A Coding Theoretic Study of Homogeneous Markovian Predictive Games

Takara Nomura and Akio Fujiwara

***Abstract*—This paper explores a predictive game in which a Forecaster announces odds based on a time-homogeneous Markov kernel, establishing a game-theoretic law of large numbers for the relative frequencies of occurrences of all finite strings. A key feature of our proof is a betting strategy inspired by a universal coding scheme, drawing on the martingale convergence theorem and algorithmic randomness theory, without relying on a diversified betting approach that involves countably many operating accounts. We apply these insights to thermodynamics, offering a game-theoretic perspective on Leó Szilárd's thought experiment.**

***Index Terms*—Game-theoretic probability, martingale, universal coding, Szilárd's engine, entropy.**

I. Introduction

GAME-THEORETIC probability theory [1], proposed by Shafer and Vovk in 2001, offers a framework for studying stochastic behavior without relying on traditional concept of probability. To illustrate this approach, we begin by recalling a fundamental result from game-theoretic probability theory.

Let $\Omega := \{1, 2, \ldots, A\}$ be a finite alphabet, and let Ω^n, Ω^*, and Ω^∞ denote the sets of sequences over Ω of length n, finite length, and infinite (one-sided) length, respectively. The empty string is denoted by λ. An element of Ω^n is represented symbolically as ω^n. We also introduce the notation $\omega_i^j := \omega_i \omega_{i+1} \cdots \omega_j$ for $i \le j$, denoting the substring from the ith to the jth coordinates of a longer sequence $\omega_1 \omega_2 \cdots \omega_n \cdots$. By convention, if $i > j$, we set $\omega_i^j := \lambda$. For $x \in \Omega^*$ and $y \in \Omega^* \cup \Omega^\infty$, we write $x \sqsubset y$ to indicate that x is a prefix of y.

Let us introduce the following set:

$$\mathcal{P}(\Omega) := \left\{ p : \Omega \to (0,1) \;\middle|\; \sum_{\omega \in \Omega} p(\omega) = 1 \right\}.$$

Given a $p \in \mathcal{P}(\Omega)$, consider the following game.

Simple predictive game
Players: Skeptic and Reality.
Protocol: $K_0 = 1$.
FOR $n \in \mathbb{Z}_{>0}$:
$\beta_n \in \mathbb{R}^\Omega$.
Reality announces $\omega_n \in \Omega$.

$$K_n := K_{n-1} + \left\{ \beta_n(\omega_n) - \sum_{a \in \Omega} \beta_n(a) p(a) \right\}.$$

END FOR.

This protocol can be understood as a betting game in which Skeptic predicts Reality's "stochastic" move, regarding $p(a)$ as the "probability" of the occurrence of $a \in \Omega$. Further, β_n and K_n denote Skeptic's bet and capital at step n, respectively, with the recursion formula specifying how the capital evolves. Namely, the formula states that, at step n, after Skeptic announces $\beta_n = (\beta_n(1), \beta_n(2), \ldots, \beta_n(A))$ and Reality announces $\omega_n \in \Omega$, Skeptic gains $\beta_n(\omega_n)$ and loses $\sum_{a \in \Omega} \beta_n(a) p(a)$ in assets. This loss term $\sum_{a \in \Omega} \beta_n(a) p(a)$ reflects the "expected value" of the Skeptic's stakes under the odds p and can be regarded as a participation cost that must be paid in advance, regardless of the actual outcome. Note that β_n can take negative values. Since β_n can depend on Reality's past move ω_1^{n-1}, we identify Skeptic's strategy $\{\beta_n\}_n$ with a map $\beta : \Omega^* \to \mathbb{R}^\Omega$ as $\beta_n(a) := (\beta(\omega_1^{n-1}))(a)$.

Apparently, this game is in favor of Reality because Reality announces ω_n *after* knowing Skeptic's bet β_n, preventing Skeptic from becoming rich. However, Shafer and Vovk showed the following surprising result.

Theorem 1 (Game-Theoretic Law of Large Numbers): In the simple predictive game, Skeptic has a prudent strategy $\beta : \Omega^* \to \mathbb{R}^\Omega$ that ensures $\lim_{n\to\infty} K_n = \infty$ unless

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_a(\omega_i) = p(a)$$

for all $a \in \Omega$, where δ_a denotes the Kronecker delta. Here, a strategy is called prudent if $K_n > 0$ for all $n \in \mathbb{Z}_{>0}$ and every sequence $\omega_1^n \in \Omega^n$ chosen by Reality.

The theorem implies that there exists a betting strategy β_n that guarantees Skeptic becomes infinitely rich if Reality's moves deviate from the "law of large numbers," all while avoiding the risk of bankruptcy.[1] Note that Skeptic's capital may diverge even while the empirical frequencies of Reality's

Received 3 February 2025; revised 10 July 2025; accepted 30 October 2025. Date of publication 11 November 2025; date of current version 23 December 2025. This work was supported in part by the Japan Science and Technology Agency (JST) Exploratory Research for Advanced Technology (ERATO) under Grant JPMJER2402; and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 22340019, Grant 17H02861, Grant 23H01090, and Grant 23K25787. *(Corresponding author: Takara Nomura.)*

The authors are with the Department of Mathematics, The University of Osaka, Toyonaka, Osaka 560-0043, Japan (e-mail: u004883d@ecs.osaka-u.ac.jp; fujiwara@math.sci.osaka-u.ac.jp).

Communicated by V. Kostina, Associate Editor for Communications.
Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2025.3631705.
Digital Object Identifier 10.1109/TIT.2025.3631705

[1]If Reality's objective is to prevent Skeptic from becoming infinitely rich, then Theorem 1 can be rephrased as follows: *Skeptic can force the event* $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_a(\omega_i) = p(a)$, which corresponds to the original statement by Shafer and Vovk [1].

moves converge; the two phenomena are not mutually exclusive. For more information, see Appendix C.

After the publication of Shafer and Vovk's seminal book [1], game-theoretic probability theory has been widely studied across various contexts and applications. Notable developments include the game-theoretic proof of Lévy's zero-one law [2], [3], investigations into Reality's strategic choices [4], [5], and the game-theoretic formulation of Jeffrey's law [6], among others. These contributions are further extended in Shafer and Vovk's second book [7], which builds upon the foundations established in the first, deepening the analysis and exploring new frontiers in game-theoretic probability. In recent years, many studies inspired by game-theoretic probability have explored new applications, including statistical inference [8], statistical testing [9] and machine learning [10]. These works have significantly advanced the field, addressing both theoretical and practical challenges within the framework of game-theoretic probability. However, despite these developments, the connection between game-theoretic probability theory and information theory, particularly in the context of universal coding, has received limited attention; see, for example, [11] and [12] for related work primarily addressing sequential testing problems.

This paper aims to elucidate the coding theoretic aspects underpinning the theory by presenting an alternative proof of Theorem 1 and its generalizations. Our approach offers new insights into game-theoretic probability by employing a single, coding theoretic (pure) strategy, in contrast to Shafer and Vovk's original proof, which relies on a diversified ("mixed") betting strategy using countably many operating accounts [1], and to subsequent alternative proofs grounded in statistical estimation [13], [14]. We also note that Feder [15] previously proposed a closely related strategy. Although the aims and scope differ, we cite this work for completeness and to acknowledge the prior contribution.

Generalized predictive game
Players: Forecaster, Skeptic, and Reality.
Protocol: $K_0 = 1$.
FOR $n \in \mathbb{Z}_{>0}$:
Forecaster announces $p_n \in \mathcal{P}(\Omega)$.
Skeptic announces $\beta_n \in \mathbb{R}^\Omega$.
Reality announces $\omega_n \in \Omega$.
$$K_n := K_{n-1} + \left\{\beta_n(\omega_n) - \sum_{a\in\Omega}\beta_n(a)p_n(a)\right\}.$$
END FOR.

In order to explicate our coding theoretic approach, consider the following generalized game in which a Forecaster comes into play to announce a possibly "non-i.i.d." process.

Let us identify Skeptic's betting strategy $\beta_n \in \mathbb{R}^\Omega$ with $\alpha_n \in \mathbb{R}^\Omega$ that satisfies

$$\beta_n(a) = K_{n-1}\cdot\alpha_n(a), \qquad (a\in\Omega).$$

Then, the recursion formula for the capital is rewritten as

$$K_n = K_{n-1}\left\{1 + \alpha_n(\omega_n) - \sum_{a\in\Omega}\alpha_n(a)p_n(a)\right\}. \tag{1}$$

We shall call α_n a betting strategy as well, and call it prudent if the corresponding β_n is prudent. We also identify $\{\alpha_n\}_n$ with a map $\alpha : \Omega^* \to \mathbb{R}^\Omega$ as $\alpha_n(a) = (\alpha(\omega_1^{n-1}))(a)$.

Now, associated with a prudent strategy α_n is the following quantity:

$$Q(\omega \mid \omega_1^{n-1}) := \left\{1 + \alpha_n(\omega) - \sum_{a\in\Omega}\alpha_n(a)p_n(a)\right\} p_n(\omega). \tag{2}$$

Since α_n is prudent, we see that $Q(\omega \mid \omega_1^{n-1}) > 0$ for all $\omega \in \Omega$. Moreover,

$$\begin{aligned}\sum_{\omega\in\Omega} & Q(\omega \mid \omega_1^{n-1}) \\ &= \sum_{\omega\in\Omega}\left\{1 + \alpha_n(\omega) - \sum_{a\in\Omega}\alpha_n(a)p_n(a)\right\} p_n(\omega) \\ &= 1 + \sum_{\omega\in\Omega}\alpha_n(\omega)p_n(\omega) - \sum_{a\in\Omega}\alpha_n(a)p_n(a) \\ &= 1.\end{aligned}$$

Therefore, the quantity $Q(\omega \mid \omega_1^{n-1})$ defined by (2) can be regarded as a conditional probability. Conversely, for any conditional probability $Q(\omega \mid \omega_1^{n-1})$, there exists a prudent strategy α_n (although not unique) that satisfies (2): for instance, let $\alpha_n(a) := Q(a \mid \omega_1^{n-1})/p_n(a)$ for each $a \in \Omega$. Thus, the role of Skeptic in the above predictive game is regarded as announcing a conditional probability $Q(\omega \mid \omega_1^{n-1})$.

Now, suppose that Forecaster happens to have a predetermined probability measure P on $(\Omega^\infty, \mathscr{F})$, with $\mathscr{F} := \sigma(\{\Gamma_x\}_{x\in\Omega^*})$ being the σ-algebra generated by the cylinder sets $\Gamma_x := \{y \in \Omega^\infty : x \sqsubset y\}$, and announces each function p_n as the conditional probability, given the past data ω_1^{n-1}, as follows:

$$p_n(a) := P(a \mid \omega_1^{n-1}) := P(a \mid \Gamma_{\omega_1^{n-1}}), \quad (a\in\Omega).$$

Then, we have from (1) and (2) that

$$\begin{aligned}\frac{K_n}{K_{n-1}} &= \left\{1 + \alpha_n(\omega) - \sum_{a\in\Omega}\alpha_n(a)p_n(a)\right\} \\ &= \frac{Q(\omega_n \mid \omega_1^{n-1})}{P(\omega_n \mid \omega_1^{n-1})},\end{aligned}$$

and hence

$$K_n = K_0\prod_{i=1}^{n}\frac{K_i}{K_{i-1}} = \frac{Q(\omega_1^n)}{P(\omega_1^n)}. \tag{3}$$

Put differently, the capital process K_n is nothing but the likelihood ratio process between P and Q. This perspective is consistent with the interpretation developed in [7, Section 10.5] and further elaborated in [16, Section 2.2 and 2.3].

