (i.e. the transition probabilities between two integers are nonzero if and only if the two integers are nearest neighbors). Suppose that $D \subset \mathbb{Z}$ is a finite subset of $\mathbb{Z}$. Suppose that X_n is a simple Markov chain whose transition probabilities coincide with those of W_n outside of the subset D (we refer to such a Markov chain as a simple D-perturbation of W_n). The main result of this paper is the following.

Theorem 1.1. *Let W_n be a simple Markov chain on the integers. Suppose that $D \subset \mathbb{Z}$ is a finite set and that X_n is a simple D-perturbation of W_n. Let $M = \max\{|i| : i \in D\} + 1$. Then the Markov chain W_n and the joint distributions of the first hitting time and first hitting place*

Received 22 August 2006; revision received 28 July 2008.
* Postal address: Department of Statistics, Columbia University, New York, NY 10027, USA.
Email address: vp@stat.columbia.edu
** Postal address: Centro de Finanzas, IESA Caracus, Venezuela. Email address: henryk.gzyl@iesa.edu.ve
*** Postal address: Division of Natural Science, New College of Florida, Sarasota, FL 34243, USA.
Email address: mcdonald@ncf.edu

of X_n starting at the origin for the sets $\{-M, M\}$ and $\{-(M+1), (M+1)\}$ determine the transition probabilities of X_n.

We view Theorem 1.1 as an inverse result. Viewed in this context, there are a great number of directions in which we can seek generalizations. These directions include generalizations of the state space (discrete problems on other graphs, continuous problems, etc.), generalizations of the process (Markov chains on graphs, diffusions, etc.), and generalizations of the probabilistic data which one is given (moments of hitting times, etc.).

Our result can be further sharpened: from the proof of Theorem 1.1 in Section 3, it is clear that we need the joint distribution of time and place for $n \leq 3M-1$, and given such information, there is an algorithm which constructs the required transition probabilities (cf. Corollary 3.1). In addition, our result does not depend on the initial state of the Markov chain. Given that this is the case, we can formulate a version of 'diffusion tomography' in which we study properties of a given inaccessible region using particles which interact with the region via diffusion. Grünbaum and his colleagues (who coined the term 'diffusion tomography') have investigated related problems in the context of medical imaging (for a survey of optical tomography and associated inverse problems, see [1]; for applications of Markov chains to optical tomography, see [5], [6], [7], and [8]). The present work differs from those cited above in several ways. In particular, our work demonstrates that (in the language of the above papers) detailed 'time-of-flight' data suffices to solve the highly nonlinear inverse problem in one space dimension.

Because the joint distributions cited in Theorem 1.1 determine the underlying D-perturbation, Theorem 1.1 provides a context in which we can define and study meaningful problems in applied statistics. We develop general statistical tools for such problems and discuss several applications.

Our proof of Theorem 1.1 involves an analysis of the pathspace associated to our Markov chain.

The paper is organized as follows. In Section 2 we develop the notation used throughout the paper, the preliminary material needed for the proof of our main result, and an example that illustrates how our result can be used in practice to derive consistent estimators for the unknown transition probabilities. In Section 3 we provide a proof of our main result and corollaries which provide an explicit algorithm for constructing the required transition probabilities (cf. Corollary 3.1). In Section 4 we formalize and discuss our statistical results.

2. Background and notation

As in the introduction, let W_n be a simple Markov chain on the integers, and denote the transition probabilities of W_n by

$$w_i = \mathrm{P}(W_{n+1} = i+1 \mid W_n = i).$$

For notational convenience, we will write $w_i^* = 1 - w_i$ for the backward hopping probabilities (see Figure 1).

Throughout this paper, we will view the integers as defining a bidirected graph with edges given by connecting nearest neighbors and a natural orientation given by the ordering of the integers. With these conventions, by a *path of length k* in the integers we will mean a sequence of k integers connected by edges.

