The Birthday Problem and Zero-Error List Codes

Parham Noorzad, Michelle Effros, Michael Langberg, and Victoria Kostina

Abstract

As an attempt to bridge the gap between classical information theory and the combinatorial world of zero-error information theory, this paper studies the performance of randomly generated codebooks over discrete memoryless channels under a zero-error constraint. This study allows the application of tools from one area to the other. Furthermore, it leads to an information-theoretic formulation of the birthday problem, which is concerned with the probability that in a given population, a fixed number of people have the same birthday. Due to the lack of a closed-form expression for this probability when the distribution of birthdays is not uniform, the resulting computation is not feasible in some applications; the information-theoretic formulation, however, can be analyzed for all distributions.

I. Introduction

Finding channel capacity under a zero-error constraint requires fundamentally different tools and ideas from the study of capacity under an asymptotically negligible error constraint; the former is essentially a graph-theoretic problem [1], while the latter mainly relies on probabilistic arguments. To obtain a better understanding of the contrast between zero-error information theory and classical information theory, we apply probabilistic tools to the study of zero-error channel coding.

The random code construction of Shannon [2] shows that for the discrete memoryless channel $(\mathcal{X}, W(y|x), \mathcal{Y})$, a sequence of rate-$R$ codebooks $(\mathscr{C}_n)_{n=1}^{\infty}$, randomly generated according to distribution $P(x)$, achieves

$$\lim_{n\to\infty} \mathbb{E}\big[P_e^{(n)}(\mathscr{C}_n)\big] = 0, \tag{1}$$

if $R < I(X;Y)$. In (1), $P_e^{(n)}(\mathscr{C}_n)$ is the average probability of error of codebook $\mathscr{C}_n$. From Markov's inequality, it follows that $R < I(X;Y)$ suffices to ensure that

$$\forall\, \epsilon \in (0,1)\colon \lim_{n\to\infty} \Pr\big\{P_e^{(n)}(\mathscr{C}_n) \le \epsilon\big\} = 1. \tag{2}$$

Our aim is to understand the behavior of randomly generated codebooks when $\epsilon = 0$ in (2). Specifically, we seek to find necessary and sufficient conditions on R in terms of the channel W and input distribution $P(x)$, such that the sequence of randomly generated codebooks $(\mathscr{C}_n)_{n=1}^{\infty}$ satisfies

$$\lim_{n\to\infty} \Pr\big\{P_e^{(n)}(\mathscr{C}_n) = 0\big\} = 1.$$

This material is based upon work supported by the National Science Foundation under Grant Numbers 1321129, 1527524, and 1526771.

P. Noorzad, M. Effros, and V. Kostina are with the California Institute of Technology, Pasadena, CA 91125 USA (emails: parham@caltech.edu, effros@caltech.edu, vkostina@caltech.edu).

M. Langberg is with the State University of New York at Buffalo, Buffalo, NY 14260 USA (email: mikel@buffalo.edu).

In other words, our goal is to quantify the performance of randomly generated codebooks under the zero-error constraint. The study of this problem leads to an information-theoretic formulation of the birthday problem, which we next describe.

The classical birthday problem studies the probability that a fixed number of individuals in a population have the same birthday under the assumption that the birthdays are independent and identically distributed (i.i.d.). While this probability is simple to calculate when the i.i.d. distribution is uniform, it is computationally difficult in the nonuniform case. Specifically, in the nonuniform case, the mentioned probability is given by a sum that grows exponentially in the number of possible birthdays. Methods that approximate this sum under various simplifying assumptions are given in [3]–[5].

Here we study the mentioned probability in an information-theoretic setting. Consider the problem of channel coding over an identity channel, which is a channel over a finite alphabet where the output equals the input. A given code over such a channel is a "zero-error L-list code" if and only if no group of $L+1$ messages are mapped to the same codeword. Associating codewords with birthdays, we obtain an information-theoretic formulation of the birthday problem: Given a randomly generated codebook (set of birthdays), what is the probability that $L+1$ codewords (birthdays) are identical? We study this problem in Section IV.

While the birthday problem corresponds to zero-error list coding over the identity channel, in Section III we consider the general setting of zero-error list coding over an arbitrary discrete memoryless channel. Similar to the zero-error list capacity problem [6], the problem we study here can be solved using only knowledge of the distinguishability hypergraph of the channel. We discuss hypergraphs and their application to zero-error list codes in the next section.

II. Hypergraphs and Zero-Error List Codes

A discrete channel is a triple

$$\big(\mathcal{X}, W(y|x), \mathcal{Y}\big),$$

where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, and for each $x \in \mathcal{X}$, $W(\cdot|x)$ is a probability mass function on $\mathcal{Y}$.

A hypergraph $G = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E} \subseteq 2^{\mathcal{V}}$, where $2^{\mathcal{V}}$ denotes the collection of subsets of $\mathcal{V}$. We assume that each edge has cardinality at least two.

The "distinguishability hypergraph" of channel W, denoted by $G(W)$, is a hypergraph with vertex set $\mathcal{X}$ and an edge set $\mathcal{E} \subseteq 2^{\mathcal{X}}$ which contains collections of inputs that are "distinguishable" at the decoder. Formally, $\mathcal{E}$ consists of all subsets $e \subseteq \mathcal{X}$ that satisfy

$$\forall\, y \in \mathcal{Y}\colon \prod_{x \in e} W(y|x) = 0; \tag{3}$$

that is, $e \subseteq \mathcal{X}$ is an edge if no $y \in \mathcal{Y}$ is reachable from all $x \in e$. Note that $G(W)$ has the property that the superset of any edge is an edge; that is, if $e \in \mathcal{E}$ and $e \subseteq e' \subseteq \mathcal{X}$, then $e' \in \mathcal{E}$.

