Chapter 9

Channel Reliability Function

Po-Ning Chen

Department of Communications Engineering

National Chiao-Tung University

Hsin Chu, Taiwan 30050

Motivations

- $R < C \Rightarrow Pe(n) \to 0$.
- For example,
  - if $C = 1$ bit, and there are two optimal codes with rates 0.5 bits and 0.25 bits respectively, is it possible for one to consider to use the latter code even if it results in lower information transmission rate?
  - The answer is affirmative if a higher error exponent is required.

Random-coding exponent

Definition 9.1 (channel reliability function) Let $Pe(n, R)$ be the minimum probability error achievable for block codes of length n with rate no less than R. Then the channel reliability function $E(R)$ is defined as the error exponent of $Pe(n, R)$, i.e.,

$$E(R) \stackrel{\triangle}{=} \lim_{n\to\infty} -\frac{1}{n}\log Pe(n, R).$$

Definition 9.2 (random-coding exponent) The random coding exponent for DMC with generic distribution $P_{Y|X}$ is defined by

$$E_r(R) \stackrel{\triangle}{=} \max_{0\le s\le 1}\max_{P_X}\left[-sR - \log\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}^{1/(1+s)}(y|x)\right)^{1+s}\right].$$

Theorem 9.3 (random coding exponent) For DMC with generic transition probability $P_{Y|X}$,

$$E(R) \ge E_r(R).$$

(I.e., $Pe(n, R) \le e^{-nE_r(R)}$, for sufficiently large n.)

Proof: Similar to the proof of Channel capacity, the code is randomly selected according to some input distribution $P_{\tilde{X}}$.

Random-coding exponent

Step 1: Maximum likelihood decoder. Let $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{M_n}\} \in \mathcal{X}^n$ denote the set of n-tuple block codewords selected, and let the decoding partition for symbol m (namely, the set of channel outputs that classify to m) be

$$\mathcal{U}_m = \{y^n \ : \ P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) > P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'}), \quad \text{for all } m' \neq m\}.$$

Those channel outputs that are on the boundary, i.e.,

$$\text{for some } m \text{ and } m', \quad P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) = P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'}),$$

will be arbitrarily assigned to either m or m'.

Step 2: Property of indicator function for $s > 0$. Let ϕ_m be the indicator function of $\mathcal{U}_m$. Then for all $s > 0$.

$$1 - \phi_m(y^n) \leq \left\{ \sum_{m' \neq m, 1 \leq m' \leq M_n} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/(1+s)} \right\}^s.$$

Step 3: Probability of error given codeword $\boldsymbol{c}_m$ is transmitted. Let

Random-coding exponent

$P_{e|m}$ denote the probability of error given codeword $\boldsymbol{c}_m$ is transmitted. Then

$$\begin{aligned}
P_{e|m} &\leq \sum_{y^n \notin \mathcal{U}_m} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \\
&= \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)[1 - \phi_m(y^n)] \\
&\leq \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \left\{ \sum_{m' \neq m, 1 \leq m' \leq M_n} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/(1+s)} \right\}^s \\
&= \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m) \left\{ \sum_{m' \neq m, 1 \leq m' \leq M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'}) \right\}^s
\end{aligned}$$

Step 4: Expectation of $P_{e|m}$.

$$\begin{aligned}
& E[P_{e|m}] \\
\le\ & E\left[\sum_{y^n\in\mathcal{Y}^n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\left\{\sum_{m'\neq m, 1\le m'\le M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right\}^s\right] \\
=\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\left\{\sum_{m'\neq m, 1\le m'\le M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right\}^s\right] \\
=\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\right] E\left[\left\{\sum_{m'\neq m, 1\le m'\le M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right\}^s\right]
\end{aligned}$$

where the latter step follows because $\{\boldsymbol{c}_m\}_{1\le m\le M}$ are independent random variables.

Step 5: Jensen's inequality for $0 < s \le 1$, and bounds on probability of error. By Jensen's inequality, when $0 < s \le 1$,

$$E[t^s] \le (E[t])^s.$$

Random-coding exponent

Therefore,

$$
\begin{aligned}
& E[P_{e|m}] \\
\le\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\right] E\left[\left\{\sum_{m'\neq m, 1\le m'\le M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right\}^s\right] \\
\le\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\right] \left(E\left[\sum_{m'\neq m, 1\le m'\le M_n} P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right]\right)^s \\
=\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\right] \left(\sum_{m'\neq m, 1\le m'\le M_n} E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_{m'})\right]\right)^s
\end{aligned}
$$

Since the codewords are selected with identical distribution, the expectations

$$E\left[P_{Y^n|X^n}^{1/(1+s)}(y^n|\boldsymbol{c}_m)\right]$$

should be the same for each m. Hence,

$$\begin{aligned}
&E[P_{e|m}] \\
\le\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_m)\right]\left(\sum_{m'\neq m, 1\le m'\le M_n} E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})\right]\right)^s \\
=\ & \sum_{y^n\in\mathcal{Y}^n} E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_m)\right]\left((M_n-1)E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_m)\right]\right)^s \\
=\ & (M_n-1)^s \sum_{y^n\in\mathcal{Y}^n}\left(E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_m)\right]\right)^{1+s} \\
\le\ & M_n^s \sum_{y^n\in\mathcal{Y}^n}\left(E\left[P^{1/(1+s)}_{Y^n|X^n}(y^n|\boldsymbol{c}_m)\right]\right)^{1+s} \\
=\ & M_n^s \sum_{y^n\in\mathcal{Y}^n}\left(\sum_{x^n\in\mathcal{X}^n} P_{\widetilde{X}^n}(x^n)P^{1/(1+s)}_{Y^n|X^n}(y^n|x^n)\right)^{1+s}
\end{aligned}$$

Since the upper bound of $E[P_{e|m}]$ is no longer dependent on m, $E[P_e]$ can certainly be bounded by the same bound, namely,

$$E[Pe] \le M_n^s \sum_{y^n\in\mathcal{Y}^n}\left(\sum_{x^n\in\mathcal{X}^n} P_{\widetilde{X}^n}(x^n)P^{1/(1+s)}_{Y^n|X^n}(y^n|x^n)\right)^{1+s}$$

By using the fact that $P_{\widetilde{X}^n}$ and $P_{Y^n|X^n}$ are product distributions with identical marginal, and taking logarithmic operation for both sides of the above inequality, we have desired result. Note that $\limsup_{n\to\infty}(1/n)\log_2 M_n = R$. $\square$

The properties of random coding exponent

Definition 9.4 (random-coding exponent) The random coding exponent for DMC with generic distribution $P_{Y|X}$ is defined by

$$E_r(R) \stackrel{\triangle}{=} \max_{0 \le s \le 1} \left[-sR + E_0(s) \right],$$

where

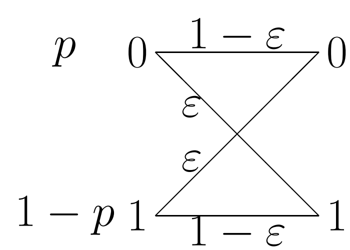
$$E_0(s) \stackrel{\triangle}{=} \max_{P_X} \left[-\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}^{1/(1+s)}(y|x) \right)^{1+s} \right].$$

The properties of $E_r(R)$ can be realized via the analysis of function $E_0(s)$ as follows.

