# Information and complexity

## Channel coding

Dimitri Petritis

Institut de recherche mathématique de Rennes
Université de Rennes 1 et CNRS (UMR 6625)

September 2022

# Channels
## Markovian modelling

- **Channel** = general notion. Initially meant to study transmission of a coded message through noisy medium.
- Presently, to mean arbitrary transformation of a word of a finite alphabet to another word of a (may be different) finite alphabet.
- **Input:** random word $\mathbf{X} \in \mathbb{X}^+$, where $\mathbb{X}$ input alphabet.
- **Output:** random word $\mathbf{Y} \in \mathbb{Y}^+$, where $\mathbb{Y}$ output alphabet..
- **Transmission probability:** conditional probability $\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x})$.
- Assume (for simplicity)
  - input and output words of same lenght; i.e. $|\mathbf{x}| = |\mathbf{y}|$ and
  - input symbols emitted by independent source.

  Both hypotheses can be relaxed at the price of more complicated formulæ.

# Channels
## Ideal and realistic

- Transmit 1 bit of information = transport the precise state — of physical system encoding the bit — through a physical medium or process — the channel:

| Transmission vector | Ideal channel | Realistic channel |
|---|---|---|
| electric current | ideal cooper wire | copper wire with $> 0$ resistivity |
| Hertzian beam | empty space | atmosphere |
| laser beam | empty space or fiber | fiber not 100% transparent |
| photon | empty space or fiber | fiber not 100% transparent |
| DNA | cellular mitosis/meiosis | mutations |
| ... | ... | ... |

- $\Rightarrow$ transmission errors.

# Channels

## Markovian modelling (cont'd)

### Definition

$(p_n)_{n\in\mathbb{N}}$, $n \in \mathbb{N}$, sequence defined by

$$\mathbb{X}^n \times \mathbb{Y}^n \ni (\mathbf{x}, \mathbf{y}) \to p_n(\mathbf{x}, \mathbf{y}) := \mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{x} = \mathbf{x}) \in [0, 1].$$

- Triple $(\mathbb{X}, \mathbb{Y}, (p_n)_{n\in\mathbb{N}})$ **discrete channel**.
- Channel is **memoryless** if exists stochastic matrix $P : \mathbb{X} \times \mathbb{Y} \to [0, 1]$ s.t. for all $n \in \mathbb{N}$, $\mathbf{x} \in \mathbb{X}^n$, and $\mathbf{y} \in \mathbb{Y}^n$, conditional probability reads $p_n(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^n P(x_i, y_i)$.

Memoryless channel identified with triple $(\mathbb{X}, \mathbb{Y}, P)$, where $P$ a $|\mathbb{X}| \times |\mathbb{Y}|$-stochastic matrix.

3-A: Examples and illustration of the generality of the notion of channel.

eg1: $\mathcal{X} = \mathcal{Y} = \{0,1\}$ $\quad |\mathcal{X}| = |\mathcal{Y}| = 2$

$$P = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-e_0 & e_0 \\ 1 & e_1 & 1-e_1 \end{array} \qquad e_0, e_1 \in [0,1]$$

$$\mathbb{P}(\underline{Y} = 110 \mid \underline{X} = 010) = P(0,1)\,P(1,1)\,P(0,0) = e_1(1-e_1)(1-e_0)$$

eg: $\mathcal{X} = \{a, b, c, d\}$ $\quad \mathcal{A} = \{0,1\}$ $\quad C: \mathcal{X} \to \mathcal{A}^+$

| $x$ | $C(x)$ |
|---|---|
| a | 0 |
| b | 10 |
| c | 110 |
| d | 111 |

instantaneous code can be viewed as $(\mathcal{X}, \mathcal{Y}, P)$ channel $\quad \mathcal{Y} = C(\mathcal{X}) = \{0, 10, 110, 111\}$

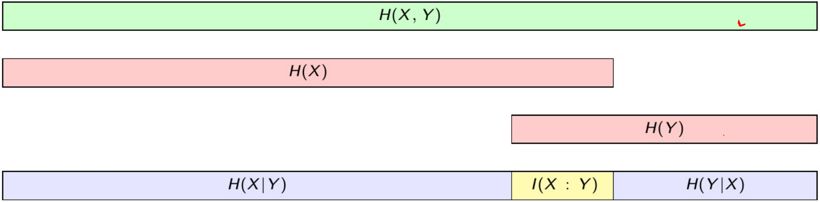
$$P = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{cccc} 0 & 10 & 110 & 111 \\ \hline 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{array}$$

