Exponential Bounds for List Size Moments and Error Probability

Emre Telatar
Room 2C-174, Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, USA
Email: telatar@bell-labs.com

Abstract — **We consider list decoding with a variable list size for discrete memoryless channels. We obtain simultaneous upper bounds on the error probability and the moments of list size.**

I. Introduction

In list decoding the decoder is not restricted to declaring a single estimate of the transmitted message. It may declare a list of messages containing none, one, or more than one estimate. In addition to being theoretically useful, these decoders may simplify the overall encoder and decoder design if they are made a part of a concatenated code: the inner code can employ a list decoder and the outer decoder can decide on a member of the list based on contextual information or redundancy introduced at the outer encoder [1].

A list decoder is said to make an error if the list it declares does not contain the transmitted message. If we were happy with a decoder that declares lists of unbounded size, it would be a simple matter to design list decoders that make no errors: simply declare the set of all messages. However, this is rather unsatisfactory and thus we should investigate the tradeoff between the list size and the probability of error.

There are several possible ways to measure how large a list is. In this paper will consider moments of the cardinality of the decoded list as our list size. Accordingly, we will fix a positive real number ρ and define $E\big[|\mathcal{L}|^\rho\big]$ be our list size where $\mathcal{L}$ is the decoded list (a random variable), $|\mathcal{L}|$ its cardinality, and E the expectation over the transmitted messages and received words. Note that the special case of $\rho = 1$ corresponds to the classical definition and is treated in [1].

Suppose we are given a discrete memoryless channel with input alphabet X, output alphabet Y, and transition probabilities $\{P(y|x),\, x \in X,\, y \in Y\}$. The extension of P to blocks of n inputs and outputs will be denoted by P^n. A block code of block length n is a collection $C \subset X^n$. The rate of this code is $R = (\ln|C|)/n$. A list decoder for C is a mapping that assigns to every channel output word $\mathbf{y} \in Y^n$ a subset of codewords $\mathcal{L}(\mathbf{y}) \subset C$. The average error probability and the list size are then given by

$$P_e = \frac{1}{|C|} \sum_{\mathbf{c}\in C} \sum_{\mathbf{y}\in Y^n} P^n(\mathbf{y}|\mathbf{c}) 1\big(\mathbf{c} \notin \mathcal{L}(\mathbf{y})\big)$$

and

$$E\big[|\mathcal{L}|^\rho\big] = \frac{1}{|C|} \sum_{\mathbf{c}\in C} \sum_{\mathbf{y}\in Y^n} P^n(\mathbf{y}|\mathbf{c}) |\mathcal{L}(\mathbf{y})|^\rho,$$

respectively. For any rate R less than the Shannon capacity of the channel, it can be shown that by appropriate coding P_e and $E\big[|\mathcal{L}|^\rho\big] - 1$ can be made to decrease exponentially in block length. We will say that a pair (E_e, E_ℓ) of real numbers is an *achievable error and* ρ^{th} *moment list exponent pair at rate* R (an *achievable pair* for short) if there is a sequence of codes of rate at least R for which for sufficiently large n

$$P_e \le \exp -nE_e \quad \text{and} \quad E\big[|\mathcal{L}|^\rho\big] \le 1 + \exp -nE_\ell.$$

II. Results

We will only consider codes of a fixed composition, i.e., all codewords will be of a given type $Q \in T^n(X)$. We will consider list decoding rules of the following form: Fix a set $\mathcal{D} \subset \mathcal{T}^n(Y|X) \times \mathcal{T}^n(Y|X)$ where $\mathcal{T}^n(Y|X)$ is the set of conditional types. Upon receiving a $\mathbf{y} \in Y^n$, compute its conditional type $V_\mathbf{c}$ with respect to each codeword $\mathbf{c} \in C$. Put $\mathbf{c}$ in $\mathcal{L}(\mathbf{y})$ is $(V_\mathbf{c}, V_{\mathbf{c}'}) \in \mathcal{D}$ for all $\mathbf{c}' \in C \setminus \{\mathbf{c}\}$. This is a decoding rule in the spirit of [2] and our results can be considered a generalization of the results in [3].

Theorem. *Given a channel* P *and a* $\rho > 0$,

$$\Big(\min_{\substack{V,W:(V,W)\notin\mathcal{D}\\ VQ=WQ}} \big[D(V\|P|Q) + |I(Q,W) - R|^+\big], \\ \min_{\substack{V,W:(W,V)\in\mathcal{D}\\ VQ=WQ}} \big[D(V\|P|Q) + \rho(I(Q,W) - R)\big]\Big)$$

is an achievable pair at rate R *for any* Q *and* $\mathcal{D}$. *In the expression above* V *and* W *range over conditional distributions,* VQ *denotes the distribution at the output of a channel* V *when* Q *is the input distribution, and* D *and* I *are the informational divergence and mutual information respectively.*

An appropriate choice of $\mathcal{D}$ yields the results in [4]. A more complicated choice of $\mathcal{D}$ yields results that are tighter than those obtained in [1] for $\rho = 1$.

III. Remarks

The threshold rule given in [1, Eq. (15)] is no longer the optimal criterion for inclusion in a list when ρ is different from 1. Nor is the rule considered above, not even for $\rho = 1$. It yields tighter results only because we restrict the codes to be of a fixed composition.

References

[1] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory,* vol. IT-14, no. 2, pp. 206–220, March, 1968.

[2] I. Csiszár and J. Körner", "Graph decomposition: a new key to coding theorems," *IEEE Transactions on Information Theory,* vol. IT-27, no. 1, pp. 5–12, January, 1981.

[3] E. Telatar and R. G. Gallager, "New exponential upper bounds to error and erasure probabilities," *Proceeding of the IEEE International Symposium on Information Theory,* Trondheim, Norway, p. 379, June–July 1994.

[4] E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Transactions on Information Theory,* vol. IT-43, no. 6, pp. 1977–1982, November, 1997.