Lemma 9.5 (properties of $E_r(R)$)

1. $E_r(R)$ is non-increasing.
2. $E_r(R)$ is convex in R. (Note that the first two properties imply that $E_r(R)$ is strictly decreasing.)
3. $E_r(C) = 0$, where C is channel capacity, and $E_r(C-\delta) > 0$ for all $0 < \delta < C$.
4. There exists R_{cr} such that for $0 < R < R_{cr}$, the slope of $E_r(R)$ is -1.

The properties of random coding exponent



BSC channel with crossover probability ε and input distribution $(p, 1-p)$.

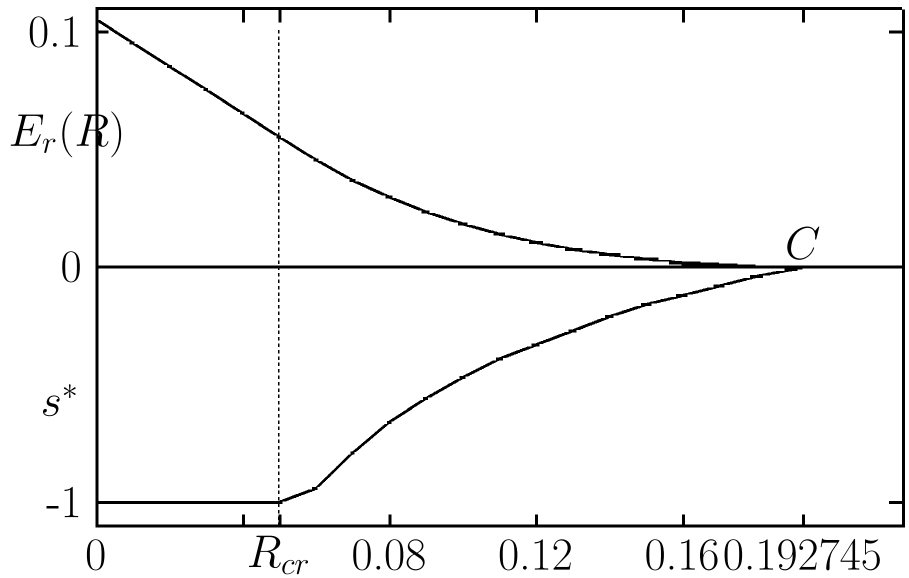
Example 9.6 For BSC with crossover probability ε, the random coding exponent becomes

$$E_r(R) = \max_{0\le p\le 1}\max_{0\le s\le 1}\left\{-sR - \log\left[\left(p(1-\varepsilon)^{1/(1+s)} + (1-p)\varepsilon^{1/(1+s)}\right)^{(1+s)} + \left(p\varepsilon^{1/(1+s)} + (1-p)(1-\varepsilon)^{1/(1+s)}\right)^{(1+s)}\right]\right\},$$

where $(p, 1-p)$ is the input distribution.

Note that the input distribution achieving $E_r(R)$ is uniform, i.e., $p = 1/2$. The random coding exponent for $\varepsilon = 0.2$ is depicted in Figure 9.3.

The properties of random coding exponent



Random coding exponent for BSC with crossover probability 0.2.

Expurgated exponent

- Random coding exponent is derived based on *random coding technique.*
- Since the codewords selection is *unbiased*, both the "good" codes and "bad" codes contribute the same when computing the expectation of error probability.
- Therefore, if, to some extent, we can reduce the contribution of the "bad" codes, a better bound on channel reliability function may be found.

Key 1:

- In stead of randomly selecting M_n codewords, expurgated approach first draws $2M_n$ codewords to form a codebook $\mathscr{C}_{2M_n}$, and sorts these codewords in ascending order in terms of $P_{e|m}(\mathscr{C}_{2M_n})$, which is the error probability given codeword m is transmitted.
- After that, it chooses the first M_n codewords (whose $P_{e|m}(\mathscr{C}_{2M_n})$ is smaller) to form a new codebook, $\mathscr{C}_{M_n}$.
- It can be expected that for $1 \leq m \leq M_n$ and $M_n < m' \leq 2M_n$,

$$P_{e|m}(\mathscr{C}_{M_n}) \leq P_{e|m}(\mathscr{C}_{2M_n}) \leq P_{e|m'}(\mathscr{C}_{2M_n}).$$

  Hence, a better codebook is obtained.

Expurgated exponent

Key 2:

- *Lyapounov's inequality*:

$$E^{1/\alpha}[|X|^\alpha] \le E^{1/\beta}[|X|^\beta], \quad 0 < \alpha \le \beta.$$

- Note that by Lyapounov's inequality,

$$E^{1/\rho}\left[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})\right] \le E\left[P_{e|m}(\mathscr{C}_{2M_n})\right].$$

for $0 < \rho \le 1$.

Expurgated exponent

Lemma 9.7 (existence of "good" code for expurgated exponent) For a sequence of code size M_n satisfying $\limsup_{n\to\infty}(1/n)\log_2 M_n = R$, there exists one block code with size M_n and

$$P_{e|m}(\mathscr{C}_{M_n}) \le 2^{1/\rho} E^{1/\rho}\left[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})\right]. \tag{9.2.1}$$

Proof: Randomly draw $2M_n$ codewords according to some input distribution $P_{\tilde{X}^n}$, and let the codebook be denoted by

$$\mathscr{C}_{2M_n} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_{2M_n}\}.$$

Let $\phi(\cdot)$ be the indicator function of the set

$$\{t \in \Re \ : \ t < 2^{1/\rho} E^{1/\rho}[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})]\},$$

for some $\rho > 0$. Hence, $\phi(\,P_{e|m}(\mathscr{C}_{2M_n})\,) = 1$ if

$$P_{e|m}(\mathscr{C}_{2M_n}) < 2^{1/\rho} E^{1/\rho}[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})].$$