# Sources and channels

## Reminder

If source law is $\mu \in \mathrm{PV}_{\mathbb{X}}$ and channel transmission matrix is $P$, can compute

- source entropy $H(X)$, $= -\sum \mu(x) \log \mu(x)$
- joint law of input-output $\kappa(x, y) = \mathbb{P}(X = x, Y = y) = \mu(x)P(x, y)$ (hence joint entropy $H(X, Y)$), $= -\sum_{x,y} \kappa(x,y) \log \kappa(x,y)$
- output law $\nu(y) = \sum_{x\in\mathbb{X}} \kappa(x, y) = \sum_{x\in\mathbb{X}} \mu(x)P(x, y)$ $= (\mu P)(y)$ (hence output entropy $H(Y)$),
- conditional entropies $H(X|Y)$ and $H(Y|X)$ and mutual information $I(X : Y)$.



# Channels
## Channel classification

**Channels without loss**: caracterised by $H(X|Y) = 0$; if output is known, no residual uncertainty on input. Equivalently $I(X : Y) = H(X) - H(X|Y) = H(X)$.

**Deterministic channels**: their transmission matrix is deterministic i.e.

$$\forall x \in \mathbb{X}, \exists! y := y_x \in \mathbb{Y}, P(x, y_x) = 1.$$

If $\mu$ source law,

$$\mathbb{X} \times \mathbb{Y} \ni (x, y) \mapsto \kappa(x, y) = \mu(x) P(x, y) = \mu(x) \delta_{y_x, y}.$$

Hence $H(X, Y) = H(\kappa) = H(X)$, $H(Y|X) = H(X, Y) - H(X) = 0$ and $I(X : Y) = H(Y) - H(Y|X) = H(Y)$, i.e. if input is known, no residual uncertainty on output.

**Noiseless channels:** without loss ($H(X|Y) = 0$) and deterministic ($H(Y|X) = 0$). Hence $I(X : Y) = H(X) = H(Y)$.

**Useless channels:** $\forall \mu \in \mathrm{PV}_{\mathbb{X}}$, $I(X : Y) = 0$. Hence

$$0 = I(X : Y) = H(X) - H(X|Y) = 0 \Rightarrow H(X) = H(X|Y),$$

i.e. input, $X$, and output, $Y$, variables independent.

**Symmetric channels:** continues to next slide . . .

# Channels
## Channel classification (cont'd)

**Symmetric channels (cont(d):** $\mathbb{S}_n$ permutation group on $n$ objects and $(\mathbb{X}, \mathbb{Y}, P)$ memoryless channel. $|\mathbb{X}| = n$

**Definition**

Assume $\exists \mathbf{p} \in \mathrm{PV}_{\mathbb{Y}}$ and $\exists \mathbf{z} \in [0,1]^{|\mathbb{X}|}$, s.t.

1. $\forall x \in \mathbb{X}, \exists \sigma_x \in \mathbb{S}_{|\mathbb{Y}|} : \forall y \in \mathbb{Y}, P(x,y) = p(\sigma_x y)$ and
2. $\forall y \in \mathbb{X}, \exists \sigma_y \in \mathbb{S}_{|\mathbb{X}|} : \forall x \in \mathbb{X}, P(x,y) = z(\sigma_y x)$.

Then channel is **symmetric**.

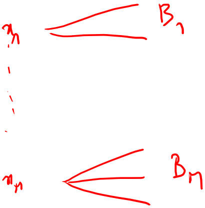
3-B: Examples of various types of channels.

d) Lossless $H(X|Y) = 0 \Rightarrow I(X;Y) = H(X)$.

