International Journal of
Mathematical Modelling & Computations
Vol. 02, No. 03, 2012, 221- 229

An Application of Trajectories Ambiguity in Two-State Markov Chain

M. Khodabin *

Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj, Iran

Received: 12 February 2012; Accepted: 19 June 2012.

Abstract. In this paper, the ambiguity of finite state irreducible Markov chain trajectories is reminded and is obtained for two state Markov chain. I give an applicable example of this concept in President election.

Keywords: Markov chain; Shannon's entropy; Shannon's entropy rate; Stationary distribution; Trajectory of a Markov chain; Stirling's approximation.

Index to information contained in this paper

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. For more details see, for example, [1–3, 6].

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

*Corresponding author. Email: m-khodabin@kiau.ac.ir

continuous or indiscrete time. They form one of the most important classes of random processes and has many applications in :

Queueing theory, statistics, physics, the world's mobile telephone systems, speech recognition, bio informatics, reinforcement learning, internet applications, generating sequences of random numbers, Finance and Economics, dynamic macroeconomics, biological modelling, simulations of brain function, games of chance modelling, algorithmic music composition, advanced baseball analysis, to generate superficially "real-looking" text given a sample document and so forth. For more details I refer to [7].

Laura. Ekroot and Thomas M. Cover[4] have introduced the entropy of Markov trajectory. T. M. and J.A. Thomas [2] has shown an initial link between information theory and some of concepts in Markov chain. Here, I use these concepts to introduce a connection between information theory and Markov chain.

This paper is organized as follows:

Section 2 reminds some short definitions in information theory and Markov chain. Section 3, studies the ambiguity of finite state irreducible Markov chain. In Section 4 an application of trajectories ambiguity in two-state Markov chain is introduced. Finally, Section 5 gives some brief conclusion.

2. Preliminaries

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 $\{X_n; n \geqslant 0\}$.

DEFINITION 2.1 *([5, 7]). A discrete stochastic process $\{X_n; n \geqslant 0\}$ 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

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

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

A probability distribution on the S, such that the distribution at time n is the same as the distribution at the time $(n+1)$, i.e., $\pi = \pi P$ is called a stationary distribution. Note that if the finite state Markov chain is irreducible and aperiodic, then the stationary distribution is unique, and from any starting distribution, the distribution of X_n tends to the stationary distribution as $n \to \infty$.

2.2 *Entropy and entropy rate*

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

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

Definition 2.2 *([2, 6]). The Shannon's entropy $H(X)$ of random variable X is defined by*

$$H(X) = -\sum_{x \in \mathbb{A}} p(x) \log(p(x)), \tag{1}$$

for example the binary entropy function is obtained as

$$H(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \ . \tag{2}$$

Suppose that we have a sequence of random variables, $X_1, X_2, ..., X_n, ...$ The entropy growth rate of the sequence with n is named entropy rate. i.e.

Definition 2.3 *([2]). The entropy rate H_0 of a stochastic process $\{X_i\}$ is defined as*

$$H_0 = lim_{n \to \infty} \frac{H(X_1, X_2, ..., X_n)}{n}, \tag{3}$$

when the limit exists.

The limiting conditional entropy H_0^* of a stochastic process $\{X_i\}$ is defined [2] as

$$H_0^* = lim_{n \to \infty} H(X_n \mid X_{n-1}, H_{n-2}, ..., X_1),$$

when the limit exists.

Theorem 2.4 *([2]) For a stationary stochastic process, both limits exist and they are equal, i.e., $H_0 = H_0^*$.*

The entropy rate of Markov chain is obtained as

$$H_0 = -\sum_i \sum_j \pi_i p_{ij} \log p_{ij}. \tag{4}$$

3. Trajectories ambiguity

The number of bits of randomness in a trajectory of a Markov chain has applications in backgammon, gambling, population growth, and evolution. In this section the entropy of trajectories of finite state irreducible Markov chains is recalled from [4]. Consider a finite state irreducible Markov chain with transition matrix P and initial state $X_1 = i$.

