

Communication Complexity of Exact Sampling Under Rényi Information

Spencer Hill¹, Student Member, IEEE, Fady Alajaji², Senior Member, IEEE, and Tamás Linder³, Fellow, IEEE

Abstract—We study the problem of exact sampling under an exponential communication cost, specifically Campbell’s average codeword length $L(t)$ of order t , and Rényi’s entropy. We provide a lower bound on the Campbell cost of exact sampling that grows approximately as $D_{1/\alpha}(P||Q)$, the Rényi divergence of order $1/\alpha$, with $\alpha = 1/1 + t$. Using the Poisson functional representation of Li and El Gamal (2017), we prove an upper bound on $L(t)$ whose leading Rényi divergence term has order within ϵ of that of the lower bound. Our results reduce to the bounds of Harsha et al. (2010) as $\alpha \rightarrow 1$. We also provide numerical examples comparing the bounds in the cases of normal and Laplacian distributions, demonstrating that the upper and lower bounds are typically within 5-10 bits of each other. Our results characterize exactly the optimal asymptotic Campbell cost $L(t)$ per sample as the number of independent and identically distributed (i.i.d.) samples grows to infinity. We show that under the exponential cost, any causal sampler performs strictly worse asymptotically than noncausal samplers. This contrasts with the case of expected message length, where both causal and noncausal samplers have the same optimal asymptotic cost.

Index Terms—Communication complexity, channel simulation, Rényi entropy, Rényi divergence, α -mutual information, Rényi conditional entropy, Poisson functional representation, noncausal sampling, common randomness, variable-length codes, exponential cost.

I. INTRODUCTION

WE CONSIDER the problem of communicating a sample from a probability distribution P , given that the sender and receiver have access to a shared source of randomness $\{U_i\}_{i \geq 1}$ drawn independently according to another probability distribution Q . This problem, called exact sampling, is visualized in Fig. 1. As studied in [2], [3], and [4], among others, the key question is how to design a communication protocol such that the expected message length needed to communicate the sample from P is minimized. Naively communicating a sample from P using lossless coding is often impossible, as

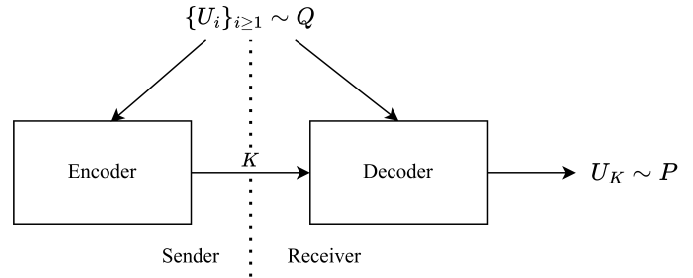


Fig. 1. Block diagram of the exact sampling problem.

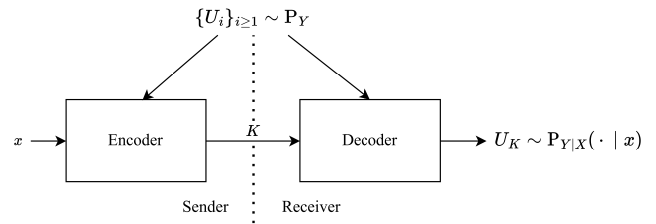


Fig. 2. Block diagram of the channel simulation problem using sampling.

in principle, the sample can have a continuous distribution. Instead, protocols aim to transmit an index $K \in \mathbb{N}$, where $\mathbb{N} = \{1, 2, \dots\}$ is the set of positive integers, such that the K th sample in the shared randomness has exact distribution P , which we write as $U_K \sim P$.

The goal is then to minimize the expected message length of K , or (almost) equivalently, to minimize $H(K)$, where $H(\cdot)$ denotes Shannon entropy. Formally, with $\mathcal{C} : \mathbb{N} \rightarrow \{0, 1\}^*$ denoting a uniquely decodable binary variable-length code and $l(\mathcal{C}(k))$ denoting the length of the codeword of k , the goal is to design a sampling algorithm and code \mathcal{C} such that $\mathbb{E}[l(\mathcal{C}(K))]$ is minimized. This is intimately related to the problem of channel simulation (visualized in Fig. 2), where X and Y are general random variables with joint distribution P_{XY} and we let $P = P_{Y|X}(\cdot | x)$ and $Q = P_Y$; i.e., for input x drawn via P_X , we wish to sample from the conditional distribution $P_{Y|X}(\cdot | x)$ given shared access to the marginal distribution P_Y .

Channel simulation, also called reverse channel coding or relative entropy coding, has recently gained significant theoretical and practical attention due to its promise as an alternative for quantization in deep learning-based compression systems [5], [6]. There has been a corresponding research effort to understand the theoretical limits of channel simulation and its related problems [7], [8].

Generalizing, one may also be concerned about costs other than the expected message length. Motivated by Campbell’s

Received 27 August 2025; revised 10 February 2026; accepted 30 March 2026. Date of publication 3 April 2026; date of current version 21 May 2026. This work was supported by the Natural Sciences and Engineering Research Council of Canada. An earlier version of this paper was presented in part at the 2025 IEEE International Symposium on Information Theory and in part at the 61st Allerton Conference on Communication, Control, and Computing. (Corresponding author: Spencer Hill.)

The authors are with the Department of Mathematics and Statistics, Queen’s University, Kingston, ON, K7L 3N6, Canada (e-mail: spencer.hill@queensu.ca; fa@queensu.ca; tamas.linder@queensu.ca).

Communicated by L. Yu, Associate Editor for Shannon Theory and Information Measures.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2026.3680805>.

Digital Object Identifier 10.1109/TIT.2026.3680805

lossless source coding theorem [1], we consider one-shot exact sampling under an exponential storage cost. Such a cost is appropriate for applications where buffer overflow can occur [9], [10], [11], [12], and therefore, the cost of long codewords is especially significant [13], [14], [15]. Here for a given $t > 0$, we aim to minimize the average codeword length of order t , also called the normalized cumulant generating function of codeword lengths [13], [14], [15], given by

$$L(t) = \frac{1}{t} \log(\mathbb{E}[2^{tI(\mathcal{C}(K))}]). \quad (1)$$

As $t \rightarrow 0$ in (1), by L'Hôpital's rule, we recover the average codeword length, $\mathbb{E}[I(\mathcal{C}(K))]$. In this work, we provide Rényi generalizations of prior results to the Campbell cost $L(t)$. In particular, we lower bound $L(t)$ for any sampling algorithm and use the Poisson functional representation to prove an upper bound on $L(t)$ whose leading Rényi divergence term has order within ϵ of the lower bound. We also describe an operational procedure for encoding K with a universal code and prove an upper bound on $L(t)$ that reduces to the bound of Harsha et al. [3] as $t \rightarrow 0$. Qualitatively, the lower and upper bounds match in shape and are within 5-10 bits of each other for most $0 < \alpha = 1/1 + t < 1$, even for distributions where $L(t)$ is very large. Our bounds admit closed forms for many distributions [16] and are otherwise easily computed numerically, unlike the natural bound $L(t) < H_\alpha(K) + 1$ (Proposition 1), whose right-hand side is computable only for special P, Q under the Poisson functional representation but quickly becomes intractable as $L(t)$ grows. We also analyze the Campbell cost in the asymptotic regime, where the goal is for the receiver to sample from the product distribution $P^{\otimes n}$ after receiving a single message. We derive exactly the optimal rate $\lim_{n \rightarrow \infty} L(t)/n$ over all sampling algorithms, and show that under mild assumptions, any causal sampler [17] (which examines the shared sequence one at a time) does strictly worse than a noncausal sampling algorithm. This differs from the case of the expected message length, where noncausal and causal algorithms achieve the same optimal asymptotic communication rate.

The remainder of the paper is organized as follows. Section II contains some preliminary definitions and a review of the relevant literature, as well as a detailed summary of our main contributions. Section III lower bounds $L(t)$ in the exact sampling problem, and Section IV describes two communication protocols using the Poisson functional representation which closely approximate this lower bound. Section V extends our results to the asymptotic regime, and Section VI gives numerical comparisons of the lower and upper bounds in the cases of normal and Laplacian distributions and a discussion of the difficulty of computing $H_\alpha(K)$.

Notation: For P, Q probability distributions, we write $P \ll Q$ to indicate that P is absolutely continuous with respect to Q , in which case $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative. For a distribution P on \mathcal{U} , we use $P^{\otimes n}$ to denote the product distribution on \mathcal{U}^n corresponding to n independent samples from P . We let $\{0, 1\}^*$ denote the collection of all finite-length binary words, i.e., $\{0, 1\}^* = \{0, 1, 00, 01, 10, 11, 000, \dots\}$. We write $\mathcal{N}(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2 . Throughout, \log is in base 2, all information measures are in bits, \ln denotes the natural logarithm, and Γ is the gamma function.

II. PRELIMINARIES AND CONTRIBUTIONS SUMMARY

A. Rényi Information Measures

Let X be a discrete random variable with alphabet \mathcal{X} and distribution P_X . For $\alpha \in (0, 1) \cup (1, \infty)$, we define the Rényi entropy of order α of X as [18]

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{x \in \mathcal{X}} P_X(x)^\alpha \right). \quad (2)$$

$H_\alpha(X)$ is nonincreasing in α , and if $H_\alpha(X) < \infty$ for some $0 < \alpha < 1$, then $\lim_{\alpha \nearrow 1} H_\alpha(X) = H(X)$, recovering the Shannon entropy. In [1], Campbell provided an operational meaning to the Rényi entropy by connecting it with the average code length of order t , $L(t)$.

Proposition 1 [1]: Suppose the discrete random variable X is encoded using a uniquely decodable binary code $\mathcal{C} : \mathcal{X} \rightarrow 0, 1^*$ such that each codeword has length $l(\mathcal{C}(x)) = n_x$, for $x \in \mathcal{X}$. Then the average code length of order $t > 0$ given by

$$L(t) = \frac{1}{t} \log \left(\sum_{x \in \mathcal{X}} P_X(x) 2^{tn_x} \right), \quad (3)$$

satisfies

$$H_\alpha(X) \leq L(t), \quad (4)$$

where $\alpha = \frac{1}{t+1}$. Moreover, there exists a uniquely decodable binary code \mathcal{C} such that

$$L(t) < H_\alpha(X) + 1. \quad (5)$$

We note that while Campbell's original proof was for X with finite alphabet, it can easily be extended to the case of a countably infinite alphabet. For simplicity, we will often refer to $L(t)$ as just the Campbell cost. $L(t)$ is a nondecreasing function of t , and is strictly increasing in t unless all the lengths n_x are equal. As a result, if X has an infinite support, then $L(t) \rightarrow \infty$ as $t \rightarrow \infty$. If X is finite-valued, then $L(t) \rightarrow \max_{x \in \mathcal{X}} n_x$ as $t \rightarrow \infty$.

Rényi also defined the divergence of order α between probability distributions P and Q , $D_\alpha(P||Q)$, as

$$D_\alpha(P||Q) = \frac{1}{\alpha-1} \log \left(\mathbb{E}_{X \sim Q} \left[\left(\frac{dP}{dQ}(X) \right)^\alpha \right] \right),$$

where $\alpha \in (0, 1) \cup (1, \infty)$ and it is assumed that $P \ll Q$. We note that $\lim_{\alpha \nearrow 1} D_\alpha(P||Q) = D(P||Q)$, i.e., the Kullback-Leibler divergence is recovered when $\alpha \nearrow 1$ [19]. There are several Rényi generalizations of the Shannon mutual information [20]; in our problem of channel simulation, Sibson's definition [21] naturally emerges. For arbitrary random variables X and Y on Polish spaces \mathcal{X} and \mathcal{Y} , with joint distribution P_{XY} and marginals P_X and P_Y , we define the Sibson α -mutual information for any $\alpha \in (0, 1) \cup (0, \infty)$ as

$$I_\alpha^s(X; Y) = \inf_{Q_Y} D_\alpha(P_{XY}||P_X \times Q_Y). \quad (6)$$

The infimum is achieved by the distribution Q_{Y_α} on \mathcal{Y} defined by

$$\frac{dQ_{Y_\alpha}(y)}{dP_Y} = \frac{\mathbb{E}_X \left[\left(\frac{dP_{XY}}{dP_X P_Y}(X, y) \right)^\alpha \right]^{\frac{1}{\alpha}}}{\mathbb{E}_Y \left[\mathbb{E}_X \left[\left(\frac{dP_{XY}}{dP_X P_Y}(X, Y) \right)^\alpha \right]^{\frac{1}{\alpha}} \right]}. \quad (7)$$

As $\alpha \rightarrow 1$, $\lim_{\alpha \nearrow 1} I_\alpha^s(X; Y) = I(X; Y)$, recovering Shannon's mutual information.

