Chapter 10

Information Theory of Networks

Po-Ning Chen

Department of Communications Engineering

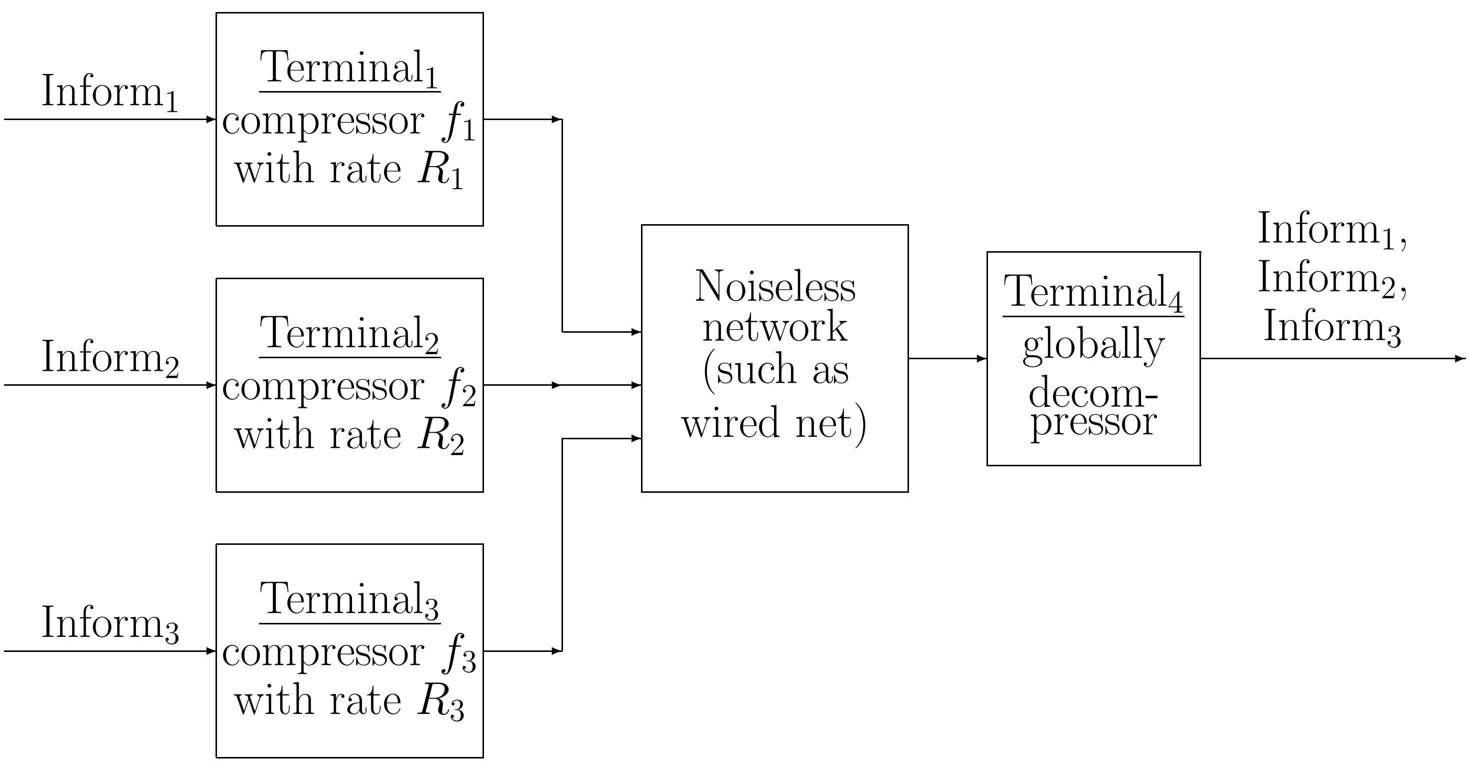
National Chiao-Tung University

Hsin Chu, Taiwan 30050

Backgrounds

- Theory regarding to the communications among many (more than three) terminals.
- This is usually named *network*.

