Shannon's Mathematical Theory of Data Transmission

(for mathematicians)

Samir M. Perlaza

INRIA
Centre de Recherche de Sophia Antipolis - Méditerranée, France

Géométrie Algébraique, Théorie des nombres et Applications – GTA 2021
University of French Polynesia
Faaa, August 17 2021

Table of Contents

Data Transmission

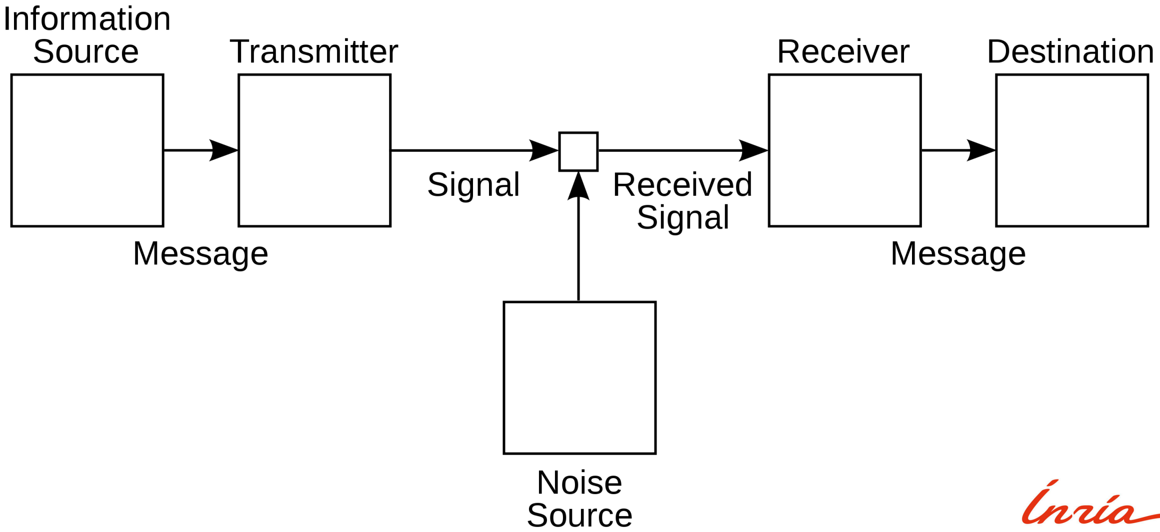


Table of Contents

What is Information?

To talk about **information**, it is necessary to define the **problem of estimation** or decision making.

The Problem of Estimation

- Consider a product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X}$ discrete. For all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,
$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i).$$
- For all $y \in \mathcal{Y}$, let $P_{Z|Y=y}$ be a **probability measure** on $(\mathcal{X}, \mathscr{F}_X)$. – (the estimator).
- $P_{Z|Y=y}(x)$ is the probability with which x is chosen given y.
- Let $f : \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \to [0, +\infty)$ be a **cost function**.
- An **estimator** $P_{Z|Y}$ is **preferred** against an estimator $Q_{Z|Y}$ if
$$\int f(x,y,z) \mathrm{d}P_{XYZ}(x,y,z) \leqslant \int f(x,y,z) \mathrm{d}Q_{XYZ}(x,y,z)$$
where,
$$P_{XYZ}(x,y,z) = P_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x) \text{ and}$$
$$Q_{XYZ}(x,y,z) = Q_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x).$$
- What is the **optimal estimator**?

The Problem of Estimation with $\mathcal{X} = \{0, 1\}$

Definition (Estimator)

Given a function $\delta : \mathcal{Y} \to [0, 1]$, the output of the estimator is 1 given the input $y \in \mathcal{Y}$ with probability $\delta(y)$.

Any estimator leads to **estimation errors** of two forms:

Definition (Error Probability with $\mathcal{X} = \{0, 1\}$)

Given a decision rule formed by the sets $\mathcal{Y}_0$ and $\mathcal{Y}_1$,

- the **type-I error probability** is $\int \delta(y) \mathrm{d}P_{Y|X=0}(y)$, and
- the **type-II error probability** is $\int (1 - \delta(y)) \, dP_{Y|X=1}(y)$.