Definition 2.1. Let W_n be a simple Markov chain on $\mathbb{Z}$, and suppose that $D \subset \mathbb{Z}$ is a finite set. We say that a Markov chain X_n on $\mathbb{Z}$ is a D-perturbation if the transition probabilities of X_n

https://doi.org/10.1239/jap/1222441820 Published online by Cambridge University Press

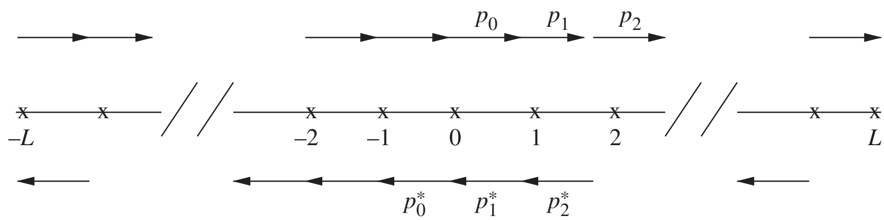

FIGURE 1: Transition probabilities for the one-dimensional problem.

coincide with those of W_n on $\mathbb{Z} \setminus D$. We say that X_n is a simple D-perturbation of W_n if X_n is a simple Markov chain which is a D-perturbation of W_n.

If L is a positive integer, we will denote by D_L the subset of $\mathbb{Z}$ defined by

$$D_L = \{i \in \mathbb{Z} \colon |i| \leq L\}.$$

Given an arbitrary finite $D \subset \mathbb{Z}$, let $L = \max\{|i| \colon i \in D\}$. Then $D \subset D_L$. The sets D_L play a fundamental role in the sequel.

If X_n is a simple D-perturbation of W_n, we will denote by p_i the transition probabilities of X_n. If, as above, $L = \max\{|i| \colon i \in D\}$, we can write

$$p_i = \begin{cases} \mathrm{P}(X_{n+1} = i + 1 \mid X_n = i) & \text{if } |i| \leq L, \\ w_i & \text{otherwise.} \end{cases}$$

We will denote by P^l the probability measure associated to X_n which charges paths beginning at l.

Given $D \subset D_L$ and $M > L$, we will denote by $\eta = \eta_{-M,M}$ the first hitting time of X_n for $\{-M, M\}$:

$$\eta_{-M,M} = \inf\{n \geq 0 \colon X_n \in \{-M, M\}\}.$$

We will write the joint distribution of the first hitting time and first hitting place as

$$\mathrm{P}_M^0(k, -) = \mathrm{P}^0(\eta = k,\ X_\eta = -M),$$
$$\mathrm{P}_M^0(k, +) = \mathrm{P}^0(\eta = k,\ X_\eta = M).$$

The first nontrivial example of the inverse problem described in Theorem 1.1 occurs for the set $D = D_1$. We let X_n be a simple D_1-perturbation of W_n and we study this problem in detail.

Suppose that we are given the joint distribution of the first hitting time and first hitting place for the set $\{-2, 2\}$. With the notation as above, we have

$$\mathrm{P}_2^0(2, +) = p_0 p_1, \qquad (2.1)$$
$$\mathrm{P}_2^0(2, -) = p_0^* p_{-1}^*.$$

For all positive integers k, it is clear that

$$\mathrm{P}_2^0(2k + 1) = 0.$$

In addition, every path of length $2k$ starting at 0 and having first hitting time of $\{-2, 2\}$ given by $2k$ and first hitting place given by 2 has an associated occurrence probability of the form

$$(p_0^* p_{-1})^{l_1} (p_0 p_1^*)^{l_2} p_0 p_1,$$

where $l_1 + l_2 + 1 = k$. Thus,

$$P_2^0(2k, +) = \Big(\sum c_{l_1,l_2}(p_0^* p_{-1})^{l_1}(p_0 p_1^*)^{l_2}\Big) p_0 p_1, \tag{2.2}$$

where the coefficient c_{l_1,l_2} counts the number of distinct paths occurring for the partition of k given by $l_1 + l_2 + 1 = k$. Thus,

$$c_{l_1,l_2} = \binom{k-1}{l_1},$$

and we can sum:

$$P_2^0(2k, +) = (p_0^* p_{-1} + p_0 p_1^*)^{k-1} P_2^0(2, +). \tag{2.3}$$

But, $p_1^* = (1 - p_1)$ and $p_{-1} = (1 - p_{-1}^*)$. From this we conclude that

$$p_0^* p_{-1} + p_0 p_1^* = -(p_0 p_1 + p_0^* p_{-1}^* - 1).$$

This together with a trivial algebraic computation proves the following lemma.