An "independent set" of a hypergraph $G = (\mathcal{V}, \mathcal{E})$ is a subset $\mathcal{I} \subseteq \mathcal{V}$ such that no subset of $\mathcal{I}$ is in $\mathcal{E}$. In zero-error coding over the channel $(\mathcal{X}, W(y|x), \mathcal{Y})$, an independent set corresponds to a collection of inputs $\mathcal{I} \subseteq \mathcal{X}$ for which there exists an output $y \in \mathcal{Y}$ that is reachable from any input in the collection. The hypergraph G is "complete

multipartite" if there exists a partition $\{\mathcal{I}_j\}_{j=1}^k$ of $\mathcal{V}$ such that each $\mathcal{I}_j$ is an independent set, and for every subset $e \subseteq \mathcal{V}$, either $e \in \mathcal{E}$, or $e \subseteq \mathcal{I}_j$ for some $1 \le j \le k$.

As an example, consider a deterministic channel $(\mathcal{X}, W(y|x), \mathcal{Y})$, where for some mapping $\varphi : \mathcal{X} \to \mathcal{Y}$,

$$W(y|x) = \mathbf{1}\{y = \varphi(x)\}.$$

For this channel, $G(W)$ is a complete multipartite hypergraph. Specifically, the sets $\{\varphi^{-1}(y)\}_{y\in\mathcal{Y}}$ are the independent components of G, where for $y \in \mathcal{Y}$,

$$\varphi^{-1}(y) := \big\{x \in \mathcal{X} \big| \varphi(x) = y\big\}.$$

For positive integers i and j with $j \ge i$, $[i : j]$ denotes the set $\{i, \ldots, j\}$. When $i = 1$, we denote $[1 : j]$ by $[j]$. For example, $[1] = \{1\}$ and $[2 : 4] = \{2, 3, 4\}$. For any set $\mathcal{A}$ and nonnegative integer $k \le |\mathcal{A}|$, define the set

$$\binom{\mathcal{A}}{k} := \big\{\mathcal{B} \big| \mathcal{B} \subseteq \mathcal{A}, |\mathcal{B}| = k\big\}.$$

An (M, L) list code for the channel W [7] consists of an encoder

$$f : [M] \to \mathcal{X},$$

and a decoder

$$g : \mathcal{Y} \to \bigcup_{\ell=1}^{L} \binom{[M]}{\ell}.$$

The pair (f, g) is an (M, L) *zero-error* list code for channel W if for every $m \in [M]$ and $y \in \mathcal{Y}$ satisfying $W(y|f(m)) > 0$, we have $m \in g(y)$.

Proposition 1 provides a necessary and sufficient condition for the existence of an (M, L) zero-error list code for W in terms of its distinguishability hypergraph $G(W)$. The proof, which follows directly from the definitions, appears in Subsection V-A.

Proposition 1. *An (M, L) zero-error list code for W exists if and only if there exists an encoder*

$$f : [M] \to \mathcal{X},$$

such that the image of every $(L + 1)$-subset $\{m_\ell\}_{\ell=1}^{L+1}$ of $[M]$ under f is an edge of $G(W)$.

Henceforth, we say a mapping $f : [M] \to \mathcal{X}$ is an (M, L) zero-error list code for W if it satisfies the condition stated in Proposition 1.

For each positive integer n, the nth extension channel of W is the channel

$$\big(\mathcal{X}^n, W^n(y^n|x^n), \mathcal{Y}^n\big),$$

where

$$W^n(y^n|x^n) := \prod_{i=1}^{n} W(y_i|x_i).$$

An (M, L) zero-error list code for W^n is referred to as an (M, n, L) zero-error list code for W. The distinguishability hypergraph of W^n, $G(W^n)$, equals $G^n(W)$, the nth "co-normal power" of $G(W)$ [8]. For any positive integer n

and any hypergraph $G = (\mathcal{V}, \mathcal{E})$, the nth co-normal power of G, denoted by G^n, is defined on the set of nodes $\mathcal{V}^n$ as follows. For each $k \geq 2$, the k-subset $e = \{v_1^n, \ldots, v_k^n\} \subseteq \mathcal{V}^n$ is an edge of G^n if for at least one $t \in [n]$, $\{v_{1t}, \ldots, v_{kt}\} \in \mathcal{E}$. This definition is motivated by the fact that k codewords are distinguishable if and only if their components are distinguishable in at least one dimension.

III. Main Result

Fix a sequence of probability distributions $(P_n(x^n))_{n=1}^{\infty}$, where P_n is defined over $\mathcal{X}^n$. Our aim in this work is to study the performance of the sequence of random codes $F_n : [M_n] \to \mathcal{X}^n$ over channel W, where

$$F_n(1), \ldots, F_n(M_n)$$

are M_n i.i.d. random variables, where

$$\forall\, m \in [M_n] \colon \Pr\left\{F_n(m) = x^n\right\} := P_n(x^n).$$

We seek to find conditions on the sequence $(M_n)_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \Pr\left\{F_n \text{ is an } (M_n, n, L) \text{ zero-error list code for } W\right\} = 1.$$

Theorem 3, below, provides the desired conditions. The conditions rely on a collection of functions of the pair $(G^n(W), P_n)$, denoted by

$$(\theta_{L+1}^{(\ell)}(G^n(W), P_n))_{\ell=1}^{L+1},$$

which we next define.

Fix a hypergraph $G = (\mathcal{V}, \mathcal{E})$. For any positive integer k, let $v_{[k]} = (v_1, \ldots, v_k)$ denote an element of $\mathcal{V}^k$. For all $v_{[k]} \in \mathcal{V}^k$ and every nonempty subset $S \subseteq [k]$, define $v_S := (v_j)_{j\in S}$. Let P be a probability mass function on $\mathcal{V}$ and set

$$P(v_S) := \prod_{j\in S} P(v_j).$$

In addition, for each positive integer $k \geq 2$, define the mapping $\sigma_k : \mathcal{V}^k \to 2^{\mathcal{V}}$ as

$$\sigma_k(v_{[k]}) := \{v_1, \ldots, v_k\}.$$

In words, σ_k maps each vector $v_{[k]} \in \mathcal{V}^k$ to the set containing its distinct components. For example, if $v_{[k]} = (v, \ldots, v)$ for some $v \in \mathcal{V}$, then $\sigma_k(v_{[k]}) = \{v\}$. When the value of k is clear from context, we denote σ_k with σ.