Let $\mathscr{F}_n := \sigma(\{\Gamma_{x^n}\}_{x^n\in\Omega^n})$ for each $n \in \mathbb{Z}_{>0}$, Then the capital process (3) is a P-martingale relative to the natural filtration $\{\mathscr{F}_n\}_n$, in that

$$\begin{aligned}\mathbb{E}\left[\frac{Q(\omega_1^n)}{P(\omega_1^n)}\,\middle|\,\mathscr{F}_{n-1}\right] &= \sum_{\omega_n\in\Omega}\frac{Q(\omega_1^n)}{P(\omega_1^n)}P(\omega_n \mid \omega_1^{n-1}) \\ &= \sum_{\omega_n\in\Omega}\frac{Q(\omega_1^n)}{P(\omega_1^{n-1})} = \frac{Q(\omega_1^{n-1})}{P(\omega_1^{n-1})}.\end{aligned}$$

It then follows from the martingale convergence theorem [17] that the capital process K_n converges almost surely to some nonnegative value under the probability measure P. Game-theoretic probability provides a reciprocal description of this mechanism by asserting that K_n can diverge "on a P-null set," which, in the context of predictive games, is interpreted as "if Reality does not align with Forecaster P."

Furthermore, due to (3), the logarithm of the capital process is written as

$$\log_A K_n = \left\{-\log_A P(\omega_1^n)\right\} - \left\{-\log_A Q(\omega_1^n)\right\}. \tag{4}$$

This expression is simply the difference between the Shannon codelengths for P and Q, evoking the notion of *randomness deficiency* in algorithmic randomness theory [18]. For a computable probability measure P on Ω^∞, an infinite sequence $\omega_1^\infty \in \Omega^\infty$ is Martin-Löf P-random if and only if the sequence

$$-\log_A P(\omega_1^n) - \mathcal{K}(\omega_1^n) \tag{5}$$

of randomness deficiencies is bounded from above, where $\mathcal{K}(\omega_1^n)$ is the prefix Kolmogorov complexity of $\omega_1^n \in \Omega^*$. Obviously, (4) and (5) are similar in form, as both represent differences in codelengths. However, it is crucial to note that (5) contains an uncomputable quantity $\mathcal{K}(\omega_1^n)$. This observation motivates the design of the conditional probability Q in (2), or equivalently, a betting strategy α_n in the generalized predictive game, by employing ideas and techniques from computable universal coding schemes.

Time-homogeneous k th-orderMarkovian predictive game
Players: Forecaster, Skeptic, and Reality.
Protocol: $K_0 = 1$.
FOR $n \in \mathbb{Z}_{>0}$:
Forecaster announces $p_n \in \mathcal{P}(\Omega)$ such that $p_n(a) := M(a \mid \omega_{n-k}^{n-1})$ for $n > k$.
Skeptic announces $\alpha_n \in \mathbb{R}^\Omega$.
Reality announces $\omega_n \in \Omega$.

$$K_n := K_{n-1}\left\{1 + \alpha_n(\omega_n) - \sum_{a\in\Omega}\alpha_n(a)p_n(a)\right\}.$$

END FOR.

Nevertheless, due to the difficulty of addressing fully general predictive games, we focus here on the time-homogeneous Markovian predictive game[2] as a first step toward developing methods applicable to the general stationary ergodic case. Suppose Forecaster has a kth-order Markov kernel $M : \Omega \times \Omega^k \to (0, 1) : (a, \omega^k) \mapsto M(a \mid \omega^k)$ that satisfies

$$\sum_{a\in\Omega} M(a \mid \omega^k) = 1$$

for all $\omega^k \in \Omega^k$.

This protocol can be understood as a variant of the generalized predictive game in which Forecaster announces $p_n : \Omega \to (0, 1)$ for $n > k$ according to a time-homogeneous Markov kernel M as $p_n(a) = M(a \mid \omega_{n-k}^{n-1})$ based on Reality's past moves. Note that for $n \le k$, $p_n(a)$ can be arbitrary as long as $p_n \in \mathcal{P}(\Omega)$.

For each $\omega^\ell \in \Omega^\ell$ with $\ell \in \mathbb{Z}_{\ge k}$, let $P(\omega^\ell)$ be defined by

$$P(\omega^\ell) := \pi(\omega^k)\prod_{i=k+1}^{\ell} M(\omega_i \mid \omega_{i-k}^{i-1}),$$

where $\pi : \Omega^k \to (0, 1)$ is the stationary distribution associated with the Markov kernel M. The main result of this paper is the following:

Theorem 2: In the time-homogeneous kth-order Markovian predictive game, Skeptic has a prudent strategy $\alpha : \Omega^* \to \mathbb{R}^\Omega$ that ensures $\lim_{n\to\infty} K_n = \infty$ unless

$$\lim_{n\to\infty}\frac{S_n(a^\ell)}{n} = P(a^\ell)$$

for all $\ell \in \mathbb{Z}_{\ge k}$ and $a^\ell \in \Omega^\ell$, where $S_n(a^\ell)$ denotes the number of occurrences of a^ℓ in ω^n.

Theorem 2 establishes that there exists a betting strategy that guarantees Skeptic can become infinitely rich if Reality's moves do not align with the Markovian Forecaster's announcements.[3] Specifically, this happens when the relative frequency of occurrences of some string of length ℓ ($\ge k$) fails to converge to the stationary joint distribution associated with the kth-order Markov kernel.

This paper is organized as follows. Section II introduces a betting strategy based on the incremental parsing scheme of Ziv and Lempel [20], and presents several lemmas that lay the groundwork for proving the main result. For improved readability, the proofs of these lemmas are deferred to Appendix A. In Section III, we prove Theorem 2 by incorporating properties of Lempel-Ziv incremental parsing established in the previous section, thereby bringing ideas inspired by universal coding into game-theoretic probability. Section IV explores applications of Theorem 2 to thermodynamics, specifically a game-theoretic interpretation of Szilárd's engine and a discussion of entropy in predictive games. Finally, Section V provides concluding remarks. For the reader's convenience, additional information on stationary distributions of Markov chains and an alternative proof of Theorem 1 are provided in Appendices B and C, respectively.

II. Preliminaries

In this section, we develop a betting strategy using the technique of incremental parsing and establish several lemmas in preparation for the proof of Theorem 2.

A. Betting Strategy Inspired by Lempel-Ziv Coding Scheme

We outline an algorithm for incremental parsing [20], which divides a string into substrings separated by slashes, with each substring being the shortest one not previously encountered. The algorithm runs as follows: Start with an initial slash. After each slash, scan the input sequence until the shortest string that has not yet been marked off is identified. Since

[2]The Markovian predictive game introduced in this paper is entirely distinct from the Markov game commonly used in the field of operations research [19].

[3]Using Shafer and Vovk's terminology [1] once again, Theorem 2 can be restated as follows: *Skeptic can force the event* $\lim_{n\to\infty} S_n(a^\ell)/n = P(a^\ell)$ *for all* $\ell \in \mathbb{Z}_{\ge k}$ *and* $a^\ell \in \Omega^\ell$.

this string is the shortest unseen string, all its prefixes must have appeared earlier in the sequence. For example, a sequence 1000011101011 of length 13 is decomposed into

$$/1/0/00/01/11/010/11.$$

Suppose a sequence ω_1^n is parsed as

$$\omega_1^n = /\omega_{n_0+1}^{n_1}/\omega_{n_1+1}^{n_2}/\cdots/\omega_{n_{T-1}+1}^{n_T}/\omega_{n_T+1}^{n},$$

where $n_0 = 0$, $n_1 = 1$, and all parsed substrings except the last one, $\omega_{n_T+1}^n$, are distinct. In what follows, the substrings $\omega_{n_0+1}^{n_1}, \omega_{n_1+1}^{n_2}, \ldots, \omega_{n_{T-1}+1}^{n_T}$ are referred to as parsed phrases. Note that the number T of parsed phrases depends on the sequence ω_1^n, and the last string $\omega_{n_T+1}^n$ may be empty.

We now construct a betting strategy at step n, given Reality's past moves ω_1^{n-1} ($n \geq 2$). Using incremental parsing, we decompose ω_1^{n-1} into

$$\omega_1^{n-1} = /\omega_{n_0+1}^{n_1}/\omega_{n_1+1}^{n_2}/\cdots/\omega_{n_{T-1}+1}^{n_T}/\omega_{n_T+1}^{n-1}.$$

Next, we define the set

$$V\left(\omega_1^{n_T}\right) := \left\{ \xi \in \Omega^* \;\middle|\; \begin{array}{l} \xi \neq \omega_{n_{j-1}+1}^{n_j} \text{ for all } j \in \{0,1,\ldots,T\}, \text{ and} \\ \xi = \omega_{n_{j-1}+1}^{n_j} b \text{ for some } j \in \{0,1,\ldots,T\} \\ \text{and } b \in \Omega \end{array} \right\},$$

where $\omega_{n_{-1}+1}^{n_0} = \lambda$ is the empty string. Equivalently, $V(\omega_1^{n_T})$ consists of all one-symbol extensions of already parsed phrases that have not yet appeared as phrases; it is the "frontier" of the current dictionary, whose elements are candidates to become new phrases at some later step (not necessarily the next). The size of this set is given by

$$|V(\omega_1^{n_T})| = A + T(A-1).$$

This can be shown by induction on T: For $T = 0$, we have $n = 1$, and

$$V(\omega_1^{n-1}) = V(\lambda) = \Omega.$$

For $T \geq 1$, the set $V(\omega_1^{n_T})$ is constructed as

$$V(\omega_1^{n_T}) = \left(V(\omega_1^{n_{T-1}}) \setminus \{\omega_{n_{T-1}+1}^{n_T}\}\right) \cup \{\omega_{n_{T-1}+1}^{n_T} b \mid b \in \Omega\},$$

which yields a recursive formula $|V(\omega_1^{n_T})| = |V(\omega_1^{n_{T-1}})| - 1 + A$, ensuring the desired result.

Finally, we define the conditional probability $Q_{LZ}(\omega_n \mid \omega_1^{n-1})$, which determines the betting strategy $\alpha_n(\omega_n)$, as follows. For $n = 1$, let $Q_{LZ}(a) := 1/A$. For $n \geq 2$, we define

$$Q_{LZ}(a \mid \omega_1^{n-1}) := \frac{|\{\xi \in V(\omega_1^{n_T}) \mid \omega_{n_T+1}^{n-1} a \sqsubset \xi\}|}{|\{\xi \in V(\omega_1^{n_T}) \mid \omega_{n_T+1}^{n-1} \sqsubset \xi\}|}. \tag{6}$$

Note that

$$\{\xi \in V(\omega_1^{n_T}) \mid \omega_{n_T+1}^{n-1} \sqsubset \xi\} = \bigsqcup_{a \in \Omega} \{\xi \in V(\omega_1^{n_T}) \mid \omega_{n_T+1}^{n-1} a \sqsubset \xi\},$$

which follows from the definition of $V(\omega_1^{n_T})$. Moreover, the set

$$\{\xi \in V(\omega_1^{n_T}) \mid \omega_{n_T+1}^{n-1} a \sqsubset \xi\}$$

is nonempty for any $a \in \Omega$. Thus, we conclude that

$$\sum_{a \in \Omega} Q_{LZ}(a \mid \omega_1^{n-1}) = 1 \quad \text{and} \quad Q_{LZ}(a \mid \omega_1^{n-1}) > 0.$$

The motivation behind the definition (6) is now in order. When $n - 1 = n_T$ (i.e., when $\omega_{n_T+1}^{n-1} = \lambda$), we have

$$Q_{LZ}(a \mid \omega_1^{n_T}) = \frac{|\{\xi \in V(\omega_1^{n_T}) \mid a \sqsubset \xi\}|}{|V(\omega_1^{n_T})|}.$$

Thus, for each $\xi = \xi_1^\tau \in V(\omega_1^{n_T})$,

$$Q_{LZ}(\xi \mid \omega_1^{n_T}) := Q_{LZ}(\xi_1 \mid \omega_1^{n_T}) \prod_{t=2}^{\tau} Q_{LZ}(\xi_t \mid \omega_1^{n_T} \xi_1^{t-1}) = \frac{1}{|V(\omega_1^{n_T})|}.$$

In other words, the conditional probability $Q_{LZ}(a \mid \omega_1^{n-1})$ is designed to induce the uniform distribution over $V(\omega_1^{n_T})$. This observation also implies that

$$Q_{LZ}(\omega_{n_T+1}^n \mid \omega_1^{n_T}) > \frac{1}{|V(\omega_1^{n_T})|} \tag{7}$$

whenever $n > n_T$.