Lemma 2.1. *Let W_n be a simple Markov chain on the integers, and suppose that X_n is a simple D_1-perturbation of W_n. Then the joint distribution of the first hitting time and first hitting place of X_n starting at the origin for the set $\{-2, 2\}$ does not determine the transition probabilities of X_n.*

In fact, the proof provides a method for constructing two different perturbations with first hitting times and first hitting places having the same joint distributions:

$$(p_{-1}, p_0, p_1) = \left(\tfrac{3}{4}, \tfrac{1}{2}, \tfrac{1}{4}\right),$$
$$(p_{-1}, p_0, p_1) = \left(\tfrac{1}{2}, \tfrac{3}{4}, \tfrac{1}{6}\right).$$

We note that the conclusion of Lemma 2.1 depends on the location of the starting point: it is trivial to verify that other starting points give hitting time probabilities which determine the transitions.

Suppose that we know the joint distribution of the first hitting time and first hitting place for the set $\{-2, 2\}$ and the set $\{-3, 3\}$. Then

$$P_3^0(3, +) = p_0 p_1 w_2 \tag{2.4}$$

and

$$P_3^0(5, +) = P_3^0(3, +)(p_0 p_1^* + p_0^* p_{-1} + p_1 w_2^*). \tag{2.5}$$

From (2.3), (2.4), and (2.5), we conclude that

$$p_1 = \frac{1}{w_2^*}\left(\frac{P_3^0(5, +)}{P_3^0(3, +)} - \frac{P_2^0(4, +)}{P_2^0(2, +)}\right), \tag{2.6}$$

from which it follows that the transition probabilities are determined. In particular, we have proven the following lemma.

Lemma 2.2. *Let W_n be a simple Markov chain on the integers, and suppose that X_n is a simple D_1-perturbation of W_n. Then the joint distributions of the first hitting time and first hitting place of X_n starting at the origin for the sets $\{-2, 2\}$ and $\{-3, 3\}$ determine the transition probabilities of X_n.*

From Lemma 2.2, it follows that there is a natural statistical problem associated to a simple D_1-perturbation of W_n. We formalize this problem as follows. From (2.1), (2.2), (2.6), and their analogs for p_{-1}, and using the fact that $p_i^* = 1 - p_i$ for all i, we can see that the transition probabilities p_0, p_1, and p_{-1} can be obtained from the 8-vector $\bar{\mathrm{P}}$ given by

$$\bar{\mathrm{P}} = ((\mathrm{P}_2^0(2, \pm)), (\mathrm{P}_3^0(3, \pm)), (\mathrm{P}_2^0(4, \pm)), (\mathrm{P}_3^0(5, \pm))).$$

Let $\{(X_1^i, X_2^i, \dots, X_5^i)\}_{i=1}^m$ be m independent copies of our Markov chain up to time $n = 5$. The natural estimator for the first component of the vector $\bar{\mathrm{P}}$ is given by

$$\mathrm{P}_2^0(2, +, m) = \frac{1}{m} \sum_{i=1}^{m} \xi_{\{\eta_{-2,2}=2,\, X_2^i=2\}}, \tag{2.7}$$

where $\xi_{\{\eta_{-2,2}=2, X_2^i=2\}}$ is the indicator function for the given event. By the strong law of large numbers, (2.7) is a consistent statistic for $\mathrm{P}_2^0(2, +)$. Repeating the construction for each component of $\bar{\mathrm{P}}$, we obtain a strongly consistent vector estimator for $\bar{\mathrm{P}}$. Using the continuous mapping theorem (if all the original probabilities in our random vector are nonzero), we obtain a vector $(p_{-1,m}, p_{0,m}, p_{1,m})$ which is a strongly consistent estimator for (p_{-1}, p_0, p_1) (cf. Section 4, below).

Observing that we have used eight parameters to determine three unknowns, it is natural to question whether there is a more efficient algorithm which recovers the perturbation. With this in mind, consider the set $\{-2, 3\}$, introduce the obvious notation, and follow the argument used to establish Lemma 2.2, to obtain

$$p_1 = \frac{1}{w_2^*} \left(\frac{\mathrm{P}^0_{-2,3}(5, +)}{\mathrm{P}^0_{-2,3}(3, +)} - \frac{\mathrm{P}^0_{-2,3}(4, -)}{\mathrm{P}^0_{-2,3}(2, -)} \right),$$

establishing that the transition probabilities of X_n are determined by W_n and the joint distribution of the first hitting time and first hitting place of X_n starting at the origin for the set $\{-2, 3\}$. While we suspect that this example is an anomaly (i.e. for $L > 1$, we cannot recover the transition probabilities using data from a single interval), we are able to prove only that our algorithm for solving the inverse problem fails.

