A Proof of the Generalized Markov Lemma with Countable Infinite Sources

Pablo Piantanida
Dept. of Telecommunications
SUPELEC
Gif-sur-Yvette, France
Email: pablo.piantanida@supelec.fr

Leonardo Rey Vega
School of Engineering
University of Buenos Aires and CONICET
Buenos Aires, Argentina
Email: lrey@fi.uba.ar

Alfred O. Hero III
Dept. of Electrical Eng. & CompSci
University of Michigan
Ann Arbor, MI, USA
Email: hero@umich.edu

***Abstract*—The *Generalized Markov Lemma* has been used in the proofs of several multiterminal source coding theorems for finite alphabets. An alternative approach to extend this result to countable infinite sources is proposed. We establish sufficient conditions to guarantee the joint typicality of reproduction sequences of random descriptions that have not been necessarily generated from the product of probability measures. Compared to existing proofs for finite alphabets, our technique is simpler and self-contained. It also offers bounds on the asymptotic tail probability of the typicality event providing a scaling law for a large number of source encoders.**

I. Introduction

Consider the distributed source coding problem where two memoryless sources are compressed separately and sent to a common decoder over rate-limited links. The decoder wishes to obtain a lossy estimate of the sources with a fidelity criterion. This fundamental problem arises in several applications, e.g. distributed storage, sensor networks and caching in wireless networks. Berger and Tung [1], [2] derived a general inner bound on its rate-distortion region. Each encoder quantizes its own observation using a noisy reproduction and these values are then communicated to the decoder using lossless Slepian-Wolf compression, which recovers them and thus constructs the sources estimates. The question of optimality of Berger-Tung's inner bound continues to be of great interest. Berger and Yeung [3] show that the bound is optimal if at least one of the sources must be reproduced losslessly. Wagner *et al.* [4] shown that it is optimal for two-Gaussian sources. Han and Kobayashi [5] have established direct coding theorems for several separate encoders. In spite of many works, the exact characterization of the rate-distortion region still remains open.

The proof of coding theorems for multiterminal source coding involve two technical results. The one referred to as the *Mutual Packing Lemma* [6] that provides necessary conditions to guarantee the success of joint decoding of all reproduction sequences at the decoder. The second is a central result that is referred to as *Generalized Markov Lemma* (GML), which guarantees the joint typicality between all involved sequences. In particular, since encoding is distributed the GML enforces two fundamental constraints: i) the test channels must be conditionally independent of each other, and ii) the reproductions must be formed from the typicality mappings (quantization), which yields "non-product" probability measures (pms).

The first *Markov lemma* for "product" pms (memoryless reproductions) was introduced by Wyner [7] and recently extended to countable infinite alphabets [8] via a different notion of typicality. Although most of the literature attributed the GML for "non-product" pms to [1], the original proof is given in [2]. This involves a combination of rather sophisticated algebraic and combinatorial arguments over finite alphabets. An alternative proof was also provided in [6] which strongly relies on a result by Uhlmann [9] used to bound the probability of the central error event. Based on source coding and probabilistic arguments of strong typical sets, Han and Kobayashi [5] have extended the GML to multiple sources over finite alphabets. By following this approach, Oohama in [10] derived a GML for scalar Gaussian sources.

In this paper, we develop another alternative approach that extends GML to countable infinite sources for multiple source encoders and side information at the decoder. For this extension we cannot rely on the standard properties of strong typical sets –in contrast to the proofs in [5], [6]– We approach this problem via a large deviation analysis of the multinomial characterization of empirical probability measures, which allows us to bound the probability of the relevant typicality event. Our technique is simpler and does not require external tools. It also offers bounds on the asymptotic tail of the probability of the typicality event, which provides us a scaling law for large number of source encoders.

Notations: Boldface letters x^n and upper-case letters X^n are used to denote vectors and random vectors of length n, respectively. Let $\mathcal{X}$ be a countable infinite alphabet and let $\sigma(\mathcal{X})$ be a σ-field, i.e. $2^{\mathcal{X}}$. Let Λ denote the set of all probability measures (pm) p on $\sigma(\mathcal{X})$. For each $n \in \mathbb{Z}_+$ let $\Pi_n \triangleq \{k/n : k = 1, \dots, n\}$. The *empirical* pm of a sample $x^n \in \mathcal{X}^n$ is given by

$$\hat{p}_{x^n}(\mathcal{B}) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[x_i \in \mathcal{B}],\ \mathcal{B} \in \sigma(\mathcal{X}) \ . \qquad (1)$$

We regard the n-Cartesian power of $(\mathcal{X}^n, \sigma(\mathcal{X}^n))$ where p_X^n is the n-th Cartesian power of p_X. Let p and q be pms where

The work of L. Rey Vega was partially supported by project UBACyT 2002010200250. The work of A. Hero was partially supported by a DIGITEO Chair from 2008 to 2013.

