**Communications and Signal Processing Laboratory (CSPL) Technical Report No. 374,
University of Michigan at Ann Arbor, July 25, 2006.**

On Large Deviation Analysis of Sampling from Typical Sets

Dinesh Krithivasan and S. Sandeep Pradhan

***Abstract*— Consider a pair of correlated sources (X, Y). With the collection of typical sequences of X and Y, one can associate a nearly semi-regular bipartite graph. The typical sequences of X and Y form vertices, and two sequences are connected by an edge if they are jointly typical. In this work, we study the structural properties of these graphs. In particular, we study regularity and scarcity of this graph by considering the asymptotic properties of samples taken from this graph. These results find applications in certain frameworks for transmission of correlated sources over multiuser channels.**

I. INTRODUCTION

A fundamental concept which has been instrumental in the development of information theory is the notion of typicality. As an illustration, consider a pair (X, Y) of discrete memoryless stationary correlated sources with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and a generic probability distribution $p_{X,Y}(\cdot,\cdot)$. An n-length sequence $x^n \in \mathcal{X}^n$ is said to be typical (or individually typical) if its empirical histogram is close to the marginal distribution p_X of X. Similarly one can define typical sequences for the source Y. Further, a sequence pair $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly typical if its empirical joint histogram is close to the distribution $p_{X,Y}$. It has been well-known [1] that for sufficiently large n, (a) there are roughly $2^{nH(X)}$ (respectively $2^{nH(Y)}$) individually typical sequences in $\mathcal{X}^n$ (respectively $\mathcal{Y}^n$), (b) there are roughly $2^{nH(X,Y)}$ jointly typical sequence pairs in $\mathcal{X}^n \times \mathcal{Y}^n$, and (c) roughly for every typical sequence $\mathcal{X}^n$ (respectively $\mathcal{Y}^n$), there are $2^{nH(Y|X)}$ (respectively $2^{nH(X|Y)}$) typical sequences in $\mathcal{Y}^n$ (respectively $\mathcal{X}^n$) which are jointly typical. These observations naturally lead to an undirected bipartite nearly semi-regular [2] graph on the set of typical sequences. The individually typical sequences in $\mathcal{X}^n$ and $\mathcal{Y}^n$ form the vertex set of this graph, and a vertex pair is connected by an edge if it is a jointly typical sequence pair. This graph may be referred to as the typicality-graph [3] of (X, Y). Our goal in this paper is to get a deeper understanding of this fundamental graph associated with a pair of correlated sources. Although the size of this graph is astronomical (when n is large), one can analyze this graph by looking at the asymptotic behavior of samples taken from it.

Moreover, it turns out that the insight we get by looking at correlated sources in this way leads to some interesting frameworks for the problem of transmission of these sources over multiuser channels [3], [4], [5], [6], [7], [8], where bipartite graphs can be used to represent information in multiuser systems. In [3], [6], [7], [8], partial characterizations of the sets of all nearly semi-regular graphs that could be used to represent a pair of correlated sources for transmission over multiple-access channels and broadcast channels were obtained. Similarly, partial characterizations of the sets of all nearly semi-regular bipartite graphs whose edges can be reliably transmitted over multiple-access channels and broadcast channels were also obtained.

With these two motivations, we present two information-theoretic results in this paper regarding the regularity and the sparsity of the typicality-graph. The first result is obtained by doing the following experiment. Suppose we sample 2^{nR_1} sequences from the typical set of X independently with replacement, and similarly sample 2^{nR_2} sequences from the typical set of Y. The underlying typicality-graph induces a graph on these $2^{nR_1} + 2^{nR_2}$ sequences. We provide a characterization of the probability that this random subgraph is sparse. We consider two measures of sparsity. The first is that the graph has no edge and the other is that the number of edges is sufficiently small.

The second result is obtained by doing the following experiment. As before, suppose we sample M sequences from the typical set of X and N sequences from the typical set of Y, where M and N are fixed integers. We provide a characterization of the probability that the induced random subgraph is fully connected.

The organization of the rest of the paper is as follows.

In Section II, we present the mathematical preliminaries. These include 2 versions of Suen's correlation inequality and Lovász Local Lemma. In Section III we investigate the probability of there being no jointly typical sequences in a sampling from the respective typical sets of 2 random variables, i.e., the typicality graph has no edges. This question was treated in [9] where the authors showed that the probability of this event can be made arbitrarily small for sufficiently large n. We derive tighter upper and lower bounds for the probability of this event. We also generalize it to the case of more than 2 variables. Upper bounds on the probability of significantly less number of jointly typical sequences than expected are also derived. In Section IV, we treat the other extreme case of all pairs of sequences being jointly typical, i.e., the typicality graph is completely connected. In Section V, we discuss the applications of this result and conclude the paper.

