ISIT'2013 1569718899

Assisted Sampling of Correlated Sources

Vinod M. Prabhakaran
Tata Institute of Fundamental Research
Mumbai, India
Email: vinodmp@tifr.res.in

Anand D. Sarwate
Toyota Technological Institute at Chicago
Chicago, USA
Email: asarwate@ttic.edu

***Abstract*—We study a distributed sampling scenario in which two agents observing components of a correlated source must each generate components of a second correlated source. The agents are aided by an "omniscient" third terminal which observes the two input sources and transmits rate-limited messages to assist the terminals in generating the required correlation in their outputs. We identify two sub-cases of this problem based on how the generated sources must depend on the input sources.**

I. Introduction

Consider the following scenario: there are two actuators, A and B, which have correlated sensor readings U and V, respectively. These sensor readings are also available to centralized controller who can send rate-limited messages to the actuators A and B that can allow them to make correlated actions. In this paper we consider the effect of the rate constraint on the degree of correlation achievable by A and B in such a system. In particular, we study an information theoretic model in which the two actuators have access to U^n and V^n and must produce correlated sequences X^n and Y^n from a given distribution. The general setup for our problem is shown in Figure 1 with Alice and Bob as the two actuators. We refer to the centralized controller as a *genie*, following now-standard terminology from communications [1, p. 419].

Many authors have studied how to generate correlated random variables in a distributed setting. The earlier works focused on generating a single variable in a distributed manner with no assistance from a third terminal and no interaction between the terminals. Gács and Körner [2] studied the problem where (U^n, V^n) are independent and identically distributed (i.i.d.) from a distribution $p(u, v)$, and the two terminals must produce single variables X and Y which agree with probability $1 - \epsilon_n$, with $\epsilon_n \to 0$ as $n \to \infty$. The Gács-Körner common information characterizes the maximum entropy of the variable X normalized by n. In this non-communicative model, the original result was strengthened by Witsenhausen [3] to include impossibility results for even single-bit agreement; more recently approximate agreement was studied by Bogdanov and Mossel [4]. Kamath and Anantharam [5] studied the case where X and Y need not agree but should have a specified joint distribution.

Wyner [6] studied the problem where a third terminal (genie) sends common random bits to the two terminals who want to produce i.i.d. samples X^n, Y^n from a given distribution; the terminals do not interact. Wyner's *common information* characterizes the minimum rate required. Anantharam and Borkar [7] showed conditions under which Alice and Bob, having access to U and V as well as independent common random bits (but no interaction), cannot simulate a given joint distribution$p(x, y, u, v)$. Motivated by an application to secure computation, Prabhakaran and Prabhakaran [8], [9] generalized the setup of the Gács-Körner problem to include an omniscient genie who observes U^n, V^n and communicates to the two terminals over rate-limited links. In this "assisted common information" setup, they characterized the trade-off between the rates of communication and the normalized entropy rate of the common random variable. Our problem setup bears a close resemblance, but instead of agreeing on a common random variable we require Alice and Bob to generate correlated sequences.

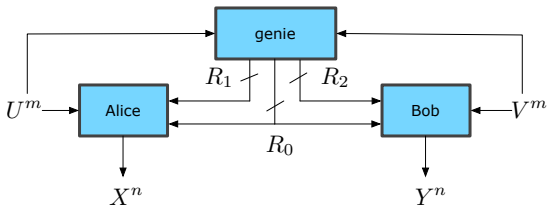

Fig. 1. Generating correlated variables from correlated sources.

There is a rich class of problems where two (or more) interacting terminals generate correlated random variables. In channel synthesis [10], [11], Alice sends rate-limited messages to Bob to simulate a noisy channel between them. Recent work has generalized this setup to interactive communication [12] and a general characterization was been recently obtained by Yassaee et al. [13]. A number of different scenarios for three terminals trying to coordinate actions was studied by Cuff et al. [14], and a recent relay model was considered by Haddadpour et al. [15] that also recovers some of the channel synthesis results [6], [11].