$p \ll q$ (*absolutely continuous*), we denote the relative entropy by $\mathcal{D}(p\|q)$ and the *total variational distance* by $\|p-q\|_{\text{TV}} \triangleq \sup_{\mathcal{A}\in\sigma(\mathcal{X})} |p(\mathcal{A})-q(\mathcal{A})|$. A RV X with *Bernoulli* probability distribution is denoted by $\text{Bern}\big(p_X(X=1)\big)$. Let X, Y and V be three RVs with pm p. If $p(x|y,v)=p(x|y)$ for each x,y,v, then they form a Markov chain, which is denoted by $X -\!\!\circ\!\!- Y -\!\!\circ\!\!- V$. The conditional product pm for i.i.d. RVs V^n given Y^n is denoted by $p^n_{V|Y}$ while a non-product pm is denoted by $p_{V^n|Y^n}$. The set of strong and conditional strong typical sequences are denoted by $T^n_\delta(V)$ and $T^n_\delta(V|y^n)$, respectively. Let $b_n = o(a_n)$ indicate $\limsup_{n\to\infty}(b_n/a_n)=0$ and $b_n=\mathcal{O}(a_n)$ indicates that $\limsup_{n\to\infty}|b_n/a_n|<\infty$.

II. Basic Definitions and Auxiliary Results

We begin with some basic definitions and auxiliary results which are required to prove the *Generalized Markov Lemma* (GML) stated in Section III.

Definition 1 (Joint typical sequences): A vector sequence $(x^n_1,\dots,x^n_N)\in\mathcal{X}^n_1\times\cdots\times\mathcal{X}^n_N$ is called *(strongly) joint δ-typical* w.r.t. $(X_1,\dots,X_N)$ (or simply *joint typical*) if

$$\big\|\hat{p}_{x^n_1\cdots x^n_N}-p_{X_1\dots X_N}\big\|_{\text{TV}}\le\delta\ , \tag{2}$$

and $\hat{p}_{x^n_1\cdots x^n_N}$ stands for empirical pms of tuples $(a_1,\dots,a_N)\in\mathcal{X}_1\times\cdots\times\mathcal{X}_N$ in $(x^n_1,\dots,x^n_N)$ and $\hat{p}_{x^n_1\cdots x^n_N}\ll p_{X_1\dots X_N}$. The set of all sequences is denoted by $T^n_\delta(X_1,\dots,X_N)$.

Definition 2 (Conditionally typical sequence): Let $x^n\in\mathcal{X}^n$. A sequence $y^n\in\mathcal{Y}^n$ is called *(strongly) δ-typical (w.r.t. Y) given x^n* if

$$\big\|\hat{p}_{x^ny^n}-\hat{p}_{x^n}p_{Y|X}\big\|_{\text{TV}}\le\delta\ , \tag{3}$$

and $\hat{p}_{x^ny^n}\ll\hat{p}_{x^n}p_{Y|X}$. $T^n_\delta(Y|x^n)$ denotes the set of all sequences.

Lemma 1 (Measure concentration): There exists a sequence $\eta_n=\mathcal{O}\big(c^{-n\delta^2_n}\big)\xrightarrow[n\to\infty]{}0$, for some constant $c>1$, s.t.

$$P^n_{X_1X_2\dots X_N}\big(T^n_{\delta_n}(X_1X_2\dots X_N)\big)\ge 1-\eta_n\ , \tag{4}$$

provided that $n\delta^2_n\to\infty$ as $n\to\infty$.

Proof: We are interested in uniform deviations of empirical averages for which we can simply use the bounded differences technique [11]. ■

Remark 1 (Stirling's approximation): The following approximation to the binomial coefficient holds for n large:

$$\frac{1}{n}\log\binom{n}{k}=H_2\left(\frac{k}{n}\right)+\mathcal{O}\left(\frac{\log n}{n}\right)\ , \tag{5}$$

for $k\le n=1,2,\dots$ where $H_2(x)$ denotes the binary entropy function.

Proposition 1 (Large deviation of Hypergeometric pm): Let $nK\in\mathbb{Z}_+$ be a random variable that follows a *Hypergeometric distribution* defined by

$$\Pr(nK=nk|m,N,N_1,n)=\binom{nN_1}{nk}\binom{n(N-N_1)}{n(m-k)}\Big/\binom{nN}{nm}$$

for $nk\in[1,nm]$, such that $n\max(0,m+N_1-N)\le nk\le n\min(N_1,m)$, where the parameters are defined as: nN is the population size; nN_1 is the number of success states in the population; nm is the number of draws; nk is the number of successes in the nm draws, and $n=1,2,3,\dots$ Then, for every $nk\in[0,nm]$,

$$\frac{1}{n}\log\Pr(nK=nk|m,N,N_1,n)=-I(k)+\mathcal{O}\left(\frac{\log n}{n}\right), \tag{6}$$

where the *rate function* $I(k)$ is defined as

$$\frac{I(k)}{N}\triangleq H_2\left(\frac{m}{N}\right)-\left[\frac{N_1}{N}H_2\left(\frac{k}{N_1}\right)+\left(1-\frac{N_1}{N}\right)H_2\left(\frac{m-k}{N-N_1}\right)\right] \tag{7}$$

Furthermore, the rate function $I(k)$ satisfies

$$I(k)\ge\frac{2}{N}\Big|k-\mathbb{E}[K]\Big|^2\quad\text{with}\quad\mathbb{E}[nK]=n\left(\frac{mN_1}{N}\right)\ , \tag{8}$$

Proof: We apply Remark 1 to each of the terms involved in $Pr(nK=nk|m,N,N_1,n)$ to obtain (6). It remains to show (8). Observe that $I(k)=\mathcal{D}(p_{XY}\|p_X\times p_Y)$, for two RVs $X,Y\in\{0,1\}$ with joint probability distribution $p_{Y|X}p_X$ where $p_X\equiv\text{Bern}\left(\frac{N_1}{N}\right)$ and $p_{Y|X}$ such that $p_{Y|X}(Y=1|X=1)\equiv\text{Bern}\left(\frac{k}{N_1}\right)$ and $p_{Y|X}(Y=1|X=0)\equiv\text{Bern}\left(\frac{m-k}{N-N_1}\right)$.

