

The Empirical Distribution of Good Codes

Shlomo Shamai (Shitz)¹ and Sergio Verdú²

¹Department of Electrical Engineering, Technion-Israel Inst. of Technology, Haifa 32000, Israel sshlomo@ee.technion.ac.il

²Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A. Verdu@Princeton.edu

Abstract — Finding the input distribution that maximizes mutual information leads, not only to the capacity of the channel, but to engineering insights that tell the designer what good codes should be like. This is due to the folk theorem: *The empirical distribution of any good code (i.e., approaching capacity with vanishing probability of error) maximizes mutual information. This paper formalizes and proves this statement.*

I. INTRODUCTION

The unique n -dimensional distribution that maximizes the n -block input-output mutual information of a binary symmetric channel (BSC) puts equal mass on all 2^n binary n -strings. Thus, common wisdom in information theory indicates that in order to approach the capacity of a BSC, a code must be such that the ensemble of its equiprobable codewords appears to be generated by a source of independent equally-likely bits. Formalizing and proving such a statement is not trivial as evidenced by the fact that the entropy rate of a source of pure bits is equal to 1 bit, whereas the entropy rate of the channel input induced by 2^{nR} equiprobable codewords is equal to R , and if the probability of error is to vanish, then $R \leq 1 - h(p) < 1$. Thus, convergence of the n -dimensional input distributions to a Bernoulli-1/2 source is ruled out. A good deal of the intuition on which the above common wisdom is grounded arises from the consideration of the input distributions of *random coding*, where not only do we average over equiprobable codewords, but over codebooks generated randomly according to the distribution maximizing mutual information. Then, the averaged input distributions of a random code are trivially equal to the capacity achieving input distributions. However, this trivial conclusion predicts nothing about the behavior of the input distributions of any particular code, which is the problem of interest.

It has been shown in [1] that for any finite-input channel that satisfies the strong converse, the *output* distribution induced by any good code sequence converges (in normalized divergence) to the (unique) output distribution induced by a capacity achieving input distribution. In certain cases (such as discrete memoryless channels with full-rank transition matrices [2]), such a result implies convergence of the input statistics. However, in general, such convergence does not follow directly from the convergence of output statistics.

II. DEFINITIONS

A. Empirical Distributions. For every codeword of a channel code we can find its first-order empirical distribution by computing the fraction of symbols in the codeword equal to each input letter. If for a given codebook we average the empirical distributions over equiprobable codewords we obtain the *first-order empirical distribution of the code*. Analogously, κ -th order empirical distributions can be defined by computing for each κ -string \mathbf{v} the fraction of κ -strings within the codeword equal to \mathbf{v} . Averaging over equiprobable codewords results in

the κ -th order empirical distribution of the code. Thus, for a code composed of M codewords of blocklength n , $\{z_{im}, i = 1 \dots n, m = 1, \dots, M\}$, the κ -th-order empirical distribution, $P_{\hat{X}^{(\kappa)}}^n$, is defined as:

$$P_{\hat{X}^{(\kappa)}}^n = \frac{1}{n - \kappa + 1} \sum_{i=1}^{n-\kappa+1} P_{\hat{X}_i^{(\kappa)}}^n$$

where

$$P_{\hat{X}_i^{(\kappa)}}^n(a_1, \dots, a_\kappa) = \frac{1}{M} \sum_{m=1}^M 1\{z_{im} = a_1\} \cdots 1\{z_{i+\kappa-1,m} = a_\kappa\}$$

B. Good Codes are channel codes whose rate is close to the channel capacity and whose decoding error probability vanishes with blocklength. More precisely, a *good code-sequence* for a channel with capacity C is a sequence of (n, M, λ_n) codes such that:

$$\lambda_n \rightarrow 0, \\ \liminf_{n \rightarrow \infty} \frac{\log M}{n} = C.$$

III. DISCRETE MEMORYLESS CHANNELS

We have obtained results for a variety of channels, including channels with memory and continuous-alphabet channels. Our main result for discrete memoryless channels (DMC) is

Theorem 1 *Consider any good code sequence which does not use any symbol having zero mass under every input distribution that maximizes the single-letter mutual information. Then, the κ -order empirical distribution of such a code sequence satisfies:*

$$\lim_{n \rightarrow \infty} \min_{P_{\mathcal{X}} \text{ s.t. } I(\mathcal{X}; \mathcal{Y}) = C} \mathbb{D}\left(P_{\hat{X}^{(\kappa)}}^n \| P_{\mathcal{X}} \times \cdots \times P_{\mathcal{X}}\right) = 0.$$

where C is the channel capacity.

Note that the existence of a good code sequence satisfying the approximation property in Theorem 1 for any fixed κ is predicted by the optimality of constant-composition codes. But, in fact, this result holds for *any* good code sequence because of Theorem 1. A refinement of Theorem 1 entails letting κ grow with n . We have shown that any growth faster than $\log n$ destroys convergence.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the U.S.-Israel Binational Science Foundation. Fruitful discussions with Professor Amir Dembo are acknowledged.

REFERENCES

- [1] T. S. Han and S. Verdú, "Approximation Theory of Output Statistics," *IEEE Trans. Inform. Theory*, vol. 39, No. 3, pp. 752-772, May 1993.
- [2] T. S. Han and S. Verdú, "Spectrum Invariance under Output Approximation for Discrete Memoryless Channels with Full Rank," *Problems of Information Transmission*, pp. 101-118, April - June, 1993.