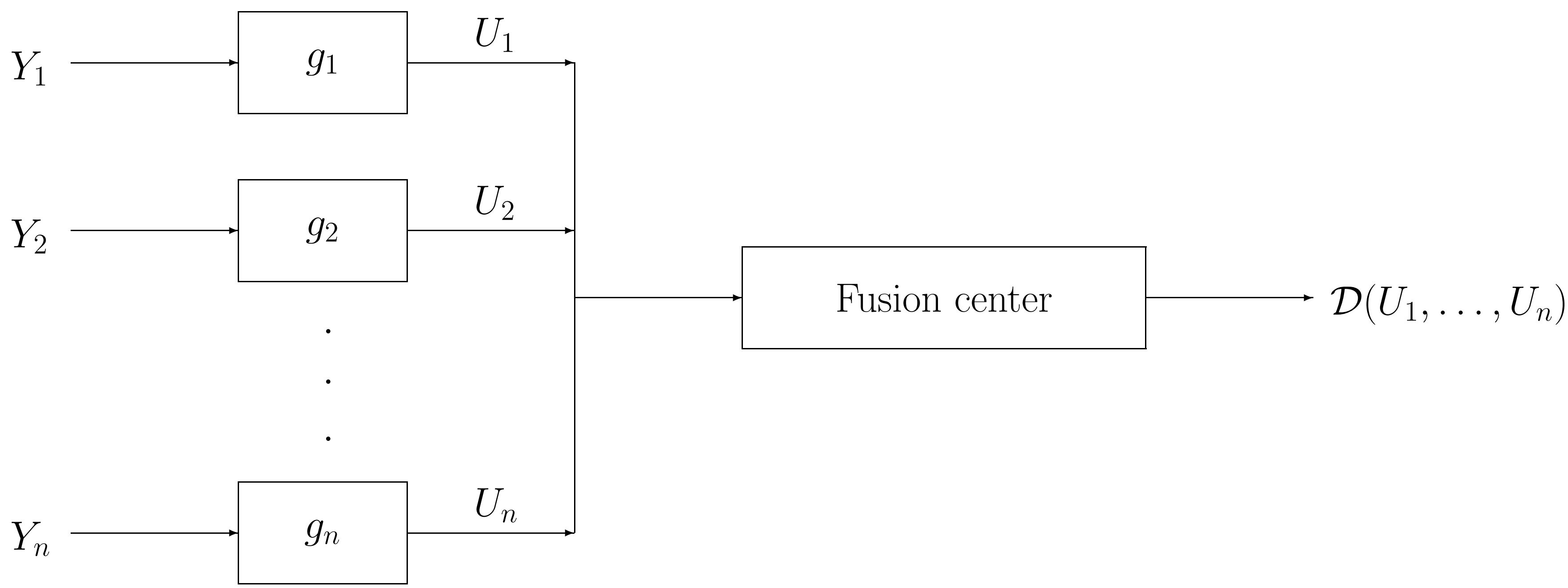
Example 10.1 (multi-access channels)



Backgrounds

Example 10.2 (broadcast channel)

Example 10.3 (distributed detection)



Distributed detection with n senders. Each observations Y_i may come from one of two categories. The final decision $\mathcal{D} \in \{H_0, H_1\}$.

Example 10.4 (some other examples)

1) *Relay channel.* This channel consists of one source sender, one destination receiver, and several intermediate sender-receiver pairs that act as relays to facilitate the communication between the source sender and destination receiver.

2) *Interference channel.* Several senders and several receivers communicate simultaneously on common channel, where interference among them could introduce degradation on performance.

3) *Two-way communication channel.* Instead of conventional one-way channel, two terminals can communicate in a two-way fashion (full duplex).

Lossless data compression over distributed sources for block codes

Definition 10.5 (independent encoders among distributed sources)

- There are several sources

$$X_1, X_2, \ldots, X_m$$

  (may or may not be independent) which are respectively obtained by m terminals.

- Before each terminal transmits its local source to the receiver, a block encoder f_i with rate R_1 and block length n is applied

$$f_i \;:\; \mathcal{X}_i^n \to \{1, 2, \ldots, 2^{nR_i}\}.$$

- It is assumed that there is no conspiracy among block encoders.

Definition 10.6 (global decoder for independently compressed sources)
A global decoder $g(\cdot)$ will recover the original sources after receiving all the independently compressed sources, i.e.,

$$g \;:\; \{1, \ldots, 2^{R_1}\} \times \cdots \times \{1, \ldots, 2^{R_m}\} \to \mathcal{X}_1^n \times \cdots \times \mathcal{X}_m^n.$$

Lossless data compression over distributed sources for block codes

Definition 10.7 (probability of error) The probability of error is defined as

$$P_e(n) \triangleq Pr\{g(f_1(X_1^n), \ldots, f_m(X_m^n)) \neq (X_1^n, \ldots, X_m^n)\}.$$

Definition 10.8 (achievable rates) A rates $(R_1, \ldots, R_m)$ is said to be achievable if there exists a sequence of block codes such that

$$\limsup_{n\to\infty} P_e(n) = 0.$$

Definition 10.9 (achievable rate region for distributed sources) The achievable rate region for distributed sources is the set of all achievable rates.

Observation 10.10 The achievable rate region is convex.

Theorem 10.11 (Slepian-Wolf) For distributed sources consisting of two random variables X_1 and X_2, the achievable region is

$$\begin{aligned} R_1 &\geq H(X_1|X_2) \\ R_2 &\geq H(X_2|X_1) \\ R_1 + R_2 &\geq H(X_1, X_2). \end{aligned}$$

Proof:

1. Achievability Part: We need to show that for any (R_1, R_2) satisfying the constraint, a sequence of code pairs for X_1 and X_2 with asymptotically zero error probability exists.