Using Pinsker's inequality and by means of standard manipulations, it is easy to check that

$$\mathcal{D}(p_{XY}\|p_X\times p_Y)\ge\frac{2}{N^2}\Big|k-\mathbb{E}[K]\Big|^2\ . \tag{9}$$

This concludes the proof of the proposition. ■

III. Generalized Markov Lemma (GML): Two-Encoders and Side Information

In this section, we present an alternative proof of the *Generalized Markov Lemma* (GML) [2], [6] for the case of two encoders with side information and countable infinite alphabets. We approach this problem via a large deviation analysis of the multinomial distribution which allows us to bound the probability events of the form $\{\hat{p}_{S^n}\in\Omega\subset\Lambda\}$.

Lemma 2 (GML with side information at the decoder): Let $q_{U^nV^nX^nY^nZ^n}=p_{U^n|X^n}p_{V^n|Y^n}p_{X^nY^nZ^n}$ be any (non-product) probability measures defined on $(\mathcal{U}^n,\sigma(\mathcal{U}^n))$, $(\mathcal{V}^n,\sigma(\mathcal{V}^n))$ and $(\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n,\sigma(\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n))$, respectively, and let (U,X,Z,Y,V) be random variables defined on countable infinite alphabets $\mathcal{U}\times\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}\times\mathcal{V}$ with joint probability p_{UXZYV} satisfying the Markov conditions:

$$(U,X,Z) -\!\!\circ\!\!- Y -\!\!\circ\!\!- V\quad\text{and}\quad(V,Y,Z) -\!\!\circ\!\!- X -\!\!\circ\!\!- U\ .$$

For every tuple $(x^n,z^n,y^n)\in T^n_\delta(X,Z,Y)$, it holds that

$$\Pr\Big\{(U^n,V^n)\notin\mathcal{T}^n_\epsilon(U,V|x^n,z^n,y^n)\,\Big|\,U^n\in T^n_\delta(U|x^n),\ V^n\in T^n_\delta(V|y^n),x^n,y^n\Big\}=\mathcal{O}\left(c^{-n}\right)\xrightarrow[n\to\infty]{}0\,, \tag{10}$$

for some constant $c>1$, provided that $p_{U^n|X^n}$ and $p_{V^n|Y^n}$ are uniformly distributed over the sets $T^n_\delta(U|x^n)$ and $T^n_\delta(V|y^n)$, respectively and where $\epsilon\xrightarrow[\delta\to 0]{}0$.

Remark 2: The result in Lemma 2 says that, under the Markov condition assumed, *marginal typicality* between the side information z^n, the source sequence x^n and its description u^n implies *joint typicality* with the source y^n and description v^n. Furthermore, this result holds for a large class of typical sequences as long as the class is stronger and hence implies definitions 1 and 2 (e.g. unified jointly typical sets [8]). Indeed, Lemma 2 can be used to systematically extend rate-distortion regions from discrete to countable infinite sources.

Remark 3: As a matter of fact, this result provides the GML needed for the inner bound derived in [12].

A. Proof of the Generalized Markov Lemma 2

In order to simplify the notation, for values $(u,x,z,y,v) \in \mathcal{U}\times\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}\times\mathcal{V}$ and given tuples of sequences $(x^n,z^n,y^n) \in T_\delta^n(X,Z,Y)$, we define –using (1)– the counting measures:

$$K_{V^n}(x,z,y,v) \triangleq \hat{p}_{x^n z^n y^n V^n}(x,z,y,v) \ , \tag{11}$$

$$\hat{p}_{V^n y^n}(v,y) = \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} K_{V^n}(v,x,z,y) \ , \tag{12}$$

$$K_{U^n V^n}(u,x,z,y,v) \triangleq \hat{p}_{U^n x^n z^n y^n V^n}(u,x,z,y,v) \ , \tag{13}$$

$$\hat{p}_{U^n x^n}(u,x) = \sum_{v\in\mathcal{V}}\sum_{z\in\mathcal{Z}}\sum_{y\in\mathcal{Y}} K_{U^n V^n}(u,x,z,y,v) \ , \tag{14}$$

where for convenience we dropped the sequences (x^n,z^n,y^n).

