Error Exponents in a Source Coding Problem of Korner *

V. Anantharam
School of Electrical Engineering
Cornell University
Ithaca, NY 14853, USA

Abstract

Consider a discrete memoryless stationary information source with finite alphabet. Suppose that the letters emitted by the source are not all distinguishable. This might be the case, for example, if the symbols were handwritten letters. Distinguishability between letters is a symmetric relation and thus determines a graph on the source alphabet. Similarly, one talks of distinguishability between fixed length strings from the alphabet. Korner considered the problem of block coding for such a source with the requirement that only indistinguishable strings can be assigned the same codeword. He found a fundamental quantity, the graph entropy, which gives the minimum rate at which coding can be done with vanishingly small error probability. In this note we compute the error exponents for this source coding problem.

* Research supported by NSF PYI award NCR 8857731, an IBM Faculty Development Award and by BellCore Inc.

Error Exponents in a Source Coding Problem of Korner *

1. Introduction

Consider a discrete memoryless stationary information source (X, P) with finite alphabet X and marginal distribution P. Suppose that the letters emitted by the source are not all distinguishable. This might be the case, for example, if the symbols were handwritten letters, with slight differences in handwriting making no difference in the meaning of the symbol. Distinguishability between letters determines a symmetric relation on the source alphabet X. If we draw an edge between each pair of distinguishable letters, we get a graph $G = (X, E)$ with vertex set X and edge set E. The triple $[X, E, P]$ is called a *probabilistic graph*. Now, two sequences $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$ and $\mathbf{y} = (y_1, \ldots, y_n) \in X^n$ are distinguishable iff they are distinguishable in at least one co-ordinate. The *co-normal product* G^n is the graph on X^n which has an edge between $\mathbf{x}$ and $\mathbf{y}$ iff there is an edge between x_i and y_i for some $1 \leq i \leq n$. Clearly, the co-normal product describes the distinguishability relation on X^n.

We are interested in a special kind of source coding problem for the information source (X, P) with distinguishability structure $G = (X, E)$. We would like to code n-sequences from the source so that the same codeword is assigned to a pair of n-sequences iff they are indistinguishable. If the code uses M_n codewords, it is said to have *rate* $\frac{\log M_n}{n}$. Here and in the rest of the paper logarithms are taken to base 2. The *error probability* of the code is the probability of the the set of source strings which are not coded. (Alternately they may be coded by means of a special error symbol δ). A source coding scheme is a sequence of source codes for each string length n. It is said to have *asymptotic rate* R if

$$\lim_{n \to \infty} \frac{\log M_n}{n} = R \ .$$

A source coding scheme may thus be described by functions $(e_n)_{n=1}^{\infty}$ where $e_n \ : \ X^n \to [M_n] \cup \{\delta\}$, with $[M_n]$ denoting $\{1, \ldots, M_n\}$. The restriction is that the strings mapping into any $m \in [M_n]$ form an *independent set* in G^n, i.e. no pair of such strings can share an edge. The error probability of the code on blocks of length n is then

$$\mathrm{Err}(e_n) = P^n(\mathbf{x} \in X^n \ : \ e_n(\mathbf{x}) = \delta) \ .$$

The source coding problem described above was first considered by Korner, [5]. The main result of Korner is the existence and characterization of a fundamental quantity $H(G, P)$, the *graph entropy* of P relative to the graph structure G, which plays a role analogous to the usual entropy in conventional source coding, where all the symbols from the alphabet are assumed to be distinguishable from one another. For any $0 < \lambda < 1$, let

* Research supported by NSF PYI award NCR 8857731, an IBM Faculty Development Award and by BellCore Inc.

$M(n,\lambda)$ denote the minimum number of codewords needed to encode n-sequences from the source with error probability less than λ. Then

$$\lim_{n\to\infty} \frac{1}{n} \log M(n,\lambda) = H(G,P) \qquad (1.1) \ .$$

Equation (1.1) says that the indicated limit exists and is independent of λ. In particular, one can make the following statements :

(A) Consider any source coding scheme $(e_n)_{n=1}^{\infty}$ which uses M_n codewords to code strings of length n and having error probability $\mathrm{Err}(e_n)$. If

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n < H(G,P)$$

then

$$\lim_{n\to\infty} \mathrm{Err}(e_n) = 1 \ .$$

(B) There is a source coding scheme having

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = H(G,P)$$

with

$$\lim_{n\to\infty} \mathrm{Err}(e_n) = 0 \ .$$

Thus, in order to do source coding with vanishingly small error probability, we must have rate at least as large as the graph entropy.