$\mathcal{X} = \{x_1 \dots x_M\}$ $\mathcal{Y} = \bigsqcup_{i=}^{M} B_i$ $B_i \neq \emptyset$

$(P(x,y))$

$$\mathbb{P}(Y \in B_1 \mid X = x_i) = \sum_{y \in B_i} P(x_i, y) = 1$$




$$\mathbb{P}(X = x_i \mid Y \in B_i) = 1$$

Deterministic

52 cards $= \mathcal{X}_1 \times \mathcal{X}_2 = \mathcal{X}$ $\quad \mathcal{X}_1 = \{1, \ldots 10, J, Q, K\}$

$\mathcal{Y} = \mathcal{X}_2 = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$

$X \in \mathcal{X}$

$Y \in \mathcal{Y}$

$$I(X;Y) = \underbrace{H(Y)}_{2} - \underbrace{H(Y|X)}_{0}$$

Noiseless $\qquad$ lossless + det

$a \longrightarrow a'$

$b \longrightarrow b'$

$I(X;Y) = H(X) = H(Y)$

Symmetrie:

$$P_1 = \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{bmatrix}$$

$\mathcal{X} = \{0,1\}$

$\mathcal{Y} = \{a, b, c, d\}$

$$P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix} = \underline{P}$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - H(\underline{p})$$

# Channel capacity

## Definition

For fixed channel (i.e. fixed transmission matrix $P$), **capacity of the channel** is the quantity

$$\mathrm{cap} := \mathrm{cap}(P) = \sup_{\mu \in \mathcal{M}_1(\mathbb{X})} I(X : Y).$$

## Remark

For the moment significance of capacity unclear. Two main results can be established:

1. if transmission rate $R < \mathrm{cap}$, possible to transmit information with arbitrarily small error;
2. if $R > \mathrm{cap}$, impossible to make transmission error vanish.

 +  = no problem

 +  = no problem

 +  = no problem

 +  = problem.

UNIVERSITÉ DE RENNES 1

# Channel capacity (cont'd)

**Proposition**

$(\mathbb{X}, \mathbb{Y}, P)$ *noiseless channel with capacity* cap.

1. $\mathrm{cap} \geq 0$.
2. $\mathrm{cap} \leq \log \mathrm{card}\mathbb{X}$.
3. $\mathrm{cap} \leq \log \mathrm{card}\mathbb{Y}$.

Fix reasonable decoding rule $\Delta : \mathbb{Y} \to \mathbb{X}$ guessing .

**Definition**

Channel $(\mathbb{X}, \mathbb{Y}, P)$. The decision rule = guessing the emitted symbol $x \in \mathbb{X}$ when the received symbol is $y \in \mathbb{Y}$. The rule

$$\Delta(y) \in \arg\max_{z \in \mathbb{X}} \mathbb{P}(X = z | Y = y), y \in \mathbb{Y}$$

is called **of maximum likelihood** decision rule.

# Channel capacity
## Example

**Example**

Channel $(\mathbb{X}, \mathbb{Y}, P)$ with $\mathbb{X} = \{x_1, x_2, x_3\}$, $\mathbb{Y} = \{y_1, y_2, y_3\}$, and

$$P_{xy} := \mathbb{P}(Y = y | X = x) = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}.$$

3-C: Work out this example.

**Definition**

A **decision rule** is a stochastic kernel

$$K_\Delta : \mathbb{Y} \times \mathbb{X} \to [0, 1]$$

assigning to every received symbol $y \in \mathbb{Y}$ a probability $K_\Delta(y, x)$ for every possibly emitted symbol $x \in \mathbb{X}$.

$\Delta(y) = \underset{z}{\operatorname{argmax}} \quad \mathbb{P}(X = z \mid Y = y)$

Caution: $\neq P_{zy}, \neq P_{yz}$

$$P = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

$\qquad\quad 1 \qquad 0.9 \qquad 1.1$

$$\hat{P}_{yz} = \mathbb{P}(X = z \mid Y = y) = \frac{\mathbb{P}(X = z, Y = y)}{\mathbb{P}(Y = y)} = \frac{\mu(z)\, P_{zy}}{\sum_w \mu(w)\, P_{wy}}$$

Let $\mu$ = unif $\qquad \hat{P}_{yz} = \dfrac{P_{zy}}{\sum_w P_{wy}} = (\hat{P}^t)_{zy}$

$$\hat{P}^t = \begin{bmatrix} 0.5 & 0.333.. & 0.1818.. \\ 0.2 & 0.33.. & 0.4545.. \\ 0.3 & 0.33.. & 0.3636 \end{bmatrix} \qquad \hat{P} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 1/3 & 1/3 & 1/3 \\ 0.18 & 0.45 & 0.36 \end{bmatrix}$$

| $y$ | $\operatorname{argmax} \hat{P}_{zy}$ |
|---|---|
| $y_1$ | $\{x_1\}$ |
| $y_2$ | $\{x_1, x_2, x_3\}$ |
| $y_3$ | $\{x_2\}$ |