Expurgated exponent

By Markov's inequality,

$$
\begin{aligned}
& E\left[\sum_{1\le m\le 2M_n} \phi(P_{e|m}(\mathscr{C}_{2M_n}))\right] \\
= & \sum_{1\le m\le 2M_n} E\left[\phi(P_{e|m}(\mathscr{C}_{2M_n}))\right] \\
= & \; 2M_n E\left[\phi(P_{e|m}(\mathscr{C}_{2M_n}))\right] \\
= & \; 2M_n Pr\left\{P_{e|m}(\mathscr{C}_{2M_n}) < 2^{1/\rho} E^{1/\rho}[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})]\right\} \\
= & \; 2M_n Pr\left\{P^{\rho}_{e|m}(\mathscr{C}_{2M_n}) < 2E[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})]\right\} \\
\ge & \; 2M_n\left(1 - \frac{E[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})]}{2E[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})]}\right) \\
= & \; M_n.
\end{aligned}
$$

Therefore, there exist at lease one codebook such that

$$\sum_{1\le m\le 2M_n} \phi(P_{e|m}(\mathscr{C}_{2M_n})) \ge M_n.$$

Hence, by selecting M_n codewords from this codebook with $\phi(P_{e|m}(\mathscr{C}_{2M_n})) = 1$, a new codebook is formed, and it is obvious that (9.2.1) holds for this new codebook.

Expurgated exponent

$\square$

Expurgated exponent

Definition 9.8 (expurgated exponent) The expurgated exponent for DMC with generic distribution $P_{Y|X}$ is defined by

$$E_{ex}(R) \stackrel{\triangle}{=} \max_{s'\geq 1} \max_{P_X} \left[-s'R - s' \log \sum_{x\in\mathcal{X}} \sum_{x'\in\mathcal{X}} P_X(x)P_X(x') \left(\sum_{y\in\mathcal{Y}} \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|x')} \right)^{1/s'} \right].$$

Theorem 9.9 (expurgated exponent) For DMC with generic transition probability $P_{Y|X}$,

$$E(R) \geq E_{ex}(R).$$

Proof: Randomly select $2M_n$ codewords according to some input distribution $P_{\tilde{X}}$.

Step 1: Maximum likelihood decoder. Let $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{2M_n}\} \in \mathcal{X}^n$ denote the set of n-tuple block codewords selected, and let the decoding partition for symbol m (namely, the set of channel outputs that classify to m) be

$$\mathcal{U}_m = \{y^n \;:\; P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) > P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'}), \quad \text{for all } m' \neq m\}.$$

Those channel outputs that are on the boundary, i.e.,

$$\text{for some } m \text{ and } m', \quad P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) = P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'}),$$

will be arbitrarily assigned to either m or m'.

Step 2: Property of indicator function for $s = 1$. Let ϕ_m be the indicator function of $\mathcal{U}_m$. Then for all $s > 0$.

$$1 - \phi_m(y^n) \le \left\{ \sum_{m' \ne m} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/(1+s)} \right\}^s.$$

(Note that this step is the same as random coding exponent, except only $s = 1$ is considered.) By taking $s = 1$, we have

$$1 - \phi_m(y^n) \le \left\{ \sum_{m' \ne m} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/2} \right\}.$$

Step 3: Probability of error given codeword $\boldsymbol{c}_m$ is transmitted. Let $P_{e|m}(\mathscr{C}_{2M_n})$ denote the probability of error given codeword $\boldsymbol{c}_m$ is transmitted.

Then

$$
\begin{aligned}
& P_{e|m}(\mathscr{C}_{2M_n}) \\
\le\ & \sum_{y^n \notin \mathcal{U}_m} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \\
=\ & \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)[1-\phi_m(y^n)] \\
\le\ & \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \left\{ \sum_{m' \neq m, 1 \le m' \le 2M_n} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/2} \right\} \\
\le\ & \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \left\{ \sum_{1 \le m' \le 2M_n} \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/2} \right\} \\
=\ & \sum_{1 \le m' \le 2M_n} \left\{ \sum_{y^n \in \mathcal{Y}^n} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \left[\frac{P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)} \right]^{1/2} \right\} \\
=\ & \sum_{1 \le m' \le 2M_n} \left\{ \sum_{y^n \in \mathcal{Y}^n} \sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})} \right\}
\end{aligned}
$$

Step 4: Standard inequality for $s' = 1/\rho \ge 1$. It is known that for any

Expurgated exponent

$0 < \rho = 1/s' \le 1$, we have

$$\left(\sum_i a_i\right)^\rho \le \sum_i a_i^\rho,$$

for all non-negative sequence a_i.[1]

Hence,

$$\begin{aligned} P^\rho_{e|m}(\mathscr{C}_{2M_n}) &\le \left(\sum_{1\le m'\le 2M_n}\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}\right)^\rho \\ &\le \sum_{1\le m'\le 2M_n}\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^\rho \end{aligned}$$

[1] *Proof:* Let $f(\rho) = (\sum_i a_i^\rho)/\left(\sum_j a_j\right)^\rho$. Then we need to show $f(\rho) \ge 1$ when $0 < \rho \le 1$. Let $p_i = a_i/(\sum_k a_k)$, and hence, $a_i = p_i(\sum_k a_k)$. Take it to the numerator of $f(\rho)$:

$$\begin{aligned} f(\rho) &= \frac{\sum_i p_i^\rho(\sum_k a_k)^\rho}{(\sum_j a_j)^\rho} \\ &= \sum_i p_i^\rho \end{aligned}$$

Now since $\partial f(\rho)/\partial\rho = \sum_i \log(p_i)p_i^\rho \le 0$ (which implies $f(\rho)$ is non-increasing in ρ) and $f(1) = 1$, we have the desired result.

Expurgated exponent

Step 5: Expectation of $P_{e|m}^{\rho}(\mathscr{C}_{2M_n})$.

$$
\begin{aligned}
& E[P_{e|m}^{\rho}(\mathscr{C}_{2M_n})] \\
\le\ & E\left[\sum_{1\le m'\le 2M_n}\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{\rho}\right] \\
=\ & \sum_{1\le m'\le 2M_n} E\left[\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{\rho}\right] \\
=\ & 2M_n E\left[\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{\rho}\right].
\end{aligned}
$$

