arXiv:1508.03349v2 [cs.IT] 21 Jan 2016

THE MULTIVARIATE COVERING LEMMA AND ITS CONVERSE

PARHAM NOORZAD, MICHELLE EFFROS, AND MICHAEL LANGBERG

Abstract. The multivariate covering lemma states that given a collection of k codebooks, each of sufficiently large cardinality and independently generated according to one of the marginals of a joint distribution, one can with probability arbitrarily close to one choose one codeword from each codebook such that the resulting k-tuple of codewords is jointly typical with respect to the joint distribution. Prior proofs of the multivariate covering lemma primarily employ strong typicality. We give a proof of this lemma for weakly typical sets. This allows achievability proofs that rely on the covering lemma to go through for continuous (e.g., Gaussian) channels without the need for quantization. The covering lemma and its converse are widely used in information theory, including in rate-distortion theory and in achievability results for multi-user channels.

1. Introduction

The covering lemma and its extensions play a crucial role in achievability results in network information theory. Covering lemmas are useful for enabling network nodes to transmit codewords that "look like" they are generated from a dependent distribution, whereas in reality, they are carefully selected from sufficiently large codebooks that are independently generated. This allows nodes to obtain the benefits of both independent and dependent codewords: like independent codewords, such codewords can be decoded in different locations; like dependent codewords they have the potential to achieve rates higher than those achieved by independent codewords. This benefit, however, comes at a cost in rate. Thus the strategy is useful when the benefit transmitting dependent codewords exceeds its cost.

In the context of the covering lemma, the concept of "looking like" dependent codewords is captured by the notion of being jointly typical with respect to a dependent distribution. As there are various ways to define the typical set (here we specifically focus on weakly typical [2] and strongly typical sets [3]), one may ask whether a specific version of the covering lemma holds for a given definition of the typical set. The weakly typical set has two advantages over the strongly typical set. First, it is easily defined for continuous (e.g., Gaussian) distributions. Second, the weakly typical set has a simple one-shot counterpart, which allows proofs using the weakly typical set to be written in the one-shot framework in a simple manner. On the other hand, some results hold for the strongly typical set that do not hold for the weakly typical set. Thus it is helpful to review the covering lemma and its extensions and see for which definition of the typical set each result is currently known to hold.

The simplest case of the covering lemma is the situation where given a random vector and an independently generated codebook, a node looks for a codeword in the codebook that is jointly typical (with respect to a dependent distribution) with the given random vector. The result obtained in this case, simply referred to as the "covering lemma", appears in the achievability proof of the rate distortion theorem using weakly typical sets [2]. The second case, called the "mutual covering lemma," treats the case where given two independently generated codebooks, a node looks for a jointly typical pair of codewords, where each codeword is from one of the codebooks. This result is used in Marton's inner bound for the two-user broadcast channel and is proved for strongly typical sets [4, 7]. Recently, by extending the proof of [2], the authors of [6, 8] prove a one-shot version of the mutual covering lemma. This proof can be used to show the validity of the mutual covering lemma for weakly typical sets in the asymptotic setting. The proof in [6, 8], however, requires stronger independence assumptions on the codebooks than the proof using strongly typical sets in [3, 4]. Finally, the "multivariate covering lemma" is the extension of the mutual covering lemma to k independently generated codebooks, and can be used to obtain an inner bound on the broadcast channel with k users [3]. As stated in [3], one can show this result holds for strongly typical sets by extending the proof of the mutual covering lemma [4].

In this work, using the general strategy of El Gamal and Van der Meulen [4] and some ideas regarding weakly typical sets from Koetter, Effros, and Médard [5], we give a proof of the multivariate covering lemma for weakly typical sets. We also provide a converse, a special case of which is usually referred to as the packing lemma [3]. We remark that while similar to the argument in [4], we use Chebyshev's inequality for the direct result (Section 4), it is also possible to use the Cauchy-Schwarz inequality (see Appendix A), which leads to a more accurate upper bound.