In what follows, we refer to the betting strategy α_n based on the conditional probability $Q_{LZ}(a \mid \omega_1^{n-1})$ as the *Lempel-Ziv betting strategy*. Since α_n is canonically constructed via $\alpha_n(a) = Q_{LZ}(a \mid \omega_1^{n-1})/p_n(a)$, we may identify α_n with $Q_{LZ}(\cdot \mid \omega_1^{n-1})$ and, for simplicity, also refer to $Q_{LZ}(\cdot \mid \omega_1^{n-1})$ itself as the Lempel-Ziv betting strategy.

Example: Let $\Omega = \{0,1\}$, and suppose that Skeptic has observed the sequence of Reality's outcomes $\omega_1^{13} = /1/0/00/01/11/010/11$. At this stage, the current set of candidate phrases is

$$V(\omega_1^{11}) = \{10, 000, 001, 011, 110, 111, 0100, 0101\},$$

and the conditional probabilities $Q_{LZ}(a \mid \omega_1^{13})$ at step $n = 14$ are given by

$$Q_{LZ}(0 \mid \omega_1^{13}) = \frac{|\{110\}|}{|\{110, 111\}|} = \frac{1}{2},$$
$$Q_{LZ}(1 \mid \omega_1^{13}) = \frac{|\{111\}|}{|\{110, 111\}|} = \frac{1}{2}.$$

Upon hearing Forecaster's announcement $p_{14}(a)$, Skeptic announces the bet

$$\alpha_{14}(a) = \frac{Q_{LZ}(a \mid \omega_1^{13})}{p_{14}(a)} = \frac{1}{2\, p_{14}(a)}, \quad (a \in \Omega).$$

Suppose further that Reality's announcement at step $n = 14$ is $\omega_{14} = 0$. Skeptic then updates the set of candidate phrases accordingly:

$$V(\omega_1^{14}) = \{10, 000, 001, 011, 111, 0100, 0101, 1100, 1101\},$$

and the updated strategy becomes

$$Q_{LZ}(0 \mid \omega_1^{14}) = \frac{|\{000, 001, 011, 0100, 0101\}|}{|V(\omega_1^{14})|} = \frac{5}{9},$$
$$Q_{LZ}(1 \mid \omega_1^{14}) = \frac{|\{10, 111, 1100, 1101\}|}{|V(\omega_1^{14})|} = \frac{4}{9}.$$

Remark: From the above discussion,

$$-\log Q_{LZ}(\omega_1^{n_T}) = \sum_{j=0}^{T-1} \log(A + j(A-1)), \tag{8}$$

where

$$Q_{LZ}(\omega_1^n) := \prod_{i=1}^{n} Q_{LZ}(\omega_i \mid \omega_1^{i-1}).$$

On the other hand, the Lempel-Ziv codelength ℓ_{LZ} [20] is given by

$$\ell_{LZ}(\omega_1^n) = \sum_{j=1}^{T+1} \lceil \log_A(jA) \rceil.$$

B. Properties of Lempel-Ziv Betting Strategy

In this section, we outline several fundamental properties of the Lempel-Ziv betting strategy. All proofs are deferred to Appendix A.

Define the complexity $c(\omega_1^n)$ of a sequence $\omega_1^n \in \Omega^n$ as the total number T of parsed phrases obtained through its incremental parsing [21, p. 448]. The quantity $c(\omega_1^n)\log c(\omega_1^n)$ is known to play an essential role in Lempel-Ziv coding, and it is also important in this paper as a bridge between several key quantities. The following lemma provides an asymptotic lower bound for the difference between $c(\omega_1^n)\log c(\omega_1^n)$ and $-\log Q_{LZ}(\omega_1^n)$.

Lemma 3: For any $\omega_1^\infty \in \Omega^\infty$, the following inequality holds:

$$\liminf_{n\to\infty} \frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\} \geq 0.$$

For $n, \ell \in \mathbb{Z}_{>0}$ with $n > \ell$, let $T_n(a_1^\ell)$ denote the number of occurrences of $a_1^\ell \in \Omega^\ell$ in the cyclically extended word $\omega_1^n\omega_1^{\ell-1}$ of length $n+\ell-1$. Similarly, let $T_n(b \mid a_1^\ell)$ represent the number of occurrences of b immediately following $a_1^\ell \in \Omega^\ell$ in the extended word $\omega_1^n\omega_1^\ell$ of length $n+\ell$. In other words, $T_n(b \mid a_1^\ell) = T_n(a_1^\ell b)$.

Note that, by definition,

$$\sum_{a_1^\ell \in \Omega^\ell} T_n(a_1^\ell) = n.$$

Moreover, the quantity $T_n(a_1^\ell)$ is asymptotically equivalent to $S_n(a_1^\ell)$, the number of occurrences of a_1^ℓ in ω_1^n, in the sense that

$$\lim_{n\to\infty} \frac{1}{n}\left(T_n(a_1^\ell) - S_n(a_1^\ell)\right) = 0.$$

The following lemma states that the marginals of conditional counts $T_n(b \mid a_1^\ell)$ correspond to $T_n(a_1^\ell)$ and $T_n(a_2^\ell b)$.

Lemma 4: Let $n, \ell \in \mathbb{Z}_{>0}$ with $n > \ell$. Then, for all $a_1^\ell \in \Omega^\ell$ and $b \in \Omega$, the following identities hold:

$$\sum_{b\in\Omega} T_n(b \mid a_1^\ell) = T_n(a_1^\ell) \quad \text{and} \quad \sum_{a_1\in\Omega} T_n(b \mid a_1^\ell) = T_n(a_2^\ell b).$$

Let us now fix $\ell \in \mathbb{Z}_{>0}$ arbitrarily. Given ω_1^n with $n > \ell$, we define

$$\hat{M}_n^\ell(b \mid a^\ell) := \frac{T_n(b \mid a^\ell)}{T_n(a^\ell)}$$

for $a^\ell \in \Omega^\ell$ satisfying $T_n(a^\ell) > 0$. Due to Lemma 4, we observe that

$$\sum_{b\in\Omega} \hat{M}_n^\ell(b \mid a^\ell) = \sum_{b\in\Omega} \frac{T_n(b \mid a^\ell)}{T_n(a^\ell)} = 1.$$

Thus, with an appropriate definition of $\hat{M}_n^\ell(b \mid a^\ell)$ for $a^\ell \in \Omega^\ell$ satisfying $T_n(a^\ell) = 0$, $\hat{M}_n^\ell$ can be regarded as an ℓth-order Markov kernel formally associated with ω_1^n. The following lemma provides a condition that ensures $T_n(a^k)/n$ converges to the stationary distribution π of the kth-order Markov kernel M.

Lemma 5: If

$$\lim_{n\to\infty} \hat{M}_n^k(b \mid a_1^k) = M(b \mid a_1^k)$$

for all $a_1^k \in \Omega^k$ and $b \in \Omega$, then,

$$\lim_{n\to\infty} \frac{T_n(a_1^k)}{n} = \pi(a_1^k)$$

for all $a_1^k \in \Omega^k$.

Next, for $n, \ell \in \mathbb{Z}_{>0}$ with $n > \ell$, we define

$$\hat{R}_n^\ell(\omega_1^n) := \left\{\prod_{i=1}^{\ell} \hat{M}_n^\ell(\omega_i \mid \omega_{n-\ell+i}^n \omega_1^{i-1})\right\} \times \left\{\prod_{i=\ell+1}^{n} \hat{M}_n^\ell(\omega_i \mid \omega_{i-\ell}^{i-1})\right\}.$$

In the first factor, which represents $\hat{R}_n^\ell(\omega_1^\ell)$, we formally introduce concatenated strings $\omega_{n-\ell+i}^n\omega_1^{i-1}$ of length ℓ for $i = 1, \ldots, \ell$, to facilitate the application of the ℓth Markov kernel $\hat{M}_n^\ell$. The following lemma presents an important inequality relating $c(\omega_1^n)\log c(\omega_1^n)$ to $-\log \hat{R}_n^\ell(\omega_1^n)$. While its proof resembles Ziv's inequality and the asymptotic optimality of the Lempel-Ziv algorithm [21, Section 13.5.2], our setting differs in that $\hat{M}_n^\ell$ depends solely on Reality's moves and does not assume any underlying stochastic process.

Lemma 6: For any $n, \ell \in \mathbb{Z}_{>0}$ with $n > \ell$ and $\omega_1^n \in \Omega^n$, the following inequality holds:

$$\frac{1}{n} c(\omega_1^n)\log c(\omega_1^n) \leq -\frac{1}{n}\log \hat{R}_n^\ell(\omega_1^n) + \delta_\ell(n),$$

where $\delta_\ell(n) \to 0$ as $n \to \infty$.

Finally, for $n \in \mathbb{Z}_{\geq k}$, let us introduce

$$\tilde{P}(\omega_1^n) := \left\{\prod_{i=1}^{k} p_i(\omega_i)\right\} \cdot \left\{\prod_{i=k+1}^{n} M(\omega_i \mid \omega_{i-k}^{i-1})\right\},$$

which represents Forecaster's announcements. The next lemma establishes the relationship between $\tilde{P}$ and $\hat{R}_n^\ell$.

Lemma 7: For any $n, \ell \in \mathbb{Z}_{\geq k}$ with $n > \ell$, the following identity holds:

$$\begin{aligned} &-\log \tilde{P}(\omega_1^n) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^n \omega_1^{i-1}) \\ &\quad + \sum_{a_1^\ell \in \Omega^\ell} T_n(a_1^\ell) \cdot D\left(\hat{M}_n^\ell(\,\cdot \mid a_1^\ell) \,\middle\|\, M(\,\cdot \mid a_{\ell-k+1}^\ell)\right), \end{aligned}$$

where $D(\,\cdot \,\|\, \cdot\,)$ is the Kullback-Leibler divergence.

III. Proof of Theorem 2

Let $K^\alpha : \Omega^* \to \mathbb{R}$ represent the capital process associated with a betting strategy α, so that $K_n = K^\alpha(\omega_1^n)$ for $\omega_1^n \in \Omega^*$. The following lemma demonstrates that the limit supremum for the capital process can be replaced by a limit.

Lemma 8: Suppose a strategy α satisfies

$$\limsup_{n\to\infty} K^\alpha(\omega_1^n) = \infty$$

for a specific sequence of Reality's moves $\{\omega_1^n\}_{n\in\mathbb{Z}_{>0}}$. Then there exists another strategy α_* such that

$$\lim_{n\to\infty} K^{\alpha_*}(\omega_1^n) = \infty$$

for the same sequence $\{\omega_1^n\}_{n\in\mathbb{Z}_{>0}}$.

Proof: The required strategy α_* can be defined as follows: Strategy α_* uses α as long as the capital K_n retains below 2. Once K_n reaches or exceeds 2, α_* transfers the net gain $\Delta K := K_n - 1\,(\geq 1)$ into an external storage. It then restart the game with a capital of 1, employing the strategy α again. ■

Lemma 9: In the time-homogeneous kth-order Markovian predictive game, the Lempel-Ziv betting strategy Q_{LZ} ensures $\lim_{n\to\infty} K_n = \infty$ unless

$$\lim_{n\to\infty} \hat{M}_n^k(b \mid a_1^k) = M(b \mid a_1^k)$$

for all $a_1^k \in \Omega^k$ and $b \in \Omega$.