Definition 3.1 *A trajectory t_{ij} from state i to state j of a Markov chain is a path with initial state i, final state j , and no intervening state equal to j.*

The probability $p(t_{ij})$ of a trajectory $t_{ij} = ix_2x_3 \dots x_kj$ is given by

$$p(t_{ij}) = P_{ix_2} P_{x_2x_3} \dots P_{x_kj}$$

Irreducibility of P implies that

$$\sum_{t_{ij}\in\tau_{ij}} p(t_{ij}) = 1,$$

where τ_{ij} is the set of all trajectories from i to j. So, the entropy H_{ij} of the trajectories from i to j is defined by

$$H_{ij} = -\sum_{t_{ij}\in\tau_{ij}} p(t_{ij})\log p(t_{ij}).$$

Theorem 3.2 *([4]) For an irreducible Markov chain, the entropy H_{ii} of the random trajectory from state i back to state i is given by*

$$H_{ii} = \frac{H_0}{\pi_i}, \tag{5}$$

where π_i is the stationary probability for state i and H_0 is given in (3).

Suppose that :

- $P_{i.}$ denote the ith row of the Markov transition matrix P.
- $H = [H_{ij}]$ denote the matrix of trajectory entropies.
- $H^* = \begin{pmatrix} H(P_{1.}) & H(P_{1.}) & \dots & H(P_{1.}) \\ H(P_{2.}) & H(P_{2.}) & \dots & H(P_{2.}) \\ \vdots & \vdots & \vdots & \vdots \\ H(P_{m.}) & H(P_{m.}) & \dots & H(P_{m.}) \end{pmatrix}$ denote the matrix of first step entropies.
- $A = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_m \\ \pi_1 & \pi_2 & \dots & \pi_m \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_m \end{pmatrix}$ denote the matrix of stationary distribution.
- $H_\Delta = \begin{pmatrix} H_{11} & 0 & \dots & 0 \\ 0 & H_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & H_{mm} \end{pmatrix}$ denote the diagonal matrix associated with H.
- $K = (I - P + A)^{-1}(H^* - H_\Delta)$
- $\tilde{K} = \begin{pmatrix} K_{11} & K_{22} & \dots & K_{mm} \\ K_{11} & K_{22} & \dots & K_{mm} \\ \vdots & \vdots & \vdots & \vdots \\ K_{11} & K_{22} & \dots & K_{mm} \end{pmatrix}$ i.e. $\tilde{K}_{ij} = K_{jj}$ for all i , j .

Then, we have the following general theorem.

Theorem 3.3 *([4]) The matrix H of trajectory entropies is given by*

$$H = K - \tilde{K} + H_\Delta \tag{6}$$

4. Application in President example

Suppose that the President of one country tells person A his intention to run or not to run in the next election. Then A relays the news to B, who in turn relays the message to C, and so forth, always to some new person. We assume that there is a probability α that a person will change the answer from "yes" to "no" when transmitting it to the next person and a probability β that he will change it from "no" to "yes". We choose as states the message, either yes or no. Obviously, this process is a homogeneous two-state Markov chain with state space $S = \{\ yes\ ,\ no\ \}$ and transition probability matrix

$$P = \begin{pmatrix} P_{yes,yes} & P_{yes,no} \\ P_{no,yes} & P_{no,no} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \quad 0 < \alpha < 1, \ \ 0 < \beta < 1.$$

By $\pi_0 = [\pi_0(yes), \pi_0(no)] = [a, b]$, where $0 \leqslant a \leqslant 1$ and $0 \leqslant b \leqslant 1$ and $a + b = 1$, it follows that

$$\pi_n = \pi_0 P^n = [a, b] \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n \end{pmatrix}.$$