We next define functions of the pair (G, P) that are instrumental in characterizing the performance of random codebooks over channels with zero-error constraints. For $k \geq 2$, define the quantity $I_k(G, P)$ as

$$I_k(G, P) := -\frac{1}{k-1} \log \sum_{v_{[k]} : \sigma(v_{[k]}) \notin \mathcal{E}} P(v_{[k]}). \tag{4}$$

Note that in (4),

$$\sum_{v_{[k]} : \sigma(v_{[k]}) \notin \mathcal{E}} P(v_{[k]})$$

equals the probability of selecting k vertices in G, with replacement, that are indistinguishable. The negative sign in (4) results in the nonnegativity of $I_k(G,P)$; division by $k-1$ makes it comparable to the Rényi entropy of order k [9], which is defined as

$$H_k(P) := -\frac{1}{k-1}\log\sum_x \big(P(x)\big)^k.$$

We now define the sequence of functions $(\theta_k^{(j)}(G,P))_{j\in[k]}$. This sequence arises from the application of a second moment bound in the proof of Theorem 3 given in Subsection V-C. Set $\theta_k^{(k)}(G,P) := I_k(G,P)$, and for $j \in [k-1]$, let

$$\theta_k^{(j)}(G,P) := 2I_k(G,P) + \frac{1}{k-1}\log\sum_{v_{[j]}} P(v_{[j]})\left[\sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\right]^2. \tag{5}$$

The following proposition describes the properties the sequence $(\theta_k^{(j)}(G,P))_{j\in[k]}$ satisfies. Its proof appears in Subsection V-B.

Proposition 2. *For every hypergraph $G=(\mathcal{V},\mathcal{E})$, probability mass function P on $\mathcal{V}$, and positive integer $k\geq 2$, the following statements hold.*

(i) For all $j\in[k]$,

$$0 \leq \theta_k^{(j)}(G,P) \leq I_k(G,P).$$

(ii) We have

$$0 \leq I_k(G,P) \leq H_k(P).$$

Let $\operatorname{supp}(P)$ *denote the support of P. Then*

$$I_k(G,P)=0 \iff \forall\, e\subseteq \operatorname{supp}(P)\colon \big(2\leq|e|\leq k \implies e\notin\mathcal{E}\big)$$
$$I_k(G,P)=H_k(P) \iff \forall\, e\subseteq \operatorname{supp}(P)\colon \big(2\leq|e|\leq k \implies e\in\mathcal{E}\big).$$

(iii) For every positive integer $n\geq 2$, define the probability mass function P^n on $\mathcal{V}^n$ as

$$\forall\, v^n\in\mathcal{V}^n\colon P^n(v^n) := \prod_{i=1}^n P(v_i).$$

Then for all $j\in[k]$,

$$\theta_k^{(j)}(G^n,P^n) = n\theta_k^{(j)}(G,P).$$

When $k=2$, $I_k(G,P)$ has further properties which we discuss in the Appendix.

We next state our main result which provides upper and lower bounds on the cardinality of a randomly generated codebook that has zero error.

Theorem 3. *Consider a channel $(\mathcal{X},W(y|x),\mathcal{Y})$ and a sequence of distributions $(P_n(x^n))_{n=1}^\infty$. If*

$$\lim_{n\to\infty} M_n^{L+1}2^{-LI_{L+1}(G^n(W),P_n)} = 0, \tag{6}$$

then

$$\lim_{n\to\infty}\Pr\big\{F_n \text{ is an } (M_n,n,L) \text{ zero-error list code}\big\} = 1. \tag{7}$$

Conversely, assuming (7), then for some $\ell \in [L+1]$,

$$\lim_{n\to\infty} M_n^{\ell} 2^{-L\theta_{L+1}^{(\ell)}(G^n(W),P_n)} = 0. \tag{8}$$

In Theorem 3, if a channel W and a sequence of distributions $(P_n)_{n=1}^{\infty}$ satisfy

$$\max_{\ell\in[L+1]} \frac{1}{\ell}\theta_{L+1}^{(\ell)}(G^n(W),P_n) = \frac{1}{L+1} I_{L+1}(G^n(W),P_n), \tag{9}$$

for every $n \geq 1$, then (6), in addition to being sufficient for (7), is necessary as well. In the next corollary, we give two scenarios under which (9) holds. One case is when for all n, P_n is the uniform distribution on a "clique" $\mathcal{C}_n$ of $G^n(W)$. For any hypergraph $G = (\mathcal{V}, \mathcal{E})$, a subset $\mathcal{C} \subseteq \mathcal{V}$ is a clique of G if every subset of $\mathcal{C}$ with cardinality greater than or equal to two is in $\mathcal{E}$. In zero-error coding, a clique corresponds to a collection of inputs for which every output is reachable from at most one input in the collection.

Corollary 4. *Consider a channel $(\mathcal{X}, W(y|x), \mathcal{Y})$ and a sequence of distributions $(P_n(x^n))_{n=1}^{\infty}$. We have*

$$\lim_{n\to\infty} \Pr\left\{F_n \text{ is an } (M_n, n, L) \text{ zero-error list code}\right\} = 1$$
$$\iff \lim_{n\to\infty} M_n^{L+1} 2^{-LI_{L+1}(G^n(W),P_n)} = 0,$$

if either of the conditions below hold:

(1) The hypergraph $G(W)$ is complete multipartite.

(2) For each n, P_n is the uniform distribution on $\mathcal{C}_n$, where $\mathcal{C}_n$ is a clique of $G^n(W)$.

In Theorem 3, fix an input distribution P and a rate $R \geq 0$. Then setting $P_n := P^n$ and $M_n := \lfloor 2^{nR} \rfloor$ for all positive integers n, and applying parts (i) and (iii) of Proposition 2 leads to the following corollary.