Bayesian Method

For all $(i, j) \in \{0, 1\}^2$,

- let $C_{ij} \in [0, +\infty)$ be the **cost** of choosing i when $X = j$, with
- $0 \leqslant C_{0,0} < C_{1,0}$ and $0 \leqslant C_{1,1} < C_{0,1}$.

Given an estimator $\delta : \mathcal{Y} \to [0, 1]$, **the expected cost** is

$$C(\delta) = P_X(0)\left(C_{00}\int (1-\delta(y))\,\mathrm{d}P_{Y|X=0}(y) + C_{10}\int \delta(y)\mathrm{d}P_{Y|X=0}(y)\right) + P_X(1)\left(C_{11}\int \delta(y)\mathrm{d}P_{Y|X=1}(y) + C_{01}\int (1-\delta(y))\,\mathrm{d}P_{Y|X=1}(y)\right)$$

Example ($C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$)

The expected cost

$$C(\delta) = P_X(0)\int \delta(y)\mathrm{d}P_{Y|X=0}(y) + P_X(1)\int (1-\delta(y))\,\mathrm{d}P_{Y|X=1}(y)$$

Bayesian Method: Optimal Decision Rules

Theorem

An estimator $\delta : \mathcal{Y} \to [0,1]$ *that satisfies for all* $y \in \mathcal{Y}$,

$$\delta(y) = \begin{cases} 1 & \text{if} \quad y \in \mathcal{Y}_1 \\ 0 & \text{if} \quad y \in \mathcal{Y}_0 \\ \alpha & \text{otherwise} \end{cases}$$

where $\alpha \in [0,1]$, *and*

$$\mathcal{Y}_1 = \left\{ y \in \mathcal{Y} : \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) > \frac{P_X(0)\,(C_{10} - C_{00})}{P_X(1)\,(C_{01} - C_{11})} \right\} \text{ and}$$

$$\mathcal{Y}_0 = \left\{ y \in \mathcal{Y} : \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) < \frac{P_X(0)\,(C_{10} - C_{00})}{P_X(1)\,(C_{01} - C_{11})} \right\},$$

is optimal.

Bayesian Method: (Sketch of) Proof of Optimality

Given an estimator δ, **the expected cost** is

$$\begin{aligned}
C(\delta) &= \int P_X(0)\left(C_{00}(1-\delta(y)) + C_{10}\delta(y)\right)\mathrm{d}P_{Y|X=0}(y) + \int P_X(1)\left(C_{11}\delta(y) + C_{01}(1-\delta(y))\right)\mathrm{d}P_{Y|X=1}(y) \\
&= \int P_X(0)\left(C_{00}(1-\delta(y)) + C_{10}\delta(y)\right)\mathrm{d}P_{Y|X=0}(y) \\
&\quad + \int P_X(1)\left(C_{11}\delta(y) + C_{01}(1-\delta(y))\right)\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\mathrm{d}P_{Y|X=0}(y) \\
&= \int \Bigg((1-\delta(y))\left(P_X(0)\,C_{00} + P_X(1)\,C_{01}\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\right) \\
&\quad + \delta(y)\left(P_X(0)\,C_{10} + P_X(1)\,C_{11}\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\right)\Bigg)\mathrm{d}P_{Y|X=0}(y),
\end{aligned}$$

where the last equality involves a convex combination of two terms.

Connections to *Maximum à Posteriori* (MAP) estimation

Theorem

Under the assumptions that $C_{00} = C_{11} = 0$ *and* $C_{10} = C_{01} = 1$ *and* $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, *with* $\mathcal{X}$ *and* $\mathcal{Y}$ *discrete, is a probability measure space, an estimator* $\delta : \mathcal{Y} \rightarrow [0, 1]$ *that satisfies for all* $y \in \mathcal{Y}$,

$$\delta(y) = \begin{cases} 1 & \text{if} \quad y \in P_{X|Y=y}(1) > P_{X|Y=y}(0) \\ 0 & \text{if} \quad y \in P_{X|Y=y}(1) < P_{X|Y=y}(0) \\ \alpha & \text{otherwise}, \end{cases}$$

with $\alpha > 0$, **is optimal**.