π_0 represents the President's choice. From irreducibly and finitely state space and by

$$\pi = \pi P,$$

we get

$$\pi = [\pi(yes) = \frac{\beta}{\alpha+\beta}\ ,\ \pi(no) = \frac{\alpha}{\alpha+\beta}], \tag{7}$$

as an unique stationary distribution. Via (7) we have

$$A = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix},$$

on one hand, the entropy rate is

$$H_0 = \frac{\alpha\beta}{\alpha+\beta} \log(\frac{(1-\alpha)^{1-\frac{1}{\alpha}}(1-\beta)^{1-\frac{1}{\beta}}}{\alpha\beta}), \tag{8}$$

and from it the matrix of first step entropies results

$$H^* = \begin{pmatrix} \log(\frac{(1-\alpha)^{\alpha-1}}{\alpha^\alpha}) & \log\frac{(1-\alpha)^{\alpha-1}}{\alpha^\alpha}) \\ \log\frac{(1-\beta)^{\beta-1}}{\beta^\beta}) & \log\frac{(1-\beta)^{\beta-1}}{\beta^\beta}) \end{pmatrix}.$$

Specifically, by (7) and (8) the diagonal matrix associated with H is obtained as

$$H_\Delta = \begin{pmatrix} \alpha\log(\frac{(1-\alpha)^{1-\frac{1}{\alpha}}(1-\beta)^{1-\frac{1}{\beta}}}{\alpha\beta}) & 0 \\ 0 & \beta\log(\frac{(1-\alpha)^{1-\frac{1}{\alpha}}(1-\beta)^{1-\frac{1}{\beta}}}{\alpha\beta}) \end{pmatrix}.$$

Then, it follows that

$$K = (I - P + A)^{-1}(H^* - H_\Delta) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{(\alpha+\beta)^2} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{(\alpha+\beta)^2} \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{(\alpha+\beta)^2} & \frac{\beta}{\alpha+\beta} + \frac{\alpha}{(\alpha+\beta)^2} \end{pmatrix} \times$$

$$\begin{pmatrix} \alpha \log \beta - \frac{\alpha\beta-\alpha}{\beta} \log(1-\beta) & \log \frac{(1-\alpha)^{\alpha-1}}{\alpha^\alpha} \\ \log \frac{(1-\beta)^{\beta-1}}{\beta^\beta} & \beta \log \alpha - \frac{\alpha\beta-\beta}{\alpha} \log(1-\alpha) \end{pmatrix}.$$

Finally, via these relations, we can obtain the matrix H of trajectory entropies by Theorem 3.3.

For more illustrations, see the below tables and diagrams that shows the matrix H components for some values of α and β. Here "0=yes" and "1=no".

$\alpha = \beta$	H_{00}	H_{01}	H_{10}	H_{11}
0.0001	0.0002887	1.443	1.443	0.0002887
0.001	0.002894	1.447	1.447	0.002894
0.01	0.029	1.469	1.469	0.029
0.1	**0.307**	**1.534**	**1.534**	**0.307**
0.2	0.608	1.52	1.52	0.608
0.3	0.877	1.461	1.461	0.877
0.4	1.096	1.37	1.37	1.096
0.5	1.25	1.25	1.25	1.25
0.6	1.323	1.102	1.102	1.323
0.7	1.294	0.925	0.925	1.294
0.8	1.135	0.709	0.709	1.135
0.9	0.788	0.438	0.438	0.788
0.99	0.147	0.074	0.074	0.147
0.999	0.021	0.011	0.011	0.021
0.9999	0.002802	0.001401	0.001401	0.002802

Table 1. Trajectory entropies ($\alpha = \beta$)

It's observe that $H_{00} = H_{11}$ and $H_{01} = H_{10}$.
The most value of randomness is seen in 4th line, whither

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix},$$

and

$$H = \begin{pmatrix} 0.307 & 1.534 \\ 1.534 & 0.307 \end{pmatrix}.$$

So, the most value of ambiguity are in the ($yes \to no \to \ldots \to no$) and ($no \to yes \to \ldots \to yes$) trajectories, i.e., when we know that the persons relays the reality message (no by no and yes by yes) with high probability, we have least ambiguity.