Graph entropy has turned out to be of importance in a number of problems in communication theory, graph theory, and computer science, [2], [4], [8], [9]. For more on the source coding problems that originally motivated its introduction, see [5], [6], [7].

The purpose of this note is to discuss the form of the error exponent in the above source coding problem of Korner. Namely, we are interested in studying how rapidly the error probability can be made to decrease, given an upper bound to the asymptotic rate at which coding has to be done. For a probability distribution Q on X, let $K(Q,P)$ denote the Kullback-Leibler information discrimination of Q relative to P, given by

$$K(Q,P) = \sum_{x\in X} Q(x) \log \frac{Q(x)}{P(x)} \ .$$

Then the main result of this note is the following :

Theorem 1.1 : Let $0 \le R \le \sup_P H(G,P)$. Then there exists a source coding scheme $(e_n)_{n=1}^{\infty}$ using M_n codewords to code source strings of length n such that

$$\lim_{n\to\infty} \frac{\log M_n}{n} = R$$

with

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Err}(e_n) \leq - \inf_{Q \,:\, H(G,Q) \geq R} K(Q,P) \tag{1.2}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log(1 - \mathrm{Err}(e_n)) \geq - \inf_{Q \,:\, H(G,Q) < R} K(Q,P) \ . \tag{1.3}$$

Further, this result is optimal in the sense that for any source coding scheme having

$$\limsup_{n\to\infty} \frac{\log M_n}{n} \leq R$$

we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Err}(e_n) \geq - \inf_{Q \,:\, H(G,Q) > R} K(Q,P) \tag{1.4}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log(1 - \mathrm{Err}(e_n)) \leq - \inf_{Q \,:\, H(G,Q) \leq R} K(Q,P) \ . \tag{1.5}$$

Finally, for $0 < R < \sup_P H(G,P)$ the right hand sides of (1.2) and (1.4) are equal, as are the right hand sides of (1.3) and (1.5).

Remarks : (1). The proof of (1.4) and (1.5) is by counting. The existence of codes satisfying (1.2) and (1.3) is proved by random coding arguments - it is an interesting question whether one can find universal codes (i.e. codebooks that are independent of P, for a fixed graph structure G) with the same error exponent.

(2). The original motivation of the investigation was to see if it would unearth a fundamental quantity standing in the same relation to Kullback-Leibler information discrimination that graph entropy bears to usual entropy. This continues to be an interesting question.

2. Preliminaries

Before proceeding with the proof of Theorem 1.1, we first need the information theoretic characterization of graph entropy due to Korner, [5]. An *independent set* in $G = (X,E)$ is a subset $A \subseteq X$ such that no pair of vertices from A share an edge. A *kernel* is a maximal independent subset. Let $\mathcal{A}$ denote the collection of kernels of X. Let $X \circ \mathcal{A}$ denote the subset of $X \times \mathcal{A}$ consisting of the pairs (x,a) such that $x \in a$. Let $\mathcal{M}(X)$ denote the collection of probability distributions on X. Given a probability distribution $Q \in \mathcal{M}(X)$ a *Q-admissible* probability distribution $\underline{Q}$ on $X \circ \mathcal{A}$ is one for which $\sum_{a \ni x} \underline{Q}(x,a) = Q(x)$. We let $\mathcal{M}_Q(X \circ \mathcal{A})$ denote the collection of Q-admissible probability distributions on $X \circ \mathcal{A}$ and consistently use $\underline{Q}$ to denote a generic element of $\mathcal{M}_Q(X \circ \mathcal{A})$. We also consistently use $\tilde{Q}$ to denote the probability distribution on $\mathcal{A}$ given by $\tilde{Q}(a) = \sum_{x \in a} \underline{Q}(x,a)$, where $\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})$.

Following Korner, [5], we define random variables $\underline{X} \ : \ X \circ \mathcal{A} \to X$ and $\underline{Y} \ : \ X \circ \mathcal{A} \to \mathcal{A}$ giving the first and second co-ordinates respectively. The following information theoretic

characterization of graph entropy is proved in [5] :

$$H(G,Q) = \min_{\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})} I(\underline{X} \wedge \underline{Y})$$

where $I(\underline{X} \wedge \underline{Y})$ denotes the mutual information between $\underline{X}$ and $\underline{Y}$. We let $\mathcal{A}^n$ denote the set of n-sequences of elements from $\mathcal{A}$. We observe that $\mathcal{A}^n$ is precisely the collection of kernels of G^n, see e.g. the last paragraph of Section 2 of [5] for a proof.