Neyman-Pearson Method

- Probability of choosing 1 when $x = 1$ (detection) is

$$\int \delta(y) \mathrm{d}P_{Y|X=1}(y)$$

- Probability of choosing 0 when $x = 1$ (false alarm) is

$$\int (1 - \delta(y)) \mathrm{d}P_{Y|X=1}(y)$$

Choose the estimator that achieves the maximum detection probability and a probability of false alarm that is **not bigger than** p, with $p > 0$.

$$\max_{\delta} \int \delta(y) \mathrm{d}P_{Y|X=1}(y)$$
$$\text{s.t.} \int (1 - \delta(y)) \mathrm{d}P_{Y|X=1}(y) \leqslant p.$$

Neyman-Pearson Method: Optimal Estimator

Theorem

An estimator $\delta : \mathcal{Y} \to [0,1]$ *that satisfies for all* $y \in \mathcal{Y}$,

$$\delta(y) = \begin{cases} 1 & \textit{if} \quad \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) > \eta \\ 0 & \textit{if} \quad \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) < \eta \\ \alpha & \textit{if} \quad \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) = \eta, \end{cases}$$

where $\alpha \in [0,1]$ *and* $\eta \geqslant 0$, *are chosen such that*

$$\int (1 - \delta(y)) \mathrm{d}P_{Y|X=1}(y) = p,$$

is optimal.

What is **information**? – Comments

To talk about **information**, we need to realize that:

- There is a choice to be made (**estimation/decision** problem)
- The choice is based on three elements:
  - an **observation** $y \in \mathcal{Y}$; and
  - a **product measure space** $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$; and
  - a **cost function**.
- **Underlying assumption**:
  - The **product measure space** and the **cost function** are both <u>known</u>.

What is **information**? – Definition with $\mathcal{X} = \{0, 1\}$

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0, 1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right).$$

Note that

$$\iota(y) = \log_2 \left(\left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right)^{-1} \right) = \log_2 \left(\frac{\mathrm{d}P_{Y|X=0}}{\mathrm{d}P_{Y|X=1}}(y) \right).$$

What if $P_{Y|X=0}$ and $P_{Y|X=1}$ are not **absolutely continuous** w.r.t each other?

What is **information**? – Definition with $\mathcal{X} = \{0, 1\}$

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0, 1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right).$$

The information brought by observing y on whether $x = 1$ or $x = 0$ is

$$-\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right) \textit{bits}.$$

A rule of the form $\iota(y) > \tau$ is enough for decision making, $\tau > 0$.

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0,1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A})P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2\left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\right).$$

The **expectation of the information**

- with respect to $P_{Y|X=1}$ is

$$\int \log_2(\iota(y))\,\mathrm{d}P_{Y|X=1}(y) = -\int \frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\log_2\left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y)\right)\mathrm{d}P_{Y|X=0}(y)$$

What is **information**? – Definition with $\mathcal{X} = \{0, 1\}$

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0, 1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right).$$

The **expectation of the information**

- with respect to $P_{Y|X=1}$, when $P_{Y|X=0}$ is a **counting measure**, is

$$-\int \log_2 (\iota(y)) \, \mathrm{d}P_{Y|X=1}(y) = -\sum_{y \in \mathcal{Y}} p(y) \log_2 p(y),$$

  which is the **Shannon entropy** of the probability mass function p (of the measure $P_{Y|X=0}$)

What is **information**? – Definition with $\mathcal{X} = \{0,1\}$

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0,1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right).$$

The **expectation of the information**

- with respect to $P_{Y|X=1}$, when $P_{Y|X=0}$ is the **Lebesgue measure**, is

$$-\int \log_2 (\iota(y)) \, \mathrm{d}P_{Y|X=1}(y) = -\int_{\mathcal{Y}} f(y) \log_2 f(y) \, \mathrm{d}y,$$

which is the **Shannon (differential) entropy** of the probability density function f (of the measure $P_{Y|X=0}$)

What is **information**? – Definition with $\mathcal{X} = \{0, 1\}$

Definition (Information)

Given the product measure space $(\mathcal{X} \times \mathcal{Y}, \mathscr{F}_X \times \mathscr{F}_Y, P_{XY})$, with $\mathcal{X} = \{0, 1\}$ and for all $i \in \mathcal{X}$ and $\mathcal{A} \in \mathscr{F}_Y$,

$$P_{XY}(i, \mathcal{A}) = P_{Y|X=i}(\mathcal{A}) P_X(i),$$

the information function $\iota : \mathcal{Y} \to \mathbb{R}$ is for all $y \in \mathcal{Y}$

$$\iota(y) = -\log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right).$$

The **expectation of the information**

- with respect to $P_{Y|X=0}$ is

$$-\int \log_2 \left(\frac{\mathrm{d}P_{Y|X=1}}{\mathrm{d}P_{Y|X=0}}(y) \right) \mathrm{d}P_{Y|X=0}(y).$$

What is **information**? – Definition with $\mathcal{X} = \{0, 1\}$

What is so special about
counting measures and **the Lebesgue Measure**?