3. Proof of Theorem 1.1

We begin with a reduction.

Lemma 3.1. *Theorem 1.1 is true if and only if, for all* $L > 0$, *Theorem 1.1 is true for* $D = D_L$.

Proof. Let D be an arbitrary finite subset of $\mathbb{Z}$, and let X_n be a simple D-perturbation of W_n. Let $L = \max\{|i| : i \in D\}$, and let $M = L + 1$. Then $D \subset D_L$ and X_n is a simple D_L-perturbation of W_n with $\max\{|i| : i \in D_L\} + 1 = M$. Assuming that the transition probabilities of X_n are determined on D_L, we find that the transition probabilities of X_n are determined on D. This completes the proof.

To prove Theorem 1.1 for sets of the form D_L, we will give a careful analysis of the structure of paths beginning at the origin and having certain prescribed hitting properties. To this end, let $L > 0$, let $m > 0$, and define (for the remainder of the paper)

$$T = (L + m + 1) + 2m. \tag{3.1}$$

https://doi.org/10.1239/jap/1222441820 Published online by Cambridge University Press

We consider paths beginning at the origin and having T as the first hitting time for $L+m+1$. More precisely, we define

$$\Gamma = \{\gamma\colon \gamma(0)=0,\ \gamma(T)=L+m+1,\ \gamma(j)<L+m+1 \text{ for all } j<T\}. \tag{3.2}$$

Elements of Γ have a nice lower bound.

Lemma 3.2. *For Γ as in (3.2), if $\gamma \in \Gamma$ then $\gamma(j) > -(L+m)$ for all $j \le T$.*

Proof. Let γ be a curve satisfying $\gamma(0)=0$ and $\gamma(j)=-(L+m)$. Then $j>(L+m-1)$. Denote by τ the first time that γ visits $L+m+1$. Then

$$\begin{aligned}\tau &\ge j+(L+m)+(L+m+1)\\ &> L+m-1+(L+m)+(L+m+1)\\ &= (L+m+L)+L+2m.\end{aligned}$$

We conclude that $\tau > T$, which completes the proof.

We partition Γ by the first hitting times of $L+m$.

Lemma 3.3. *Let Γ be as in (3.2), and define*

$$\Gamma_k = \{\gamma \in \Gamma\colon \gamma(T-(2k-1)) = L+m,\ \gamma(j) < L+m \text{ for all } j<T\}. \tag{3.3}$$

Then

(a) $\Gamma_i \cap \Gamma_j = \varnothing$ *if* $i \ne j$,

(b) $\bigcup_{k=1}^{m+1} \Gamma_k = \Gamma$.

Proof. If $i \ne j$, elements of Γ_i and Γ_j have different first hitting times of $L+m$, and, thus, $\Gamma_i \cap \Gamma_j = \varnothing$. Since every element of Γ begins at 0 and hits $L+m$ by time T, $\bigcup_{k=1}^{m+1} \Gamma_k = \Gamma$. This completes the proof.

Initial segments of paths in Γ_k define paths with nice first hitting properties. We make this precise in the following lemma.

Lemma 3.4. *Let $1 \le k \le m+1$. For $\gamma \in \Gamma_k$, the path $T_k(\gamma)$ obtained by truncating γ at time $T-(2k-1)$ satisfies*

(a) $T_k(\gamma)(0) = 0$,

(b) $T_k(\gamma)(T-(2k-1)) = L+m$,

(c) $-(L+m) < T_k(\gamma)(j) < L+m$ *for all* $j < T-(2k-1)$.

Proof. By definition, for $\gamma \in \Gamma_k$, the first hitting time of $L+m$ is $T-(2k-1)$, from which statement (b) and the right-hand side inequality of statement (c) follow. Statement (a) is trivial, and the left-hand side inequality of statement (c) follows from Lemma 3.2. This completes the proof.