In our problem, Alice and Bob observe U^n and V^n, respectively, where $(U^n, V^n) \sim p(u, v)$ i.i.d. They must generate sequences X^n and Y^n using the setup in Figure 1. We identify two different cases. In the *conditional sampling* case, the two terminals must generate random variables such that the induced joint distribution of X^n, Y^n, U^n, V^n is close to i.i.d. sampling from a given $p(x, y|u, v)p(u, v)$. Closeness is in total variation sense throughout. For this case we provide upper

and lower bounds on the rates required for Alice and Bob to simulate the joint distribution. In the *unconditional* case, the two terminals must generate variables (X^n, Y^n) whose induced marginal distribution $P(x^n, y^n)$ is close to a i.i.d. sampling from a given $p(x, y)$. We completely characterize the rate region when $X = Y = Z$. For space reasons, the full proofs are deferred to the full version of this work.

II. Problem Statement

The general problem is shown in Figure 1. Let $\{(U_i, V_i) : i = 1, 2, \ldots\}$ be discrete memoryless source generated according to the joint distribution $p(u, v)$ over $\mathcal{U} \times \mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are finite discrete alphabets. There are three terminals in the system. Alice takes as input m symbols U^m and two rate-limited messages, Bob takes as input m-symbols V^m and two rate-limited messages, and the third terminal takes as input the pair (U^m, V^m) and outputs three rate-limited messages. Alice must output a length-n sequence X^n from a finite discrete alphabet $\mathcal{X}$ and Bob must output a length-n sequence Y^n from a finite discrete alphabet $\mathcal{Y}$. We study the $m = n$ case.

Definition 1: An (n, R_0, R_1, R_2) **genie** is a set of deterministic maps (g_0, g_1, g_2) with a random variable θ_G taking values in a space Θ_G such that:

$$g_0 : \mathcal{U}^n \times \mathcal{V}^n \times \Theta_G \to [2^{nR_0}]$$
$$g_1 : \mathcal{U}^n \times \mathcal{V}^n \times \Theta_G \to [2^{nR_1}]$$
$$g_2 : \mathcal{U}^n \times \mathcal{V}^n \times \Theta_G \to [2^{nR_2}].$$

These maps take a pair of vectors to side-information messages to the two agents. The third argument $\theta \in \Theta_G$ is private randomness for the genie that it can use to randomize the mapping from sequences to messages.

Definition 2: An (n, R_0, R_1, R_2) **source-generation code** is a pair of maps (f_A, f_B) and a pair of random variables (θ_A, θ_B) taking values in $\Theta_A \times \Theta_B$ such that

$$f_A : \mathcal{U}^n \times [2^{nR_1}] \times [2^{nR_0}] \times \Theta_A \to \mathcal{X}^n$$
$$f_B : \mathcal{V}^n \times [2^{nR_2}] \times [2^{nR_0}] \times \Theta_B \to \mathcal{Y}^n.$$

The arguments in Θ_A and Θ_B represent independent sources of private randomness that the terminals can use to randomize the mapping from messages to sequences.

Definition 3: An (n, R_0, R_1, R_2) **source-generation system** is a tuple $(f_A, f_B, g_0, g_1, g_2)$ with three independent random variables $\boldsymbol{\theta} = (\theta_A, \theta_B, \theta_G)$ where (f_A, f_B) is an (n, R_0, R_1, R_2), source-generation code with randomness (θ_A, θ_B) and (g_0, g_1, g_2) is an (n, R_0, R_1, R_2) genie with randomness θ_G. Define the average conditional distribution

$$\begin{aligned}&\hat{p}_n(x^n, y^n | u^n, v^n)\\ &= \mathbb{P}_{\boldsymbol{\theta}}\Big(f_A(u^n, g_1(u^n, v^n, \theta_G), g_0(u^n, v^n, \theta_G), \theta_A) = x^n,\\ &\quad f_B(v^n, g_2(u^n v^n, \theta_G), g_0(u^n, v^n, \theta_G), \theta_B) = y^n \big| u^n, v^n\Big),\end{aligned}$$