Proof: Using (3) and applying Lemma 6, we obtain

$$\begin{aligned}\frac{\log K_n}{n} &= \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\}\\ &= \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^n) - c(\omega_1^n)\log c(\omega_1^n)\right\}\\ &\quad + \frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\}\\ &\geq \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^n) + \log \hat{R}_n^k(\omega_1^n) - n\delta_k(n)\right\}\\ &\quad + \frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\}.\end{aligned}$$

Applying Lemma 7 with $\ell = k$, we further evaluate K_n as

$$\begin{aligned}&\frac{\log K_n}{n}\\ &\geq \sum_{a_1^k\in\Omega^k} \frac{T_n(a_1^k)}{n}\cdot D\left(\hat{M}_n^k(\,\cdot\mid a_1^k)\,\middle\|\, M(\,\cdot\mid a_1^k)\right)\\ &+ \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^k) + \log\prod_{i=1}^k M(\omega_i \mid \omega_{n-k+i}^n\omega_1^{i-1})\right\}\\ &- \delta_k(n) + \frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\}.\end{aligned}$$

From the definition of $\tilde{P}$,

$$\lim_{n\to\infty}\frac{1}{n}\left\{-\log \tilde{P}(\omega_1^k) + \log\prod_{i=1}^k M(\omega_i \mid \omega_{n-k+i}^n\omega_1^{i-1})\right\} = 0.$$

Furthermore, by Lemma 6, we know that $\delta_k(n) \to 0$ as $n \to \infty$, and by Lemma 3,

$$\liminf_{n\to\infty}\frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\} \geq 0.$$

In light of Lemma 8, therefore, it now suffices to prove the following claim: For a given $\omega_1^\infty \in \Omega^\infty$, if there exists $b_1^{k+1} \in \Omega^{k+1}$ such that $\hat{M}_n^k(b_{k+1} \mid b_1^k)$ does not converge to $M(b_{k+1} \mid b_1^k)$ as $n \to \infty$, then

$$\limsup_{n\to\infty}\sum_{a_1^k\in\Omega^k}\frac{T_n(a_1^k)}{n}\cdot D\left(\hat{M}_n^k(\,\cdot\mid a_1^k)\,\middle\|\, M(\,\cdot\mid a_1^k)\right) > 0.$$

We prove this claim by contradiction. Assume that for a given $\omega_1^\infty \in \Omega^\infty$, there exists $b_1^{k+1} \in \Omega^{k+1}$ such that $\hat{M}_n^k(b_{k+1} \mid b_1^k)$ does not converge to $M(b_{k+1} \mid b_1^k)$ as $n \to \infty$, and yet

$$\lim_{n\to\infty}\sum_{a_1^k\in\Omega^k}\frac{T_n(a_1^k)}{n}\cdot D\left(\hat{M}_n^k(\,\cdot\mid a_1^k)\,\middle\|\, M(\,\cdot\mid a_1^k)\right) = 0. \tag{9}$$

From this assumption, there exists a subsequence $(n_i)_i \subset (n)$ such that

$$\lim_{i\to\infty} D\left(\hat{M}_{n_i}^k(\,\cdot\mid b_1^k)\,\middle\|\, M(\,\cdot\mid b_1^k)\right) > 0.$$

Combining this with (9), we obtain

$$\lim_{i\to\infty}\frac{T_{n_i}(b_1^k)}{n_i} = 0. \tag{10}$$

Further, using (9) and (10), we can deduce that for all $d_k \in \Omega$,

$$\lim_{i\to\infty}\frac{T_{n_i}(d_k b_1^{k-1})}{n_i} = 0. \tag{11}$$

To establish (11), we first recall the second identity in Lemma 4, which yields

$$T_n(da_2^\ell b) \leq \sum_{a_1\in\Omega} T_n(a_1 a_2^\ell b) = T_n(a_2^\ell b)$$

for any $d \in \Omega$ and $a_2^\ell b \in \Omega^\ell$. Thus, for any $d_k \in \Omega$,

$$0 \leq \frac{T_{n_i}(d_k b_1^k)}{n_i} \leq \frac{T_{n_i}(b_1^k)}{n_i}.$$

From (10), it follows that

$$\lim_{i\to\infty}\frac{T_{n_i}(d_k b_1^k)}{n_i} = 0. \tag{12}$$

Let us prove (11) by contradiction. Suppose there exists $d_k \in \Omega$ such that

$$\limsup_{i\to\infty}\frac{T_{n_i}(d_k b_1^{k-1})}{n_i} > 0.$$

Then, there exists a subsequence $(n_{i_j})_j \subset (n_i)_i$ such that

$$\lim_{j\to\infty}\frac{T_{n_{i_j}}(d_k b_1^{k-1})}{n_{i_j}} > 0. \tag{13}$$

Consequently, for sufficiently large j, we have

$$\begin{aligned}\frac{T_{n_{i_j}}(d_k b_1^k)}{n_{i_j}} &= \frac{T_{n_{i_j}}(b_k \mid d_k b_1^{k-1})}{n_{i_j}}\\ &= \frac{T_{n_{i_j}}(d_k b_1^{k-1})}{n_{i_j}}\frac{T_{n_{i_j}}(b_k \mid d_k b_1^{k-1})}{T_{n_{i_j}}(d_k b_1^{k-1})}\\ &= \frac{T_{n_{i_j}}(d_k b_1^{k-1})}{n_{i_j}}\hat{M}_{n_{i_j}}^k(b_k \mid d_k b_1^{k-1}),\end{aligned}$$

and thus, combining (12) and (13), we find $\lim_{j\to\infty}\hat{M}_{n_{i_j}}^k(b_k \mid d_k b_1^{k-1}) = 0$. Recalling $M(b_k \mid d_k b_1^{k-1}) > 0$, we find that

$$\lim_{j\to\infty}\hat{M}_{n_{i_j}}^k(b_k \mid d_k b_1^{k-1}) \neq M(b_k \mid d_k b_1^{k-1}). \tag{14}$$

Since (13) and (14) contradict (9), we have (11).

By repeatedly applying this reasoning, adding d_{k-i+1} to the front and removing b_{k-i+1} from the rear of the word b_1^k in (10) for $i = 1, \ldots, k$, we conclude that

$$\lim_{i\to\infty} \frac{T_{n_i}(d_1^k)}{n_i} = 0$$

for all $d_1^k \in \Omega^k$. This contradicts the fact that

$$\sum_{d_1^k \in \Omega} T_n(d_1^k) = n$$

for all $n\ (\geq k)$. Hence, the claim is proved. ■

Now we are ready to prove Theorem 2.

Proof of Theorem 2: We prove the assertion by induction in $\ell\ (\geq k)$. For $\ell = k$, the assertion holds by Lemmas 5 and 9.

Assume that the assertion holds for some $\ell\ (\geq k)$, namely,

$$\lim_{n\to\infty} \frac{T_n(a_1^\ell)}{n} = P(a_1^\ell) = \pi(a_1^k) \prod_{i=k+1}^{\ell} M(a_i \mid a_{i-k}^{i-1})\ (> 0)$$

for all $a_1^\ell \in \Omega^\ell$. Since

$$\frac{T_n(a_1^{\ell+1})}{n} = \frac{T_n(a_1^\ell)}{n} \hat{M}_n^\ell(a_{\ell+1} \mid a_1^\ell),$$

we see that

$$\lim_{n\to\infty} \frac{T_n(a_1^{\ell+1})}{n} = P(a_1^{\ell+1})$$

holds if and only if

$$\lim_{n\to\infty} \hat{M}_n^\ell(a_{\ell+1} \mid a_1^\ell) = M(a_{\ell+1} \mid a_{\ell-k+1}^\ell).$$

By a similar evaluation in the proof of Lemma 9, we see that

$$\begin{aligned}
&\frac{\log K_n}{n} \\
&\geq \sum_{a_1^\ell \in \Omega^\ell} \frac{T_n(a_1^\ell)}{n} \cdot D\left(\hat{M}_n^\ell(\,\cdot \mid a_1^\ell) \,\|\, M(\,\cdot \mid a_{\ell-k+1}^\ell)\right) \\
&+ \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^n \omega_1^{i-1})\right\} \\
&- \delta_\ell(n) + \frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\}.
\end{aligned}$$

By the assumption of induction, $T_n(a_1^\ell)/n$ converges to a positive number $P(a_1^\ell)$ for all $a_1^\ell \in \Omega^\ell$ as $n \to \infty$. Therefore, if $\hat{M}_n^\ell(a_{\ell+1} \mid a_1^\ell)$ does not converge to $M(a_{\ell+1} \mid a_{\ell-k+1}^\ell)$ as $n \to \infty$, we have $\limsup_{n\to\infty} K_n = \infty$. This completes the proof. ■

IV. Applications

In the previous section, we demonstrated that adopting a coding-theoretic idea leads to a new way of understanding game-theoretic probability. This methodological innovation motivates further investigation into its broader implications. Here we present two applications of our main result. The first concerns Szilárd's engine, which bridges thermodynamics and information theory, while the second raises questions related to stationary ergodic games and the role of entropy within game-theoretic contexts.

A. Szilárd's Engine Game

We begin by applying the framework of predictive games to thermodynamics, conceptualizing a thermodynamic cyclic as a betting game between Scientist and Nature.[4] As a fundamental prototype, we consider a work-extracting game inspired by Leó Szilárd's thought experiment [23],[5] which has been widely discussed in the context of the second law of thermodynamics and its relationship to information theory.

Consider the following work-extracting game played on a hypothetical engine illustrated in Figure 1:

(i) A partition, connected to two containers by inextensible strings, is placed at a specific position within a cylinder and fixed in place.

(ii) Scientist places a weight $m(0)$ on the left container and another weight $m(1)$ on the right container.

(iii) Nature inserts a single molecule into one side of the partition, announces whether the molecule is in the left chamber ($\omega = 0$) or the right chamber ($\omega = 1$), and then releases the partition.

(iv) If $\omega = 0$, the molecule pushes the partition to the right, and Scientist gains potential energy $m(0)g\ell_1 - m(1)g\ell_1$, where g is the gravitational acceleration and ℓ_1 is the displacement of the weights. If $\omega = 1$, on the other hand, Scientist instead gains potential energy $m(1)g\ell_0 - m(0)g\ell_0$.

(v) Once the partition reaches the end of the cylinder, it is reset to its original position as in step (i).

Szilárd's engine game
Players: Scientist and Nature.
Protocol: $W_0 = 1$.
  FOR $n \in \mathbb{Z}_{>0}$:
    Scientist announces $m_n = (m_n(0), m_n(1)) \in \mathbb{R}^\Omega$.
    Nature announces $\omega_n \in \Omega$.
    $W_n := W_{n-1} + (m_n(1) - m_n(0))g(\ell_0 + \ell_1)(\omega_n - r)$.
  END FOR.

Letting $\Omega := \{0, 1\}$ and

$$r := \frac{\ell_1}{\ell_0 + \ell_1} \in (0, 1),$$

the above procedure can be formulated as a game-theoretic process:

At first glance, this game appears to favor Nature, as Nature announces ω_n after Scientist has set the weights. However, we can prove the following.

Theorem 10: In Szilárd's engine game, Scientist has a prudent strategy $\{m_n\}_n$ that ensures $\lim_{n\to\infty} W_n = \infty$ unless

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_1(\omega_i) = r.$$

[4]A related perspective is discussed in [22], which aims to give a game-theoretic characterization of Gibbs' distribution. Our approach differs by emphasizing the coding theoretic aspect of Szilárd's engine while adhering closely to the original formalism of the Shafer-Vovk theorem.

[5]The original work [23] is in German; an English translation is available [24].

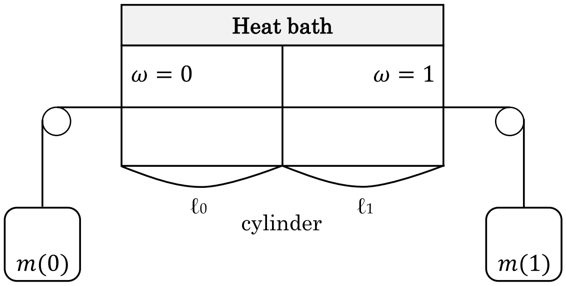

Fig. 1. Szilárd's engine game.

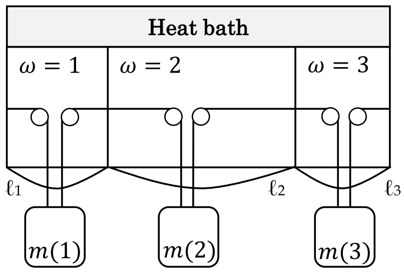

Fig. 2. Generalized Szilárd's engine game having three chambers. The pulleys can move horizontally and are assumed to be negligibly small.

Proof: Observe that the recurrence relation is rewritten as

$$W_n := W_{n-1}\left[1+\sum_{a\in\Omega}\alpha_n(a)\left(\delta_{\omega_n}(a)-p(a)\right)\right],$$

where $p := (p(0),\ p(1)) := (1-r,\ r)$ and

$$\alpha_n(a) := \frac{m_n(a)g(\ell_0+\ell_1)}{W_{n-1}}.$$

Thus, Theorem 10 is an immediate consequence of Theorem 1. ∎

The implication of Theorem 10 is as follows: If Nature does not behave in accordance with the expected statistical law, Scientist can extract an infinite amount of work from the engine. A distinctive feature of this finding is that it does not require invoking Maxwell's demon [25] or employing any measurement scheme to determine a molecule's position before setting weights. Instead, Scientist only needs to detect deviations in Nature's behavior from the law of large numbers.