Corollary 5. *Consider a channel $(\mathcal{X}, W(y|x), \mathcal{Y})$ and an input distribution P. If*

$$R < \frac{L}{L+1} I_{L+1}(G, P),$$

then

$$\lim_{n\to\infty} \Pr\left\{F_n \text{ is an } (2^{nR}, n, L) \text{ zero-error list code}\right\} = 1. \tag{10}$$

Conversely, assuming (10), then

$$R < LI_{L+1}(G, P).$$

IV. The Birthday Problem

In this section, we study an information-theoretic version of the birthday problem. Assuming an i.i.d. distribution on the birthdays of a given population, the birthday problem studies the probability that a fixed number of people have the same birthday. Stated abstractly, the birthday problem studies the probability that a randomly generated mapping takes the same value for a fixed number of inputs. Based on this viewpoint, we present an information-theoretic formulation of this problem.

Fix an integer $k \geq 2$, a finite set $\mathcal{X}$, and a sequence of probability mass functions $(P_n(x^n))_{n=1}^{\infty}$. For each n, let $F_n : [M_n] \to \mathcal{X}^n$ be a random mapping with i.i.d. values and distribution $P_n(x^n)$; that is,

$$\forall\, m \in [M_n]\colon \Pr\big\{F_n(m) = x^n\big\} = P_n(x^n).$$

Our aim is to find conditions on $(M_n)_{n=1}^{\infty}$ so that

$$\lim_{n\to\infty} \Pr\big\{\exists\, m_1, \ldots, m_k \colon F_n(m_1) = \cdots = F_n(m_k)\big\} = 0. \tag{11}$$

Let $W = (\mathcal{X}, \mathbf{1}\{y = x\}, \mathcal{X})$ denote the identity channel on $\mathcal{X}$. Note that every subset e of $\mathcal{X}$ with $|e| \geq 2$ is an edge of $G(W)$. Thus for $n \geq 2$, every $e \subseteq \mathcal{X}^n$ with $|e| \geq 2$ is an edge of $G^n(W)$. Therefore, for distinct messages $m_1, \ldots, m_k \in [M_n]$, we have $F_n(m_1) = \cdots = F_n(m_k)$ if and only if

$$\big(F_n(m_1), \ldots, F_n(m_k)\big) \text{ is not an edge in } G^n(W).$$

Hence Proposition 1 implies that (11) holds if and only if

$$\lim_{n\to\infty} \Pr\big\{F_n \text{ is an } (M_n, n, k-1) \text{ zero-error list code}\big\} = 1. \tag{12}$$

Now from Corollary 4, it follows that (12) holds if and only if

$$\lim_{n\to\infty} M_n^k 2^{-(k-1)H_k(P_n)} = 0.$$

In words, to guarantee the absence of collisions of kth order, the population size M_n must be negligible compared to $2^{-\frac{k-1}{k}H_k(P_n)}$.

V. Proofs

A. *Proof of Proposition 1*

Let (f, g) be an (M, L) zero-error list code for W. If $[M]$ has a subset of cardinality $L+1$, say $\{m_\ell\}_{\ell=1}^{L+1}$, such that $\{f(m_\ell)\}_{\ell=1}^{L+1}$ is not an edge in G, then for some $y \in \mathcal{Y}$,

$$\prod_{\ell \in [L+1]} W(y|f(m_\ell)) > 0.$$

Thus for every $\ell \in [L+1]$, $W(y|f(m_\ell)) > 0$, which implies $m_\ell \in g(y)$. Thus $g(y)$ contains at least $L+1$ distinct elements, which is a contradiction.

Conversely, suppose we have an encoder $f : [M] \to \mathcal{X}$ such that for each $y \in \mathcal{Y}$, the set

$$\mathcal{M}_y \coloneqq \Big\{m \in [M] \Big| W(y|f(m)) > 0\Big\}$$

has cardinality at most L. Then if we define the decoder as

$$\forall\, y \in \mathcal{Y}\colon g(y) = \mathcal{M}_y,$$

then the pair (f, g) is an (M, L) zero-error list code.

B. Proof of Proposition 2

(i) We prove the nonnegativity of $\theta_k^{(j)}(G,P)$ first for $j=k$ and then for arbitrary $j\in[k-1]$. Recall that $\theta_k^{(k)}(G,P)=I_k(G,P)$. We have

$$\sum_{v_{[k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[k]}) \le \sum_{v_{[k]}} P(v_{[k]}) = \Big(\sum_{v} P(v)\Big)^k = 1,$$

which implies

$$\theta_k^{(k)}(G,P) = I_k(G,P) = -\frac{1}{k-1}\log \sum_{v_{[k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[k]}) \ge 0.$$

For $j\in[k-1]$, we rewrite $\theta_k^{(j)}(G,P)$ as

$$\theta_k^{(j)}(G,P) = \frac{1}{k-1}\log \frac{\sum_{v_{[j]}} P(v_{[j]})\Big[\sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\Big]^2}{\Big[\sum_{v_{[k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[k]})\Big]^2}.$$

Note that

$$\sum_{v_{[j]}} P(v_{[j]}) = \Big(\sum_{v} P(v)\Big)^j = 1,$$

and thus by the Cauchy-Schwarz inequality,

$$\begin{aligned}\sum_{v_{[j]}} P(v_{[j]})\Big[\sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\Big]^2 &\ge \Big[\sum_{v_{[j]}} P(v_{[j]}) \sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\Big]^2 \\ &\ge \Big[\sum_{v_{[k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[k]})\Big]^2,\end{aligned}$$

which implies $\theta^{(j)}(G,P)\ge 0$.