the joint distribution

$$\hat{p}_n(u^n, v^n, x^n, y^n) = \hat{p}_n(x^n, y^n | u^n, v^n) \prod_{i=1}^n p(u_i, v_i),$$

and the marginal distribution

$$\hat{p}_n(x^n, y^n) = \sum_{u^n, v^n} \hat{p}_n(u^n, v^n, x^n, y^n).$$

Note that the probability is taken over $\boldsymbol{\theta}$ – these distributions are averaged over the private random variables $\boldsymbol{\theta}$. We say these distributions are **generated** by the source-generation system.

We identify two different kinds of simulation structures. In *conditional sampling* the goal is for the outputs to be approximately distributed according to some conditional distribution $p(x, y|u, v)$. That is, we would like the i-th output samples (X_i, Y_i) to depend on the i-th input samples (U_i, V_i). In *unconditional sampling* the goal is for the outputs to be approximately i.i.d. according to a given $p(x, y)$. The following two definitions make this more precise.

Definition 4: The distribution $p(x, y, u, v)$ is said to be **conditionally samplable** with rates (R_0, R_1, R_2) if, for every $\epsilon > 0$, there exists a sequence of $(n, R_0 + \epsilon, R_1 + \epsilon, R_2 + \epsilon)$ source-generation systems $\{(f_A^{(n)}, f_B^{(n)}, g_0^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ such that

$$\lim_{n\to\infty} \left\| \prod_{i=1}^n p(x_i, y_i, u_i, v_i) - \hat{p}_n(x^n, y^n, u^n, v^n) \right\|_1 = 0,$$

where the joint distribution $\hat{p}_n(x^n, y^n, u^n, v^n)$ is generated by $\{(f_A^{(n)}, f_B^{(n)}, g_0^{(n)}, g_1^{(n)}, g_2^{(n)})\}$.

Definition 5: The distribution $p(x, y)$ is said to be **unconditionally samplable** with rates (R_0, R_1, R_2) if, for every $\epsilon > 0$, there exists a sequence of $(n, R_0 + \epsilon, R_1 + \epsilon, R_2 + \epsilon)$ source-generation systems $\{(f_A^{(n)}, f_B^{(n)}, g_0^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ such that

$$\lim_{n\to\infty} \left\| \prod_{i=1}^n p(x_i, y_i) - \hat{p}_n(x^n, y^n) \right\|_1 = 0,$$

where the marginal distribution $\hat{p}_n(x^n, y^n)$ is generated by $\{(f_A^{(n)}, f_B^{(n)}, g_0^{(n)}, g_1^{(n)}, g_2^{(n)})\}$.

III. Conditional sampling

We first turn to results on conditional sampling.

A. Achievability

Theorem 1: The distribution $p(x, y, u, v)$ is conditionally samplable with rates (R_0, R_1, R_2) if there exists a random variables Q in a set $\mathcal{Q}$ with $|\mathcal{Q}| \leq |\mathcal{U}| \cdot |\mathcal{V}| \cdot |\mathcal{X}| \cdot |\mathcal{Y}| + 1$ with $p(q, x, y|u, v) = p(q|u, v)p(x|q, u)p(y|q, v)$ such that

$$\sum_q p(q|u, v)p(x|q, u)p(y|q, v) = p(x, y|u, v),$$

and (R_0, R_1, R_2) satisfy the following constraints.

$$R_0 + R_1 \geq I(X, Y, V; Q|U),$$
$$R_0 + R_2 \geq I(X, Y, U; Q|V).$$

Proof sketch: The proof generalizes the idea behing the achievability proof in [6] and employs the elegant proof technique of [16]. The rough intuition is as follows: a random codebook of rate R_Q is constructed $\sim p(q)$. On observing (u^n, v^n), the genie picks uniformly at random a codeword