This result closely resembles Kelvin's formulation of the second law of thermodynamics, which asserts that it is impossible to extract any net amount of work from a thermodynamic system while leaving the system in the same state. To clarify the position of our framework within existing theory, it is helpful to compare it with the traditional information-theoretic interpretation of Maxwell's demon. In that framework, the paradox is resolved by introducing the costs of information acquisition and erasure, where "information" serves as a compensating quantity [26], [27]. In contrast, our formulation is more flexible: it does not forbid *any* net conversion of heat into work from a single heat bath by an arbitrary cyclic process, but rather forbids the *unbounded* extraction of work from the system.

Extending the previous argument to the case when the outcome space Ω is an arbitrary finite set is straightforward. Consider a device illustrated in Figure 2, corresponding to the case when $\Omega = \{1, 2, 3\}$. The cylinder contains two partitions, dividing it into three chambers labeled by $\omega = 1, 2, 3$. Each partition is connected to two containers by inextensible strings and negligibly small pulleys that can move horizontally. Weights can be placed on these containers. The containers correspond one-to-one with the chambers and are labeled accordingly.

A generalized Szilard's engine game for $\Omega = \{1, 2, 3\}$ runs as follows:

(i) Each of two partitions is placed at a specific position within the cylinder and fixed in place.

(ii) Scientist places a weight $m(a)$ on each container a for $a = 1, 2, 3$.

(iii) Nature places a single molecule in one of the three chambers, announces its label ω, and releases the partitions.

(iv) The molecule pushes the partitions at the boundaries of chamber ω, causing the chamber to expand until all the partitions are pressed against the end(s) of the cylinder, and Scientist gains potential energy as follows: If $\omega = 1$, the work extracted is

$$m(1)g\frac{\ell_2+\ell_3}{2} - m(2)g\frac{\ell_2}{2} - m(3)g\frac{\ell_3}{2}.$$

If $\omega = 2$, the work extracted is

$$m(2)g\frac{\ell_3+\ell_1}{2} - m(3)g\frac{\ell_3}{2} - m(1)g\frac{\ell_1}{2}.$$

If $\omega = 3$, the work extracted is

$$m(3)g\frac{\ell_1+\ell_2}{2} - m(1)g\frac{\ell_1}{2} - m(2)g\frac{\ell_2}{2}.$$

(v) Once the partitions come to rest, they return to their original positions as in step (i).

In a single round of the game, Scientist extracts the following amount of work:

$$\sum_{a=1}^{3}\frac{m(a)g}{2}(\ell_1+\ell_2+\ell_3)\left(\delta_\omega(a)-\frac{\ell_a}{\ell_1+\ell_2+\ell_3}\right).$$

Generalized Szilárd's engine game
Players: Scientist and Nature.
Protocol: $W_0 = 1$.
FOR $n \in \mathbb{Z}_{>0}$:
Scientist announces $m_n \in \mathbb{R}^\Omega$.
Nature announces $\omega_n \in \Omega$.

$$W_n := W_{n-1} + \sum_{a\in\Omega}\frac{m_n(a)g}{2}(\ell_1+\cdots+\ell_A)\left(\delta_{\omega_n}(a)-p(a)\right).$$

END FOR.

Generalizing Szilárd's engine game to an arbitrary finite set $\Omega = \{1, 2, \ldots, A\}$ is straightforward: one simply increases the number of chambers illustrated in Figure 2. Defining

$$p(a) := \frac{\ell_a}{\ell_1+\cdots+\ell_A}, \quad (a\in\Omega),$$

we can formulate the generalized work-extracting protocol as follows.

Now, we extend Theorem 10 to this generalized setting:

Theorem 11: In generalized Szilárd's engine game, Scientist has a prudent strategy $\{m_n\}_n$ that ensures $\lim_{n\to\infty} W_n = \infty$ unless

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_a(\omega_i) = p(a), \quad (\forall a \in \Omega).$$

We can further generalize the game described above by allowing the chamber size ratios $p(a)$ to vary in each round n, introducing a Forecaster who announces these ratios. In this extended protocol, Theorem 2 provides a generalization of Theorem 11, incorporating a time-homogeneous finite-order Markovian Forecaster.

B. Entropy

Given the pivotal role of universal coding schemes in establishing game-theoretic law of large numbers, it is natural to expect that the protocol of a predictive game is also intertwined with the concept of entropy. The next proposition formalizes this connection, where we continue to assume that M is a kth-order Markov kernel with strictly positive entries and π is the stationary distribution of M.

Proposition 12: In the time-homogeneous kth-order Markovian predictive game, Skeptic has a prudent strategy $\alpha : \Omega^* \to \mathbb{R}^\Omega$ that ensures $\lim_{n\to\infty} K_n = \infty$ unless

$$\lim_{n\to\infty} \frac{-\log Q_{LZ}(\omega_1^n)}{n} = H(M), \tag{15}$$

where

$$H(M) := -\sum_{a_1^k \in \Omega^k} \pi(a_1^k) \sum_{b\in\Omega} M(b \mid a_1^k) \log M(b \mid a_1^k)$$

is the entropy rate.

Proof: We observed in the proof of Theorem 2 that, under the Lempel-Ziv betting strategy Q_{LZ}, the boundedness of the capital process, i.e., $\limsup_{n\to\infty} K_n < \infty$, guarantees not only that

$$\frac{\log K_n}{n} = \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^n) - (-\log Q_{LZ}(\omega_1^n))\right\} \longrightarrow 0$$

but also that

$$\hat{M}_n^k(\,\cdot \mid a_1^k) \longrightarrow M(\,\cdot \mid a_1^k) \quad \text{and} \quad \frac{T_n(a_1^k)}{n} \longrightarrow \pi(a_1^k)$$

for all $a_1^k \in \Omega^k$ as $n \to \infty$.

As a consequence, using a similar computation as in the proof of Lemma 7, we obtain

$$\begin{aligned} &-\frac{1}{n}\log \tilde{P}(\omega_1^n) \\ &= -\sum_{a_1^k\in\Omega^k} \frac{T_n(a_1^k)}{n} \sum_{b\in\Omega} \hat{M}_n^k(b \mid a_1^k) \log M(b \mid a_1^k) \\ &\quad + \frac{1}{n}\left\{-\log \tilde{P}(\omega_1^k) + \log \prod_{i=1}^{k} M(\omega_i \mid \omega_{n-k+i}^n \omega_1^{i-1})\right\} \\ &\longrightarrow H(M). \end{aligned}$$

Combining these asymptotic properties, (15) follows immediately. ■

The implication of Proposition 12 is as follows: To prevent Skeptic from becoming infinitely rich, Reality must ensure that the asymptotic compression rate of its moves coincides with the entropy rate.

Note that Proposition 12 bears a close resemblance to Lempel-Ziv's theorem [20]

$$\lim_{n\to\infty} \frac{\ell_{LZ}(\omega_1^n)}{n} = H_A(P), \quad \text{P-a.s.}$$

as well as Brudno's theorem [28]

$$\lim_{n\to\infty} \frac{\mathcal{K}(\omega_1^n)}{n} = H_A(P), \quad \text{P-a.s.}$$

for the prefix Kolmogorov complexity $\mathcal{K}(\omega_1^n)$ when data ω_1^n are drawn according to a stationary ergodic probability measure P on Ω^∞, where

$$H_A(P) := \lim_{n\to\infty} \mathbb{E}\left[-\frac{1}{n}\log_A P(\omega_1^n)\right]$$

is the entropy rate to the base A. In addition, the last asymptotic property in the proof of Proposition 12 corresponds to the Shannon-McMillan-Breiman theorem [21]

$$\lim_{n\to\infty} \left\{-\frac{1}{n}\log P(\omega_1^n)\right\} = H(P), \quad \text{P-a.s.}$$

which also has a counterpart in algorithmic randomness theory [29], [30], [31]. These observations prompt us to call a Forecaster *stationary ergodic* if they announce predictions according to the prescription

$$p_n(\omega_n) := P(\omega_n \mid \omega_1^{n-1}), \quad (\omega_1^n \in \Omega^n),$$

where P represents a predetermined stationary ergodic probability measure.

If we were to discover a compression algorithm capable of efficiently compressing Reality's moves within the game-theoretic context, we could define the entropy of a game as the asymptotic data compression rate, assuming that Reality faithfully follows the predictions of a stationary ergodic Forecaster and thereby prevents Skeptic from becoming infinitely rich.

However, this definition, which relies on the existence of a stationary ergodic Forecaster, may not be fully satisfactory from the perspective of Dawid's prequential principle [32], as the prequential framework does not impose any structural assumptions on the underlying data-generating mechanisms but instead evaluates predictive performance based solely on the observed data sequence. The validation and further exploration of the concepts of game entropy and stationary ergodic Forecasters remain topics for future investigation.

V. Concluding Remarks

In this paper, we established a generalization of the game-theoretic law of large numbers in a time-homogeneous kth-order Markovian predictive game. By constructing a Lempel-Ziv-inspired strategy based on incremental parsing and the martingale properties of the game, we provided new insights into the relationship between game-theoretic randomness and coding theory.

We also explored applications to thermodynamics by formulating a game-theoretic version of Szilard's engine. Our results demonstrated that Nature must behave stochastically, satisfying the law of large numbers, to avoid violating the second law of thermodynamics. Furthermore, we introduced the concept of entropy in predictive games, associating it to the codelength of universal coding.

Despite these advances, several important challenges remain. For instance, integrating additional thermodynamic concepts such as thermal equilibrium, thermal contact, temperature, and free energies into a game-theoretic framework remains a significant open problem. Additionally, extending the framework to non-Markovian processes could provide deeper insights into the dynamics of predictive games.

Appendix A
Proof of Lemmas

In this appendix, we provide detailed proofs of the lemmas stated in Section II-B.

A. Proof of Lemma 3

Since

$$\begin{aligned}&-\log Q_{LZ}(\omega_1^n)\\&\quad=-\log Q_{LZ}(\omega_1^{n_T}\omega_{n_T+1}^n)\\&\quad=-\log Q_{LZ}(\omega_1^{n_T})-\log Q_{LZ}(\omega_{n_T+1}^n\mid\omega_1^{n_T})\end{aligned}$$

for $n>n_T$, it follows from (7) and (8) that, for any n,

$$\begin{aligned}-\log Q_{LZ}(\omega_1^n)&\le-\log Q_{LZ}(\omega_1^{n_T})+\log|V(\omega_1^{n_T})|\\&=\sum_{j=0}^{c(\omega_1^n)}\log(A+j(A-1))\\&<\sum_{j=0}^{c(\omega_1^n)}\log(A+c(\omega_1^n)(A-1))\\&=(c(\omega_1^n)+1)\log(A+c(\omega_1^n)(A-1)).\end{aligned}$$

Consequently,

$$\begin{aligned}&\frac{1}{n}\left\{c(\omega_1^n)\log c(\omega_1^n)-(-\log Q_{LZ}(\omega_1^n))\right\}\\&\quad>\frac{1}{n}c(\omega_1^n)\log c(\omega_1^n)\\&\qquad-\frac{1}{n}(c(\omega_1^n)+1)\log(A+c(\omega_1^n)(A-1))\\&\quad=\frac{c(\omega_1^n)}{n}\log\frac{c(\omega_1^n)}{A+c(\omega_1^n)(A-1)}\\&\qquad-\frac{\log(A+c(\omega_1^n)(A-1))}{n}.\end{aligned}$$

Thus, the following Lemma 13 proves the claim.

Lemma 13: For sufficiently large n and for all $\omega_1^n\in\Omega^n$,

$$(0<)\ c(\omega_1^n)<\frac{n}{(1-\varepsilon_n)\log_A n},$$

where $\varepsilon_n\to0$ as $n\to\infty$. Specifically, $c(\omega_1^n)/n\to0$ as $n\to\infty$.

