Distributed Task Encoding

Annina Bracher
Group Risk Management
Swiss Re, 8022 Zurich, Switzerland
Email: annina.bracher@gmail.com

Amos Lapidoth and Christoph Pfister
Signal and Information Processing Laboratory
ETH Zurich, 8092 Zurich, Switzerland
Email: {lapidoth,pfister}@isi.ee.ethz.ch

***Abstract*—The rate region of the task-encoding problem for two correlated sources is characterized using a novel parametric family of dependence measures. The converse uses a new expression for the ρ-th moment of the list size, which is derived using the relative α-entropy.**

I. Introduction

We extend the task-encoding problem introduced by Bunte and Lapidoth [1] to the distributed setting depicted in Figure 1. A source generates a sequence of pairs $\{(X_i, Y_i)\}_{i=1}^n$ over the finite alphabet $\mathcal{X} \times \mathcal{Y}$. Using the functions

$$f_n : \mathcal{X}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R}_\mathsf{X}} \rfloor\}, \tag{1}$$
$$g_n : \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R}_\mathsf{Y}} \rfloor\}, \tag{2}$$

the sequence X^n is described by one of $\lfloor 2^{n\mathsf{R}_\mathsf{X}} \rfloor$ labels and the sequence Y^n by one of $\lfloor 2^{n\mathsf{R}_\mathsf{Y}} \rfloor$ labels. The decoder outputs the list of all pairs of sequences that could have produced the given pair of labels. The size of this list is

$$L(X^n, Y^n) \triangleq |\{(x', y') \in \mathcal{X}^n \times \mathcal{Y}^n : f_n(x') = f_n(X^n) \wedge g_n(y') = g_n(Y^n)\}|. \tag{3}$$

For a fixed $\rho > 0$, a rate pair $(\mathsf{R}_\mathsf{X}, \mathsf{R}_\mathsf{Y})$ is called achievable if there exists a sequence of task encoders $\{(f_n, g_n)\}_{n=1}^\infty$ such that the ρ-th moment of the list size tends to one as n tends to infinity, i.e., if

$$\lim_{n\to\infty} \mathrm{E}[L(X^n, Y^n)^\rho] = 1. \tag{4}$$

Our main contribution is Theorem 1, which states that rate pairs $(\mathsf{R}_\mathsf{X}, \mathsf{R}_\mathsf{Y})$ in the interior of the following region are achievable, while those outside the region are not:

$$\mathsf{R}_\mathsf{X} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n)}{n}, \tag{5}$$
$$\mathsf{R}_\mathsf{Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(Y^n)}{n}, \tag{6}$$
$$\mathsf{R}_\mathsf{X} + \mathsf{R}_\mathsf{Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n, Y^n) + K_{\tilde{\rho}}(X^n; Y^n)}{n}, \tag{7}$$

where $H_{\tilde{\rho}}$ denotes the Rényi entropy, $K_{\tilde{\rho}}$ is a dependence measure defined in Section II, and throughout the paper

$$\tilde{\rho} \triangleq \frac{1}{1+\rho}. \tag{8}$$

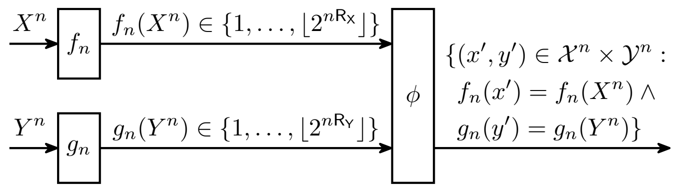

Fig. 1. Distributed task encoding with encoders f_n and g_n and a decoder ϕ.

In the IID case, (5)–(7) reduce to

$$\mathsf{R}_\mathsf{X} \geq H_{\tilde{\rho}}(P_X), \tag{9}$$
$$\mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(P_Y), \tag{10}$$
$$\mathsf{R}_\mathsf{X} + \mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(P_{XY}) + K_{\tilde{\rho}}(X; Y). \tag{11}$$

Compared to Slepian–Wolf coding [2], we notice two major differences. First, the constraints (9) and (10) only depend on the marginal PMFs P_X and P_Y, so the information that X reveals about Y and vice-versa has no influence on these constraints. Second, the constraint on the sum rate includes a term $K_{\tilde{\rho}}$, which is not present in the single-source setting [1, Theorem I.2]. (The term $K_{\tilde{\rho}}$ is always nonnegative and zero if and only if X and Y are independent [3, Theorem 2].)

Task encoding is related to the Massey–Arikan guessing experiment [4], [5], where the decoder repeatedly guesses until correct. While the guessing problem and the task-encoding problem lead to the same asymptotics in the single-source setting [6], this is no longer the case in the distributed setting [7]: except if X and Y are independent, the guessing region from Section VI is strictly larger than the task-encoding region (9)–(11).