$q^n(m)$ from among the codewords which are jointly typical with the observed vectors; notice that, in the random coding argument, the number of eligible codewords will be roughly over $2^{n(R_Q - I(U,V;Q))}$. The identity of the codeword is conveyed to Alice and Bob by Slepian-Wolf binning the codebook to take advantage of their side information. In particular, the rates of communication must satisfy

$$R_0 + R_1 \geq R_Q - I(U;Q), \quad R_0 + R_2 \geq R_Q - I(V;Q).$$

After decoding m, Alice passes $(q^n(m), u^n)$ through the memoryless $p(x|q,u)$ channel that she simulates using her private randomness and outputs the resulting x^n. Similarly, Bob produces y^n by passing $(q^n(m), v^n)$ through the memoryless channel $p(y|q,v)$. It can be shown that for sufficiently large n there is a codebook such that the induced $p(x^n, y^n|u^n, v^n)$ distribution is within any desired total variation distance from the desired distribution provided

$$R_Q \geq I(X,Y;Q|U,V) + I(U,V;Q) = I(X,Y,U,V;Q).$$

The details are deferred to a full version of this paper. ■

B. Outerbound

Theorem 2: If (R_0, R_1, R_2) is achievable, there is a $p(q_1,q_2|u,v)p(x|q_1,u)p(y|q_2,v)$ such that

$$\sum_{q_1,q_2} p(q_1,q_2|u,v)p(x|q_1,u)p(y|q_2,v) = p(x,y|u,v),$$

and the rates satisfy

$$R_0 + R_1 \geq I(X,Y,V;Q_1|U),$$
$$R_0 + R_2 \geq I(X,Y,U;Q_2|V).$$

Proof sketch: This proof sketch is restricted to the case when the outputs are required to be perfectly memoryless $\sim p(x,y|u,v)$. Below, the parts which only hold approximately under the original problem setting are indicated by a $\approx$, the details for this case are deferred to the full paper. Let W_0 denote the common message and $W_i, i = 1,2$ the private messages from the genie. Let W_{0i} denote $(W_0, W_i), i = 1,2$ and let $T \sim \text{Unif}\{1,\ldots,n\}$ be independent of $(U^n, V^n, W_0, W_1, W_2)$. We will use $U_{\bar{i}}$ below to denote (U^i, U_{i+1}^n).

$$\begin{aligned}
n(R_0+R_1) &\geq I(X^n, Y^n, V^n; W_{01}|U^n) \\
&= \sum_{i=1}^n I(X_i, Y_i, V_i; W_{01}|U^n, X^{i-1}, Y^{i-1}, V^{i-1}) \\
&\approx \sum_{i=1}^n I(X_i, Y_i, V_i; W_{01}, X^{i-1}, Y^{i-1}, V^{i-1}, U_{\bar{i}}|U_i) \\
&\geq \sum_{i=1}^n I(X_i, Y_i, V_i; W_{01}, U_{\bar{i}}|U_i) \\
&= nI(X_T, Y_T, V_T; W_{01}, U_{\bar{T}}|U_T, T) \\
&\approx nI(X_T, Y_T, V_T; W_{01}, U_{\bar{T}}, T|U_T) \\
&= nI(X_T, Y_T, V_T; Q_1|U_T),
\end{aligned}$$

where $Q_1 = (W_{01}, U_{\bar{T}}, T)$. Similarly, if $Q_2 = (W_{02}, V_{\bar{T}}, T)$,

$$R_0 + R_2 \geq I(X_T, Y_T, U_T; Q_2|V_T).$$

It is also easy to verify that

$$(Q_2, X_T, V_T) - (Q_1, U_T) - X_T, \text{ and}$$
$$(Q_1, X_T, U_T) - (Q_2, V_T) - Y_T$$

are Markov chains. (This holds exactly even for the original problem setting). Since $(U_T, V_T, X_T, Y_T) \sim p(u,v)p(x,y|u,v)$ (only approximately for the original problem setting), we have the theorem. ■

C. Special cases

a) When $X = Y = Z$: Suppose we want a common output sequence with conditional distribution $p(z|u,v)$.