2. Problem Statement

For every positive integer n, define the set $[n] = \{1, \dots, n\}$. Let k be a positive integer and

$$p(u_0, u_1, \dots, u_k, u_{k+1})$$

be a probability distribution on the set

$$\prod_{j=0}^{k+1} \mathcal{U}_j.$$

For every nonempty $S \subseteq [k]$ define

$$\mathcal{U}_S = \prod_{j \in S} \mathcal{U}_j.$$

For every $j \in [k]$, let M_j be a nonnegative integer. For every nonempty $S \subseteq [k]$, define the set $\mathcal{M}_S$ as

$$\mathcal{M}_S = \prod_{j \in S} [M_j].$$

and let $\mathcal{M} = \mathcal{M}_{[k]}$. For every $\mathbf{m} = (m_1, \dots, m_k) \in \mathcal{M}$, let the random vector

$$(U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1})$$

have distribution

$$p(u_0) \prod_{j=1}^{k+1} p(u_j|u_0),$$

where $p(u_0)$ and each $p(u_j|u_0)$ are the conditional marginals of $p(u_0, \ldots, u_{k+1})$. In addition, let $\mathcal{F}$ be an arbitrary subset of $\mathcal{U}_0 \times \mathcal{U}_{[k+1]}$. We want to find upper and lower bounds on the probability

$$\mathbf{P}\Big\{\forall \mathbf{m} \in \mathcal{M} : \big(U_0, U_1(m_1), \ldots, U_k(m_k), U_{k+1}\big) \notin \mathcal{F}\Big\}.$$

We derive the lower bound (Section 3) using the union bound, which does not depend on the statistical dependencies of the vectors

$$\big(U_0, U_1(m_1), \ldots, U_k(m_k), U_{k+1}\big)$$

for different values of $\mathbf{m}$. For the upper bound (Section 4), which leads to the multivariate covering lemma, we require a stronger assumption, which we next describe.

Let $\mathbf{m} = (m_j)_{j\in[k]}$ and $\mathbf{m}' = (m'_j)_{j\in[k]}$ be in $\mathcal{M}$. Define the set $S_{\mathbf{m},\mathbf{m}'}$ as

$$S_{\mathbf{m},\mathbf{m}'} = \big\{j \in [k] : m_j = m'_j\big\}.$$

When $\mathbf{m}$ and $\mathbf{m}'$ are clear from context, we denote $S_{\mathbf{m},\mathbf{m}'}$ with S. In the proof of the upper bound we require

$$\mathbf{P}\Big\{\forall j \in [k] : U_j(m_j) = u_j \text{ and } U_j(m'_j) = u'_j \big| U_0 = u_0, U_{k+1} = u_{k+1}\Big\}$$
$$= \prod_{j=1}^{k} p(u_j|u_0) \times \prod_{j\in S^c} p(u'_j|u_0),$$

for all u_0 and all $(u_j)_j$ and $(u'_j)_j$ such that if $j \in S$, then $u_j = u'_j$ (Assumption I). Note that if there exists a $j \in S$ where $u_j \neq u'_j$ then the probability on the left hand side equals zero.

In the corresponding asymptotic problem (Section 5), we apply our bounds to

$$\mathbf{P}\Big\{\forall \mathbf{m} : \big(U_0^n, U_1^n(m_1), \ldots, U_k^n(m_k), U_{k+1}^n\big) \notin A_\delta^{(n)}\Big\},$$

where for every $\mathbf{m}$,

$$\big(U_0^n, U_1^n(m_1), \ldots, U_k^n(m_k), U_{k+1}^n\big)$$

is simply n i.i.d. copies of the original random vector

$$\big(U_0, U_1(m_1), \ldots, U_k(m_k), U_{k+1}\big),$$

(Assumption II) and $A_\delta^{(n)}$ is the weakly typical set for the distribution $p(u_0, u_1, \ldots, u_k, u_{k+1})$. Our main result follows.

Theorem 1 (Multivariate Covering Lemma)**.** *Suppose Assumptions (I) and (II) hold for the joint distribution of*

$$U_0^n, \big\{U_1^n(m_1), \ldots, U_k^n(m_k)\big\}_{\mathbf{m}}, U_{k+1}^n.$$

For the direct part, suppose for all $j \in [k]$, $M_j \geq e^{nR_j}$. *If for all nonempty* $S \subseteq [k]$,

$$\sum_{j\in S} R_j > \sum_{j\in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) + (8k - 2|S| + 10)\delta, \tag{1}$$

then

$$\lim_{n\to\infty} \mathbf{P}\Big\{\exists \mathbf{m} : \big(U_0^n, U_1^n(m_1), \ldots, U_k^n(m_k), U_{k+1}^n\big) \in A_\delta^{(n)}\Big\} = 1. \tag{2}$$

For the converse, assume for all $j \in [k]$, $M_j \leq e^{nR_j}$. If Equation (2) holds, then

$$\sum_{j\in S} R_j \geq \sum_{j\in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) - 2(|S|+1)\delta,$$

for all nonempty $S \subseteq [k]$.

