


Article

Some Generalized Entropy Ergodic Theorems for Nonhomogeneous Hidden Markov Models

Qifeng Yao ^{*,†} , Longsheng Cheng [†], Wenhe Chen and Ting Mao

School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China; cheng_longsheng@njust.edu.cn (L.C.); chenwenhe@njust.edu.cn (W.C.); maoting@njust.edu.cn (T.M.)

* Correspondence: yao_qifeng@njust.edu.cn

[†] These authors contributed equally to this work

Abstract: Entropy measures the randomness or uncertainty of a stochastic process, and the entropy rate refers to the limit of the time average of entropy. The generalized entropy rate in the form of delayed averages can overcome the redundancy of initial information while ensuring stationarity. Therefore, it has better practical value. A Hidden Markov Model (HMM) contains two stochastic processes, a stochastic process in which all states can be observed and a Markov chain in which all states cannot be observed. The entropy rate is an important characteristic of HMMs. The transition matrix of a homogeneous HMM is unique, while a Nonhomogeneous Hidden Markov Model (NHMM) requires the transition matrices to be dependent on time variables. From the perspective of model structure, NHMMs are novel extensions of homogeneous HMMs. In this paper, the concepts of the generalized entropy rate and NHMMs are defined and fully explained, a strong limit theorem and limit properties of a norm are presented, and then generalized entropy ergodic theorems with an almost surely convergence for NHMMs are obtained. These results provide concise formulas for the computation and estimation of the generalized entropy rate for NHMMs.

Keywords: entropy rate; nonhomogeneous hidden Markov models; generalized entropy ergodic theorem; delayed averages

MSC: 60F15; 94A17



Citation: Yao, Q.; Cheng, L.; Chen, W.; Mao, T. Some Generalized Entropy Ergodic Theorems for Nonhomogeneous Hidden Markov Models. *Mathematics* **2024**, *12*, 605. <https://doi.org/10.3390/math12040605>

Academic Editor: Maria C. Mariani

Received: 27 January 2024

Revised: 15 February 2024

Accepted: 16 February 2024

Published: 18 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Shannon introduced entropy, originally a thermodynamic function, into information theory to measure the uncertainty of random phenomena [1]. Shannon's entropy in information theory has been successfully applied in many engineering fields, including vibration-signal feature extraction [2], chaotic image encryption [3], groundwater quality evaluation [4], etc. In order to accurately compute the average uncertainty of a stochastic process, the existence and properties of the entropy rate should be estimated [5,6]. Many approaches have been adopted to improve the theoretical integrity of the entropy rate. One of the famous works is the Shannon–McMillan–Breiman theorem or entropy ergodic theorem, or the asymptotic equipartition property (AEP), which reflects the almost surely (a.s.) convergence of the entropy rate to a constant. Liu and Yang [7] proposed an extension of the Shannon–McMillan–Breiman theorem and some limit properties for nonhomogeneous Markov chains. Yang [8] proved the AEP for a nonhomogeneous Markov information source. Ordentlich et al. [9] used the Blackwell measure to compute the entropy rate. The entropy rate for Hidden Markov Models (HMMs) has been expressed in terms of upper and lower bounds [10]. However, even if states of HMMs are finite, the finite expression of the entropy rate does not exist, which is mainly because the set of their predictive features is generically infinite. To solve this problem, Jurgens et al. [11] evolved the mixed state according to the iterated function system and sampled the entropy of the place-dependent

probability distribution at each step sequentially. Using an arbitrarily long word, the mean of these entropies can converge to the entropy rate.

The entropy rate is defined by the average of random variables throughout the entire process. In fact, we often encounter time series that appear to be “locally stationary”, so we can take an average of what has happened in some window of the recent past. The generalized entropy rate in the form of delayed averages can overcome the redundancy of initial information while ensuring stationarity and, therefore, has better practical value. Essentially, the generalized entropies are nonnegative functions defined on probability distributions that satisfy continuity, maximality, and expansibility. Delayed averages of random variables were first discussed by Zygmund [12]. Using the limiting behavior of delayed averages, Chow [13] proposed necessary and sufficient conditions for the Borel summability of independent identically distributed random variables. Lai [14] studied analogues of the law of the iterated logarithm for delayed averages of independent random variables. On this basis, Gut and Stadtmüller [15] studied the strong law of large numbers on delayed averages of random fields. Wang [16] discussed the limit theorems of delayed averages for row-wise conditionally independent stochastic arrays and a class of asymptotic properties of moving averages for Markov chains in Markovian environments. From these studies, it can be seen that the limit of delayed averages has laid a solid theoretical foundation and is naturally applied to the entropy rate.

Combining the generalized entropy rate and nonhomogeneous Markov chains, Wang and Yang [17,18] studied generalized entropy ergodic theorems with a.s. and \mathcal{L}_1 convergence for time nonhomogeneous Markov chains, and they obtained generalized entropy ergodic theorems for non-null stationary processes using Markov approximation. Shi et al. [19] studied the generalized AEP of higher-order nonhomogeneous Markov information sources by establishing several strong deviation theorems. The entropy rate is an important characteristic of HMMs, and the entropy rate of HMMs plays an important role in applications such as communication decoding, compression, and sorting. Therefore, theoretical research on the entropy rate of HMMs is very necessary. The classical HMMs were first introduced by Baum and Petrie, and they have been widely applied in various fields, including speech recognition, facial expression recognition, gene prediction, gesture recognition, musical composition, bio-informatics, and big data ranking [20–23]. The power of these models is that they can be very efficiently implemented and simulated. A homogeneous HMM contains two stochastic processes: the observed process is assumed to be conditionally temporally independent given the hidden process, and the hidden process is assumed to evolve according to a first-order Markov chain. A Non-homogeneous Hidden Markov Model (NHMM) provides the idea by allowing for the transition matrices of the hidden states to be related to a set of observed covariates, that is, the transition matrix of a homogeneous HMM is unique, while an NHMM requires the transition matrices to be dependent on time variables. From the perspective of model structure, NHMMs are novel extensions of homogeneous HMMs. In the last ten years, new theories on NHMMs have emerged. Yang et al. [24] stated the law of large numbers for countable NHMMs. Zhang et al. [25] studied the stability analysis and controller design for a family of nonhomogeneous hidden semi-Markov jump systems with limited information of sojourn-time probability density functions. Shahzadi et al. [26] proposed a class of Non-homogeneous Hidden Semi-Markov Models for modelling partially observed processes that do not necessarily behave in a stationary and memoryless manner.