A. Notation

A word about the notation used in this paper is in order. We use the notation of Csiszár and Körner ([1]) for types,

This work was supported by NSF grant (CAREER) CCF-0448115
D. Krithivasan and S. S. Pradhan are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48105, USA. email: `dineshk@umich.edu`, `pradhanv@eecs.umich.edu`

typical sets and conditional types. The ϵ-typical set of n-length x^n sequences is denoted $A_\epsilon^{(n)}(X)$. This is sometimes abbreviated as $A_\epsilon(X)$ or simply $A(X)$. The cardinality of set A is denoted by $\mid A \mid$.

The distance between 2 distributions defined on the same alphabet is defined as follows:

$$\mid P(x) - Q(x) \mid = \max_{a \in \mathcal{X}} \mid P(a) - Q(a) \mid \tag{1}$$

Denote by $N(a \mid x^n)$ the number of occurrences of $a \in \mathcal{X}$ in the sequence x^n. $N(a, b \mid x^n, y^n)$ is similarly defined as the number of joint occurrences of the pair $(a, b) \in \mathcal{X} \times \mathcal{Y}$ in the pair $x^n \times y^n$. We say that $y^n \in \mathcal{Y}^n$ has conditional type V given $x^n \in \mathcal{X}^n$ if $\forall a \in \mathcal{X}, b \in \mathcal{Y}$

$$N(a, b \mid x^n, y^n) \quad = \quad N(a \mid x^n) V(b \mid a) \tag{2}$$

For any given $x^n \in \mathcal{X}^n$ and stochastic matrix $V : \mathcal{X} \to \mathcal{Y}$, the set of sequences $y^n \in \mathcal{Y}^n$ having conditional type V given x^n will be called the V-shell of x^n, denoted by $T_V(x^n)$.

We use the notation $\sum_{\sim x_i}(\cdot)$ to denote that the summation is over all variables other than x_i.

II. MATHEMATICAL PRELIMINARIES

In this section, we present the mathematical tools used in the subsequent sections.

A. Suen's inequalities

Suen's inequalities provide bounds for the probabilities of sums of (possibly dependent) indicator random variables ([10], [11], [12]). They use the concept of a dependency graph. For a given finite family of indicator random variables, the dependency graph is constructed by denoting each random variable I_i with vertex i and an edge between vertices i and j if the indicator random variables I_i and I_j are dependent. One version of the inequality can be stated as below.

Theorem 2.1: Let $I_i \in \text{Be}(p_i)$, $i \in \mathcal{I}$ be a family of Bernoulli random variables having a dependency graph L with vertex set $\mathcal{I}$ and edge set $E(L)$. Let $X = \sum_i I_i$ and $\lambda = \mathbb{E}X = \sum_i p_i$. Moreover, write $i \sim j$ if $(i, j) \in E(L)$ and let $\Delta = \frac{1}{2} \sum \sum_{i \sim j} \mathbb{E}(I_i I_j)$ and $\delta = \max_i \sum_{k \sim i} p_k$. Then

$$P(X = 0) \quad \leq \quad \exp\left\{ -\min\left(\frac{\lambda^2}{8\Delta}, \frac{\lambda}{2}, \frac{\lambda}{6\delta} \right) \right\} \tag{3}$$

Under the same assumptions as in Theorem 2.1, it is also possible to obtain upper bounds for the lower tail of the distribution of X.

Theorem 2.2: With assumptions as in Theorem 2.1 and $0 \leq a \leq 1$, we have

$$\begin{aligned} & P(X \leq a\lambda) \\ & \leq \exp\left\{ -\min\left((1-a)^2 \frac{\lambda^2}{8\Delta + 2\lambda}, (1-a)\frac{\lambda}{6\delta} \right) \right\} \end{aligned} \tag{4}$$

B. Lovász Local Lemma

Lovász Local Lemma ([13]) provides a sufficient condition for showing that the probability that none of the events from a given collection of events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ occurs is positive.

Theorem 2.3: Let L be a dependency graph for events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ in a probability space. Suppose there exists $x_i \in [0, 1]$ for $1 \leq i \leq n$ such that

$$P(\mathcal{E}_i) \quad \leq \quad x_i \prod_{(i,j) \in E(L)} (1 - x_j) \tag{5}$$

Then, the probability that none of the events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ occurs is lower bounded by

$$P(\cap_{i=1}^n \bar{\mathcal{E}}_i) \quad \geq \quad \prod_{i=1}^n (1 - x_i) \tag{6}$$

Further improvement can be made by a more careful usage of Lovász Local Lemma as given by Janson in [11].