Step 6: Lemma 9.7. From Lemma 9.7, there exists one codebook with size M_n

Expurgated exponent

such that

$$
\begin{aligned}
& P_{e|m}(\mathscr{C}_{M_n}) \\
\le\ & 2^{1/\rho} E^{1/\rho}[P^{\rho}_{e|m}(\mathscr{C}_{2M_n})] \\
\le\ & 2^{1/\rho}\left(2M_n E\left[\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{\rho}\right]\right)^{1/\rho} \\
=\ & (4M_n)^{1/\rho}E^{1/\rho}\left[\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{\rho}\right] \\
=\ & (4M_n)^{s'}E^{s'}\left[\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|\boldsymbol{c}_m)P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'})}\right\}^{1/s'}\right] \\
=\ & (4M_n)^{s'}\left(\sum_{x^n\in\mathcal{X}^n}\sum_{(x')^n\in\mathcal{X}^n}P_{\widetilde{X}^n}(x^n)P_{\widetilde{X}^n}((x')^n)\right. \\
& \left.\times\left\{\sum_{y^n\in\mathcal{Y}^n}\sqrt{P_{Y^n|X^n}(y^n|x^n)P_{Y^n|X^n}(y^n|(x')^n)}\right\}^{1/s'}\right)^{s'}
\end{aligned}
$$

Expurgated exponent

By using the fact that $P_{\widetilde{X}^n}$ and $P_{Y^n|X^n}$ are product distributions with identical marginal, and taking logarithmic operation for both sides of the above inequality, we have desired result. $\square$

The properties of expurgated exponent

Definition 9.10 (expurgated exponent) The expurgated exponent for DMC with generic distribution $P_{Y|X}$ is defined by

$$E_{ex}(R) \stackrel{\triangle}{=} \max_{s \geq 1} \left[-sR + E_x(s) \right],$$

where

$$E_x(s) \stackrel{\triangle}{=} \max_{P_X} \left[-s \log \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} P_X(x) P_X(x') \left(\sum_{y \in \mathcal{Y}} \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|x')} \right)^{1/s} \right].$$

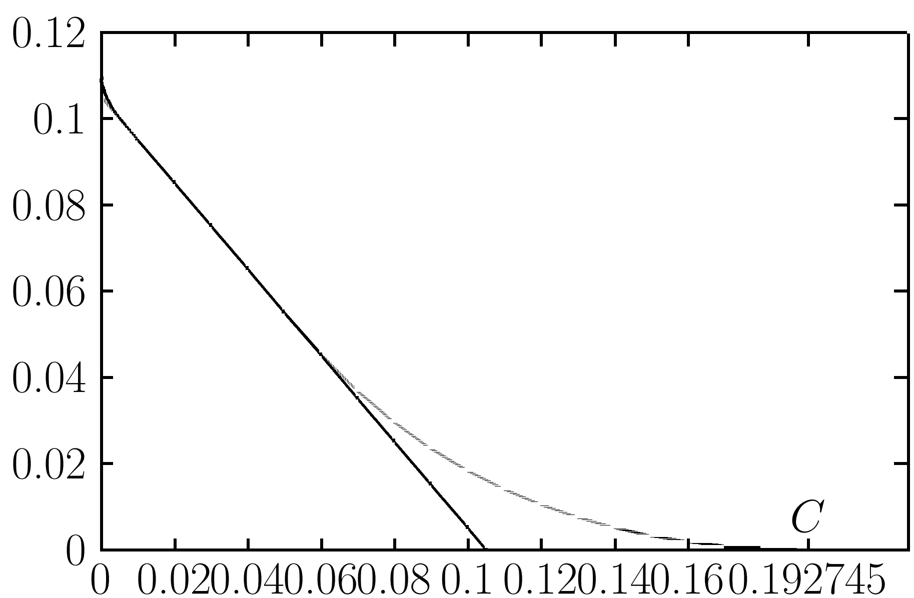
Lemma 9.11 (properties of $E_{ex}(R)$)

1. *$E_{ex}(R)$ is non-increasing.*
2. *$E_{ex}(R)$ is convex in R. (Note that the first two properties imply that $E_{ex}(R)$ is strictly decreasing.)*
3. *There exists R_{cr} such that for $R > R_{cr}$, the slope of $E_r(R)$ is -1.*

Example 9.12 For BSC with crossover probability ε, the expurgated exponent becomes

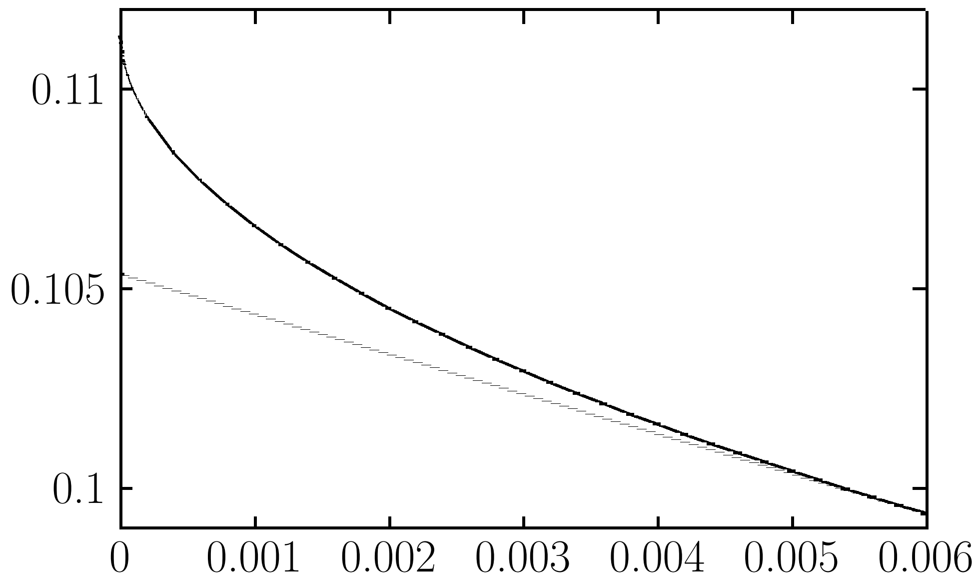
$$E_{ex}(R) = \max_{1\geq p\geq 0}\max_{s\geq 1}\left\{-sR - s\log\left[p^2 + 2p(1-p)\left(2\sqrt{\varepsilon(1-\varepsilon)}\right)^{1/s} + (1-p)^2\right]\right\}$$

where $(p, 1-p)$ is the input distribution. Note that the input distribution achieving $E_{ex}(R)$ is uniform, i.e., $p = 1/2$. The expurgated exponent, as well as random coding exponent, for $\varepsilon = 0.2$ is depicted in Figures 9.4 and 9.5.



Expurgated exponent (solid line) and random coding exponent (dashed line) for BSC with crossover probability 0.2 (over the range of $(0, 0.192745)$).

The properties of expurgated exponent



Expurgated exponent (solid line) and random coding exponent (dashed line) for BSC with crossover probability 0.2 (over the range of $(0, 0.006)$).