$x^* \in \operatorname{argmax}$

$\Delta(y) = x^*$

# Coding of a noisy channel

Want to transmit $n$-letter words over $\mathbb{X}$ through memoryless channel $(\mathbb{X}, \mathbb{Y}, P)$. Consider

- either channel $(\mathbb{X}, \mathbb{Y}, P)$ trasmitting (sequentially, i.e. letter by letter) *random words* $\mathbf{X} = X_1 \cdots X_n \in \mathbb{X}^n$ towards *random words* $\mathbf{Y} = Y_1 \cdots Y_n \in \mathbb{Y}^n$ according to the transmission matrix $P$,

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} P(x_i, y_i) =: Q_n(\mathbf{x}, \mathbf{y}),$$

$\underline{x} = (x_1 \cdots x_n)$

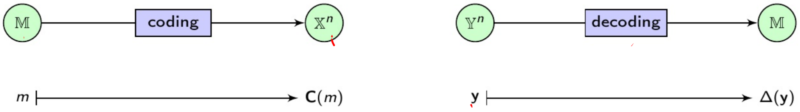
- or extended channel to $(\mathbb{X}^n, \mathbb{Y}^n, Q_n)$, where $Q_n$ is the transmission matrix between $\mathbb{X}^n$ and $\mathbb{Y}^n$ trasmitting (globally) *random words* $\mathbf{X} \in \mathbb{X}^n$ towards *random words* $\mathbf{Y} \in \mathbb{Y}^n$ according to the transmission matrix $Q_n$,

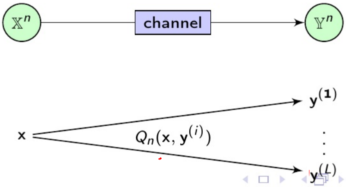
$$\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) =: Q_n(\mathbf{x}, \mathbf{y}),$$

# Coding of a noisy channel (cont'd)

## Actions considered separately

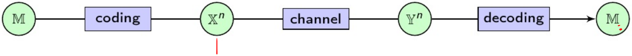
- Sets of words $\mathbb{X}^n$ and $\mathbb{Y}^n$ not of direct interest.
- Set of messages $\mathbb{M}$ encoded into words of $\mathbb{X}^n$ and words of $\mathbb{Y}^n$ decoded into messages of $\mathbb{M}$.
- Two mappings
    - coding $\mathbb{M} \ni m \mapsto \mathbf{C}(m) \in \mathbb{X}^n$ and
    - decoding (deterministic or random) rule $\mathbb{Y}^n \ni \mathbf{y} \mapsto \Delta(\mathbf{y}) \in \mathbb{M}$.

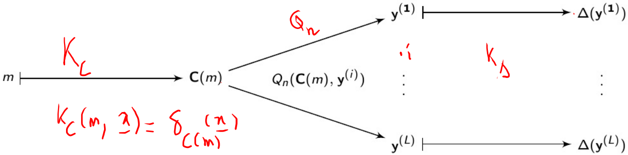


# Coding of a noisy channel (cont'd)

## Actions considered sequentially




$\nu^{\delta_m}(m') = \sum_{v \in M} \sum_{x \in \mathbb{X}^n} \sum_{y \in \mathbb{Y}^n} \delta_m(v) K_C(v, x) \times Q_n(x, y) K_\Delta(y, m')$

Net effect of channel: transform input message $M = m$ (distributed according to $\mu = \delta_m$) into random variable $M' \in \mathbb{M}$ of law $\nu^{\delta_m}$,

$$
\begin{aligned}
\nu^{\delta_m}(m') &= \mathbb{P}_m(M' = m') = \mathbb{P}(M' = m' | M = m) \\
&= \sum_{v \in \mathbb{M}} \mathbb{P}(M' = m' | M = v) \delta_m(v) = K_{\mathbf{C}} Q_n K_\Delta(m, m') \\
&= \sum_{\mathbf{x} \in \mathbb{X}^n} \sum_{\mathbf{y} \in \mathbb{Y}^n} K_{\mathbf{C}}(m, \mathbf{x}) Q_n(\mathbf{x}, \mathbf{y}) K_\Delta(\mathbf{y}, m') \\
&= \sum_{\mathbf{y} \in \mathbb{Y}^n} Q_n(\mathbf{C}(m), \mathbf{y}) \delta_{\Delta(\mathbf{y}), m'}.
\end{aligned}
$$

3-D : Explain steps.