Consider that in the estimation problem:

- The set $\mathcal{Y}$ is discrete; and
- The measure $P_{Y|X=0}$ is a counting measure of the form

$$P_{Y|X=0}(\mathcal{A}) = \frac{|\mathcal{A}|}{|\mathcal{Y}|}, \text{ with } \mathcal{A} \in \mathscr{F}_Y.$$

Hence,

- The measure **does not depend** on the exact element y.
- For all observations $y \in \mathcal{Y}$, $P_{Y|X=0}(y) = \frac{1}{|\mathcal{Y}|}$.

Table of Contents

The Problem of Data Transmission

- **Equiprobable** source symbols:

$$\mathcal{W} = \{1, 2, \ldots, M\}, \text{ with } M \in \mathbb{N}.$$

- **Channel**
  - Input symbols: $\mathcal{X}$
  - Output symbols: $\mathcal{Y}$
  - Input-Output **Random** Transformation:

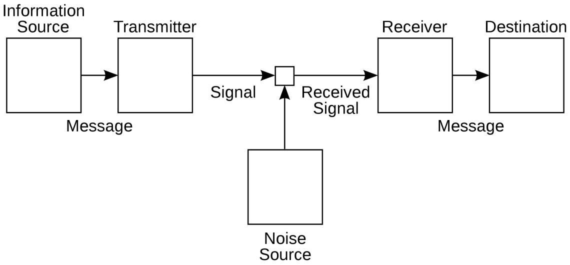
$$\forall x \in \mathcal{X}, \quad Y \sim P_{Y|X=x}$$

- **Transmission Duration**:

$$n \text{ channel uses}$$

- **Transmission Rate:**

$$R = \frac{\log_2 M}{n} \text{ bits per channel use.}$$

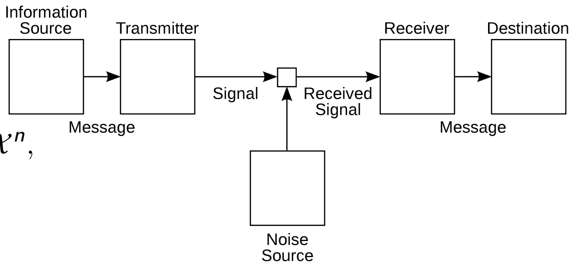


The Problem of Data Transmission

- Transmitter uses M **codewords**:

$$\begin{aligned}\boldsymbol{u}(1) &= (u_1(1), u_2(1), \ldots, u_n(1)) \in \mathcal{X}^n \\ \boldsymbol{u}(2) &= (u_1(2), u_2(2), \ldots, u_n(2)) \in \mathcal{X}^n \\ &\vdots \\ \boldsymbol{u}(M) &= (u_1(M), u_2(M), \ldots, u_n(M)) \in \mathcal{X}^n,\end{aligned}$$

  with $n \in \mathbb{N}$, the **block length**.
- To transmit symbol $i \in \mathcal{W}$,
  at each channel use $t \in \{1, 2, \ldots, n\}$,
  - **Channel Input:** $u_t(i)$; and
  - **Channel Output:** $Y_t \sim P_{Y|X=u_t(i)}$.