B. Problem Setup

Let P and Q be probability distributions on the Polish space \mathcal{U} with $P \ll Q$. Suppose the sender and receiver have shared access to the i.i.d. sequence $\{U_i\}_{i \geq 1}$ distributed according to Q ; this sequence is referred to as *common randomness*. The sender transmits an index K , encoded using the uniquely decodable binary variable-length code $\mathcal{C} : \mathbb{N} \rightarrow 0, 1^*$ with codeword lengths $l(\mathcal{C}(K))$, such that $U_K \sim P$. We refer to Q as the *proposal* distribution and P as the *target* distribution; it is always assumed that $P \ll Q$. This problem will henceforth be referred to as *exact sampling* with proposal Q and target P . In this setup, we wish to minimize the Campbell cost, $L(t)$, of the message, $\mathcal{C}(K)$, over all protocols such that $U_K \sim P$. We also study the *asymptotic* setting, where the proposal and target distributions are the product distributions $Q^{\otimes n}$ and $P^{\otimes n}$, respectively. For any $t > 0$, let $L_n^*(t)$ be the infimum of the set of achievable Campbell costs of order t among all protocols for proposal $Q^{\otimes n}$ and target $P^{\otimes n}$. The goal is to characterize the optimal asymptotic rate $\limsup_{n \rightarrow \infty} L_n^*(t)/n$ [22].

An important subclass of sampling algorithms is causal samplers [17]. A sampling algorithm is called causal if the index K is a stopping time, i.e., the event $\{K \geq k\}$ is independent of $\{U_{k+1}, U_{k+2}, \dots\}$. Intuitively, a causal sampler examines the shared randomness $\{U_i\}_{i \geq 1}$ sequentially. Any sampler which is not causal is called noncausal. Examples of causal samplers include greedy rejection sampling [3], [23] and greedy Poisson rejection sampling [24], while the Poisson functional representation [2] is a noncausal sampler.

C. Prior Work

The problem of channel simulation was first studied by Wyner [25], who considered the asymptotic version without common randomness but allowing error between the simulated and target distributions. Winter [26] showed that, by allowing a shared source of unlimited randomness, it is possible to simulate the channel with small error at an asymptotic cost equal to the mutual information. Bennett et al. [27] showed that the same result holds even for exact channel simulation. More recently, Cuff [28], Bennett et al. [29], and Yu and Tan [30] studied the tradeoff between communication rate and amount of common randomness in the asymptotic regime. Sriramu and Wagner [8] and Flamich et al. [31] examined the optimal rate of exact channel synthesis in the asymptotic case. There has also been recent work, motivated by applications in deep learning-based compression, on developing algorithms which are computationally efficient, e.g. [6], [23], [24], [32], [33], and [34]. We refer the reader to [22] for a comprehensive survey of channel simulation methods and results.

In the problem of exact sampling, it is not hard to show that a lower bound on the expected message length is $\mathbb{E}[l(\mathcal{C}(K))] \geq D(P\|Q)$ [3]. The first achievability bound is due to Harsha et al. [3], who described a communication protocol for discrete P, Q for which

$$\mathbb{E}[l(\mathcal{C}(K))] \leq D(P\|Q) + (1 + \epsilon) \log(D(P\|Q) + 1) + c_\epsilon,$$

with c_ϵ a constant depending on ϵ . Their proof effectively amounts to showing that $\mathbb{E}[\log K] \leq D(P\|Q) + 2 \log e$ and encoding K using a universal code. As noted in [2], by instead using a power law code one can bound $\mathbb{E}[l(\mathcal{C}(K))] \leq$

$D(P\|Q) + \log(D(P\|Q) + 1) + 7.78$ for discrete P, Q . The Poisson functional representation, first proposed by Li and El Gamal in [2], has since been used to prove a tighter upper bound on $\mathbb{E}[l(\mathcal{C}(K))]$ [35].

Definition 1 (Poisson functional representation [36]): Let P and Q be the target and proposal distributions, respectively, of an exact sampling problem. Let $\{T_i\}_{i \geq 1}$ be a rate-one Poisson process and let $\{U_i\}_{i \geq 1}$ be an i.i.d. sequence distributed according to Q and independent of $\{T_i\}_{i \geq 1}$. Define

$$K := \arg \min_{i \geq 1} \frac{T_i}{\frac{dP}{dQ}(U_i)}. \quad (8)$$

It is shown in [36] that under this definition, $U_K \sim P$. For P and Q general distributions and K generated according to Definition 1, [35] showed that

$$\mathbb{E}[l(\mathcal{C}(K))] \leq D(P\|Q) + \log(D(P\|Q) + 2) + 3.$$

Moreover, [2] showed that there exist distributions for which $\mathbb{E}[l(\mathcal{C}(K))] \geq D(P\|Q) + \log(D(P\|Q) + 1) - 1$, i.e., the log term in the upper bound cannot be removed in general. Recent improvements on these bounds are given in [7] and [31], where the *channel simulation divergence* $D_{CS}(P\|Q)$ (see [7] for the definition) is used to show

$$\begin{aligned} D(P\|Q) &\leq D_{CS}(P\|Q) \\ &\leq \mathbb{E}[l(\mathcal{C}(K))] \\ &\leq D_{CS}(P\|Q) + \log(e + 1) + 1. \end{aligned}$$

D. Contributions

In this work, we study the Campbell cost, $L(t)$, of exact sampling algorithms. Although exact sampling and its associated minimum communication cost are well-studied problems, to the best of our knowledge, this natural extension (from linear cost to an exponential cost in message lengths) has not been investigated in the literature. Our main contributions can be summarized as follows:

- We provide a lower bound on the $L(t)$ cost of encoding K using any (possibly noncausal) sampling algorithm that grows approximately as $D_{1/\alpha}(P\|Q)$, where $\alpha = 1/1 + t$. We also prove the lower bound $L(t) \geq D_{2-\alpha}(P\|Q) + \frac{1}{1-\alpha} \log\left(\frac{1}{2-\alpha}\right)$, which is a tighter bound when α is close to 1 or $D_\alpha(P\|Q)$ is small.
- We use the Poisson functional representation to prove the upper bound $L(t) \leq (1 + \epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) + c_1(\alpha, \epsilon)$, for any $0 < \alpha < 1$ and $\epsilon > 0$, where $c_1(\alpha, \epsilon)$ is a constant depending on α and ϵ . Using numerical examples, we demonstrate that this upper bound is within 5-10 bits of the $D_{1/\alpha}(P\|Q)$ lower bound, even for distributions where $L(t)$ is large.
- We use our results to upper bound the Rényi entropy of channel simulation by the Sibson α -mutual information [21], establishing a Rényi-generalized version of the strong functional representation lemma [2].
- We describe an operational procedure for encoding K using a universal code that gives $L(t) \leq D_{\frac{2-\alpha}{\alpha}}(P\|Q) + (1 + \epsilon) \log(D(P\|Q) + 1) + c_2(\epsilon)$, for any $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$, where $c_2(\epsilon)$ is a constant. We show that this upper bound reduces to the bound of Harsha et al. [3] as $\alpha \rightarrow 1$.

- We characterize the minimum asymptotic Campbell cost of exact sampling between product distributions of dimension n , showing that $\lim_{n \rightarrow \infty} L_n^*(t)/n = D_{1/\alpha}(P\|Q)$, where $\alpha = \frac{1}{1+t}$.
- We show that under mild assumptions, a causal sampler has $\liminf_{n \rightarrow \infty} L_n^*/n \geq D_\beta(P\|Q)$, where $\beta > \frac{1}{\alpha}$, and thus does strictly worse than a noncausal sampler in the asymptotic Campbell cost.

Our main technical tool is to construct an encoding of K that allows us to invoke the upper and lower bounds on the moments of K in [17] and [36]. This requires, among other technical details, nontrivially extending the lower bound of [17, Thm. 7] to any bijection of K and the moment upper bound [36, Prop. 4] beyond concave functions of K .

III. LOWER BOUND ON THE CAMPBELL COST

We present our main results as bounds on the Campbell cost $L(t)$; but, using Proposition 1, we can write them as bounds on $H_\alpha(K)$ within one bit (this will be formalized in Corollary 1). In this section, we will prove a lower bound on the Campbell cost of any sampling algorithm. Note that for all $0 < \alpha < 1$ and with $t = \frac{1-\alpha}{\alpha}$, one has the trivial lower bound $L(t) \geq H_\alpha(K) \geq H(K) \geq D(P\|Q)$, where the final inequality follows similar to [3, Prop. IV.3]. However, such a lower bound loses the dependency on α , which is key to understanding $H_\alpha(K)$ and $L(t)$.

Theorem 1: Let K be the output of any exact sampling algorithm with proposal distribution Q , target distribution P , and common randomness $\{U_i\}_{i \geq 1}$. Suppose K is encoded using a uniquely decodable binary code \mathcal{C} . Then, for any $0 < \alpha < 1$ with $t = \frac{1-\alpha}{\alpha}$,

$$L(t) \geq \max \text{LB}_1^\alpha, \text{LB}_2^\alpha, \quad (9)$$

with $\text{LB}_1^\alpha := D_{\frac{1}{\alpha}}(P\|Q) + \frac{\alpha}{1-\alpha} \log(\alpha) - 1$ and $\text{LB}_2^\alpha := D_{2^{-\alpha}}(P\|Q) + \frac{1}{1-\alpha} \log\left(\frac{1}{2-\alpha}\right)$.

Theorem 1 lower bounds the Campbell cost for any exact sampling problem. The constant term in LB_1^α ranges between -1 (at $\alpha = 0$) and $\frac{-1}{\ln 2} - 1$ (at $\alpha = 1$) while the constant term in LB_2^α ranges between -1 (at $\alpha = 0$) and $\frac{-1}{\ln 2}$ (at $\alpha = 1$). Thus, as $\alpha \rightarrow 1$ in (9) we recover the bound $\mathbb{E}[n_K] \geq D(P\|Q) - \frac{1}{\ln(2)}$, the lower bound in [3] with a small penalty term of $\frac{-1}{\ln(2)}$. Except for P and Q close together (in the sense of Rényi divergence) or α close to 1, in (9), the dominant term will typically be $D_{1/\alpha}(P\|Q)$ and LB_1^α will yield the tighter bound. We include LB_2^α only to improve the lower bound for such situations. Numerical examples comparing LB_1^α and LB_2^α , shown in Section VI, support this conclusion. Before proving Theorem 1, we state a lemma to be used in the proof.

Lemma 1: Let K be the output of any sampling algorithm with proposal distribution Q , target distribution P , and common randomness $\{U_i\}_{i \geq 1}$. Then, for any $0 < \alpha < 1$ and bijection $g: \mathbb{N} \rightarrow \mathbb{N}$,

$$\mathbb{E}[(g(K))^\alpha] \geq \frac{1}{1+\alpha} 2^{\alpha D_{\alpha+1}(P\|Q)}. \quad (10)$$

Proof: The lower bound (10) was originally proved in [17] for countable \mathcal{U} and $g(k) = k$, we extend the argument to Polish spaces and all bijections $g: \mathbb{N} \rightarrow \mathbb{N}$. The proof can be found in Appendix A. \square

Proof of Theorem 1: We will separately show that $L(t) \geq \text{LB}_1^\alpha$ and $L(t) \geq \text{LB}_2^\alpha$. We first prove that $\text{LB}_2^\alpha \leq H_\alpha(K)$; the desired inequality ($\text{LB}_2^\alpha \leq L(t)$) is immediate by the bound $H_\alpha(K) \leq L(t)$ from Proposition 1. Define $g: \mathbb{N} \rightarrow \mathbb{N}$ to be a bijection which maps k to its index in $\{\mathbb{P}(K = k)\}_{k \geq 1}$ sorted in nonincreasing order (with ties broken arbitrarily), so that with $\tilde{K} := g(K)$ and $\tilde{p}(k) := \mathbb{P}(\tilde{K} = k)$ one has $\tilde{p}(k) \geq \tilde{p}(k+1)$ for all $k \in \mathbb{N}$. Note that $H_\alpha(K) = H_\alpha(\tilde{K})$. For any $n \in \mathbb{N}$,

$$1 = \sum_{k=1}^{\infty} \tilde{p}(k) \geq \sum_{k=1}^n \tilde{p}(k) \geq \sum_{k=1}^n \tilde{p}(n) = n\tilde{p}(n), \quad (11)$$

implying $\tilde{p}(n) \leq \frac{1}{n}$ for all n . Then, because $1 - \alpha > 0$,

$$\begin{aligned} \sum_{k=1}^{\infty} \tilde{p}(k)^\alpha &= \sum_{k=1}^{\infty} \tilde{p}(k) \left(\frac{1}{\tilde{p}(k)} \right)^{1-\alpha} \\ &\geq \sum_{k=1}^{\infty} \tilde{p}(k) k^{1-\alpha} = \mathbb{E}[\tilde{K}^{1-\alpha}]. \end{aligned} \quad (12)$$

Taking logarithms and dividing by $1 - \alpha$ gives

$$\begin{aligned} H_\alpha(K) &= H_\alpha(\tilde{K}) \\ &\geq \frac{1}{1-\alpha} \log(\mathbb{E}[\tilde{K}^{1-\alpha}]) \\ &\geq \frac{1}{1-\alpha} \log\left(\frac{1}{2-\alpha} 2^{(1-\alpha)D_{2-\alpha}(P\|Q)}\right) \\ &= D_{2-\alpha}(P\|Q) + \frac{1}{1-\alpha} \log\left(\frac{1}{2-\alpha}\right), \end{aligned} \quad (13)$$

where (13) follows from Lemma 1.

