Conference Paper

Refinement of the random coding bound

Author(s):
Altuğ, Yücel; Wagner, Aaron B.

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-007052679

Rights / License:
In Copyright - Non-Commercial Use Permitted
Abstract—We provide an improved random coding bound for a class of discrete memoryless channels that improves the pre-factor in front of the exponent. Specifically, while the tightest known bounds have a constant pre-factor, the presented bound has a pre-factor of order $O(1/\sqrt{N})$, where $N$ is the blocklength.

I. INTRODUCTION

Characterizing the interplay between the data rate, error probability, and blocklength of the best codes for a given discrete memoryless channel (D.M.C.) is a central problem in information theory. Although this interplay has been investigated starting from the early days of the field [1], [2], [3], [4], [5], [6], [7], [8], [9], it is still an active research topic [10], [11], [12], [13], [14]. Due to the difficulty of the problem, it is customary to consider formulations that are asymptotic in the blocklength to characterize this interplay.

One of the best-known formulations is “error exponents,” in which the rate is held fixed below the channel capacity and one characterizes the rate of decay of the error probability, which is exponential in $N$ for most channels [3], [4], [5], [6]. Another asymptotic characterization is the “normal approximation” in which the error probability is fixed at a positive constant in $(0, 1/2)$ and one characterizes the speed with which the data rate converges to capacity [2], [13]. Between these two characterizations, one can consider the setup in which the error probability tends to zero and the rate simultaneously approaches capacity, with the goal of characterizing the speed of one convergence as a function of the other [14]. We call these three asymptotics the “small error probability,” “large error probability,” and “medium error probability” regimes, respectively. It should be noted that although there is an underlying relation between these three asymptotics via the minimum error probability of the $(N, R)$ codes, it is not possible to directly derive one regime’s conclusion from any of the other two.

In this paper, we consider upper bounds on the error probability of the best $(N, R)$ code in the small error probability regime. We prove a refinement of the random coding bound that has a pre-factor of $O(1/\sqrt{N})$ for a broad class of D.M.C.s, which includes positive channels with positive dispersion. This improves the best known pre-factor of $O(1)$ due to Fano [3] and Gallager [4]. It should be remarked that the improvement in the pre-factor is significant for rates close to the capacity, where the optimal error exponent itself is small and the pre-factors are therefore important. Although some improved finite-$N$ bounds could easily be extracted from the proofs in this paper, we defer the task of optimizing these bounds and numerically comparing them to the existing bounds for future work.

To better understand the asymptotic regimes mentioned above, noting the analogy between channel coding and the analysis of sums of i.i.d. random variables in probability theory is instructive. In particular, the small, medium, and large error probability regimes of channel coding correspond to large deviations, moderate deviations, and central limit theory, respectively, for i.i.d. sums. With this analogy in mind, the problem we consider in this paper is reminiscent of the “exact asymptotics” problem [18], [21, Theorem 3.7.4] in large deviations. The aim of that problem is to determine the pre-factor of the exponentially vanishing term in the large deviations theorem. Bahadur and Rao [18] characterized this pre-factor, including the constant, under some regularity conditions. Their result gives some hope that it may be possible to determine the optimal pre-factor, or at least its order, for channel coding in the small error probability regime. The current best pre-factors in the sphere packing bound [22] and random coding bound (this paper) do not coincide, however, and the order of the optimal pre-factor is only known for special cases such as the binary symmetric [15], binary erasure [15], and power-constrained A.W.G.N. [16] channels.

II. NOTATION, DEFINITIONS AND STATEMENT OF THE RESULT

A. Notation

Boldface letters denote vectors, regular letters with subscripts denote individual elements of vectors. Furthermore, capital letters represent random variables and lowercase letters denote individual realizations of the corresponding random variable. Throughout the paper, all logarithms are base-$e$. For a finite set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the set of all probability measures on $\mathcal{X}$. Similarly, for two finite sets $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ denotes the set of all stochastic matrices from $\mathcal{X}$ to $\mathcal{Y}$, $\mathbb{R}$, $\mathbb{R}^+$ and $\mathbb{R}_+$ denote the set of real, positive real and non-negative real numbers, respectively. $\mathbb{Z}^+$ denotes the set of positive integers.