Truncation provides for a decomposition of paths in Γ_k: each such path consists of an initial segment which has nice first hitting properties, followed by an end segment which never visits the vertex L. We make this precise in the following lemma.

https://doi.org/10.1239/jap/1222441820 Published online by Cambridge University Press

Lemma 3.5. *For* $1 \leq k < m+1$,

$$P^0(\Gamma_k) = P^0_{L+m}(T-(2k-1),+)\chi_k,$$

where χ_k *is an expression which involves only the transition probabilities* w_j, $L+1 \leq j \leq L+m+1$.

Proof. By Lemma 3.4, each $\gamma \in \Gamma_k$ can be decomposed as a path $T_k(\gamma)$ starting at 0 with the first hitting time of $\{-(L+m), (L+m)\}$ occurring at time $T-(2k-1)$ and position $L+m$, followed by a path of length $(2k-1)$ which begins at $L+m$ and ends when it makes its first visit to $L+m+1$. We will write

$$\tilde{\Gamma}_k = \{\gamma : \gamma(0) = L+m,\ \gamma(2k-1) = L+m+1,\ \gamma(j) < L+m+1 \text{ for all } j < 2k-1\}.$$

By choice of k, if $\gamma \in \tilde{\Gamma}_k$, $\gamma(j) > L$ for all $0 \leq j \leq 2k-1$. Thus, if $\gamma \in \tilde{\Gamma}_k$, we can compute $P^{L+m}(\{\gamma\})$ in terms of the transition probabilities w_j, $L < j \leq L+m$. Summing over all elements $\tilde{\Gamma}_k$ gives an expression

$$\chi_k = P^{L+m}(\tilde{\Gamma}_k),$$

which involves only the transition probabilities w_j, $L < j \leq L+m$. Finally, we compute

$$P^0(\Gamma_k) = P^0_{L+m}(T-(2k-1),+)\chi_k,$$

as required. This completes the proof.

The following lemma is the essential step in establishing Theorem 1.1.

Lemma 3.6. *Suppose that* W_n *is a simple Markov chain on* $\mathbb{Z}$. *For* $L > 0$, *let* X_n *be a simple* D_L*-perturbation of* W_n. *Let* $m > 0$. *Then* p_L *and* p_{-L} *are determined by the Markov chain* W_n *and the joint distributions of the first hitting time and first hitting place of* X_n *starting from the origin for the sets* $\{-(L+m), (L+m)\}$ *and* $\{-(L+m+1), (L+m+1)\}$.

Proof. From Lemma 3.3 and Lemma 3.5, we have

$$\begin{aligned} P^0_{L+m+1}(T,+) &= P^0(\Gamma) \\ &= \sum_{k=1}^{m} P^0_{L+m}(T-(2k-1),+)\chi_k + P^0(\Gamma_{m+1}). \end{aligned} \tag{3.4}$$

We let γ_* be the element of Γ_{m+1}, which changes direction exactly twice. From (3.1) and (3.3), γ_* is the path which starts at 0, moves right $L+m$ units, moves left m units, and moves right $m+1$ units. Thus, we can explicitly compute the probability that γ_* occurs:

$$P^0(\{\gamma_*\}) = \chi_* p_L, \tag{3.5}$$

where

$$\chi_* = P^0_{L+m+1}(L+m+1,+)\prod_{i=L+1}^{L+m-1} w_i \prod_{j=L+1}^{L+m}(1-w_j).$$

If $\gamma \in \Gamma_{m+1} \setminus \{\gamma_*\}$ then, as in Lemma 3.5, we may view γ as a truncation followed by a path which never visits L. Thus, as in Lemma 3.5, we can write

$$P^0(\Gamma_{m+1} \setminus \{\gamma_*\}) = P^0_{L+m}(L+m,+)\chi_{m+1}, \tag{3.6}$$

where χ_{m+1} depends only on the transition probabilities w_j, $L < j \leq L + m$. Using (3.4), (3.5), and (3.6), we can solve for p_L:

$$p_L = \frac{1}{\chi_*}\left(\mathrm{P}^0_{L+m+1}(T, +) - \sum_{k=1}^{m+1} \mathrm{P}^0_{L+m}(T - (2k-1), +)\chi_k\right). \tag{3.7}$$

By symmetry, there is a similar formula for p_{-L}. This completes the proof.

Proof of Theorem 1.1. By Lemma 3.1, it suffices to consider the case in which $D = D_L$, where $L > 0$ is arbitrary. We proceed by induction on L.