# Channel capacity
## Transmission error

- **Individual transmission error**

$$e(m) := e^{(n)}(m) = \mathbb{P}_{\delta_m}(M' \neq m) = \sum_{m' \neq m} \sum_{\mathbf{y} \in \mathbb{Y}^n} Q_n(\mathbf{C}(m), \mathbf{y}) \delta_{\Delta(\mathbf{y}), m'}$$

$$= \sum_{\mathbf{y} \in \mathbb{Y}^n} Q_n(\mathbf{C}(m), \mathbf{y}) \mathbb{1}_{\mathbb{M} \setminus \{m\}}(\Delta(\mathbf{y})).$$

- **Maximal transmission error**

$$e_{\max} := e^{(n)}_{\max} = \max_{m \in \mathbb{M}} e^{(n)}(m).$$

- **Mean transmission error**

$$\overline{e} := \overline{e}^{(n)} = \sum_{m \in \mathbb{M}} \mu(m) e^{(n)}(m).$$

# Channel capacity
## Bloc codes

### Definition

An $[n, k]$-**bloc code** (with $k$ and $n$ integers $\geq 1$) for a discrete memoryless channel $(\mathbb{X}, \mathbb{Y}, P)$ is the triple $(\mathbb{M}, \mathbf{C}, \Delta)$, where

- $\mathbb{M}$ is the set of messages with $\operatorname{card}\mathbb{M} = k$,
- $\mathbf{C} : \mathbb{M} \to \mathbb{X}^n$ is the bloc coding of size $n$,
- $\Delta : \mathbb{Y}^n \to \mathbb{M}$ is the decoding.

Denote $\mathcal{K}$, more precisely $\mathcal{K}(n, k)$ (or simpy $[n, k]$), such a code. The image $\mathbf{C}(\mathbb{M}) \subseteq \mathbb{X}^n$ is the **glossary of the code** $\mathcal{K}$.

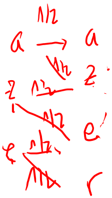
### Definition

Let $\mathcal{K}$ an $[n, k]$ bloc code.

- **Transmission rate** $R$
$$R := R[\mathcal{K}] = \frac{\log_{|\mathbb{X}|} k}{n}.$$
- A transmission rate $R$ is **attainable** if exists sequence $(\mathcal{K}_\ell)_{\ell \in \mathbb{N}}$ of $[n_\ell, k_\ell]$-bloc codes, such that
$$\lim_{l \to \infty} \frac{\log_{|\mathbb{X}|} k_\ell}{n_\ell} \to R \text{ and } \lim_{l \to \infty} e_{\max}[\mathcal{K}_\ell] = 0.$$

Layout of Serbian keyboard
(at least as simulated on my computer)

$M = \{a, z, e, r, \dots\}$




$M = X \to Y = M$

Layout of French keyboard


Original messages

$\{a, e, t, \dots\}$

By restricting the set $M$ the channel becomes lossless

Error correction

$p = 0.1$

$0 \longmapsto 000$ 

$1 \longmapsto 111$

010 $\to 0$

001 $\to 0$

110 $\to 1$

Residual error $R_3$ code

if $p = 0.1$ is 0.03

$R_{61} \to 10^{-15}$

# Channel capacity
## Fundamental theorem of transmission

**Theorem (Shannon theorem for transmission)**

*Let* $(\mathbb{X}, \mathbb{Y}, P)$ *be a memoryless channel with capacity* $\mathrm{cap} := \mathrm{cap}(P)$.