We follow the notation of Csiszár-Körner [9] for fundamental information theoretic quantities.

B. Definitions

Throughout the paper, let $W$ be a D.M.C. from $\mathcal{X}$ to $\mathcal{Y}$, $E_n(R, W)$ and $E_n(\rho, Q, W)$ denotes the random coding exponent (cf. [7, eq. (5.6.16)]) and the Gallager’s function (cf. [7, eq. (5.6.14)]), respectively. For any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{X})$ the ensemble average error probability (resp. ensemble average error probability conditioned on the
message \( m \) of a random code with codewords generated by using \( Q \) along with a maximum likelihood decoder\(^1\) is denoted by \( \bar{P}_e(Q) \). \( \sigma^2(W) \) denotes the dispersion (cf. [2], [13]) of \( W \in P(\mathcal{Y}|\mathcal{X}) \).

Further,

\[
S_{Q,W} := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : Q(x)W(y|x) > 0 \}, \tag{1}
\]

\[
\tilde{S}_{Q,W} := \{(x, y, z) : Q(x)W(y|x)Q(z|W(y|z)) > 0 \}, \tag{2}
\]

\[
\mathcal{X}_y(W) := \{ x \in \mathcal{X} : W(y|x) > 0 \}. \tag{3}
\]

Given a \((Q, W) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}|\mathcal{X})\) pair, the following property amounts to saying that “feasibility decoding is optimal” (F.D.I.O.) when a random code with distribution \( Q \) is used for transmission over the D.M.C. \( W \).

**Property 1:** (F.D.I.O.) For all \((x, y, z) \in \tilde{S}_{Q,W}, W(y|x) = W(y|z)\).

**C. Statement of the result**

**Theorem 1:** Let \( W \in P(\mathcal{Y}|\mathcal{X}) \) be arbitrary with \( \sigma^2(W) > 0 \) and\(^2\) \( R \in [R_{cr}(W), C(W)] \). If there exists \( Q \in P(\mathcal{X}) \) achieving \( E_o(R, W) \) such that the F.D.I.O. property does not hold, then for all \( N \in \mathbb{Z}^+ \)

\[
\bar{P}_e(Q) \leq \frac{K_1}{N} e^{-N E_o(R, W)}, \tag{4}
\]

where \( K_1 \) is a positive constant that only depends on \( W \) and \( R \).

**Remark 1:** Using the standard expurgation arguments, Theorem 1 ensures the existence of an \((N, R)\) code with maximal error probability upper bounded by the right side of (4), with a different constant.

**Remark 2:** The proof of Theorem 1 is based on Fano’s proof of the random coding bound [3, pp. 324–331] with a refinement to prove the smaller pre-factor. In the literature, Fano’s argument has been superceded by Gallager’s, which is simpler and gives the same exponent. Fano’s argument, however, lends itself more readily to obtaining improved pre-factors.

**Remark 3:** Note that the assumption related to the F.D.I.O. property in Theorem 1 holds for a reasonably large class of channels, such as the class of positive channels with positive dispersion. Yet this assumption does not exclude only degenerate cases; note that the binary erasure channel (B.E.C.) is excluded. Moreover, the assumption is not necessary in order to have an upper bound on the error probability with a pre-factor of \( O(1/\sqrt{N}) \). Indeed, Elias [15] proves such an upper bound for B.E.C.