Another contribution concerns the ρ-th moment of the list size in the single-source setting. Let $\mathcal{X}$ be a finite set of tasks from which a task X is drawn at random according to the PMF P and then mapped to one of M labels by a task encoder $f : \mathcal{X} \to \{1, \ldots, \mathsf{M}\}$. Given a task x, we denote by

$$L(x) \triangleq |\{x' \in \mathcal{X} : f(x') = f(x)\}| \tag{12}$$

the size of the list, i.e., the number of tasks that have the same label as x. In Lemma 1, we show that the ρ-th moment of the list size can be expressed as

$$\mathrm{E}[L(X)^\rho] = 2^{\rho[H_{\tilde{\rho}}(P) + \Delta_{\tilde{\rho}}(P||Q) - \log \mathsf{M}']}, \tag{13}$$

where $H_{\tilde{\rho}}$ and $\Delta_{\tilde{\rho}}$ are the Rényi entropy of order $\tilde{\rho}$ and the relative $\tilde{\rho}$-entropy, respectively, which will be defined in

Section II; Q is an auxiliary PMF that depends only on the task encoder f; and M' equals the number of used labels. The analogy between (13) and a similar expression in classical fixed-to-variable length source coding is discussed at the end of Section III.

The remainder of this paper is organized as follows. In Section II, we define Rényi's information measures and review some of their properties. In Section III, we prove (13) and draw the analogy between (13) and a similar expression in classical fixed-to-variable length source coding. In Section IV, we show that (5)–(7) characterize the region of achievable rate pairs for distributed task encoding. In Section V, we compare (12) with the related setting where the decoder's list only contains tasks with positive posterior probability. In Section VI, we discuss the guessing problem for two correlated sources.

II. Rényi's Information Measures

All logarithms in this paper are to base two. The Rényi entropy of order α was introduced by Rényi [8] and is defined for $\alpha > 0$ and $\alpha \neq 1$ as

$$H_\alpha(P) \triangleq \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha, \tag{14}$$

where P is a PMF. It is a generalization of Shannon entropy because $\lim_{\alpha\to 1} H_\alpha(P) = H(P)$. If the PMF of X is P_X, we also use $H_\alpha(X)$ to denote $H_\alpha(P_X)$.

The Rényi divergence of order α was also introduced by Rényi [8] and is defined for $\alpha > 0$ and $\alpha \neq 1$ as

$$D_\alpha(P||Q) \triangleq \frac{1}{\alpha-1} \log \sum_x P(x)^\alpha Q(x)^{1-\alpha}, \tag{15}$$

where P and Q are PMFs and where we use the convention that for $\alpha > 1$, we read $P(x)^\alpha Q(x)^{1-\alpha}$ as $\frac{P(x)^\alpha}{Q(x)^{\alpha-1}}$ and say that $\frac{0}{0} = 0$ and $\frac{p}{0} = \infty$ for $p > 0$. It is a generalization of Kullback–Leibler divergence because $\lim_{\alpha\to 1} D_\alpha(P||Q)$ is equal to $D(P||Q)$.

The relative α-entropy was defined by Sundaresan [9], [10] for $\alpha > 0$ and $\alpha \neq 1$ as

$$\Delta_\alpha(P||Q) \triangleq \frac{\alpha}{1-\alpha} \log \sum_x P(x) Q(x)^{\alpha-1} + \log \sum_x Q(x)^\alpha - \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha, \tag{16}$$

where P and Q are PMFs and where we use the convention that for $\alpha < 1$, we read $P(x)Q(x)^{\alpha-1}$ as $\frac{P(x)}{Q(x)^{1-\alpha}}$ and say that $\frac{0}{0} = 0$ and $\frac{p}{0} = \infty$ for $p > 0$. It is also a generalization of Kullback–Leibler divergence because $\lim_{\alpha\to 1} \Delta_\alpha(P||Q)$ is equal to $D(P||Q)$.

Relative α-entropy and Rényi divergence are related as follows [3, Lemma 1]:

$$\Delta_\alpha(P||Q) = D_{\frac{1}{\alpha}}(\hat{P}||\hat{Q}), \tag{17}$$

where the transformed PMFs $\hat{P}$ and $\hat{Q}$ are given by

$$\hat{P}(x) \triangleq \frac{P(x)^\alpha}{\sum_{x'} P(x')^\alpha}, \tag{18}$$

$$\hat{Q}(x) \triangleq \frac{Q(x)^\alpha}{\sum_{x'} Q(x')^\alpha}. \tag{19}$$

For a fixed $\alpha > 0$, this transformation is bijective on the set of all PMFs because for all $x \in \mathcal{X}$,

$$Q(x) = \frac{\hat{Q}(x)^{1/\alpha}}{\sum_{x'} \hat{Q}(x')^{1/\alpha}}. \tag{20}$$

The measure of dependence $K_\alpha(X;Y)$ was introduced in [3] and is defined as

$$K_\alpha(X;Y) \triangleq \min_{Q_X, Q_Y} \Delta_\alpha(P_{XY}||Q_X Q_Y), \tag{21}$$

where P_{XY} is the joint PMF of X and Y and the minimization is over all PMFs Q_X and Q_Y. It is a generalization of the mutual information because $\lim_{\alpha\to 1} K_\alpha(X;Y) = I(X;Y)$.