Corollary 3: The set of all achievable (R_0, R_1, R_2) are given by

$$R_0 + R_1 \geq H(Z|U), \qquad R_0 + R_2 \geq H(Z|V).$$

Proof: Achievability follows from taking $Q = Z$ in the innerbound theorem. Converse follows from the general outerbound theorem above. Plugging in $X = Y = Z$, we have

$$\begin{aligned}
R_0 + R_1 &\geq I(Z,V;Q_1|U) \geq I(Z;Q_1|U) \\
&= H(Z|U) - H(Z|Q_1,U).
\end{aligned}$$

But, we have the Markov chain $(Q_2, Z, V) - (Q_1, U) - Z$ which gives $H(Z|Q_1,U) = 0$. Hence, $R_0 + R_1 \geq H(Z|U)$. Similarly, $R_0 + R_2 \geq H(Z|V)$. ■

b) When U, V are deterministic: This is a slight generalization of Wyner [6]. We now also allow private links (in addition to a common link) from the genie. Recall that Wyner's common information of X, Y is $K_{\mathsf{W}}(X;Y) = \min_{X-Q-Y} I(X,Y;Q)$.

Corollary 4: The set of all achievable (R_0, R_1, R_2) is characterized by:

$$R_0 + \min(R_1, R_2) \geq K_{\mathsf{W}}(X;Y).$$

Proof: Achievability is by Wyner's argument [6]. It also follows from specializing the innerbound to this case. Specializing the general outerbound, we have

$$R_0 + R_1 \geq I(X,Y;Q_1) \geq K_{\mathsf{W}}(X;Y),$$

where the second inequality follows from the fact that $X - Q_1 - Y$ is required to be a Markov chain. Similarly, $R_0 + R_2 \geq K_{\mathsf{W}}(X;Y)$. ■

c) When X, V are deterministic: This is the channel simulation of Cuff [17]. It is clear that Alice and the genie may be treated as co-located. Then the problem reduces to simulating the channel $p(y|u)$ from Alice-genie to Bob using a bit-pipe of rate $R_0 + R_2$ between then.

Corollary 5: The set of all achievable (R_0, R_1, R_2) is characterized by

$$R_0 + R_2 \geq K_{\mathsf{W}}(U;Y).$$

Proof: Achievability follows directly from specializing the inner bound. Also see [17]. The converse follows from specializing the outer bound

$$R_0 + R_2 \geq I(Y, U; Q_2) \geq K_{\mathsf{W}}(U; Y),$$

where the second inequality follows from the fact that $U - Q_2 - Y$ is required to be a Markov chain. ∎

IV. Unconditional sampling

We now turn to the case of *unconditional simulation* in which Alice and Bob must produce sequences X^n and Y^n whose joint distribution approaches (in total variation distance) independent copies of a specified $p(x, y)$ as $n \to \infty$. In this case we prove results for the special case where $X = Y = Z$. That is, Alice and Bob should produce a nearly identical sequences Z_1^n and Z_2^n which are close to i.i.d. samples from a distribution $p(z)$.

Theorem 6: The distribution $p(z)$ is unconditionally samplable with rates (R_0, R_1, R_2) if and only if there exists a random variable Q on a set $\mathcal{Q}$ with $|\mathcal{Q}| \leq |\mathcal{U}| \cdot |\mathcal{V}| + 2$ and conditional distribution $p(q|u, v)$ and $R \geq 0$ such that

$$R_0 + R_1 \geq I(Q; V|U) + R \tag{1}$$
$$R_0 + R_2 \geq I(Q; U|V) + R \tag{2}$$
$$H(Z) \leq I(Q; U, V) + R. \tag{3}$$

If $H(Z) \geq H(U, V)$ then we may take $Q = (U, V)$. If $H(Z) \leq H(U, V)$ then we may take $R = 0$.