**Remark 4:** The above theorem applies to the ensemble average error probability at rates above the critical rate. For rates below the critical rate, the order of the pre-factor turns out to hinge on the F.D.I.O property, a fact that was heretofore overlooked in the literature. The next result, Theorem 2, corrects a small oversight by Gallager [17], who claims that for any D.M.C. \( W \) and rate \( R < R_{cr}(W) \)

\[
\min_{Q \in P(\mathcal{X})} \bar{P}_e(Q) \sim \frac{g}{\sqrt{N}} e^{-N E_o(R)}, \tag{5}
\]

for some constant \( g \). In fact, his argument is valid only for those \( W \in P(\mathcal{Y}|\mathcal{X}) \) such that there exists \( Q \in P(\mathcal{X}) \) with \( E_o(1, Q, W) = \max_{P \in P(\mathcal{X})} E_o(1, P, W) \) and the F.D.I.O. property is not true. For the remaining channels, we have the following result.

**Theorem 2:** Consider an arbitrary \( W \in P(\mathcal{Y}|\mathcal{X}) \) with \( C(W) > 0 \) and \( R < R_{cr}(W) \). If the F.D.I.O. property holds for all \( Q \in P(\mathcal{X}) \) with \( E_o(1, Q, W) = \max_{P \in P(\mathcal{X})} E_o(1, P, W) \), then for all \( N \in \mathbb{Z}^+ \)

\[
K_2 e^{-N E_o(R, W)} \leq \bar{P}_e(Q) \leq e^{-N E_o(R, W)}, \tag{6}
\]

for some \( 0 < K_2 < 1 \) that depends on \( W \) and \( R \).

**III. PROOF OF THEOREM 1**

Let \( W \in P(\mathcal{Y}|\mathcal{X}) \) with \( \sigma^2(W) > 0 \) be arbitrary. Fix an arbitrary \( R \in [R_{cr}(W), C(W)] \). Consider an arbitrary \( Q \in P(\mathcal{X}) \) with \( E_o(R, Q, W) = E_o(R, W) \), such that the F.D.I.O. property does not hold, whose existence is guaranteed by the assumption of the theorem. For the sake of notational convenience, let \( \tilde{S}_Q, \mathcal{S}_Q \) and \( \mathcal{X}_y \) denote \( \tilde{S}_{Q,W}, \mathcal{S}_{Q,W} \) and \( \mathcal{X}_y(W) \) as given in (1), (2) and (3), respectively. Define

\[
P_{X,Y,Z}(x, y, z) := Q(x)W(y|x)Q(z), \tag{7}
\]

for all \((x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \). Also, let

\[
\tilde{P}_{X,Y,Z}(x, y, z) := \begin{cases} \frac{P_{X,Y,Z}(x, y, z)}{P_{X,Y,Z}(\tilde{S}_Q)} & \text{if } (x, y, z) \in \tilde{S}_Q \\ 0 & \text{else} \end{cases} \tag{8}
\]

Define

\[
P_{N,X,Y,Z}^{m}(x, y, z) := \prod_{n=1}^{N} P_{X,Y,Z}(x_n, y_n, z_n) \quad \text{and} \quad \mathcal{S}_Q^N \quad \text{(resp. } \tilde{S}_Q^N \text{)} \quad \text{denote the } N \text{-fold cartesian product of } \mathcal{S}_Q \text{ (resp. } \tilde{S}_Q \text{)}. \quad \text{Hence,}
\]

\[
P_{N,X,Y,Z}^{m}(x, y, z | \mathcal{S}_Q^N) = \tilde{P}_{N,X,Y,Z}(x, y, z | \mathcal{S}_Q^N) := \prod_{n=1}^{N} \tilde{P}_{X,Y,Z}(x_n, y_n, z_n). \tag{9}
\]

One can show that

\[
\exists ! \rho^* \in [0, 1], \quad \text{s.t.} \quad \frac{\partial E_o(\rho, Q, W)}{\partial \rho} \bigg|_{\rho = \rho^*(Q, W)} = R. \tag{9}
\]