Recalling (8) and (14), we obtain from (16)

$$\Delta_{\tilde{\rho}}(P||Q) = \frac{1}{\rho} \log \sum_x P(x) Q(x)^{-\rho\tilde{\rho}} + \log \sum_x Q(x)^{\tilde{\rho}} - H_{\tilde{\rho}}(P). \tag{22}$$

III. Moments of the List Size

Lemma 1. *Let P be a PMF on the finite set $\mathcal{X}$, let M be a positive integer, let f be a function from $\mathcal{X}$ to $\{1, \ldots, \mathsf{M}\}$, let L be defined as in (12), and let $\rho > 0$. Define the PMF Q as*

$$Q(x) \triangleq \frac{L(x)^{-(1+\rho)}}{\sum_{x'} L(x')^{-(1+\rho)}}. \tag{23}$$

If X is distributed according to P, then

$$\mathrm{E}[L(X)^\rho] = 2^{\rho[H_{\tilde{\rho}}(P) + \Delta_{\tilde{\rho}}(P||Q) - \log \mathsf{M}']}, \tag{24}$$

where M' denotes the number of labels that are actually used (as opposed to allowed), i.e.,

$$\mathsf{M}' \triangleq |\{f(x) : x \in \mathcal{X}\}| \leq \mathsf{M}. \tag{25}$$

Proof. Since $L(x) \geq 1$ for all $x \in \mathcal{X}$, Q is well-defined and indeed a PMF. Rearranging (23), we get

$$L(x) = \beta Q(x)^{-\tilde{\rho}} \tag{26}$$

for some positive β. Let $\mathcal{M}' \triangleq \{f(x) : x \in \mathcal{X}\}$ be the set of labels that are used, and observe that

$$\begin{aligned} \mathsf{M}' &= \sum_{m\in\mathcal{M}'} 1 && (27)\\ &= \sum_{m\in\mathcal{M}'} \sum_{x: f(x)=m} L(x)^{-1} && (28)\\ &= \sum_{m\in\mathcal{M}'} \sum_{x: f(x)=m} \beta^{-1} Q(x)^{\tilde{\rho}} && (29)\\ &= \beta^{-1} \sum_x Q(x)^{\tilde{\rho}}, && (30) \end{aligned}$$

where (27) holds because $\mathsf{M}' = |\mathcal{M}'|$; (28) holds because for all $x \in \mathcal{X}$ with $f(x) = m$, $L(x) = |\{x' \in \mathcal{X} : f(x') = m\}|$; (29) follows from (26); and (30) holds because each x appears exactly once on the RHS of (29). Consequently, β can be expressed as

$$\beta = \frac{1}{\mathsf{M}'} \sum_x Q(x)^{\tilde{\rho}}, \tag{31}$$

and

$$\begin{aligned}
&\mathrm{E}[L(X)^\rho] \\
&\quad= \sum_x P(x)L(x)^\rho && (32)\\
&\quad= \beta^\rho \sum_x P(x)Q(x)^{-\rho\tilde{\rho}} && (33)\\
&\quad= 2^{\rho[-\log \mathsf{M}' + \log \sum_x Q(x)^{\tilde{\rho}} + \frac{1}{\rho}\log \sum_x P(x)Q(x)^{-\rho\tilde{\rho}}]} && (34)\\
&\quad= 2^{\rho[H_{\tilde{\rho}}(P) + \Delta_{\tilde{\rho}}(P||Q) - \log \mathsf{M}']}, && (35)
\end{aligned}$$

where (33) follows from (26); (34) follows from (31); and (35) follows from (22). ∎

Remark 1. *For every binary fixed-to-variable length source code, we have [2, (5.25)]*

$$\mathrm{E}[L'(X)] = H(P) + D(P||Q) + \log \frac{1}{\alpha}, \qquad (36)$$

where $L'(x)$ is the length of the codeword for symbol x; P is the PMF of the source; α is defined as $\sum_x 2^{-L'(x)}$; and the PMF Q is given by $Q(x) \triangleq \frac{1}{\alpha}2^{-L'(x)}$. The expected codeword length is thus determined by three terms: an entropy term that depends only on the source; a divergence term that measures how well the code is matched to the source; and an inefficiency term that depends only on the code. (For uniquely decodable codes, $\alpha \leq 1$ by Kraft's inequality.)