The Problem of Data Transmission

- After n channel uses, at the receiver

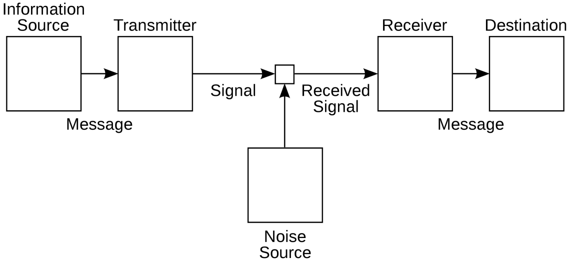
$$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n) \sim P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}.$$

where for all $\boldsymbol{y} \in \mathcal{Y}^n$

$$P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}_t}(\boldsymbol{y}) = \prod_{t=1}^{n} P_{Y|X=u_t(i)}(y_t).$$

- **Estimation:**
  - Given the observation $\boldsymbol{y} \in \mathcal{Y}^n$,
  - A partition of $\mathcal{Y}^n$: $\mathcal{D}_1$, $\mathcal{D}_2$, ..., $\mathcal{D}_M$
  - Symbol $i \in \mathcal{W}$ is the receiver output if

$$\boldsymbol{Y} \in \mathcal{D}_i,$$



The Problem of Data Transmission

- Average **decoding error probability**:

$$\lambda = \frac{1}{M}\sum_{i=1}^{M} \Pr\left[\boldsymbol{Y} \notin \mathcal{D}_i | \boldsymbol{X} = \boldsymbol{u}(i)\right]$$

- **Objective:**
  Choose the sets $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M$ for minimizing the
  average decoding error probability.

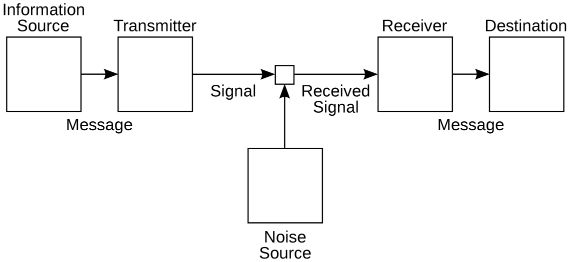


Table of Contents

Fundamental Limits – Main Questions

Given an information rate (fixed n and M),

- What is the **minimum decoding error probability**?

Given a tolerable decoding error probability,

- What is the **maximum information rate**?
- What is the **shortest code** for transmitting M symbols?

Fundamental Limits: $(n,\ M,\ \lambda)$-Codes

Definition ($(n,\ M,\ \lambda)$-Code)

Given a tuple $(M,\ n,\ \lambda) \in \mathbb{N}^2 \times [0,1]$, an $(n,\ M,\ \lambda)$-code for the random transformation $P_{\boldsymbol{Y}|\boldsymbol{X}}$ is a system

$$\left\{ \left(\boldsymbol{u}(1), \mathcal{D}_1 \right), \left(\boldsymbol{u}(2), \mathcal{D}_2 \right), \ldots, \left(\boldsymbol{u}(M), \mathcal{D}_M \right) \right\},$$

where for all $(j, \ell) \in \mathcal{W}^2$, with $j \neq \ell$:

$$\begin{aligned} \boldsymbol{u}(j) &= (u_1(j), u_2(j), \ldots, u_n(j)) \in \mathcal{X}^n, \\ \mathcal{D}_j \cap \mathcal{D}_\ell &= \emptyset, \\ \bigcup_{j \in \mathcal{W}} \mathcal{D}_j &\subseteq \mathcal{Y}^n, \text{ and} \\ \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}} \left[\mathbb{1}_{\{Y \notin \mathcal{D}_i\}} \right] &\leqslant \lambda. \end{aligned}$$

Fundamental Limits – Main Questions

Definition (Minimum decoding error probability (DEP))

Given a pair $(n, M) \in \mathbb{N}^2$, the minimum average DEP for the memoryless channel $(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}})$, denoted by $\lambda^*(n, M)$, is given by

$$\lambda^*(n, M) = \min \left\{ \lambda \in [0, 1] : \exists (n, M, \lambda)\text{-code} \right\}.$$

Fundamental Limits: (n, M, λ)-Codes

A lower bound on the average error probability.

Theorem (Verdu-Han-1994)

Every (n, M, λ)-code for a random transformation $P_{\boldsymbol{Y}|\boldsymbol{X}}$ satisfies

$$\lambda > \sup_{\beta > 0} \left[\Pr\left[\iota(\boldsymbol{X}, \boldsymbol{Y}) \leqslant \log_2(\beta)\right] - \frac{\beta}{M} \right],$$

where $P_{\boldsymbol{X}}$ is the empirical input distribution of the code. For all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$,

$$P_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{x}) = P_{\boldsymbol{X}}(\boldsymbol{x}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}),$$

and

$$\iota(\boldsymbol{x}, \boldsymbol{y}) = \log_2\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})} \right).$$

Fundamental Limits: (n, M, λ)-Codes

An upper bound on the average error probability.