In the direct part of Theorem 1, we can weaken the lower bound on $\sum_{j\in S} R_j$ when $S = [k]$. Specifically, we can replace Equation (1) with

$$\sum_{j=1}^{k} R_j > \sum_{j=1}^{k} H(U_j|U_0) - H(U_{[k]}|U_0, U_{k+1}) + 2(k+1)\delta.$$

for $S = [k]$.

3. The Lower Bound

For every $S \subseteq [k]$, define $\mathcal{F}_S$ as the projection of $\mathcal{F}$ on $\mathcal{U}_0 \times \mathcal{U}_S \times \mathcal{U}_{k+1}$. Then for every $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$, let $\mathcal{F}(u_0, u_S, u_{k+1})$ be the set of all u_{S^c} such that $(u_0, u_{[k]}, u_{k+1}) \in \mathcal{F}$. In addition, for every nonempty $S \subseteq [k]$, let α_S and β_S be constants such that

$$\alpha_S \leq \log \frac{p(u_S|u_0, u_{k+1})}{\prod_{j\in S} p(u_j|u_0)}$$

for all $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$ and

$$\beta_S \leq \log \frac{p(u_S|u_0, u_{S^c}, u_{k+1})}{\prod_{j\in S} p(u_j|u_0)}$$

for all $(u_0, u_S, u_{S^c}, u_{k+1}) \in \mathcal{F}$. Furthermore, let the constant γ satisfy

$$\gamma \geq \log \frac{p(u_{[k]}|u_0, u_{k+1})}{\prod_{j\in [k]} p(u_j|u_0)}$$

for all $(u_0, u_{[k]}, u_{k+1}) \in \mathcal{F}$.

For every $\mathbf{m} = (m_1, \ldots, m_k) \in \mathcal{M}$, define the random variable $Z_{\mathbf{m}}$ as

$$Z_{\mathbf{m}} = \mathbf{1}\Big\{\big(U_0, U_1(m_1), \ldots, U_k(m_k), U_{k+1}\big) \in \mathcal{F}\Big\}$$

and set

$$Z = \sum_{\mathbf{m}\in\mathcal{M}} Z_{\mathbf{m}}.$$

Our aim is to find a lower bound for $\mathbf{P}\{Z = 0\}$. Note that for every nonempty $S \subseteq [k]$,

$$\begin{aligned}
\mathbf{P}\left\{\exists \mathbf{m} : Z_{\mathbf{m}} = 1\right\} &= \mathbf{P}\Big\{\exists \mathbf{m} : \big(U_0, U_1(m_1), \ldots, U_k(m_k), U_{k+1}\big) \in \mathcal{F}\Big\} \\
&\leq \mathbf{P}\Big\{\exists \mathbf{m} : \big(U_0, \big(U_j(m_j)\big)_{j\in S}, U_{k+1}\big) \in \mathcal{F}_S\Big\} \\
&\leq |\mathcal{M}_S| \sum_{\mathcal{F}_S} p(u_0, u_{k+1}) \prod_{j\in S} p(u_j|u_0) \\
&\leq |\mathcal{M}_S| e^{-\alpha_S} \sum_{\mathcal{F}_S} p(u_0, u_S, u_{k+1}) \\
&\leq |\mathcal{M}_S| e^{-\alpha_S}.
\end{aligned}$$

Thus

$$\begin{aligned}\mathbf{P}\{Z=0\} &= 1-\mathbf{P}\left\{\exists \mathbf{m} : Z_{\mathbf{m}}=1\right\} \\ &\geq 1-\min_{|S|\neq\emptyset}|\mathcal{M}_S|e^{-\alpha_S}. \end{aligned} \tag{3}$$