Define

\[
D_0 := \lambda' \left( \frac{\rho^*}{1 + \rho^*} \right) \tag{10}
\]

\[
\lambda_1(s, t) := \log E_{\tilde{P}_{X,Y,Z}} \left[ e^{s \log \frac{W(y|x)}{W(x,y)}} + t \log \frac{W(x,y)}{W(y|x)} \right], \tag{11}
\]

for any \((s, t) \in \mathbb{R}^2 \). Observe that owing to the fact that \( \log W(y|x) \leq \log W(x,y) \in \mathbb{R} \), for all \((x, y, z) \in \tilde{S}_Q, \lambda_1(\cdot) \) is infinitely differentiable on \( \mathbb{R}^2 \).
Finally, for any \( N \in \mathbb{Z}^+ \), let
\[
D_N := \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \frac{f(Y_n)}{W(Y_n|X_n)} > D_0 \right\}.
\]

From the standard random coding arguments, one can show that
\[
\hat{P}_{\epsilon,1}(Q) \leq P_{X,Y,Z}^N \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \frac{f(Y_n)}{W(Y_n|X_n)} \leq D_0, \right.
\]
\[
\left. \frac{1}{N} \sum_{n=1}^{N} \log \frac{W(Y_n|X_n)}{W(Y_n|Z_n)} \leq 0 \right\} \cdot e^{NR} + P_{X,Y}^N \{ D_N \}.
\]

(12)

In order to conclude the proof, we need to upper bound the two terms on the right side of (12).

We begin with the second term. To this end, define
\[
\Lambda^*(a) := \sup_{\lambda \in \mathbb{R}} \{ a\lambda - \Lambda(\lambda) \},
\]
for any \( a \in \mathbb{R} \). By noting the convexity of \( \Lambda(\cdot) \); (10), (13) and some algebra imply that
\[
\Lambda^*(D_0) = E_r(R, W).
\]

(14)

Owing3 to a result of Ney [19, eq. (2)], we deduce4 that
\[
P_{X,Y}^N \{ D_N \} \leq \frac{k_1}{\sqrt{N}} e^{-N \Lambda^*(D_0)} = \frac{k_1}{\sqrt{N}} e^{-N E_r(R, W)},
\]

for some \( k_1 \in \mathbb{R}^+ \) and for all sufficiently large \( N \), where the equality follows from (14).

Next, we upper bound the first term on the right side of (12). First, since for any \( (x, y, z) \notin \hat{S}_Q \),
\[
\log \frac{W(y|x)}{W(y|z)} = \infty, \quad Q \times W \times Q - \text{a.s.}
\]
we have
\[
\alpha_N := \frac{P_{X,Y,Z}^N}{P_{X,Y}^N} \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \frac{f(Y_n)}{W(Y_n|X_n)} \leq D_0, \right.
\]
\[
\left. \frac{1}{N} \sum_{n=1}^{N} \log \frac{W(Y_n|X_n)}{W(Y_n|Z_n)} \leq 0 \right\} \cdot \hat{S}_Q^N \hat{\alpha}_N,
\]

(16)

where, in (16) we define
\[
\hat{\alpha}_N := \frac{\tilde{P}_{X,Y,Z}^N}{P_{X,Y,Z}^N} \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \frac{f(Y_n)}{W(Y_n|X_n)} \leq D_0, \right.
\]
\[
\left. \frac{1}{N} \sum_{n=1}^{N} \log \frac{W(Y_n|X_n)}{W(Y_n|Z_n)} \leq 0 \right\}.
\]

(17)

Define
\[
t^* := -1/(1 + \rho^*), \quad s^* := 1 + 2t^*.
\]

(18)

One can show that
\[
\frac{\partial \Lambda_1(s, t)}{\partial s}\bigg|_{s=s^*} = D_0, \quad \frac{\partial \Lambda_1(s, t)}{\partial \tilde{t}}\bigg|_{t=t^*} = 0.
\]

(19)