- *For every* $R < \mathrm{cap}$, *exists sequence* $(\mathcal{K}_\ell)_{\ell\in\mathbb{N}}$ *of* $[n_\ell, k_\ell]$*-codes, with transmission rates* $R_\ell := \frac{\log_{|\mathbb{X}|} k_\ell}{n_\ell} \to R$, *such that* $\lim_{\ell\to\infty} \overline{\mathrm{e}}[\mathcal{K}_\ell] = 0$.
- *Conversely, for every* $R > \mathrm{cap}$ *and every sequence* $(\mathcal{K}_\ell)_{\ell\in\mathbb{N}}$ *of* $[n_\ell, k_\ell]$*-codes with blocs of increasing size (i.e.* $n_1 < n_2 < n_3 < \ldots$*) and transmission rates* $R_\ell \geq R$, *we have* $\lim_{\ell\to\infty} \overline{\mathrm{e}}[\mathcal{K}_\ell] = 1$.

# Channel capacity
## Fundamental theorem of transmission (cont'd)

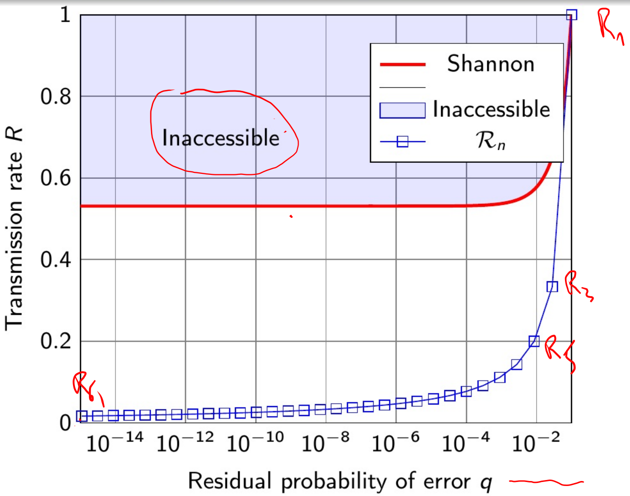



Figure: For binary symmetric channel, with error rate $p = 0.1$, red curve is the Shannon boundary. Blue marks = transmission rates vs. residual error probability for the family of repetition ECC $\mathcal{R}_n$, with $n = 1, 3, 5, \ldots, 61$.

# Channel capacity

## Exercise: capacity of the "sum" of two channels

**Exercise**

- Let $\mathcal{K}_i = (\mathbb{X}_i, \mathbb{Y}_i, P_i)$, $i = 1, 2$ be two channels.
- Denote by $\mathbb{X} = \mathbb{X}_1 \boxplus \mathbb{X}_2$ (if $\mathbb{X}_1$ and $\mathbb{X}_2$ are distinct then $\mathbb{X}_1 \boxplus \mathbb{X}_2 = \mathbb{X}_1 \sqcup \mathbb{X}_2$; else, start by distinguishing artificially the elements of $\mathbb{X}_1$ and $\mathbb{X}_2$ before taking their union.)
- Similarly for $\mathbb{Y} = \mathbb{Y}_1 \boxplus \mathbb{Y}_2$.
- Transmission matrix of the "sum" is the bloc matrix $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$.
- $X$ a $\mathbb{X}$-valued r.v. whose law described by $\pi \in \mathrm{PV}_{\mathbb{X}}$ and $Y$ a $\mathbb{Y}$-valued r.v. whose law determined by the cannel.

1. Compute $H(\pi)$. *Hint:* Let $p = \sum_{x\in\mathbb{X}_1} \pi(x)$ (hence $1 - p = \sum_{x\in\mathbb{X}_2} \pi(x)$.)
2. Compute $H(X|Y)$.
3. Consider r.v. $X_1$ and $X_2$ with values in $\mathbb{X}_1$ and $\mathbb{X}_2$ and laws $\rho_1$ and $\rho_2$; denote by $Y_1$ and $Y_2$ the restrictions to $\mathbb{Y}_1$ and $\mathbb{Y}_2$ of $Y$. Show that $H(X|Y) = pH(X_1|Y_1) + (1-p)H(X_2|Y_2)$ and conclude that
$$C(p) := \sup_{\pi : \sum_{x\in\mathbb{X}_1} \pi(x) = p} I(X : Y) = H(p, 1-p) + pC_1 + (1-p)C_2.$$
4. Show that $\arg\max_p C(p)$ is $p = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}$.
5. Conclude that capacity of "sum" channel reads $2^C = 2^{C_1} + 2^{C_2}$.