Step 1: Random coding.

Step 2: Error probability.

2. Converse Part: We have to prove that if a sequence of code pairs for X_1 and X_2 has asymptotically zero error probability, then its rate pair (R_1, R_2) satisfies the constraint. □

Corollary 10.12 Given sequences of several (correlated) discrete memoryless sources $X_1, \ldots, X_m$ which are obtained from different terminals (and are to be encoded independently), the achievable code rate region satisfies

$$\sum_{i \in I} R_i \geq H(X_I | X_{L-I}),$$

for any index set $I \subset L \triangleq \{1, 2, \ldots, m\}$, where X_I represents $(X_{i_1}, X_{i_2}, \ldots)$ for $\{i_1, i_2, \ldots\} = I$.

Example. $m = 3$.

$$\begin{aligned}
R_1 + R_2 + R_3 &\geq H(X_1, X_2, X_3) \\
R_1 + R_2 &\geq H(X_1, X_2 | X_3) \\
R_2 + R_3 &\geq H(X_2, X_3 | X_1) \\
R_3 + R_1 &\geq H(X_3, X_1 | X_2) \\
R_1 &\geq H(X_1 | X_2, X_3) \\
R_2 &\geq H(X_2 | X_1, X_3) \\
R_3 &\geq H(X_3 | X_1, X_2)
\end{aligned}$$

Partial decoding of the original sources

- **Full decoding versus Partial decoding.**
  - The receiver intends to fully reconstruct all the original information transmitted, $X_1, \ldots, X_m$.
  - The receiver may only want to reconstruct part of the original information, say X_i for $i \in I \subset \{1, \ldots, m\}$ or X_I.
- Since it is in general assumed that $X_1, \ldots, X_m$ are **dependent**, the remaining information, X_i for $i \notin I$, should be helpful in the re-construction of X_I.
- Accordingly, these remain information are usually named the *side information* for lossless data compression.

Definition 10.13 (reconstructed information and side information) Let $L \triangleq \{1, 2, \ldots, m\}$ and I is any proper subset of L. Denote X_I as the sources X_i for $i \in I$, and similar notation is applied to X_{L-I}.

In the data compression with side-information, X_I is the information needs to be re-constructed, and X_{L-I} is the side-information.

Partial decoding of the original sources

Definition 10.14 (independent encoders among distributed sources) There are several sources $X_1, X_2, \ldots, X_m$ (may or may not be independent) which are respectively obtained by m terminals. Before each terminal transmits its local source to the receiver, a block encoder f_i with rate R_i and block length n is applied

$$f_i \ : \ \mathcal{X}_i^n \to \{1, 2, \ldots, 2^{nR_i}\}.$$

It is assumed that there is no conspiracy among block encoders.

Definition 10.15 (global decoder for independently compressed sources) A global decoder $g(\cdot)$ will recover the original sources after receiving all the independently compressed sources, i.e.,

$$g \ : \ \{1, \ldots, 2^{R_1}\} \times \cdots \times \{1, \ldots, 2^{R_m}\} \to \mathcal{X}_I^n.$$

Definition 10.16 (probability of error) The probability of error is defined as

$$P_e(n) \stackrel{\triangle}{=} Pr\{g(f_I(X_I^n) \neq (X_I^n)\}.$$

Definition 10.17 (achievable rates) A rates

$$(R_1, \ldots, R_m)$$

is said to be achievable if there exists a sequence of block codes such that

$$\limsup_{n\to\infty} P_e(n) = 0.$$

Partial decoding of the original sources

Definition 10.18 (achievable rate region for distributed sources) The achievable rate region for distributed sources is the set of all achievable rates.

Observation 10.19 The achievable rate region is convex.

Theorem 10.20 For distributed sources with two random variable X_1 and X_2, let X_1 be the re-constructed information and X_2 be the side information, the boundary function $R_1(R_2)$ for the achievable region is

$$R_1(R_2) \geq \min_{\{Z \,:\, X_1 \to X_2 \to Z \text{ and } I(X_2;Z) \leq R_2\}} H(X_1|Z)$$

- **Interpretation.**

  $$R_1 \geq H(X_1|Z) \text{ and } R_2 \geq I(X_2;Z),$$

  for any $X_1 \to X_2 \to Z$.

