Error Exponents of Typical Random Codes

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

ISIT 2018, Vail, Colorado, U.S.A., June 2018.

Typical Random Codes

Traditional random coding error exponents are defined as

$$E_{\mathsf{r}}(R) = \lim_{n\to\infty}\left[-\frac{\ln \mathbf{E} P_{\mathsf{e}}(\mathcal{C}_n)}{n}\right].$$

We define typical–code error exponents as

$$E_{\mathsf{typ}}(R) = \lim_{n\to\infty}\left[-\frac{\mathbf{E} \ln P_{\mathsf{e}}(\mathcal{C}_n)}{n}\right].$$

- By Jensen's inequality, $E_{\mathsf{typ}}(R) \geq E_{\mathsf{r}}(R)$.
- $E_{\mathsf{r}}(R)$ – dominated by bad codes; $E_{\mathsf{typ}}(R)$ – dominated by typical codes.

Let $\mathcal{G}_E = \{\mathcal{C}_n : \quad P_{\mathsf{e}}(\mathcal{C}_n) \doteq e^{-nE}\}$.

$$\overline{P_{\mathsf{e}}(\mathcal{C}_n)} \doteq \sum_E P(\mathcal{G}_E)\cdot e^{-nE} \doteq P(\mathcal{G}_E^*)\cdot e^{-nE^*}.$$

Otoh, $E_{\mathsf{typ}}(R) = \sum_E P(\mathcal{G}_E)\cdot E = E_0$, where $P[\mathcal{G}_{E_0}] \to 1$.

Motivation

- $E_{\text{typ}}(R)$ is never worse than $E_{\text{r}}(R)$.
- Code selected once and for all: no LLN to support $\mathbf{E}P_{\text{e}}(\mathcal{C}_n)$.
- Once selected, w.h.p. $P_{\text{e}}(\mathcal{C}_n) \sim e^{-nE_0}$, forever.
- Theoretical framework for random–like codes (Battail, 1995).
- Analogy: physics of disordered sys. – quenched vs. annealed average.

Q: With all these motivations, why wasn't it explored much more before?

A: Not so easy to analyze (also in physics) ....

Related Work

- Barg & Forney (2002): i.i.d. random coding, BSC:

  At low rates: $E_{\text{typ}}(R) = E_{\text{ex}}(2R) + R$.

- Nazari (2011); Nazari, Anastasopoulos & Pradhan (2014):

  upper and lower bounds for the α–decoder.

- Stat. phys. literature: Kabashima (2008), Mora & Riviore (2006), ...:

  LDPC codes - replica analysis and cavity method.

- Battail (1995):

  random–like codes.

Contributions

We derive the exact typical–code error exponent for a class of stochastic decoders,

$$P(\hat{m} = m|\boldsymbol{y}) \propto \exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\},$$

e.g., $g(Q_{XY} = \beta\mathbf{E}_Q \ln W'(Y|X)$, $g(Q_{XY}) = \beta \cdot \alpha(Q_{XY})$, $g(Q_{XY}) = \beta I_Q(X;Y)$.
Extending Barg & Forney (2002) in several directions:

- General DMC is considered, not merely the BSC.
- Covering a wider family of decoders.
- Ensemble of constant composition codes – optimal PI distribution.
- Relation to expurgated exponent – for all R and a general decoder.
- The analysis technique is applicable also to more general scenarios.

Main Result

Let

$$\alpha(R, Q_Y) = \sup[g(Q_{XY}) - I_Q(X;Y)] + R,$$

where supremum is over $\{Q_{X|Y}:\ I_Q(X;Y) \leq R,\ Q_X = P_X\}$.

$$\Gamma(Q_{XX'}, R) = \inf_{Q_{Y|XX'}} \{D(Q_{Y|X}\|W|P_X) + I_Q(X';Y|X) + [g(Q_{XY}) \wedge \alpha(R, Q_Y) - g(Q_{X'Y})]_+\}.$$

Theorem: The typical error exponent is

$$E_{\text{typ}}(R) = \inf\{\Gamma(Q_{XX'}, R) + I_Q(X;X')\} - R,$$

where the infimum is over $\{Q_{XX'}:\ I_Q(X;X') \leq 2R,\ Q_X = Q_{X'} = P_X\}$.

ML Decoding

In ML decoding: minimization s.t.

$$\mathbf{E}_Q \ln W(Y|X') \geq \max\{\mathbf{E}_Q \ln W(Y|X), D(R, Q_Y)\},$$

$$D(R, Q_Y) = \sup\{\mathbf{E}_Q \ln W(Y|X''):\ I_Q(X''; Y) \leq R,\ (Q_Y \times Q_{X''|Y})_X = P_X\},$$

being the typical highest score of an incorrect message.

This is not a union bound of pairwise error events.