We have the same structure in (24): an entropy term that depends only on the source; a divergence term that measures how well the code is matched to the source; and an inefficiency term that depends only on the code ($\mathsf{M}' \leq \mathsf{M}$ must hold by definition).

IV. Distributed Task Encoding

Theorem 1. *Recalling the definition of an achievable rate pair from the introduction, rate pairs $(\mathsf{R_X}, \mathsf{R_Y})$ in the interior of the following region are achievable, while those outside the region are not:*

$$\mathsf{R_X} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n)}{n}, \qquad (37)$$

$$\mathsf{R_Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(Y^n)}{n}, \qquad (38)$$

$$\mathsf{R_X} + \mathsf{R_Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n, Y^n) + K_{\tilde{\rho}}(X^n; Y^n)}{n}. \qquad (39)$$

If $\{(X_i, Y_i)\}_{i=1}^\infty$ are IID P_{XY}, the region (37)–(39) reduces to

$$\mathsf{R_X} \geq H_{\tilde{\rho}}(P_X), \qquad (40)$$

$$\mathsf{R_Y} \geq H_{\tilde{\rho}}(P_Y), \qquad (41)$$

$$\mathsf{R_X} + \mathsf{R_Y} \geq H_{\tilde{\rho}}(P_{XY}) + K_{\tilde{\rho}}(X; Y). \qquad (42)$$

Proof. In both the proof of the converse and the direct part, we use the fact that the set on the RHS of (3) is a Cartesian product, so

$$L(x^n, y^n) = L_X(x^n)L_Y(y^n) \qquad (43)$$

for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, where

$$L_X(x^n) \triangleq |\{x' \in \mathcal{X}^n : f_n(x') = f_n(x^n)\}|, \qquad (44)$$

$$L_Y(y^n) \triangleq |\{y' \in \mathcal{Y}^n : g_n(y') = g_n(y^n)\}|. \qquad (45)$$

We begin with the converse, i.e., with showing that if a rate pair $(\mathsf{R_X}, \mathsf{R_Y})$ is achievable, then (37)–(39) must be satisfied. Observe that

$$\begin{aligned}
\mathrm{E}[L(X^n, Y^n)^\rho] &= \mathrm{E}[L_X(X^n)^\rho L_Y(Y^n)^\rho] && (46)\\
&\geq \mathrm{E}[L_X(X^n)^\rho] && (47)\\
&= 2^{\rho[H_{\tilde{\rho}}(X^n) + \Delta_{\tilde{\rho}}(P_{X^n}||Q) - \log \mathsf{M}']} && (48)\\
&\geq 2^{\rho n[\frac{1}{n}H_{\tilde{\rho}}(X^n) - \mathsf{R_X}]}, && (49)
\end{aligned}$$

where (46) follows from (43); (47) holds because $L_Y(y^n) \geq 1$ for all $y^n \in \mathcal{Y}^n$; (48) follows from Lemma 1 applied with the function $f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R_X}} \rfloor\}$ and the PMF P_{X^n}; and (49) holds because $\Delta_{\tilde{\rho}}$ is nonnegative [10] and because $\mathsf{M}' \leq \lfloor 2^{n\mathsf{R_X}} \rfloor \leq 2^{n\mathsf{R_X}}$. If (37) is not satisfied, then there exists a $\gamma > 0$ such that

$$\frac{1}{n}H_{\tilde{\rho}}(X^n) - \mathsf{R_X} \geq \gamma \qquad (50)$$

holds for infinitely many values of n. In that case, (49) implies that $\limsup_{n\to\infty} \mathrm{E}[L(X^n, Y^n)^\rho] = \infty$, which precludes the possibility that $\lim_{n\to\infty} \mathrm{E}[L(X^n, Y^n)^\rho] = 1$. Thus, (37) is necessary for the rate pair $(\mathsf{R_X}, \mathsf{R_Y})$ to be achievable. The necessity of (38) follows by swapping the role of X and Y in the above proof. To see that (39) is necessary, introduce the PMFs

$$Q_{X^n}(x^n) \triangleq \frac{L_X(x^n)^{-(1+\rho)}}{\sum_{x'\in\mathcal{X}^n} L_X(x')^{-(1+\rho)}}, \qquad (51)$$

$$Q_{Y^n}(y^n) \triangleq \frac{L_Y(y^n)^{-(1+\rho)}}{\sum_{y'\in\mathcal{Y}^n} L_Y(y')^{-(1+\rho)}}, \qquad (52)$$

and observe that

$$\begin{aligned}
&\mathrm{E}[L(X^n, Y^n)^\rho] \\
&\quad= 2^{\rho[H_{\tilde{\rho}}(X^n, Y^n) + \Delta_{\tilde{\rho}}(P_{X^nY^n}||Q) - \log \mathsf{M}']} && (53)\\
&\quad= 2^{\rho[H_{\tilde{\rho}}(X^n, Y^n) + \Delta_{\tilde{\rho}}(P_{X^nY^n}||Q_{X^n}Q_{Y^n}) - \log \mathsf{M}']} && (54)\\
&\quad\geq 2^{\rho n[\frac{1}{n}H_{\tilde{\rho}}(X^n, Y^n) + \frac{1}{n}K_{\tilde{\rho}}(X^n; Y^n) - (\mathsf{R_X} + \mathsf{R_Y})]}, && (55)
\end{aligned}$$