We let $\mathcal{M}_n(X)$ denote the collection of probability distributions on X for which the probability of any source symbol is a rational number with denominator dividing n. Given an n-sequence $\mathbf{x} \in X^n$, its *empirical distribution* or *type* is the probability distribution $(\frac{N(x|\mathbf{x})}{n}, x \in X)$, where $N(x \mid \mathbf{x})$ denotes the number of occurence of the symbol x in the string $\mathbf{x}$. Note that this belongs to $\mathcal{M}_n(X)$. Thus $\mathcal{M}_n(X)$ is also called the collection of n-types. It is easy to see that

$$\| \mathcal{M}_n(X) \| \leq (n+1)^{\|X\|}$$

where $\|X\|$ denotes the cardinality of X.

Fix $K > 0$. Given $Q \in \mathcal{M}(X)$, an n-sequence $\mathbf{x} \in X^n$ is said to be Q-typical if, for every $x \in X$

$$\mid N(x \mid \mathbf{x}) - nQ(x) \mid \leq K\sqrt{nQ(x)} \ .$$

The following lemma is Lemma 3 of Korner, [5]. See also Csiszar and Korner, [1].

Lemma 2.1 : Let $T^n(Q) \subseteq X^n$ denote the set of Q-typical sequences. Then, for every $\mathbf{x} \in T^n(Q)$

$$2^{-nH(Q)-C\sqrt{n}} \leq Q^n(\mathbf{x}) \leq 2^{-nH(Q)+C\sqrt{n}}$$

and

$$2^{nH(Q)-C\sqrt{n}} \leq \|T^n(Q)\| \leq 2^{-nH(Q)+C\sqrt{n}} \ .$$

Here C depends only on K and $\|X\|$, and in particular is independent of n and Q. Throughout this note $K > 0$ will be fixed once and for all, as will P, the marginal distribution of source letters. Q will be used for a generic element of $\mathcal{M}(X)$ and C will denote a finite constant that is independent of n and Q, but may vary from line to line.

We also need to discuss the special situation of probabilistic graphs $[X, E, Q]$ where the relation $x \rho y \Leftrightarrow \{$ either $(x,y) \in E$ or $x = y\}$ determines an equivalence relation on X. Clearly then $G = (X, E)$ is the union of pairwise disjoint complete subgraphs. Let $H(Q \mid \rho)$ denote the conditional entropy given the equivalence class ρ, namely

$$H(Q \mid \rho) = \sum_{x \in X} Q(x) \log \frac{\sum_{y \ : \ x \rho y} Q(y)}{Q(x)}$$

Let $\mathcal{R}$ denote the collection of equivalence classes under ρ. Let us write Q_ρ for the probability distribution on $\mathcal{R}$ given by $Q_\rho(r) = \sum_{x \in r} Q(x)$. Then the following lemma is easy to prove :

Lemma 2.2 : The number of Q-typical n-sequences in a Q_ρ-typical n-sequence of equivalence classes is bounded above by $2^{nH(Q|\rho)+C\sqrt{n}}$ and below by $2^{nH(Q|\rho)-C\sqrt{n}}$.

Proof : Let $\mathbf{r} = (r_1, \ldots, r_n)$ be a Q_ρ-typical n-sequence. Namely,, for each $r \in \mathcal{R}$

$$\mid N(r \mid \mathbf{r}) - nQ_\rho(r) \mid \leq K\sqrt{nQ_\rho(r)} \ .$$

Then we have, for each $r \in \mathcal{R}$,

$$nQ_\rho(r) \leq \max(4K^2, 2N(r \mid \mathbf{r})) \tag{2.1}$$

because if $nQ_\rho(r) \geq 4K^2$ then $nQ_\rho(r) \geq 2K\sqrt{nQ_\rho(r)}$, and so

$$N(r \mid \mathbf{r}) \geq nQ_\rho(r) - K\sqrt{nQ_\rho(r)} \geq \frac{nQ_\rho(r)}{2} \ .$$