We next prove the upper bound on $\theta^{(j)}(G,P)$. Note that

$$\begin{aligned}\sum_{v_{[j]}} P(v_{[j]})\Big[\sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\Big]^2 &\le \sum_{v_{[j]}} P(v_{[j]})\Big[\sum_{v_{[j+1:k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[j+1:k]})\Big] \\ &= 2^{-(k-1)I_k(G,P)}.\end{aligned}$$

Thus

$$\theta_k^{(j)}(G,P) \le 2I_k(G,P) - I_k(G,P) = I_k(G,P).$$

(ii) The inequality $I_k(G,P)\ge 0$ is proved in (i). Equality holds if and only if

$$\forall\, v_{[k]}\in \big(\mathrm{supp}(P)\big)^k\colon \sigma(v_{[k]})\notin\mathcal{E},$$

which is equivalent to

$$\forall\, e\subseteq \mathrm{supp}(P)\colon \big(2\le|e|\le k \implies e\notin\mathcal{E}\big).$$

We next prove the upper bound on $I_k(G,P)$. Since each edge of G has cardinality at least two, for all $v\in\mathcal{V}$, $\{v\}\notin\mathcal{E}$. Thus

$$\sum_{v_{[k]}:\sigma(v_{[k]})\notin\mathcal{E}} P(v_{[k]}) \ge \sum_{v} \big(P(v)\big)^k,$$

which implies

$$I_k(G,P) \le H_k(P),$$

where $H_k(P)$ is the Rényi entropy of order k. Equality holds if and only if

$$\forall\, v_{[k]} \in \big(\mathrm{supp}(P)\big)^k \colon \big(v_1 = \cdots = v_k\big) \vee \big(\sigma(v_{[k]}) \in \mathcal{E}\big),$$

which is equivalent to

$$\forall\, e \subseteq \mathrm{supp}(P) \colon \big(2 \le |e| \le k \implies e \in \mathcal{E}\big).$$

(iii) Fix a positive integer n. Let $\mathcal{E}_n$ denote the set of edges of G^n. Let $v_{[k]}^n$ denote the vector

$$v_{[k]}^n := \big(v_1^n, \ldots, v_k^n\big),$$

and $\sigma_k(v_{[k]}^n)$ denote the set

$$\sigma_k(v_{[k]}^n) := \big\{v_1^n, \ldots, v_k^n\big\}.$$

Furthermore, let $\mathcal{S} \subseteq \mathcal{V}^k$ denote the set

$$\mathcal{S} := \big\{v_{[k]} \big| \sigma_k(v_{[k]}) \notin \mathcal{E}\big\}.$$

Note that for each $v_{[k]}^n$, $\sigma_k(v_{[k]}^n) \notin \mathcal{E}_n$ if and only if

$$\forall\, t \in [n] \colon \big\{v_{1t}, \ldots, v_{kt}\big\} \notin \mathcal{E}.$$

Thus

$$\big\{v_{[k]}^n \big| \sigma_k(v_{[k]}^n) \notin \mathcal{E}_n\big\} = \mathcal{S}^n,$$

which implies

$$\begin{aligned}
\sum_{v_{[k]}^n : \sigma_k(v_{[k]}^n) \notin \mathcal{E}_n} P^n(v_{[k]}^n) &= \sum_{v_{[k]}^n \in \mathcal{S}^n} P^n(v_{[k]}^n) \\
&= \sum_{v_{[k]}^n \in \mathcal{S}^n} \prod_{t \in [n]} P(v_{[k]t}) \\
&= \prod_{t \in [n]} \sum_{v_{[k]t} \in S} P(v_{[k]t}) \\
&= \Big(\sum_{v_{[k]} \in \mathcal{S}} P(v_{[k]}) \Big)^n.
\end{aligned}$$

Therefore,

$$I_k(G^n, P^n) = n I_k(G,P).$$

For $j \in [k-1]$, we can write $\theta_k^{(j)}(G^n, P^n)$ as

$$\theta_k^{(j)}(G^n, P^n) = 2I_k(G^n, P^n) + \frac{1}{k-1} \log \sum_{\substack{(v_{[j]}^n, v_{[j+1:k]}^n, \bar{v}_{[j+1:k]}^n) : \\ (v_{[j]}^n, v_{[j+1:k]}^n) \in \mathcal{S}^n \\ (v_{[j]}^n, \bar{v}_{[j+1:k]}^n) \in \mathcal{S}^n}} P^n\big(v_{[j]}^n\big) P^n\big(v_{[j+1:k]}^n\big) P^n\big(\bar{v}_{[j+1:k]}^n\big).$$

Using a similar argument as above, it follows that for all $j \in [k-1]$,

$$\theta_k^{(j)}(G^n, P^n) = n \theta_k^{(j)}(G,P).$$

C. Proof of Theorem 3

We start by finding upper and lower bounds on the probability that a random mapping from $[M]$ to $\mathcal{X}$ is an (M, L) zero-error list code for the channel W. The theorem then follows from applying our bounds to the channel W^n for every positive integer n.

Consider the random mapping $F : [M] \to \mathcal{X}$, where $(F(m))_{m\in[M]}$ is a collection of i.i.d. random variables and each $F(m)$ has distribution

$$\Pr\big\{F(m) = x\big\} := P(x).$$

For every $S \in \binom{[M]}{L+1}$, define

$$Z_S := \mathbf{1}\Big\{\{F(m)\}_{m\in S} \notin \mathcal{E}\Big\};$$

that is, Z_S is the indicator of the event that $\{F(m)\}_{m\in S}$ is not an edge of the distinguishability hypergraph $G(W)$. Let[1]

$$Z := \sum_{S\in\binom{[M]}{L+1}} Z_S.$$

Note that by Proposition 1, F is an (M, L) zero-error list code if and only if $Z = 0$. The rest of the proof consists of computing a lower and an upper bound for $\Pr\{Z = 0\}$.