4. The Upper Bound

In deriving our upper bound on $\mathbf{P}\{Z=0\}$, we apply conditioning and Chebyshev's inequality. Thus, the factor

$$\frac{1}{\big(\mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\}\big)^2}$$

appears, where

$$\begin{aligned}\mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\} &= \mathbf{P}\left\{U_{[k]}\in\mathcal{F}(u_0,u_{k+1})|U_0=u_0,U_{k+1}=u_{k+1}\right\} \\ &= \sum_{u_{[k]}\in\mathcal{F}(u_0,u_{k+1})} p(u_{[k]}|u_0,u_{k+1})\end{aligned}$$

and $\mathcal{F}(u_0,u_{k+1})$ (Section 3) is simply the set of all $u_{[k]}$'s that satisfy $(u_0,u_{[k]},u_{k+1})\in\mathcal{F}$. Thus to get a reasonably accurate upper bound, we require $\mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\}$ to be large. However, as we cannot guarantee this for all (u_0,u_{k+1}), we partition the (u_0,u_{k+1}) pairs into "good" and "bad" sets, corresponding to large and small values of $\mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\}$, respectively. The probability of the good set is large when $\mathbf{P}\{(U_0,U_{[k]},U_{k+1})\in\mathcal{F}\}$ is sufficiently large. To see this, fix $\epsilon>0$ and following Appendix III of [5], define the set $\mathcal{G}\subseteq\mathcal{U}_0\times\mathcal{U}_{k+1}$ as

$$\mathcal{G}=\big\{(u_0,u_{k+1}) : \mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\}\geq 1-\epsilon\big\},$$

Note that $\mathcal{G}$ is the set of all good (u_0,u_{k+1}) pairs as defined above. We have

$$\begin{aligned}\mathbf{P}\left\{(U_0,U_{[k]},U_{k+1})\in\mathcal{F}\right\} &= \sum_{u_0,u_{k+1}}\sum_{u_{[k]}\in\mathcal{F}(u_0,u_{k+1})} p(u_0,u_{k+1})p(u_{[k]}|u_0,u_{k+1}) \\ &= \sum_{u_0,u_{k+1}} p(u_0,u_{k+1})\,\mathbf{P}\{\mathcal{F}(u_0,u_{k+1})\} \\ &\leq (1-\epsilon)\,\mathbf{P}\{(U_0,U_{k+1})\notin\mathcal{G}\}+\mathbf{P}\{(U_0,U_{k+1})\in\mathcal{G}\} \\ &= 1-\epsilon\,\mathbf{P}\{(U_0,U_{k+1})\notin\mathcal{G}\}.\end{aligned}$$

Thus

$$\mathbf{P}\{(U_0,U_{k+1})\notin\mathcal{G}\}\leq\frac{1}{\epsilon}\,\mathbf{P}\left\{(U_0,U_{[k]},U_{k+1})\notin\mathcal{F}\right\}. \tag{4}$$

Our aim is to find an upper bound for $\mathbf{P}\{Z=0\}$. To do this, we write

$$\begin{aligned}\mathbf{P}\{Z=0\} &= \sum_{u_0,u_{k+1}} p(u_0,u_{k+1})\,\mathbf{P}\{Z=0|u_0,u_{k+1}\} \\ &\leq \frac{1}{\epsilon}\,\mathbf{P}\left\{(U_0,U_{[k]},U_{k+1})\notin\mathcal{F}\right\}+\sum_{(u_0,u_{k+1})\in\mathcal{G}} p(u_0,u_{k+1})\,\mathbf{P}\{Z=0|u_0,u_{k+1}\},\end{aligned} \tag{5}$$

where the inequality follows from Equation (4). Therefore, to find an upper bound on $\mathbf{P}\{Z=0\}$, it suffices to find an upper bound on $\mathbf{P}\{Z=0|U_0=u_0,,U_{k+1}=u_{k+1}\}$ for all $(u_0,u_{k+1})\in\mathcal{G}$. Fix $(u_0,u_{k+1})\in\mathcal{G}$. We use Chebyshev's inequality to find an upper bound on $\mathbf{P}\{Z=0|U_0=u_0,U_{k+1}=u_{k+1}\}$. Thus we need to calculate $\mathbb{E}[Z|U_0 = u_0, U_{k+1} = u_{k+1}]$ and $\mathbb{E}[Z^2|U_0 = u_0, U_{k+1} = u_{k+1}]$. For a given $\mathbf{m}$, from the definition of γ (Section 3) it follows