Although there have been fruitful achievements in the two fields of generalized entropy ergodic theorems and NHMMs, research on generalized entropy ergodic theorems for NHMMs is limited. Therefore, we consider extending the application scenarios of the generalized entropy rate. Motivated by the above work, the main focus of this paper is to obtain a strong limit theorem of delayed averages of real-number functions and generalized entropy ergodic theorems with an almost surely convergence for NHMMs. These results provide a general idea for the relevant theoretical proof and concise formulas for the computation and estimation of the generalized entropy rate for NHMMs, and they lay the

necessary mathematical, theoretical foundation for the reliability of model applications. The rest of this paper is organized as follows: in Section 2, a detailed description of NHMMs and related definitions are introduced. Section 3 presents some limit properties that are used in Section 4. In Section 4, the main results and the proofs are given. Section 5 summarizes the important content and discusses the significance of the research.

2. Preliminaries

This section provides preliminaries for the subsequent important contents. Continuing with the introduction, we state basic concepts and properties of NHMMs, point out the relationship between NHMMs and homogeneous HMMs, and define the generalized entropy, the generalized entropy rate, and other commonly used symbols on this basis.

Firstly, we give the definition and properties of NHMMs. Let $X = \{X_n, n \geq 0\}$ be a nonhomogeneous Markov chain defined on the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ taking values in a finite state space $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$, and $Y = \{Y_n, n \geq 0\}$ a stochastic process defined on the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ taking values in a finite state space $\mathcal{L} = \{l_1, l_2, \dots, l_M\}$. $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is called an NHMM if and only if it meets the following forms and conditions:

1. The initial distribution of the nonhomogeneous Markov chain $X = \{X_n, n \geq 0\}$ is

$$q^{(0)} = (q_0(s_1), q_0(s_2), \dots, q_0(s_N)), \quad q_0(s_i) = \mathbb{P}(X_0 = s_i), \quad 0 \leq i \leq N \quad (1)$$

and the transition matrices are

$$Q_n = [q_n(s_i; s_j)], \quad (2)$$

where

$$q_n(s_i; s_j) = \mathbb{P}(X_n = s_j | X_{n-1} = s_i), \quad n \geq 1, \quad s_i, s_j \in \mathcal{S}, \quad 1 \leq i, j \leq N. \quad (3)$$

2. For any n

$$\mathbb{P}\{(Y_0, Y_1, \dots, Y_n) = (y_0, y_1, \dots, y_n) | X\} = \prod_{i=0}^n \mathbb{P}\{Y_i = y_i | X_i\}, \quad a.s. \quad (4)$$

where (y_0, y_1, \dots, y_n) are the realizations of (Y_0, Y_1, \dots, Y_n) .

If for any n , $Q_n = Q$, and the conditional probabilities $\mathbb{P}(Y_n = y_n | X_n = x_n)$ are independent on n , where Q is a stochastic matrix, x_n and y_n are realizations of X_n and Y_n , then $X = \{X_n, n \geq 0\}$ is a homogeneous Markov chain and $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is a homogeneous HMM.

Next, we list some properties for NHMMs, which are also equivalent definitions of the model and play a role in proving the theorems. For any $0 \leq m < k$, let $X_m^k = (X_m, X_{m+1}, \dots, X_k)$, $Y_m^k = (Y_m, Y_{m+1}, \dots, Y_k)$, and let x_m^k and y_m^k be the realizations of X_m^k and Y_m^k , respectively. In addition, For any $m, k \geq 0$, let $X_{m,k} = (X_m, X_{m+1}, \dots, X_{m+k})$, $Y_{m,k} = (Y_m, Y_{m+1}, \dots, Y_{m+k})$.

- $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is an NHMM if and only if for any n ,

$$\mathbb{P}\{Y_0^n = y_0^n | X_0^n = x_0^n\} = \prod_{i=1}^n \mathbb{P}\{Y_i = y_i | X_i = x_i\}. \quad (5)$$

- $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is an NHMM if and only if for any n ,

$$\begin{aligned} & \mathbb{P}\{X_0^n = x_0^n, Y_0^n = y_0^n\} \\ &= \mathbb{P}\{X_0 = x_0\} \mathbb{P}\{Y_0 = y_0 | X_0 = x_0\} \prod_{i=1}^n \mathbb{P}\{X_i = x_i | X_{i-1} = x_{i-1}\} \mathbb{P}\{Y_i = y_i | X_i = x_i\}. \end{aligned} \tag{6}$$

- $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is an NHMM if and only if for any n ,

$$\mathbb{P}\{X_n = x_n | X_0^{n-1} = x_0^{n-1}, Y_0^{n-1} = y_0^{n-1}\} = \mathbb{P}\{X_n = x_n | X_{n-1} = x_{n-1}\} \tag{7}$$

and

$$\mathbb{P}\{Y_n = y_n | X_0^n = x_0^n, Y_0^{n-1} = y_0^{n-1}\} = \mathbb{P}\{Y_n = y_n | X_n = x_n\}. \tag{8}$$

By Equations (7) and (8), we have

$$\begin{aligned} & \mathbb{P}\{X_{n+1} = x_{n+1}, Y_{n+1} = y_{n+1} | X_0^n = x_0^n, Y_0^n = y_0^n\} \\ &= \mathbb{P}\{X_{n+1} = x_{n+1}, Y_{n+1} = y_{n+1} | X_n = x_n\}. \end{aligned} \tag{9}$$

