A Converse Bound on the Mismatched Distortion-Rate Function

Maël Le Treust, CNRS IRISA Rennes

joint work with Tristan Tomala, HEC Paris

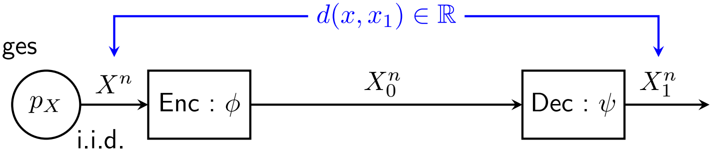
Entropies et divergences : modélisation . statistique . algorithmique

Caen, 16th May 2024

Lossy Source Coding Problem - Case of a Noiseless Channel

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$|\mathcal{X}_0|$: finite number of available messages

distortion fonction :



(problem) $$D^n = \min_{\phi,\psi} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d(X_t, X_{1,t})\right]$$

(solution) $$D^\star = \min_{\substack{p_{X_1|X} \\ I(X;X_1) \leq \log_2 |\mathcal{X}_0|}} \mathbb{E}\Big[d(X, X_1)\Big]$$

distortion-rate function

with $I(X;X_1) = D_{KL}\big(p_X p_{X_1|X} \big|\big| p_X p_{X_1}\big)$
Kullback-Leibler divergence

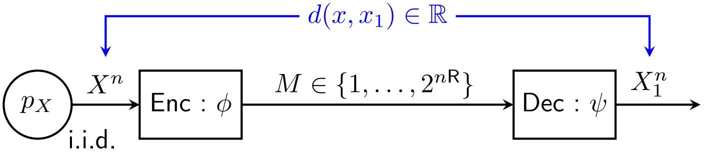
Theorem [Shannon 1959]

$$\lim_{n \to +\infty} D^n = \inf_{n \in \mathbb{N}} D^n = D^\star$$

Lossy Source Coding Problem - Distortion-Rate Trade-off

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$\mathsf{R} \geq 0$: growth rate number of messages

distortion fonction :



(problem) $$D^n(\mathsf{R}) = \min_{\phi,\psi} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d(X_t, X_{1,t})\right]$$

(solution) $$D^\star(\mathsf{R}) = \min_{\substack{p_{X_1|X} \\ I(X;X_1) \leq \mathsf{R}}} \mathbb{E}\Big[d(X, X_1)\Big]$$

distortion-rate function

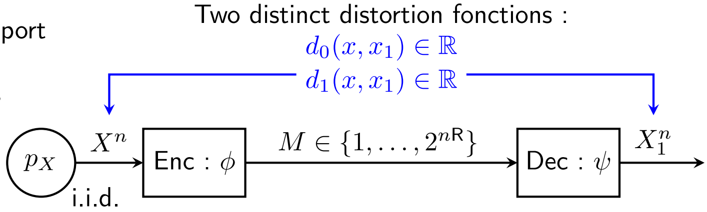
with $I(X;X_1) = D_{KL}\big(p_X p_{X_1|X} \big|\big| p_X p_{X_1}\big)$
Kullback-Leibler divergence

Theorem [Shannon 1959]

$$\forall\, \mathsf{R} \geq 0, \qquad \lim_{n \to +\infty} D^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D^n(\mathsf{R}) = D^\star(\mathsf{R})$$

Mismatch Distortion-Rate Problem

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$\mathsf{R} \geq 0$: growth rate number of messages



(problem) $D_1^n(\mathsf{R}) = \inf_{\psi} \max_{\phi \in \mathcal{A}(\psi)} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_1(X_t, X_{1,t})\right]$ $\qquad$ $\mathcal{A}(\psi) = \operatorname{argmin} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_0(X_t, X_{1,t})\right]$

Theorem [Lapidoth 1997] : achievability bound

$$\forall\ \mathsf{R} \geq 0, \qquad \lim_{n \to +\infty} D_1^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D_1^n(\mathsf{R}) \leq D_1^\star(\mathsf{R})$$