We now prove that $L(t) \geq \text{LB}_1^\alpha$ by considering an encoding of K which is one-to-one and not necessarily uniquely decodable. Here, a one-to-one code is an injective function $f: \mathbb{N} \rightarrow 0, 1^*$. Consider a bijection $g: \mathbb{N} \rightarrow \mathbb{N}$ such that $\tilde{K} := g(K)$ has probability distribution $\tilde{p}(k) := \mathbb{P}(\tilde{K} = k)$ which is nonincreasing in k . The minimum Campbell costs of encoding K and \tilde{K} are the same, and the same is true for any one-to-one encoding of \tilde{K} . Therefore, because uniquely decodable codes are contained in the class of all one-to-one codes, showing that $L(t) \geq D_{\frac{1}{\alpha}}(P\|Q) + \frac{\alpha}{1-\alpha} \log(\alpha) - 1 = \text{LB}_1^\alpha$ for any one-to-one encoding of \tilde{K} will prove the theorem. Since the probability distribution of \tilde{K} is nonincreasing in k ; therefore, by the rearrangement inequality (see, e.g., [37]) the optimal one-to-one code f^* under the Campbell cost must have nondecreasing codeword lengths, i.e., $l(f^*(k)) \leq l(f^*(k+1))$ for all k . Therefore, an f^* that maps k to the k th element of $0, 1^*$, i.e., $f^*(1) = 0, f^*(2) = 1, \text{etc.}$, is optimal. This optimal f^* has codeword lengths $l(f^*(k)) = \lceil \log(k+1) \rceil$, and hence

$$\begin{aligned} L(t) &= \frac{1}{t} \log\left(\mathbb{E}\left[2^{tl(f(\tilde{K}))}\right]\right) \\ &\geq \frac{1}{t} \log\left(\mathbb{E}\left[2^{tl(f^*(\tilde{K}))}\right]\right) \\ &= \frac{1}{t} \log\left(\sum_{k=1}^{\infty} \tilde{p}(k) 2^{t \lceil \log(k+1) \rceil}\right) \\ &\geq \frac{1}{t} \log\left(\sum_{k=1}^{\infty} \tilde{p}(k) 2^{t \log(k-1)}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{t} \log(\mathbb{E}[\tilde{K}^t]) - 1 \\
 &\geq \frac{1}{t} \log\left(\frac{1}{1+t} 2^{tD_{1+\alpha}(P\|Q)}\right) - 1 \\
 &= D_{\frac{1}{\alpha}}(P\|Q) + \frac{\alpha}{1-\alpha} \log(\alpha) - 1,
 \end{aligned} \tag{14}$$

where (14) follows from Lemma 1. \square

IV. UPPER BOUNDS USING THE POISSON FUNCTIONAL REPRESENTATION

Theorem 1 tells us that the minimum Campbell cost of any sampling algorithm grows at least as $D_{\frac{1}{\alpha}}(P\|Q)$. The natural question is whether any one-shot sampling algorithms exist that achieve this lower bound. Suppose that we could bound $L(t) \leq D_{1/\alpha}(P\|Q) + C$ for some constant C and all distributions P, Q . Then, taking $\alpha \rightarrow 1$ (equivalently $t \rightarrow 0$) would yield $\mathbb{E}[L(\mathcal{C}(K))] \leq D(P\|Q) + C$, a bound shown to be invalid by an explicit counter-example [2]. Therefore, it is reasonable to conjecture that a tight upper bound in the form $L(t) \leq D_{1/\alpha}(P\|Q) + c$, which is valid for all P and Q , does not exist for the Campbell cost, either. Instead, we have the following upper bound on $L(t)$ using the Poisson functional representation.

Theorem 2: Let K be the output of the Poisson functional representation given proposal distribution Q and target distribution P . Then, for any $0 < \alpha < 1$ and $\epsilon > 0$, there exists a uniquely decodable encoding of K such that

$$L(t) \leq (1 + \epsilon)D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) + c_1(\alpha, \epsilon), \tag{15}$$

with $t = \frac{1-\alpha}{\alpha}$ and $c_1(\alpha, \epsilon)$ a constant term defined as

$$c_1(\alpha, \epsilon) := \begin{cases} (1 + \epsilon) \log e + 1 + \log\left(1 + \frac{1}{\epsilon}\right), & \frac{1}{2} < \alpha < 1 \text{ and } 0 < \epsilon < \frac{2\alpha-1}{1-\alpha} \\ \frac{\alpha}{1-\alpha} \log\left(\Gamma\left(\frac{1+\epsilon(1-\alpha)}{\alpha}\right)\right) + 4 & 0 < \alpha < \frac{1}{2} \\ + 3\epsilon - \frac{2\alpha}{1-\alpha} + \log\left(1 + \frac{1}{\epsilon}\right), & \text{or } \epsilon \geq \frac{2\alpha-1}{1-\alpha}. \end{cases} \tag{16}$$

We will refer to the upper bound in (15) as $\text{UB}_{1+\epsilon}^{\alpha} := (1 + \epsilon)D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) + c_1(\alpha, \epsilon)$. We note that $c_1(\alpha, \epsilon)$ is finite for all $0 < \alpha < 1$, but tends to ∞ as $\alpha \rightarrow 0$. We can upper bound $c_1(\alpha, \epsilon)$ by a simpler function which more clearly shows the dependence on α and ϵ .

Proposition 2: Let $c_1(\alpha, \epsilon)$ be defined as in (16). Then, for any $\epsilon > 0$ and $0 < \alpha < 1$,

$$\begin{aligned}
 c_1(\alpha, \epsilon) &< \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon\right) \log\left(\frac{1}{\alpha}\right) \\
 &\quad + \log\left(\frac{1}{\epsilon}\right) + 1.5\epsilon^2 + 4.5\epsilon + 2.6 =: \tilde{c}_1(\alpha, \epsilon).
 \end{aligned}$$

Proof: The proof can be found in Appendix B. \square

Fig. 3 shows the value of $c_1(\alpha, \epsilon)$ and the looser bound in Proposition 2, minimized over $\epsilon > 0$ for each $0.2 < \alpha < 1$. Numerically, we see that $\min_{\epsilon > 0} \tilde{c}_1(\alpha, \epsilon)$ of Proposition 2 is approximately 2-3 bits looser than $\min_{\epsilon > 0} c_1(\alpha, \epsilon)$.

Before proving Theorem 2, we state two lemmas which will be instrumental in upper-bounding the moments of K .

Lemma 2: Let K be the output of the Poisson functional representation given proposal distribution Q and target

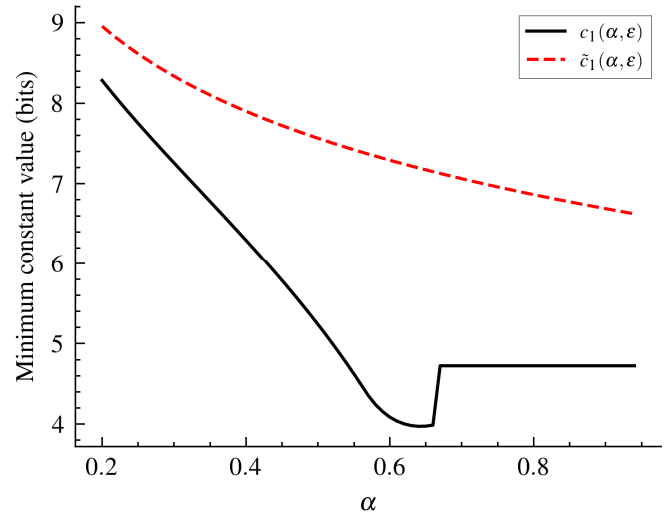


Fig. 3. $\min_{\epsilon > 0} c_1(\alpha, \epsilon)$ and $\min_{\epsilon > 0} \tilde{c}_1(\alpha, \epsilon)$, as α varies between 0.2 and 1.

distribution P . Then,

$$\mathbb{E}[\log K] \leq D(P\|Q) + 1, \tag{17}$$

and for $0 < \alpha < 1$,

$$\mathbb{E}[K^\alpha] \leq 2^{\alpha D_{\alpha+1}(P\|Q)} + \alpha. \tag{18}$$

Equation (17) was proved in [35], and the upper bound in (18) follows from [36, Prop. 4] after substituting $j = 1$. We note that for $0 < \alpha < 1$ and $g(k) = k$, the Poisson functional representation in (18) almost achieves the lower bound on the α -moment of any sampling algorithm in Lemma 1.

Lemma 3: Let $X \sim \text{Geo}(p)$ be a geometric random variable with parameter $0 < p < 1$. Then, for any $r \geq 1$,

$$\mathbb{E}[X^r] \leq 2^{r-1} \left(\frac{\Gamma(r+1)}{p^r} + 1\right).$$

Proof: The proof can be found in Appendix C. \square

Proof of Theorem 2: We will prove Theorem 2 in two cases. In both cases, we will relate the Campbell cost $L(t)$ to the r th moment of K , with $r = \frac{1-\alpha}{\alpha}(1 + \epsilon)$. For $r < 1$ (equivalently $\epsilon < \frac{2\alpha-1}{1-\alpha}$), we will bound the moments of K using (18) in Lemma 2, while in the case $r \geq 1$ (equivalently $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$) we will use the bound on the moments of a geometric random variable in Lemma 3. In particular, we will show that for any $1/2 < \alpha < 1$ and $0 < \epsilon < \frac{2\alpha-1}{1-\alpha}$, with $t = \frac{1-\alpha}{\alpha}$,

$$\begin{aligned}
 L(t) &\leq (1 + \epsilon)D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) \\
 &\quad + (1 + \epsilon) \log e + 1 + \log\left(1 + \frac{1}{\epsilon}\right).
 \end{aligned} \tag{19}$$

Then, for any $0 < \alpha < 1$, for $t = \frac{1-\alpha}{\alpha}$ and $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$ (so that $t(1 + \epsilon) \geq 1$) we will show that

$$\begin{aligned}
 L(t) &\leq (1 + \epsilon)D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) \\
 &\quad + \frac{\alpha}{1-\alpha} \log\left(\Gamma\left(\frac{1 + \epsilon(1-\alpha)}{\alpha}\right)\right) \\
 &\quad + 4 + 3\epsilon - \frac{2\alpha}{1-\alpha} + \log\left(1 + \frac{1}{\epsilon}\right).
 \end{aligned} \tag{20}$$

Taking (19) and (20) together yields the statement of Theorem 2.

Fix $1/2 < \alpha < 1$ and $0 < \epsilon < \frac{2\alpha-1}{1-\alpha}$. Then, by the Kraft inequality there exists a uniquely decodable encoding of K with codeword lengths $n_k \leq (1 + \epsilon) \log k + 1 + \log(1 + \frac{1}{\epsilon})$; this is shown in Appendix D. Let $t = \frac{1-\alpha}{\alpha}$. Then, $0 < t < 1$ and $0 < \epsilon < \frac{2\alpha-1}{1-\alpha} = \frac{1-t}{t}$ imply that $t(1 + \epsilon) < 1$, so we can apply the upper bound on $\mathbb{E}[K^\alpha]$ in (18) to get

$$\begin{aligned} L(t) &= \frac{1}{t} \log \left(\sum_{k=1}^{\infty} p(k) 2^{m_k} \right) \\ &\leq \frac{1}{t} \log \left(\sum_{k=1}^{\infty} p(k) 2^{t(1+\epsilon) \log k + t + \log(1 + \frac{1}{\epsilon})} \right) \\ &= \frac{1}{t} \log(\mathbb{E}[K^{(1+\epsilon)t}]) + 1 + \log \left(1 + \frac{1}{\epsilon} \right) \\ &\leq \frac{1}{t} \log \left(2^{(1+\epsilon)t D_{(1+\epsilon)t+1}(P\|Q)} + (1 + \epsilon)t \right) \\ &\quad + 1 + \log \left(1 + \frac{1}{\epsilon} \right) \end{aligned} \quad (21)$$

$$\begin{aligned} &\leq (1 + \epsilon) D_{(1+\epsilon)t+1}(P\|Q) + (1 + \epsilon) \log e \\ &\quad + 1 + \log \left(1 + \frac{1}{\epsilon} \right). \end{aligned} \quad (22)$$

Here, (21) follows from (18) and (22) from $\log(x + a) \leq \log(x) + a \log e$ for all $x \geq 1$ and $a > 0$. Noting that $(1 + \epsilon)t + 1 = \frac{1+\epsilon(1-\alpha)}{\alpha}$ gives (19).