Define the function $\varphi(x)$, $0 \leq x \leq e^{-1}$ to be the smallest root of the equation

$$\varphi(x) = e^{x\varphi(x)} \tag{7}$$

$\varphi(x)$ is well-defined in $[0, e^{-1}]$ and in particular $\varphi(x) = 1 + x + O(x^2)$. With λ and δ defined as in Theorem 2.1 and defining $\tau \triangleq \max_i P(\mathcal{E}_i)$, we have the following version of Lovász Local Lemma.

Theorem 2.4: With definitions as before,

$$P(\cap_{i=1}^n \mathcal{E}_i) \quad \geq \quad \exp\{-\lambda\varphi(\delta + \tau)\} \tag{8}$$

III. SPARSE GRAPHS

We present the main result below. Without loss of generality, we can assume that $R_1 \geq R_2$ for the rest of the paper.

A. Summary of Results

Theorem 3.1: Suppose X and Y are two correlated finite-alphabet random variables with joint distribution $p(x, y)$. For any $\epsilon > 0$ and any positive real numbers R_1 and R_2 such that $R_1 + R_2 > I(X; Y)$, if two collections of sequences $\mathcal{C}_X$ and $\mathcal{C}_Y$ are generated with uniform distribution (with replacement) on the typical sets $A_\epsilon^{(n)}(X)$ and $A_\epsilon^{(n)}(Y)$ of size $\theta_1 = 2^{nR_1}$ and $\theta_2 = 2^{nR_2}$ respectively, then the number of jointly typical sequences U in this collection satisfy the following relation:

$$\begin{aligned} & \lim_{n \to \infty} \frac{1}{n} \log\log \frac{1}{P(U = 0)} \\ & \qquad \geq \min(R_2, R_1 + R_2 - I(X; Y)) \end{aligned} \tag{9}$$

In fact, this inequality can be shown to be satisfied with equality in the case of $R_2 \leq R_1 < I(X; Y)$.

One can also derive an upper bound on the tail of the distribution of the number of jointly typical sequences as follows.

Theorem 3.2: For any $\gamma > 0$, we have

$$\begin{aligned} & \lim_{n \to \infty} \frac{1}{n} \log\log \left[P\left(\frac{\mathbb{E}(U) - U}{\mathbb{E}(U)} \geq e^{-n\gamma} \right) \right]^{-1} \\ & \geq \begin{cases} R_1 + R_2 - I(X; Y) - \gamma & \text{if } R_1 < I(X; Y) \\ R_2 - \gamma & \text{if } R_1 > I(X; Y) \end{cases} \end{aligned} \tag{10}$$

B. Proof of the upper bound

We first prove the upper bound on the probability of non-existence (equation (9)) using Suen's inequality.

Proof: Let $X^n(i)$ and $Y^n(j)$ denote the i^{th} and j^{th} codewords in the random codebooks $\mathcal{C}_X$ and $\mathcal{C}_Y$ respectively.

For $1 \leq i \leq \theta_1$, $1 \leq j \leq \theta_2$ define the indicator random variables as follows.

$$U_{ij} = \begin{cases} 1 & (X^n(i), Y^n(j)) \in A_\epsilon^{(n)}(X,Y) \\ 0 & \text{else} \end{cases} \tag{11}$$

Then the number of jointly typical sequences becomes

$$U \triangleq \sum_{i=1}^{\theta_1} \sum_{j=1}^{\theta_2} U_{ij} \tag{12}$$

We are interested in the event $\{U = 0\}$. By using a suitable dependency graph, this probability can be bounded using Suen's inequality (equation (3)).

It is clear that the indicator random variables U_{ij} and $U_{i'j'}$ $(i' \neq i, j' \neq j)$ are independent since the codebooks are generated by picking sequences i.i.d with replacement. However the random variable U_{ij} is dependent on all other indicator random variables of the form $U_{i'j}$ and $U_{ij'}$. Hence the dependency graph has a very regular structure with vertices indexed by the ordered pair (i,j), $1 \leq i \leq \theta_1, 1 \leq j \leq \theta_2$. Each vertex (i,j) is connected to exactly $\theta_1 + \theta_2 - 2$ vertices all of which share one of the two indices i or j. If vertices (i,j) and (k,l) are connected (i.e. $\{i = k\}$ or $\{j = l\}$), we denote it by $\{i,j\} \sim \{k,l\}$.