$$\begin{aligned}\mathbb{E}[Z_{\mathbf{m}}|u_0, u_{k+1}] &= \mathbf{P}\Big\{\big(U_1(m_1), \ldots, U_k(m_k)\big) \in \mathcal{F}(u_0, u_{k+1})\big|u_0, u_{k+1}\Big\}\\ &= \sum_{\mathcal{F}(u_0,u_{k+1})} p(u_1|u_0)\ldots p(u_k|u_0)\\ &\geq \sum_{\mathcal{F}(u_0,u_{k+1})} e^{-\gamma} p(u_{[k]}|u_0, u_{k+1})\\ &= e^{-\gamma}\,\mathbf{P}\{\mathcal{F}(u_0, u_{k+1})\} \geq (1-\epsilon)e^{-\gamma}.\end{aligned}$$

where the last inequality follows from the fact that $(u_0, u_{k+1}) \in \mathcal{G}$. Thus, by linearity of expectation,

$$\mathbb{E}[Z|U_0 = u_0, U_{k+1} = u_{k+1}] \geq |\mathcal{M}|e^{-\gamma}(1-\epsilon). \tag{6}$$

Next, we find an upper bound on $\mathbb{E}[Z^2|U_0 = u_0, U_{k+1} = u_{k+1}]$. We have

$$Z^2 = \sum_{\mathbf{m}} Z_{\mathbf{m}}^2 + \sum_{\mathbf{m}\neq\mathbf{m}'} Z_{\mathbf{m}}Z_{\mathbf{m}'} = Z + \sum_{\mathbf{m}\neq\mathbf{m}'} Z_{\mathbf{m}}Z_{\mathbf{m}'},$$

since $Z_{\mathbf{m}}^2 = Z_{\mathbf{m}}$ and $Z = \sum_{\mathbf{m}} Z_{\mathbf{m}}$. Thus

$$\mathbb{E}[Z^2|u_0, u_{k+1}] = \mathbb{E}[Z|u_0, u_{k+1}] + \mathbb{E}\Big[\sum_{\mathbf{m}\neq\mathbf{m}'} Z_{\mathbf{m}}Z_{\mathbf{m}'}\big|u_0, u_{k+1}\Big]$$

For any pair of distinct $\mathbf{m}$ and $\mathbf{m}'$ with nonempty $S = S_{\mathbf{m},\mathbf{m}'}$, we have

$$\begin{aligned}&\mathbb{E}\big[Z_{\mathbf{m}}Z_{\mathbf{m}'}|u_0, u_{k+1}\big]\\ &\quad= \sum_{\mathcal{F}_S(u_0,u_{k+1})} \prod_{i\in S} p(u_i|u_0)\Big(\sum_{u_{S^c}\in\mathcal{F}(u_0,u_S,u_{k+1})} \prod_{j\in S^c} p(u_j|u_0)\Big)^2\\ &\quad\leq e^{-\alpha_S - 2\beta_{S^c}} \sum_{\mathcal{F}_S(u_0,u_{k+1})} p(u_S|u_0, u_{k+1})\Big(\sum_{u_{S^c}\in\mathcal{F}(u_0,u_S,u_{k+1})} p(u_{S^c}|u_0, u_S, u_{k+1})\Big)^2\\ &\quad\leq e^{-\alpha_S - 2\beta_{S^c}},\end{aligned}$$

where $\mathcal{F}_S(u_0, u_{k+1})$ is the set of all u_S that satisfy $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$. On the other hand, if $S = S_{\mathbf{m},\mathbf{m}'}$ is empty, then $Z_{\mathbf{m}}$ and $Z'_{\mathbf{m}}$ are independent given $(U_0, U_{k+1}) = (u_0, u_{k+1})$, and

$$\mathbb{E}\big[Z_{\mathbf{m}}Z_{\mathbf{m}'}|u_0, u_{k+1}\big] = \big(\mathbb{E}[Z_{\mathbf{m}}|u_0, u_{k+1}]\big)^2.$$