Relation to Expurgated Exponent

Defining

$$E_0(R,S) = \inf\{\Gamma(Q_{XX'},S) + I_Q(X;X')\} - R,$$

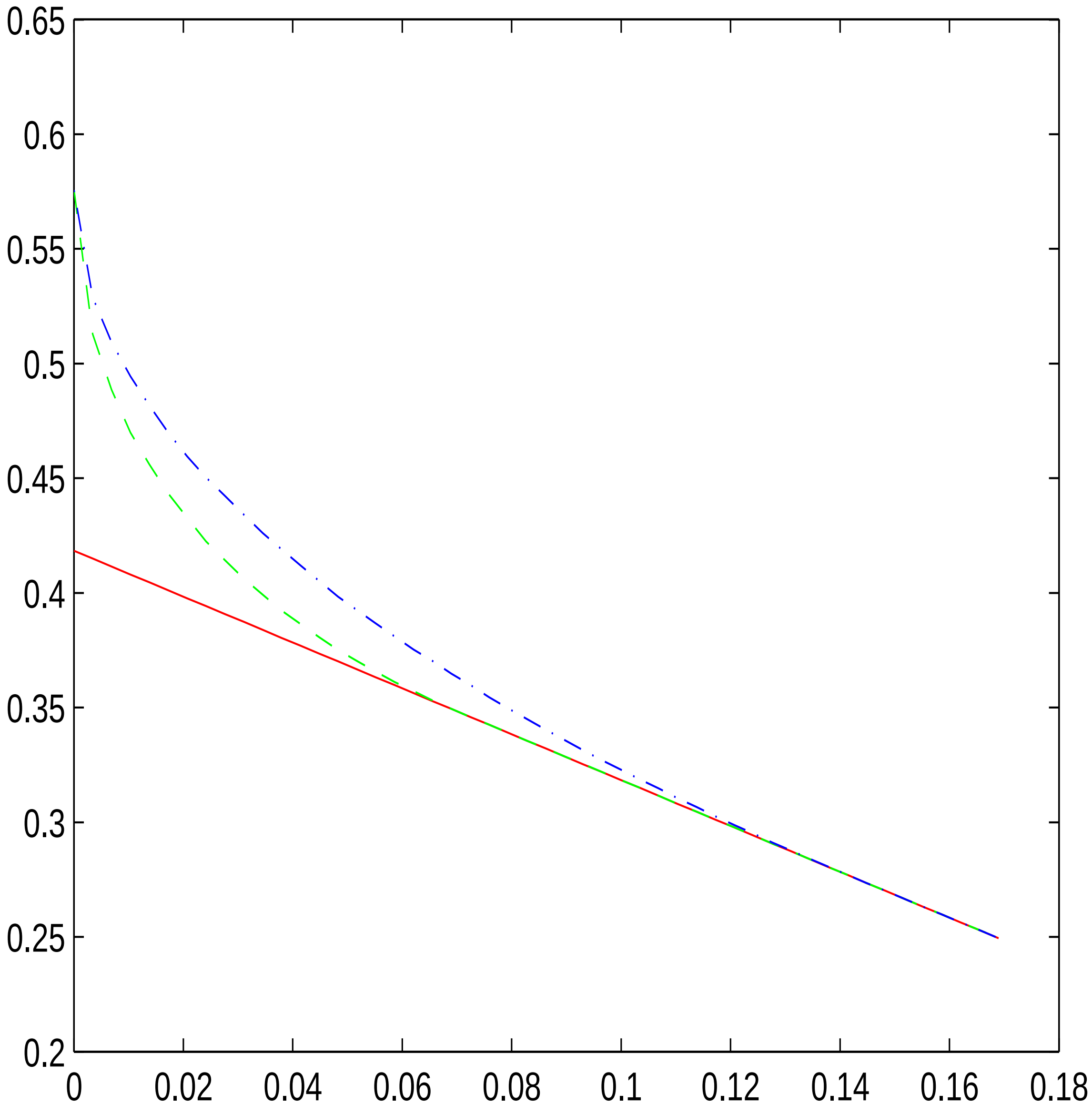
the infimum being over $\{Q_{XX'} : \; I_Q(X;X') \leq R, \; Q_X = Q_{X'} = P_X\}$, we have for all R:

$$E_{\text{ex}}(R) = E_0(R,R); \qquad E_{\text{typ}}(R) = E_0(2R,R) + R.$$

In general,

$$E_{\text{typ}}(R) \leq E_{\text{ex}}(2R) + R,$$

because here the expurgated exponent is improved.



rand. coding, expurg. and typical exponents for z–channel w. crossover 0.1.

A Few Words on the Analysis

- Using the identity $\mathbf{E} \ln P_{\mathrm{e}}(\mathcal{C}) = \lim_{\rho\to\infty} \rho \ln \mathbf{E}[P_{\mathrm{e}}(\mathcal{C})]^{1/\rho}$.
- Using the method of type class enumerators.
- Main idea behind the analysis: handling summations of exponentially many fractions with random denominators – exploit concentration properties.

$$\mathbf{E}\left[\frac{1}{M}\sum_{m}\sum_{m'\neq m}\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{X}_m)\cdot\frac{e^{ng(\boldsymbol{X}_{m'},\boldsymbol{y})}}{e^{ng(\boldsymbol{X}_m,\boldsymbol{y})}+\sum_{\tilde{m}\neq m} e^{ng(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})}}\right]^{\rho}.$$

In paritcular, with very high probability,

$$\sum_{\tilde{m}\neq m} e^{ng(\boldsymbol{X}_{\tilde{m}},\boldsymbol{y})} \geq e^{n\alpha(R,\hat{P}\boldsymbol{y})}.$$

- Showing that the reversed inequality holds for most terms w.h.p.

Other Applications

The same techniques are applicable in other scenarios:

- List decoding (fixed list size): involves the notion of multi–information.
- Decoding with an erasure option.

The details are in the paper.

Future Directions

- Analogues in source coding (e.g., Slepian–Wolf).
- Source–channel coding.
- Multi-user situations: MAC, BC, etc.
- Other (more structured) ensembles: allowing dependencies.
- Universal decoding.
- Continuous alphabets (Gaussian channel).