Define the following quantities.

$$\alpha_{ij} \triangleq P(U_{ij} = 1) = P((X^n(i), Y^n(j)) \in A_\epsilon^{(n)}(X,Y)) \tag{13}$$

$$\beta_{\{ij\}\{kl\}} \triangleq \mathbb{E}(U_{ij}U_{kl}) \quad \text{where } \{i,j\} \sim \{k,l\} \tag{14}$$

We can easily derive uniform bounds for these quantities.

$$\alpha \triangleq 2^{-n(I(X;Y)+\epsilon_1)} \leq \alpha_{ij} \leq 2^{-n(I(X;Y)-\epsilon_1)} \triangleq \alpha' \tag{15}$$

where $\epsilon_1(\epsilon)$ is a continuous positive function of ϵ such that $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$. The upper bound for α_{ij} is valid for all n while the lower bound is valid for sufficiently large n (see [14]).

Similarly $\beta_{\{ij\}\{kl\}}$ can be bounded uniformly (see Appendix) to give

$$2^{-2n(I(X;Y)+2\epsilon_2)} \leq \beta_{\{ij\}\{kl\}} \leq 2^{-2n(I(X;Y)-2\epsilon_2)} \triangleq \beta \tag{16}$$

The quantities involved in equation (3) can now be estimated as below.

$$\lambda \triangleq \mathbb{E}(U) \geq \theta_1\theta_2\alpha \tag{17}$$

$$\begin{aligned} \Delta &\triangleq \frac{1}{2} \sum_{\{i,j\}} \sum_{\{k,l\}\sim\{i,j\}} \mathbb{E}(U_{ij}U_{kl}) \\ &\leq \frac{1}{2}\theta_1\theta_2(\theta_1 + \theta_2 - 2)\beta \end{aligned} \tag{18}$$

$$\begin{aligned} \delta &\triangleq \max_{\{i,j\}} \sum_{\{k,l\}\sim\{i,j\}} \mathbb{E}(U_{kl}) \leq \max_{\{i,j\}}(\theta_1 + \theta_2 - 2)\alpha' \\ &= (\theta_1 + \theta_2 - 2)\alpha' \end{aligned} \tag{19}$$

The three terms in the exponent in equation (3) can be bounded to give

$$\frac{\lambda^2}{8\Delta} \geq \frac{1}{8}2^{n(R_2 - \epsilon')} \tag{20}$$

$$\frac{\lambda}{2} \geq \frac{1}{2}2^{n(R_1+R_2-I(X;Y)-\epsilon_1)} \tag{21}$$

$$\frac{\lambda}{6\delta} \geq \frac{1}{12}2^{n(R_2-2\epsilon_1)} \tag{22}$$

where $\epsilon' \triangleq 2(\epsilon_1 + 2\epsilon_2)$.

Substitution of these bounds into equation (3) gives us equation (9).

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log\log \frac{1}{P(U=0)} \\ &\qquad \geq \min(R_2, R_1 + R_2 - I(X;Y)) \end{aligned} \tag{23}$$

C. Proof of the lower bound

We will now derive a lower bound for the probability of non-existence of any jointly typical sequences using Lovász Local Lemma. We employ Lovász Local Lemma (Theorem 2.3) to the $\theta_1\theta_2$ events as defined by the equation (11). Because of symmetry, we can set all x_i in equation (5) to be the same. Then, Theorem 2.3 yields the following. Suppose there exists $0 \leq x \leq 1$ such that

$$\alpha \leq P(U_{ij} = 1) \leq x(1-x)^{(\theta_1+\theta_2-2)} \tag{24}$$

Then

$$P(U = 0) \geq (1-x)^{\theta_1\theta_2} \tag{25}$$

Simple Calculus can be used to show that the maximum value of the function $f(x) = x(1-x)^n$ is less than $(ne)^{-1}$. Hence for equation (24) to have a solution at all, we need $\alpha \leq ((\theta_1 + \theta_2 - 2)e)^{-1}$. This is equivalent to the assumption that $R_2 \leq R_1 < I(X;Y)$. Under this assumption, choosing $x = {\theta_1}^{-1}$ satisfies equation (24) and hence from equation (25), we have

$$P(U=0) \geq \exp(-(\theta_2 + 1)) \tag{26}$$

Further improvement of the lower bound can be achieved by using the other version of Lovász Local Lemma given in Theorem 2.4. λ and δ are as defined in equations (17) and (19). In this case, $\tau = \max_{\{ij\}} P(U_{ij} = 1)$ and this is upper bounded by α' from equation (15). Using Theorem 2.4, we can write

$$P(U=0) \geq \exp(-\lambda\varphi(\delta + \tau)) \tag{27}$$

Combining equations (26) and (27) we have

$$\begin{aligned} -\log P(U=0) &\leq \min\left(\theta_2 + 1, \lambda\varphi(\delta+\tau)\right) \\ &\leq \min(\theta_2 + 1, \theta_1\theta_2\alpha'\varphi(\delta+\tau)) \end{aligned} \tag{28}$$

Under the assumption $R_2 < R_1 < I(X;Y)$, $\delta + \tau \leq (\theta_1 + \theta_2 - 1)\alpha' \to 0$ as $n \to \infty$ and hence $\varphi(\delta + \tau) \to 1$. Taking logarithms across equation (28) and letting $n \to \infty$ gives us

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n} \log\log P(U=0) \\ &\qquad \leq \min(R_2, R_1 + R_2 - I(X;Y)) \end{aligned} \tag{29}$$

Comparing equations (23) and (29), we see that under the assumption $R_2 < R_1 < I(X;Y)$, the result of Theorem 3.1 is tight.