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a Q-typical n-sequence in $\mathbf{r}$, i.e. $x_i \in r_i$, $1 \leq i \leq n$. Then, for each $x \in X$

$$\mid N(x \mid \mathbf{x}) - nQ(x) \mid \leq K\sqrt{nQ(x)}$$

by Q-typicality of $\mathbf{x}$. We claim that for each $r \in \mathcal{R}$ the restriction of $\mathbf{x}$ to those co-ordinates having $r_i = r$ is $(\frac{Q(x)}{Q_\rho(r)}; x \in r)$-typical, with a different K. Indeed, for $x \in r$, we have

$$\begin{aligned}
\mid N(x \mid \mathbf{x}) - N(r \mid \mathbf{r})\frac{Q(x)}{Q_\rho(r)} \mid &\leq \mid N(x \mid \mathbf{x}) - nQ(x) \mid + \mid nQ(x) - N(r \mid \mathbf{r})\frac{Q(x)}{Q_\rho(r)} \mid \\
&\leq K\sqrt{nQ(x)} + \frac{Q(x)}{Q_\rho(r)}K\sqrt{nQ_\rho(r)} \\
&= K[\sqrt{\frac{Q(x)}{Q_\rho(r)}} + \frac{Q(x)}{Q_\rho(r)}]\sqrt{nQ_\rho(r)} \ .
\end{aligned}$$

Using (2.1) gives

$$\begin{aligned}
\mid N(x \mid \mathbf{x}) - N(r \mid \mathbf{r})\frac{Q(x)}{Q_\rho(r)} \mid &\leq \max(3K^2, \sqrt{2K})[\sqrt{\frac{Q(x)}{Q_\rho(r)}} + \frac{Q(x)}{Q_\rho(r)}]\sqrt{N(r \mid \mathbf{r})} \\
&\leq 2\max(3K^2, \sqrt{2K})\sqrt{\frac{N(r \mid \mathbf{r})Q(x)}{Q_\rho(r)}} \ .
\end{aligned}$$

It follows that $2^{n\sum_{r\in\mathcal{R}} \frac{N(r|\mathbf{r})}{n} H(Q|\rho=r)+C\sqrt{n}}$ and $2^{n\sum_{r\in\mathcal{R}} \frac{N(r|\mathbf{r})}{n} H(Q|\rho=r)-C\sqrt{n}}$ are respectively an upper bound and a lower bound for number of Q-typical n-sequences $\mathbf{x}$ in $\mathbf{r}$, where $H(Q \mid \rho = r)$ denotes $\sum_{x\in r} \frac{Q(x)}{Q_\rho(r)} \log \frac{Q_\rho(r)}{Q(x)}$. But the Q_ρ-typicality of $\mathbf{r}$ ensures that we can further bound this above by $2^{nH(Q|\rho)+C\sqrt{n}}$ and below by $2^{nH(Q|\rho)-C\sqrt{n}}$ where, as before, C depends only on K and $\|X\|$.

Finally, we need to discuss the continuity properties of the functions $Q \to H(G,Q)$ and $Q \to K(Q,P)$ on $\mathcal{M}(X)$. Give $\mathcal{M}(X)$ the topology of pointwise convergence, which is the same as its induced topology as the unit simplex in $\mathcal{R}_+^{\|X\|}$. Then $K(Q,P)$ is a bounded continuous function of $Q \in \mathcal{M}(X)$, which is also convex and nonnegative, and is zero if and only if $Q = P$. See Ellis, [3], for proofs of these statements; here we are assuming without loss of generality that $P(x) > 0$ for all $x \in X$. Regarding the graph entropy, we require two lemmas :

Lemma 2.3 : $H(G,Q)$ is a continuous function of Q.

Proof : Given $Q_n \in \mathcal{M}(X)$ with $Q_n \to Q$ as $n \to \infty$. let $\underline{Q}^*$ achieve the minimum in the information theoretic characterization of $H(G,Q)$ as $\min_{\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})} I(\underline{X} \wedge \underline{Y})$. Consider $\underline{Q}_n \in \mathcal{M}_{Q_n}(X \circ \mathcal{A})$ given by

$$\underline{Q}_n(x,a) = Q_n(x)\underline{Q}^*(a \mid x) = Q_n(x)\frac{\underline{Q}^*(x,a)}{Q(x)}$$