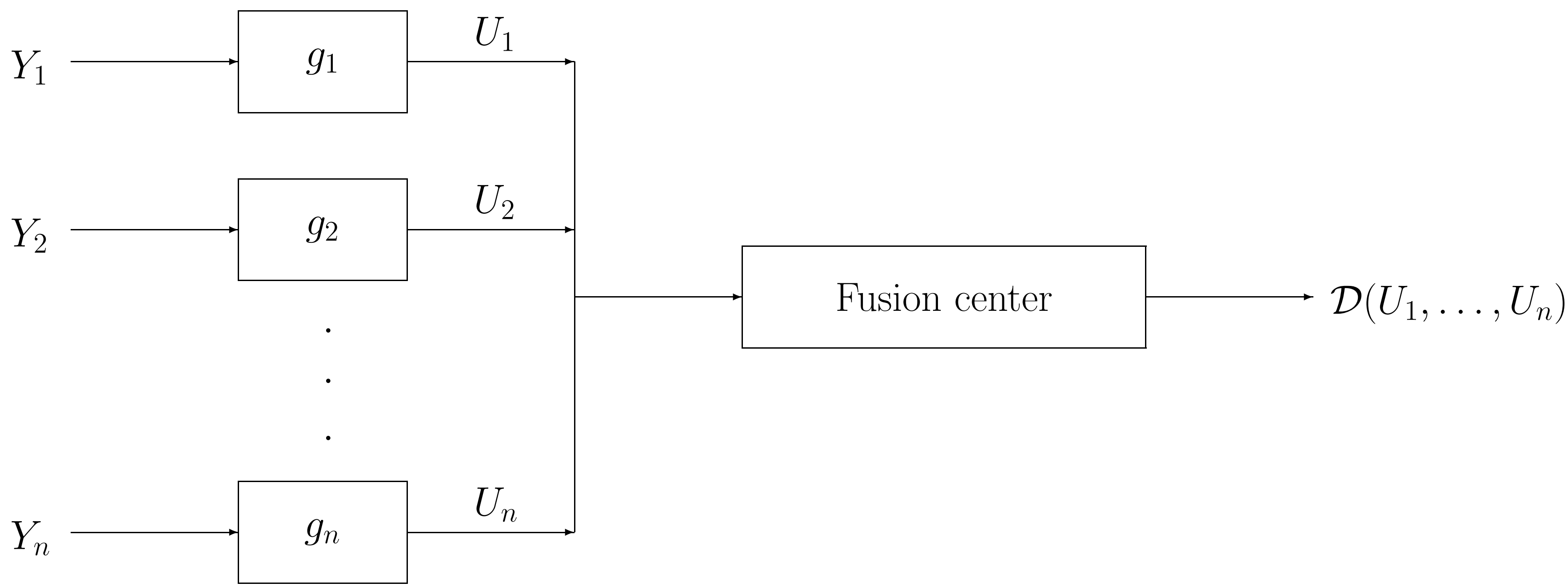
- Z = the coding outputs of X_2, received by the decoder, and is used by the receiver as a side information to reconstruct X_1.
- Hence, $I(X_2;Z)$ is the transmission rate from sender X_2 to the receiver.
- For all $f_2(X_2) = Z$ that has the same transmission rate $I(X_2;Z)$, the one that minimize $H(X_1|Z)$ will yield the minimum compression rate for X_1.

Distributed detection

- **Motivations.** Instead of re-construction of the original information, the decoder of a multiple sources system may only want to classify the sources into one of finitely many categories. This problem is usually named *distributed detection.*

Definition 10.21 (distributed system $\mathcal{S}_n$) A distributed detection system $\mathcal{S}_n$, as depicted in Fig. 10.3, consists of n geographically dispersed sensors, noiseless one-way communication links, and a fusion center. Each sensor makes an observation (denoted by Y_i) of a random source, quantizes Y_i into an m-ary message $U_i = g_i(Y_i)$, and then transmits U_i to the fusion center. Upon receipt of $(U_1, \ldots, U_n)$, the fusion center makes a global decision $\mathcal{D}(U_1, \ldots, U_n)$ about the nature of the random source.

Distributed detection



Distributed detection in $\mathcal{S}_n$.

Distributed detection

- The optimal design of $\mathcal{S}_n$ entails choosing quantizers $g_1, \ldots, g_n$ and a global decision rule $\mathcal{D}$ so as to optimize a given performance index.
- Binary hypothesis testing under the (classical) Neyman-Pearson and Bayesian formulations.

History.

- The joint optimization of entities $g_1, \ldots, g_n$ and $\mathcal{D}$ in $\mathcal{S}_n$ is a hard computational task, except in trivial cases (such as when the observations Y_i lie in a set of size no greater than m).
- It has been shown that whenever $Y_1, \ldots, Y_n$ are independent given each hypothesis, an optimal solution can be found in which $g_1, \ldots, g_n$ are threshold-type functions of the local likelihood ratio (possibly with some randomization for Neyman-Pearson testing).
- Still, we should note that optimization of $g_1, \ldots, g_n$ over the class of threshold-type likelihood-ratio quantizers is prohibitively complex when n is large.

Let us reduce the problem to the simplest statistics: i.i.d.

Question: whether a symmetric optimal solution exists in which the quantizers g_i are identical?