In the following text, probability measures are widely used. Therefore, for any $k \geq 1$, denote

$$p_k(x_k; y_k) = \mathbb{P}(Y_k = y_k | X_k = x_k), \tag{10}$$

$$p_k(x_{k-1}; x_k, y_k) = \mathbb{P}(Y_k = y_k, X_k = x_k | X_{k-1} = x_{k-1}), \tag{11}$$

$$p(x_0, y_0, \dots, x_k, y_k) = \mathbb{P}(X_0 = x_0, Y_0 = y_0, \dots, X_k = x_k, Y_k = y_k), \tag{12}$$

and

$$\begin{aligned} & p(x_0, y_0, \dots, x_{k-1}, y_{k-1}; x_k, y_k) \\ &= \mathbb{P}(X_k = x_k, Y_k = y_k | X_0 = x_0, Y_0 = y_0, \dots, X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1}). \end{aligned} \tag{13}$$

A delayed average is one of the widely known technical indicators used to predict future data in time series analysis, and we define the entropy and the entropy rate in the form of delayed averages. Let $(f(n))_{n=0}^\infty$ be a sequence of non-negative integers such that $f(n) \rightarrow \infty$ as $n \rightarrow \infty$.

For simplicity, we use the natural logarithm here; thus, the generalized entropy can be measured. According to information theory, the definition of the entropy of an NHMM $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is

$$H(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n) = E[-\log \mathbb{P}(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n)], \tag{14}$$

where $E[\cdot]$ represents the expectation, and the definition of the entropy rate of an NHMM $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ is

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n). \tag{15}$$

Combining the concept of the generalized entropy with the properties of NHMMs, we have

$$\begin{aligned}
 & H(X_{n,f(n)}, Y_{n,f(n)}) \\
 &= H(X_n, Y_n, \dots, X_{n+f(n)}, Y_{n+f(n)}) \\
 &= H(X_n, Y_n) + \sum_{k=n+1}^{n+f(n)} H(X_k, Y_k | X_n, Y_n, \dots, X_{k-1}, Y_{k-1}) \\
 &= H(X_n, Y_n) - \sum_{k=n+1}^{n+f(n)} \sum_{x_n, \dots, x_k \in \mathcal{S}} \sum_{y_n, \dots, y_k \in \mathcal{L}} p(x_n, y_n, \dots, x_k, y_k) \log p(x_n, y_n, \dots, x_{k-1}, y_{k-1}; x_k, y_k) \\
 &= H(X_n, Y_n) - \sum_{k=n+1}^{n+f(n)} \sum_{x_{k-1} \in \mathcal{S}} q_k(x_{k-1}) \sum_{x_k \in \mathcal{S}} \sum_{y_k \in \mathcal{L}} p_k(x_{k-1}; x_k, y_k) \log p_k(x_{k-1}; x_k, y_k) \\
 &= H(X_n, Y_n) - \sum_{k=n+1}^{n+f(n)} \sum_{x_{k-1} \in \mathcal{S}} q_k(x_{k-1}) \sum_{x_k \in \mathcal{S}} \sum_{y_k \in \mathcal{L}} q_k(x_{k-1}; x_k) p_k(x_k; y_k) \log q_k(x_{k-1}; x_k) p_k(x_k; y_k) \\
 &= H(X_n, Y_n) + \sum_{k=n+1}^{n+f(n)} H(X_k, Y_k | X_{k-1}), \tag{16}
 \end{aligned}$$

where $X_{n,f(n)}$ and $Y_{n,f(n)}$ denotes $(X_n, X_{n+1}, \dots, X_{n+f(n)})$ and $(Y_n, Y_{n+1}, \dots, Y_{n+f(n)})$, respectively, and we use the distribution

$$q^{(k)} = (q_k(s_1), q_k(s_2), \dots, q_k(s_N)), \quad q_k(s_i) = \mathbb{P}(X_k = s_i), \quad 0 \leq i \leq N, \tag{17}$$

which is the same below.

To prepare for the following text, we introduce concepts that may frequently appear.

Definition 1. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM defined as above. Define the generalized entropy density as

$$\begin{aligned}
 F_{n,f(n)}(\omega) &= -\frac{1}{f(n)} \log \mathbb{P}(X_{n,f(n)}, Y_{n,f(n)}) \\
 &= -\frac{1}{f(n)} \left\{ \log \mathbb{P}(X_n, Y_n) + \sum_{k=n+1}^{n+f(n)} \log \mathbb{P}(X_k, Y_k | X_{k-1}, Y_{k-1}) \right\}, \tag{18}
 \end{aligned}$$

where $\omega \in \Omega$.

In the following text, we prove that the limit of the entropy density is the entropy rate.

Definition 2 ([21]). Let $X = \{X_n, n \geq 0\}$ be a homogeneous Markov chain. Let Q be a transition matrix of $X = \{X_n, n \geq 0\}$. Q is called strongly ergodic if there exists a probability distribution $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ in \mathcal{S} satisfying

$$\sup_{q^{(0)}} \|q^{(0)} Q^n - \pi\| \rightarrow 0, \quad \text{as } n \rightarrow \infty \tag{19}$$

where $q^{(0)}$ is a starting vector. Obviously, Equation (19) implies $\pi Q = \pi$, and π is called the stationary distribution determined by Q .