Suppose now that $0 < \alpha < 1$ and $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$, so that with $t = \frac{1-\alpha}{\alpha}$, $(1 + \epsilon)t \geq 1$. From [36, Eq. 29] (after substituting $j = 1$), for any $u \in \mathcal{U}$,

$$K | U_K = u \sim \text{Geo}(\beta(u)), \quad (23)$$

for $\beta(u) = \mathbb{E}_{U \sim Q} \left[\max \left\{ \frac{dP}{dQ}(u), \frac{dP}{dQ}(U) \right\} \right]^{-1}$. By Lemma 3, for $r = (1 + \epsilon)t \geq 1$,

$$\begin{aligned} &\mathbb{E}[K^r | U_K = u] \\ &\leq 2^{r-1} \left(\frac{\Gamma(r+1)}{\beta(u)^r} + 1 \right) \\ &= 2^{r-1} \Gamma(r+1) \mathbb{E}_{U \sim Q} \left[\max \left\{ \frac{dP}{dQ}(u), \frac{dP}{dQ}(U) \right\} \right]^r + 2^{r-1} \\ &\leq 2^{r-1} \Gamma(r+1) \left(\frac{dP}{dQ}(u) + \mathbb{E}_{U \sim Q} \left[\frac{dP}{dQ}(U) \right] \right)^r + 2^{r-1} \\ &= 2^{r-1} \Gamma(r+1) \left(\frac{dP}{dQ}(u) + 1 \right)^r + 2^{r-1} \\ &\leq \Gamma(r+1) 2^{2r-2} \left(\left(\frac{dP}{dQ}(u) \right)^r + 1 \right) + 2^{r-1} \end{aligned} \quad (24)$$

where (24) follows again by $(x + 1)^r \leq 2^{r-1}(x^r + 1)$ for any $r \geq 1$ and $x > 0$. Taking expectation with respect to $U_K \sim P$, we get that

$$\mathbb{E}[K^r] \leq \Gamma(r+1) 2^{2r-2} (2^{r D_{r+1}(P\|Q)} + 1) + 2^{r-1}. \quad (25)$$

Again encoding K using a uniquely decodable code with codeword lengths $n_k \leq (1 + \epsilon) \log k + 1 + \log(1 + \frac{1}{\epsilon})$ and writing $r = (1 + \epsilon)t$ and $c(\epsilon) = 1 + \log(1 + \frac{1}{\epsilon})$, we have that

$$\begin{aligned} L(t) &\leq \frac{1}{t} \log(\mathbb{E}[K^r]) + c(\epsilon) \\ &\leq \frac{1}{t} \log(\Gamma(r+1) 2^{2r-2} (2^{r D_{r+1}(P\|Q)} + 1) + 2^{r-1}) \\ &\quad + c(\epsilon) \end{aligned} \quad (26)$$

$$\begin{aligned} &\leq \frac{1}{t} \log(\Gamma(r+1) 2^{2r-2} (2^{r D_{r+1}(P\|Q)} + 1)) \\ &\quad + \frac{1}{t} \log(2^{r-1}) + c(\epsilon) \end{aligned} \quad (27)$$

$$\begin{aligned} &= \frac{1}{t} \log(2^{r D_{r+1}(P\|Q)} + 1) + \frac{1}{t} \log(\Gamma(r+1)) \\ &\quad + \frac{2r-2}{t} + \frac{r-1}{t} + c(\epsilon) \\ &\leq \frac{1}{t} \log(2^{r D_{r+1}(P\|Q)}) + \frac{1}{t} \log(\Gamma(r+1)) \\ &\quad + \frac{2r-2}{t} + \frac{1}{t} + \frac{r-1}{t} + c(\epsilon) \\ &= \frac{r}{t} D_{r+1}(P\|Q) + \frac{1}{t} \log(\Gamma(r+1)) \\ &\quad + \frac{2r-2}{t} + \frac{1}{t} + \frac{r-1}{t} + c(\epsilon). \end{aligned} \quad (28)$$

Here, (26) follows by (25), while (27) and (28) both follow from the inequality $\log(x + 1) \leq \log(x) + 1$ for all $x \geq 1$. Substituting $r = (1 + \epsilon)t$ and $t = \frac{1-\alpha}{\alpha}$ gives (20) and the statement of the theorem. \square

The Rényi divergence term in UB_1^α has order $\frac{1+\epsilon(1-\alpha)}{\alpha}$, which as $\epsilon \rightarrow 0$ goes to $\frac{1}{\alpha}$, the order of the Rényi divergence term in the (generally) tighter LB_1^α . However, $c_1(\alpha, \epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 0$, thus UB_1^α is not tight. Numerical examples in Section VI demonstrate that, after minimizing over $\epsilon > 0$, UB_1^α is typically within 5-10 bits of LB_1^α . As $\alpha \rightarrow 1$, we obtain the upper bound on the expected message length $\mathbb{E}[l(C(K))]$ $\leq (1 + \epsilon) D(P\|Q) + (1 + \epsilon) \log e + 1 + \log(1 + \frac{1}{\epsilon})$, for any $\epsilon > 0$. This bound is generally looser than the bound from the strong functional representation lemma [2], but better than the upper bound $\mathbb{E}[l(C(K))] \leq D(P\|Q) + (1 + \epsilon) \log(D(P\|Q) + 1) + c_\epsilon$ by Harsha et al. [3]. Encoding K using the Elias omega code [38], as done in [3], yields the following generalization of their result, which is strictly worse than UB_1^α .

Theorem 3: Let K be the output of the Poisson functional representation given proposal distribution Q and target distribution P . Then, for any $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$, encoding K using a universal code with codeword lengths $n_k = \log k + (1 + \epsilon) \log \log(k + 1) + O(1)$ gives

$$L(t) \leq D_{\frac{2-\alpha}{\alpha}}(P\|Q) + (1 + \epsilon) \log(D(P\|Q) + 1) + c_2(\epsilon), \quad (29)$$

with $t = \frac{1-\alpha}{\alpha}$ and $c_2(\epsilon) = 3 + \epsilon + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right)$.

Proof: Let $2/3 < \alpha < 1$ and $0 < \epsilon \leq \frac{3\alpha-2}{2-2\alpha}$. We will bound $L(t)$ using a universal coding of the natural numbers. In particular, using the prefix-free encoding described in [39, Ex. 1.11.16] (see also [3]) we can set $n_k \leq \log k + (1 + \epsilon) \log \log(k + 1) + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right)$ for any $\epsilon > 0$; this is proved in Appendix D. Since $t = \frac{1-\alpha}{\alpha}$, $0 < t < 1/2$. Then,

$$\begin{aligned} L(t) &= \frac{1}{t} \log(\mathbb{E}[2^{m_k}]) \\ &\leq \frac{1}{t} \log(\mathbb{E}[K^t \log(K + 1)^{(1+\epsilon)t}]) \\ &\quad + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right) \\ &\leq \frac{1}{t} \log\left(\sqrt{\mathbb{E}[K^{2t}] \mathbb{E}[\log(K + 1)^{2(1+\epsilon)t}]}\right) \\ &\quad + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right) \end{aligned} \quad (30)$$

$$= \frac{1}{2t} (\log(\mathbb{E}[K^{2t}]) + \log(\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}])) \\ + 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right). \quad (31)$$

Here, (30) follows from the Cauchy-Schwartz inequality. We will bound each of the two terms in (31) separately, beginning with $\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}]$. As $0 < t < \frac{1}{2}$, choosing $0 < \epsilon \leq \frac{1-2t}{2t} = \frac{3\alpha-2}{2-2\alpha}$ gives that $2(1+\epsilon)t \leq 1$, and so Jensen's inequality can be applied as

$$\frac{1}{2t} \log(\mathbb{E}[\log(K+1)^{2(1+\epsilon)t}]) \\ \leq \frac{1}{2t} \log(\mathbb{E}[\log(K+1)]^{2(1+\epsilon)t}) \\ = (1+\epsilon) \log(\mathbb{E}[\log(K+1)]) \\ \leq (1+\epsilon) \log(\mathbb{E}[\log K] + 1) \quad (32)$$

$$\leq (1+\epsilon) \log(D(P\|Q) + 1 + 1) \quad (33)$$

$$\leq (1+\epsilon) \log(D(P\|Q) + 1) + 1 + \epsilon. \quad (34)$$

Here, (32) and (34) both follow from the inequality $\log(x+1) \leq \log(x) + 1$ for all $x \geq 1$ and (33) follows from the upper bound (17) on $\mathbb{E}[\log K]$ in Lemma 2. In the first term of (31), because $0 < t < 1/2$ by assumption, we can apply the upper bound (18) on $\mathbb{E}[K^\alpha]$ in Lemma 2 to get that

$$\frac{1}{2t} \log(\mathbb{E}[K^{2t}]) \leq \frac{1}{2t} \log(2^{2tD_{2t+1}(P\|Q)} + 2t) \\ \leq D_{2t+1}(P\|Q) + 1, \quad (35)$$

where (35) holds again by $\log(x+a) \leq \log(x) + a$. Combining (34) and (35), and noting $2t+1 = \frac{2-\alpha}{\alpha}$, we obtain

$$L(t) \leq D_{\frac{2-\alpha}{\alpha}}(P\|Q) + (1+\epsilon) \log(D(P\|Q) + 1) + c_2(\epsilon),$$

with $c_2(\epsilon) = 2 + \epsilon + \log\left(\frac{\ln 2}{\epsilon} + \frac{3}{2}\right)$. \square

We will refer to the upper bound in Theorem 3 as $\text{UB}_2^\alpha := D_{\frac{2-\alpha}{\alpha}}(P\|Q) + (1+\epsilon) \log(D(P\|Q) + 1) + c_2(\epsilon)$. As $\alpha \rightarrow 1$, UB_1^α reduces to $\mathbb{E}[\mathcal{I}(\mathcal{C}(K))] \leq D(P\|Q) + (1+\epsilon) \log(D(P\|Q) + 1) + c_2(\epsilon)$ for any $\epsilon > 0$, the same upper bound as in [3]. Note that UB_2^α is valid only for $2/3 < \alpha < 1$, whereas UB_1^α is valid for all $0 < \alpha < 1$. Moreover, for all $2/3 < \alpha < 1$, $\text{UB}_1^\alpha < \text{UB}_2^\alpha$. To see why, observe that for $\epsilon < 1$, $\frac{1+\epsilon(1-\alpha)}{\alpha} < \frac{2-\alpha}{\alpha}$, i.e., the order of the leading Rényi divergence term is strictly less in UB_1^α . Since $D_\alpha(P\|Q)$ is nondecreasing in α and is strictly increasing unless $P = Q(\cdot|A)$ for some event A (see, e.g., [19, Thm. 3]) one has $D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) < D_{\frac{2-\alpha}{\alpha}}(P\|Q)$ for general P and Q . For such distributions, $(1+\epsilon)D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P\|Q) < D_{\frac{2-\alpha}{\alpha}}(P\|Q)$ for $\epsilon > 0$ small enough. As $c_1(\alpha, \epsilon) < c_2(\epsilon)$ for all $2/3 < \alpha < 1$, UB_1^α is better even for small $D_\alpha(P\|Q)$. Numerical results, shown in Section VI, illustrate this conclusion.

While Theorems 1 to 3 are bounds on the Campbell cost, we can use Proposition 1 to obtain bounds on $H_\alpha(K)$.

Corollary 1: Let K be the output of any exact sampling algorithm with proposal distribution Q , target distribution P , and common randomness $\{U_i\}_{i \geq 1}$. For any $0 < \alpha < 1$, $H_\alpha(K) > \max\{\text{LB}_1^\alpha, \text{LB}_2^\alpha\} - 1$, with LB_1^α and LB_2^α given in Theorem 1. Moreover, for K the output of the Poisson functional representation, for any $0 < \alpha < 1$ and $\epsilon > 0$ we have that $H_\alpha(K) \leq \text{UB}_1^\alpha$, with UB_1^α given in Theorem 2.

Proof: The upper bound is immediate after substitution in (4) in Proposition 1. The lower bound follows from (5) in

Proposition 1, specifically that for any sampling algorithm outputting K , there exists an encoding of K such that $L(t) < H_\alpha(K) + 1$. Then, the lower bound $L(t) \geq \max\{\text{LB}_1^\alpha, \text{LB}_2^\alpha\}$ of Theorem 1 applies to this sampling algorithm and encoding, meaning $H_\alpha(K) > \max\{\text{LB}_1^\alpha, \text{LB}_2^\alpha\} - 1$. \square

In the problem of exact channel simulation used to motivate this paper, we have the following corollaries to Theorem 2.

Corollary 2: Let X and Y be arbitrary random variables with Polish alphabets \mathcal{X} and \mathcal{Y} . The sender is provided with $x \in \mathcal{X}$ generated according to P_X and wishes to transmit a message K so that the receiver can generate $y \in \mathcal{Y}$ with distribution $P_{Y|X}(\cdot|x)$. We assume the sender and receiver have shared access to the sequence $\{U_i\}_{i \geq 1}$ i.i.d. according to the proposal distribution Q_{Y_α} defined in (7). Let K be the output of the Poisson functional representation. Then, for any $0 < \alpha < 1$ and $\epsilon > 0$,

$$H_\alpha(K) \leq (1+\epsilon)I_{\frac{1+\epsilon(1-\alpha)}{\alpha}}^S(X; Y) + c_1(\alpha, \epsilon), \quad (36)$$

where $I_\alpha^S(X; Y)$ is the Sibson α -mutual information defined in (6) and $c_1(\alpha, \epsilon)$ is as in Theorem 2.

Remark 1: The distribution Q_{Y_α} minimizes $D_\alpha(P_{XY}\|P_X \times Q_Y)$ over all proposal distributions. Notably, the minimizer is not the marginal distribution P_Y . This differs from the Shannon setting, where selecting the proposal distribution to be P_Y is known to be optimal.

Remark 2: Corollary 2 can be interpreted as a Rényi-generalized version of the strong functional representation lemma [2]. Indeed, for $Z = \{U_i\}_{i \geq 1}$ the common randomness, we have that

$$H_\alpha(Y|Z) \leq H_\alpha(K, Z|Z) = H_\alpha(K|Z) \leq H_\alpha(K).$$

As such, (36) upper bounds the Rényi conditional entropy $H_\alpha(Y|Z)$. Here, the definition of the Rényi conditional entropy can be any which satisfies both conditioning reduces entropy and the data processing inequality; examples include the Arimoto [40] and Hayashi-Skoric conditional entropies [41], [42], [43].