■

D. Tail Estimates

Using Theorem 2.2, one can also compute an upper bound on the probability that there exists less number of jointly typical sequences than expected. The terms being minimized in the exponent in equation (4) can be approximated as

$$\begin{aligned} \frac{\lambda^2}{8\Delta + 2\lambda} &\geq \frac{\theta_1\theta_2\alpha^2}{2\alpha' + 4(\theta_1+\theta_2-2)\beta} \\ &\geq \frac{1}{8}\frac{\theta_1\theta_2\alpha^2}{\alpha' + \theta_1\beta} \end{aligned} \tag{30}$$

$$\geq \begin{cases} \frac{1}{16}2^{n(R_1+R_2-I(X;Y)-3\epsilon_1)} & \text{if } R_1 < I(X;Y) \\ \frac{1}{16}2^{n(R_2-(2\epsilon_1+4\epsilon_2))} & \text{if } R_1 > I(X;Y) \end{cases} \tag{31}$$

$$\frac{\lambda}{6\delta} \geq \frac{\theta_1\theta_2\alpha}{6(\theta_1+\theta_2-2)\alpha'} \geq \frac{1}{12}2^{n(R_2-2\epsilon_1)} \tag{32}$$

For $0 \leq a \leq 1$, the exponent can be calculated as

$$\begin{aligned} &-\log P(U \leq a\lambda) \\ &\geq \begin{cases} (\frac{1-a^2}{16})2^{n(R_1+R_2-I(X;Y)-3\epsilon_1)} & \text{if } R_1 < I(X;Y) \\ (\frac{1-a}{16})2^{n(R_2-(2\epsilon_1+4\epsilon_2))} & \text{if } R_1 > I(X;Y) \end{cases} \end{aligned} \tag{33}$$

Note that we are usually interested in the event that there are significantly less number of jointly typical sequences than expected. Keeping this in mind, choose $a = 1 - e^{-n\gamma}$. Then, we have that

$$\lim_{n\to\infty} \frac{1}{n}\log\log\left[P\left(\frac{\mathbb{E}(U)-U}{\mathbb{E}(U)} \geq e^{-n\gamma}\right)\right]^{-1}$$

$$\geq \begin{cases} R_1 + R_2 - I(X;Y) - \gamma & \text{if } R_1 < I(X;Y) \\ R_2 - \gamma & \text{if } R_1 > I(X;Y) \end{cases} \tag{34}$$

E. Generalization to several random variables

The results of the previous sections can be easily generalized to the case of more than 2 random variables. In this case, the quantity $A(X_1;X_2;\ldots;X_n)$ defined below plays a role similar to that of mutual information in the 2 variable case.

$$A(X_1;\ldots;X_n) \triangleq \sum_{i=1}^{n} H(X_i) - H(X_1,\ldots,X_n) \tag{35}$$

The case of 3 random variables is illustrated below. We state the following result without proof. The result follows from using Theorem 2.1 in the same way as it was used to prove Theorem 3.1.

Theorem 3.3: Suppose X, Y and Z are three correlated finite-alphabet random variables with joint distribution $p(x,y,z)$. For any $\epsilon > 0$ and any positive real numbers $R_1 > R_2 > R_3$ such that $R_1 + R_2 + R_3 > A(X;Y;Z)$, if three collections of sequences $\mathcal{C}_X$, $\mathcal{C}_Y$ and $\mathcal{C}_Z$ are generated with uniform distribution (with replacement) on the typical sets $A_\epsilon^{(n)}(X)$, $A_\epsilon^{(n)}(Y)$ and $A_\epsilon^{(n)}(Z)$ of size $\theta_1 = 2^{nR_1}$, $\theta_2 = 2^{nR_2}$ and $\theta_3 = 2^{nR_3}$ respectively, then the number of jointly typical sequences U in this collection satisfy the following relation:

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n}\log\log\frac{1}{P(U=0)} \\ &\qquad \geq \min(R_1+R_2+R_3 - A(X;Y;Z), R_3, \Gamma) \end{aligned} \tag{36}$$

where

$$\begin{aligned} \Gamma \triangleq\ & R_1 + R_2 + R_3 - \\ & \max(R_1+R_2, R_1+R_3, R_2+R_3, R_1+I(Y;Z), \\ & \qquad R_2 + I(X;Z), R_3 + I(X;Y)) \end{aligned} \tag{37}$$

Upper bounds on the tail probabilities can be derived similarly.