Distributed detection

- If so, then the optimal system design is considerably simplified.
- The answer is negative in general.

The general problem is as follows.

- System $\mathcal{S}_n$ is used for testing $H_0 : P$ versus $H_1 : Q$, where P and Q are one-dimensional marginals of the i.i.d. data $Y_1, \ldots, Y_n$.
- As n tends to infinity, both the minimum type II error probability $\beta_n^*(\alpha)$ (as function of the type I error probability bound α) and the Bayes error probability $\gamma_n^*(\pi)$ (as function of the prior probability π of H_0) vanish at an exponential rate.
- It thus becomes legitimate to adopt a measure of asymptotic performance based on the error exponents

$$e_{\text{NP}}^*(\alpha) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log \beta_n^*(\alpha)$$
$$e_{\text{B}}^*(\pi) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log \gamma_n^*(\pi) .$$

- It was shown by Tsitsiklis that, under certain assumptions on the hypotheses P and Q, it is possible to achieve the same error exponents using identical quantizers.

- Thus if $\beta_n^\diamond(\alpha)$, $\gamma_n^\diamond(\pi)$, $e_{\mathrm{NP}}^\diamond(\alpha)$ and $e_{\mathrm{B}}^\diamond(\pi)$ are the counterparts of $\beta_n^*(\alpha)$, $\gamma_n^*(\pi)$, $e_{\mathrm{NP}}^*(\alpha)$ and $e_{\mathrm{B}}^*(\pi)$ under the constraint that the quantizers $g_1, \ldots, g_n$ are identical, then

$$(\forall \alpha \in (0,1)) \qquad e_{\mathrm{NP}}^\diamond(\alpha) = e_{\mathrm{NP}}^*(\alpha)$$

and

$$(\forall \pi \in (0,1)) \qquad e_{\mathrm{B}}^\diamond(\pi) = e_{\mathrm{B}}^*(\pi) .$$

(Of course, for all n, $\beta_n^\diamond(\alpha) \geq \beta_n^*(\alpha)$ and $\gamma_n^\diamond(\pi) \geq \gamma_n^*(\pi)$.)

- This result provides some justification for restricting attention to identical quantizers when designing a system consisting of a large number of sensors.

Here we will focus on two issues.

- The first issue is the exact asymptotics of the minimum error probabilities achieved by the absolutely optimal and best identical-quantizer systems.
- the ratio $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi)$.
  - Note that equality in the error exponents of $\gamma_n^*(\pi)$ and $\gamma_n^\diamond(\pi)$ does not in itself guarantee that for any given n, the values of $\gamma_n^*(\pi)$ and $\gamma_n^\diamond(\pi)$ are in any sense close.

Definition 10.25 (divergence) The (Kullback-Leibler, informational) divergence, or relative entropy, of P relative to Q is defined by

$$D(P\|Q) \stackrel{\triangle}{=} E_P[X] \;=\; \int \log \frac{dP}{dQ}(y)\, dP(y)\ .$$

Lemma 10.27 (Neyman-Pearson type II error exponent of fixed test level) The optimal Neyman-Pearson error exponent in testing P versus Q at any level $\alpha \in (0,1)$ based on the i.i.d. observations $Y_1, \ldots, Y_n$ is $D(P\|Q)$.

Definition 10.28 (moment generation function of log-likelihood ratio) $\Psi(\theta)$ is the moment generation function of X under Q:

$$\Psi(\theta) \stackrel{\triangle}{=} E_Q[\exp\{\theta X\}] \;=\; \int \left(\frac{dP}{dQ}(y)\right)^{\theta} dQ(y)\ .$$

Lemma 10.29 (concavity of $\Psi(\theta)$)

1. For fixed $\theta \in [0,1]$, $\Psi(\theta)$ is a finite-valued concave functionals of the pair (P, Q) with the property $P \equiv Q$.
2. For fixed (P, Q) with $P \equiv Q$, $\Psi(\theta)$ is finite and convex in $\theta \in [0,1]$.

This last property, together with the fact that $\Psi(0) = \Psi(1) = 1$, guarantees that $\Psi(\theta)$ has a minimum value which is less than or equal to unity, achieved by some $\theta^* \in (0,1)$.

Definition 10.30 (Chernoff exponent) We define the Chernoff exponent

$$\rho(P,Q) \triangleq -\log \Psi(\theta^*) \ = \ -\log \left[\min_{\theta \in (0,1)} \Psi(\theta) \right].$$

Lemma 10.31 The Chernoff exponent coincides with the Bayes error exponent.