Partitioning upper bound for channel reliability

Upper bounds:

- *partitioning* bound
- *sphere-packing* bound
- *straight-line* bound

Keys:

- Hypothesis testing
- Model: the receiver end can be modeled as:

$$\begin{aligned} H_0 &: \boldsymbol{c}_m = \text{codeword transmitted} \\ H_1 &: \boldsymbol{c}_m \neq \text{codeword transmitted}, \end{aligned}$$

  where m is the final decision made by receiver upon receipt of some channel outputs.

- The channel decoding error given that codeword m is transmitted becomes the type II error, which can be computed using the theory of binary hypothesis testing.

Partitioning upper bound for channel reliability

Definition 9.13 (partitioning bound) For a DMC with marginal $P_{Y|X}$,

$$E_p(R) \stackrel{\triangle}{=} \max_{P_X} \min_{\{P_{\tilde{Y}|X} \,:\, I(P_X, P_{\tilde{Y}|X}) \le R\}} D(P_{\tilde{Y}|X} \| P_{Y|X} | P_X).$$

Observation 9.14 (partitioning bound) If $P_{\tilde{X}}$ and $P_{\tilde{Y}|X}$ are distributions that achieves $E_p(R)$, and $P_{\tilde{Y}}(y) = \sum_{x \in \mathcal{X}} P_{\tilde{X}}(x) P_{\tilde{Y}|X}(y|x)$, then

$$E_p(R) \stackrel{\triangle}{=} \max_{P_X} \min_{\{P_{\hat{Y}|X} \,:\, D(P_{\hat{Y}|X} \| P_{\tilde{Y}} | P_X) \le R\}} D(P_{\hat{Y}|X} \| P_{Y|X} | P_X).$$

In addition, the distribution $P_{\hat{Y}|X}$ that achieves $E_p(R)$ is a tilted distribution between $P_{Y|X}$ and $P_{\tilde{Y}}$, i.e.,

$$E_p(R) \stackrel{\triangle}{=} \max_{P_X} D(P_{Y_\lambda|X} \| P_{Y|X} | P_X),$$

where $P_{Y_\lambda|X}$ is a tilted distribution[2] between $P_{Y|X}$ and $P_{\tilde{Y}}$, and λ is the solution of

$$D(P_{Y_\lambda} \| P_{\tilde{Y}} | P_X) = R.$$

[2] $$P_{Y_\lambda|X}(y|x) \stackrel{\triangle}{=} \frac{P_{\tilde{Y}}^{\lambda}(y) P_{Y|X}^{1-\lambda}(y|x)}{\sum_{y' \in \mathcal{Y}} P_{\tilde{Y}}^{\lambda}(y') P_{Y|X}^{1-\lambda}(y'|x)}.$$

Partitioning upper bound for channel reliability

Theorem 9.15 (partitioning bound) For a DMC with marginal $P_{Y|X}$, for any $\varepsilon > 0$ arbitrarily small,

$$E(R+\varepsilon) \le E_p(R).$$

Lemma 9.16

$$\begin{aligned} E_p(R) &= \max_{P_X} \max_{s\ge 0} \left[-sR - \log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x)^{1/(1+s)} \right)^{1+s} \right] \\ &= \max_{s\ge 0} \left[-sR - E_0(s) \right]. \end{aligned}$$

- Recall: the random coding exponent is

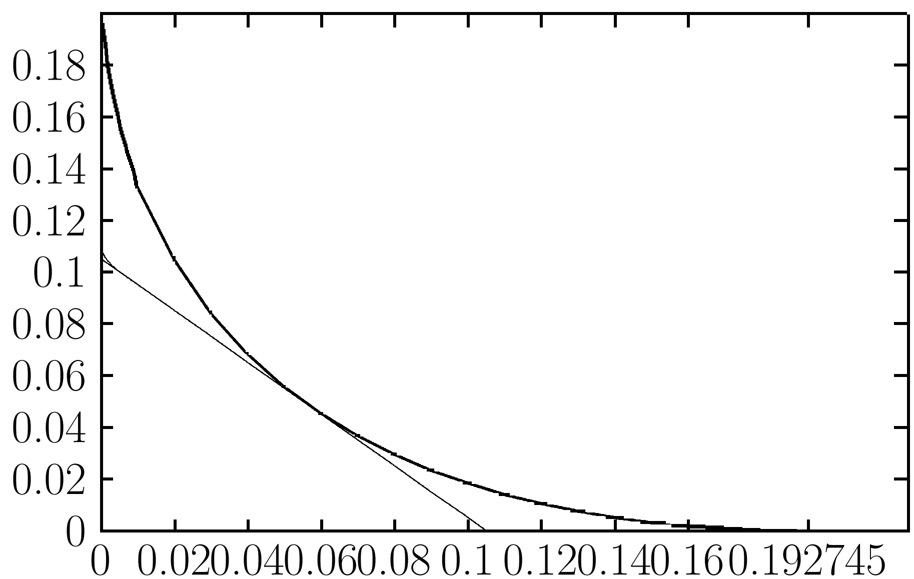
$$\max_{0<s\le 1} [-sR - E_0(s)].$$

- the optimizer s^* is the slope of $E_r(R)$ times -1.
- Hence, the channel reliability $E(R)$ satisfies

$$\max_{0<s\le 1} [-sR - E_0(s)] \le E(R) \le \max_{s\ge 0} [-sR - E_0(s)],$$

  and for optimizer $s^* \in (0,1]$ (i.e., $R_{cr} \le R \le C$), the upper bound meets the lower bound.

Partitioning upper bound for channel reliability



Partitioning exponent (thick line), random coding exponent (thin line) and expurgated exponent (thin line) for BSC with crossover probability 0.2.

Sphere-packing upper bound

- **Ball or sphere**:
  - For a given space $\mathcal{A}$
  - and a given distance measures $d(\cdot,\cdot)$ for elements in $\mathcal{A}$,
  - a *ball* or *sphere* centered at a with radius r is defined as

$$\{b \in \mathcal{A} \;:\; d(b,a) \le r\}.$$

- **The problem of sphere-packing:**
  - to find the minimum radius if M spheres need to be packed into the space $\mathcal{A}$.
  - Its dual problem is to find the maximum number of balls with fixed radius r that can be packed into space $\mathcal{A}$.

Relation of sphere-packing and coding

- To find the best codebook which yields minimum decoding error is one of the main research issues in communications.
- Roughly speaking, if two codewords are *similar*, they should be more vulnerable to noise.
- Hence, a good codebook should be a set of codewords, which look very *different* from others.
- In mathematics, such "codeword resemblance" can be modeled as a distance function.
- We can then say if the distance between two codewords is large, they are more "different", and more robust to noise or interference.
- Accordingly, a good code book becomes a set of codewords whose minimum distance among codewords is largest.
- This is exactly the sphere-packing problem.