IV. FULLY CONNECTED GRAPH

In this section, we investigate the other extreme case of every pair of sequences picked from the respective typical sets being jointly typical. Suppose we pick M sequences from the typical set $A_\epsilon^{(n)}(X)$ and N sequences from the typical set $A_\epsilon^{(n)}(Y)$(independently with replacement). We are interested in the probability that all MN pairs of sequences are jointly typical.

A. Summary of Results

Theorem 4.1: Let FC be the event that all MN pairs of sequences are jointly typical. The main result of this section is the following upper bound on the probability of the event FC.

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n}\log P(FC) &\geq (M+N-1)I(X;Y) \\ &+ \min_{\mathcal{P}} (N-1)I(Y;X_2,\ldots X_M \mid X_1) \\ &+ A(X_1;\ldots;X_m \mid Y) \end{aligned} \tag{38}$$

where X_i, $1 \leq i \leq M$ are random variables with distribution P_X. The minimization is over the family $\mathcal{P}$ of conditional distributions $P_{X_1,\ldots,X_M|Y}$.

B. Proofs

Let $X^n(i)$ be the x^n sequence picked at the i^{th} drawing. Let P_{XY} be the joint distribution of X and Y with marginals as P_X, P_Y. We derive an upper bound for the event FC by successively conditioning on the conditional type class with respect to the earlier picked sequences. This is illustrated below.

$$P(FC) = \sum_{\bar{P}\in\mathcal{G}_1} \frac{P(FC \mid P_{x^n} = \bar{P}) \mid T_{\bar{P}} \mid}{\mid A(X) \mid} \tag{39}$$

The summation in equation (39) runs over all type classes $\bar{P}$ which belong to the collection $\mathcal{G}_1$ defined below.

$$\mathcal{G}_1 \triangleq \{P : \mid P - P_X \mid \leq \epsilon\} \tag{40}$$

We now condition on the conditional type of the second sequence picked with respect to the first sequence. Note that the V- shell given $X^n(1)$ in which $X^n(2)$ lies must be such that $P_{X^n(2)}$ is close to the distribution P_X. This can be expressed as the following constraint on V_2.

$$\mathcal{G}_2 \triangleq \{V_2 \ : | \sum_{\sim x_1} \bar{P}V_2 - P_X | \leq \epsilon\} \tag{41}$$

Equation (39) can now be further conditioned as

$$P(FC) = \frac{1}{| A(X) |} \sum_{\bar{P} \in \mathcal{G}_1} | T_{\bar{P}} | \sum_{V_2 \in \mathcal{G}_2} \frac{| T_{V_2} |}{| A(X) |} \times P(FC \mid \bar{P}, V_2) \tag{42}$$

In the above equation $P(FC \mid \bar{P}, V_2)$ denotes the conditional probability of the event FC given the first sequence belongs to type class $\bar{P}$ and the second sequence has conditional type V_2 given the first sequence.

This conditioning process can be continued till we have conditioned on $V_2, \ldots, V_M$. Each of these has to satisfy a constraint similar to equation (41). For $2 \leq i \leq M$, the general form of the constraint can be expressed as

$$\mathcal{G}_i \triangleq \{V_i \ : | \sum_{\sim x_i} \bar{P}V_2 \ldots V_i - P_X | \leq \epsilon\} \tag{43}$$

After the conditioning on $\bar{P}, V_2, \ldots, V_M$, the probability can be written as

$$P(FC) = \frac{1}{| A(X) |} \sum_{\bar{P} \in \mathcal{G}_1} | T_{\bar{P}} | \sum_{V_2 \in \mathcal{G}_2} \frac{| T_{V_2} |}{| A(X) |} \cdots \cdots \sum_{V_M \in \mathcal{G}_M} \frac{| T_{V_M} |}{| A(X) |} \times P(FC \mid \bar{P}, V_2, \ldots, V_M) \tag{44}$$

Now, we turn to the question of evaluating $P(FC \mid \bar{P}, V_2, \ldots, V_M)$. For the event FC to occur, each of the N y^n sequences must be picked from a conditional type class V that ensures that the y^n sequence picked is jointly typical with all the M x^n sequences. This requirement places the following system of constraints on the conditional type class V.

$$\mathcal{G} \triangleq \{V \ : | \sum_{\sim x_i} \bar{P}V_2 \ldots V_M V - P_{XY} | \leq \epsilon \quad 1 \leq i \leq M\} \tag{45}$$