We have

$$H(G,Q_n) \leq \sum_{(x,a) \in X \circ \mathcal{A}} \underline{Q}_n(x,a) \log \frac{\underline{Q}_n(x,a)}{Q_n(x)\tilde{Q}_n(a)}$$

where $\tilde{Q}_n(a) = \sum_{x \in a} Q_n(x)\underline{Q}^*(a \mid x)$. Clearly, $\underline{Q}_n \to \underline{Q}^*$ as $n \to \infty$ in the topology of pointwise convergence on $\mathcal{M}(X \circ \mathcal{A})$ and also $\tilde{Q}_n \to \tilde{Q}$ in $\mathcal{M}(\mathcal{A})$. It follows that $H(G,Q)$ is upper semicontinuous as a function of $Q \in \mathcal{M}(X)$. For the converse, let $\underline{Q}_n^* \in \mathcal{M}_{Q_n}(X \circ \mathcal{A})$ achieve the minimum in the information theoretic characterization $\min_{\underline{Q}_n \in \mathcal{M}_{Q_n}(X \circ \mathcal{A})} I(\underline{X} \wedge \underline{Y})$ of $H(G,Q_n)$. By restricting to a subsequence if necessary, we can assume that $\underline{Q}_n$ converges to some $\underline{Q} \in \mathcal{M}(X \circ \mathcal{A})$ as $n \to \infty$. Clearly $\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})$ and we have

$$\begin{aligned} H(G,Q) &\leq \sum_{(x,a) \in X \circ \mathcal{A}} \underline{Q}(x,a) \log \frac{\underline{Q}(x,a)}{Q(x)\tilde{Q}(a)} \\ &= \lim_{n \to \infty} H(G,Q_n) \end{aligned}$$

which completes the proof of the lemma.

The next lemma is most easily proved by using a geometric characterization of graph entropy due to Csiszar, Korner, Lovasz, Marton and Simonyi, [2]. From the graph G we construct a convex subset $VP(G)$ of the orthant $\mathbf{R}_+^{\|X\|}$ called the *vertex packing polytope* of G. $VP(G)$ is the convex hull of the indicator vectors of the independent sets of G. Then we have

$$H(G,P) = \min_{\underline{a} \in VP(G)} (-\sum_{x \in X} P(x) \log a(x)) \ , \tag{2.2}$$

see [2], Lemma 3.1.

Lemma 2.4 : Fix any R, $0 \leq R < \sup_P H(G,P)$. Then $\{P \ : \ H(G,P) \geq R\}$ is the closure of $\{P \ : \ H(G,P) > R\}$.

Proof : Clearly it is enough to prove that every local maximum of $H(G,P)$ on $\mathcal{M}(X)$ is a global maximum. Since (2.2) characterizes $H(G,P)$ as the minmum of a family of linear functions on $\mathcal{M}(X)$, which is a convex subset of the orthant, this is elementary.

Lemma 2.3 and Lemma 2.4, together with the continuity of $K(Q,P)$ in Q verify the claim in Theorem 1.1 that for $0 < R < \sup_P H(G,P)$ the right hand sides of (1.2) and (1.4) are equal and the right hand sides of (1.3) and (1.5) are equal.

3. Error Exponents

We proceed to complete the proof of Theorem 1.1, proving first the statements (1.4) and (1.5) and then (1.2) and (1.3).

To prove (1.4) and (1.5), let $(e_n)_{n=1}^{\infty}$ be an arbitrary source coding scheme which uses M_n codewords to code n-sequences of source symbols, and with $\limsup_{n\to\infty} \frac{\log M_n}{n} \leq R$. Let $Q \in \mathcal{M}(X)$ and let $S \subset X^n$ be the set of n-sequences of letters from the alphabet which are received error free. Then we have

$$P^n(S \cap T^n(Q)) \leq M_n \max_{A \in \mathcal{A}^n} P^n(A \cap T^n(Q)) \tag{3.1}$$

because we can write S as the union of at most M_n subsets consisting of strings which are coded as the codeword m, $1 \leq m \leq M_n$, and each such set of strings is an independent set in G^n. Also

$$P^n(A \cap T^n(Q)) \leq \max_{\mathbf{x} \in T^n(Q)} P^n(\mathbf{x}) \max_{A \in \mathcal{A}^n} \|A \cap T^n(Q)\| \ . \tag{3.2}$$

From Lemma 2.1, we have

$$\begin{aligned} \max_{\mathbf{x} \in X^n} P^n(\mathbf{x}) \leq & \max_{\mathbf{x} \in T^n(Q)} \prod_{x \in X} P(x)^{N(x|\mathbf{x})} \\ & \leq \prod_{x \in X} P(x)^{nQ(x) - K\sqrt{nQ(x)}} \\ & = 2^{n \sum_{x \in X} Q(x) \log P(x) - K\sqrt{n} \sum_{x \in X} \log P(x)} \end{aligned} \tag{3.3}$$