Proof: As in the proof of [2, Thm. 1], we condition on $X = x$ and take expectation of the upper bound on $\mathbb{E}[K^r | X = x]$ in (18) and (25). Choosing $P = P_{Y|X}(\cdot|x)$ and $Q = Q_{Y_\alpha}$, we obtain

$$\mathbb{E}[K^r | X = x] \leq \gamma_1(r)2^{rD_{r+1}(P_{Y|X}(\cdot|x)\|Q_{Y_\alpha})} + \gamma_2(r),$$

where $r > 0$ and $\gamma_1(r), \gamma_2(r)$ are constants as in (18) for $0 < r < 1$ and (25) for $r \geq 1$. Taking expectation with respect to X gives

$$\mathbb{E}[K^r] \leq \gamma_1(r)\mathbb{E}_X[2^{rD_{r+1}(P_{Y|X}(\cdot|X)\|Q_{Y_\alpha})}] + \gamma_2(r) \\ = \gamma_1(r)2^{rI_{r+1}^S(X; Y)} + \gamma_2(r).$$

Encoding K as in the proof of Theorem 2, combined with this unconditional moment upper bound, gives (36). \square

Corollary 3: In the case of channel simulation with arbitrary input, i.e., where the distribution P_X is not known, for any $0 < \alpha < 1$ and $\epsilon > 0$, we have

$$\max_{x \in \mathcal{X}} H_\alpha(K | X = x) \leq (1+\epsilon)C_{\frac{1+\epsilon(1-\alpha)}{\alpha}}^S + c_1(\alpha, \epsilon), \quad (37)$$

where

$$C_\alpha^S := \max_{P_X} I_\alpha^S(X; Y) = \min_{Q_Y} \max_{x \in \mathcal{X}} D_\alpha(P_{Y|X}(\cdot|x)\|Q_Y)$$

is the Sibson capacity of the channel $X \rightarrow Y$ [44].

Proof: Fix $x \in \mathcal{X}$, $P = P_{Y|X}(\cdot | x)$, and $Q = Q_Y$. Theorem 2 gives that

$$\begin{aligned} H_\alpha(K | X = x) \\ \leq (1 + \epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P_{Y|X}(\cdot | x) \| Q_Y) + c_1(\alpha, \epsilon). \end{aligned}$$

Maximizing over $x \in \mathcal{X}$ and minimizing over Q_Y gives (37). \square

As $\alpha \rightarrow 1$ in (36), the leading term becomes $I(X; Y)$, Shannon's mutual information. Similarly, the Sibson capacity in (37) recovers the Shannon capacity as $\alpha \rightarrow 1$.

V. ASYMPTOTIC RESULTS

A natural extension of the one-shot exact sampling problem is the asymptotic problem, where the common randomness is the i.i.d. sequence $\{U_i\}_{i \geq 1} \sim Q^{\otimes n}$ and we wish to sample from the product distribution $P^{\otimes n}$ by communicating a single message. In the setting of expected message length, letting R_n^* be the minimum expected message length over all exact samplers with proposal $Q^{\otimes n}$ and target $P^{\otimes n}$, the optimal asymptotic communication rate (bits per sample) is known to be [3]

$$\lim_{n \rightarrow \infty} \frac{R_n^*}{n} := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[l(\mathcal{C}(K))] = D(P \| Q).$$

Similarly, in the problem of i.i.d. channel simulation, where upon input x^n drawn via P_{X^n} one simulates a sample from the channel $P_{Y^n|X^n}(\cdot | x^n)$, the achievable rate is $\lim_{n \rightarrow \infty} R_n^*/n = I(X; Y)$. We can derive analogous asymptotic generalizations for the Campbell cost, $L(t)$.

Theorem 4: Given distributions $P \ll Q$, for any $t > 0$ let $L_n^*(t)$ be the minimum achievable Campbell cost among all exact sampling algorithms with target $P^{\otimes n}$ and common randomness $\{U_i\}_{i \geq 1} \sim Q^{\otimes n}$. Then, with $\alpha = \frac{1}{t+1}$,

$$\lim_{n \rightarrow \infty} \frac{L_n^*(t)}{n} = D_{\frac{1}{\alpha}}(P \| Q). \quad (38)$$

Proof: For any $n, t > 0$, and sampling algorithm, Theorem 1 yields the lower bound

$$\frac{L(t)}{n} \geq \frac{1}{n} D_{\frac{1}{\alpha}}(P^{\otimes n} \| Q^{\otimes n}) - \frac{1}{n} \left(\frac{\alpha}{1-\alpha} \log \alpha - 1 \right).$$

The additivity of the Rényi divergence for product distributions [19] gives $D_\alpha(P^{\otimes n} \| Q^{\otimes n}) = n D_\alpha(P \| Q)$, which implies the lower bound $\liminf_{n \rightarrow \infty} L_n^*(t)/n \geq D_{1/\alpha}(P \| Q)$. The upper bound on $L(t)$ from Theorem 2 gives that

$$\frac{L(t)}{n} \leq \frac{(1 + \epsilon)}{n} D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P^{\otimes n} \| Q^{\otimes n}) + \frac{c_1(\alpha, \epsilon)}{n},$$

for any $n, t > 0$, and $\epsilon > 0$. Taking $n \rightarrow \infty$, and again applying additivity, we obtain

$$\limsup_{n \rightarrow \infty} \frac{L_n^*(t)}{n} \leq (1 + \epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P \| Q). \quad (39)$$

As $\epsilon > 0$ in (39) is arbitrary, we can take $\epsilon \rightarrow 0$ to get that $\limsup_{n \rightarrow \infty} L_n^*(t)/n \leq D_{1/\alpha}(P \| Q)$. Here, we have used the continuity of the Rényi divergence in the order to conclude that $\lim_{\epsilon \rightarrow 0} D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P \| Q) = D_{1/\alpha}(P \| Q)$ [19]. The result follows. \square

As $\alpha \rightarrow 1$ in (38) we recover the optimal communication rate $\lim_{n \rightarrow \infty} R_n^*/n = D(P \| Q)$. In [17], Liu and Verdú showed that a causal sampler has a strictly worse expected α -moment of K than a noncausal sampler. In particular, they showed

(under some regularity assumptions) that for a causal sampler and any $\alpha \in (0, 1)$, $\mathbb{E}[K^\alpha] \approx 2^{\alpha D_{\frac{1}{1-\alpha}}(P \| Q)}$. In contrast, for a noncausal sampler, $\mathbb{E}[K^\alpha] \approx 2^{\alpha D_{\alpha+1}(P \| Q)}$. Under mild assumptions, we see a similar fundamental gap in the asymptotic Campbell cost between causal and noncausal sampling algorithms.

Theorem 5: Given distributions $P \ll Q$, for any $t > 0$ let $L_n^*(t)$ be the minimum achievable Campbell cost among all causal sampling algorithms between proposal $Q^{\otimes n}$ and target $P^{\otimes n}$ such that the transmitted index always has a nonincreasing probability distribution. If $t < 1$, we assume that $D_{\frac{1}{t} + \delta}(P \| Q)$ is finite for some $\delta > 0$. Then, with $\alpha = \frac{1}{t+1}$ we have that

$$\liminf_{n \rightarrow \infty} \frac{L_n^*(t)}{n} \geq D_\beta(P \| Q), \quad (40)$$

where

$$\beta = \begin{cases} \frac{\alpha}{2\alpha-1}, & \alpha \in (\frac{1}{2}, 1) \\ \infty, & \alpha \in (0, \frac{1}{2}]. \end{cases} \quad (41)$$

Proof: The proof can be found in Appendix E. \square

The assumption in Theorem 5 that K have a nonincreasing probability distribution is far less restrictive than it may first appear. All state-of-the-art exact causal sampling algorithms exhibit this property, including rejection sampling, greedy rejection sampling [3], [23], and greedy Poisson rejection sampling [24].

In the case of expected message length, causal and noncausal samplers achieve the same optimal communication rate of $D(P \| Q)$. With β as in (41), we have that $\beta > 1/\alpha$ and thus $D_\beta(P \| Q) > D_{1/\alpha}(P \| Q)$ in general. Interestingly, noncausal samplers have a strictly lower optimal asymptotic Campbell cost than causal samplers, and the gap is often significant. For example, if P and Q are Gaussians with different means, for $\alpha \leq 1/2$ a noncausal sampler can achieve an asymptotic cost of $D_{1/\alpha}(P \| Q)$, while the asymptotic Campbell cost per sample of a causal sampler, such as greedy rejection sampling or greedy Poisson rejection sampling, is infinite.

Intuitively, a noncausal sampler outperforms a causal sampler by “looking ahead” in the sequence of samples and resisting the temptation to select the current sample if a better one appears shortly afterwards. The Poisson functional representation appears to realize the advantage of noncausal samplers, and is near-optimal in the Campbell cost setting. Moreover, our results suggest the conjecture that the Poisson functional representation may even be optimal for exact sampling independent of the chosen cost.

VI. NUMERICAL EXAMPLES

In this section, we give numerical examples comparing LB_1^α , LB_2^α , UB_1^α , and UB_2^α from Theorems 1 to 3. Let P and Q be normal distributions on \mathbb{R} , specifically, for $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 \in \mathbb{R}_{\geq 0}$, let $P = \mathcal{N}(\mu_1, \sigma_1)$ and $Q = \mathcal{N}(\mu_2, \sigma_2)$. Trying to losslessly communicate a sample from a normal distribution is not possible; however, by allowing access to a shared source of randomness, we can communicate the sample with finite cost. Fig. 4 compares the bounds for three different choices of the normal distribution parameters; note that the true minimum $L(t)$ is guaranteed to lie between UB_1^α and $\max\{\text{LB}_1^\alpha, \text{LB}_2^\alpha\}$. We have only shown the bounds for $0.2 < \alpha < 1$ for display reasons, as both LB_1^α and UB_1^α go to ∞ as $\alpha \rightarrow 0$.

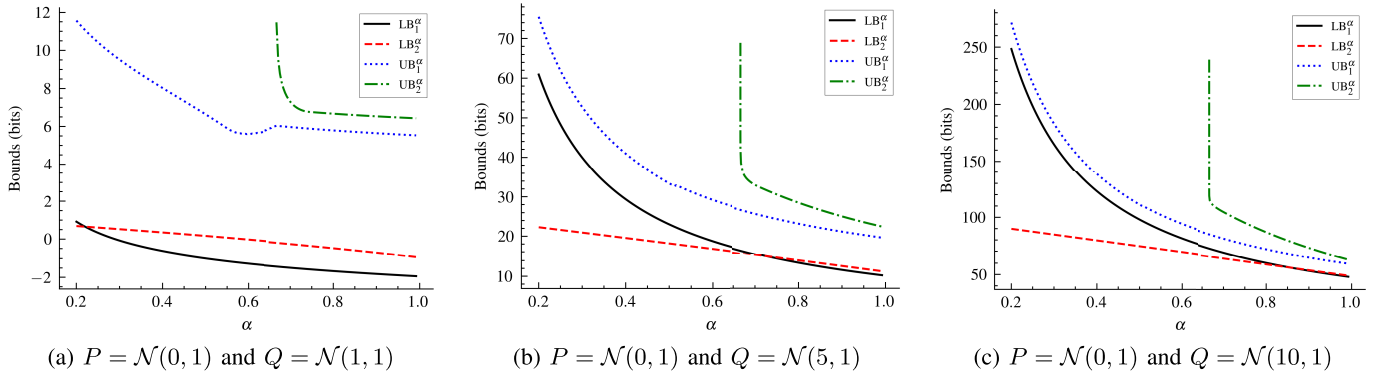


Fig. 4. Comparison of the bounds on the Campbell cost for two normal distributions, for $0.2 < \alpha < 1$.

In both upper bounds, for each value of α , the bound is optimized over ϵ to find the minimum value. As discussed in Section IV, in all cases UB_1^α is tighter than UB_2^α for all $2/3 < \alpha < 1$. Since $P = \mathcal{N}(\mu_1, \sigma_1)$ and $Q = \mathcal{N}(\mu_2, \sigma_2)$, we have that

$$D_\alpha(P||Q) = \ln\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\ln\left(\frac{\sigma_2^2}{(\sigma_2^2)_\alpha^*}\right)}{2(\alpha-1)} + \frac{1}{2} \frac{\alpha(\mu_1 - \mu_2)^2}{(\sigma_2^2)_\alpha^*},$$

for $(\sigma_2^2)_\alpha^* = \alpha\sigma_2^2 + (1-\alpha)\sigma_1^2$ [45, p. 45]; see also [16]. For $\mu_1 \neq \mu_2$, $D_{1/\alpha}(P||Q) \rightarrow \infty$ as $\alpha \rightarrow 0$, while $\lim_{\alpha \rightarrow 0} D_{2-\alpha}(P||Q) = D_2(P||Q) < \infty$. Hence, LB_1^α matches the asymptotic behaviour of $L(t)$ as $\alpha \rightarrow 0$, while LB_2^α does not. As discussed in Section III, the lesser constant term in LB_1^α compared to LB_2^α means that the lower bound in LB_2^α is tighter for α close to one and small $D_\alpha(P||Q)$; this can be seen in Fig. 4.