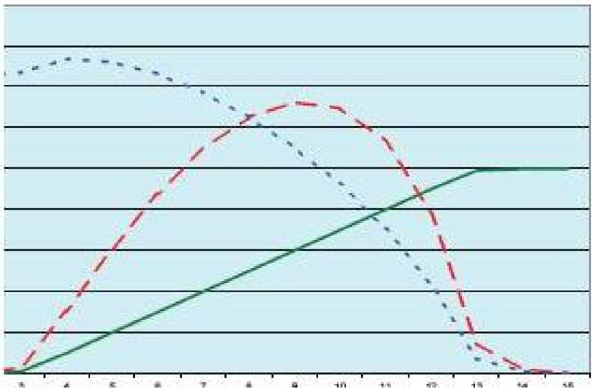

Figure 1. Trajectory entropies for some value ($\alpha = \beta$): the green (solid) line represent α, the red (dash) line represent H_{00} and the blue (dot) line represent H_{01}..

α	β	H_{00}	H_{01}	H_{10}	H_{11}
0.01	0.99	0.016	1.553	0.016	1.553
0.1	0.9	0.207	1.862	0.207	1.862
0.2	**0.8**	**0.467**	**1.866**	**0.467**	**1.866**
0.3	0.7	0.74	1.726	0.74	1.726
0.4	0.6	1.006	1.509	1.006	1.509
0.5	0.5	1.25	1.25	1.25	1.25
0.6	0.4	1.457	0.972	1.457	0.972
0.7	0.3	1.614	0.692	1.614	0.692
0.8	0.2	1.701	0.425	0.425	1.701
0.9	0.1	1.687	0.187	1.687	0.187
0.99	0.01	1.502	0.015	1.502	0.015

Table 2. Trajectory entropies ($\beta = 1 - \alpha$)

It's observe that $H_{00} = H_{10}$ and $H_{01} = H_{11}$.
The most value of randomness is seen in 3th line, whither

$$P = \begin{pmatrix} 0.8\ 0.2 \\ 0.8\ 0.2 \end{pmatrix},$$

and

$$H = \begin{pmatrix} 0.467\ 1.866 \\ 0.467\ 1.866 \end{pmatrix}.$$

So, the most value of ambiguity are in the ($yes \to no \to \ldots \to no$) and ($no \to yes \to \ldots \to no$) trajectories, i.e., when we know that the persons relays the reality message (yes by yes)or the unreality message (no by no) with high probability, we have least ambiguity.

Finally, In table 3 we consider $\beta \neq \alpha$, that are selected randomly.

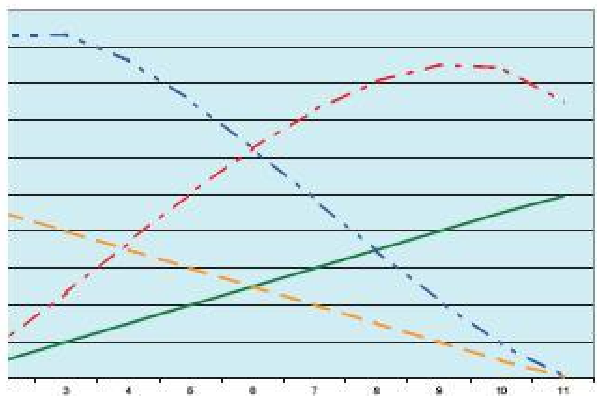

Figure 2. Trajectory entropies for some value ($\beta = 1 - \alpha$): the green (solid) line represent α, light orange (dash)represent β, the red (dash-dot) line represent H_{00} and the blue (dash-dot-dot) line represent H_{01}.