The main idea underlying the proof of (10) is to show that for a given $\delta>0$ and every tuple of sequences $(x^n,z^n,y^n) \in T_\delta^n(X,Z,Y)$ we have

$$\begin{aligned}
&\Pr\Big\{(U^n,V^n) \notin T_\delta^n(U,V|x^n,z^n,y^n)\,\Big|\,U^n \in T_\delta^n(U|x^n),\\
&\qquad V^n \in T_\delta^n(V|y^n),x^n,y^n\Big\}\\
&= \Pr\Big\{\,\big|K_{U^nV^n}(u,x,z,y,v) - \hat{p}_{x^nz^ny^n}(x,z,y)p_{U|X}(u|x)\\
&\qquad \times p_{V|Y}(v|y)\big| > \delta \quad \text{for some } (u,x,z,y,v) \in\\
&\qquad \mathcal{U}\times\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}\times\mathcal{V}\,\Big|\,U^n \in T_\delta^n(U|x^n),\\
&\qquad V^n \in T_\delta^n(V|y^n),x^n,y^n\Big\} = \mathcal{O}\left(c^{-n}\right) \xrightarrow[n\to\infty]{} 0
\end{aligned} \tag{15}$$

for a constant $c>1$. Indeed, (15) holds if for $(u,v,x,y,z) \in \text{supp}(\hat{p}_{x^nz^ny^n}p_{UV|XY})$ we can bound expressions:

$$\begin{aligned}
&\Pr\Big\{\,\big|K_{U^nV^n}(u,x,z,y,v) - \hat{p}_{x^nz^ny^n}(x,z,y)p_{U|X}(u|x)\times\\
&p_{V|Y}(v|y)\big| > \delta\,\Big|\,U^n \in T_\delta^n(U|x^n), V^n \in T_\delta^n(V|y^n),x^n,y^n\Big\}\\
&\leq \Pr\big(\{K_{U^nV^n} \in \mathcal{B}_n\} \cap \{K_{V^n} \in \mathcal{A}_n^c\}\big|\\
&\qquad U^n \in T_\delta^n(U|x^n), V^n \in T_\delta^n(V|y^n),x^n,y^n\big) \qquad (16)\\
&\quad + \Pr\big(\{K_{V^n} \in \mathcal{A}_n\}\big|V^n \in T_\delta^n(V|y^n),x^n,y^n\big) \ , \qquad (17)
\end{aligned}$$

where the above sets of rational measures are defined by

$$\mathcal{A}_n(v,x,y,z) \triangleq \Big\{\mu \in \mathit{\Pi}_n\Big|\ |\mu - \hat{p}_{x^nz^ny^n}(x,z,y) \times p_{V|Y}(v|y)\big| > \delta\Big\} \ , \tag{18}$$

$$\mathcal{B}_n(u,v,x,y,z) \triangleq \Big\{\eta \in \mathit{\Pi}_n\Big|\ |\eta - \hat{p}_{x^nz^ny^n}(x,z,y)p_{U|X}(u|x) \times p_{V|Y}(v|y)\big| > \delta\Big\} \ , \tag{19}$$

for each tuple $(u,v,x,y,z) \in \text{supp}(\hat{p}_{x^nz^ny^n}p_{UV|XY})$.

1) Bounding the probability event $\{K_{V^n} \in \mathcal{A}_n\}$ in expression (17)*:* We first show that there exists $c_1>1$ satisfying

$$\begin{aligned}
&\Pr\big(\{K_{V^n} \in \mathcal{A}_n\}\big|\,V^n \in T_\delta^n(V|y^n),x^n,y^n\big)\\
&= \sum_{\mu\in\mathcal{A}_n} \Pr\left(n\times K_{V^n} = n\mu\right|\,V^n \in T_\delta^n(V|y^n),x^n,y^n\big) \qquad (20)\\
&= \mathcal{O}\left(c_1^{-n}\right) \xrightarrow[n\to\infty]{} 0\,. \qquad (21)
\end{aligned}$$

For given sequences $(x^n,z^n,y^n) \in T_\delta^n(X,Z,Y)$ and $v^n \in T_\delta^n(V|y^n)$, we emphasize that $\hat{p}_{y^n} = p_Y + \epsilon_n$, $\hat{p}_{x^nz^ny^n} = p_{XZY} + \epsilon_n'$ and $\hat{p}_{v^ny^n} = p_{VY} + \epsilon_n''$, where $\epsilon_n \equiv \epsilon_n' \equiv \epsilon_n'' \equiv o(1/n) \to 0$ as $n\to\infty$ provided $\delta \equiv \delta_n \to 0$ as $n\to\infty$, and for convenience we have dropped (x,z,y,v).

The variable $n\times K_{V^n}(x,z,y,v)$ corresponds to the number of occurrences of the symbol "v", which satisfies

$$\sum_{v\in\mathcal{V}} K_{V^n}(x,z,y,v) = \hat{p}_{x^nz^ny^n}(x,z,y) \ . \tag{22}$$