In some special cases, one can directly compute the distribution of K and thus $H_\alpha(K)$, and then bound the cost $L(t)$ as $L(t) < H_\alpha(K) + 1$. Indeed, (23) gives the explicit formula $K | U_K = u \sim \text{Geo}(\beta(u))$, with $\beta(u) = \mathbb{E}_{U \sim Q} \left[\max \left\{ \frac{dP}{dQ}(u), \frac{dP}{dQ}(U) \right\} \right]^{-1}$. When $P = \mathcal{N}(\mu_1, \sigma_1)$ and $Q = \mathcal{N}(\mu_2, \sigma_2)$ are normal distributions with respective densities p and q , we can write for $u \in \mathbb{R}$ that

$$\begin{aligned} \beta(u)^{-1} &= \mathbb{E}_{U \sim Q} \left[\max \left\{ \frac{dP}{dQ}(u), \frac{dP}{dQ}(U) \right\} \right] \\ &= \int_{-\infty}^{\infty} \frac{dP}{dQ}(\min\{x, u\}) q(x) dx \\ &= \int_{-\infty}^u \frac{dP}{dQ}(x) q(x) dx + \int_u^{\infty} \frac{dP}{dQ}(u) q(x) dx \\ &= \int_{-\infty}^u p(x) dx + \frac{dP}{dQ}(u) \int_u^{\infty} q(x) dx \\ &= \Phi_P(u) + \frac{dP}{dQ}(u) (1 - \Phi_Q(u)), \end{aligned}$$

with Φ_P and Φ_Q the cumulative distribution functions of P and Q , respectively. We then obtain that

$$\begin{aligned} \mathbb{P}(K = k) &= \mathbb{E}_{U_K \sim P} [\mathbb{P}(K = k | U_K)] \\ &= \int_{-\infty}^{\infty} (1 - \beta(u))^{k-1} \beta(u) p(u) du. \end{aligned} \quad (42)$$

Eq. (42) gives an explicit formula for the distribution of K , which can be computed using numerical quadrature. From

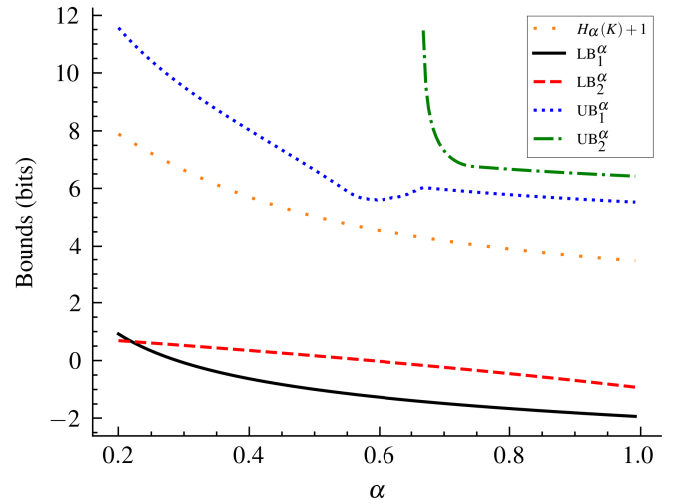


Fig. 5. Comparison of the bounds on the Campbell cost and $H_\alpha(K)$ for $P = \mathcal{N}(0, 1)$ and $Q = \mathcal{N}(1, 1)$, for $0.2 < \alpha < 1$.

this expression, we can closely approximate the Rényi entropy $H_\alpha(K)$ as

$$H_\alpha(K) \approx \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N \mathbb{P}(K = k)^\alpha \right), \quad (43)$$

for some large N . Fig. 5 compares the bounds with $H_\alpha(K) + 1$, approximated as in (43) with $N = 1000$.¹ As seen in Fig. 5, the obtained upper bound is tighter than UB_1^α or UB_2^α . However, this is a special case where P and Q are close together, and therefore the probability distribution of K is concentrated on low values. In this example, $1 - \sum_{k=1}^{1000} \mathbb{P}(K = k) < 2 \times 10^{-8}$, whereas for $P = \mathcal{N}(0, 1)$ and $Q = \mathcal{N}(5, 1)$, $1 - \sum_{k=1}^{1000} \mathbb{P}(K = k) > 0.83$. The distribution of K becomes exceedingly heavy-tailed, making accurately estimating $H_\alpha(K)$ computationally intractable. In contrast, UB_1^α and UB_2^α can be computed instantaneously, even for complex distributions that do not admit a simple form of $\mathbb{P}(K = k)$.

As another numerical example, let P, Q be Laplacian distributions, i.e., for $\theta_1, \theta_2 \in \mathbb{R}$ and $\lambda_1, \lambda_2 > 0$ we have $P = \text{Laplace}(\theta_1, \lambda_1)$ and $Q = \text{Laplace}(\theta_2, \lambda_2)$, where $\text{Laplace}(\theta, \lambda)$

¹The error in this approximation of $H_\alpha(K)$, for $P = \mathcal{N}(0, 1)$ and $Q = \mathcal{N}(1, 1)$, is less than 0.02 bits for all values of α .

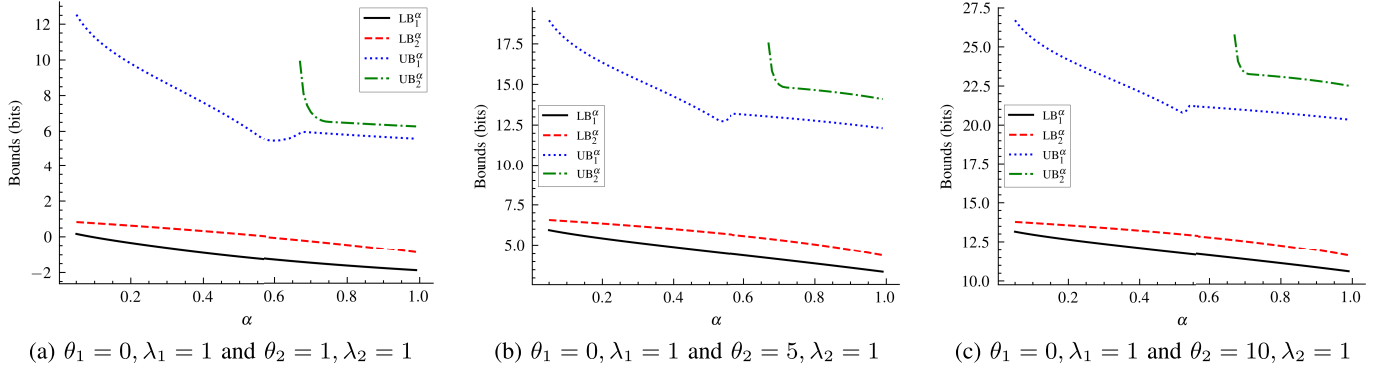


Fig. 6. Comparison of the bounds on the Campbell cost for $P = \text{Laplace}(\theta_1, \lambda_1)$ and $Q = \text{Laplace}(\theta_2, \lambda_2)$, for $0.05 < \alpha < 1$.

denotes a Laplacian distribution with mean θ and variance $2\lambda^2$. For $\alpha > 0$ with $\alpha \neq \frac{\lambda_1}{\lambda_1 + \lambda_2}$, [16] gives the formula for $D_\alpha(P\|Q)$ as

$$D_\alpha(P\|Q) = \begin{cases} \ln \frac{\lambda_2}{\lambda_1} + \frac{1}{\alpha-1} \ln \left(\frac{\lambda_1 \lambda_2^2 g(\alpha)}{\alpha^2 \lambda_2^2 - (1-\alpha)^2 \lambda_1^2} \right), & \alpha \lambda_2 + (1-\alpha)\lambda_1 > 0, \\ \infty, & \alpha \lambda_2 + (1-\alpha)\lambda_1 \leq 0, \end{cases} \quad (44)$$

where

$$g(\alpha) := \frac{\alpha}{\lambda_1} \exp\left(-\frac{(1-\alpha)|\theta_1 - \theta_2|}{\lambda_2}\right) - \frac{1-\alpha}{\lambda_2} \exp\left(-\frac{\alpha|\theta_1 - \theta_2|}{\lambda_1}\right).$$

Fig. 6 compares the bounds of Theorems 1 to 3 for the Laplacian distributions P and Q . Observe that, in each case, LB_2^α is tighter, despite having a leading Rényi divergence term of $D_{2-\alpha}(P\|Q)$. As seen in (44), the value of α has less impact on its Rényi divergence (than in the case of normal distributions), and therefore, the larger constant term makes LB_2^α tighter than LB_1^α . Moreover, in contrast to when P and Q are normal distributions, for $P = \text{Laplace}(\theta_1, \lambda_1)$ and $Q = \text{Laplace}(\theta_2, \lambda_2)$, if $\lambda_2 \geq \lambda_1$ then $\lim_{\alpha \rightarrow \infty} D_\alpha(P\|Q) < \infty$. Hence, the asymptotics of LB_1^α and LB_2^α align more closely as $\alpha \rightarrow 0$ than in the case of Gaussians. In all three cases in Fig. 6, observe that LB_2^α and UB_1^α are still within 5-10 bits for most α , diverging only as $\alpha \rightarrow 0$.

VII. CONCLUSION

In this paper, we have generalized bounds on the expected communication cost of one-shot exact sampling to the Campbell cost and Rényi's entropy. Such bounds on the exponential cost are useful in situations where one wishes to disproportionately penalize long codewords, such as applications with buffer overflow. Several interesting directions for future work stem from these results. Most obviously, it is an open question if UB_1^α can be tightened to be in the spirit of [2], specifically so (15) reduces to $\mathbb{E}[I(\mathcal{C}(K))] \leq D(P\|Q) + \log(D(P\|Q) + 1) + O(1)$ as $\alpha \rightarrow 1$. We also conjecture that the asymptotic gap between causal and noncausal samplers shown in Section V still holds in the one-shot case and without the requirement of a nonincreasing probability distribution. In the broader field of channel simulation, the upper bound in Corollary 2 suggests the question of whether $I_{1/\alpha}^\alpha(X; Y)$ is a lower bound

on the Campbell cost of *any* channel simulation algorithm. While most state-of-the-art channel simulation algorithms rely on sampling, there are other, fundamentally different, protocols [34]. Proving such a lower bound would characterize the optimal asymptotic Campbell rate of channel simulation under known and worst-case input. More generally, the strong functional representation lemma has been applied to several problems outside of one-shot exact channel simulation, most notably one-shot variable-length lossy source coding and multiple description coding [2]. It would be interesting to see if the strong functional representation lemma can play a role in information-theoretic coding problems using the Campbell cost or Rényi entropy.

APPENDIX

A. Proof of Lemma 1

We first let $g(k) = k$, i.e., we show that

$$E[K^\alpha] \geq \frac{1}{1+\alpha} 2^{\alpha D_{\alpha+1}(P\|Q)}. \quad (45)$$

To do so, we will first show that for any $k \in \mathbb{N}$, $\mathbb{P}(K = k | U_K = u) \leq \frac{dP}{dQ}(u)$ for P -almost every $u \in \mathcal{U}$. Define ν to be the probability measure generated by the joint distribution of K and U_K , i.e., $\nu(A) := \mathbb{P}((K, U_K) \in A)$ for each Borel set $A \subseteq \mathbb{N} \times \mathcal{U}$. Then, defining $\nu_k(B) := \nu(k \times B)$, for any Borel set $B \subseteq \mathcal{U}$ we get that

$$\begin{aligned} \int_B 1 dQ(u) &= Q(B) \\ &= \sum_{k=1}^{\infty} \nu_k(B) \\ &\geq \nu_k(B) \\ &= \int_B \mathbb{P}(K = k | U_K = u) dP(u) \\ &= \int_B \mathbb{P}(K = k | U_K = u) \frac{dP}{dQ}(u) dQ(u). \end{aligned}$$

Thus,

$$\mathbb{P}(K = k | U_K = u) \leq \frac{1}{\frac{dP}{dQ}(u)} \quad (46)$$

for P -almost every $u \in \mathcal{U}$. Let $E \subseteq \mathcal{U}$ be the set of all u where (46) holds. For $u \in E$, writing $p_u(k) = \mathbb{P}(K = k | U_K = u)$ and

$c_u = (dP/dQ)(u)$, we have that $p_u(k) \leq \frac{1}{c_u}$ for all k . Then, for $c_u := c_u - \lfloor c_u \rfloor$ the fractional part of c_u , we claim

$$\sum_{k=1}^{\infty} k^\alpha p_u(k) \geq \frac{1}{c_u} \left(\sum_{k=1}^{\lfloor c_u \rfloor} k^\alpha + \{c_u\} (\lfloor c_u \rfloor)^\alpha \right). \quad (47)$$

If $p_u(k)$ does not have distribution $p_u(k) = \frac{1}{c_u}$ for $k \in \{1, \dots, \lfloor c_u \rfloor\}$ and $p_u(\lfloor c_u \rfloor) = c_u / c_u$, we can move any excess probability from indices $k \geq \lfloor c_u \rfloor$ until it has such a distribution. Since we shift probabilities from indices $k \geq \lfloor c_u \rfloor$ to $k \leq \lfloor c_u \rfloor$ and k^α is increasing in k , (47) must hold. Hence, for all $u \in E$,

$$\begin{aligned} \mathbb{E}[K^\alpha | U_K = u] &= \sum_{k=1}^{\infty} k^\alpha p_u(k) \\ &\geq \frac{1}{c_u} \left(\sum_{k=1}^{\lfloor c_u \rfloor} k^\alpha + \{c_u\} (\lfloor c_u \rfloor)^\alpha \right) \\ &\geq \frac{1}{c_u} \left(\int_0^{c_u} t^\alpha dt \right) \\ &= \frac{1}{1+\alpha} c_u^\alpha = \frac{1}{1+\alpha} \left[\frac{dP}{dQ}(u) \right]^\alpha, \end{aligned} \quad (48)$$

where the inequality in (48) follows because $f(t) = t^\alpha$ is an increasing function of t . Taking expectation over $U \sim P$ gives, as $P(E) = 1$, the lower bound in (45).