Proof: See Lemma 13.5.3 of [21]. ∎

B. Proof of Lemma 4

The first identity follows from

$$\begin{aligned}&\sum_{b\in\Omega}T_n(b\mid a_1^\ell)\\&\quad=\sum_{b\in\Omega}(\text{number of occurrences of } a_1^\ell b \text{ in } \omega_1^n\omega_1^\ell)\\&\quad=(\text{number of occurrences of } a_1^\ell \text{ in } \omega_1^n\omega_1^{\ell-1})=T_n(a_1^\ell).\end{aligned}$$

On the other hand, observe that

$$\begin{aligned}&\sum_{a_1\in\Omega}T_n(b\mid a_1^\ell)\\&\quad=\sum_{a_1\in\Omega}(\text{number of occurrences of } a_1^\ell b \text{ in } \omega_1^n\omega_1^\ell)\\&\quad=(\text{number of occurrences of } a_2^\ell b \text{ in } \omega_2^n\omega_1^\ell)\\&\quad=(\text{number of occurrences of } a_2^\ell b \text{ in } \omega_1^n\omega_1^{\ell-1})+\Delta_1+\Delta_\ell,\end{aligned}$$

where

$$\begin{aligned}\Delta_1:&=(\text{adjustment for the effect of adding } \omega_1\\&\qquad\text{to the head of } \omega_2^n\omega_1^\ell)\\&=\begin{cases}-1&(\omega_1^\ell=a_2^\ell b),\\0&(\text{otherwise}),\end{cases}\end{aligned}$$

and

$$\begin{aligned}\Delta_\ell:&=(\text{adjustment for the effect of removing } \omega_\ell\\&\qquad\text{from the tail of } \omega_2^n\omega_1^\ell)\\&=\begin{cases}1&(\omega_1^\ell=a_2^\ell b),\\0&(\text{otherwise}).\end{cases}\end{aligned}$$

Since $\Delta_1+\Delta_\ell=0$, the second identity holds.

C. Proof of Lemma 5

The assumption $\hat{M}_n^k\to M$ ensures that for sufficiently large n, we have $\hat{M}_n^k(b\mid a_1^k)>0$ for all $a_1^k\in\Omega^k$ and $b\in\Omega$. For $\omega^k\in\Omega^k$, define $\hat{q}_n(\omega^k):=T_n(\omega^k)/n$. Then, by Lemma 4, the empirical distribution $\hat{q}_n$ is stationary under the kth Markov kernel $\hat{M}_n^k$ [33]:

$$\begin{aligned}\sum_{a_1\in\Omega}\hat{M}_n^k(b\mid a_1^k)\hat{q}_n(a_1^k)&=\sum_{a_1\in\Omega}\frac{T_n(b\mid a_1^k)}{T_n(a_1^k)}\frac{T_n(a_1^k)}{n}\\&=\frac{1}{n}\sum_{a_1\in\Omega}T_n(b\mid a_1^k)=\frac{T_n(a_2^kb)}{n}\\&=\hat{q}_n(a_2^kb).\end{aligned}$$

Thus, by Lemma 16 in Appendix B, $\hat{q}_n$ is the unique stationary distribution of Markov matrix $\hat{M}_n^k$.

Consider a convergent subsequence $\{\hat{q}_{n_i}-\pi\}_i$ of $\{\hat{q}_n-\pi\}_n$, which converges to some $r\in[-1,1]^{\Omega^k}$. By the assumption $\hat{M}_n^k\to M$, we obtain

$$\begin{aligned}r&=\lim_{i\to\infty}(\hat{q}_{n_i}-\pi)=\lim_{i\to\infty}(\hat{M}_{n_i}^k\hat{q}_{n_i}-M\pi)\\&=\lim_{i\to\infty}\left\{\hat{M}_{n_i}^k(\hat{q}_{n_i}-\pi)+(\hat{M}_{n_i}^k-M)\pi\right\}=Mr.\end{aligned}$$

By the Perron-Frobenius theorem (Section 4.4 of [34]), there exists a constant $c \in \mathbb{R}$ satisfying $r = c\pi$. Furthermore,

$$\sum_{a_1^k \in \Omega^k} r(a_1^k) = \lim_{i\to\infty} \sum_{a_1^k \in \Omega^k} (\hat{q}_{n_i}(a_1^k) - \pi(a_1^k)) = 0.$$

Thus, we must have $c = 0$, completing the proof.

D. Proof of Lemma 6

Suppose the sequence ω_1^n is parsed into C_n distinct substrings as

$$\omega_1^n = /\omega_{n_0+1}^{n_1}/\omega_{n_1+1}^{n_2}/\cdots/\omega_{n_{C_n-1}+1}^{n_{C_n}},$$

where $n_0 = 0$ and $n_{C_n} = n$. For example, if ω_1^n is parsed using the incremental parsing algorithm as

$$\omega_1^n = /\omega_{n_0+1}^{n_1}/\omega_{n_1+1}^{n_2}/\cdots/\omega_{n_{T-1}+1}^{n_T}/\omega_{n_T+1}^{n},$$

we define $C_n = T$ and set $\omega_{n_{C_n-1}+1}^{n_{C_n}} := \omega_{n_{T-1}+1}^{n_T}\omega_{n_T+1}^{n}$.

Now, define $s_i := \omega_{i-\ell}^{i-1}$ for $\ell + 1 \le i \le n$, and extend them cyclically for $1 \le i \le \ell$ as follows:

$$s_1 := \omega_{n-\ell+1}^n, \quad s_2 := \omega_{n-\ell+2}^n\omega_1, \quad \ldots, \quad s_\ell := \omega_n\omega_1^{\ell-1}.$$

For $m \in \mathbb{Z}_{>0}$ and $s \in \Omega^\ell$, let $c_{m,s}$ denote the number of occurrences of the word $\omega_{n_{j-1}+1}^{n_j}$ of length m such that $s_{n_{j-1}+1} = \omega_{n_{j-1}-\ell+1}^{n_{j-1}} = s$ among the C_n substrings $\omega_{n_0+1}^{n_1}, \omega_{n_1+1}^{n_2}, \ldots, \omega_{n_{C_n-1}+1}^{n_{C_n}}$, i.e.,

$$c_{m,s} := \left|\left\{ j \in \{1,2,\ldots,C_n\} \,\middle|\, |\omega_{n_{j-1}+1}^{n_j}| = m,\ s_{n_{j-1}+1} = s \right\}\right|.$$

Letting $\mathcal{U} := \{(m,s) \in \mathbb{Z}_{>0} \times \Omega^\ell \mid c_{m,s} > 0\}$, we have

$$\sum_{(m,s)\in\mathcal{U}} c_{m,s} = C_n \quad\text{and}\quad \sum_{(m,s)\in\mathcal{U}} m \cdot c_{m,s} = n.$$

For each $(m,s) \in \mathcal{U}$, let

$$\mathcal{J}_{m,s} = \{ j \in \{1,2,\ldots,C_n\} \mid n_j - n_{j-1} = m,\ s_{n_{j-1}+1} = s\}.$$

With a slight abuse of notation, we define, for any $m \in \mathbb{Z}_{>\ell}$ and $a_1^m \in \Omega^m$,

$$\hat{M}_n^\ell(a_{\ell+1}^m \mid a_1^\ell) := \prod_{i=\ell+1}^{m} \hat{M}_n^\ell(a_i \mid a_{i-\ell}^{i-1}).$$

Then, we can evaluate $\log \hat{R}_n^\ell(\omega_1^n)$ as follows:

$$\begin{aligned}
&\log \hat{R}_n^\ell(\omega_1^n)\\
&= \sum_{j=1}^{C_n} \log \hat{M}_n^\ell(\omega_{n_{j-1}+1}^{n_j} \mid s_{n_{j-1}+1})\\
&= \sum_{(m,s)\in\mathcal{U}} \sum_{j\in\mathcal{J}_{m,s}} \log \hat{M}_n^\ell(\omega_{n_{j-1}+1}^{n_j} \mid s_{n_{j-1}+1})\\
&= \sum_{(m,s)\in\mathcal{U}} c_{m,s} \sum_{j\in\mathcal{J}_{m,s}} \frac{1}{c_{m,s}} \log \hat{M}_n^\ell(\omega_{n_{j-1}+1}^{n_j} \mid s_{n_{j-1}+1})\\
&\le \sum_{(m,s)\in\mathcal{U}} c_{m,s} \log\left(\frac{1}{c_{m,s}} \sum_{j\in\mathcal{J}_{m,s}} \hat{M}_n^\ell(\omega_{n_{j-1}+1}^{n_j} \mid s_{n_{j-1}+1})\right).
\end{aligned}$$

In the last inequality, we used Jensen's inequality. Since the parsed substrings $\{\omega_{n_{j-1}+1}^{n_j}\}_{1\le j\le C_n}$ are distinct, we have

$$\sum_{j\in\mathcal{J}_{m,s}} \hat{M}_n^\ell(\omega_{n_{j-1}+1}^{n_j} \mid s_{n_{j-1}+1}) \le 1$$

for all $(m,s) \in \mathcal{U}$. As a consequence,

$$\begin{aligned}
&\log \hat{R}_n^\ell(\omega_1^n)\\
&\le -\sum_{(m,s)\in\mathcal{U}} c_{m,s} \log c_{m,s}\\
&= -c(\omega_1^n)\log c(\omega_1^n) - c(\omega_1^n) \sum_{(m,s)\in\mathcal{U}} \frac{c_{m,s}}{c(\omega_1^n)} \log \frac{c_{m,s}}{c(\omega_1^n)},
\end{aligned}$$

where $c(\omega_1^n) = C_n$. Writing $\pi_{m,s} := c_{m,s}/c(\omega_1^n)$, we have

$$\sum_{(m,s)\in\mathcal{U}} \pi_{m,s} = 1 \quad\text{and}\quad \sum_{(m,s)\in\mathcal{U}} m \cdot \pi_{m,s} = \frac{n}{c(\omega_1^n)}.$$

We now define the random variables U and V as follows:

$$\Pr(U = m,\ V = s) := \pi_{m,s}.$$

From the above bound on $\log \hat{R}_n^\ell(\omega_1^n)$, it follows that

$$-\frac{1}{n}\log \hat{R}_n^\ell(\omega_1^n) \ge \frac{c(\omega_1^n)}{n}\log c(\omega_1^n) - \frac{c(\omega_1^n)}{n} H(U,V),$$

where

$$H(U,V) := -\sum_{(m,s)\in\mathcal{U}} \frac{c_{m,s}}{c(\omega_1^n)} \log \frac{c_{m,s}}{c(\omega_1^n)}.$$

By the subadditivity of entropy, we have

$$H(U,V) \le H(U) + H(V).$$

Since the expectation of U is given by

$$\mathbb{E}[U] = \frac{n}{c(\omega_1^n)},$$

applying Lemma 14 below, we can bound $H(U)$ as

$$\begin{aligned}
H(U) &\le (\mathbb{E}[U]+1)\log(\mathbb{E}[U]+1) - (\mathbb{E}[U])\log(\mathbb{E}[U])\\
&= \log \frac{n}{c(\omega_1^n)} + \left(\frac{n}{c(\omega_1^n)} + 1\right)\log\left(\frac{c(\omega_1^n)}{n} + 1\right).
\end{aligned}$$

On the other hand, since $H(V) \le \log|\Omega|^\ell = \ell \log A$, we obtain

$$\begin{aligned}
\delta_\ell(n) :&= \frac{c(\omega_1^n)}{n} H(U,V)\\
&\le \frac{c(\omega_1^n)}{n}\log\frac{n}{c(\omega_1^n)} + \left(1 + \frac{c(\omega_1^n)}{n}\right)\log\left(\frac{c(\omega_1^n)}{n} + 1\right)\\
&\quad + \frac{c(\omega_1^n)}{n}\ell\log A.
\end{aligned}$$

Since $c(\omega_1^n)/n \to 0$ as $n \to \infty$ by Lemma 13, it follows that $\delta_\ell(n) \to 0$ as $n \to \infty$. This completes the proof.