Definition 3. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ be a vector, and then the norm of α is defined by

$$\|\alpha\| = \sum_{i=1}^N |\alpha_i|. \tag{20}$$

Let $A = (a_{ij})_{N \times N}$ be a square matrix, and then the norm of A is defined by

$$\|A\| = \max_i \sum_{j=1}^N |a_{ij}|. \tag{21}$$

3. Some Limit Properties

In this section, we give some lemmas, which are used to prove the main conclusions. These lemmas include a strong limit theorem for NHMMs and limit properties of a norm. Lemma 1 provides a strong limit theorem for bounded functions of NHMMs. Lemma 2 gives the convergence of bounded functions of real sequences on averages. Lemma 3 points out that under the condition of the convergence of the transition matrices of NHMMs, the counting averages of Markov chains converge to a stationary distribution of irreducible matrices. Lemma 4 points out the convergence of the transition matrix multiplication average. Lemma 5 proves the relationship between the convergence of vector sequences on averages and the convergence of subsequences.

Lemma 1. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM. Let $g_n(\cdot, \cdot, \cdot)$ be a sequence of bounded real number functions defined on $S^2 \times \mathcal{L}$. If for any $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \exp\{-\epsilon f(n)\} < \infty, \tag{22}$$

and there exists a positive number γ such that for any $s \in S, l \in \mathcal{L}$,

$$\limsup_{n \rightarrow \infty} \max_{t \in S} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} g_k(t, s, l) p_k(t; s, l) \exp\{\gamma |g_k(t, s, l)|\} = C_\gamma(s, l) < \infty, \tag{23}$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \{g_k(X_{k-1}, X_k, Y_k) - E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]\} = 0, \text{ a.s.} \tag{24}$$

where $E[\cdot | \cdot]$ represents conditional expectation.

Proof. Let u be a nonzero real number, and define

$$\Lambda_{n,f(n)}(u, \omega) = \frac{\exp\{u \sum_{k=n+1}^{n+f(n)} g_k(X_{k-1}, X_k, Y_k)\}}{\prod_{k=n+1}^{n+f(n)} E[e^{g_k(X_{k-1}, X_k, Y_k)} | X_{k-1}]}, \quad n = 1, 2, \dots \tag{25}$$

where $\omega \in \Omega$ and $E[\cdot | \cdot]$ represents conditional expectation. By the properties of conditional expectations, we have

$$\begin{aligned} & E\Lambda_{n,f(n)}(u, \omega) \\ &= E[E[\Lambda_{n,f(n)}(u, \omega) | X_{0,n+f(n)-1}]] \\ &= E[E[\Lambda_{n,f(n)-1}(u, \omega) \frac{e^{u g_{n+f(n)}(X_{n+f(n)-1}, X_{n+f(n)}, Y_{n+f(n)})}}{E[e^{g_{n+f(n)}(X_{n+f(n)-1}, X_{n+f(n)}, Y_{n+f(n)})} | X_{n+f(n)-1}]} | X_{0,n+f(n)-1}]] \\ &= E[\frac{\Lambda_{n,f(n)-1}(u, \omega) E[e^{u g_{n+f(n)}(X_{n+f(n)-1}, X_{n+f(n)}, Y_{n+f(n)})} | X_{n+f(n)-1}]}{E[e^{g_{n+f(n)}(X_{n+f(n)-1}, X_{n+f(n)}, Y_{n+f(n)})} | X_{n+f(n)-1}]}] \\ &= E\Lambda_{n,f(n)-1}(u, \omega) = \dots = E\Lambda_{n,1}(u, \omega) = 1. \end{aligned} \tag{26}$$

For any $\epsilon > 0$, by Markov inequality and Equation (25), we have

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}[f^{-1}(n) \log \Lambda_{n,f(n)}(u, \omega) \geq \epsilon] &= \sum_{n=1}^{\infty} \mathbb{P}[\Lambda_{n,f(n)}(u, \omega) \geq \exp(\epsilon f(n))] \\ &\leq \sum_{n=1}^{\infty} 1 \cdot \exp\{-\epsilon f(n)\} < \infty. \end{aligned} \tag{27}$$

Combining the Borel–Cantelli Lemma and the arbitrariness of ϵ , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{f(n)} \log \Lambda_{n,f(n)}(u, \omega) \leq 0. \text{ a.s.} \tag{28}$$

Expanding Equation (28), we have

$$\limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \{u g_k(X_{k-1}, X_k, Y_k) - \log E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]\} \leq 0. \text{ a.s.} \tag{29}$$

Let $0 < u < \gamma$. Using the inequalities $\log x \leq x - 1 (x > 0)$ and $0 \leq e^x - 1 - x \leq \frac{1}{2}x^2 e^{|x|} (x \in \mathbb{R})$, we have

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \{g_k(X_{k-1}, X_k, Y_k) - E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]\} \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \left\{ \frac{1}{u} \log E[e^{u g_k(X_{k-1}, X_k, Y_k)} | X_{k-1}] - E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}] \right\} \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \left\{ \frac{E[e^{u g_k(X_{k-1}, X_k, Y_k)} - 1 - u g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]}{u} \right\} \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} E[g_k^2(X_{k-1}, X_k, Y_k) e^{u |g_k(X_{k-1}, X_k, Y_k)|} | X_{k-1}] \\ &\leq \frac{1}{2} u C_{\gamma}(s, l) < \infty. \text{ a.s.} \end{aligned} \tag{30}$$

Let $u \downarrow 0$ in Equation (30), then

$$\limsup_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \{g_k(X_{k-1}, X_k, Y_k) - E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]\} \leq 0. \text{ a.s.} \tag{31}$$

Similarly, let $-\gamma < u < 0$, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \{g_k(X_{k-1}, X_k, Y_k) - E[g_k(X_{k-1}, X_k, Y_k) | X_{k-1}]\} \geq 0. \text{ a.s.} \tag{32}$$

Equation (24) follows immediately from inequalities (31) and (32). \square

The key to prove Lemma 1 is to construct the likelihood ratio. By using the approximation of the upper and lower limits, it is possible to obtain an almost surely convergence of bounded functions. This lemma is used to prove Theorem 1.