A. Achievability

The achievability proof is based on the assisted common information problem [9], and rely mostly on a straightforward double binning scheme along arguments from Wyner-Ziv coding.

Proof sketch: There are two cases to consider, depending on whether $R_0 \leq R$ or $R_0 > R$. We describe the $R_0 \leq R$ case for illustration. The genie generates $2^{n\bar{R}}$ length n-sequences $\{q^n(k) : k \in [2^{n\bar{R}}]\}$ according to the distribution $p(q|u, v)$ and bins them into 2^{nR_1} and 2^{nR_2} bins for Alice and Bob, respectively. Given the pair (u^n, v^n) the genie finds an index k such that $q^n(k)$ is jointly typical with (u^n, v^n) and sends the corresponding bin indices to Alice and Bob, along with additional (common) uniformly random bits at rate R. By the same arguments as in the Wyner-Ziv theorem, if

$$\begin{aligned} \bar{R} &\geq I(Q; U, V) \\ \bar{R} - R_1 &\leq I(Q; U) \\ \bar{R} - R_2 &\leq I(Q; V), \end{aligned}$$

then Alice and Bob can recover the index k. This gives the bounds (1) and (2). From this q^n and the additional common randomness they can synthesize a common Z^n as long as (3) holds. Note that in this sketch we have not argued that the $nI(Q; U, V)$ bits of the Wyner-Ziv codeword decoded by Alice and Bob are uniform, nor have we described how these common bits are used to synthesize Z^n. ∎

B. Converse

Proof: Let Z_1^n and Z_2^n, resp., be the outputs of Alice and Bob, and let (W_0, W_1, W_2) correspond to the messages of rates (R_0, R_1, R_2) generated by the genies, with $W_{01} = (W_0, W_1)$, $W_{02} = (W_0, W_2)$ and $W_{012} = (W_0, W_1, W_2)$. Also, let $T \sim \text{Unif}\{1, \ldots, n\}$ be independent of $(U^n, V^n, W_{012}, Z_1^n, Z_2^n)$. By the definition of unconditional samplability, for any $\epsilon > 0$ there exists a sufficiently large n such that $\mathbb{P}(Z_1^n \neq Z_2^n) \leq \epsilon$, so by Fano's inequality,

$$H(Z_1^n|Z_2^n) \leq h(\epsilon) + \epsilon(n \log |\mathcal{Z}| - 1) \stackrel{\triangle}{=} n\delta, \tag{4}$$

where $\delta \to 0$ as $\epsilon \to 0$. We have

$$\begin{aligned} n(R_0 + R_1 + \epsilon) &\geq H(W_{01}) \geq H(W_{01}|U^n) \\ &= I(W_{01}; V^n|U^n) + H(W_{01}|U^n, V^n). \end{aligned} \tag{5}$$

But using the fact that $Z_1^n - W_{01}U^n - V^n$ is a Markov Chain, and the fact that (U^n, V^n) are drawn i.i.d.,

$$\begin{aligned} I(W_{01}; V^n|U^n) &= I(Z_1^n, W_{01}; V^n|U^n), \\ &\geq I(Z_1^n; V^n|U^n) \qquad (6) \\ &= \sum_{i=1}^n I(Z_1^n, U^{i-1}, U_{i+1}^n, V^{i-1}; V_i|U_i) \\ &\geq \sum_{i=1}^n I(Z_1^n, U^{i-1}, V^{i-1}; V_i|U_i) \\ &= nI(Z_1^n, U^{T-1}, V^{T-1}, T; V_T|U_T). \qquad (7) \end{aligned}$$