Given a tuples of sequences $(x^n,z^n,y^n) \in T_\delta^n(X,Z,Y)$, the problem of finding the pm to evaluate (20) becomes clearly equivalent to a conventional counting problem in which we have a population of size $n\,\hat{p}_{y^n} = n\,(p_Y+\epsilon_n)$, of which $n\,\hat{p}_{x^nz^ny^n} = n\,(p_{XZY}+\epsilon_n')$ are "v" symbols and $n\,(\hat{p}_{y^n} - \hat{p}_{x^nz^ny^n})$ are not. Then, we pick without replacement $n\,\hat{p}_{v^ny^n} = n\,(p_{VY}+\epsilon_n'')$ samples of such population. For instance, the probability measure in expression (20) is the probability that there are $n\times K_{V^n} = n\mu$ symbols "v" amount the $n\,\hat{p}_{v^ny^n}$ samples drawn. This probability can be easily computed based on basic counting arguments [2]. Therefore, the probability becomes equal to:

$$\begin{aligned}
&\Pr\left(n\times K_{V^n} = n\mu\right|\,V^n \in T_\delta^n(V|y^n),x^n,y^n\big) =\\
&\binom{n\,\hat{p}_{x^nz^ny^n}}{n\mu}\binom{n\,(\hat{p}_{y^n} - \hat{p}_{x^nz^ny^n})}{n\,(\hat{p}_{v^ny^n} - \mu)}\Big/\binom{n\,\hat{p}_{y^n}}{n\,\hat{p}_{v^ny^n}} \ ,
\end{aligned} \tag{23}$$

the support of K_{V^n} is bounded by $0 \leq \text{supp}(K_{V^n}) \leq \min\left(\hat{p}_{v^ny^n}, \hat{p}_{x^nz^ny^n}\right)$, in order to form a complete system of events, and the conditional mean is given by

$$\mathbb{E}[K_{V^n}] = p_{XZY}p_{V|Y} + o(1/n) \ . \tag{24}$$

The probability measure in (23) is the *Hypergeometric distribution* defined in proposition 1. Hence, we can apply Proposition 1 by setting: $N = \hat{p}_{y^n}$, $N_1 = \hat{p}_{x^nz^ny^n}$ and $m = \hat{p}_{v^ny^n}$:

$$\begin{aligned}
&\frac{1}{n}\log\Pr\left(n\times K_{V^n} = n\mu\right|\,V^n \in T_\delta^n(V|y^n),x^n,y^n\big)\\
&\quad = -I_{K_V}(\mu) + \mathcal{O}\left(n^{-1}\log n\right) \xrightarrow[n\to\infty]{} -I_{K_V}(\mu) \ ,
\end{aligned} \tag{25}$$

for each $\mu \in [0,p_{VY}]$, where the *rate function* $I_{K_V}(\mu)$ is:

$$\begin{aligned}
\frac{I_{K_V}(\mu)}{p_Y} &\triangleq H_2\left(p_{V|Y}\right) - \Bigg[p_{XZ|Y}H_2\left(\frac{\mu/p_Y}{p_{XZ|Y}}\right)\\
&\quad + \left(1-p_{XZ|Y}\right)H_2\left(\frac{p_{V|Y} - \mu/p_Y}{1-p_{XZ|Y}}\right)\Bigg] + o(1) \ .
\end{aligned} \tag{26}$$

Furthermore, Proposition 1 also guarantees:

$$\begin{aligned} I_{K_V}(\mu) &\geq 2\Big|\mu - \mathbb{E}[K_{V^n}]\Big|^2 + \hat{\epsilon}_n \\ &= 2\left|\mu - p_{XZY}p_{V|Y} - o(1/n)\right|^2 + \hat{\epsilon}_n \ . \end{aligned} \tag{27}$$

By using the above bound (27) in expression (20), we obtain

$$\begin{aligned} &\Pr\left(\{K_{V^n} \in \mathcal{A}_n\}\middle|\, V^n \in T_\delta^n(V|y^n), x^n, y^n\right) \\ &\quad \leq \sum_{\mu \in \mathcal{A}_n} \exp\left[-n\left(I_{K_V}(\mu) + \mathcal{O}(n^{-1}\log n)\right)\right] \quad (28) \\ &\leq \exp\left[-n\left(\min_{\mu\in\mathcal{A}_n} 2\left|\mu - p_{XZY}p_{V|Y} - o(1/n)\right|^2 + o(1)\right)\right] \quad (29) \end{aligned}$$

where (29) follows by minimizing the exponent with respect to $\mu \in \mathcal{A}_n$ and noting that $\|\mathcal{A}_n\| \leq n$. On the other hand, since $(x^n, z^n, y^n) \in T_\delta^n(X,Z,Y)$ then $\hat{p}_{x^n z^n y^n} = p_{XYZ} + \epsilon_n$, which from definition (18) implies

$$\left|\mu - p_{XZY}p_{V|Y} - o(1/n)\right| > \delta \ , \tag{30}$$

for all $\mu \in \mathcal{A}_n$, and thus

$$p_{XZY}p_{V|Y} + o(1/n) \notin \mathcal{A}_n \ . \tag{31}$$

Provided condition (31) holds, by using (30), we now bound expression (29) as follows:

$$\min_{\mu\in\mathcal{A}_n} \left|\mu - p_{XZY}p_{V|Y} - o(1/n)\right|^2 \geq \delta^2 \ . \tag{32}$$

Finally, our claim in (21) simply follows by combining the upper bound in (29) together with the bound in (32), which yields the desired bound:

$$\Pr\left(\{K_{V^n} \in \mathcal{A}_n\}\middle| V^n \in T_\delta^n(V|y^n), x^n, y^n\right) \leq \mathcal{O}\left(c_1^{-n}\right) \ ,$$

provided by $n\delta^2 \xrightarrow[n\to\infty]{} \infty$, where $c_1 \triangleq \exp\left(2\delta^2 + o(1)\right)$.