Lemma 2 ([17]). Let $h(z)$ be a bounded function defined on an real-number interval D and $\{z_n\}_{n=0}^\infty$ a sequence in D . If $\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |z_k - z| = 0$, $h(z)$ is continuous at point z , and Equation (22) holds, then

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |h(z_k) - h(z)| = 0. \tag{33}$$

Lemma 3 ([17]). Let Q be an irreducible transition matrix. If Equation (22) holds, and for any $t, s \in \mathcal{S}$,

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |q_k(s; t) - q(s; t)| = 0, \tag{34}$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n}^{n+f(n)-1} 1_{\{s_i\}}(X_k) = \pi_i \text{ a.s.} \tag{35}$$

holds for any $s_i \in \mathcal{S}, 1 \leq i \leq N$, where $1_{\{s_i\}}(X_k)$ is an indicative function and $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ is the unique stationary distribution determined by Q .

These two lemmas serve as support for theorem 2. It should be emphasized that Q in Lemma 3 is irreducible, so this lemma still holds for ergodic matrices.

Lemma 4. Let $X = \{X_n, n \geq 0\}$ be a nonhomogeneous Markov chain with transition matrices $\{Q_n, n \geq 1\}$. Let Q be a periodic strong ergodic stochastic matrix. Assume that $c = (c_1, c_2, \dots, c_N)$ is a left eigenvector of Q , the unique solution of equations $cQ = c$. Let B be a constant random matrix, where each row of B is c . If Equation (22) holds and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|Q_k - Q\| = 0, \tag{36}$$

then

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} Q^{(m, m+k-1)} - B \right\| = 0 \tag{37}$$

holds for any $m \in \mathbb{N}$, where $Q^{(m, m+k-1)} = Q_m \dots Q_{m+k-1}$, $Q^{(m, m)} = Q_m$, $Q_0 = I$ (I is the identical matrix).

The proof process of Lemma 4 is similar to theorem 1 of [27]. It should be emphasized that the matrices appearing in Lemma 4 are composed of constants, so the norm can be calculated. Lemma 4 is one of the prerequisites for Theorem 3.

Lemma 5. Let $\{\beta_n\}_{n=1}^\infty$ and β be column vectors with real entries. If Equation (22) holds and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_k - \beta\| = 0, \tag{38}$$

then there exists a subsequence $\{\beta_{n_k}\}_{k=1}^\infty$ of $\{\beta_n\}_{n=1}^\infty$ such that

$$\lim_{k \rightarrow \infty} \|\beta_{n_k} - \beta\| = 0. \tag{39}$$

Proof. Constructing an inequality, there exists $m \in \mathbb{N}$

$$\frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_{m+k} - \beta\| \leq \frac{m+f(n)}{f(n)} \cdot \frac{1}{m+f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_k - \beta\|. \tag{40}$$

From Equation (40), it can be concluded that there exists $m \in \mathbb{N}$, such that

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_{m+k} - \beta\| = 0. \tag{41}$$

holds. Choose a positive sequence $\{b_k\} \downarrow 0$, and by Equation (41), $\exists u, v \in \mathbb{N}$,

$$\frac{1}{v} \sum_{k=u+1}^{u+v} \|\beta_{m+k} - \beta\| \leq b_1. \tag{42}$$

Therefore, there exists $u \leq n_1 \leq u + v$ such that $\|g_{n_1} - g\| \leq b_1$. By (40), we have

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_{n_1+k} - \beta\| = 0. \tag{43}$$

Similarly, there exists $n_2 \geq n_1$ such that $\|\beta_{n_2} - \beta\| \leq a_1$. Generally, we can get a subsequence $\{\beta_{n_k}\}_{k=1}^\infty$ of $\{\beta_n\}_{n=1}^\infty$ such that $\|\beta_{n_k} - \beta\| \leq b_k, k = 1, 2, 3, \dots$. Equation (39) follows immediately. \square

The key to the proof of Lemma 5 is to construct inequality (40). According to the properties of the convergent sequence, even if the sequence is not monotonic, there exists $m \in \mathbb{N}$, and inequality (40) holds. Lemma 5 is also one of the prerequisites for Theorem 3.

4. Generalized Entropy Ergodic Theorems

In this section, we give the main results and their proofs. Generalized entropy ergodic theorems with an almost surely convergence for NHMMs are presented. These results provide concise formulas for the computation and estimation of the generalized entropy rate for NHMMs.

Theorem 1. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM. If for any $\epsilon > 0$,

$$\sum_{n=1}^\infty \exp\{-\epsilon f(n)\} < \infty, \tag{44}$$

and there exists a positive number γ , such that for any $s \in \mathcal{S}, l \in \mathcal{L}$,

$$\limsup_{n \rightarrow \infty} \max_{t \in \mathcal{S}} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} g_k(t, s, l) p_k(t; s, l) \exp\{\gamma |g_k(t, s, l)|\} = C_\gamma(s, l) < \infty, \tag{45}$$

then

$$\lim_{n \rightarrow \infty} F_{n, f(n)}(\omega) = \lim_{n \rightarrow \infty} -\frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p_k(X_{k-1}; s, l) \log p_k(X_{k-1}; s, l). \text{ a.s.} \tag{46}$$

Proof. Set $g_k(X_{k-1}; s, l) = -\log p_k(X_{k-1}; s, l)$ and $\gamma = \frac{1}{2}$ in Lemma 3. Using the inequality $(\log x)^2 x^{\frac{1}{2}} \leq e^{-2} (0 \leq x \leq 1)$, we can conclude that for any $s \in \mathcal{S}, l \in \mathcal{L}$,

$$\begin{aligned}
 & C_{\frac{1}{2}}(s, l) \\
 &= \limsup_{n \rightarrow \infty} \max_{t \in \mathcal{S}} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} [\log p_k(t; s, l)]^2 p_k(t; s, l) \exp\left\{\frac{1}{2} |\log p_k(t; s, l)|\right\} \\
 &\leq 16 \exp\{-2\}.
 \end{aligned} \tag{47}$$