where (53) follows from Lemma 1 by viewing the distributed task encoder as a function that maps pairs (x^n, y^n) to one of $\lfloor 2^{n\mathsf{R_X}} \rfloor \cdot \lfloor 2^{n\mathsf{R_Y}} \rfloor$ labels; (54) holds because plugging (43) into (23) leads to $Q(x^n, y^n) = Q_{X^n}(x^n)Q_{Y^n}(y^n)$ for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$; and (55) holds since the definition (21) implies $K_{\tilde{\rho}}(X^n; Y^n) \leq \Delta_{\tilde{\rho}}(P_{X^nY^n}||Q_{X^n}Q_{Y^n})$ and because we have $\mathsf{M}' \leq \lfloor 2^{n\mathsf{R_X}} \rfloor \cdot \lfloor 2^{n\mathsf{R_Y}} \rfloor \leq 2^{n(\mathsf{R_X}+\mathsf{R_Y})}$. By the same argument as in (50), (55) implies that (39) is necessary for the achievability of the rate pair $(\mathsf{R_X}, \mathsf{R_Y})$.

We next turn to the direct part and show that a rate pair $(\mathsf{R_X}, \mathsf{R_Y})$ is achievable whenever (37)–(39) all hold with strict inequalities. We first use the methods from [1, Section III-B] to obtain task encoders f_n and g_n that are based on auxiliary PMFs Q_{X^n} and Q_{Y^n}, respectively, and we give bounds on the ρ-th moment of the list size. We then show how to choose

Q_{X^n} and Q_{Y^n} to ensure that (4) is satisfied. Throughout the proof of the direct part, we assume

$$\lfloor 2^{n\mathsf{R_X}} \rfloor - n\log|\mathcal{X}| - 2 > 0, \tag{56}$$
$$\lfloor 2^{n\mathsf{R_Y}} \rfloor - n\log|\mathcal{Y}| - 2 > 0. \tag{57}$$

This entails no loss of generality since we are only interested in the large-n asymptotic performance of our scheme, and because $\mathsf{R_X}$ and $\mathsf{R_Y}$ are positive, there exists some n_0 such that (56) and (57) hold for all $n \geq n_0$.

Using [1, Proposition III.2] twice, we obtain task encoders $f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R_X}} \rfloor\}$ and $g_n \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R_Y}} \rfloor\}$ satisfying

$$L_X(x^n) \leq \lambda_X(x^n), \tag{58}$$
$$L_Y(y^n) \leq \lambda_Y(y^n) \tag{59}$$

for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, where

$$\beta_X \triangleq \frac{2\sum_{x^n\in\mathcal{X}^n} Q_{X^n}(x^n)^{\tilde\rho}}{\lfloor 2^{n\mathsf{R_X}} \rfloor - n\log|\mathcal{X}| - 2}, \tag{60}$$
$$\beta_Y \triangleq \frac{2\sum_{y^n\in\mathcal{Y}^n} Q_{Y^n}(y^n)^{\tilde\rho}}{\lfloor 2^{n\mathsf{R_Y}} \rfloor - n\log|\mathcal{Y}| - 2}, \tag{61}$$
$$\lambda_X(x^n) \triangleq \begin{cases} \lceil \beta_X Q_{X^n}(x^n)^{-\tilde\rho} \rceil & \text{if } Q_{X^n}(x^n) > 0, \\ \infty & \text{if } Q_{X^n}(x^n) = 0, \end{cases} \tag{62}$$
$$\lambda_Y(y^n) \triangleq \begin{cases} \lceil \beta_Y Q_{Y^n}(y^n)^{-\tilde\rho} \rceil & \text{if } Q_{Y^n}(y^n) > 0, \\ \infty & \text{if } Q_{Y^n}(y^n) = 0, \end{cases} \tag{63}$$

which are well-defined thanks to (56) and (57). The condition of [1, Proposition III.2] on λ_X is fulfilled with $\alpha = 2$ because

$$\begin{aligned} &2\sum_{x^n\in\mathcal{X}^n} \frac{1}{\lambda_X(x^n)} + \log|\mathcal{X}^n| + 2 \\ &\quad\leq 2\sum_{x^n\in\mathcal{X}^n} \frac{Q_{X^n}(x^n)^{\tilde\rho}}{\beta_X} + n\log|\mathcal{X}| + 2 && (64)\\ &\quad= \lfloor 2^{n\mathsf{R_X}} \rfloor, && (65)\end{aligned}$$