Each y^n sequence can be picked from any conditional type class $V \in \mathcal{G}$ and be jointly typical with all M x^n sequences. Thus

$$P(FC \mid \bar{P}V_2, \ldots, V_M) = \left(\frac{\sum_{V \in \mathcal{G}} | T_V |}{| A(Y) |} \right)^N \tag{46}$$

Substituting equation (46) in (44), we have the final expression.

$$P(FC) = \frac{1}{| A(X) |} \sum_{\bar{P} \in \mathcal{G}_1} | T_{\bar{P}} | \sum_{V_2 \in \mathcal{G}_2} \frac{| T_{V_2} |}{| A(X) |} \cdots \cdots \sum_{V_M \in \mathcal{G}_M} \frac{| T_{V_M} |}{| A(X) |} \times \left(\frac{\sum_{V \in \mathcal{G}} | T_V |}{| A(Y) |} \right)^N \tag{47}$$

We now use the fact that there are only polynomial number of type classes and conditional type classes to bound equation (47) as

$$P(FC) \leq \frac{2^{n\epsilon'} \max | T_{\bar{P}} || T_{V_2} | \cdots | T_{V_M} | | T_V |^N}{| A(X) |^M | A(Y) |^N} \tag{48}$$

where the maximum is over all the types and conditional types $\bar{P} \in \mathcal{G}_1, V_i \in \mathcal{G}_i, V \in \mathcal{G}$ and $\epsilon' \to 0$ as $n \to \infty$.

We use the following bounds for the different conditional typical sets (see [1]).

$$\begin{aligned} | T_{V_i}(x_i^n \mid x_1^n, \ldots, x_{i-1}^n) | &\leq 2^{n(H(V_i|\bar{P}V_2 \ldots V_{i-1})+\delta)} \\ | T_V(y^n \mid x_1^n, \ldots, x_M^n) | &\leq 2^{n(H(V|\bar{P}V_2 \ldots V_M)+\delta)} \end{aligned} \tag{49}$$

Using these bounds, equation (48) can be written as

$$\begin{aligned} P(FC) \leq & \frac{1}{| A(X) |^M | A(Y) |^N} 2^{n\epsilon'} \max 2^{n(H(\bar{P})+\delta)} \\ & \times 2^{n(H(V_2|\bar{P})+\delta)} \ldots 2^{n(H(V_M|\bar{P}V_2 \ldots V_{M-1})+\delta)} \\ & \times 2^{n(H(V|\bar{P}V_2 \ldots V_M)+\delta)} \end{aligned} \tag{50}$$

Define random variables $X_1, \ldots, X_M$ such that X_i has distribution P_X for $1 \leq i \leq M$. Further, let the joint distribution of X_i and Y be as P_{XY} for all $1 \leq i \leq M$. Denote the distribution $P_{X_1,\ldots,X_M|Y}$ by $\mathcal{P}$.

By continuity of entropy, equation (50) can be written in terms of these random variables as

$$\begin{aligned} P(FC) \leq & \frac{1}{| A(X) |^M | A(Y) |^N} 2^{n\epsilon'} \max_{\mathcal{P}} 2^{n(H(X_1)+\delta')} \\ & \times 2^{n(H(X_2|X_1)+\delta')} \ldots 2^{n(H(X_M|X_1 \ldots X_{M-1})+\delta')} \\ & \times 2^{n(H(Y|X_1,\ldots,X_M)+\delta')} \end{aligned} \tag{51}$$

where the maximization is over all distributions in $\mathcal{P}$.

Using the bound $| A(X) | \geq 2^{n(H(X)-\epsilon)}$ and a similar bound for $| A(Y) |$, we can simplify equation (51) as

$$\begin{aligned} P(FC) \leq \quad & \max_{\mathcal{P}} \quad 2^{n\delta''} \times 2^{-n[H(X_2)-H(X_2|X_1)]} \\ \times \quad & 2^{-n[H(X_3)-H(X_3|X_1X_2)]} \ldots \\ \times \quad & 2^{-n[H(X_M)-H(X_M|X_1,\ldots X_{M-1})]} \\ \times \quad & 2^{-nN[H(Y)-H(Y|X_1,\ldots,X_M)]} \end{aligned} \tag{52}$$

where $\delta'' \triangleq (M + N - 1)(\delta' + \epsilon)$ and the maximization is over the distributions $P_{X_i|X_1 \ldots X_{i-1}}$ for $1 \leq i \leq M$ and $P_{Y|X_1 \ldots X_M}$.