Example 9.17 (Hamming distance versus BSC) For BSC, the source alphabet and output alphabet are both $\{0,1\}^n$. The Hamming distance between two elements in $\{0,1\}$ is given by

$$d_H(x,\hat{x}) = \begin{cases} 0, & \text{if } x = \hat{x}; \\ 1, & \text{if } x \neq \hat{x}. \end{cases}$$

Its extension definition to n-tuple is

$$d_H(x^n,\hat{x}^n) = \sum_{i=1}^{n} d_H(x_i,\hat{x}_i).$$

It is known that the best decoding rule is the maximum likelihood ratio decoder, i.e.,

$$\phi(y^n) = m, \text{ if } P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) \geq P_{Y^n|X^n}(y^n|\boldsymbol{c}_{m'}) \text{ for all } m'.$$

Since for BSC with crossover probability ε,

$$\begin{aligned} P_{Y^n|X^n}(y^n|\boldsymbol{c}_m) &= \varepsilon^{d_H(y^n,\boldsymbol{c}_m)}(1-\varepsilon)^{n-d_H(y^n,\boldsymbol{c}_m)} \\ &= (1-\varepsilon)^n \left(\frac{\varepsilon}{1-\varepsilon}\right)^{d_H(y^n,\boldsymbol{c}_m)}. \end{aligned} \tag{9.4.2}$$

Therefore, the best decoding rule can be re-written as:

$$\phi(y^n) = m, \text{ if } d_H(y^n,\boldsymbol{c}_m) \geq d_H(y^n,\boldsymbol{c}_{m'}) \text{ for all } m'.$$

Relation of sphere-packing and coding

As a result, if two codewords are too close in Hamming distance, the number of bits of outputs that can be used to classify its origin will be less, and therefore, will result in a poor performance.

From the above example, two observations can be made.

- First, if the distance measure between codewords can be written as a function of the transition probability of channel, such as (9.4.2), one may regard the probability of decoding error with the distances between codewords.
- Secondly, the coding problem in some cases can be reduced to a sphere-packing problem.
- As a consequence, solution of the sphere-packing problem can be used to characterize the error probability of channels.

Theorem 9.18 Let $\mu_n(\cdot,\cdot)$ be the Bhattacharya distance[3] between two elements in $\mathcal{X}^n$. Denote by $d_{n,M}$ the largest minimum distance among M selected codewords of length n. (Obviously, the largest minimum radius among M disjoint spheres in a code space is half of $d_{n,M}$.) Then

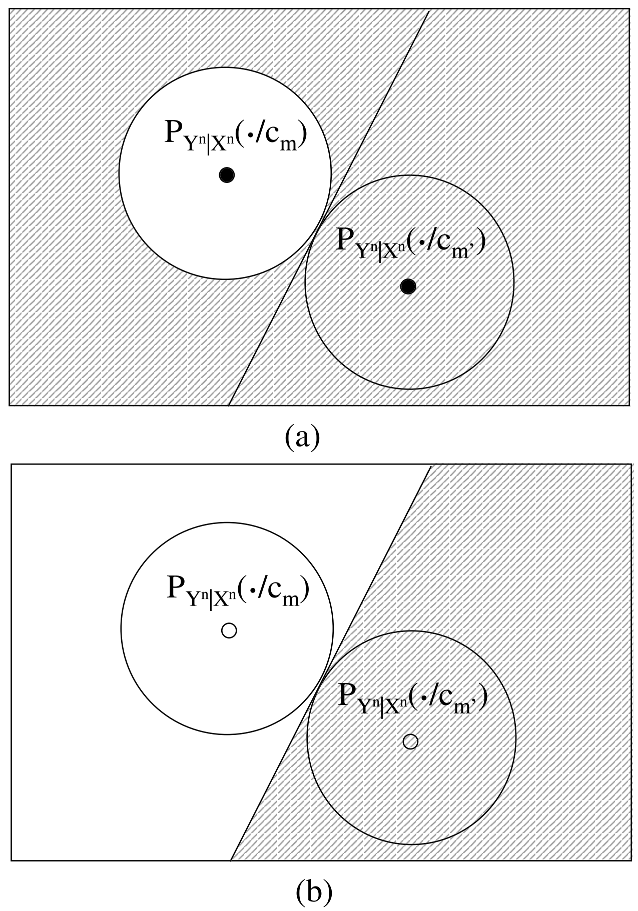
$$\limsup_{n\to\infty} -\frac{1}{n}\log Pe(n,R) \le \limsup_{n\to\infty} \frac{1}{n} d_{n,M=e^{nR}}.$$

Since, according the above theorem, the largest minimum distance can be used to formulate an upper bound on channel reliability, this quantity becomes essential. We therefore introduce its general formula in the next subsection.

[3]The Bhattacharya distance (for channels $P_{Y^n|X^n}$) between elements x^n and $\hat{x}^n$ is defined by

$$\mu_n(x^n,\hat{x}^n) \stackrel{\triangle}{=} -\log \sum_{y^n\in\mathcal{Y}^n} \sqrt{P_{Y^n|X^n}(y^n|x^n)P_{Y^n|X^n}(y^n|\hat{x}^n)}.$$

Relation of sphere-packing and coding



(a) The shaded area is $\mathcal{U}_m^c$; (b) The shaded area is $\mathcal{A}_{m,m'}^c$.

The largest minimum distance of block codes

- History: Varshamov-Gilbert lower bound, Elias bound, McEliece bound
- If the size of the code alphabet, q, is an even power of a prime, satisfying $q \geq 49$, and the distance measure is the Hamming distance, a better lower bound can be obtained through the construction of the Algebraic-Geometric code, of which the idea was first proposed by Goppa.
- Later, Zinoviev and Litsyn proved that a better lower bound than the Varshamov-Gilbert bound is actually possible for any $q \geq 46$.

Distance-spectrum formula

Notations

- The n-tuple code alphabet is denoted by $\mathcal{X}^n$.
- $\mu_n(\hat{x}^n, x^n)$ denotes the distance.
- A codebook with block length n and size M is represented by

$$\mathscr{C}_{n,M} \stackrel{\triangle}{=} \left\{ \boldsymbol{c}_0^{(n)}, \boldsymbol{c}_1^{(n)}, \boldsymbol{c}_2^{(n)}, \ldots, \boldsymbol{c}_{M-1}^{(n)} \right\},$$

  where $\boldsymbol{c}_m^{(n)} \stackrel{\triangle}{=} (c_{m1}, c_{m2}, \ldots, c_{mn})$, and each c_{mk} belongs to $\mathcal{X}$.