2) Bounding the probability event $\{K_{U^nV^n} \in \mathcal{B}_n\}$ *in* (16): We now show that there exists a constant $c_2 > 1$ satisfying

$$\begin{aligned} &\Pr\left(\{K_{U^nV^n} \in \mathcal{B}_n\} \cap \{K_{V^n} \in \mathcal{A}_n^c\}\middle|\, U^n \in T_\delta^n(U|x^n),\right. \\ &\hspace{14em} \left. V^n \in T_\delta^n(V|y^n), x^n, y^n\right) \\ &\leq \Pr\left(\{K_{U^nV^n} \in \mathcal{B}_n\}\middle|\, \{K_{V^n} \in \mathcal{A}_n^c\}, U^n \in T_\delta^n(U|x^n),\right. \\ &\hspace{14em} \left. V^n \in T_\delta^n(V|y^n), x^n, y^n\right) \\ &= \sum_{\eta\in\mathcal{B}_n} \Pr\left(n \times K_{U^nV^n} = n\eta\middle|\, U^n \in T_\delta^n(U|x^n),\right. \\ &\hspace{8em} \left. V^n \in T_\delta^n(V|x^n, z^n, y^n), x^n, y^n\right) \quad (33) \\ &= \mathcal{O}\left(c_2^{-n}\right) \xrightarrow[n\to\infty]{} 0 \, . \quad (34) \end{aligned}$$

Here we will proceed similarly as before, for which we need to compute the probability measure involved in (33). For given sequences $(x^n, z^n, y^n) \in T_\delta^n(X,Z,Y)$, and $u^n \in T_\delta^n(U|x^n)$, and $v^n \in T_\delta^n(V|x^n, z^n, y^n)$, we denote: $\hat{p}_{x^n} = p_X + \epsilon_n$, $\hat{p}_{x^n z^n y^n v^n} = p_{XZY}p_{V|Y} + \epsilon_n''$ and $\hat{p}_{u^n x^n} = p_{UX} + \epsilon_n'$, where $\epsilon_n \equiv \epsilon_n' \equiv \epsilon_n'' \equiv o(1/n) \to 0$ as $n \to \infty$ provided we take $\delta \equiv \delta_n \to 0$ as $n \to \infty$, and for convenience we have dropped (u, x, z, y, v) from the definitions.

The random variable $K_{U^nV^n}(u, x, z, y, v)$ corresponds to the number of occurrences of symbols "(u, v)" satisfies:

$$\sum_{u\in\mathcal{U}} K_{U^nV^n}(u, x, z, y, v) = K_{V^n}(x, z, y, v) \ . \tag{35}$$

Given a tuples of sequences $(x^n, z^n, y^n) \in T_\delta^n(X,Z,Y)$, the involved pm can be found via a simple counting problem, as previously addressed. Here we have a population of size $n\,\hat{p}_{x^n}(x) = n\,(p_X + \epsilon_n)$, of which $n \times K_{V^n}(x, z, y, v) = n\left(p_{XZY}p_{V|Y} + \epsilon_n''\right)$ are "u" symbols and $n\,\left(\hat{p}_{x^n}(x) - K_{V^n}(x, z, y, v)\right)$ are not. Then, we pick without replacement $n\,\hat{p}_{u^n x^n} = n\,(p_{UX} + \epsilon_n')$ samples of such population. For instance, the probability measure in expression (33) is the probability that there are $n \times K_{U^nV^n} = n\eta$ symbols "(u, v)" amount the $n\,(p_{UX} + \epsilon_n')$ samples drawn. This probability can be easily computed as before and reads

$$\begin{aligned} &\Pr\left(n \times K_{U^nV^n} = n\eta\middle|\, U^n \in T_\delta^n(U|x^n),\right. \\ &\hspace{8em} \left. V^n \in T_\delta^n(V|x^n, z^n, y^n), x^n, y^n\right) \\ &= \binom{n\,\hat{p}_{x^n z^n y^n v^n}}{n\eta}\binom{n\,(\hat{p}_{x^n} - \hat{p}_{x^n z^n y^n v^n})}{n\,(\hat{p}_{u^n x^n} - \eta)}\Big/\binom{n\,\hat{p}_{x^n}}{n\,\hat{p}_{u^n x^n}} \end{aligned} \tag{36}$$

where the support of $K_{U^nV^n}$ is bounded by $0 \leq \text{supp}(K_{U^nV^n}) \leq \min\left(\hat{p}_{u^n x^n}, \hat{p}_{x^n z^n y^n v^n}\right)$, in order to form a complete system of events. The conditional mean is

$$\mathbb{E}[K_{U^nV^n}] = p_{XZY}p_{U|X}p_{V|Y} + o(1/n) \ . \tag{37}$$

The probability measure in (36) becomes again a *Hypergeometric distribution* and hence from Proposition 1 by setting $N = \hat{p}_{x^n}$, $N_1 = \hat{p}_{x^n z^n y^n v^n}$ and $m = \hat{p}_{u^n x^n}$. Then, a large deviation principle holds