Before proceeding further, we define
\[
B := (-\infty, D_0) \times (-\infty, 0], \quad a_B := [D_0, 0]^T, \quad \Lambda_1^*(b) := \sup_{a \in \mathbb{R}^2} \{ \langle b, a \rangle - \Lambda_1(a) \}.
\]

(20)

for any \( b \in \mathbb{R}^2 \). Note that owing to the convexity of \( \Lambda_1(\cdot) \), one can show that
\[
\Lambda_1^*(a_B) = s^* D_0 - \Lambda_1(s^*, t^*),
\]

(22)

where \( s^* \) and \( t^* \) are as defined in (18).

Moreover,
\[
\Lambda_0(0) := \log E_{P_{X,Y,Z}} \left[ e^{\Lambda \log \frac{W(Y|X)}{W(Y|Z)}} \right] > 0.
\]

(24)

A result of Ney [19, Theorem] implies that5 there exists a unique dominating point (cf. [19, Definition]) \( d_B \in B \), which also satisfies (cf. [20, eq. (3.3)]),
\[
\Lambda_1^*(d_B) = \inf_{b \in B} \sup_{a \in \mathbb{R}^2} \{ \langle b, a \rangle - \Lambda_1(a) \}.
\]

(25)

Moreover, [19, eq. (2)] implies that
\[
\frac{c_1}{\sqrt{N}} e^{-N \Lambda_1^*(d_n)} \geq \hat{\alpha}_N \leq \frac{c_2}{\sqrt{N}} e^{-N \Lambda_1^*(a_B)},
\]

(26)

for some \( c_1, c_2 \in \mathbb{R}^+ \) and all sufficiently large \( N \). Next, one can prove that
\[
\Lambda_1^*(a_B) = \Lambda_1^*(d_B).
\]

(27)

Plugging (27) into (26) yields
\[
\hat{\alpha}_N \leq \frac{c_2}{\sqrt{N}} e^{-N \Lambda_1^*(a_B)},
\]

(28)

for some \( c_2 \in \mathbb{R}^+ \) and all sufficiently large \( N \). Moreover, it is possible to check that
\[
\Lambda_1(s^*, t^*) = -\log P_{X,Y,Z} \left\{ \tilde{S}_Q \right\} + 2\Lambda \left( \frac{\rho^*}{1 + \rho^*} \right).
\]

(29)

3Because of the fact that the random variables in \( D_N \) are i.i.d., we might have used the stronger `exact asymptotics’ result [18] by considering lattice and non-lattice cases separately. However, the application of such a result has little effect on the ultimate bound since the second term of (12) cannot be handled by an analogous theorem.

4To justify the application of the aforementioned result, one can check that \( \Lambda(\cdot) \) satisfies the regularity conditions stated in [19, Theorem]. Moreover, one can verify that \( \Lambda^* \left( \frac{1}{N} \log \frac{W(Y|X)}{W(Y|Z)} \right) > \Lambda^*(0) = E_{P_{X,Y,Z}} \left[ \log \frac{W(Y|X)}{W(Y|Z)} \right] \cdot \hat{S}Q \). This guarantees the “mean does not belong to the set” regularity condition that \( D_N \) satisfy. The remaining regularity conditions on \( D_N \) are easy to check. Therefore the existence of a unique “dominating point” (cf. [19, Definition]) is guaranteed owing to [19, Theorem]. However, one can check that this dominating point is \( D_0 \), (e.g. [20, pg. 159]). Lastly, [20, eq. (3.3)] gives the exponential factor in (15).