Example 10.32 (counterexample to $\gamma_n^*(\pi)/\gamma_n^\diamond(\pi) \to 1$) Consider a ternary observation space $\mathcal{Y} = \{a_1, a_2, a_3\}$ with binary quantization. The two hypotheses are assumed equally likely, with

y	a_1	a_2	a_3
$P(y)$	$1/12$	$1/4$	$2/3$
$Q(y)$	$1/3$	$1/3$	$1/3$
$(dP/dQ)(y)$	$1/4$	$3/4$	2

There are only two nontrivial deterministic LRQ's: $\hat{g}$, which partitions $\mathcal{Y}$ into $\{a_1\}$ and $\{a_2, a_3\}$; and $\bar{g}$, which partitions $\mathcal{Y}$ into $\{a_1, a_2\}$ and $\{a_3\}$.

$$\limsup_{k\to\infty} \frac{\gamma_{2k}^*(1/2)}{\gamma_{2k}^\diamond(1/2)} \ \le \ \frac{23}{24}\,.$$

Assumption 10.33 (boundedness assumption) There exists $\delta \geq 0$ for which

$$\sup_{g \in \mathcal{G}_m} E_P[|X_g|^{2+\delta}] < \infty, \tag{10.2.5}$$

where $\mathcal{G}_m$ is the set of all possible m-ary quantizers.

Theorem 10.34 *The boundedness assumption is equivalent to*

$$\limsup_{t \to \infty} E_P[|X_{\tau_t}|^{2+\delta}] \; < \; \infty \; , \tag{10.2.6}$$

where τ_t is defined as

$$\tau_t \triangleq ((-\infty, t] \, , \, (t, \infty)). \tag{10.2.7}$$

We now distinguish between three cases.

Case A. $$\limsup_{t \to \infty} E_P[X_{\tau_t}] \; = \; \infty \; .$$

Case B. $$0 \; < \; \limsup_{t \to \infty} E_P[X_{\tau_t}] \; < \; \infty \; .$$

Case C. $$\limsup_{t \to \infty} E_P[X_{\tau_t}] \; = \; 0 \quad \text{and} \quad \limsup_{t \to \infty} E_P[X_{\tau_t}^2] \; = \; \infty \; .$$

Example Let the observation space be the unit interval $(0, 1]$ with its Borel field. For $a > 0$, define the distributions P and Q by

$$P\{Y \le y\} = y \ , \qquad Q\{Y \le y\} = \exp\left\{\frac{a+1}{a}\left(1 - \frac{1}{y^a}\right)\right\} \ .$$

The pdf of Q is strictly increasing in y, and thus the likelihood ratio $(dP/dQ)(y)$ is strictly decreasing in y. Hence the event $\{X > t\}$ can also be written as $\{Y < y_t\}$, where $y_t \to 0$ as $t \to \infty$. Using this equivalence, we can examine the limiting behavior of $E_P[X_{\tau_t}]$ and $E_P[X_{\tau_t}^2]$ to obtain:

a. $a > 1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case A)

b. $a = 1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = 2, \ \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case B)

c. $1/2 < a < 1$: $\lim_{t\to\infty} E_P[X_{\tau_t}] = 0, \ \lim_{t\to\infty} E_P[X_{\tau_t}^2] = \infty$ (Case C)

d. $a \le 1/2$: $\lim_{t\to\infty} E_P[X_{\tau_t}^2] < \infty$ (Assumption 10.33 is satisfied) .

□

Theorem 10.37 (result for Case A) If $\limsup_{t\to\infty} E_P[X_{\tau_t}] = \infty$, then for all $m \geq 2$ and $\alpha \in (0,1)$,

$$e^{*}_{\mathrm{NP}}(\alpha) = e^{\diamond}_{\mathrm{NP}}(\alpha) = \infty .$$

Theorem 10.38 (result for Case B) Consider hypothesis testing with m-ary quantization, where $m \geq 2$. If

$$0 \ < \ \limsup_{t\to\infty} E_P[X_{\tau_t}] \ < \ \infty \ , \tag{10.2.11}$$

then there exist:

1. an increasing sequence of integers $\{n_k, k \in \mathbf{N}\}$ and a function $L \ : \ (0,1) \mapsto (0,\infty)$ which is monotonically increasing to infinity, such that