$$\begin{aligned} &\frac{1}{n}\log\Pr\left(n \times K_{U^nV^n} = n\eta\middle|\, V^n \in T_\delta^n(U|x^n), V^n \in\right. \\ &\left. T_\delta^n(V|x^n, z^n, y^n), x^n, y^n\right) = -I_{K_{UV}}(\eta) + \mathcal{O}\left(n^{-1}\log n\right) \\ &\hspace{12em} \xrightarrow[n\to\infty]{} -I_{K_{UV}}(\eta) \ , \end{aligned} \tag{38}$$

for all $\eta \in [0, p_{UX}]$, where the *rate function* $I_{K_{UV}}(\eta)$ is

$$\begin{aligned} \frac{I_{K_{UV}}(\eta)}{p_X} &\triangleq H_2\left(p_{U|X}\right) - \left[p_{ZY|X}p_{V|Y}H_2\left(\frac{\eta/p_X}{p_{ZY|X}p_{V|Y}}\right)\right. \\ &\left. + \left(1 - p_{ZY|X}p_{V|Y}\right)H_2\left(\frac{p_{U|X} - \eta/p_X}{1 - p_{ZY|X}p_{V|Y}}\right)\right] + o(1) \ . \end{aligned} \tag{39}$$

Furthermore, the rate function $I_{K_{UV}}(\eta)$ satisfies

$$I_{K_{UV}}(\eta) \geq 2\left|\eta - p_{XZY}p_{U|X}p_{V|Y} - o(1/n)\right|^2 + \hat{\epsilon}_n \ . \tag{40}$$

By using the bound (40) in expression (33), we obtain

$$\begin{aligned} &\Pr\left(\{K_{U^nV^n} \in \mathcal{B}_n\}\middle|\, \{K_{V^n} \in \mathcal{A}_n^c\}, U^n \in T_\delta^n(U|x^n),\right. \\ &\hspace{12em} \left. V^n \in T_\delta^n(V|y^n), x^n, y^n\right) \leq \\ &\exp\left[-n\left(\min_{\eta\in\mathcal{B}_n} 2\left|\eta - p_{XZY}p_{U|X}p_{V|Y} - o(1/n)\right|^2 + o(1)\right)\right] \\ &\quad \leq \exp\left[-n\left(2\delta^2 + o(1)\right)\right] = \mathcal{O}\left(c_2^{-n}\right) \xrightarrow[n\to\infty]{} 0 \end{aligned} \tag{41}$$

where (41) follows by minimizing the exponent with respect to $\eta \in \mathcal{B}_n$ and noting that $\|\mathcal{B}_n\| \leq n$. On the other hand, we know that $(x^n, z^n, y^n) \in T^n_\delta(X, Z, Y)$ and the event $\{K_{V^n} \in \mathcal{A}_n^c\}$ implies $V^n \in T^n_\delta(V|x^n, z^n, y^n)$, which leads to $\hat{p}_{x^n z^n y^n V^n} = p_{XYZ} p_{V|Y} + \epsilon''_n$. From here it is not difficult to show –similarly as before– the desired uniform bound in (41). The result, finally follows by setting $c_2 \triangleq \exp\left(2\delta^2 + o(1)\right)$, provided $n\delta^2 \to \infty$ as $n \to \infty$. It is worth mentioning here that c_1 and c_2 may differ in the constants of $o(1)$, and thus by setting $c \triangleq \min(c_1, c_2)$ the proof of the lemma is finished.

IV. Generalized Markov Lemma (GML): Multiple Encoders and Side Information

In this section, we provide an alternative proof to the *Generalized Markov Lemma* (GML) [5] for the case of multiple source encoders with side information and countable infinite alphabets. We exploit again the multinomial characterization of the empirical probability measures to extend the proof of Lemma 2 to multiple encoders. Moreover, a scaling law for large number of source encoders is also investigated.

Lemma 3 (GML for multiple source encoders): Let

$$q_{V_1^n \cdots V_K^n | Y_1^n \cdots Y_K^n Z^n} = \prod_{k=1}^{K} p_{V_k^n | Y_k^n} \ , \tag{42}$$

be (non-product) pms defined on $(\mathcal{V}_k^n, \sigma(\mathcal{V}_k^n))$, for $k = \{1, \ldots, K\}$, and $(\mathcal{Y}_1^n \times \cdots \times \mathcal{Y}_K^n \times \mathcal{Z}^n, \sigma(\mathcal{Y}_1^n \times \cdots \times \mathcal{Y}_K^n \times \mathcal{Z}^n))$, respectively, and let $(V^K, Y^K, Z) = (V_1, \ldots, V_K, Y_1, \ldots, Y_K, Z)$ be random variables defined on countable infinite sources with joint pm $p_{V^K Y^K Z}$ satisfying the Markov conditions:

$$(V^{k-1}, Y^{k-1}, V_{k+1}^K, Y_{k+1}^K, Z) -\!\!\circ\!\!- Y_k -\!\!\circ\!\!- V_k \ , \tag{43}$$