Achievability bound of [Lapidoth 1997]

$$\mathsf{D}_1^\star(\mathsf{R}) = \inf_{p_{X_0},\, p_{X_1|X_0}} \max_{\substack{f_{XX_0} \\ \in \mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R})}} \mathbb{E}\Big[d_1(X, X_1)\Big]$$

where X_0 codebook random variable

$$\mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R}) = \underset{f_{XX_0} \in \mathcal{D}(p_{X_0}, \mathsf{R})}{\operatorname{argmin}} \mathbb{E}\Big[d_0(X, X_1)\Big]$$

$$\mathcal{D}(p_{X_0}, \mathsf{R}) = \Big\{ f_{XX_0} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}_0), \quad f_X = p_X, \quad f_{X_0} = p_{X_0}, \quad I(X; X_0) \leq \mathsf{R} \Big\}$$

Random coding argument :

- Fix distributions p_{X_0} f_{XX_0} and $p_{X_1|X_0}$
- Generate codebook $(X_0^n(m))_{m \in \{1, \ldots, 2^{n\mathsf{R}'}\}}$ i.i.d. according to p_{X_0}
- Decoding function ψ : generates X_1^n i.i.d. according to $p_{X_1|X_0}$ for $X_0^n(m)$
- Knowing ψ and x^n, encoding function ϕ selects $m \in \operatorname{argmin} \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^n d_0(x_t, X_{1,t})\Big]$

$\rightarrow$ Main difficulty : study of the empirical distribution induced by this specific encoding function ϕ

Strategic Communication - Sender Receiver Game

Four scenarios :

1. Cooperative scenario :
   joint selection (ϕ, ψ) (Shannon 1959)

2. Encoder decides first : Bayesian persuasion (Kamenica and Gentzkow 2011)

3. Decoder decides first : Mechanism design (Jackson and Sonnenschein 2007)

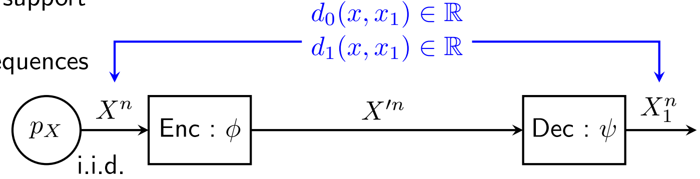
4. Simultaneous decision : (ϕ, ψ) form Nash equilibrium (Crawford and Sobel 1982)

Les tricheurs, Le Caravage, 1595

Large Information Rate - Mechanism Design in Game Theory

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$\mathsf{R} = \log_2 |\mathcal{X}|$: #messages equal #source sequences

Two distinct distortion fonctions :



(solution) $$\overline{D_1} = \min_{p_{X_1|X} \in \overline{\mathcal{G}}} \mathbb{E}\Big[d_1(X, X_1)\Big]$$

where $\overline{\mathcal{G}}$ set of distributions $p_{X_1|X}$ that satisfy $\forall f_{X'|X}$ s.t. $\sum_x p_X(x) f_{X'|X}(\cdot|x) = p_X$:

$$\sum_{x,x_1} p_X(x) p_{X_1|X}(x_1|x) d_0(x, x_1) \leq \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) p_{X_1|X}(x_1|x') d_0(x, x_1)$$

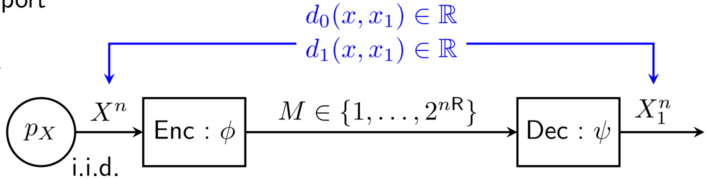
Theorem [Jackson and Sonnenschein 2007]

$$\mathsf{R} = \log_2 |\mathcal{X}| \quad \Longrightarrow \quad \lim_{n \to +\infty} D_1^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D_1^n(\mathsf{R}) = \overline{D_1}$$

