List Decoding – Random Coding Exponents and Expurgated Exponents

Neri Merhav

Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, Israel

ISIT 2014, Honolulu, Hawaii, June–July, 2014.
First introduced by Elias (1957) and Wozencraft (1958).

Decoder outputs a list of $L$ candidate messages (finalists).

Application: inner decoder of a concatenated code.

Error event: correct message not on the list.

Most of the literature: algorithmic issues concerning structured codes.

This talk: error exponents (random coding, sphere–packing, expurgated).
Background (Cont’d)

There are two classes of list decoders, according to the nature of list size $L$:

- $L$ is a random variable (that depends on the channel output).
- $L$ is deterministic.

The second category is further divided to:

- Fixed list size regime (FLS): $L = \text{const.}$, independent of $n$.
- Exponential list size regime (ELS): $L = e^{\lambda n}$, with $\lambda > 0$ fixed.

In this talk, we consider the second category under both regimes.
System Model and Problem Definition

A code \( C = \{x_0, x_1, \ldots, x_{M-1}\}, M = e^{nR}, \) is selected at random.

The marginal of each codeword \( x_i \in X^n \) is \( \text{Unif}\{T(Q)\} \).

The channel \( P(y|x) \) is a DMC.

The index \( I \) of the transmitted message \( x_I \) is \( \text{Unif}\{0, 1, \ldots, M - 1\} \).

The decoder outputs the indices of the \( L \) most likely messages.

Error event: \( I \) is not on the list.

Objective: characterize error exponents.
Some Well–Known Results

The following is given as an exercise, in the books of Gallager and Viterbi & Omura:

\[
\overline{P_e} \leq \min_{0 \leq \rho \leq L} M^\rho \sum_{y \in \mathcal{Y}^n} \left[ \sum_{x \in \mathcal{X}^n} P(x) P(y|x)^{1/(1+\rho)} \right]^{1+\rho}.
\]

In the \textit{fixed list–size regime}, with a product–form random coding distribution \(Q\), this yields

\[
E_r(R, L) = \sup_{0 \leq \rho \leq L} \sup_Q \left[ E_0(\rho, Q) - \rho R \right],
\]

where

\[
E_0(\rho, Q) = -\ln \left( \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q(x) P(y|x)^{1/(1+\rho)} \right]^{1+\rho} \right).
\]

Thus, \(E_r(R, 1) \equiv E_r(R)\) is the ordinary random coding exponent.
Some Well-Known Results (Cont’d)

In the exponential list-size regime, \( L = e^{\lambda n} \) [Shannon–Gallager–Berlekamp 1967]:

\[
\overline{P_e} \geq \exp\{-nE_{sp}(R - \lambda)\},
\]

where

\[
E_{sp}(R) = \sup_{\rho \geq 0} \sup_Q [E_0(\rho, Q) - \rho R],
\]

or, equivalently,

\[
E_{sp}(R) = \sup_Q \inf_{\{\tilde{P}_Y|X: \tilde{I}(X;Y) \leq R\}} D(\tilde{P}_Y|X \parallel P_Y|X|Q),
\]

In the book by Csiszár and Körner, the reader is asked to prove that \( E_r(R - \lambda) \) is achievable.
A General Non–Asymptotic Upper Bound

**Theorem:** The average probability of list error, $\overline{P_e}$, associated with the optimal list decoder, is upper bounded by

$$\overline{P_e} \leq \sum_{x,y} P(x)P(y|x) \exp \left\{ -nL \left[ \hat{I}_{xy}(X;Y) + \frac{\ln L}{n} - R - O \left( \frac{\log n}{n} \right) \right]_+ \right\},$$

where $P(x)$ is the uniform distribution over $T(Q)$ and $\hat{I}_{xy}(X;Y)$ is the empirical mutual information induced by $(x,y)$.

The proof is by a careful large deviations analysis of the binomial random variable

$$N(x, y) = \sum_{m=1}^{M-1} \mathcal{I}\{P(y|X_m) \geq P(y|x)\}.$$
The Fixed List Size Regime

The dependence on $L$ appears twice:

$$
\overline{P_e} \leq \sum_{x,y} P(x)P(y|x) \exp \left\{ -nL \left\lfloor \hat{I}_{xy}(X;Y) + \frac{\ln L}{n} - R - O \left( \log \frac{n}{n} \right) \right\} \right\},
$$

In the FLS regime, $\frac{\ln L}{n} \to 0$, and averaging $\exp\{-nL[\hat{I}_{xy}(X;Y) - R]_+\}$ yields

$$
\overline{P_e} \leq e^{-nE(R,L,Q)}, \quad \text{where}
$$

$$
E(R, L, Q) \triangleq \min_{\tilde{P}_Y|X} \left\{ D(\tilde{P}_Y|X || P_Y|X|Q) + L \cdot [\tilde{I}(X;Y) - R]_+ \right\},
$$

The best exponent is obtained by maximizing over $Q$ to yield

$$
E(R, Q) = \max_Q E(R, L, Q).
$$
This result has been obtained also in [D’yachkov 1980]. In the paper, we also show that:

- This upper bound is exponentially tight.
- It (exponentially) agrees with the expression of Gallager/Viterbi–Omura:

\[
\overline{P_e} \leq \min_{0 \leq \rho \leq L} M^\rho \sum_{y \in Y^n} \left[ \sum_{x \in X^n} P(x) P(y|x)^{1/(1+\rho)} \right]^{1+\rho},
\]

with \( P(x) = \text{Unif}\{T(Q)\} \).
- The MMI list decoder universally achieves \( E(R, L, Q) \).
The Exponential List Size Regime

\[ \overline{P_e} \leq \sum_{x,y} P(x)P(y|x) \exp \left\{ -nL \left[ \hat{I}_{x,y}(X;Y) + \frac{\ln L}{n} - R - O \left( \frac{\log n}{n} \right) \right] + \right\}, \]

In the ELS regime, \( \frac{\ln L}{n} = \lambda \). By defining

\[ \mathcal{E} = \left\{ (x,y) : \hat{I}_{x,y}(X;Y) + \lambda - R \geq \epsilon \right\}. \]

we see that the contribution of \( \mathcal{E} \) is \( \leq \exp(-n\epsilon e^{\lambda n}) = e^{-n\epsilon} \), and so,

\[ \overline{P_e} \leq \Pr\{\mathcal{E}^c\} = \exp \left\{ -n \min_{\{\tilde{P}_Y|X : \hat{I}(X;Y) \leq R - \lambda\}} D(\tilde{P}_Y|X \parallel P_Y|X|Q) \right\} \]

\[ \triangleq \exp\{-nE_{sp}(R - \lambda, Q)\} \]

which, for the optimum \( Q \), becomes \( \exp\{-nE_{sp}(R - \lambda)\} \).
The Exponential List Size Regime (Cont’d)

- The SGB lower bound is achieved – the gap with $E_r(R - \lambda)$ is closed.
- The reliability function of the ELS regime is characterized exactly.
- The universal MMI list decoder achieves the optimum exponent.
- For $\lambda = 0$, $E_{sp}(R)$ is achieved for $L \geq \rho^*(R)$, the achiever of $E_{sp}(R)$.
- Moments of $N(X, Y)$ (related to the guessing problem):

$$\lim_{n \to \infty} \inf \frac{\ln E\{N(X_0, Y)^{\rho}\}}{n} \geq \begin{cases} -E_{sp}(R) & \rho \leq \rho^*(R) \\ \rho R - E_0(\rho) & \rho > \rho^*(R) \end{cases}$$

and the bound is tight at least for large enough $\rho$. 
Define the multi–variate “Bhattacharyya distance”:

\[
d(x_0, x_1, \ldots, x_L) = -\ln \left[ \sum_{y \in \mathcal{Y}} \prod_{i=0}^{L} P(y|x_i)^{1/(L+1)} \right]
\]

and the multi–information:

\[
I(X_0; X_1; \ldots; X_L) = \sum_{i=0}^{L} H(X_i) - H(X_0, X_1, \ldots, X_L)
\]

Next, define

\[
\mathcal{A}(R, Q) \triangleq \{ P_{X_0X_1\ldots X_L} : I(X_0; X_1; \ldots; X_L) \leq LR, P_{X_0} = P_{X_1} = \ldots = P_{X_L} = Q \}.
\]
**Expurgated Exponents (Cont’d)**

**Theorem:** There exists a sequence of rate–$R$ codes for which

$$
\lim_{n \to \infty} \left[ -\frac{\ln \max_m \mathcal{P}_e \mid m}{n} \right] \geq E_{\text{ex}}(R, L), \quad \text{where}
$$

$$
E_{\text{ex}}(R, L) \triangleq \sup_Q \inf \{ \mathcal{P}_0 \mathcal{X}_1 \ldots \mathcal{X}_L \in \mathcal{A}(R, Q) \}
\left[ E_d(X_0, X_1, \ldots, X_L) + I(X_0; X_1; \ldots; X_L) \right] - LR,
$$
Expurgated Exponents (Comments)

This is an extension of the Csiszár–Körner–Marton expurgated exponent of ordinary decoding ($L = 1$).

Similarly as in the case $L = 1$, $E_{\text{ex}}(R, L)$ is given by the “distortion–rate” function:

$$D(R) = \min_{P_{X_0 X_1 \ldots X_L} \in A(R, Q)} \mathbb{E}\{d(X_0, X_1, \ldots, X_L)\}$$

for $R \leq I^*(X_0; X_1; \ldots; X_L)/L$ and by the tangential straight–line of slope $-L$ for $R > I^*(X_0; X_1; \ldots; X_L)/L$, where $I^*(X_0; X_1; \ldots; X_L)$ is induced by $P_{X_0 X_1 \ldots X_L}^*$, the achiever of $E_{\text{ex}}(\infty, L)$.

Modification to the Gaussian case: the optimum $P_{X_0 X_1 \ldots X_L}$ is always a multivariate Gaussian with zero–mean, unit–variance components whose correlation coefficients are all the same (by symmetry).
Summary of Results

- A general, non-asymptotic upper bound on the probability of list error.
- Particularizing this bound to the FLS and ELS regimes.
  - FLS: exponentially tight bound, in agreement with Gallager/Viterbi–Omura and D’yachkov.
  - ELS: established $E_{sp}(R - \lambda)$ as the reliability function.
- Both regimes: MMI list decoding achieves these exponents.
- We characterized moments of $N(X, Y)$ with relation to guessing.
- We derived an expurgated bound.