Thus (assume $|\mathcal{M}_\emptyset| = 1$)

$$\begin{aligned}\mathbb{E}[Z^2|u_0, u_{k+1}] &= \mathbb{E}[Z|u_0, u_{k+1}] + \sum_{S\subset[k]} |\mathcal{M}_S| \prod_{j\in S^c} \big(|\mathcal{M}_j|^2 - |\mathcal{M}_j|\big)\mathbb{E}[Z_{\mathbf{m}}Z_{\mathbf{m}'}|u_0, u_{k+1}]\\ &\leq \mathbb{E}[Z|u_0, u_{k+1}] + \big(\mathbb{E}[Z|u_0, u_{k+1}]\big)^2 + \sum_{\emptyset\subset S\subset[k]} |\mathcal{M}_S||\mathcal{M}_{S^c}|^2 e^{-\alpha_S - 2\beta_{S^c}},\end{aligned} \tag{7}$$

where the notation $\emptyset \subset S \subset [k]$ means that S is a nonempty proper subset of $[k]$. We have

$$\begin{aligned}\mathbf{P}\left\{Z=0|u_0,u_{k+1}\right\} &\le \mathbf{P}\left\{\big|Z-\mathbb{E}[Z|u_0,u_{k+1}]\big| \ge \mathbb{E}[Z|u_0,u_{k+1}]\Big|u_0,u_{k+1}\right\}\\ &\overset{(a)}{\le} \frac{\mathrm{Var}(Z|u_0,u_{k+1})}{\big(\mathbb{E}[Z|u_0,u_{k+1}]\big)^2} = \frac{\mathbb{E}[Z^2|u_0,u_{k+1}]}{\big(\mathbb{E}[Z|u_0,u_{k+1}]\big)^2} - 1\\ &\overset{(b)}{\le} \frac{1}{1-\epsilon}|\mathcal{M}|^{-1}e^{\gamma} + \frac{1}{(1-\epsilon)^2}\sum_{\emptyset\subset S\subset[k]} |\mathcal{M}_S|^{-1}e^{-\alpha_S-2\beta_{S^c}+2\gamma},\end{aligned}$$

where (a) follows from Chebyshev's inequality and (b) follows from Equations (6) and (7). Now using Equation (5), we get

$$\mathbf{P}\{Z=0\} \le \frac{1}{\epsilon}\,\mathbf{P}\{\mathcal{F}^c\} + \frac{1}{1-\epsilon}|\mathcal{M}|^{-1}e^{\gamma} + \frac{1}{(1-\epsilon)^2}\sum_{\emptyset\subset S\subset[k]} |\mathcal{M}_S|^{-1}e^{-\alpha_S-2\beta_{S^c}+2\gamma}. \tag{8}$$

5. The Asymptotic Result

In this section, using our lower and upper bounds, we prove Theorem 1. We first prove the direct part using our upper bound from Section 4. Set $\mathcal{F} = A_\delta^{(n)}$ and for every $j \in [k]$, choose an integer $M_j \ge e^{nR_j}$. Choose a sequence $\{\epsilon_n\}_n$ such that

$$\lim_{n\to\infty} \frac{1}{\epsilon_n}\,\mathbf{P}\left\{(A_\delta^{(n)})^c\right\} = 0.$$

This is simple to do, since $\mathbf{P}\left\{(A_\delta^{(n)})^c\right\}$ decays exponentially in n (see Appendix B). Fix a nonempty $S \subseteq [k]$. Notice that if $\big(U_0^n, (U_j^n)_{j\in S}, U_{k+1}^n\big) \in \mathcal{F}_S$, then

$$\Big|\log \frac{p(u_S^n|u_0^n,u_{k+1}^n)}{\prod_{j\in S} p(u_j^n|u_0^n)} - n\Big(\sum_{j\in S} H(U_j|U_0) - H(U_S|U_0,U_{k+1})\Big)\Big| \le 2n(|S|+1)\delta.$$

Thus we may choose

$$\alpha_S = n\Big(\sum_{j\in S} H(U_j|U_0) - H(U_S|U_0,U_{k+1}) - 2(|S|+1)\delta\Big)$$

and

$$\gamma = n\Big(\sum_{j=1}^{k} H(U_j|U_0) - H(U_{[k]}|U_0,U_{k+1}) + 2(k+1)\delta\Big).$$