Our Contribution : New Converse Bound

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$\mathsf{R} \geq 0$: growth rate number of messages

Two distinct distortion fonctions :



(problem) $D_1^n(\mathsf{R}) = \inf_{\psi} \max_{\phi \in \mathcal{A}(\psi)} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_1(X_t, X_{1,t})\right]$ $\qquad \mathcal{A}(\psi) = \operatorname{argmin} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_0(X_t, X_{1,t})\right]$

Theorem [Le Treust and Tomala 2024] : converse bound

$$\forall\, \mathsf{R} \geq 0, \qquad \widetilde{\mathsf{D}}_1(\mathsf{R}) \leq \lim_{n \to +\infty} D_1^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D_1^n(\mathsf{R}) \leq D_1^\star(\mathsf{R})$$

New Converse Bound

$$\widetilde{\mathsf{D}}_1(\mathsf{R}) = \min_{p_{X_1|X} \in \mathcal{G}(\mathsf{R})} \mathbb{E}_{p_{X_1|X}}\Big[d_1(X, X_1)\Big]$$

$\mathcal{G}(\mathsf{R})$ set of distributions $p_{X_1|X}$ that satisfy :

1. Information constraint :

$$I(X; X_1) \leq \mathsf{R}$$

2. Mismatched encoding constraint :

$$\sum_{x,x_1} p_X(x) p_{X_1|X}(x_1|x) d_0(x, x_1) \leq \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) p_{X_1|X}(x_1|x') d_0(x, x_1), \qquad (1)$$

$$\forall f_{X'|X} \text{ s.t. } \sum_x p_X(x) f_{X'|X}(\cdot|x) = p_X$$

3. Tie-breaking rule : $\forall f_{X'|X}$ satisfying (1) with equality :

$$\sum_{x,x_1} p_X(x) p_{X_1|X}(x_1|x) d_1(x, x_1) \geq \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) p_{X_1|X}(x_1|x') d_1(x, x_1)$$

Sketch of Proof of the Converse Bound

Fix blocklength n, decoding ψ, and encoding $\phi \in \mathcal{A}(\psi)$ that maximizes $\mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} d_1\big(X_t, X_{1,t}\big)\Big]$

Average distribution :

$$\forall (x, x_1), \qquad q_{XX_1}(x, x_1) = \frac{1}{n}\sum_{t=1}^{n} p_{X_t X_{1,t}}(x, x_1) \qquad \leftarrow \text{ marginal of } p_X^{\otimes n}(x^n)\phi(m|x^n)\psi(x_1^n|m)$$

1. Information constraint : $n\mathsf{R} \geq I(X^n; M) \geq I(X^n; X_1^n) \geq \sum_{t=1^n} I(X_t; X_{1,t}) \geq nI(X; X_1)$
2. Take $f_{X'|X}$ s.t. $\sum_x p_X(x) f_{X'|X}(\cdot|x) = p_X$ and construct encoding $\phi' = \phi \circ f_{X'|X}^{\otimes n}$

$$\sum_{x,x_1} p_X(x) q_{X_1|X}(x_1|x) d_0(x, x_1) = \mathbb{E}_\phi\Bigg[\frac{1}{n}\sum_{t=1}^{n} d_0\big(X_t, X_{1,t}\big)\Bigg]$$

$$\leq \mathbb{E}_{\phi'}\Bigg[\frac{1}{n}\sum_{t=1}^{n} d_0\big(X_t, X_{1,t}\big)\Bigg] = \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) q_{X_1|X}(x_1|x') d_0(x, x_1) \qquad (2)$$

3. Tie-breaking rule $\forall f_{X'|X}$ satisfying (2) with equality :

$$\sum_{x,x_1} p_X(x) q_{X_1|X}(x_1|x) d_1(x, x_1) \geq \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) q_{X_1|X}(x_1|x') d_1(x, x_1)$$