Lower Bound. By Markov's inequality,

$$\begin{aligned}
&\Pr\big\{F \text{ is an } (M, L) \text{ zero-error list code}\big\}\\
&\quad = \Pr\{Z = 0\}\\
&\quad = 1 - \Pr\big\{Z \geq 1\big\}\\
&\quad \geq 1 - \mathbb{E}[Z].
\end{aligned}$$

For any $S \in \binom{[M]}{L+1}$,

$$\mathbb{E}[Z_S] = \sum_{x_{[L+1]}:\sigma(x_{[L+1]})\notin\mathcal{E}} P(x_{[L+1]}) = 2^{-LI_{L+1}(G,P)}. \tag{13}$$

By linearity of expectation,

$$\mathbb{E}[Z] = \binom{M}{L+1} 2^{-LI_{L+1}(G,P)}. \tag{14}$$

Thus

$$\Pr\big\{Z = 0\big\} \geq 1 - \binom{M}{L+1} 2^{-LI_{L+1}(G,P)}.$$

Upper Bound. We apply the second moment method. By the Cauchy-Schwarz inequality,

$$\mathbb{E}[Z] = \mathbb{E}\big[Z\mathbf{1}_{\{Z\geq 1\}}\big] \leq \sqrt{\mathbb{E}[Z^2] \times \Pr\{Z \geq 1\}},$$

thus

$$\Pr\{Z \geq 1\} \geq \frac{\big(\mathbb{E}[Z]\big)^2}{\mathbb{E}[Z^2]}$$

[1]For results regarding the distribution of Z in the classical birthday problem scenario, we refer the reader to the work of Arratia, Goldstein, and Gordon [10], [11]. A direct application of the bounds in [10], [11] to $\Pr\{Z = 0\}$ leads to weaker results than those we present here.

or

$$\Pr\{Z=0\} \le 1 - \frac{\big(\mathbb{E}[Z]\big)^2}{\mathbb{E}[Z^2]}.$$

To evaluate the upper bound on $\Pr\{Z=0\}$, we calculate $\mathbb{E}[Z^2]$. We have

$$\begin{aligned} Z^2 &= \left[\sum_{S\in\binom{[M]}{L+1}} Z_S\right]^2 \\ &= \sum_{S\in\binom{[M]}{L+1}} Z_S^2 + \sum_{\substack{S,S'\in\binom{[M]}{L+1}\\ S\neq S'}} Z_S Z_{S'} \\ &= \sum_{S\in\binom{[M]}{L+1}} Z_S + \sum_{\ell=0}^{L} \sum_{\substack{S,S'\in\binom{[M]}{L+1}\\ |S\cap S'|=\ell}} Z_S Z_{S'}. \end{aligned}$$

For all $\ell \in \{0,1,\dots,L\}$, fix sets $S_\ell, S'_\ell \in \binom{[M]}{L+1}$ such that $|S_\ell \cap S'_\ell| = \ell$. When $\ell \in [L]$, $(F(m))_{m\in S_\ell}$ and $(F(m))_{m\in S'_\ell}$ are independent given $(F(m))_{m\in S_\ell\cap S'_\ell}$. Thus for $\ell\in[L]$,

$$\begin{aligned} \mathbb{E}[Z_{S_\ell} Z_{S'_\ell}] &= \sum_{x_{[\ell]}} P(x_{[\ell]}) \left[\sum_{\substack{x_{[\ell+1:L+1]}:\\ \sigma(x_{[L+1]})\notin\mathcal{E}}} P(x_{[\ell+1:L+1]})\right]^2 \\ &= 2^{L(\theta^{(\ell)}_{L+1}(G,P) - 2I_{L+1}(G,P))}, \end{aligned} \tag{15}$$

where in (15), we use the definition (5). Note that when $\ell = 0$, Z_{S_0} and $Z_{S'_0}$ are independent. Thus by (13),

$$\mathbb{E}[Z_{S_0} Z_{S'_0}] = \big(\mathbb{E}[Z_{S_0}]\big)^2 = 2^{-2LI_{L+1}(G,P)}.$$

Therefore,

$$\begin{aligned} \mathbb{E}[Z^2] &= \binom{M}{L+1} 2^{-LI_{L+1}(G,P)} \\ &+ \binom{M}{0, L+1, L+1} 2^{-2LI_{L+1}(G,P)} \\ &+ \sum_{\ell=1}^{L} \binom{M}{\ell, L+1-\ell, L+1-\ell} 2^{L(\theta^{(\ell)}_{L+1}(G,P) - 2I_{L+1}(G,P))}, \end{aligned} \tag{16}$$

where in (16), for $\ell\in\{0,1,\dots,L\}$,

$$\begin{aligned} &\binom{M}{\ell, L+1-\ell, L+1-\ell} \\ &\quad := \binom{M}{\ell}\binom{M-\ell}{L+1-\ell}\binom{M-L-1}{L+1-\ell}, \end{aligned}$$

equals the number of pairs (S, S'), where $S, S' \in \binom{[M]}{L+1}$ and $|S \cap S'| = \ell$. Combining (14) with (16) now gives

$$\begin{aligned}
\frac{\mathbb{E}[Z^2]}{\big(\mathbb{E}[Z]\big)^2} &= \binom{M}{L+1}^{-1} 2^{LI_{L+1}(G,P)} \\
&+ \binom{M}{L+1}^{-2} \binom{M}{0, L+1, L+1} \\
&+ \binom{M}{L+1}^{-2} \sum_{\ell=1}^{L} \binom{M}{\ell, L+1-\ell, L+1-\ell} 2^{L\theta^{(\ell)}_{L+1}(G,P)} \\
&\le \binom{M}{L+1}^{-1} 2^{LI_{L+1}(G,P)} + 1 \\
&\quad + \sum_{\ell=1}^{L} \binom{L+1}{\ell}^2 \binom{M}{\ell}^{-1} 2^{L\theta^{(\ell)}_{L+1}(G,P)},
\end{aligned}$$

where in the last inequality, we apply the fact that for all $\ell \in \{0, 1, \ldots, L\}$,

$$\begin{aligned}
\binom{M}{L+1}^{-2} \binom{M}{\ell, L+1-\ell, L+1-\ell} &= \binom{L+1}{\ell}^2 \binom{M}{\ell}^{-1} \times \frac{(M-L-1)!(M-L-1)!}{(M-\ell)!(M-2L-2+\ell)!} \\
&= \binom{L+1}{\ell}^2 \binom{M}{\ell}^{-1} \prod_{j=\ell}^{L} \Big(\frac{M-L-1+\ell-j}{M-j}\Big) \\
&\le \binom{L+1}{\ell}^2 \binom{M}{\ell}^{-1}.
\end{aligned}$$

This completes the proof of the upper bound. The asymptotic result (8) follows from the fact that for all $\ell \in [M]$,

$$\binom{M}{\ell} \ge \Big(\frac{M}{\ell}\Big)^{\ell}.$$

D. Proof of Corollary 4

For each of the conditions, we verify (9).