for every $k = \{1, \ldots, K\}$. For every tuple of sequences $(y_1^n, \ldots, y_K^n, z^n) \in T^n_\delta(Y^K, Z)$, it holds that

$$\begin{aligned} &\Pr\Big\{(V_1^n, \ldots, V_K^n) \notin T^n_\delta(V_1^n, \ldots, V_K^n | y_1^n, \ldots, y_K^n, z^n) \,\big| \\ &V_k^n \in T^n_\delta(V_k | y_k^n), y_1^n, \ldots, y_K^n\Big\} = \mathcal{O}\left(c^{-n}\right) \xrightarrow[n\to\infty]{} 0 \,, \end{aligned} \tag{44}$$

for some constant $c > 1$, provided that $p_{V_k^n|Y_k^n}$ are uniform probability measures over the sets $T^n_\delta(V_k|y_k^n)$, respectively, for every $k = \{1, \ldots, K\}$.

Sketch of the Proof: For lack of space, we only provide a brief idea of the probability analysis of all typically events involved in this proof [13]. First, we bound the probability of the union of non-typical sets in (44) as the summation of probabilities of K-different events. This is done by following a similar approach to that used in the proof of Lemma 2 (see expression (17)). Then, by systematically applying Proposition (1) and using induction, it is not difficult to see that the probability of each of these events can be upper bounded by $\mathcal{O}\left(c_k^{-n}\right)$, for some constants $\{c_k > 1\}$ with $k = \{1, \ldots, K\}$.

Remark 4 (Scaling law for large number of encoders): Consider the same assumptions than Lemma 3 and let K_n be an increasing number of source encoders. The *optimal scaling exponent* is the largest positive number β such that

$$\begin{aligned} &\liminf_{n\to\infty} \frac{\log K_n}{n} > \beta \ , \quad \text{and} \\ &\Pr\Big\{(V_1^n, \ldots, V_K^n) \notin T^n_\delta(V_1^n, \ldots, V_K^n | y_1^n, \ldots, y_K^n, z^n) \,\big| \\ &\qquad V_k^n \in T^n_\delta(V_k | y_k^n), y_1^n, \ldots, y_K^n\Big\} \xrightarrow[n\to\infty]{} 0. \end{aligned} \tag{45}$$

Then, it holds that any exponent $0 \leq \beta < 1$ satisfies the previous conditions.

Sketch of the Proof: This remark simply follows by looking at the asymptotic behavior of expression (44) with both n and K_n. By carefully selecting $\delta_n \to 0$ as $n \to \infty$, it is easy to check that expression (44) behaves as $\mathcal{O}\left(K_n \beta^{-n}\right) \to 0$ as $n \to \infty$ for any $0 \leq \beta < 1$.

V. Summary and Discussion

The *Generalized Markov Lemma* (GML) was shown over countable infinite sources with multiple source encoders and side information. We approached this problem via a large deviation analysis of the multinomial characterization of the empirical probability measures, which offers bounds on the asymptotic tail of the probability of the typically event. Although the tools employed here apply only to countable alphabets, this extension is theoretically important since it can be shown to be the most natural way of extending –under some fairly assumptions– the GML to continuous alphabets [13].

References

[1] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. Series CISM Courses and Lectures Springer-Verlag, New York, 1978, vol. 229, pp. 171–231.

[2] S. Y. Tung, "Multiterminal source coding," Ph.D. Dissertation, Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[3] T. Berger and R. Yeung, "Multiterminal source encoding with one distortion criterion," *Information Theory, IEEE Trans. on*, vol. 35, no. 2, pp. 228–236, 1989.

[4] S. Tavildar, P. Viswanath, and A. Wagner, "The gaussian many-help-one distributed source coding problem," *Information Theory, IEEE Trans. on*, vol. 56, no. 1, pp. 564–581, 2010.

[5] T. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *Information Theory, IEEE Trans. on*, vol. 26, no. 3, pp. 277–288, 1980.

[6] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[7] A. Wyner, "On source coding with side information at the decoder," *Information Theory, IEEE Trans. on*, vol. 21, no. 3, pp. 294–300, 1975.

[8] S.-W. Ho, "Markov lemma for countable alphabets," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, 2010, pp. 1448–1452.

[9] W. Uhlmann, "Vergleich der hypergeometrischen mit der binomial-verteilung," *Metrika*, vol. 10, no. 1, pp. 145–158, 1966.

[10] Y. Oohama, "The rate-distortion function for the quadratic gaussian ceo problem," *Information Theory, IEEE Trans. on*, vol. 44, no. 3, pp. 1057–1070, 1998.

[11] C. McDiarmid, "On the method of bounded differences," *Surveys in Combinatorics, 1989, J. Siemons ed., London Mathematical Society Lecture Note Series 141, Cambridge University Press*, no. 4, pp. 148–188, 1989.

[12] M. Gastpar, "The wyner-ziv problem with multiple sources," *Information Theory, IEEE Trans. on*, vol. 50, no. 11, pp. 2762–2768, 2004.

[13] P. Piantanida, L. Rey Vega, and A. Hero III, "A proof to the generalized markov lemma for continuous alphabets," *Information Theory, IEEE Trans. on*, 2014, (to be submitted).