By Lemma 2.2, the result is true when $L = 1$. When $L = n + 1$, we can use Lemma 3.6 with $m = 1$ to determine p_{n+1} and $p_{-(n+1)}$. Applying Lemma 3.6 n times, incrementing m while decreasing L at each repetition, completes the proof.

From the proofs of Lemma 3.6 and Theorem 1.1, we note that, with M as defined in Theorem 1.1, we only require values of the joint distribution of the exit time and place for time $t \leq M + 3$. It is then the case that (3.7) gives an explicit value for the transition probability p_L. From this we deduce the following corollary.

Corollary 3.1. *Under the hypotheses of Theorem 1.1, there is an algorithmic procedure for explicitly determining the transition probabilities for any given simple D-perturbation. With M as in Theorem 1.1, the algorithm depends on data from the joint distribution of the exit time and place for time $t \leq 3L + 2$.*

Proof. From the proof of Lemma 3.6 we use (3.7) L times, each time decreasing L by 1 and increasing m by 1. This has the effect of changing the value of T by 2 at each step. The corollary follows from (3.1).

From the proof of Theorem 1.1, it is clear that the starting point of the process need not be an element of the region which defines the perturbation, nor need the starting point be the origin. From the point of view of application, this fact allows us to develop a 'probabilistic scattering' framework.

Finally, we note that there is an alternate proof of Theorem 1.1 using the strong Markov property. Our proof, however, provides an algorithm for constructing the required transition probabilities.

4. Associated statistics

As in the introduction, let W_n be a simple Markov chain on the integers. Suppose that $D \subset \mathbb{Z}$ is a finite set and that X_n is a simple D-perturbation of W_n. Let $M = \max\{|i| : i \in D\} + 1$ and, for $L = M - 1$, view X_n as a simple D_L-perturbation of W_n. By Theorem 1.1 and its proof, the transition probabilities associated to X_n can be reconstructed from the components of the $2(2L + 2)$-vector (cf. Corollary 3.1):

$$\bar{\mathrm{P}} = ((\mathrm{P}^0_{L+1}(L+1, \pm)), (\mathrm{P}^0_{L+2}(L+2, \pm)), \dots, (\mathrm{P}^0_{L+2}(3L+2, \pm))).$$

Let $\{(X^i_1, X^i_2, \dots, X^i_{3L+2})\}_{i=1}^m$ be m independent copies of our random walk up to time $n = 3L + 2$. The natural estimator for the first component of the vector $\bar{\mathrm{P}}$ is given by

$$\mathrm{P}^0_{L+1}(L+1, +, m) = \frac{1}{m}\sum_{i=1}^{m} \xi_{\{\eta_{-(L+1),L+1} = L+1,\, X^i_{L+1} = L+1\}}, \tag{4.1}$$

where $\xi_{\{\eta_{-(L+1),L+1} = L+1,\, X^i_{L+1} = L+1\}}$ is the indicator function for the given event.

Theorem 4.1. *Let X_n be a simple D_L-perturbation of a given Markov chain W_n. Suppose that, for $L+1 \le j, l \le 3L+2$, $\mathrm{P}^0_{L+1}(j, \pm, m)$ and $\mathrm{P}^0_{L+2}(l, \pm, m)$ are the natural estimators defined as analogs of (4.1). Let $p_{i,m}$ denote the estimator of p_i obtained using the estimators $\mathrm{P}^0_{L+1}(j, \pm, m)$ and $\mathrm{P}^0_{L+2}(l, \pm, m)$, and the algorithm of Corollary 3.1. Then, for each i, $p_{i,m}$ is a consistent estimator for p_i.*

Proof. By the strong law of large numbers, (4.1) is a consistent statistic for each component of $\bar{\mathrm{P}}$, the vector determining the D_L-perturbation X_n. By the proof of Theorem 1.1, it is clear that each of the p_i defining the perturbation is a rational function of the components of $\bar{\mathrm{P}}$. By the continuous mapping theorem (if all the original probabilities in our random vector are nonzero), we obtain a vector $(p_{-L,m}, p_{-L+1,m}, \dots, p_{L,m})$ which is a strongly consistent estimator for $(p_{-L}, p_{-L+1}, \dots, p_L)$. This completes the proof.

There are a number of pertinent observations to be made concerning Theorem 4.1.