where (64) follows from (62) and (65) follows from (60). The same arguments show that the respective condition on λ_Y is also fulfilled. We upperbound the ρ-th moment of the list size as follows (we neglect the issue that $\lambda_X(x^n)$ and $\lambda_Y(y^n)$ can be infinite, but it is possible to show that (70) continues to hold without this simplification):

$$\begin{aligned} &\mathrm{E}[L(X^n,Y^n)^\rho] \\ &\quad= \sum_{x^n,y^n} P(x^n,y^n) L_X(x^n)^\rho L_Y(y^n)^\rho && (66)\\ &\quad\leq \sum_{x^n,y^n} P(x^n,y^n) \lambda_X(x^n)^\rho \lambda_Y(y^n)^\rho && (67)\\ &\quad= \sum_{x^n,y^n} P(x^n,y^n) \left\lceil \frac{\beta_X}{Q_{X^n}(x^n)^{\tilde\rho}} \right\rceil^\rho \left\lceil \frac{\beta_Y}{Q_{Y^n}(y^n)^{\tilde\rho}} \right\rceil^\rho && (68)\\ &\quad\leq \sum_{x^n,y^n} P(x^n,y^n) \Bigg\{ 1 + \frac{2^\rho \beta_X^\rho}{Q_{X^n}(x^n)^{\rho\tilde\rho}} + \frac{2^\rho \beta_Y^\rho}{Q_{Y^n}(y^n)^{\rho\tilde\rho}} \\ &\qquad\qquad + \frac{4^\rho [\beta_X\beta_Y]^\rho}{[Q_{X^n}(x^n)Q_{Y^n}(y^n)]^{\rho\tilde\rho}} \Bigg\} && (69)\\ &\quad= 1 + 2^{\rho[H_{\tilde\rho}(X^n) + \Delta_{\tilde\rho}(P_{X^n}||Q_{X^n}) - n\mathsf{R_X} + \delta_1]} \\ &\qquad + 2^{\rho[H_{\tilde\rho}(Y^n) + \Delta_{\tilde\rho}(P_{Y^n}||Q_{Y^n}) - n\mathsf{R_Y} + \delta_2]} \\ &\qquad + 2^{\rho[H_{\tilde\rho}(X^n,Y^n) + \Delta_{\tilde\rho}(P_{X^nY^n}||Q_{X^n}Q_{Y^n}) - n(\mathsf{R_X}+\mathsf{R_Y}) + \delta_3]}, && (70)\end{aligned}$$

where (66) follows from (43); (67) follows from (58) and (59); (68) follows from (62) and (63); (69) follows from the inequality $\lceil \xi \rceil^\rho < 1 + 2^\rho \xi^\rho$ from [1, (26)], which holds for all $\rho > 0$ and $\xi \geq 0$; and (70) follows from (22), a longer computation, and the definitions

$$\delta_1 \triangleq \log \frac{4 \cdot 2^{n\mathsf{R_X}}}{\lfloor 2^{n\mathsf{R_X}} \rfloor - n\log|\mathcal{X}| - 2}, \tag{71}$$
$$\delta_2 \triangleq \log \frac{4 \cdot 2^{n\mathsf{R_Y}}}{\lfloor 2^{n\mathsf{R_Y}} \rfloor - n\log|\mathcal{Y}| - 2}, \tag{72}$$
$$\delta_3 \triangleq \delta_1 + \delta_2. \tag{73}$$

From (17) we know that

$$\Delta_{\tilde\rho}(P_{X^n}||Q_{X^n}) = D_{1+\rho}(\hat P_{X^n}||\hat Q_{X^n}) \tag{74}$$
$$\Delta_{\tilde\rho}(P_{Y^n}||Q_{Y^n}) = D_{1+\rho}(\hat P_{Y^n}||\hat Q_{Y^n}) \tag{75}$$
$$\Delta_{\tilde\rho}(P_{X^nY^n}||Q_{X^n}Q_{Y^n}) = D_{1+\rho}(\hat P_{X^nY^n}||\hat Q_{X^n}\hat Q_{Y^n}), \tag{76}$$

where (76) holds because the transformation (19) of a product is the product of the transformations. Let $Q^*_{X^n}$ and $Q^*_{Y^n}$ be PMFs that achieve equality in (21), so

$$\Delta_{\tilde\rho}(P_{X^nY^n}||Q^*_{X^n}Q^*_{Y^n}) = K_{\tilde\rho}(X^n;Y^n). \tag{77}$$