Following the arguments on pp. 418 -419 of Korner, [5], it is easy to see that

$$\max_{A \in \mathcal{A}^n} \|A \cap T^n(Q)\| \leq 2^{n \max_{\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})} H(\underline{Q}|\alpha) + C\sqrt{n}} \tag{3.4}$$

where α is the equivalence relation on $X \circ \mathcal{A}$ given by

$$(x,a)\alpha(y,b) \Leftrightarrow a = b$$

and $H(\underline{Q} \mid \alpha)$ denotes the relative entropy of $\underline{Q}$ given the α equivalence class. A sketch of the reasoning underlying (3.4) is as follows : Let $A^0 = (a_1^0, \ldots, a_n^0)$ achieve $\max_{A \in \mathcal{A}^n} \|A \cap$

$T^n(Q)\|$. Pick $\mathbf{x}^0 \in T^n(Q)$. We construct $\underline{Q}^0 \in \mathcal{M}_Q(X \circ \mathcal{A})$ using the conditional distribution $\underline{Q}^0(a \mid x)$ given by the empirical distribution of kernels in $(a_1^0, \ldots, a_n^0)$ at those co-ordinates where $\mathbf{x}^0$ has symbol x. Then the n-string $(x_1^0, a_1^0), \ldots, (x_n^0, a_n^0)$ is $\underline{Q}^0$-typical and the n-string $(a_1^0, \ldots, a_n^0)$ is $\tilde{Q}^0$-typical. This shows that $\max_{A \in \mathcal{A}^n} \|A \cap T^n(Q)\|$ is no larger than the sum over $\underline{Q} \in \mathcal{M}_Q^n(X \circ \mathcal{A})$ of the number of $\underline{Q}$-typical n-strings whose $\mathcal{A}$ co-ordinates is $\tilde{Q}$-typical. We can count these as the number of n-strings consistent with a specified type of n-string of equivalence classes under the equivalence relation given by $(x,a)\alpha(y,b) \Leftrightarrow a = b$.

On the other hand, we have

$$\begin{aligned} P^n(T^n(Q) \geq & \min_{\mathbf{x} \in T^n(Q)} P^n(\mathbf{x}) \|T^n(Q)\| \\ \geq & 2^{nK(Q,P) - C\sqrt{n} + K\sqrt{n} \sum_{x \in X} \log P(x)} \end{aligned} \tag{3.5}$$

From (3.1) -(3.5), we get

$$\begin{aligned} \frac{P^n(S \cap T^n(Q))}{P^n(T^n(Q))} \leq & 2^{n[\frac{\log M_n}{n} + \sum_{x \in X} Q(x) \log P(x) + \max_{\underline{Q} \in \mathcal{M}_Q(X \circ \mathcal{A})} H(\underline{Q}|\alpha) - K(Q,P)] - 2K\sqrt{n} \sum_{x \in X} \log P(x) + C\sqrt{n}} \\ \leq & 2^{n[\frac{\log M_n}{n} - H(G,Q)] + C\sqrt{n}} \end{aligned} \tag{3.6}$$

Now suppose $H(G,Q) = R + \epsilon > R$. Then (3.6) yields

$$\lim_{n \to \infty} \frac{P^n(S \cap T^n(Q))}{P^n(T^n(Q))} = 0 \ . \tag{3.7}$$

Let $\mathcal{E} = X^n - S$ denote the set of strings that are received in error (i.e. coded as δ). Then (3.7) is equivalent to

$$\lim_{n \to \infty} \frac{P^n(\mathcal{E} \cap T^n(Q))}{P^n(T^n(Q))} = 1 \ . \tag{3.8}$$

But we have

$$\mathrm{Err}(e_n) \geq P^n(\mathcal{E} \cap T^n(Q))$$

So

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathrm{Err}(e_n) \geq \liminf_{n \to \infty} \frac{1}{n} \log P^n(T^n(Q)) = -K(Q,P) \ .$$

It follows that

$$\liminf \frac{1}{n} \log \mathrm{Err}(e_n) \geq - \inf_{Q \ : \ H(G,Q) > R} K(Q,P)$$

which establishes (1.4).