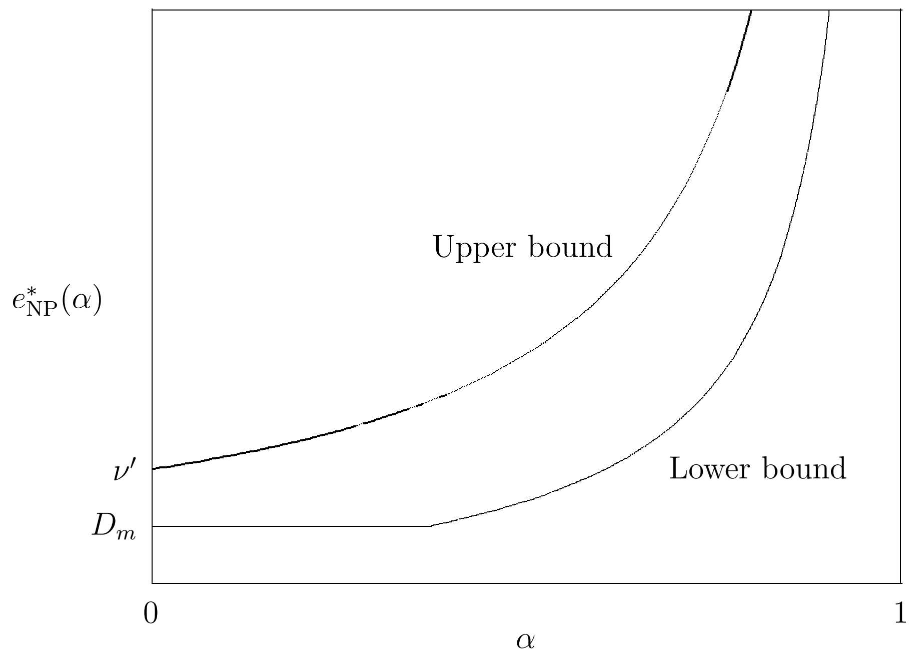
$$\liminf_{k\to\infty} -\frac{1}{n_k} \log \beta^{\diamond}_{n_k}(\alpha) \ \geq \ L(\alpha) \vee D_m \ ;$$

2. a function $M \ : \ (0,1) \mapsto (0,\infty)$ which is monotonically increasing to infinity and is such that

$$\limsup_{n\to\infty} -\frac{1}{n} \log \beta^{*}_{n}(\alpha) \ \leq \ M(\alpha)$$

where

$$L(\alpha) \triangleq \frac{\limsup_{t\to\infty} E_P[X_{\tau_t}]}{\log(1/\alpha)} \quad \text{and} \quad M(\alpha) \triangleq \frac{\sup_{g\in\mathcal{G}_m} E_P[|X_g|]}{1-\alpha}.$$



Upper and lower bounds on $e^*_{\mathrm{NP}}(\alpha)$ in Case B.

Theorem 10.39 (result for Case C) In Case C,

$$e^*_{\mathrm{NP}}(\alpha) \;=\; e^{\diamond}_{\mathrm{NP}}(\alpha) \;=\; D(P\|Q).$$

Theorem 10.40 (result under boundedness assumption) Let $\delta \le 1$ satisfy (10.2.5). If $\alpha \le 1/2$, or if $\alpha > 1/2$ and observation space $\mathcal{Y}$ is finite, then

$$\frac{\beta^*_n(\alpha)}{\beta^{\diamond}_n(\alpha)} \;\ge\; \exp\{-c'(\delta,\alpha)n^{\frac{1-\delta}{2}}\}\ .$$

In particular, if (10.2.5) holds for $\delta \ge 1$, then the ratio $\beta^*_n(\alpha)/\beta^{\diamond}_n(\alpha)$ is bounded from below.

Theorem 10.41 In Bayes testing with m-ary quantization,

$$\liminf_{n\to\infty} \frac{\gamma_n^*(\pi)}{\gamma_n^\diamond(\pi)} > 0 \tag{10.2.13}$$

for all $\pi \in (0,1)$.

Capacity region of multiple access channels

Definition 10.42 (discrete memoryless multiple access channel) A discrete memoryless multiple access channel contains several senders

$$(X_1, X_2, \ldots, X_m)$$

and one receiver Y, which are respectively defined over finite alphabet ($\mathcal{X}_1$,$\mathcal{X}_2$,...) and $\mathcal{Y}$. Also given is the transition probability $P_{Y|X_1,X_2,\ldots,X_m}$.

For simplicity, we will focus on the system with only two senders. The block code for this simple multiple access channel is defined below.

Definition 10.43 (block code for multiple access channels) A block code

$$(n, M_1 M_2)$$

for multiple access channel has block length n and rates $R_1 = (1/n)\log_2 M_1$ and $R_2 = (1/n)\log_2 M_2$ respectively for each sender as:

$$f_1 \ : \ \{1, \ldots, M_1\} \to \mathcal{X}_1^n,$$

and

$$f_2 \ : \ \{1, \ldots, M_2\} \to \mathcal{X}_2^n.$$

Upon receipt of the channel output, the decoder is a mapping

$$g \ : \ \mathcal{Y}^n \to \{1, \ldots, M_1\} \times \{1, \ldots, M_2\}.$$

Capacity region of multiple access channels

Theorem 10.44 (capacity region of memoryless multiple access channel) The capacity region for memoryless multiple access channel is the convex set of the set

$$\Big\{(R_1, R_2) \in (\Re^+ \cup \{0\})^2 \;:\; R_1 \le I(X_1; Y | X_2), R_2 \le I(X_2; Y | X_1)$$
$$\text{and } R_1 + R_2 \le I(X_1, X_2; Y)\Big\}.$$

Degraded broadcast channel

Definition 10.45 (broadcast channel) A broadcast channel consists of one input alphabet $\mathcal{X}$ and two (or more) output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$. The noise is defined by the conditional probability $P_{Y_1,Y_2|X}(y_1, y_2|x)$.