We now show how to choose Q_{X^n} and Q_{Y^n}. Even in the IID case, these will typically not be product distributions. We consider the mixture distributions

$$\hat Q_{X^n}(x^n) = \tfrac12 \hat P_{X^n}(x^n) + \tfrac12 \hat Q^*_{X^n}(x^n), \tag{78}$$
$$\hat Q_{Y^n}(y^n) = \tfrac12 \hat P_{Y^n}(y^n) + \tfrac12 \hat Q^*_{Y^n}(y^n), \tag{79}$$

and use the inverse transformation (20) to obtain Q_{X^n} and Q_{Y^n}. Consequently,

$$\begin{aligned} &\Delta_{\tilde\rho}(P_{X^n}||Q_{X^n}) \\ &\quad= D_{1+\rho}(\hat P_{X^n}||\tfrac12 \hat P_{X^n} + \tfrac12 \hat Q^*_{X^n}) && (80)\\ &\quad= \frac{1}{\rho}\log\sum_{x^n} \hat P_{X^n}(x^n)^{1+\rho} \left[\frac{\hat P_{X^n}(x^n) + \hat Q^*_{X^n}(x^n)}{2}\right]^{-\rho} && (81)\\ &\quad\leq \frac{1}{\rho}\log\sum_{x^n} \hat P_{X^n}(x^n)^{1+\rho} \left[\frac12 \hat P_{X^n}(x^n)\right]^{-\rho} && (82)\\ &\quad= 1, && (83)\end{aligned}$$

where (80) follows from (74) and (78); (81) follows from the definition (15); and (83) holds because $\hat P_{X^n}$ is a PMF. In the same way, we obtain $\Delta_{\tilde\rho}(P_{Y^n}||Q_{Y^n}) \leq 1$ and

$$\begin{aligned} &\Delta_{\tilde\rho}(P_{X^nY^n}||Q_{X^n}Q_{Y^n}) \\ &\quad= D_{1+\rho}\big(\hat P_{X^nY^n}||(\tfrac12 \hat P_{X^n} + \tfrac12 \hat Q^*_{X^n})(\tfrac12 \hat P_{Y^n} + \tfrac12 \hat Q^*_{Y^n})\big) && (84)\\ &\quad\leq \frac{1}{\rho}\log\sum_{x^n,y^n} \hat P(x^n,y^n)^{1+\rho} \left[\frac{\hat Q^*_{X^n}(x^n)\hat Q^*_{Y^n}(y^n)}{4}\right]^{-\rho} && (85)\\ &\quad= D_{1+\rho}(\hat P_{X^nY^n}||\hat Q^*_{X^n}\hat Q^*_{Y^n}) + 2 && (86)\\ &\quad= K_{\tilde\rho}(X^n;Y^n) + 2, && (87)\end{aligned}$$

where (87) follows from (17) and (77). Plugging these results into (70), we finally arrive at

$$\begin{aligned}
&\mathrm{E}[L(X^n,Y^n)^\rho] \\
&\quad \le 1 + 2^{\rho n[\frac{1}{n}H_{\tilde\rho}(X^n)-\mathsf{R_X}]}\cdot 2^{\rho(\delta_1+1)} \\
&\qquad + 2^{\rho n[\frac{1}{n}H_{\tilde\rho}(Y^n)-\mathsf{R_Y}]}\cdot 2^{\rho(\delta_2+1)} \\
&\qquad + 2^{\rho n[\frac{1}{n}H_{\tilde\rho}(X^n,Y^n)+\frac{1}{n}K_{\tilde\rho}(X^n;Y^n)-(\mathsf{R_X}+\mathsf{R_Y})]}\cdot 2^{\rho(\delta_3+2)}, \quad (88)
\end{aligned}$$

which tends to one as n tends to infinity: since $(\mathsf{R_X},\mathsf{R_Y})$ is in the interior of (37)–(39), the expressions in square brackets will be smaller than or equal to γ for some $\gamma<0$ and n large enough; and we have $\lim_{n\to\infty}\delta_1=\lim_{n\to\infty}\delta_2=2$ and also $\lim_{n\to\infty}\delta_3=4$.

We finish with the specialization of the region (37)–(39) for an IID source with PMF P_{XY}. In this case, (40)–(42) readily follow from (37)–(39) because

$$H_{\tilde\rho}(X^n)=nH_{\tilde\rho}(P_X), \quad (89)$$

$$H_{\tilde\rho}(Y^n)=nH_{\tilde\rho}(P_Y), \quad (90)$$

$$H_{\tilde\rho}(X^n,Y^n)=nH_{\tilde\rho}(P_{XY}), \quad (91)$$

$$K_{\tilde\rho}(X^n;Y^n)=nK_{\tilde\rho}(X;Y), \quad (92)$$

where (89)–(91) follow from the definition (14) and simple computations; and (92) follows from the repeated application of [3, Theorem 2, Property 3]. ■

V. On the Definition of the List

To appreciate the subtleties in defining the list, let us first consider the single-source case and compare (12) with the case where the decoder's list is only required to contain tasks whose probability, conditional on the observed label, is positive. The list size in this case is

$$L'(x)\triangleq|\{x'\in\mathcal{X}:P(x')>0\ \wedge\ f(x')=f(x)\}|. \quad (93)$$

