States recognition in random walk Markov chain via binary Entropy

Morteza Khodabin

(1)Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj, Iran

Abstract
In this paper, a new method for specification of recurrence or transient of states in one and two dimensional simple random walk based on upper and lower bounds of \( r \)-combinations from a set of \( m \) elements (\( C^m_r \)) via binary entropy is introduced.

Keywords: Markov chain; Shannon’s entropy; Stirling’s approximation.

1 Introduction

Information theory is a branch of applied mathematics and electrical engineering involving the quantification of information. A key measure of information in the theory is known as entropy, which is usually expressed by the average number of bits needed for storage or communication. The concept of entropy plays a major part in communication theory. Intuitively, entropy quantifies the uncertainty involved when encountering a random variable. The field is at the intersection of mathematics, statistics, computer science, physics, neurobiology, and electrical engineering. See [2] and [6] for more details.

On one hand, modern probability theory studies chance processes for which the knowledge of previous outcomes influences predictions for future experiments. In mathematics, a Markov chain, is a stochastic process with the Markov property, i.e., given the present state, future states are independent of the past states. Markov processes are a central topic in applied probability and statistics. The reason is that many real problems can be modeled by this kind of stochastic processes in continuous or indiscrete time. They form one of the most important classes of random processes and has many applications in:


T. M. and J.A. Thomas [2] have shown an initial link between information theory and some of concepts in Markov chain. Here, these concepts are used to introduce a new method for specification of states type in one and two dimensional simple random walk.

The paper is organized as follows:

*Corresponding authors. E-mail address: m-khodabin@kiau.ac.ir, Tel:+982634418143
Section 2 reminds some short definitions in information theory and Markov chain. Section 3 discusses about binary entropy approximation of $C_m^n$. In Section 4 a new method for specification of states type in one and two dimensional simple random walk based on entropy approximation is introduced. Finally, section 5 gives brief conclusion.

2 Preliminaries and notations

This section introduces the basic definitions that is used in the next sections and subsections. Here assumes that all random variables are discrete and "log" is to the base 2. Hence, entropy is expressed in bits.

2.1 Markov chain

A stochastic process is a system $\{X_t; t \in T\}$ of real random variables with time parameter $t \in T$. In the following we assume that the stochastic process is a discrete time and is denoted by $\{X_n; n \geq 0\}$.

Definition 2.1. ([3]). A discrete stochastic process $\{X_n; n \geq 0\}$ is said to be a Markov chain with state space $S = \{x_0, x_1, \ldots, x_n, \ldots\}$ if for $n = 0, 1, 2, \ldots$

$$Pr\{X_{n+1} = y | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} = Pr\{X_{n+1} = y | X_n = x_n\},$$

for all $x_0, x_1, \ldots, x_n, y \in S$. In this case, the probability transition matrix is given by

$$P = (p_{x_i, x_j})_{i,j \in \{0,1,\ldots,n\}},$$

where, $p_{x_i, x_j} = Pr\{X_{n+1} = x_j | X_n = x_i\}$.

We say that the state $j$ is accessible from $i$ if there exists an $n > 0$ such that $p^{(n)}_{ij} > 0$. We write $i \rightarrow j$.

If $i$ is accessible from $j$, and $j$ is accessible from $i$, we say that the states $i$ and $j$ communicate, or that they are communicating. We write $i \leftrightarrow j$.

A Markov chain is said to be irreducible if all the states communicate.

Definition 2.2. ([3]). The state $i$ is said to be recurrent if and only if

$$\sum_{n=0}^{\infty} p^{(n)}_{ii} = \infty,$$

and is transient if and only if

$$\sum_{n=0}^{\infty} p^{(n)}_{ii} < \infty.$$

2.2 Entropy

The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Let $X$ be a random variable with probability mass function

$$p(x) = Pr(X = x), \quad x \in A.$$

Definition 2.3. ([2]). The Shannon’s entropy $H(X)$ of random variable $X$ is defined by

$$H(X) = - \sum_{x \in A} p(x) \log(p(x)), \quad (2.1)$$

for example the binary entropy function is obtained as

$$H(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}. \quad (2.2)$$
3 Upper and lower bounds of $C^m_r$