For an arbitrary bijection $g : \mathbb{N} \rightarrow \mathbb{N}$, note that for every $k \in \mathbb{N}$, (46) implies

$$\begin{aligned} \mathbb{P}(g(K) = k | U_K = u) &= \mathbb{P}(K = g^{-1}(k) | U_K = u) \\ &\leq \frac{1}{\frac{dP}{dQ}(u)}, \end{aligned}$$

for P -almost every $u \in \mathcal{U}$. Following the same steps as above, we can conclude that $\mathbb{E}[(g(K))^\alpha] \geq \frac{1}{1+\alpha} 2^{\alpha D_{\alpha+1}(P\|Q)}$. \square

B. Proof of Proposition 2

We aim to show that, for any $\epsilon > 0$ and $0 < \alpha < 1$,

$$\begin{aligned} c_1(\alpha, \epsilon) &< \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \log \left(\frac{1}{\alpha} \right) \\ &\quad + \log \left(\frac{1}{\epsilon} \right) + 1.5\epsilon^2 + 4.5\epsilon + 2.6. \end{aligned}$$

First let $1/2 < \alpha < 1$ and $\epsilon < \frac{2\alpha-1}{1-\alpha}$ (we are in the first branch of $c_1(\alpha, \epsilon)$) and note that

$$\begin{aligned} &(1+\epsilon) \log e + 1 + \log \left(1 + \frac{1}{\epsilon} \right) \\ &= \log \left(\frac{1}{\epsilon} \right) + \log(\epsilon + 1) + (1+\epsilon) \log e + 1 \\ &\leq \log \left(\frac{1}{\epsilon} \right) + 2\epsilon \log e + \log e + 1 \\ &< \log \left(\frac{1}{\epsilon} \right) + 3\epsilon + 2.6, \end{aligned} \quad (49)$$

where (49) follows from $\log(x+1) \leq x \log e$ for any $x > 0$. As $\epsilon^2 > 0$ and $\log(\frac{1}{\alpha}) > 0$ for $\alpha > 1/2$, the proposition holds for this case.

Let now $0 < \alpha < 1/2$ or $\epsilon \geq \frac{2\alpha-1}{1-\alpha}$. Let $x = \frac{1+\epsilon(1-\alpha)}{1-\alpha}$. Note that the domain assumption on α and ϵ guarantees $x \geq 2$. Using

the Taylor series bound $\log \Gamma(x) < (x-1/2) \log x + (1-x) \log e$, valid for $x \geq 1$, we obtain that

$$\begin{aligned} &\frac{\alpha}{1-\alpha} \log \Gamma(x) \\ &< \frac{\alpha}{1-\alpha} \left[\left(\frac{1}{\alpha} + \epsilon \frac{1-\alpha}{\alpha} - \frac{1}{2} \right) \log \left(\frac{1+\epsilon(1-\alpha)}{\alpha} \right) \right. \\ &\quad \left. + \left(1 - \frac{1}{\alpha} - \epsilon \frac{1-\alpha}{\alpha} \right) \log e \right] \\ &= \left(\frac{1+1-\alpha}{2(1-\alpha)} + \epsilon \right) \left(\log \left(\frac{1}{\alpha} \right) + \log(1+\epsilon(1-\alpha)) \right) \\ &\quad + \left(\frac{\alpha}{1-\alpha} - \frac{1}{1-\alpha} - \epsilon \right) \log e \\ &\leq \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \log \left(\frac{1}{\alpha} \right) \\ &\quad + \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \epsilon(1-\alpha) \log e - (1+\epsilon) \log e \\ &= \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \log \left(\frac{1}{\alpha} \right) \\ &\quad - \log e - \frac{\alpha\epsilon}{2} \log e + (1-\alpha)\epsilon^2 \log e. \end{aligned} \quad (50)$$

where (50) again follows from $\log(x+1) \leq x \log e$. Combining this bound with the remaining terms of $c_1(\alpha, \epsilon)$ gives, after collecting like terms,

$$\begin{aligned} c_1(\alpha, \epsilon) &< \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \log \left(\frac{1}{\alpha} \right) + \log \left(\frac{1}{\epsilon} \right) \\ &\quad + \left(4 - \log e - \frac{2\alpha}{1-\alpha} \right) \\ &\quad + \epsilon \left(3 + \left(1 - \frac{\alpha}{2} \right) \log e \right) + \epsilon^2(1-\alpha) \log e \\ &< \left(\frac{1}{2(1-\alpha)} + \frac{1}{2} + \epsilon \right) \log \left(\frac{1}{\alpha} \right) + \log \left(\frac{1}{\epsilon} \right) \\ &\quad + 1.5\epsilon^2 + 4.5\epsilon + 2.6. \end{aligned}$$

The result follows. \square

C. Proof of Lemma 3

The case $r = 1$ holds trivially, as $\mathbb{E}[X] = \frac{1}{p}$. Let $r > 1$. We know that $X = \lceil Y \rceil$, where Y is an exponential random variable with parameter $\lambda = -\ln(1-p)$. Thus, $X \leq Y + 1$ and we have

$$\mathbb{E}[Y^r] = \int_0^\infty y^r \lambda e^{-\lambda y} dy = \int_0^\infty \left(\frac{x}{\lambda} \right)^r e^{-x} dx = \frac{\Gamma(r+1)}{\lambda^r}.$$

Further, note that for any $r > 1$ and $x > 0$, $(x+1)^r \leq 2^{r-1}(x^r + 1)$. We can therefore upper bound $\mathbb{E}[X^r]$ as

$$\begin{aligned} \mathbb{E}[X^r] &\leq \mathbb{E}[(Y+1)^r] \leq 2^{r-1} (\mathbb{E}[Y^r] + 1) \\ &= 2^{r-1} \left(\frac{\Gamma(r+1)}{\lambda^r} + 1 \right) \\ &= 2^{r-1} \left(\frac{\Gamma(r+1)}{[-\ln(1-p)]^r} + 1 \right) \\ &\leq 2^{r-1} \left(\frac{\Gamma(r+1)}{p^r} + 1 \right), \end{aligned} \quad (51)$$

where (51) follows because $p \leq -\ln(1-p)$ for any $0 < p < 1$ and $x \mapsto x^r$ is an increasing function. \square

D. Constant Term Analysis

We herein prove the existence of the uniquely decodable encodings of K used in the proofs of Theorems 2 and 3. First, in Theorem 2, we claim that there exists a prefix code with codeword lengths $n_k = (1 + \epsilon) \log k + c_\epsilon$, for any $\epsilon > 0$ and $c_\epsilon \leq 1 + \log(1 + \frac{1}{\epsilon})$ a constant such that n_k is an integer and the code is a valid prefix code. The existence of such a code is guaranteed by the Kraft inequality, since for all $\epsilon > 0$ there exists c_ϵ such that

$$\sum_{k=1}^{\infty} \frac{1}{k^{(1+\epsilon)}} \leq 2^{c_\epsilon}. \quad (52)$$

We can upper bound the sum in (52) by

$$\sum_{k=1}^{\infty} \frac{1}{k^{(1+\epsilon)}} \leq 1 + \int_1^{\infty} x^{-(1+\epsilon)} dx = 1 + \frac{1}{\epsilon}.$$

Returning to the requirement that n_k be an integer, we can take the ceiling of our codeword length, imposing at most a + 1 penalty, to get that

$$c_\epsilon \leq 1 + \log\left(1 + \frac{1}{\epsilon}\right),$$

as claimed. Similarly, in Theorem 3 we consider a universal code with codeword lengths $n_k = \log k + (1 + \epsilon) \log \log(k + 1) + c_\epsilon$, for any $\epsilon > 0$ and $c_\epsilon \leq 1 + \log\left(\frac{\ln(2)}{\epsilon} + \frac{3}{2}\right)$. By the Kraft inequality, we require that

$$1 \geq \sum_{k=1}^{\infty} 2^{-n_k} = \sum_{k=1}^{\infty} \frac{1}{k \log^{(1+\epsilon)}(k+1) 2^{c_\epsilon}}.$$

Hence, setting

$$c_\epsilon = \log\left(\sum_{k=1}^{\infty} \frac{1}{k \log^{(1+\epsilon)}(k+1)}\right) < \infty, \quad (53)$$

the code satisfies the Kraft inequality with equality. We can upper bound the sum in (53) by

$$\begin{aligned} & \sum_{k=1}^{\infty} \frac{1}{k \log^{(1+\epsilon)}(k+1)} \\ &= \sum_{k=1}^{\infty} \left[\frac{1}{(k+1) \log^{\epsilon+1}(k+1)} + \frac{1}{k(k+1) \log^{1+\epsilon}(k+1)} \right] \\ &\leq \frac{1}{2} + \int_1^{\infty} \frac{1}{(x+1) \log^{(1+\epsilon)}(x+1)} dx + \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \\ &\leq \frac{1}{2} + \frac{\ln(2)}{\epsilon} + 1 = \frac{3}{2} + \frac{\ln(2)}{\epsilon}. \end{aligned}$$

Again imposing at most a + 1 penalty to ensure n_k is an integer, we obtain that

$$c_\epsilon \leq 1 + \log\left(\frac{3}{2} + \frac{\ln(2)}{\epsilon}\right).$$

E. Proof of Theorem 5

First assume that $t < 1$ (consequently $\alpha \in (1/2, 1)$). Let K_n be the communicated index when the proposal and target

distributions are $P^{\otimes n}$ and $Q^{\otimes n}$, respectively. We first wish to prove that for any $\epsilon > 0$,

$$\liminf_{n \rightarrow \infty} \frac{1}{nt} \log(\mathbb{E}[K_n^t]) \geq D_{\frac{1}{1-t}}(P||Q) - \frac{2\epsilon}{t}. \quad (54)$$

Our proof follows closely [17], but contains details omitted there. Define the tilted distribution R_n by

$$\frac{dR_n}{dQ^{\otimes n}} := \frac{\exp\left(\frac{1}{1-t} \log \frac{dP^{\otimes n}}{dQ^{\otimes n}}\right)}{\int \exp\left(\frac{1}{1-t} \log \frac{dP^{\otimes n}}{dQ^{\otimes n}}\right) dQ^{\otimes n}}.$$

Define the random variable

$$X_n = \log \frac{dP^{\otimes n}}{dQ^{\otimes n}}(Z_n) = \sum_{i=1}^n \log \frac{dP}{dQ}(Z_i),$$

for $Z_n \sim R_n$ and $Z \sim R$ (the one-dimensional tilted distribution). Fix $\epsilon > 0$ and define

$$\begin{aligned} \tau_n &:= \mathbb{E}[X_n] - \epsilon n; \\ q_n &:= \int \exp\left(\frac{1}{1-t} \log \frac{dP^{\otimes n}}{dQ^{\otimes n}}\right) dQ^{\otimes n} \\ &= \exp\left(\frac{t}{1-t} D_{\frac{1}{1-t}}(P^{\otimes n}||Q^{\otimes n})\right). \end{aligned}$$

Then,

$$\begin{aligned} & Q^{\otimes n} \left(\left\{ \tau_n < \log \frac{dP^{\otimes n}}{dQ^{\otimes n}} \right\} \right) \\ &\geq Q^{\otimes n} \left(\left\{ \tau_n < \log \frac{dP^{\otimes n}}{dQ^{\otimes n}} < \tau_n + 2\epsilon n \right\} \right) \\ &\geq q_n \exp\left(-\frac{\tau_n + 2\epsilon n}{1-t}\right) \times \\ &\quad \times R_n \left(\left\{ \tau_n < \log \frac{dP^{\otimes n}}{dQ^{\otimes n}} < \tau_n + 2\epsilon n \right\} \right) \\ &= q_n \exp\left(-\frac{\tau_n + 2\epsilon n}{1-t}\right) \mathbb{P}(|X_n - \mathbb{E}[X_n]| < \epsilon n). \end{aligned}$$

Theorem 3 in [17] proves that

$$\mathbb{E}[K_n^t] \geq (1-t)t^t \sup_{\tau \in \mathbb{R}} \exp(\tau) \left[Q^{\otimes n} \left(\log \frac{dP^{\otimes n}}{dQ^{\otimes n}} > \tau \right) \right]^{1-t}.$$

Choosing $\tau = \tau_n$, noting that we require $\mathbb{E}[|X_n|] < \infty$ (which will be established subsequently), we obtain that

$$\begin{aligned} \mathbb{E}[K_n^t] &\geq (1-t)t^t \exp(\tau_n) q_n^{1-t} \exp\left(-\frac{\tau_n + 2\epsilon n}{1-t} \cdot 1-t\right) \times \\ &\quad \times [\mathbb{P}(|X_n - \mathbb{E}[X_n]| < \epsilon n)]^{1-t} \\ &= (1-t)t^t \exp\left(t D_{\frac{1}{1-t}}(P^{\otimes n}||Q^{\otimes n})\right) \exp(-2\epsilon n) \times \\ &\quad \times [\mathbb{P}(|X_n - \mathbb{E}[X_n]| < \epsilon n)]^{1-t}. \end{aligned}$$

Therefore, taking logarithms and dividing by nt we get that (using the additivity of the Rényi divergence for product distributions)

$$\begin{aligned} & \frac{1}{nt} \log(\mathbb{E}[K_n^t]) \\ &\geq D_{\frac{1}{1-t}}(P||Q) + \frac{1}{nt} \log((1-t)t^t) - \frac{2\epsilon}{t} \\ &\quad + \frac{1-t}{t} \log(\mathbb{P}(|X_n - \mathbb{E}[X_n]| < \epsilon n)). \end{aligned} \quad (55)$$

We can rewrite the probability in (55) as

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{i=1}^n \log \frac{dP}{dQ}(Z_i) - n \mathbb{E}_{Z \sim R} \left[\log \frac{dP}{dQ}(Z) \right] \right| < \epsilon n \right) \\ &= \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \log \frac{dP}{dQ}(Z_i) - \mathbb{E}_{Z \sim R} \left[\log \frac{dP}{dQ}(Z) \right] \right| < \epsilon \right). \end{aligned}$$

Again with the assumption that $\mathbb{E}[|X_n|] < \infty$, the weak law of large numbers (e.g. [46, Cor. 5.4.11]) gives that the probability in (55) tends to 1; (54) follows.