5Similar to the previous footnote, (24) ensures the fulfillment of the “mean does not belong to the set” condition of the theorem. The remaining conditions are easy to check.
Finally, plugging (18), (22) and (29) into (28) along with (10), (14) and some algebra yields
\[
\tilde{\alpha}_N \leq P_{X,Y,Z} \left\{ \tilde{S}_Q \right\}^{-N} \frac{c_2}{\sqrt{N}} e^{-N[\tilde{E}_c(R,W)+R]}.
\] (30)

By plugging (30) into (16) and noting \( P_{X,Y,Z}^{N} \left\{ \tilde{S}_Q ^N \right\} = P_{X,Y,Z} \left\{ \tilde{S}_Q \right\}^N \), we deduce that
\[
\alpha_N \leq \frac{c_2}{\sqrt{N}} e^{-N[\tilde{E}_c(R,W)+R]}.
\] (31)

Lastly, (12), (15) and (31) imply that
\[
P_{e,1}(Q) \leq \frac{K_1}{\sqrt{N}} e^{-N\tilde{E}_e(R,W)},
\] (32)
for some \( K_1 \in \mathbb{R}^+ \) and all sufficiently large \( N \).

\section*{IV. PROOF OF THEOREM 2}

Let \( W \in \mathcal{P}(\mathcal{Y}|\mathcal{X}) \) with \( C(W) > 0 \) and \( R > R_{ca}(W) \) be arbitrary. Assume that for all \( Q \in \mathcal{P}(\mathcal{X}) \) with \( E_0(1,Q,W) = \max_{P \in \mathcal{P}^N(\mathcal{X})} E_0(1,P,W) \), the F.D.I.O. property holds. Consider any such \( Q \in \mathcal{P}(\mathcal{X}) \). For this \( (Q,W) \) pair, let \( P_{X,Y,Z} \) and \( \tilde{P}_{X,Y,Z} \) be as given in (7) and (8), respectively. For the sake of notational convenience, let \( \tilde{S}_Q \) and \( \tilde{X}_y \) denote \( \tilde{S}_{Q,W} \) and \( \tilde{X}_y(W) \) as given in (2) and (3), respectively.

Owing to the fact that the F.D.I.O. property holds, one can show that
\[
\log P_{X,Y,Z} \left\{ \tilde{S}_Q \right\} = -E_0(1,Q,W).
\] (33)

By using the standard random coding arguments, we have
\[
P_{e,1}(Q) \leq e^{-NR} P_{X,Y,Z} \left\{ \tilde{S}_Q \right\}^N
\] (34)
\[
= e^{-N[1-R+E_0(1,Q,W)]}
\] (35)
\[
= e^{-N\tilde{E}_e(R,W)},
\] (36)

where (34) follows from the fact that for any \((x,y,z) \notin \tilde{S}_Q\),
\[
\log \frac{W(y|x)}{W(y|z)} = \infty, \quad Q \times W \times Q - (a.s.), \quad \text{and the F.D.I.O. property is satisfied.}
\]
(35) follows from (33) and (36) is true owing to the choice of \( Q \in \mathcal{P}(\mathcal{X}) \) and the fact that \( R < R_{ca}(W) \) (e.g. [17, pg. 245]). Hence, the upper bound of (6) follows.

In order to establish the lower bound of (6), one can use Gallager’s arguments [17, pp. 245-246] by noting
\[
P_{X,Y,Z}^{N} \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \frac{W(Y_n|X_n)}{W(Y_n|Z_n)} \leq 0 \right\} = e^{-NE_0(1,Q,W)}.
\]

\section*{V. COMPARISON WITH THE SPHERE-PACKING BOUND}

Our main result is an improvement of the random coding bound for rates greater than the critical rate, obtained by improving the pre-factor from \( O(1) \) to \( O(1/\sqrt{N}) \) for a broad class of channels.

Even for the symmetric channels, however, this pre-factor does not match its lower bound counterpart [22], namely \( O(N^{-2(1+\tilde{E}_c(R,W))}) \), where \( \tilde{E}_c(R) \) is the slope of \( \text{Esp}(R) \). It is worth noting, however, that for rates close to capacity, \( \tilde{E}_c(R) \) is nearly zero, so the order of the pre-factor is nearly determined in this regime.

\section*{REFERENCES}