The exponent of the probability can be lower bounded as

$$\begin{aligned} -\frac{1}{n} \log P(FC) \geq \quad & \min_{\mathcal{P}} \quad I(X_1; X_2) + I(X_1X_2; X_3) + \ldots \\ + \quad & I(X_1X_2 \ldots X_{M-1}; X_M) \\ + \quad & NI(Y; X_1X_2 \ldots X_M) - \delta'' \end{aligned} \tag{53}$$

Further simplification yields

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n} \log P(FC) \quad \geq \quad & (M + N - 1)I(X; Y) \\ + \quad & \min_{\mathcal{P}} (N - 1)I(Y; X_2, \ldots X_M \mid X_1) \\ + \quad & A(X_1; \ldots ; X_M \mid Y) \end{aligned} \tag{54}$$

V. CONCLUSIONS

The main results of this paper are Theorem 3.1, Theorem 3.2 and Theorem 4.1. With Theorem 3.1, we show that the probability of there being no jointly typical sequence in a sampling from the respective typical sets of 2 random variables goes to 0 double exponentially. In Theorem 3.2, we bound the probability of there being significantly less number of jointly typical sequences than expected. In Theorem 4.1, we investigate the contrasting case where all pairs of sequences in a sampling are jointly typical.

VI. ACKNOWLEDGEMENT

The authors would like to thank Prakash Ishwar of Boston University for discussions.

VII. APPENDIX

We derive a uniform bound on $\beta_{\{ij\}\{kl\}}$ where $\{i,j\} \sim \{k,l\}$. Without loss of generality, let us derive the bound of $\beta_{\{ij\}\{il\}}$.

We define $A_\epsilon(x^n)$ as the set of all sequences y^n such that $(x^n, y^n) \in A_\epsilon(X,Y)$

$$\begin{aligned}\beta_{\{ij\}\{il\}} &= \mathbb{E}(U_{ij}U_{il}) \\ &= P(U_{ij}=1 \text{ and } U_{il}=1)\end{aligned}$$

$$\begin{aligned}&= \sum_{x^n \in A_\epsilon(X)} \frac{P(Y^n(j), Y^n(l) \in A_\epsilon(x^n) \mid X^n(i) = x^n)}{\mid A(X) \mid} \\ &= \frac{1}{\mid A_\epsilon(X) \mid} \sum_{x^n \in A_\epsilon(X)} \left(\frac{\mid A_\epsilon(x^n) \mid}{\mid A_\epsilon(Y) \mid} \right)^2 \qquad (55)\end{aligned}$$

¿From [1], we have a uniform bound on $\mid A_\epsilon(x^n) \mid$ where $x^n \in A_\epsilon(X)$.

$$2^{n(H(Y|X)-\tilde{\epsilon})} \leq \mid A_\epsilon(x^n) \mid \leq 2^{n(H(Y|X)+\tilde{\epsilon})} \qquad (56)$$

where $\tilde{\epsilon}(\epsilon)$ is a continuous positive function of ϵ such that $\tilde{\epsilon}(\epsilon) \to 0$ as $\epsilon \to 0$. The upper bound is true for all n while the lower bound is true for sufficiently large n.

Substituting equation (56) in equation (55) gives us equation (16).

REFERENCES

[1] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press Inc. Ltd., 1981.

[2] J. H. van Lint and R. M. Wilson, *A course in combinatorics*. Cambridge University Press, 2003.

[3] S. S. Pradhan, S. Choi, and K. Ramchandran, "A graph-based framework for transmission of correlated sources over multiple access channels," *submitted to IEEE Trans. Inform. Theory*, January 2006.

[4] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.

[5] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 396–412, May 1983.

[6] S. S. Pradhan, S. Choi, and K. Ramchandran, "Achievable rates for multiple-access channels with correlated messages," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT), Chicago, IL*, June 2004.

[7] S. Choi and S. S. Pradhan, "Transmission of correlated messages over the broadcast channel," in *Proc. Conf. on Inform. Sciences and Systems (CISS), Baltimore, MD*, January 2005.

[8] S. Choi and S. S. Pradhan, "A graph-based framework for transmission of correlated sources over broadcast channels," in *Proc. Information Theory and Applications Workshop (ITA), San Diego, CA*, February 2006.

[9] A. E. Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 120–122, January 1981.

[10] W. C. S. Suen, "A correlation inequality and a poisson limit theorem for nonoverlapping balanced subgraphs of a random graph," *Random Structures Algorithms*, vol. 1, pp. 231–242, 1990.

[11] S. Janson, "New versions of suen's correlation inequality," *Random Structures Algorithms*, vol. 13, pp. 467–483, 1998.

[12] S. Janson, T. Luczak, and A. Ruciński, *Random Graphs*. John Wiley and Sons, Inc., 2000.

[13] N. Alon and J. H. Spencer, *The Probabilistic Method*. John Wiley and Sons, Inc., 1992.

[14] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, Inc., 1990.