Theorem (Polyanskiy-Poor-Verdu-2010)

For all $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$, *there always exists an* (n, M, λ)*-code such that*

$$\lambda < \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}\left[\min\{1, (M-1)\}\Pr\left[\iota\left(\bar{\boldsymbol{X}}, \boldsymbol{Y}\right) \geqslant \iota\left(\boldsymbol{X}, \boldsymbol{Y}\right)\right]\right],$$

where for all $(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{X}^n \times \mathcal{Y}^n$,

$$P_{\bar{\boldsymbol{X}}\boldsymbol{Y}\boldsymbol{X}}(\bar{\boldsymbol{x}}, \boldsymbol{y}, \boldsymbol{x}) = P_{\bar{\boldsymbol{X}}}(\bar{\boldsymbol{x}}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) P_{\boldsymbol{X}}(\boldsymbol{x}) \quad (1)$$

and

$$\iota(\boldsymbol{x}, \boldsymbol{y}) = \log_2\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right).$$

Fundamental Limits – Main Questions

Is it easy to calculate these bounds?

Fundamental Limits – Main Questions

Is it easy to calculate these bounds?

Even in the case in which for all $\boldsymbol{y} \in \mathcal{Y}^n$ and for all $i \in \{1, 2, \ldots, M\}$,

$$P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}_t}(\boldsymbol{y}) = \prod_{t=1}^{n} P_{Y|X=u_t(i)}(y_t) \quad ?$$

Fundamental Limits – Main Questions

In **memoryless channels**, for all $\boldsymbol{y} \in \mathcal{Y}^n$ and for all $i \in \{1, 2, \ldots, M\}$,

$$P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}_t}(\boldsymbol{y}) = \prod_{t=1}^{n} P_{Y|X=u_t(i)}(y_t),$$

it follows that

$$\iota(\boldsymbol{x}, \boldsymbol{y}) = \log_2\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{X}}(\boldsymbol{x})}\right) = \sum_{t=1}^{n} \log_2\left(\frac{P_{Y|X}(y_t|x_t)}{P_X(x_t)}\right) = \sum_{t=1}^{n} \iota(x_t, y_t),$$

and thus, the random variable $\iota(\boldsymbol{X}, \boldsymbol{Y})$, with $(\boldsymbol{X}, \boldsymbol{Y}) \sim P_{\boldsymbol{XY}}$ is formed by the **sum of** n **i.i.d. random variables** $\iota(X, Y)$, with $(X, Y) \sim P_{XY}$.

Fundamental Limits – Sums of i.i.d. Random Variables

Theorem (Berry-Esseen (Feller-1971))

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d random variables with probability distribution P_Y, mean μ, variance ν, and third absolute central moment ξ. Then, the CDF of the random variable $X_n = Y_1 + Y_2 + \ldots + Y_n$, denoted by F_{X_n}, satisfies

$$\sup_{a \in \mathbb{R}} |F_{X_n}(a) - F_{Z_n}(a)| \leqslant \min\left(1, \frac{c\,\xi}{\sqrt{nV^3}}\right),$$

where $c = 0.476$ (Shevtsova-2011) and F_{Z_n} is the CDF of a Gaussian random variable with the same mean and variance as X_n.

$\frac{c\,\xi}{\sqrt{nV^3}} \geqslant 4.10^{-3}$ for $n = 10^4$ (URLLC targets errors of order 10^{-6}).

Fundamental Limits – Sums of i.i.d. Random Variables

How to solve the precision problem ?