Consider the Poisson distribution with mean $\lambda$,

$$Pr(X = r) = P(r) = e^{-\lambda} \frac{\lambda^r}{r!}, \quad r \in \{0, 1, 2, \ldots\}.$$ 

For large $\lambda$, this distribution is well approximated at least in the vicinity of $r \simeq \lambda$ by a Gaussian distribution with mean and variance $\lambda$:

$$e^{-\lambda} \frac{\lambda^r}{r!} \simeq \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(r-\lambda)^2}{2\lambda}}.$$ 

Let’s plug $r = \lambda$ into this formula:

$$e^{-\lambda} \frac{\lambda^\lambda}{\lambda!} \simeq \frac{1}{\sqrt{2\pi\lambda}}.$$ 

Then, we get

$$\lambda! \simeq \lambda e^{-\lambda} \sqrt{2\pi\lambda}. \quad (3.3)$$

This is Stirling’s approximation for factorial function.

**Theorem 3.1.** Let $f(m, r) = \frac{1}{\sqrt{2\pi(m-r)}} 2^{mH(\frac{r}{m})}$, then we have

$$f(m, r) \left(1 + \frac{1}{4r}\right)^{1}(1 + \frac{1}{4(m-r)})^{-1} < C^m_r < f(m, r)(1 + \frac{1}{4m}). \quad (3.4)$$

**Proof.** The sharp form of Stirling’s formula (see e.g. [1], Vol 1, P 36, [5]) is given by,

$$\lambda^\lambda e^{-\lambda} \sqrt{2\pi\lambda} < \lambda! < \lambda^\lambda e^{-\lambda} \sqrt{2\pi\lambda} \left(1 + \frac{1}{4\lambda}\right). \quad (3.5)$$

As $\lambda \to \infty$, we achieve (3.3). By taking logarithm from (3.5) we get

$$\lambda \log \lambda - \lambda + \frac{1}{2} \log 2\pi \lambda < \log \lambda! < \lambda \log \lambda - \lambda + \frac{1}{2} \log 2\pi \lambda + \log \left(1 + \frac{1}{4\lambda}\right). \quad (3.6)$$

Now, via combination formula and (3.6) we have

$$h(m, r) - \log \left(1 + \frac{1}{4r}\right)\left(1 + \frac{1}{4(m-r)}\right) < \log C^m_r < h(m, r) + \log(1 + \frac{1}{4m}), \quad (3.7)$$

where,

$$h(m, r) = (m-r) \log \frac{m}{m-r} + r \log \frac{m}{r} - \frac{1}{2} \log \frac{2\pi(m-r)r}{m}. \quad$$

Put $p = \frac{r}{m}$ in (2.2), we get

$$(m-r) \log \frac{m}{m-r} + r \log \frac{m}{r} = mH\left(\frac{r}{m}\right),$$

then we can rewrite (3.7) as

$$\log C^m_r < mH\left(\frac{r}{m}\right) - \frac{1}{2} \log \frac{2\pi(m-r)r}{m} + \log \left(1 + \frac{1}{4m}\right),$$

and

$$\log C^m_r > mH\left(\frac{r}{m}\right) - \frac{1}{2} \log \frac{2\pi(m-r)r}{m} - \log \left(1 + \frac{1}{4r}\right)\left(1 + \frac{1}{4(m-r)}\right),$$

which is the desired result.