For the second term in (5), note first that

$$\begin{aligned} &H(Z_1^n|W_{01}U^nV^n) \\ &\quad = H(Z_1^n|Z_2^n, W_{01}, U^n, V^n) + I(Z_1^n; Z_2^n|W_{01}, U^n, V^n) \\ &\quad \leq H(Z_1^n|Z_2^n) + I(Z_1^n; Z_2^n|W_{01}, U^n, V^n) \leq n\delta, \end{aligned}$$

by (4) and the Markov chain $Z_1^n - W_{01}U^n - (V^n, Z_2^n)$. Therefore

$$\begin{aligned} H(W_{01}|U^nV^n) &\geq I(W_{01}; Z_1^n|U^n, V^n) \\ &\geq H(Z_1^n|U^n, V^n) - n\delta. \end{aligned} \tag{8}$$

Substituting (7) and (8) in (5), and using the previous argument we obtain the following:

$$\begin{aligned} R_0 + R_1 \geq\ & I(Z_1^n, U^{T-1}, V^{T-1}, T; V_T|U_T) \\ &+ n^{-1}H(Z_1^n|U^n, V^n) - \delta - \epsilon. \end{aligned} \tag{9}$$

For the other rate constraint, note that $H(Z_2^n|U^n, V^n) \geq H(Z_1^n, Z_2^n|U^n, V^n) - H(Z_1^n|Z_2^n) \geq H(Z_1^n|U^n, V^n) - \delta$. Further, the term corresponding to (6), namely, $I(Z_2^n; U^n|V^n)$ can be written as $I(Z_2^n; U^n|V^n) = I(Z_1^n, Z_2^n; U^n|V^n) - I(Z_1^n; U^n|Z_2^n, V^n) \geq I(Z_1^n; U^n|V^n) - H(Z_1^n|Z_2^n) \geq I(Z_1^n; U^n|V^n) - \delta$.

$$\begin{aligned} R_0 + R_2 \geq\ & I(Z_1^n, U^{T-1}, V^{T-1}, T; U_T|V_T) \\ &+ n^{-1}H(Z_1^n|U^n, V^n) - 3\delta - \epsilon. \end{aligned} \tag{10}$$

Let $Q = (Z_1^n, U^{T-1}, V^{T-1}, T)$ and $R = n^{-1}H(Z_1^n|U^n, V^n)$ to recover the first two constraints of the rate region.

For the last constraint,

$$\begin{aligned} nH(Z) &\approx H(Z_1^n) \\ &= I(Z_1^n; U^n, V^n) + H(Z_1^n|U^n, V^n) \\ &= \sum_{i=1}^{n} I(Z_1^n; U_i, V_i|U^{i-1}, V^{i-1}) + nR \\ &= \sum_{i=1}^{n} I(Z_1^n, U^{i-1}, V^{i-1}; U_i, V_i) + nR \\ &= nI(Q; U_T, V_T) + nR. \end{aligned}$$

The cardinality bound can be shown using the support lemma [18, p. 631], [19, p. 310]. ∎

C. Relationship to $H(U, V)$

Proof: Suppose that $p(z)$ is unconditionally simulatable with rates (R_0, R_1, R_2). Then from (1) and (3) we have

$$\begin{aligned} R_0 + R_1 &\geq I(Q; V|U) + R && (11) \\ &\geq I(Q; V|U) + H(Z) - I(Q; U, V) && (12) \\ &= H(Z) - I(Q; U) && (13) \\ &\geq H(Z) - H(U). && (14) \end{aligned}$$

Similarly, $R_0 + R_2 \geq H(Z) - H(V)$.

Suppose $H(Z) \geq H(U, V)$ and we set $Q' = (U, V)$ in the achievability scheme and $R' = H(Z) - H(U, V)$. We want to check if the rate tuple (R_0, R_1, R_2) is still achievable for this choice of Q' and R'. The condition on $H(Z)$ is clearly satisfied. For this setting, we have

$$I(Q'; V|U) + R' = H(V|U) + R' = H(Z) - H(U).$$

Therefore $R_0 + R_1 \geq I(Q'; V|U) + R'$, and similarly $R_0 + R_2 \geq I(Q'; U|V) + R'$ so we can, without loss of generality, take the auxiliary random variable equal to (U, V) in this case.