To prove (1.5), observe that

$$\begin{aligned} P^n(S) \leq & P^n(\cup_{H(G,Q) < R+\epsilon} T^n(Q)) + P^n(S \cap \cup_{H(G,Q) \geq R+\epsilon} T^n(Q)) \\ \leq & (N+1)^{\|X\|} [\sup_{H(G,Q) < R+\epsilon} P^n(T^n(Q)) + \sup_{H(G,Q) \geq R+\epsilon} P^n(S \cap T^n(Q))] \end{aligned}$$

Now

$$\frac{1}{n}\log \sup_{H(G,Q)<R+\epsilon} P^n(T^n(Q)) = - \inf_{H(G,Q)<R+\epsilon} K(Q,P)$$

and, from (3.6), it follows that

$$\frac{1}{n}\log[\sup_{H(G,Q)\geq R+\epsilon} P^n(S \cap T^n(Q))] = 0 \ .$$

Hence

$$\limsup \frac{1}{n}\log P^n(S) \leq - \inf_{H(G,Q)<R+\epsilon} K(Q,P)$$

Letting $\epsilon \to 0$, using Lemma 2.4, and the continuity of $K(Q,P)$ in Q, (1.5) follows.

To prove (1.2) and (1.3), we will use a random coding argument. For $R = 0$, (1.2) and (1.3) are vacuous. Fix $Q \in \mathcal{M}_n(X)$ with $H(G,Q) < R - \epsilon$. Let $\underline{Q}^* \in \mathcal{M}_Q(X \circ \mathcal{A})$ achieve the minimum in the information theoretic characterization $\min_{\underline{Q} \in \mathcal{M}_Q(X\circ\mathcal{A})} I(\underline{X} \wedge \underline{Y})$ of $H(G,Q)$. Set $\tilde{M}_n = \lfloor \frac{2^{nR}}{(N+1)^{\|X\|}} \rfloor$. We draw $\tilde{M}_n$ elements of $\mathcal{A}^n$ independently according to the product distribution with marginal $\tilde{Q}^{*,n}$ which for notational simplicity is denoted $Q^{\times}$. Let $A_1, \ldots, A_{\tilde{M}_n}$ denote these elements, and denote their joint distribution by Q^+. Fix $\mathbf{x} \in T^n(Q)$ and let $\mathcal{A}_{\mathbf{x}} \subseteq \mathcal{A}^n$ denote the set of kernels of G^n that cover $\mathbf{x}$. Then

$$Q^+(\mathbf{x} \text{ is not covered by one of } A_1, \ldots, A_{\tilde{M}_n}) = (1 - Q^{\times}(\mathcal{A}_{\mathbf{x}}))^{\tilde{M}_n} \tag{3.9}$$

Now

$$Q^{\times}(\mathcal{A}_{\mathbf{x}}) \geq \min_{A \in T^n(\tilde{Q}^*)} Q^{\times}(A) \cdot \|\mathcal{A}_{\mathbf{x}} \cap T^n(\tilde{Q}^*)\| \tag{3.10}$$

By Lemma , we have

$$\min_{A \in T^n(\tilde{Q}^*)} Q^{\times}(A) \geq 2^{-nH(\tilde{Q}^*) - C\sqrt{n}} \tag{3.11}$$

Further, $\|\mathcal{A}_{\mathbf{x}} \cap T^n(\tilde{Q}^*)\|$ is bounded below by the number of $\underline{Q}^*$ typical $X \circ \mathcal{A}$ valued n-sequences whose first co-ordinate is $\mathbf{x}$, because the second co-ordinate of every such n-sequence is a $\tilde{Q}^*$- typical n-sequence in $\mathcal{A}^n$ containing $\mathbf{x}$. By considering the equivalence relation π on $X \circ \mathcal{A}$ defined by the equality of the first co-ordinate, applying Lemma 2.2, gives the lower bound

$$\|\mathcal{A}_{\mathbf{x}} \cap T^n(\tilde{Q}^*)\| \geq 2^{nH(\underline{Q}^*|\pi) - C\sqrt{n}} \tag{3.12}$$

From (3.9) -(3.12), we get

$$\begin{aligned} Q^+(\mathbf{x} \text{ is not covered by one of } A_1, \ldots, A_{\tilde{M}_n}) \leq & (1 - 2^{-nH(G,Q) + C\sqrt{n}})^{\tilde{M}_n} \\ \leq & 2^{-\tilde{M}_n 2^{-nH(G,Q) + C\sqrt{n}}} \\ \leq & 2^{-2^{-n\epsilon + C\sqrt{n} + \|X\| \log n}} \end{aligned}$$