**Corollary 3.1.** If $m = 2n$ and $r = n$ then

$$\frac{1}{\sqrt{n\pi}} 2^{2n}(1 + \frac{1}{4n})^2 < C^n_r < \frac{1}{\sqrt{n\pi}} 2^{2n}(1 + \frac{1}{8n}). \quad (3.8)$$
4 Application in random walk Markov chain

The random walk, is a mathematical formalization of a trajectory that consists of taking successive random steps. The results of random walk analysis have been applied to computer science, physics, ecology, economics, and a number of other fields as a fundamental model for random processes in time. For example, the path traced by a molecule as it travels in a liquid or a gas, the search path of a foraging animal, the price of a fluctuating stock and the financial status of a gambler can all be modeled as random walks. Various different types of random walks are of interest. Often, random walks are assumed to be Markov chains or Markov processes, but other, more complicated walks are also of interest. Some random walks are on graphs, others on the line, in the plane, or in higher dimensions, while some random walks are on groups. Random walks also vary with regard to the time parameter. A popular random walk model is that of a random walk on a regular lattice, where at each step the walk jumps to another site according to some probability distribution. In simple random walk, the walk can only jump to neighbouring sites of the lattice. For more detail we refer the reader to \[4\] and \[7\].

A (classic or simple) random walk is a Markov chain whose state space is the set \( \mathbb{Z} \) of all integers, and for which

\[
p_{i,i+1} = p = 1 - p_{i,i-1} \quad \text{for} \quad i = \pm 0, \pm 1, \pm 2, \ldots,
\]

for some \( p \in (0,1) \). Thus, in the case of a classic random walk, the process can make a transition from a given state \( i \) only to one or the other of its two immediate neighbors: \( i - 1 \) or \( i + 1 \). Moreover, the length of the displacement is always the same, i.e., one unit. This walk can be illustrated as follows. Say you flip a coin with probability \( p \) as heads. If it lands on heads, you move one to the right on the number line. If it lands on tails (with probability \( q = 1 - p \)), you move one to the left.

Remark 4.1. In the case of a classic random walk, defined on all the integers, all the states communicate, which follows directly from the fact that the process cannot jump over a neighboring state and that it is unconstrained, that is, there are no boundaries (absorbing or else). The chain being irreducible, the states are either all recurrent or all transient. So from irreducibly, it’s sufficient that one consider 0 state.

Lemma 4.1. Let \( \{X_n; n \geq 1\} \) be a simple random walk on the integers started at the origin, then the probability of returning to the origin at the \( 2n \)th step is given by

\[
P_{00}^{(2n)} = C_n^2 p^n q^n \quad (4.9)
\]

Proof. Since the process moves exactly one unit to the right or to the left at each transition, it is clear that if it starts from 0, then it cannot be at 0 after an uneven number of transitions. That is,

\[
P_{00}^{(2n+1)} = 0 \quad \text{for} \quad n = 0,1,2,\ldots
\]

Next, for the process to be back to the initial state 0 after \( 2n \) transitions, there must have been exactly \( n \) transitions to the right (and thus \( n \) to the left). Since the transitions are independent and the probability that the process moves to the right is always equal to \( p \), we may write that, for \( n = 1,2,\ldots \)

\[
P_{00}^{(2n)} = \Pr[\text{Binomial}(2n, p) = n] = C_n^2 p^n q^n.
\]

To determine whether 0 state (and therefore the chain) is recurrent, we consider only the sum

\[
\sum_{n=1}^{\infty} P_{00}^{(2n)} = \sum_{n=1}^{\infty} C_n^2 p^n q^n.
\]

However, we do not need to know the exact value of this sum, but rather we need only know whether it converges or not.

Lemma 4.2. ([3] and [7]). Let \( N \) be the total number of returns that the walk ever makes to the origin, then

\[
E(N) = \sum_{n=1}^{\infty} P_{00}^{(2n)} = \sum_{n=1}^{\infty} C_n^2 p^n q^n \quad (4.10)
\]
Theorem 4.1. The simple random walk chain is persistent iff \( p = q \) and equivalently is transient iff \( p \neq q \).