Cases Where the Two Bounds Match

1. Large information rate : $\mathsf{R} = H(X)$
2. Binary uniform source and binary reconstruction : $|\mathcal{X}| = |\mathcal{X}_1| = 2$
3. Distortion function only depends on the symbols X_1 : $d_0(x, x_1) = d_0(x_1)$
4. Zero-sum distortion functions : $d_0(x, x_1) + d_1(x, x_1) = 0$

Theorem [Le Treust and Tomala 2024]

Assume that one of the four above hypothesis is satisfied :

$$\forall\, \mathsf{R} \geq 0, \qquad \widetilde{\mathsf{D}}_1(\mathsf{R}) = \lim_{n \to +\infty} D_1^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D_1^n(\mathsf{R}) = D_1^\star(\mathsf{R})$$

1. Large information rate : $\mathsf{R} = H(X)$

Idea : Identify codebook random variable $X_0 = X$: the encoder truthfully reveals the source

$$\widetilde{\mathsf{D}}_1(\mathsf{R}) = \min_{p_{X_1|X} \in \mathcal{G}(\mathsf{R})} \mathbb{E}_{p_{X_1|X}} \Big[d_1(X, X_1) \Big]$$

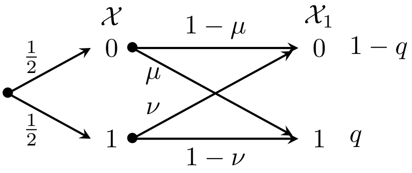
1. Take distribution $p^{\star}_{X_1|X} \in \mathcal{G}(\mathsf{R})$
2. Identify codebook random variable $X_0 = X$
3. Construct encoding ϕ : one-to-one mapping source to codeword : $X^n \to X_0^n(M)$

$$\widetilde{\mathsf{D}}_1(\mathsf{R}) \;=\; \mathbb{E}_{p^{\star}_{X_1|X}} \Big[d_1(X, X_1) \Big] = \max_{\substack{f_{X X_0} \\ \in \mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R})}} \mathbb{E}_{f_{X X_0} p_{X_1|X_0}} \Big[d_1(X, X_1) \Big] \geq \mathsf{D}_1^{\star}(\mathsf{R})$$

2. Binary Uniform Source and Binary Reconstruction : $|\mathcal{X}| = |\mathcal{X}_1| = 2$

Idea : Identify codebook random variable $X_0 = X_1$: the encoder tells decoder which symbol to output

Take $p^\star_{X_1|X} \in \mathcal{G}(\mathsf{R})$ with $(\mu, \nu) \in [0,1]^2$ such that $I(X;X_1) = \mathsf{R}$



Let $\sigma_{X'|X}$ be the permutation of the binary state.

$$\mathcal{D}(p^\star_{X_1}, \mathsf{R}) = \mathrm{conv}\left\{ \{p_X p^\star_{X_1|X}\} \cup \{p_X \sigma_{X'|X} p^\star_{X_1|X}\} \right\} = \mathrm{conv}\left\{ \begin{pmatrix} \frac{1-\mu}{2} & \frac{\mu}{2} \\ \frac{\nu}{2} & \frac{1-\nu}{2} \end{pmatrix}, \begin{pmatrix} \frac{\nu}{2} & \frac{1-\nu}{2} \\ \frac{1-\mu}{2} & \frac{\mu}{2} \end{pmatrix} \right\}$$

$f_{XX_0} = p_X p^\star_{X_1|X}$ belongs to $\mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R})$ and achieves the maximum in

$$\max_{\substack{f_{XX_0} \\ \in \mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R})}} \mathbb{E}_{f_{XX_0} \mathbb{1}_{X_1|X_0}} \Big[d_1(X, X_1) \Big] = \widetilde{\mathsf{D}}_1(\mathsf{R}) \geq \mathsf{D}^\star_1(\mathsf{R}).$$