Fundamental Limits – Sums of i.i.d. Random Variables

Theorem (Anade-Gorce-Mary-Perlaza-2020)

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables with probability distribution P_Y and cumulant generating function K_Y. Let also F_{X_n} be the CDF of the random variable $X_n = Y_1 + Y_2 + \ldots + Y_n$. Hence, for all a element of the interior of the convex hull of $\operatorname{supp} P_{X_n}$, it holds that

$$\left|F_{X_n}(a) - \hat{F}_{X_n}(a)\right| \leqslant \exp\left(nK_Y(\theta^\star) - \theta^\star a\right)\min\left(1, \frac{2c\,\xi_Y(\theta^\star)}{\left(K_Y^{(2)}(\theta^\star)\right)^{3/2}\sqrt{n}}\right),$$

with

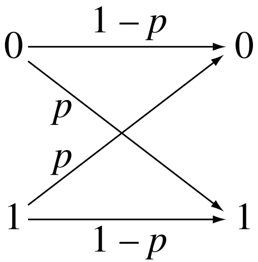
$$\hat{F}_{X_n}(a) = \mathbb{1}_{\{\theta^\star > 0\}} + (-1)^{\mathbb{1}_{\{\theta^\star > 0\}}}\exp\left(nK_Y(\theta^\star) - \theta^\star a + \frac{1}{2}\theta^{\star 2} nK_Y^{(2)}(\theta^\star)\right)Q\left(|\theta^\star|\sqrt{nK_Y^{(2)}(\theta^\star)}\right),$$

$$\xi_Y(\theta^\star) = \mathbb{E}_{P_Y}\left[\left|Y - K_Y^{(1)}(\theta^\star)\right|^3 \exp\left(\theta^\star Y - K_Y(\theta^\star)\right)\right], \text{ and } nK_Y^{(1)}(\theta^\star) = a,$$

where the functions $K_Y^{(1)}$ and $K_Y^{(2)}$ are respectively the first and second derivatives of the function K_Y; the function Q is the complementary CDF of the Gaussian random variable with mean zero and unit variance; and $c = 0.476$.

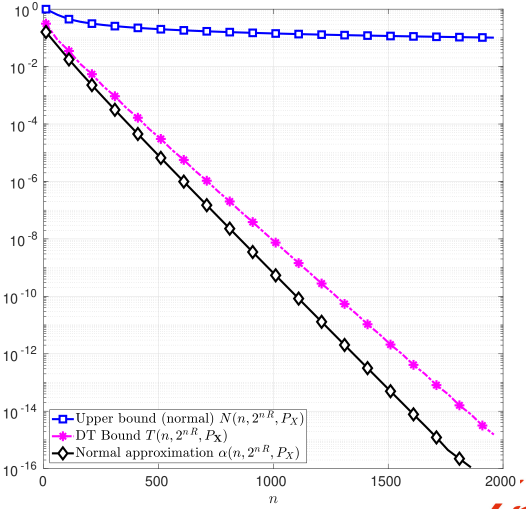
$(nK_Y(\theta^\star) - \theta^\star a)$ is a negative decreasing function of $|a - E_{P_{X_n}}[X_n]|$

Fundamental Limits – Binary Symmetric Channel

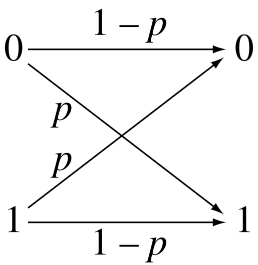


- Cross-over probability $p = 0.11$
- Information rate

$$R = \frac{\log_2(M)}{n} = 0.32 \text{ bits/ch.use}$$



Fundamental Limits – Binary Symmetric Channel



- Cross-over probability $p = 0.11$
- Information rate

$$R = \frac{\log_2(M)}{n} = 0.32 \text{ bits/ch.use}$$

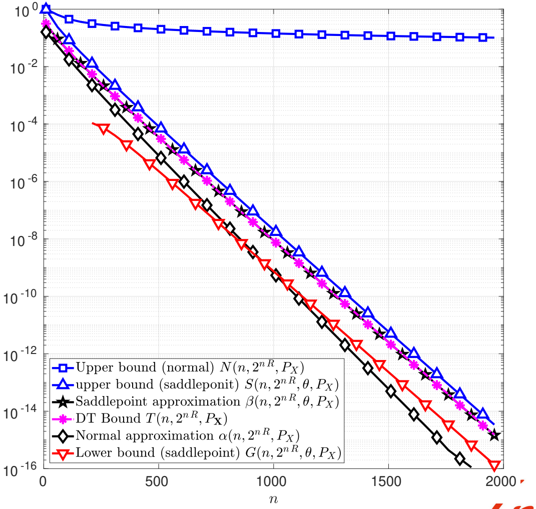


Table of Contents

Thank you for your attention. Questions ?