Lemma 14: Let Z be a nonnegative integer-valued random variable with mean μ. Then the entropy $H(Z)$ is bounded by

$$H(Z) \le (\mu+1)\log(\mu+1) - \mu\log\mu.$$

Proof: See Lemma 13.5.4 of [21]. ■

E. Proof of Lemma 7

For $n, \ell \in \mathbb{Z}_{>0}$ satisfying $n > \ell \geq k$,

$$\begin{aligned} &-\log \tilde{P}(\omega_1^n) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) - \log \tilde{P}(\omega_{\ell+1}^n \mid \omega_1^\ell) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) - \log \prod_{i=\ell+1}^{n} M(\omega_i \mid \omega_{i-k}^{i-1}) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) \\ &\quad - \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) \\ &\quad - \log \prod_{i=\ell+1}^{n} M(\omega_i \mid \omega_{i-k}^{i-1}) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) \\ &\quad - \sum_{a_1^\ell \in \Omega^\ell} \sum_{b \in \Omega} T_n(b \mid a_1^\ell) \log M(b \mid a_{\ell-k+1}^\ell) \\ &\quad + \log \hat{R}_n^\ell(\omega_1^n). \end{aligned}$$

In the last equality, we used the fact that

$$M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) = M(\omega_i \mid \omega_{n-\ell+i}^{n}\omega_1^{i-1}) \text{ and}$$
$$M(\omega_i \mid \omega_{i-k}^{i-1}) = M(\omega_i \mid \omega_{i-\ell}^{i-1}),$$

since $\ell \geq k$ and M is the kth-order Markov kernel.

Substituting the definition of $\hat{R}_n^l(\omega_1^n)$, the computation follows as:

$$\begin{aligned} &-\log \tilde{P}(\omega_1^n) + \log \hat{R}_n^\ell(\omega_1^n) \\ &= -\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) \\ &\quad + \sum_{a_1^\ell \in \Omega^\ell} \sum_{b \in \Omega} T_n(b \mid a_1^\ell) \\ &\qquad \times \left(-\log M(b \mid a_{\ell-k+1}^\ell) + \log \hat{M}_n^\ell(b \mid a_1^\ell)\right) \\ &= -\log \tilde{P}(\omega_1^\ell) + \log \prod_{i=1}^{\ell} M(\omega_i \mid \omega_{n-k+i}^{n}\omega_1^{i-1}) \\ &\quad + \sum_{a_1^\ell \in \Omega^\ell} T_n(a_1^\ell) \cdot D\left(\hat{M}_n^\ell(\,\cdot \mid a_1^{\ell-1}) \,\middle\|\, M(\,\cdot \mid a_{\ell-k+1}^\ell)\right). \end{aligned}$$

This completes the proof.

Appendix B
Stationary Distributions of Markov Chains

Given a Markov matrix $M : \Omega \times \Omega \to (0,1) : (a,b) \mapsto M(a \mid b)$ satisfying

$$\sum_{a \in \Omega} M(a \mid b) = 1$$

for all $b \in \Omega$, let $M^{(m)}$ be the mth power of M, in that $M^{(1)} := M$ and

$$M^{(m)}(a \mid b) := \sum_{c \in \Omega} M(a \mid c) M^{(m-1)}(c \mid b).$$

We recall the following well-known fact.

Lemma 15: There exists a unique probability distribution μ on Ω such that for any $a, b \in \Omega$,

$$\lim_{m \to \infty} M^{(m)}(a \mid b) = \mu(a),$$

and μ is the stationary distribution of M.

Proof: See Theorem 6 in Chapter 4 of [34]. ■

Now, consider a kth-order Markov kernel $M : \Omega \times \Omega^k \to (0,1) : (a, \omega_1^k) \mapsto M(a \mid \omega_1^k)$ that satisfies

$$\sum_{a \in \Omega} M(a \mid \omega_1^k) = 1 \qquad (\forall \omega_1^k \in \Omega^k).$$

Lemma 16: For the kth-order Markov kernel M, there is a unique stationary distribution $\pi : \Omega^k \to (0,1)$ satisfying

$$\pi(a_2^{k+1}) = \sum_{a_1 \in \Omega} M(a_{k+1} \mid a_1^k) \pi(a_1^k).$$

Proof: Define $\tilde{M} : \Omega^k \times \Omega^k \to [0,1)$ by

$$\tilde{M}(a_1^k \mid b_1^k) := \begin{cases} M(a_k \mid b_1^k) & (a_1^k = b_2^k a_k), \\ 0 & (\text{otherwise}). \end{cases}$$

Since

$$\begin{aligned} \sum_{a_1^k \in \Omega^k} \tilde{M}(a_1^k \mid b_1^k) &= \sum_{a_k \in \Omega} \sum_{a_1^{k-1} \in \Omega^{k-1}} \tilde{M}(a_1^k \mid b_1^k) \\ &= \sum_{a_k \in \Omega} M(a_k \mid b_1^k) = 1, \end{aligned}$$

we can regard $\tilde{M}$ as a first-order Markov kernel on Ω^k. Moreover it is straightforward to verify that

$$\tilde{M}^{(k)}(a_1^k \mid b_1^k) = \prod_{i=1}^{k} M(a_i \mid b_i^k a_1^{i-1}).$$

This expression ensures that $\tilde{M}^{(k)}(a_1^k \mid b_1^k) > 0$ for all $a_1^k, b_1^k \in \Omega^k$. Thus, by applying Lemma 15 to the Markov matrix $\tilde{M}^{(k)}$, we conclude that there exists a unique distribution π on Ω^k satisfying

$$\tilde{M}^{(k)} \pi = \pi.$$

Furthermore, since

$$\tilde{M}^{(k)}(M\pi) = M(\tilde{M}^{(k)}\pi) = M\pi,$$

the uniqueness of the stationary distribution for $\tilde{M}^{(k)}$ implies that $M\pi = \pi$. ■

Appendix C
Lynch-Davisson Betting Strategy for Simple Predictive Game

In this appendix, we present an alternative proof of Theorem 1 using one of the simplest universal data compression schemes [35], [36]. As a by-product, we also analyze the convergence rate of the empirical distribution.

We begin with a binary case and consider describing a binary sequence $x^n = 11001$ of length $n = 5$. For $a \in \{0, 1\}$, let $S_n(a)$ denote the number of occurrences of a in x^n. The sequence can be identified by first specifying its type (also known as the empirical distribution):

$$\hat{P}_{x^n} = \left(\frac{S_n(0)}{n}, \frac{S_n(1)}{n}\right) = \left(\frac{2}{5}, \frac{3}{5}\right)$$

and then specifying the index of this sequence among all sequences of length $n = 5$ that share this type. Thus, the given binary sequence x^n can be described by another binary sequence as follows:

$$\underbrace{\boxed{\text{specify the type}}}_{\lceil \log_2(n+1)\rceil \text{ bits}} + \underbrace{\boxed{\text{specify the sequence in the type class}}}_{\left\lceil \log_2 \binom{n}{S_n(0),\, S_n(1)} \right\rceil \text{ bits}}$$

This scheme is called the *Lynch-Davisson code*, and its codelength is given by

$$\begin{aligned}\ell_{LD}(x^n) &= \lceil \log_2(n+1)\rceil + \left\lceil \log_2 \frac{n!}{S_n(0)!\, S_n(1)!} \right\rceil \\ &= \log_2 \frac{(n+1)!}{S_n(0)!\, S_n(1)!} + O(1).\end{aligned}$$

Generalizing to a generic alphabet $\Omega = \{1, 2, \ldots, A\}$ is straightforward, and the corresponding Lynch-Davisson codelength is

$$\begin{aligned}\ell_{LD}(x^n) = &\left\lceil \log_A \frac{(n+A-1)!}{n!(A-1)!} \right\rceil \\ &+ \left\lceil \log_A \frac{n!}{S_n(1)!\, S_n(2)! \cdots S_n(A)!} \right\rceil \\ = &\ \log_A \frac{(n+A-1)!}{(A-1)!\, S_n(1)!\, S_n(2)! \cdots S_n(A)!} + O(1).\end{aligned}$$

Now, we are ready to prove Theorem 1.

Proof of Theorem 1: Let us introduce the reference probability measure P on Ω^* defined by

$$P(\omega_1^n) := \prod_{i=1}^{n} p(\omega_i),$$

and consider the "randomness deficiency" function $\mathcal{L}_{LD}(\omega_1^n)$ for the Lynch-Davisson codelength $\ell_{LD}(\omega_1^n)$ relative to the Shannon codelength $-\log_A P(\omega_1^n)$ defined by

$$\mathcal{L}_{LD}(\omega_1^n) := -\log_A P(\omega_1^n) - \ell_{LD}(\omega_1^n).$$

A crucial observation is that

$$\begin{aligned}&\mathcal{L}_{LD}(\omega_1^n) \\ &= -\log_A \prod_{i=1}^{n} p(\omega_i) \\ &\quad + \log_A \frac{(A-1)!\, S_n(1)!\, S_n(2)! \cdots S_n(A)!}{(n+A-1)!} + O(1) \\ &= -\sum_{a\in\Omega} S_n(a) \log_A p(a) \\ &\quad + \log_A \frac{(A-1)!\, S_n(1)!\, S_n(2)! \cdots S_n(A)!}{(n+A-1)!} + O(1) \qquad (16) \\ &= (\log_A e) \left\{ \sum_{a\in\Omega} S_n(a) \log \frac{S_n(a)}{p(a)} - n\log n - O(\log n) \right\} \\ &= n(\log_A e) \left\{ D(\hat{P}_{\omega_1^n} \| p) - O\left(\frac{\log n}{n}\right) \right\}, \qquad (17)\end{aligned}$$

where Stirling's formula was used in the third equality.

The relation (17) shows that $\limsup_{n\to\infty} \mathcal{L}_{LD}(\omega_1^n) = \infty$ if $\hat{P}_{\omega_1^n}$ does not converge to p. It then suffices to show that there exists a prudent betting strategy α_n that realizes

$$K_n \propto A^{\mathcal{L}_{LD}(\omega_1^n)} = \frac{(A-1)!\, S_n(1)!\, S_n(2)! \cdots S_n(A)!}{\prod_{i=1}^{n} p(\omega_i) \cdot (n+A-1)!}. \qquad (18)$$

If this were the case, then

$$\begin{aligned}\frac{K_n}{K_{n-1}} &= \frac{1}{p(\omega_n)(n+A-1)} \qquad (19) \\ &\quad \times \frac{S_n(1)!\, S_n(2)! \cdots S_n(A)!}{S_{n-1}(1)!\, S_{n-1}(2)! \cdots S_{n-1}(A)!} \\ &= \frac{1}{p(\omega_n)(n+A-1)} \cdot (S_{n-1}(\omega_n) + 1). \qquad (20)\end{aligned}$$

Comparing this with the recursion formula

$$K_n := K_{n-1} \left\{ 1 + \alpha_n(\omega_n) - \sum_{a\in\Omega} \alpha_n(a) p(a) \right\},$$

we find that

$$\alpha_n(a) := \frac{S_{n-1}(a) + 1}{p(a)(n+A-1)} \qquad (21)$$

gives a desired prudent betting strategy that satisfies (18). In fact, since

$$\begin{aligned}\sum_{a\in\Omega} \alpha_n(a) p(a) &= \frac{1}{n+A-1} \sum_{a\in\Omega} \{S_{n-1}(a) + 1\} \\ &= \frac{1}{n+A-1} \{(n-1) + A\} = 1,\end{aligned}$$

we have

$$\left\{ 1 + \alpha_n(\omega_n) - \sum_{a\in\Omega} \alpha_n(a) p(a) \right\} = \alpha_n(\omega_n),$$

which is identical to the right-hand side of (20).