(1): We first show that if G is a complete multipartite hypergraph, and P is a distribution defined on the vertices of G, then

$$\max_{\ell \in [L+1]} \frac{1}{\ell} \theta^{(\ell)}_{L+1}(G, P) = \frac{1}{L+1} I_{L+1}(G, P). \tag{17}$$

Denote the maximal independent sets forming G by $\{\mathcal{I}_j\}_{j=1}^{k}$. Define the distribution P^* on $[k]$ as

$$P^*(j) := \sum_{x \in \mathcal{I}_j} P(x).$$

In words, $P^*(j)$ is the weight assigned to the independent set $\mathcal{I}_j$ by P. Since G is a complete multipartite hypergraph,

$\sigma(v_{[L+1]}) \notin \mathcal{E}$ if and only if there exists j such that $\sigma(v_{[L+1]}) \subseteq \mathcal{I}_j$. Thus

$$\begin{aligned} I_{L+1}(G,P) &= -\frac{1}{L}\log \sum_{v_{[L+1]}:\sigma(v_{[L+1]})\notin\mathcal{E}} P(v_{[L+1]}) \\ &= -\frac{1}{L}\log \sum_{j=1}^{k} \sum_{\sigma(v_{[L+1]})\subseteq\mathcal{I}_j} P(v_{[L+1]}) \\ &= -\frac{1}{L}\log \sum_{j=1}^{k} \big(P^*(j)\big)^{L+1} \\ &= H_{L+1}(P^*), \end{aligned}$$

where H_{L+1} denotes the Rényi entropy of order $L+1$. Similarly, for all $\ell \in [L+1]$ we have

$$\theta^{(\ell)}_{L+1}(G,P) = 2H_{L+1}(P^*) - \frac{2L+1-\ell}{L} H_{2L+2-\ell}(P^*). \tag{18}$$

Using (18), we see that proving (17) is equivalent to showing that for all $\ell \in [L+1]$,

$$\Big(\sum_{j=1}^{k} P^*(j)^{2L+2-\ell}\Big)^{\frac{1}{2L+2-\ell}} \le \Big(\sum_{j=1}^{k} P^*(j)^{L+1}\Big)^{\frac{1}{L+1}},$$

which follows from the well-known fact that for all $p \ge q$, the q-norm dominates the p-norm.

It now suffices to show that if G is complete multipartite, then so is G^n for all $n \ge 2$. To see this, first note that the set of vertices of G^n is given by

$$\bigcup_{j_1,\dots,j_n\in[k]} \mathcal{I}_{j_1} \times \cdots \times \mathcal{I}_{j_n}$$

We show that an arbitrary subset of $\mathcal{V}^n$, say $\{v_1^n, \dots, v_\ell^n\}$, is an edge in G^n if and only if

$$\forall\, j_1, \dots, j_n \in [k] \colon \{v_1^n, \dots, v_\ell^n\} \not\subseteq \mathcal{I}_{j_1} \times \cdots \times \mathcal{I}_{j_n}. \tag{19}$$

By definition, $\{v_1^n, \dots, v_\ell^n\}$ is an edge in G^n if and only if for some $t \in [n]$, $\{v_{1t}, \dots, v_{\ell t}\}$ is an edge in G. Since G is complete multipartite, the latter condition holds if and only if

$$\exists\, i \in [n] \text{ such that } \forall\, j \in [k] \colon \{v_{1i}, \dots, v_{\ell i}\} \not\subseteq \mathcal{I}_j,$$

which is equivalent to (19).

(2): In this case for every n, P_n is uniform on $\mathcal{C}_n$, a clique of $G^n(W)$. Here we prove (9) for $n = 1$; a similar argument proves the result for arbitrary n. Note that for $v_{[k]} \in \mathcal{V}^k$, $\sigma(v_{[k]}) \notin \mathcal{E}$ if and only if either for some $v \in \mathcal{C}_1$, $v_1 = \cdots = v_k = v$, or for some $j \in [k]$, $v_j \notin \mathcal{C}_1$. Thus for all $\ell \in [L+1]$, $\theta^{(\ell)}_{L+1}(G(W), P_1)$ simplifies to

$$\theta^{(\ell)}_{L+1}(G(W), P_1) = \frac{\ell - 1}{L} \log |\mathcal{C}_1|,$$

which proves (9) for $n = 1$.

VI. Conclusion

From Shannon's random coding argument [2] it follows that if the rate of a randomly generated codebook is less than the input-output mutual information, the probability that the codebook has small probability of error goes to one as the blocklength goes to infinity. In this work, we find necessary and sufficient conditions on the rate so that the probability that a randomly generated codebook has *zero* probability of error goes to one as the blocklength goes to infinity. We further show that this result extends the classical birthday problem to an information-theoretic setting and provides an intuitive meaning for Rényi entropy.

Appendix
Properties of $I_2(G,P)$

In this appendix, we describe two properties of $I_2(G,P)$. In the first part, we state the Motzkin-Straus theorem, which gives the maximum of $I_2(G,P)$ over all distributions P for a fixed graph G. In the second part, we show that $I_2(G,P)$ is always less than or equal to Körner's graph entropy.