Instead, suppose that $H(Z) \leq H(U, V)$. We want to show that we can take the excess rate $R = 0$ in this case. Suppose that (1) – (3) are satisfied for some variable Q and some excess rate $R > 0$. Define the variable Q' to be Q with probability $1 - \beta$ and (U, V) with probability β along with an indicator of which of these it is. Choose $\beta = \min\left(\frac{R}{H(U,V|Q)}, 1\right)$. Then

$$\begin{aligned} I(Q'; U, V) &= H(U, V) - H(U, V|Q') \\ &= H(U, V) - (1 - \beta)H(U, V|Q) \\ &= I(Q; U, V) + \beta H(U, V|Q) \\ &= \min\left(I(Q; U, V) + R, H(U, V)\right) \\ &\geq H(Z). \end{aligned}$$

Now consider

$$\begin{aligned} I(Q'; V|U) &= H(V|U) - H(V|U, Q') \\ &= H(V|U) - (1 - \beta)H(V|U, Q) \\ &= I(Q; V|U) + \beta H(V|U, Q) \\ &\leq I(Q; V|U) + R. \end{aligned}$$

This argument implies that (1) – (3) are satisfied for the auxiliary variable Q' and excess rate 0. ∎

V. Discussion

In this paper we studied a distributed sampling problem in which a genie can help two agents observing correlated sources to output correlated sequences. This problem is related to many other problems studied previously, including non-interactive source simulation, interactive channel simulation, and assisted common information. We identified two different cases: conditional sampling and unconditional sampling. For conditional sampling we described an achievable strategy and converse argument that in general do not coincide, but do in some special cases. For the unconditional case we give a complete characterization of the problem in the case where the two terminals must generate the same sequence with high probability.

References

[1] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[2] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 119–162, 1973.

[3] H. S. Witsenhausen, "On Sequences of Pairs of Dependent Random Variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, January 1975.

[4] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Trans. on Inform. Theory*, vol. 57, no. 10, pp. 6351–6355, October 2011.

[5] S. Kamath and V. Anantharam, "Non-interactive simulation of joint distributions : The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon," in *Proc. 50th Annual Allerton Conf. on Comm., Control and Comp.*, Monticello, IL USA, October 2012.

[6] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. on Inform. Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[7] V. Anantharam and V. Borkar, "Common randomness and distributed control : A counterexample," *Systems and Control Letters*, vol. 56, no. 7-8, pp. 568–672, July 2007.

[8] V. M. Prabhakaran and M. M. Prabhakaran, "Assisted Common Information with Applications to Secure Two-Party Computation," in *Proc. IEEE Int. Symp. on Inform. Theory*, 2010.

[9] ——, "Assisted common information with an application to secure two-party sampling," ArXiV, Tech. Rep. arXiv:1206.1282 [cs.IT], June 2012. [Online]. Available: http://arxiv.org/abs/1206.1282

[10] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal., "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. on Inform. Theory*, vol. 48, no. 10, pp. 2637–2655, October 2002.

[11] P. Cuff, "Distributed channel synthesis," ArXiV, Tech. Rep. arXiv:1208.4415 [cs.IT], August 2012.

[12] A. A. Gohari and V. Anantharam, "Generating dependent random variables over networks," in *Proc. IEEE Inform. Theory Workshop*, Paraty, Brazil, October 2011, pp. 698–702.

[13] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *Proc. IEEE Int. Symp. on Inform. Theory*, 2012.

[14] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. on Inform. Theory*, vol. 56, no. 9, pp. 4181–4205, September 2010.

[15] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref, "Coordination via a relay," in *Proc. IEEE Int. Symp. on Inform. Theory*, 2012.

[16] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," ArXiV, Tech. Rep. arXiv:1203.0730 [cs.IT], 2012.

[17] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. on Inform. Theory*, 2008.

[18] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge University Press, 2011.

[19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémi Kiadó, 1982.