Now, as K_n has a nonincreasing probability distribution by assumption, a lower bound on its encoding cost is the optimal one-to-one code with codeword lengths $\lfloor \log(k+1) \rfloor$ for every $k \in \mathbb{N}$. Therefore,

$$\begin{aligned} L_n^*(t) &\geq \frac{1}{t} \log \left(\sum_{k=1}^{\infty} \mathbb{P}(K_n = k) 2^{\lfloor \log(k+1) \rfloor} \right) \\ &\geq \frac{1}{t} \log \left(\sum_{k=1}^{\infty} \mathbb{P}(K_n = k) 2^{t(\log(k+1)-1)} \right) \\ &\geq \frac{1}{t} \log(\mathbb{E}[K_n^t]) - 1. \end{aligned} \quad (56)$$

Dividing both sides by n and taking $n \rightarrow \infty$ gives that, for any $\epsilon > 0$,

$$\liminf_{n \rightarrow \infty} \frac{L_n^*(t)}{n} \geq \liminf_{n \rightarrow \infty} \frac{1}{nt} \log(\mathbb{E}[K_n^t]) \geq D_{\frac{1}{1-t}}(P||Q) - \frac{2\epsilon}{t}.$$

Noting that $\alpha = \frac{1}{1-t}$ and taking $\epsilon \rightarrow 0$ yields the asymptotic inequality.

We now turn to the assumption that $\mathbb{E}[|X_n|] < \infty$, which is equivalent to $\mathbb{E}_{Z \sim R} \left[\left| \log \frac{dP}{dQ}(Z) \right| \right] < \infty$. We will show this follows from our assumption that $D_{\frac{1}{1-t}+\delta}(P||Q) < \infty$ for some $\delta > 0$. Using a change of measure, we can write

$$\begin{aligned} \int \left| \log \frac{dP}{dQ} \right| dR &= \int \left| \log \frac{dP}{dQ} \right| \frac{\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}}}{\mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}} \right]} dQ \\ &= \frac{\mathbb{E}_Q \left[\left| \log \frac{dP}{dQ} \right| \left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}} \right]}{\mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}} \right]}. \end{aligned} \quad (57)$$

Note that the denominator in (57) causes no problems because $\mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}} \right] = \exp \left(\frac{t}{1-t} D_{\frac{1}{1-t}}(P||Q) \right) < \infty$. We see that, for any $\delta \in (0, 1]$,

$$\begin{aligned} & \mathbb{E}_Q \left[\left| \log \frac{dP}{dQ} \right| \left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}} \right] \\ &\leq \frac{1}{\delta \ln 2} \mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}+\delta} \mathbb{1}_{\left\{ \frac{dP}{dQ} \geq 1 \right\}} \right] \\ &\quad + \frac{1}{\delta \ln 2} \mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}-\delta} \mathbb{1}_{\{0 < \frac{dP}{dQ} < 1\}} \right] \\ &\leq \frac{1}{\delta \ln 2} \left(\mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\frac{1}{1-t}+\delta} \right] + 1 \right) \end{aligned}$$

$$= \frac{1}{\delta \ln 2} \left(\exp \left(\left(\delta + \frac{t}{1-t} \right) D_{\frac{1}{1-t}+\delta}(P||Q) \right) \right) < \infty \quad (58)$$

Here, (58) follows from the inequalities, for any $x \geq 0$ and $\delta \in (0, 1]$,

$$(\log x)^+ \leq \frac{x^\delta}{\delta \ln 2} \mathbb{1}_{\{x \geq 1\}}, \quad (\log x)^- \leq \frac{x^{-\delta}}{\delta \ln 2} \mathbb{1}_{\{0 < x < 1\}}.$$

Finally, for $t \geq 1$, equivalently $\alpha \in (0, 1/2]$, we use the lower bound on $\mathbb{E}[K]$ of [7, Thm. III.1] to obtain that

$$\mathbb{E}[K_n^t] \geq \mathbb{E}[K_n]^t \geq 2^{tD_\infty(P^{\otimes n}||Q^{\otimes n})} = 2^{tnD_\infty(P||Q)},$$

where the first inequality follows from Jensen's inequality. Using the same one-to-one optimal encoding of K and steps as in the case $t < 1$, the result follows. \square

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their helpful comments and feedback.

REFERENCES

- [1] L. L. Campbell, "A coding theorem and Rényi's entropy," *Inf. Control*, vol. 8, no. 4, pp. 423–429, 1965.
- [2] C. T. Li and A. E. Gamal, "Strong functional representation lemma and applications to coding theorems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 589–593.
- [3] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 438–449, Jan. 2010.
- [4] M. Braverman and A. Garg, "Public vs private coin in bounded-round information," in *Proc. Int. Colloq. Automata, Lang., Program.*, 2014, pp. 502–513.
- [5] G. Flamich, M. Havasi, and J. M. Hernández-Lobato, "Compressing images by encoding their latent representations with relative entropy coding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 16131–16141.
- [6] L. Theis and N. Y. Ahmed, "Algorithms for the communication of samples," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 21308–21328.
- [7] D. Goc and G. Flamich, "On channel simulation with causal rejection samplers," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2024, pp. 1682–1687.
- [8] S. M. Sriram and A. B. Wagner, "Optimal redundancy in exact channel synthesis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2024, pp. 1913–1918.
- [9] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 490–501, May 1968.
- [10] P. Humblet, "Generalization of Huffman coding to minimize the probability of buffer overflow (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 2, pp. 230–232, Mar. 1981.
- [11] A. C. Blumer and R. J. McEliece, "The Rényi redundancy of generalized Huffman codes," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 5, pp. 1242–1249, Sep. 1988.
- [12] A. Somazzi, P. Ferragina, and D. Garlaschelli, "On nonlinear compression costs: When Shannon meets Rényi," *IEEE Access*, vol. 12, pp. 77750–77763, 2024.
- [13] T. A. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2494–2498.
- [14] T. A. Courtade and S. Verdú, "Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 2499–2503.
- [15] S. Saito and T. Matsushima, "Non-asymptotic bounds of cumulant generating function of codeword lengths in variable-length lossy compression," *IEEE Trans. Inf. Theory*, vol. 69, no. 4, pp. 2113–2119, Apr. 2023.
- [16] M. Gil, F. Alajaji, and T. Linder, "Rényi divergence measures for commonly used univariate continuous distributions," *Inf. Sci.*, vol. 249, pp. 124–131, Nov. 2013.

- [17] J. Liu and S. Verdú, "Rejection sampling and noncausal sampling under moment constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1565–1569.
- [18] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1961, pp. 547–562.
- [19] T. Van Erven and P. Harremoës, "Rényi divergence and Kullback–Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.
- [20] S. Verdú, "A-mutual information," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 1–6.
- [21] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.
- [22] C. T. Li, *Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy*, vol. 21. Boston, MA, USA: Now Publishers, 2024.
- [23] G. Flamich and L. Theis, "Adaptive greedy rejection sampling," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2023, pp. 454–459.
- [24] G. Flamich, "Greedy Poisson rejection sampling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 37089–37127.
- [25] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.
- [26] A. Winter, "Compression of sources of probability distributions and density operators," 2002, *arXiv:quant-ph/0208131*.
- [27] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.
- [28] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.
- [29] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.
- [30] L. Yu and V. Y. F. Tan, "Exact channel synthesis," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2799–2818, May 2020.
- [31] G. Flamich, S. M. Sriramu, and A. B. Wagner, "The redundancy of non-singular channel simulation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2025, pp. 1–6.
- [32] G. Flamich, S. Markou, and J. M. Hernández-Lobato, "Fast relative entropy coding with A* coding," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 6548–6577.
- [33] G. Flamich, J. M. Hernández-Lobato, and S. Markou, "Faster relative entropy coding with greedy rejection coding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 50558–50569.
- [34] R. Barsz, E. Polito, S. Sriramu, and A. Wagner, "Fast channel simulation via error-correcting codes," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, 2024, pp. 107932–107959.
- [35] C. T. Li, "Pointwise redundancy in one-shot lossy compression via Poisson functional representation," 2024, *arXiv:2401.14805*.
- [36] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the Poisson matching lemma," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2624–2651, May 2021.
- [37] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [38] P. Elias, "Universal codeword sets and representations of the integers," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 194–203, Mar. 1975.
- [39] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, vol. 3. New York, NY, USA: Springer, 2008.
- [40] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Topics Inf. Theory*, vol. 29, no. 1, pp. 41–52, 1977.
- [41] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.
- [42] B. Škorić, C. Obi, E. Verbitskiy, and B. Schoenmakers, "Sharp lower bounds on the extractable randomness from non-uniform sources," *Inf. Comput.*, vol. 209, no. 8, pp. 1184–1196, Aug. 2011.
- [43] A. Kamatsuka and T. Yoshida, "Several representations of α -mutual information and interpretations as privacy leakage measures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2025, pp. 1–6.
- [44] I. Csiszar, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995.
- [45] F. Liese and I. Vajda, *Convex Statistical Distances*. Leipzig, Germany: Teubner, 1987.
- [46] J. S. Rosenthal, *A First Look at Rigorous Probability Theory (2nd Edition)*. Singapore: World Scientific, 2006.

Spencer Hill (Student Member, IEEE) received the B.A.Sc. degree in mathematics and engineering from Queen's University, Canada, in 2023, where he is currently pursuing the M.A.Sc. degree in mathematics and engineering. His research interests include information theory, data compression, machine learning, and reinforcement learning.

Fady Alajaji (Senior Member, IEEE) received the B.E. degree (Hons.) in electrical engineering from American University of Beirut, Lebanon, in 1988, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1990 and 1994, respectively. In 1994, he held a post-doctoral appointment with the Institute for Systems Research, University of Maryland. In 1995, he joined the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, where he is currently a Professor of mathematics and engineering. Since 1997, he has been cross appointed with the Department of Electrical and Computer Engineering, Queen's University. From 2013 to 2014, he was the Acting Head of the Department of Mathematics and Statistics. From 2003 to 2008 and from 2018 to 2019, he was the Chair of the Queen's Mathematics and Engineering Program. His research interests include information theory, digital communications, error control coding, data compression, joint source-channel coding, network epidemics, generative models, data privacy, and fairness in machine learning. He received the Premiers Research Excellence Award from the Province of Ontario. He served as an organizer and a technical program committee member for several international conferences and workshops. From 2019 to 2022, he was an Associate Editor of Shannon Theory and Information Measures for IEEE TRANSACTIONS ON INFORMATION THEORY. He served as an Area Editor (2008–2015) for Source-Channel Coding and Signal Processing and an Editor (2003–2012) for Source and Source-Channel Coding for IEEE TRANSACTIONS ON COMMUNICATIONS.

Tamás Linder (Fellow, IEEE) received the M.S. degree in electrical engineering from the Technical University of Budapest, Hungary, in 1988, and the Ph.D. degree in electrical engineering from Hungarian Academy of Sciences in 1992. He was a Post-Doctoral Researcher with the University of Hawaii in 1992, and a Visiting Fulbright Scholar with the Coordinated Science Laboratory, University of Illinois Urbana–Champaign from 1993 to 1994. From 1994 to 1998, he was a Faculty Member with the Department of Computer Science and Information Theory, Technical University of Budapest. From 1996 to 1998, he was a Visiting Research Scholar with the Department of Electrical and Computer Engineering, University of California, San Diego. In 1998, he joined Queen's University, where he is currently a Professor of mathematics and engineering with the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition. He received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellor's Research Award of Queen's University in 2003. He was an Associate Editor of Source Coding for IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2004.