In summary, the prudent betting strategy (21) ensures that

$$\limsup_{n\to\infty} \log_A K_n = \limsup_{n\to\infty} \mathcal{L}_{LD}(\omega_1^n) = \infty$$

if $\hat{P}_{\omega_1^n}$ does not converge to p as $n \to \infty$. The proof is complete. ■

Remark: In Theorem 1, the two events $K_n \to \infty$ and $\hat{P}_{\omega_1^n} \to p$ are not necessarily mutually exclusive, and both may occur simultaneously.[6] For example, suppose that $\hat{P}_{\omega_1^n}$ converges to p at the rate

$$\|\hat{P}_{\omega_1^n} - p\| = O(\sqrt{\log n / n})$$

and satisfies

$$\limsup_{n\to\infty} \frac{n}{\log n} \sum_{a\in\Omega} \frac{(\hat{P}_{\omega_1^n}(a) - p(a))^2}{p(a)} > A - 1.$$

Then, we have

$$\limsup_{n\to\infty} K_n = \infty.$$

Proof: Applying Stirling's formula

$$\log n! = \left(n + \frac{1}{2}\right) \log n - n + O(1),$$

[6] A similar argument is found in [13].

we get

$$\begin{aligned}
&\log \frac{S_n(1)!\,S_n(2)!\cdots S_n(A)!}{(n+A-1)!}\\
&= \sum_{a\in\Omega}\left\{\left(S_n(a)+\frac{1}{2}\right)\log S_n(a) - S_n(a)\right\}\\
&\quad -\left\{\left((n+A-1)+\frac{1}{2}\right)\log(n+A-1)-(n+A-1)\right\}\\
&\quad +O(1)\\
&= \sum_{a\in\Omega} S_n(a)\log S_n(a) + \frac{1}{2}\sum_{a\in\Omega}\log S_n(a)\\
&\quad -\left(n+A-\frac{1}{2}\right)\log(n+1)+O(1).
\end{aligned}$$

Thus, the randomness deficiency function $\mathcal{L}_{LD}(\omega_1^n)$ is evaluated using (16) as

$$\begin{aligned}
\frac{\mathcal{L}_{LD}(\omega_1^n)}{n\log_A e} &= -\sum_{a\in\Omega}\frac{S_n(a)}{n}\log p(a)\\
&\quad +\frac{1}{n}\log\frac{S_n(1)!\,S_n(2)!\cdots S_n(A)!}{(n+A-1)!}+O\left(\frac{1}{n}\right)\\
&= \sum_{a\in\Omega}\frac{S_n(a)}{n}(-\log p(a)+\log S_n(a))\\
&\quad +\frac{1}{2n}\sum_{a\in\Omega}\log S_n(a)\\
&\quad -\frac{1}{n}\left(n+A-\frac{1}{2}\right)\log(n+A-1)+O\left(\frac{1}{n}\right)\\
&= \sum_{a\in\Omega}\hat{P}_{\omega_1^n}(a)\left(\log\frac{\hat{P}_{\omega_1^n}(a)}{p(a)}+\log n\right)\\
&\quad +\frac{1}{2n}\sum_{a\in\Omega}\left(\log\hat{P}_{\omega_1^n}(a)+\log n\right)\\
&\quad -\left(1+\frac{A}{n}-\frac{1}{2n}\right)\log(n+A-1)+O\left(\frac{1}{n}\right)\\
&= \sum_{a\in\Omega}\hat{P}_{\omega_1^n}(a)\log\frac{\hat{P}_{\omega_1^n}(a)}{p(a)}\\
&\quad +\frac{1}{2n}\sum_{a\in\Omega}\log\hat{P}_{\omega_1^n}(a)-\frac{A-1}{2n}\log n+O\left(\frac{1}{n}\right).
\end{aligned} \tag{22}$$

Letting $Q_n(a) := \hat{P}_{\omega_1^n}(a)-p(a)$, we evaluate the first term of (22) as

$$\begin{aligned}
&\sum_{a\in\Omega}\hat{P}_{\omega_1^n}(a)\log\frac{\hat{P}_{\omega_1^n}(a)}{p(a)}\\
&= \sum_{a\in\Omega}p(a)\left(1+\frac{Q_n(a)}{p(a)}\right)\log\left(1+\frac{Q_n(a)}{p(a)}\right)\\
&= \sum_{a\in\Omega}p(a)\left\{\frac{Q_n(a)}{p(a)}+\frac{1}{2}\left(\frac{Q_n(a)}{p(a)}\right)^2-O\left(\frac{Q_n(a)}{p(a)}\right)^3\right\}\\
&= 0+\frac{1}{2}\sum_{a\in\Omega}\frac{Q_n(a)^2}{p(a)}+O\left(|Q_n|^3\right).
\end{aligned} \tag{23}$$

Combining (22) and (23), we have

$$\begin{aligned}
\frac{\mathcal{L}_{LD}(\omega_1^n)}{n\log_A e} &= \frac{1}{2}\sum_{a\in\Omega}\frac{Q_n(a)^2}{p(a)}-\frac{A-1}{2n}\log n\\
&\quad +\frac{1}{2n}\sum_{a\in\Omega}\log\hat{P}_{\omega_1^n}(a)+O\left(|Q_n|^3\right)+O\left(\frac{1}{n}\right),
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{\mathcal{L}_{LD}(\omega_1^n)}{\log_A e} &= \frac{\log n}{2}\left[\frac{n}{\log n}\sum_{a\in\Omega}\frac{Q_n(a)^2}{p(a)}-(A-1)\right]\\
&\quad +\sum_{a\in\Omega}\log\hat{P}_{\omega_1^n}(a)+nO\left(|Q_n|^3\right)+O(1).
\end{aligned} \tag{24}$$

Now, by the assumption that $|Q_n| = |\hat{P}_{\omega_1^n}-p| = O(\sqrt{\log n/n})$, we have

$$\sum_{a\in\Omega}\log\hat{P}_{\omega_1^n}(a)=O(1) \quad \text{and} \quad nO\left(|Q_n|^3\right)\to 0.$$

It then follows from (24) that

$$\limsup_{n\to\infty}\frac{n}{\log n}\sum_{a\in\Omega}\frac{(\hat{P}_{\omega_1^n}(a)-p(a))^2}{p(a)} > A-1,$$

implies $\limsup_{n\to\infty}\mathcal{L}_{LD}(\omega_1^n)=\infty$. This completes the proof. ■

Note that the quantity

$$\sum_{a\in\Omega}\frac{(\hat{P}_{\omega_1^n}(a)-p(a))^2}{p(a)}$$

corresponds to the Fisher information.

Acknowledgment

The authors would like to express their gratitude to Prof. Kenshi Miyabe for insightful discussions. They also thank the anonymous reviewers for their valuable comments and suggestions, which have greatly improved the clarity and quality of this article. Akio Fujiwara wishes to convey deep appreciation to the late Prof. Tom Cover for his enlightening and thought-provoking discussions on Maxwell's demon.

References

[1] G. Shafar and V. Vovk, *Probability and Finance: It's Only a Game!*. New York, NY, USA: Wiley, 2001.

[2] A. Takemura, V. Vovk, and G. Shafer, "The generality of the zero-one laws," *Ann. Inst. Stat. Math.*, vol. 63, no. 5, pp. 873–885, Oct. 2011.

[3] G. Shafer, V. Vovk, and A. Takemura, "Lévy's zero–one law in game-theoretic probability," *J. Theor. Probab.*, vol. 25, no. 1, pp. 1–24, Mar. 2012.

[4] K. Miyabe and A. Takemura, "Convergence of random series and the rate of convergence of the strong law of large numbers in game-theoretic probability," *Stochastic Processes their Appl.*, vol. 122, no. 1, pp. 1–30, Jan. 2012.

[5] K. Miyabe and A. Takemura, "Derandomization in game-theoretic probability," *Stochastic Processes Appl.*, vol. 125, no. 1, pp. 39–59, Jan. 2015.

[6] V. Vovk, "Merging of opinions in game-theoretic probability," *Ann. Inst. Stat. Math.*, vol. 61, no. 4, pp. 969–993, Dec. 2009.

[7] G. Shafar and V. Vovk, *Game-Theoretic Foundations for Probability and Finance*. Hoboken, NJ, USA: Wiley, 2019.

[8] A. Ramdas, P. Grünwald, V. Vovk, and G. Shafer, "Game-theoretic statistics and safe anytime-valid inference," *Stat. Sci.*, vol. 38, no. 4, pp. 576–601, Nov. 2023.

[9] S. Shekhar and A. Ramdas, "Nonparametric two-sample testing by betting," *IEEE Trans. Inf. Theory*, vol. 70, no. 2, pp. 1178–1203, Feb. 2024.
[10] F. Orabona and K.-S. Jun, "Tight concentrations and confidence sequences from the regret of universal portfolio," *IEEE Trans. Inf. Theory*, vol. 70, no. 1, pp. 436–455, Jan. 2024.
[11] A. Ramdas, J. Ruf, M. Larsson, and W. M. Koolen, "Testing exchangeability: Fork-convexity, supermartingales and e-processes," *Int. J. Approx. Reasoning*, vol. 141, pp. 83–109, Feb. 2022.
[12] G. Fields, T. Javidi, and S. Shekhar, "Sequential one-sided hypothesis testing of Markov chains," 2025, *arXiv:2501.13187*.
[13] M. Kumon, A. Takemura, and K. Takeuchi, "Capital process and optimality properties of a Bayesian skeptic in coin-tossing games," *Stochastic Anal. Appl.*, vol. 26, no. 6, pp. 1161–1180, Nov. 2008.
[14] K. Takeuchi, M. Kumon, and A. Takemura, "Multistep Bayesian strategy in coin-tossing games and its application to asset trading games in continuous time," *Stochastic Anal. Appl.*, vol. 28, no. 5, pp. 842–861, Aug. 2010.
[15] M. Feder, "Gambling using a finite state machine," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1459–1465, Sep. 1991.
[16] G. Shafer, "Testing by betting: A strategy for statistical and scientific communication," *J. Roy. Stat. Soc. Ser. A: Statist. Soc.*, vol. 184, no. 2, pp. 407–431, May 2021.
[17] D. Williams, *Probability With Martingales*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[18] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed., New York, NY, USA: Springer, 1997.
[19] E. Solan, *A Course in Stochastic Game Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2022.
[20] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.
[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Hoboken, NJ, USA: Wiley, 2006.
[22] K. Hiura, "Gibbs distribution from sequentially predictive form of the second law," *J. Stat. Phys.*, vol. 185, no. 1, p. 2, Oct. 2021.
[23] L. Szilard, "Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," *Z. Phys.*, vol. 53, no. 11, pp. 840–856, Nov. 1929.
[24] L. Szilard, "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings," *Behav. Sci.*, vol. 9, no. 4, pp. 301–310, 1964.
[25] H. S. Leff and A. F. Rex, Eds., *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing*. Bristol, U.K.: Institute of Physics Publishing, 2003.
[26] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Develop.*, vol. 5, no. 3, pp. 183–191, Jul. 1961.
[27] C. H. Bennett, "Notes on Landauer's principle, reversible computation, and Maxwell's demon," *Stud. Hist. Philosophy Sci. B, Stud. Hist. Philosophy Mod. Phys.*, vol. 34, no. 3, pp. 501–510, Sep. 2003.
[28] A. A. Brudno, "Entropy and the complexity of the trajectories of a dynamical system," *Trans. Mosc. Math. Soc.*, vol. 44, pp. 127–151, Jan. 1983.
[29] V. V. V'Yugin, "Ergodic theorems for individual random sequences," *Theor. Comput. Sci.*, vol. 207, no. 2, pp. 343–361, Nov. 1998.
[30] M. Nakamura, "Ergodic theorems for algorithmically random sequences," in *Proc. IEEE Inf. Theory Workshop*, Mar. 2005, pp. 1–4.
[31] M. Hochman, "Upcrossing inequalities for stationary sequences and applications," *Ann. Probab.*, vol. 37, no. 6, pp. 2135–2149, Nov. 2009.
[32] A. P. Dawid, "Present position and potential developments: Some personal views: Statistical theory: The prequential approach," *J. Roy. Stat. Soc. A, Gen.*, vol. 147, no. 2, pp. 278–292, Mar. 1984.
[33] L. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 431–438, Jul. 1981.
[34] R. G. Gallager, *Discrete Stochastic Processes*. Boston, MA, USA: Kluwer, 1996.
[35] T. J. Lynch, "Sequence time coding for data compression," *Proc. IEEE*, vol. 54, no. 10, pp. 1490–1491, Oct. 1966.
[36] L. D. Davisson, "Comments on 'sequence time coading for data compression,"' *Proc. IEEE*, vol. 54, no. 12, p. 2010, Dec. 1966.

Takara Nomura is currently pursuing the Ph.D. degree with the Department of Mathematics, The University of Osaka. His research interests include game-theoretic probability theory, quantum information theory, statistics, information geometry, and related topics.

Akio Fujiwara received the B.S. degree in applied physics and the M.S. and Ph.D. degrees in mathematical engineering from The University of Tokyo.

Currently, he is a Professor of mathematics with The University of Osaka. His research interests include noncommutative statistics, quantum information theory, information geometry, and algorithmic randomness.