- The minimum distance is

$$d_m(\mathscr{C}_{n,M}) \stackrel{\triangle}{=} \min_{\substack{0 \le \hat{m} \le M-1 \\ \hat{m} \ne m}} \mu_n\left(\boldsymbol{c}_{\hat{m}}^{(n)}, \boldsymbol{c}_m^{(n)} \right).$$

- The largest minimum distance is

$$d_{n,M} \stackrel{\triangle}{=} \max_{\mathscr{C}_{n,M}} \min_{0 \le m \le M-1} d_m(\mathscr{C}_{n,M}).$$

Distance-spectrum formula

- **Problem:**
$$\lim_{n\to\infty}\frac{1}{n}d_{n,M=e^{nR}}$$
for fixed R.

- **Key:** Random coding.

Theorem 9.19 (distance-spectrum formula)

$$\sup_{\boldsymbol{X}}\bar{\Lambda}_{\boldsymbol{X}}(R)\ge\limsup_{n\to\infty}\frac{d_{n,M}}{n}\ge\sup_{\boldsymbol{X}}\bar{\Lambda}_{\boldsymbol{X}}(R+\delta) \tag{9.4.3}$$

and

$$\sup_{\boldsymbol{X}}\underline{\Lambda}_{\boldsymbol{X}}(R)\ge\liminf_{n\to\infty}\frac{d_{n,M}}{n}\ge\sup_{\boldsymbol{X}}\underline{\Lambda}_{\boldsymbol{X}}(R+\delta) \tag{9.4.4}$$

for every $\delta>0$, where

$$\bar{\Lambda}_{\boldsymbol{X}}(R)\ \triangleq\ \inf\left\{a\in\Re\ :\ \limsup_{n\to\infty}\left(Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)>a\right\}\right)^{\exp\{nR\}}\right\}$$

and

$$\underline{\Lambda}_{\boldsymbol{X}}(R)\ \triangleq\ \inf\left\{a\in\Re\ :\ \liminf_{n\to\infty}\left(Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)>a\right\}\right)^{\exp\{nR\}}\right\}.$$

Distance-spectrum formula

$$E(R) \le \sup_{\boldsymbol{X}} \inf \left\{ a : \limsup_{n\to\infty} \left(Pr\left[-\frac{1}{n}\log \sum_{y^n\in\mathcal{Y}^n} P^{1/2}_{Y^n|X^n}(y^n|\hat{X}^n) P^{1/2}_{Y^n|X^n}(y^n|X^n) > a \right] \right)^{\exp\{nR\}} \right\}.$$

Properties of distance-spectrum function

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \begin{cases} = \infty & \text{for } 0 < R < \bar{R}_p(\boldsymbol{X}); \\ = \bar{D}_p(\boldsymbol{X}) & \text{at } R = R_p(\boldsymbol{X}); \\ \in (\bar{D}_0(\boldsymbol{X}), \bar{D}_p(\boldsymbol{X})] & \text{for } R_p(\boldsymbol{X}) < R < R_0(\boldsymbol{X}); \\ = \bar{D}_0(\boldsymbol{X}) & \text{for } R \geq R_0(\boldsymbol{X}). \end{cases}$$

where

$$\bar{D}_0(\boldsymbol{X}) \stackrel{\triangle}{=} \limsup_{n\to\infty} \ \text{ess inf} \frac{1}{n}\mu_n(\hat{X}^n, X^n)$$

$$\bar{R}_0(\boldsymbol{X}) \stackrel{\triangle}{=} \limsup_{n\to\infty} -\frac{1}{n}\log Pr\{\hat{X}^n = X^n\}$$

$$\bar{D}_p(\boldsymbol{X}) \stackrel{\triangle}{=} \limsup_{n\to\infty} \frac{1}{n}E[\mu_n(\hat{X}^n, X^n) | \mu_n(\hat{X}^n, X^n) < \infty]$$

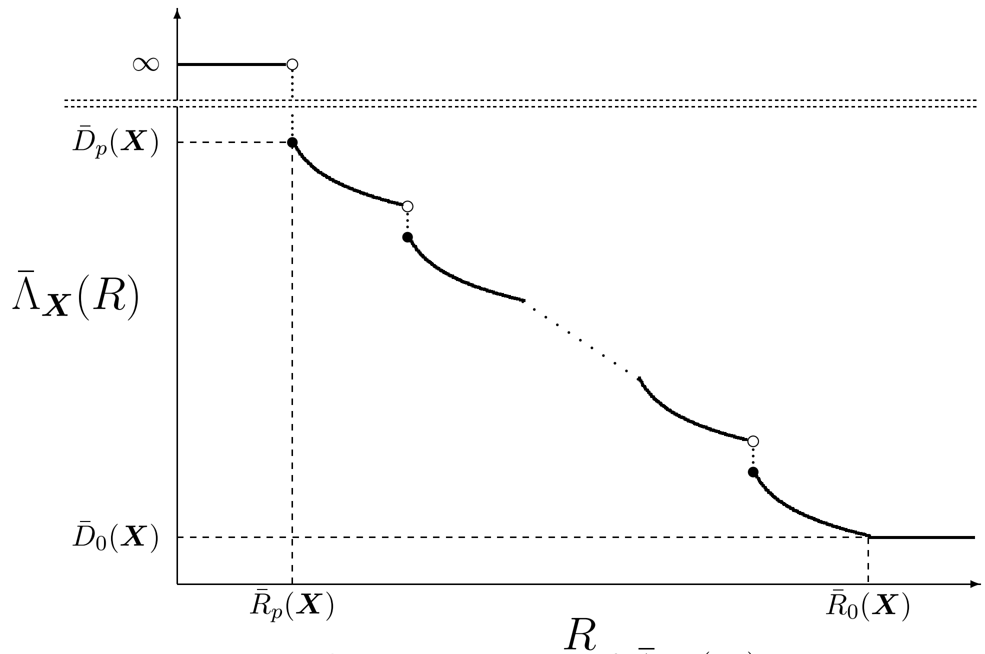
and

$$\bar{R}_p(\boldsymbol{X}) \stackrel{\triangle}{=} \limsup_{n\to\infty} -\frac{1}{n}\log Pr\{\mu_n(\hat{X}^n, X^n) < \infty\}.$$