A. The Motzkin-Straus Theorem

Consider a graph $G=(\mathcal{V},\mathcal{E})$. Motzkin and Straus [12] prove that

$$\max_P I_2(G,P) = \log \omega(G),$$

where $\omega(G)$ is the cardinality of the largest clique in G. An implication of this result is Turán's graph theorem [13], which states that

$$\omega(G) \geq \frac{|\mathcal{V}|^2}{|\mathcal{V}|^2 - |\mathcal{E}|}.$$

To see this, note that if we choose the distribution P to be uniform on the set of vertices of G, then

$$I_2(G,P) = -\log \sum_{v,v':\{v,v'\}\notin\mathcal{E}} \frac{1}{|\mathcal{V}|^2} = \log \frac{|\mathcal{V}|^2}{|\mathcal{V}|^2 - |\mathcal{E}|} \leq \log \omega(G),$$

where the inequality follows by the Motzkin-Straus theorem.

We remark that from Proposition 2 (ii) it follows that

$$\max_G I_2(G,P) = H_2(P),$$

which is achieved when G is the complete graph on the support of P.

B. Relation with Körner's Graph Entropy

Consider a graph G with vertex set $\mathcal{X}$. Let P be a probability distribution on $\mathcal{X}$ and $\mathcal{Y} \subseteq 2^{\mathcal{X}}$ be the set of all maximal independent subgraphs of G. Let $\Delta(G,P)$ denote the set of all probability distributions $P(x,y)$ on $\mathcal{X}\times\mathcal{Y}$ whose marginal on $\mathcal{X}$ equals $P(x)$, and

$$\Pr\{X \in Y\} = \sum_{(x,y):x\in y} P(x,y) = 1.$$

For the graph G and probability distribution P, Körner's graph entropy [14] is defined by

$$H_1(G,P) = \min_{\Delta(G,P)} I(X;Y). \tag{20}$$

Our aim is to define $H_2(G,P)$ in a similar manner to how $H_2(P)$, the Rényi entropy of order 2, is defined. One way to accomplish this task is through the use of Jensen's inequality. Applying Jensen's inequality to Shannon entropy gives

$$\begin{aligned} H_1(P) &= -\sum_x P(x)\log P(x) \\ &\geq -\log\Big(\sum_x \big(P(x)\big)^2\Big) = H_2(P). \end{aligned}$$

Analogously, applying Jensen's inequality to the mutual information in (20) gives

$$\begin{aligned} I(X;Y) &= -\sum_{(x,y):x\in y} P(x,y)\log\frac{P(x)P(y)}{P(x,y)} \\ &\geq -\log\Big(\sum_{(x,y):x\in y} P(x)P(y)\Big), \end{aligned}$$

Thus we define $H_2(G,P)$ as

$$\begin{aligned} H_2(G,P) &= \min_{\Delta(G,P)} -\log\Big(\sum_{(x,y):x\in y} P(x)P(y)\Big) \\ &\leq H_1(G,P). \end{aligned}$$

Our next proposition relates $H_2(G,P)$ and $I_2(G,P)$.

Proposition 6. *For any graph G and any probability distribution P defined on its vertices, $I_2(G,P) \leq H_2(G,P)$.*

Proof. Let $P(x,y) \in \Delta(G,P)$. Then

$$\sum_{x\in y} P(x)P(y) = \sum_{x,x'} P(x)P(x')\sum_y P(y|x')\mathbf{1}\big\{x,x'\in y\big\}. \tag{21}$$

Since every y is an independent subgraph of G, if $x,x'\in y$, then $(x,x')\notin\mathcal{E}$. Thus

$$\mathbf{1}\big\{x,x'\in y\big\} \leq \mathbf{1}\big\{(x,x')\notin\mathcal{E}\big\},$$

which implies

$$\sum_y P(y|x')\mathbf{1}\big\{x,x'\in y\big\} \leq \mathbf{1}\big\{(x,x')\notin\mathcal{E}\big\}.$$

By (21), we have

$$\sum_{x\in y} P(x)P(y) \leq \sum_{x,x'} P(x)P(x')\mathbf{1}\big\{(x,x')\notin\mathcal{E}\big\}.$$

Calculating the logarithm of both sides and maximizing the left hand side over $\Delta(G,P)$ gives $H_2(G,P) \geq I_2(G,P)$. □

ACKNOWLEDGMENTS

The first author thanks Ming Fai Wong for helpful discussions regarding an earlier version of Theorem 3.

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[2] ——, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[3] M. H. Gail, G. H. Weiss, N. Mantel, and S. J. O'Brien, "A solution to the generalized birthday problem with application to allozyme screening for cell culture contamination," *J. Appl. Prob.*, vol. 16, no. 2, pp. 242–251, 1979.

[4] T. S. Nunnikhoven, "A birthday problem solution for nonuniform birth frequencies," *Am. Stat.*, vol. 46, no. 4, pp. 270–274, 1992.

[5] C. Stein, "Application of Newton's identities to a generalized birthday problem and to the Poisson binomial distribution," Stanford University — Department of Statistics, Tech. Rep. 354, September 1990.

[6] J. Körner and K. Marton, "On the capacity of uniform hypergraphs," *IEEE Trans. Inf. Theory*, vol. 36, no. 1, pp. 153–156, 1990.

[7] P. Elias, "Error-correcting codes for list decoding," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 5–12, 1991.

[8] G. Simonyi, "Graph entropy: A survey," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 20, pp. 399–441, 1995.

[9] A. Rényi, "On measures of information and probability," in *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1. University of California Press, 1961, pp. 547–561.

[10] R. Arratia, L. Goldstein, and L. Gordon, "Two moments suffice for Poisson approximations: The Chen-Stein method," *Ann. Prob.*, vol. 17, no. 1, pp. 9–25, 1989.

[11] ——, "Poisson approximation and the Chen-Stein method," *Stat. Sci.*, vol. 5, no. 4, pp. 403–434, 1990.

[12] T. S. Motzkin and E. G. Straus, "Maxima for graphs and a new proof of a theorem of Turán," *Canad. J. Math.*, vol. 17, pp. 533–540, 1965.

[13] M. Aigner, "Turán's graph theorem," *Am. Math. Monthly*, vol. 102, no. 9, pp. 808–816, 1995.

[14] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Trans. 6th Prague Conf. Information Theory*. Academia, 1973, pp. 411–425.