The construction of our consistent estimator involves the use of (3.7) and its analogues. Since these equations involve probabilities in the denominators, we need to safeguard against dividing by 0. Since we are assuming that each of the transition probabilities is positive, each of the components of $\bar{\mathrm{P}}$ is nonzero. If we set

$$\tau_{L+1} = \inf_{m \ge 0}\{m \colon \xi_{\{\eta_{-(L+1),L+1}=L+1,\ X^m_{L+1}=L+1\}} > 0\}$$

then, by the strong law of large numbers, $\tau_{L+1} < \infty$ almost surely and $\mathrm{P}^0_{L+1}(L+1, +, m \vee \tau_{L+1})$ is a nonzero consistent estimator of $\mathrm{P}^0_{L+1}(L+1, +)$. The remaining components of $\bar{\mathrm{P}}$ can be treated similarly, as long as we take as our running index $m \vee \bigvee_{j=L+1}^{3L+2} \tau_j$. Given this, the estimating vector $\bar{\mathrm{P}}_{m \vee \bigvee_{j=L+1}^{3L+2} \tau_j}$ converges almost surely to $\bar{\mathrm{P}}$ and, hence, we have a consistent estimator.

We note that, while each of the entries in our vector estimator for the components of $\bar{\mathrm{P}}$ is a maximum likelihood estimate for the respective probability, it is not clear that the vector estimate for the transition probability vector defining the perturbation is a maximum likelihood estimate (it is necessary to further study the joint behavior of the vector components before making such a claim). Analogous claims hold for the derived vector $(p_{-L,m}, p_{-L+1,m}, \dots, p_{L,m})$.

We note that, as defined, the estimators $p_{i,m}$, though consistent for p_i, need not (and in fact most likely will not) be unbiased.

As sketched in the introduction, our results were developed in the context of a model involving first passage probabilities which arises in a variety of physical contexts (cf. [4]). This model involves a natural inverse problem for a continuous process. More precisely, suppose that $a \colon \mathbb{R} \to \mathbb{R}$ is a smooth function satisfying $a(x) > 0$ and $a(x) = 1$ if $|x| \ge 1$. Let L be the elliptic operator

$$L = \frac{1}{2}\left(a(x)\frac{\mathrm{d}}{\mathrm{d}x}\right)^2,$$

and suppose that X_t is a diffusion process with generator L. Suppose that we start the process at $x = 0$ and stop it at the boundary of $[-1, 1]$. Suppose that we are given the joint distribution of the first hitting time and hitting place, as well as the (spatial) derivative of the joint distribution at the boundary. Then it is natural to inquire as to whether $a(x)$ is determined. This is indeed the case. An argument using our results and convergence of the discrete approximation (as well as an independent argument using completeness theorems for products of eigenfunctions) is in preparation.

While it is easy to see that there can be no general higher-dimensional analog of our results, we have obtained completely analogous results for finite trees [3]. We expect that these results will have applications involving communication networks.

In addition to the 'standard' physical problems and those related to medical imaging, we foresee a number of other potential applications of our method. For example, it seems plausible that there will be potential applications in ecology. Releasing an animal at a central location and monitoring its arrival times and positions at a set of sites can be used to estimate how likely it is that the animal wandered in specified regions outside our direct control. Similar experiments could be conceived to analyze the flow of counterfeit money, spread of disease, and the impact of advertising on the image of a product.

References

[1] Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse Problems* **15,** R41–R93.

[2] Bal, G. and Chu, T. (2004). On the reconstruction of diffusions from first-exit time distributions. *Inverse Problems* **20,** 1053–1065.

[3] De la Peña, V., Gzyl, H. and McDonald, P. (2008). Inverse problems for random walks on trees: network tomography. To appear in *Statist. Prob. Lett.*

[4] Gardiner, C. W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 3rd edn. Springer, Berlin.

[5] Grünbaum, F. A. (1992). Diffuse tomography: the isotropic case. *Inverse Problems* **8,** 409–419.

[6] Grünbaum, F. A. (2003). Diffuse tomography as a source of challenging nonlinear inverse problems for a general class of networks. *Modern Signal Process.* **40,** 137–146.

[7] Grünbaum, F. A. and Matusevich, L. F. (2002). Explicit inversion formulas for a model in diffuse tomography. *Adv. Appl. Math.* **29,** 172–183.

[8] Patch, S. (1995). Recursive recovery of a family of Markov transition probabilites from boundary value data. *J. Math. Phys.* **36,** 3395–3412.