It is not hard to verify that

$$\begin{aligned}
 & \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \left\{g_k(X_{k-1}, X_k, Y_k) - \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} g_k(X_{k-1}, s, l) p_k(X_{k-1}; s, l)\right\} \\
 &= -\frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \left\{\log p_k(X_{k-1}; X_k, Y_k) - \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p_k(X_{k-1}; s, l) \log p_k(X_{k-1}; s, l)\right\} \\
 &= -\frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \left\{\log p_k(X_{k-1}; X_k, Y_k) - \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p_k(X_{k-1}; s, l) \log p_k(X_{k-1}; s, l)\right\} \\
 &= -\frac{1}{f(n)} \log \mathbb{P}(X_n, Y_n) + F_{n, f(n)}(\omega) \\
 &+ \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p_k(X_{k-1}; s, l) \log p_k(X_{k-1}; s, l).
 \end{aligned} \tag{48}$$

The transition probability falls within the interval $[0,1]$, so its logarithm is bounded. In addition, infinitesimal multiplication by a bounded quantity is still infinitesimal, so the initial distribution is erased. Hence the conclusion can be deduced immediately from Lemma 1. \square

Theorem 2. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM. Assume that for any $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \exp\{-\epsilon f(n)\} < \infty, \tag{49}$$

and there exists a positive number γ , such that for any $s \in \mathcal{S}, l \in \mathcal{L}$,

$$\limsup_{n \rightarrow \infty} \max_{t \in \mathcal{S}} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} g_k(t, s, l) p_k(t; s, l) \exp\{\gamma |g_k(t, s, l)|\} = C_{\gamma}(s, l) < \infty. \tag{50}$$

If for any $(t, s, l) \in \mathcal{S}^2 \times \mathcal{L}$,

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |p_k(t; s, l) - p(t; s, l)| \\
 &= \lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |q_k(t; s) p_k(s; l) - q(t; s) p(s; l)| = 0
 \end{aligned} \tag{51}$$

holds, where q represents terms of an irreducible transition matrix Q , then

$$\lim_{n \rightarrow \infty} F_{n, f(n)}(\omega) = -\sum_{t \in \mathcal{S}} \pi(t) \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p(t; s, l) \log p(t; s, l), \tag{52}$$

where $\pi(t), t \in \mathcal{S}$ belongs to $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, which is a stationary distribution of Q .

Proof. Set $h(z) = z \log z$ in Lemma 2; then, for any $(t, s, l) \in \mathcal{S}^2 \times \mathcal{L}$,

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |p_k(t; s, l) \log p_k(t; s, l) - p(t; s, l) \log p(t; s, l)| = 0. \tag{53}$$

Using absolute inequality, we have

$$\begin{aligned} & |F_{n,f(n)}(\omega) + \sum_{t \in \mathcal{S}} \pi(t) \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p(t; s, l) \log p(t; s, l)| \\ & \leq |F_{n,f(n)}(\omega) + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{t \in \mathcal{S}} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} 1_{\{t\}}(X_{k-1}) p_k(t; s, l) \log p_k(t; s, l)| \\ & \quad + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{t \in \mathcal{S}} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} 1_{\{t\}}(X_{k-1}) |p_k(t; s, l) \log p_k(t; s, l) - p(t; s, l) \log p(t; s, l)| \\ & \quad + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{t \in \mathcal{S}} [\pi(t) - 1_{\{t\}}(X_{k-1})] \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p(t; s, l) \log p(t; s, l)| \\ & \leq |F_{n,f(n)}(\omega) + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \sum_{t \in \mathcal{S}} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} 1_{\{t\}}(X_{k-1}) p_k(t; s, l) \log p_k(t; s, l)| \\ & \quad + \sum_{t \in \mathcal{S}} \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} |p_k(t; s, l) \log p_k(t; s, l) - p(t; s, l) \log p(t; s, l)| \\ & \quad + \sum_{t \in \mathcal{S}} |\pi(t) - \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} 1_{\{t\}}(X_{k-1})| \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} p(t; s, l) \log p(t; s, l) \end{aligned} \tag{54}$$

where $\pi(t), t \in \mathcal{S}$ belongs to $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, which is a stationary distribution of Q . By Lemma 3, Equation (52) follows from Equation (54). \square

The proof of Theorem 1 mainly relies on the construction of inequalities, and the proof of Theorem 2 utilizes the properties of norms. The two theorems explain that the limit of the generalized entropy density is the generalized entropy rate.

Theorem 3. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM. $Q = (q(t; s))_{N \times N}$ is another transition matrix, and assume that Q is periodic and strongly ergodic. Let

$$\beta_n(t) = \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} q_n(t; s) p_n(s; l) \log q_n(t; s) p_n(s; l), \tag{55}$$

$$\beta(t) = \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} q(t; s) p(s; l) \log q(t; s) p(s; l), \tag{56}$$

where $\beta_n(t), \beta(t), t \in \mathcal{S}$ are the elements of column vectors β_n and β , respectively, and assume that $\{\|\beta_n\|, n \geq 1\}$ are bounded. If Equation (22) holds,

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|Q_k - Q\| = 0, \tag{57}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_k - \beta\| = 0, \tag{58}$$

then the generalized entropy rate of $\{X, Y\}$ exists, and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} H(X_n, Y_n, \dots, X_{n+f(n)}, Y_{n+f(n)}) = -\pi\beta, \text{ a.s.} \tag{59}$$

where $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ is the unique stationary distribution determined by Q .