Example 10.46 Examples of broadcast channels are

- Cable Television (CATV) network;
- Lecturer in classroom;
- Code Division Multiple Access channels.

Definition 10.47 (degraded broadcast channel) A broadcast channel is said to be degraded if

$$P_{Y_1,Y_2|X}(y_1, y_2|x) = P_{Y_1|X}(y_1|x)P_{Y_2|Y_1}(y_2|y_1).$$

It can be verified that when $X \to Y_1 \to Y_2$ forms a Markov chain, in which $P_{Y_2|Y_1,X}(y_2|y_1, x) = P_{Y_2|Y_1}(y_2|y_1)$, a degraded broadcast channel is resulted. This indicates that the "parallelly" broadcast channel degrades to a "serially" broadcast channel, where the channel output Y_2 can only obtain information from channel input X through the previous channel output Y_1.

Degraded broadcast channel

Definition 10.48 (block code for broadcast channel) A block code for broadcast channel consists of one encoder $f(\cdot)$ and two (or more) decoders $\{g_i(\cdot)\}$ as

$$f \;:\; \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} \rightarrow \mathcal{X}^n,$$

and

$$\begin{aligned} g_1 &: \mathcal{X}^n \rightarrow \{1, \ldots, 2^{nR_1}\}, \\ g_2 &: \mathcal{X}^n \rightarrow \{1, \ldots, 2^{nR_2}\}. \end{aligned}$$

Definition 10.49 (error probability) Let the source index random variable be W_1 and W_2, namely $W_1 \in \{1, \ldots, 2^{nR_1}\}$ and $W_2 \in \{1, \ldots, 2^{nR_2}\}$. Then the probability of error is defined as

$$P_e \triangleq Pr\{W_1 \neq g_1[f(W_1, W_2)] \text{ or } W_2 \neq g_2[f(W_1, W_2)]\}.$$

Theorem 10.50 (capacity region for degraded broadcast channel) The capacity region for memoryless degraded broadcast channel is the convex set of

$$\bigcup_U \left\{(R_1, R_2) \;:\; R_1 \leq I(X; Y_1|U) \text{ and } R_2 \leq I(U; Y_2)\right\},$$

where the union is taking over all U satisfying $U \rightarrow X \rightarrow Y_1 Y_2$ with alphabet size $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$.

Degraded broadcast channel

Example 10.51 (capacity region for degraded BSC) Suppose $P_{Y_1|X}$ and $P_{Y_2|Y_1}$ are BSC with crossover ε_1 and ε_2, respectively. Then the capacity region can be parameterized through β as:

$$\begin{aligned} R_1 &\leq h_b(\beta \times \varepsilon_1) - h_b(\varepsilon_1) \\ R_2 &\leq 1 - h_b(\beta \times (\varepsilon_1(1-\varepsilon_2) + (1-\varepsilon_1)\varepsilon_2)), \end{aligned}$$

where $P_{X|U}(0|1) = P_{X|U}(1|0) = \beta$ and $U \in \{0, 1\}$.

Example 10.52 (capacity region for degraded AWGN channel) The channel is modeled as

$$Y_1 = X + N_1 \text{ and } Y_2 = Y_1 + N_2,$$

where the noise power for N_1 and N_2 are σ_1^2 and σ_2^2, respectively. Then the capacity region for input power constraint S should satisfy

$$\begin{aligned} R_1 &\leq \frac{1}{2}\log_2\left(1 + \frac{\alpha S}{\sigma_1^2}\right) \\ R_2 &\leq \frac{1}{2}\log_2\left(1 + \frac{(1-\alpha)S}{\alpha S + \sigma_1^2 + \sigma_2^2}\right), \end{aligned}$$

for any $\alpha \in [0, 1]$.

Gaussian multiple terminal channels

The encoder now becomes

$$\begin{aligned} f_1 &: \Re \to \Re \\ f_2 &: \Re \to \Re \\ &\vdots \\ f_m &: \Re \to \Re \end{aligned}$$

So we have now m (independent) transmitters, and one receiver in the system.

The system can be modeled as

$$Y = \sum_{i=1}^{m} X_i + N.$$

Theorem 10.53 (capacity region for AWGN multiple access channel) Suppose each transmitter has (constant) power constraint S_i. Let I denote the subset of $\{1, 2, \ldots, m\}$. Then the capacity region should be

$$\left\{ (R_1, \ldots, R_m) \; : \; (\forall\ I) \sum_{i \in I} R_i \le \frac{1}{2} \log_2 \left(1 + \frac{\sum_{i \in I} S_i}{\sigma^2} \right) \right\},$$

where σ^2 is the noise power of N.