Proof. Through (3.8) and (4.10), we attain
\[
\sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} (4pq)^n (1 + \frac{1}{4n})^{-2} < \sum_{n=1}^{\infty} p_{00}^{(2n)} < \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} (4pq)^n (1 + \frac{1}{8n}).
\] (4.11)

Let \( p = q \) and since \( 1 < 1 + \frac{1}{4n} < 2 \), then
\[
\sum_{n=1}^{\infty} p_{00}^{(2n)} < \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} (4pq)^n (1 + \frac{1}{4n})^{-2} > \frac{1}{4\sqrt{\pi}} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty.
\]

Let \( p \neq q \), since \( 4pq < 1 \) and \( 1 < 1 + \frac{1}{8n} < 2 \), then
\[
\sum_{n=1}^{\infty} p_{00}^{(2n)} < \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} (4pq)^n (1 + \frac{1}{8n}) < 2 \sum_{n=1}^{\infty} (4pq)^n = \frac{8pq}{1 - 4pq} < \infty.
\]

Consequently,
\[
E(N) = \sum_{n=1}^{\infty} p_{00}^{(2n)} \begin{cases} 0 \text{ for } p = q \\ \infty \text{ for } p \neq q \end{cases}
\]

We can generalize the random walk to the \( d \)-dimensional case, where \( d \in \mathbb{N} \). For example, a two-dimensional \((d = 2)\) random walk is a stochastic process \( \{(X_n, Y_n); n = 0, 1, \ldots\} \) for which the state space is the set
\[
S = \{(i, j); i, j \in \{0, 1, 2, \ldots\}, \}
\]
and such that
\[
P[(X_{n+1}, Y_{n+1}) = (i_{n+1}, j_{n+1}) | (X_n, Y_n) = (i_n, j_n)] = \begin{cases} \frac{1}{4} & \text{if } i_{n+1} = i_n + 1, j_{n+1} = j_n \\ \frac{1}{4} & \text{if } i_{n+1} = i_n, j_{n+1} = j_n + 1 \\ \frac{1}{4} & \text{if } i_{n+1} = i_n - 1, j_{n+1} = j_n \\ \frac{1}{4} & \text{if } i_{n+1} = i_n, j_{n+1} = j_n - 1 \\ 0 & \text{otherwise} \end{cases}
\]
\(i, j \in S\) and \( i_1 + p_2 + q_1 + q_2 = 1. \) (4.12)

Thus, the particle can only move to one of its four nearest neighbors. In three dimensions, the particle can move from a given state to one of its six nearest neighbors [the neighbors of the origin being the triplets \((\pm 1, 0, 0), (0, \pm 1, 0)\) and \((0, 0, \pm 1)\), etc.

For another example, consider symmetric simple random walk in the plane. This takes place on the points \((i, j, k)\) with integer coordinates, \(k, j \in \mathbb{Z}\). From any point \((j, k)\) the walk steps to one of the four points \(\{(j \pm 1, k), (j, k \pm 1)\}\), with equal probability.

Lemma 4.3. The probability of returning to the origin at the 2\text{nd} step, for symmetric simple random walk in the plane, started at the origin, is given by
\[
p_{00}^{(2n)} = (C_n^2)^2 \left(\frac{1}{4}\right)^{2n}
\] (4.13)

Proof.
\[
p_{00}^{(2n)} = \sum_{r=0}^{n} \frac{(2n)!}{(r! (n-r)!)^2} \left(\frac{1}{4}\right)^{2n} = \left(\frac{1}{4}\right)^{2n} \sum_{r=0}^{n} \binom{2n}{r} \binom{2n}{n-r} = \left(\frac{1}{4}\right)^{2n} \sum_{r=0}^{n} \binom{2n}{r} \binom{2n}{n-r}
\]
which is what we have desired.

Theorem 4.2. Symmetric simple random walk in the plane, returns to its starting point with probability 1. Or equivalently
\[
\sum_{n=1}^{\infty} p_{00}^{(2n)} = \infty
\] (4.14)
Proof. The proof is easy. By using (3.8) and (4.10), we have
\[
(C_n^n)^2 > \frac{1}{n\pi} 4^{2n} \left(1 + \frac{1}{4n}\right)^{-4},
\]
and
\[
\sum_{n=1}^{\infty} P_{00}^{(2n)} > \frac{1}{16\pi} \sum_{n=1}^{\infty} \frac{1}{n} = \infty.
\]

5 Conclusion

In this work, an appropriate novel method for recognize of states type in random walk chain was introduced. These results shed light on the connections between information theory and Markov chain.

References