Proof. Let $q^{(k-1)}$ be a row vector with elements $\mathbb{P}(X_{k-1} = t), t \in \mathcal{S}$. Hence, using the definition and properties of β_n , we have

$$H(X_k, Y_k | X_{k-1}) = -q^{(k-1)}\beta_k. \tag{60}$$

Simply take B as a constant random matrix whose rows are equal to π . Note that $\pi = q^{(0)}B$, where $q^{(0)}$ is an initial distribution of the Markov chain. Since

$$\begin{aligned} & \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)} - \pi \right\| \\ &= \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(0)}Q^{(0,k-1)} - q^{(0)}B \right\| \\ &\leq \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} Q^{(0,k-1)} - B \right\|, \end{aligned} \tag{61}$$

where $Q^{(0,k-1)} = Q_1 \dots Q_{k-1}$, $Q^{(0,0)} = I$ (I is the identical matrix), by Equations (57) and (61) and Lemma 5, we have

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)} - \pi \right\| = 0. \tag{62}$$

By Equation (58) and Lemma 5, there exists a subsequence $\{\beta_{n_k}\}_{k=1}^\infty$ of $\{\beta_n\}_{n=1}^\infty$ such that

$$\lim_{k \rightarrow \infty} \|\beta_{n_k} - \beta\| = 0. \tag{63}$$

Hence, $\|\beta\|$ is finite. By Equation (62) and properties of entropy, we have

$$\begin{aligned} & \left| \frac{1}{f(n)} H(X_n, Y_n, \dots, X_{n+f(n)}, Y_{n+f(n)}) + \pi\beta \right| \\ &= \left| \frac{1}{f(n)} H(X_n, Y_n) + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} H(X_k, Y_k | X_{k-1}) + \pi\beta \right| \\ &\leq \left| \frac{1}{f(n)} H(X_n, Y_n) \right| + \left| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)}\beta_k - \pi\beta \right| \\ &\leq \left| \frac{1}{f(n)} H(X_n, Y_n) \right| + \left| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)}\beta_k - \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)}\beta \right| + \left| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)}\beta - \pi\beta \right| \\ &\leq \frac{1}{f(n)} |H(X_n, Y_n)| + \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_k - \beta\| + \|\beta\| \left\| \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} q^{(k-1)} - \pi \right\| \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned} \tag{64}$$

This completes the proof of Theorem 3. \square

Theorem 3 gives a method to compute the generalized entropy rate of an NHMM under some mild conditions, and when the model degenerates into a homogeneous HMM and a nonhomogeneous Markov chain, the results are still the same. The two corollaries are existing results, which indirectly demonstrate the correctness of this theorem.

Corollary 1. Let $\{X, Y\} = \{(X_n, Y_n), n \geq 0\}$ be an NHMM with a periodic and strongly ergodic transition matrix $Q = (q(t; s))_{N \times N}$ and an emission probability matrix $P = (p(s; l))_{N \times M}$, where $t, s \in \mathcal{S}, l \in \mathcal{L}$. Let

$$\beta(t) = \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{L}} q(t; s)p(s; l) \log q(t; s)p(s; l), \tag{65}$$

where $\beta(t), t \in \mathcal{S}$ are the elements of column vector β . Assume that $\|\beta\|$ is bounded. If Equation (22) holds, then the generalized entropy rate of $\{X, Y\}$ exists, and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} H(X_n, Y_n, \dots, X_{n+f(n)}, Y_{n+f(n)}) = -\pi\beta, \text{ a.s.} \tag{66}$$

where $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ is the unique stationary distribution determined by Q .

Corollary 2. Let $X = \{X_n, n \geq 0\}$ be a nonhomogeneous Markov chain. $Q = (q(t; s))_{N \times N}$ is another transition matrix, and assume that Q is periodic and strongly ergodic, where $t, s \in \mathcal{S}$. Let

$$\beta_n(t) = \sum_{s \in \mathcal{S}} q_n(t; s) \log q_n(t; s), \tag{67}$$

$$\beta(t) = \sum_{s \in \mathcal{S}} q(t; s) \log q(t; s), \tag{68}$$

where $\beta_n(t), \beta(t), t \in \mathcal{S}$ are the elements of column vectors β_n and β , respectively. Assume that $\{\|\beta_n\|, n \geq 1\}$ are bounded. If Equation (22) holds,

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|Q_k - Q\| = 0, \tag{69}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} \sum_{k=n+1}^{n+f(n)} \|\beta_k - \beta\| = 0, \tag{70}$$

then the generalized entropy rate of X exists, and

$$\lim_{n \rightarrow \infty} \frac{1}{f(n)} H(X_n, \dots, X_{n+f(n)}) = -\pi\beta, \text{ a.s.} \tag{71}$$

where $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ is the unique stationary distribution determined by Q .

5. Conclusions

Entropy ergodic theorems reflect the almost surely convergence of the entropy rate to a constant. The generalized entropy rate in the form of delayed averages can overcome the redundancy of initial information while ensuring stationarity and, therefore, it has better practical value. An NHMM provides the idea by allowing the transition matrix of the hidden states in a homogeneous HMM to be related to a set of observed covariates. Although there have been fruitful achievements in the two fields of generalized entropy ergodic theorems and NHMMs, research on the generalized entropy ergodic theorems for NHMMs is limited. Therefore, we consider extending the application scenarios of the gen-

eralized entropy rate. In this paper, we give the basic concepts of NHMMs and the entropy rate in the form of delayed averages, list some lemmas including a strong limit theorem and limit properties of a norm, and prove some generalized entropy ergodic theorems with an almost surely convergence for NHMMs. These results, which are generalizations of previous findings of [17,27], provide concise formulas for the computation and estimation of the generalized entropy rate in the form of delayed averages. Theoretical analysis has proven the existence and limit formula of entropy rate, but its numerical calculation is also very complex and challenging. Specifically, this process requires calculating the integration on a certain measure, which is related to the parameters of HMMs. Many methods have been proposed for this, such as approximate formulas, series expansion, and statistical calculation methods. Due to the fact that approximate formulas are not applicable to any parameters, and their accuracy is difficult to estimate, the convergence conditions of the series are difficult to prove, and the sample size required for statistical convergence is difficult to estimate. Therefore, approximating with upper and lower limits is a reliable method, and its accuracy can be estimated, with relatively lenient requirements for HMM parameters. In the future, we will conduct numerical case studies based on the theoretical analysis above.