3. Distortion function only depends on the symbols $X_1 : d_0(x, x_1) = d_0(x_1)$

This is the "transparent motives" assumption in Game Theory literature

Pessimistic tie-breaking rule $\rightarrow$ the worst encoding function for $\mathbb{E}[d_1(X, X_1)]$ given p_{X_1} is selected

Take distribution $p_{X_1|X} \in \mathcal{G}(\mathsf{R})$, tie-breaking rule reformulates as

$$\sum_{x,x_1} p_X(x) p_{X_1|X}(x_1|x) d_1(x, x_1) \geq \sum_{x,x',x_1} p_X(x) f_{X'|X}(x'|x) p_{X_1|X}(x_1|x') d_1(x, x_1), \quad \forall f_{X'|X} \in \mathcal{H}$$

The solution reformulates

$$\mathsf{D}_1^\star(\mathsf{R}) = \min_{p_{X_1} \in \mathcal{P}(\mathcal{X}_1)} \mathbb{E}_{p_X p_{X_1}} \Big[d_1(X, X_1) \Big] = \widetilde{\mathsf{D}}_1(\mathsf{R}).$$

4. Zero-sum distortion functions : $d_0(x, x_1) + d_1(x, x_1) = 0$

By hypothesis, $d_0(x, x_1) = -d_1(x, x_1)$

Pessimistic tie-breaking rule $\rightarrow$ mismatched encoding constraint is equivalent to tie-breaking rule

Take distribution $p_{X_1|X} \in \mathcal{G}(\mathsf{R})$, then

$$\mathcal{F}(p_{X_0}, p_{X_1|X_0}, \mathsf{R}) = \underset{f_{XX_0} \in \mathcal{D}(p_{X_0}, \mathsf{R})}{\operatorname{argmax}} \mathbb{E}_{f_{XX_0} p_{X_1|X_0}} \Big[d_1(X, X_1) \Big]$$

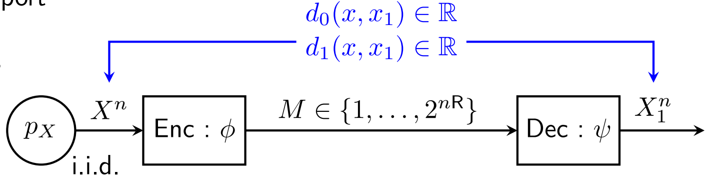
Both solutions reformulate as

$$\widetilde{\mathsf{D}}_1(\mathsf{R}) = \min_{p_{X_1} \in \mathcal{P}(\mathcal{X}_1)} \mathbb{E}_{p_X p_{X_1}} \Big[d_1(X, X_1) \Big] = \mathsf{D}_1^{\star}(\mathsf{R})$$

Take Away - Mismatch Distortion-Rate Problem

p_X : probability distribution finite support
$n \in \mathbb{N}^\star$: symbols block-length
$\mathsf{R} \geq 0$: growth rate number of messages

Two distinct distortion fonctions :



(problem) $$D_1^n(\mathsf{R}) = \inf_{\psi} \max_{\phi \in \mathcal{A}(\psi)} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_1(X_t, X_{1,t})\right] \qquad \mathcal{A}(\psi) = \operatorname{argmin} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_0(X_t, X_{1,t})\right]$$

(single-letter) $$\widetilde{\mathsf{D}}_1(\mathsf{R}) = \min_{p_{X_1|X} \in \mathcal{G}(\mathsf{R})} \mathbb{E}\Big[d_1(X, X_1)\Big]$$ lower bound mismatch distortion-rate function

Theorem [Le Treust and Tomala 2024]

Assume that one of the four above hypothesis is satisfied :

$$\forall\, \mathsf{R} \geq 0, \qquad \widetilde{\mathsf{D}}_1(\mathsf{R}) = \lim_{n \to +\infty} D_1^n(\mathsf{R}) = \inf_{n \in \mathbb{N}} D_1^n(\mathsf{R}) = D_1^\star(\mathsf{R})$$