Similarly, for every nonempty $S \subseteq [k]$, we choose β_S as

$$\beta_S = n\Big(\sum_{j\in S} H(U_j|U_0) - H(U_S|U_0,U_{S^c},U_{k+1}) - 2(|S|+1)\delta)\Big),$$

since for every $\big(U_0^n, (U_j^n)_{j\in S}, (U_j^n)_{j\in S^c}\big) \in \mathcal{F}$,

$$\Big|\log \frac{p(u_S^n|u_0^n,u_{S^c}^n,u_{k+1}^n)}{\prod_{j\in S} p(u_j^n|u_0^n)} - n\Big(\sum_{j\in S} H(U_j|U_0) - H(U_S|U_0,U_{S^c},U_{k+1})\Big)\Big| \le 2n(|S|+1)\delta.$$

From our upper bound, Equation (8), it now follows that if for all nonempty $S \subset [k]$,

$$\begin{aligned}\sum_{j\in S} R_j &> \frac{1}{n}(2\gamma - \alpha_S - 2\beta_{S^c}) \\ &= 2\sum_{j=1}^{k} H(U_j|U_0) - 2H(U_{[k]}|U_0, U_{k+1}) - \sum_{j\in S} H(U_j|U_0) + H(U_S|U_0, U_{k+1}) \\ &\quad - 2\sum_{j\in S^c} H(U_j|U_0) + 2H(U_{S^c}|U_0, U_S, U_{k+1}) + (8k - 2|S| + 10)\delta \\ &= \sum_{j\in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) + (8k - 2|S| + 10)\delta,\end{aligned}$$

and for $S = [k]$,

$$\sum_{j=1}^{k} R_j > \frac{1}{n}\gamma = \sum_{j=1}^{k} H(U_j|U_0) - H(U_{[k]}|U_0, U_{k+1}) - 2(k+1)\delta,$$

then

$$\lim_{n\to\infty} \mathbf{P}\left\{\exists \mathbf{m} : \left(U_0^n, U_1^n(m_1), \ldots, U_k^n(m_k), U_{k+1}^n\right) \in A_\delta^{(n)}\right\} = 1. \tag{9}$$

Next we prove the converse. Suppose for each $j \in [k]$, $M_j \le e^{nR_j}$ and Equation (9) holds. Then from our lower bound, Equation (3), it follows

$$\sum_{j\in S} R_j \ge \frac{1}{n}\alpha_S = \sum_{j\in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) - 2(|S|+1)\delta,$$

for all nonempty $S \subseteq [k]$.

Appendix A. Cauchy-Schwarz Inequality

Let Z be any random variable that is nonnegative with probability one and has positive first and second moments. Then

$$Z = Z\mathbf{1}\{Z > 0\}$$

almost surely. Thus

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}\big[Z\mathbf{1}\{Z > 0\}\big] \\ &\le \sqrt{\mathbb{E}[Z^2] \times \mathbf{P}\{Z > 0\}},\end{aligned}$$

where the inequality follows from Cauchy-Schwarz. Hence

$$\mathbf{P}\{Z > 0\} \ge \frac{\big(\mathbb{E}[Z]\big)^2}{\mathbb{E}[Z^2]}$$

and

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

On the other hand, using Chebyshev's inequality we get

$$\begin{aligned}\mathbf{P}\{Z = 0\} &= \mathbf{P}\left\{|Z - \mathbb{E}[Z]| \ge \mathbb{E}[Z]\right\} \\ &\le \frac{\mathrm{Var}(Z)}{\big(\mathbb{E}[Z]\big)^2} = \frac{\mathbb{E}[Z^2]}{\big(\mathbb{E}[Z]\big)^2} - 1.\end{aligned}$$

Now note that the bound resulting from Cauchy-Schwarz is stronger, since for any $t > 0$,

$$1 - t \leq \frac{1}{t} - 1.$$

Appendix B. Large Deviations

The moment generating function of a random variable X is defined as

$$M(t) = \mathbb{E}[e^{tX}]$$

for all real t for which the expectation on the right hand side is finite. If M is defined on a neighborhood of 0, say $(-t_0, t_0)$ for some $t_0 > 0$, then it has a Taylor series expansion with a positive radius of convergence [1, pp. 278-280]. In particular,

$$\frac{d}{dt} M(t)\big|_{t=0} = \mathbb{E}[X].$$

We want to find an upper bound for $\mathbf{P}\{X \geq a\}$ for some real number a. Choose $t > 0$. Using Markov's inequality, we get

$$\begin{aligned}
\mathbf{P}\{X \geq a\} &= \mathbf{P}\{tX \geq ta\} \\
&= \mathbf{P}\{e^{tX} \geq e^{ta}\} \\
&\leq e^{-ta}\mathbb{E}[e^{tX}] \\
&= e^{\log M(t) - ta}
\end{aligned}$$