From this we get

$$E_{Q^+}P^n(T^n(Q) - \cup_{i=1}^{\tilde{M}_n} A_i \leq \sum_{\mathbf{x}\in T^n(Q)} P^n(\mathbf{x})Q^+(\mathbf{x} \text{ is not covered by one of } A_1,\ldots,A_{\tilde{M}_n})$$
$$\leq 2^{-2^{-n\epsilon+C\sqrt{n}+\|X\|\log n}}$$

It follows that we can choose $\tilde{M}_n$ kernels $A_1,\ldots,A_{\tilde{M}_n}$ of G^n such that the P^n probability of all the n-sequences in $T^n(Q)$ which are not covered by $\cup_{i=1}^{\tilde{M}_n} A_i$ is at most $2^{-2^{-n\epsilon+C\sqrt{n}+\|X\|\log n}}$. Thus one can choose $\lfloor 2^{nR} \rfloor$ kernels of G^n such that the P^n probability of the totality of strings of type Q for any Q with $H(G,Q) < R-\epsilon$ is at most $(n+1)^{\|X\|}2^{-2^{-n\epsilon+C\sqrt{n}+\|X\|\log n}}$. This yields the existence of a source coding scheme $(e_n)_{n=1}^{\infty}$ for which

$$\text{Err}(e_n) \leq P^n(\cup_{H(G,Q)\geq R-\epsilon} T^n(Q)) + (n+1)^{\|X\|}2^{-2^{-n\epsilon+C\sqrt{n}+\|X\|\log n}}$$

and

$$1-\text{Err}(e_n) \geq P^n(\cup_{H(G,Q)\geq R-\epsilon} T^n(Q)) - (n+1)^{\|X\|}2^{-2^{-n\epsilon+C\sqrt{n}+\|X\|\log n}}$$

For this source coding scheme

$$\limsup \frac{1}{n}\log(\text{Err}(e_n)) \leq -\inf_{H(G,Q)\geq R-\epsilon} K(Q,P)$$

and

$$\liminf \frac{1}{n}\log(1-\text{Err}(e_n)) \geq -\inf_{H(G,Q)< R-\epsilon} K(Q,P)$$

Using Lemma 2.4 and the continuity of $K(Q,P)$ in Q, one can now use a straightforward diagonal argument to get a source coding scheme with

$$\limsup \frac{1}{n}\log(\text{Err}(e_n)) \leq -\inf_{H(G,Q)\geq R} K(Q,P)$$

and

$$\liminf \frac{1}{n}\log(1-\text{Err}(e_n)) \geq -\inf_{H(G,Q)< R} K(Q,P)$$

proving (1.2) and (1.3).

References

[1] I. Csiszar and J. Korner, *Information Theory : Coding Theorems for Discrete Memoryless Channels*, Academic Press, New York, 1982.

[2] I. Csiszar, J. Korner, L. Lovasz, K. Marton and G. Simonyi, "Entropy Splitting for antiblocking pairs and perfect graphs", *Preprint*, 1988.

[3] R. Ellis, *Entropy, large Deviations and Statistical Mechanics*, Grundlehren der Mathematischen Wissenschaften, Vol. 271, Springer, 1985.

[4] J. Korner, "Fredman-Komlos Bounds and Information Theory", *SIAM Journal of Algebraic and Discrete Methods*, Vol. 7, No. 4, pp. 560 -570, 1986.

[5] J. Korner, "Coding an Information Source having Ambiguous Alphabet and the Entropy of Graphs", in *Transactions of the 6th Prague Conference on Information Theory*, Academia, Prague, pp. 411 -425, 1973.

[6] J. Korner, "A property of conditional Entropy", *Studia Sci. math. Hung.*, Vol. 6, pp. 355 -359, 1971.

[7] J. Korner and G. Longo, "Two step Encoding for Finite Sources", *IEEE Transactions on Information Theory*, Vol. IT-19, No. 6, pp. 778 -782, 1973.

[8] J. Korner and K. Marton. "Random Access Communication and Graph Entropy", *IEEE Transactions on Information Theory*, Vol. 34, No. 2, pp. 312 -314, 1988.

[9] J. Korner and K. Marton, "New bounds for Perfect hashing using Information Theory", *European Journal of Combinatorics*, Vol. 9, pp. 523 -530, 1988.