α	β	H_{00}	H_{01}	H_{10}	H_{11}
0.14	0.93	0.283	1.962	0.061	1.881
0.43	0.58	1.077	1.443	1.062	1.453
0.61	0.26	1.549	0.918	1.622	0.66
0.29	0.72	0.708	1.754	0.688	1.758
0.16	**0.04**	**0.505**	**2.591**	**0.566**	**0.126**
0.7	0.1	1.731	0.824	1.648	0.247
0.35	0.73	0.834	1.696	0.687	1.74
0.45	0.23	1.276	1.275	1.56	0.652
0.05	0.22	0.145	0.603	2.29	0.636
0.1	0.61	0.257	1.509	1.06	1.567
0.04	0.95	0.073	1.708	0.118	1.735
0.62	0.24	1.573	0.898	1.639	0.609
0.15	0.34	0.435	1.371	1.531	0.987
0.72	0.74	1.252	0.901	0.838	1.286
0.5	0.75	1.083	1.451	0.716	1.625

Table 3. Trajectory entropies ($\alpha \neq \beta$)

The most value of randomness is seen in 5th line, whither

$$P = \begin{pmatrix} 0.84\ 0.16 \\ 0.04\ 0.96 \end{pmatrix},$$

and

$$H = \begin{pmatrix} 0.505\ 2.591 \\ 0.566\ 0.126 \end{pmatrix}.$$

So, the most value of ambiguity are in the ($yes \rightarrow no \rightarrow \ldots \rightarrow no$) trajectory, i.e., when we know that the persons relays the reality message (yes by yes and no by no) with high probability, we have least ambiguity. Furthermore, when we know that the persons relays unreality message (no by yes) with low probability, we have least ambiguity.

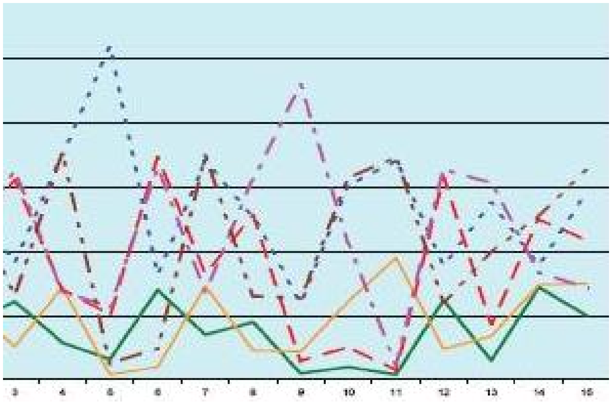

Figure 3. Trajectory entropies for some values of $\alpha \neq \beta$: the green (solid) line represent α, light orange (slim solid)represent β, the red (dashed) line represent H_{00}, the blue (dot) line represent H_{01}, the pink (dash-dot) line represent H_{10} and the brown (dash-dot-dot) line represent H_{11}.

5. Conclusion

Investigations were carried out on Markov chain properties and information theory concepts. These results shed light on the connections between information theory and Markov chain. An appropriate novel method for recognize of states type in random walk chain was proposed.

References

[1] Ash, R.B., *Information Theory* , Interscience, New York, (1965).
[2] Cover, T. M. and Thomas, J.A., *Elements of Information Theory* , John Wiley New York, (1991).
[3] Ekroot, L. and Thomas, M., Cover *The entropy of Markov trajectory*, IEEE transactions on infotmation theory, **39** (1993), 1418-1421.
[4] Kullback, S. and Leibler, R. A., *On information and sufficiency*, Ann. Math. Stat., **22** (1951), 79-86.
[5] Medhi, J., *Stochastic processes*, 3rd ed. New Age International (P) Limited, Publishers, India, (1993).
[6] Shannon, C. E., *A mathematical theory of communication*, Bell Sys. Tech. Journal, **27** (1948).
[7] Stirzaker, D., *Stochastic Processes and Models*, Oxford University Press Inc., New York, (2005).