Since $t > 0$ was arbitrary, we get

$$\mathbf{P}\{X \geq a\} \leq e^{\inf_{t>0}(\log M(t) - ta)}.$$

Define the function f as

$$f(t) = \log M(t) - ta.$$

Then $f(0) = 0$ and $f'(0) = \mathbb{E}[X] - a$. Thus if $a > \mathbb{E}[X]$,

$$\inf_{t>0} \big(\log M(t) - ta\big) < 0. \tag{10}$$

If we apply the same inequality to the random variable

$$\frac{1}{n}\sum_{i=1}^{n} X_i,$$

where the X_i's are i.i.d. copies of X, we get

$$\mathbf{P}\left\{\sum_{i=1}^{n} X_i \geq na\right\} \leq e^{n \inf_{t>0}(\log M(t) - ta)}. \tag{11}$$

Now consider a random vector $(U_1, \ldots, U_k)$ with distribution $p(u_1, \ldots, u_k)$. For every nonempty $S \subseteq [k]$, let U_S denote the random vector $(U_j)_{j \in S}$. Let $(U_1^n, \ldots, U_k^n)$ be n i.i.d. copies of $(U_1, \ldots, U_k)$. By applying inequality (11) to the random variables $\{\log \frac{1}{p(U_{Si})}\}_{i=1}^n$ and setting $a = H(U_S) + \epsilon$ for some $\epsilon > 0$, we get

$$\mathbf{P}\left\{\sum_{i=1}^{n} \log \frac{1}{p(U_{Si})} \geq n(H(U_S) + \epsilon)\right\} \leq e^{-nI_S(\epsilon)}, \tag{12}$$

where $I_S(\epsilon)$ is given by

$$I_S(\epsilon) = \inf_{t>0}\left\{t\big(H(U_S) + \epsilon\big) - \log \mathbb{E}\big[p(U_S)^{-t}\big]\right\}$$

By the union bound we get

$$\begin{aligned}\mathbf{P}\left\{(U_1^n,\ldots,U_k^n)\notin A_\epsilon^{(n)}(U_1,\ldots,U_k)\right\} &\le 2\sum_{\emptyset\subsetneq S\subseteq[k]} e^{-nI_S(\epsilon)}\\ &\le 2(2^k-1)e^{-n\min_S I_S(\epsilon)}\\ &\le e^{-nI(\epsilon)},\end{aligned}$$

where

$$I(\epsilon)=\min_{S\subseteq[k]} I_S(\epsilon)+o\Big(\frac{1}{n}\Big).$$

Finally, note that by Equation (10), each $I_S(\epsilon)$ is positive, thus so is $I(\epsilon)$.

References

[1] Patrick Billingsley, *Probability and measure*, 3rd ed., SIAM, 1995.
[2] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, 2nd ed., Wiley, 2006.
[3] Abbas El Gamal and Young-Han Kim, *Network information theory*, 2nd ed., Cambridge University Press, 2012.
[4] Abbas El Gamal and Edward C. van der Meulen, *A proof of Marton's coding theorem for the discrete memoryless broadcast channel*, IEEE Trans. Inf. Theory **IT-27** (1981), no. 1, 120–122.
[5] Ralf Koetter, Michelle Effros, and Muriel Médard, *A theory of network equivalence—Part II: Multiterminal channels*, IEEE Trans. Inf. Theory **60** (2014), no. 7, 3709–3732.
[6] Jingbo Liu, Paul Cuff, and Sergio Verdú, *One-shot mutual covering lemma and Marton's inner bound with a common message*, Proc. IEEE Int. Symp. Information Theory, 2015.
[7] Katalin Marton, *A coding theorem for the discrete memoryless broadcast channel*, IEEE Trans. Inf. Theory **IT-25** (1979), no. 3, 306–311.
[8] Sergio Verdú, *Non-asymptotic covering lemmas*, IEEE Information Theory Workshop (ITW), 2015.

California Institute of Technology
E-mail address: parham@caltech.edu

California Institute of Technology
E-mail address: effros@caltech.edu

State University of New York at Buffalo
E-mail address: mikel@buffalo.edu