In the single-source case, the two criteria lead to identical asymptotics because for every task encoder f whose ρ-th moment of the list size according to (93) is $\mathrm{E}[L'_f(X)^\rho]$, there exists a task encoder g that has the same ρ-th moment of the list size according to (12) if g is allowed to use one additional label (which is negligible in an asymptotic setting). Indeed, if

$$g(x)=\begin{cases} f(x) & \text{if } P(x)>0,\\ \mathsf{M}+1 & \text{if } P(x)=0,\end{cases} \quad (94)$$

where $\mathsf{M}+1$ denotes the additional label, then

$$\begin{aligned}
\mathrm{E}[L_g(X)^\rho] &= \sum_{x:P(x)>0}P(x)L_g(x)^\rho && (95)\\
&= \sum_{x:P(x)>0}P(x)L'_f(x)^\rho && (96)\\
&= \mathrm{E}[L'_f(X)^\rho], && (97)
\end{aligned}$$

where (96) follows from (94) since tasks with $P(x)>0$ do not share their labels with zero-probability tasks, so $L_g(x)$ is equal to $L'_f(x)$ for all $x\in\mathcal{X}$ with $P(x)>0$.

In the distributed case, the picture can change dramatically. To see why, consider an IID source with $X=Y$: under the positive posterior probability criterion, the decoder's list will only contain pairs that satisfy $x^n=y^n$, and a careful analysis shows that rate pairs $(\mathsf{R_X},\mathsf{R_Y})$ satisfying

$$\mathsf{R_X}+\mathsf{R_Y}\ge H_{\tilde\rho}(P_{XY}) \quad (98)$$

with strict inequality are achievable, while those not satisfying (98) are not. Unless X and Y are deterministic, this region is strictly larger than the region defined by (40)–(42): there are no individual constraints on $\mathsf{R_X}$ and $\mathsf{R_Y}$, and the constraint on the sum rate does not include the penalty term $K_{\tilde\rho}$.

The definition based on (12) seems easier to analyze and, unless zero-probability tasks are present, the two criteria are equivalent.

VI. Distributed Guessing

As in distributed task encoding, a source generates a sequence of pairs $\{(X_i,Y_i)\}_{i=1}^n$ over the finite alphabet $\mathcal{X}\times\mathcal{Y}$. Using the functions f_n and g_n, the sequence X^n is described by one of $\lfloor 2^{n\mathsf{R_X}}\rfloor$ labels and the sequence Y^n by one of $\lfloor 2^{n\mathsf{R_Y}}\rfloor$ labels. Given a pair of labels, the decoder repeatedly guesses (x^n,y^n) until correct. We are interested in the number of guesses that the decoder needs. For a fixed $\rho>0$, a rate pair $(\mathsf{R_X},\mathsf{R_Y})$ is called achievable if there exists a sequence of encoders $\{(f_n,g_n)\}_{n=1}^\infty$ such that the ρ-th moment of the number of guesses tends to one as n tends to infinity, i.e., if $\lim_{n\to\infty}\mathrm{E}[G(X^n,Y^n)^\rho]=1$.

In the IID case, we show in [7] that rate pairs $(\mathsf{R_X},\mathsf{R_Y})$ in the interior of the following region are achievable, while those outside the region are not:

$$\mathsf{R_X}\ge H_{\tilde\rho}(X|Y), \quad (99)$$

$$\mathsf{R_Y}\ge H_{\tilde\rho}(Y|X), \quad (100)$$

$$\mathsf{R_X}+\mathsf{R_Y}\ge H_{\tilde\rho}(X,Y), \quad (101)$$

where $H_{\tilde\rho}(X|Y)$ is the conditional Rényi entropy from [11].

References

[1] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5065–5076, Sept. 2014.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.

[3] A. Lapidoth and C. Pfister, "Two measures of dependence," in *Proc. 2016 IEEE Int. Conf. Science of Electrical Engineering*, Eilat, 2016.

[4] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE Int. Symp. Inf. Theory*, Trondheim, 1994, p. 204.

[5] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, Jan. 1996.

[6] A. Bracher, E. Hof, and A. Lapidoth, "Distributed storage for data security," in *Proc. 2014 IEEE Inf. Theory Workshop*, Hobart, 2014, pp. 506–510.

[7] A. Bracher, A. Lapidoth, and C. Pfister, "Distributed guessing and task encoding," in preparation.

[8] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. and Probability*, vol. 1, 1961, pp. 547–561.

[9] R. Sundaresan, "A measure of discrimination and its geometric properties," in *Proc. 2002 IEEE Int. Symp. Inf. Theory*, Lausanne, 2002, p. 264.

[10] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.

[11] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory*, I. Csiszár and P. Elias, Eds. Amsterdam: North-Holland Publishing, 1977, pp. 41–52.