- For a given random variable Z, its *essential infimum* is defined as

$$\text{ess inf } Z \stackrel{\triangle}{=} \sup\{z \ : \ Pr[Z \geq z] = 1\}.$$

Properties of distance-spectrum function



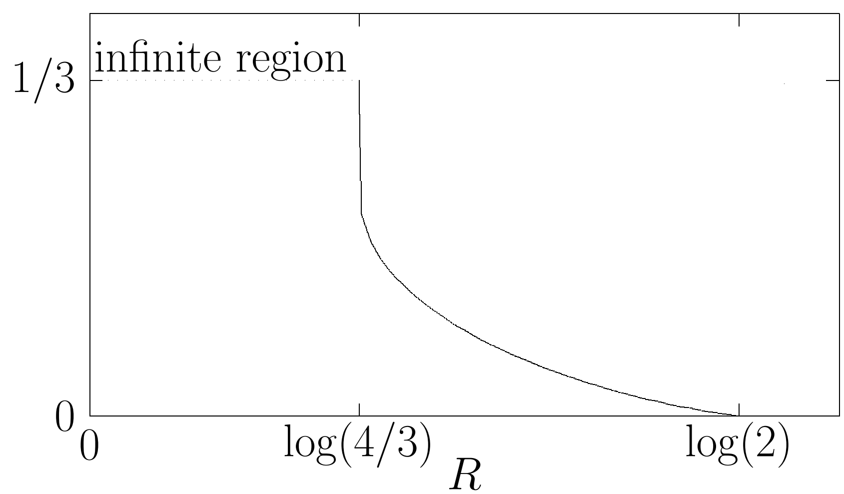
General curve of $\bar{\Lambda}_{\boldsymbol{X}}(R)$.

Properties of distance-spectrum function

Example 9.20 • $\mathcal{X} = \{0, 1\}$;

- $\mu_n(\cdot, \cdot)$ is additive with marginal distance metric $\mu(0,0) = \mu(1,1) = 0$, $\mu(0,1) = 1$ and $\mu(1,0) = \infty$;
-
$$\bar{\Lambda}_{\boldsymbol{X}}(R) = \sup_{s>0} \left\{ -sR - s \cdot \log \frac{2 + e^{-1/s}}{4} \right\}.$$
-
$$\begin{aligned}
\bar{R}_0(\boldsymbol{X}) &= -\log Pr\{\hat{X} = X\} = \log 2; \\
\bar{D}_0(\boldsymbol{X}) &= \text{ess inf} \mu(\hat{X}, X) = 0; \\
\bar{R}_p(\boldsymbol{X}) &= -\log Pr\{\mu(\hat{X}, X) < \infty\} = \log \frac{4}{3}; \\
\bar{D}_p(\boldsymbol{X}) &= E[\mu(\hat{X}, X) | \mu(\hat{X}, X) < \infty] = \frac{1}{3}.
\end{aligned}$$

Properties of distance-spectrum function



Function of $\sup_{s>0}\left\{-sR - s\log\left[\left(2+e^{-1/s}\right)/4\right]\right\}$.

- **Open questions:** *under what conditions is the Varshamov-Gilbert lower bound tight?*

Lemma 9.21 (large deviation formulas for $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and $\underline{\Lambda}_{\boldsymbol{X}}(R)$)

$$\bar{\Lambda}_{\boldsymbol{X}}(R) = \inf\left\{a \in \Re \;:\; \bar{\ell}_{\boldsymbol{X}}(a) < R\right\}$$

and

$$\underline{\Lambda}_{\boldsymbol{X}}(R) = \inf\left\{a \in \Re \;:\; \underline{\ell}_{\boldsymbol{X}}(a) < R\right\}$$

where $\bar{\ell}_{\boldsymbol{X}}(a)$ and $\underline{\ell}_{\boldsymbol{X}}(a)$ are respectively the sup- and the inf-large deviation spectrums of $(1/n)\mu_n(\hat{X}^n, X^n)$, defined as

$$\bar{\ell}_{\boldsymbol{X}}(a) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) \le a\right\}$$

and

$$\underline{\ell}_{\boldsymbol{X}}(a) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) \le a\right\}.$$

Corollary 9.22 (Varshamov-Gilbert bound)

$$\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \geq \sup_{\boldsymbol{X}} \bar{G}_{\boldsymbol{X}}(R) \quad \text{and} \quad \sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R) \geq \sup_{\boldsymbol{X}} \underline{G}_{\boldsymbol{X}}(R)$$

where

$$\bar{G}_{\boldsymbol{X}}(R) \triangleq \sup_{s>0} \left[-sR - s \cdot \underline{\varphi}_{\boldsymbol{X}}(-1/s) \right] \tag{9.4.5}$$

$$\underline{G}_{\boldsymbol{X}}(R) \triangleq \sup_{s>0} \left[-sR - s \cdot \bar{\varphi}_{\boldsymbol{X}}(-1/s) \right]. \tag{9.4.6}$$

$$\underline{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta \cdot \mu_n(\hat{X}^n, X^n)} \right]$$

and

$$\bar{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \limsup_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta \cdot \mu_n(\hat{X}^n, X^n)} \right].$$

Straight line bound

Key:

- It is conjectured that the channel reliability function is convex.
- Therefore, any non-convex upper bound can be improved by making it convex.
- This is the main idea of the *straight line bound*.

Definition 9.23 (list decoder) A *list decoder* decodes the outputs of a noisy channel by a list of candidates (possible inputs), and an error occurs only when the correct codeword transmitted is not in the list.

Definition 9.24 (maximal error probability for list decoder)

$$P_{e,max}(n, M, L) \triangleq \min_{\{\mathcal{C} \text{ with } L \text{ candidates for decoder}\}} \max_{1 \le m \le M} P_{e|m},$$

where n is the blocklength and M is the code size.

Definition 9.25 (average error probability for list decoder)

$$P_e(n, M, L) \triangleq \min_{\{\mathcal{C} \text{ with } L \text{ candidates for decoder}\}} \frac{1}{M} \sum_{1 \le m \le M} P_{e|m},$$

where n is the blocklength and M is the code size.

Straight line bound

Lemma 9.26 (lower bound on average error) For DMC,

$$P_e(n, M, 1) \geq P_e(n_1, M, L) P_{e,max}(n_2, L + 1, 1),$$

where $n = n_1 + n_2$.

Theorem 9.27 (straight-line bound)

$$E(\lambda R_1 + (1 - \lambda) R_2) \leq \lambda E_p(R_1) + (1 - \lambda) E_{sp}(R_2).$$

Proof: By applying the previous lemma with $\lambda = n_1/n$, $\log(M/L)/n_1 = R_1$ and $\log L/n_2 = R_2$, we can upper bound $P_e(n_1, M, L)$ by partitioning exponent, and $P_{e,max}(n_2, L + 1, 1)$ by sphere-packing exponent. $\square$