Author Contributions: Conceptualization, Q.Y. and L.C.; methodology, Q.Y. and W.C.; writing—original draft preparation, Q.Y. and L.C.; writing—review and editing, T.M.; funding acquisition, Q.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant number KYCX21_0358).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|------|------------------------------------|
| HMM | Hidden Markov Model |
| NHMM | Nonhomogeneous Hidden Markov Model |
| AEP | Asymptotic equipartition property |
| a.s. | Almost surely |

References

- Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *37*, 379–423. [[CrossRef](#)]
- Bafroui, H.H.; Ohadi, A. Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. *Neurocomputing* **2014**, *133*, 437–445. [[CrossRef](#)]
- Li, C.; Lin, D.; Feng, B.; Lv, J.; Hao, F. Cryptanalysis of a chaotic image encryption algorithm based on information entropy. *IEEE Access* **2018**, *6*, 75834–75842. [[CrossRef](#)]
- Hasan, M.S.U.; Rai, A.K. Groundwater quality assessment in the Lower Ganga Basin using entropy information theory and GIS. *J. Clean. Prod.* **2020**, *274*, 123077. [[CrossRef](#)]
- Jacquet, P.; Seroussi, G.; Szpankowski, W. On the entropy of a hidden Markov process. *Theor. Comput. Sci.* **2008**, *395*, 203–219. [[CrossRef](#)] [[PubMed](#)]
- Feutrill, A.; Roughan, M. A review of Shannon and differential entropy rate estimation. *Entropy* **2021**, *23*, 1046. [[CrossRef](#)] [[PubMed](#)]
- Liu, W.; Yang, W.G. An extension of Shannon-McMillan theorem and some limit properties for nonhomogeneous Markov chains. *Stoch. Proc. Appl.* **1996**, *61*, 129–145.
- Yang, W.G. The Asymptotic Equipartition Property for a Nonhomogeneous Markov Information Source. *Probab. Eng. Inform. Sci.* **1998**, *12*, 509–518. [[CrossRef](#)]
- Ordentlich, E.; Weiman, T. On the optimality of symbol by symbol filtering and denoising. *IEEE Trans. Inform. Theory* **2006**, *52*, 19–40. [[CrossRef](#)]
- Travers, N.F. Exponential bounds for convergence of entropy rate approximations in hidden Markov models satisfying a path-mergeability condition. *Stoch. Proc. Appl.* **2014**, *124*, 4149–4170. [[CrossRef](#)]
- Jurgens, A.M.; Crutchfield, J.P. Shannon entropy rate of hidden Markov processes. *J. Stat. Phys.* **2021**, *183*, 32. [[CrossRef](#)]

12. Zygmund, A. *Trigonometric Series 1*; Cambridge University Press: Cambridge, UK, 1959.
13. Chow, Y.S. Delayed sums and Borel summability for independent, identically distributed random variables. *Bull. Inst. Math. Acad. Sin.* **1972**, *1*, 286–291.
14. Lai, Z.L. Limit theorems for delayed sums. *Ann. Probab.* **1974**, *2*, 432–440. [[CrossRef](#)]
15. Gut, A.; Stradtmüller, U. On the strong law of large numbers for delayed sums and random fields. *Acta. Math. Hung.* **2010**, *129*, 182–203. [[CrossRef](#)]
16. Wang, Z.Z. A kind of asymptotic properties of moving averages for Markov chains in Markovian environments. *Commun. Stat.-Theory Methods* **2017**, *46*, 10926–10940.
17. Wang, Z.Z.; Yang, W.G. The generalized entropy ergodic theorem for nonhomogeneous Markov chains. *J. Theor. Probab.* **2016**, *29*, 761–775. [[CrossRef](#)]
18. Wang, Z.Z.; Yang, W.G. Markov approximation and the generalized entropy ergodic theorem for non-null stationary process. *Proc. Indian-Math. Sci.* **2020**, *130*, 1–13. [[CrossRef](#)]
19. Shi, Z.; Zhu, X. The generalized AEP for higher order nonhomogeneous Markov information source. *Res. Sq.* **2023**. [[CrossRef](#)]
20. Baum, L.E.; Petrie, T. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* **1966**, *37*, 1554–1563. [[CrossRef](#)]
21. Isaacson, D.; Madsen, R. *Markov Chains Theory and Applications*; Wiley: New York, NY, USA, 1976.
22. Mor, B.; Garhwal, S.; Kumar, A. A systematic review of hidden Markov models and their applications. *Arch. Comput. Methods Eng.* **2021**, *28*, 1429–1448. [[CrossRef](#)]
23. Bacha, R.E.; Zin, T.T. A Markov Chain Approach to Big Data Ranking Systems. In Proceedings of the 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), Nagoya, Japan, 24–27 October 2017; pp. 1–2.
24. Yang, G.Q.; Yang, W.G.; Wu, X.T. The strong laws of large numbers for countable non homogeneous hidden Markov models. *Commun. Stat. -Theory Methods* **2017**, *46*, 8808–8819. [[CrossRef](#)]
25. Zhang, L.; Cai, B.; Tan, T.; Shi, Y. Stabilization of non-homogeneous hidden semi-Markov jump systems with limited sojourn-time information. *Automatica* **2020**, *117*, 108963. [[CrossRef](#)]
26. Shahzadi, A.; Wang, T.; Bebbington, M.; Parry, M. Inhomogeneous hidden semi-Markov models for incompletely observed point processes. *Ann. Inst. Stat. Math.* **2023**, *75*, 253–280. [[CrossRef](#)]
27. Yang, W.G. Convergence in the Cesàro sense and strong law of large numbers for countable nonhomogeneous Markov chains. *Linear Algebra Appl.* **2002**, *354*, 275–286. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.