

# Approches comparées des méthodes statistiques mobilisant des divergences dans le cadre de l'estimation ou les tests statistiques

Emmanuelle Gautherat

CRIEG-REGARDS, Université de Reims Champagne Ardenne et MODAL'X,  
Université Paris Nanterre

Workshop : Entropies et divergences : modélisation .  
statistique . algorithmique 14-17 mai 2024

# Sommaire

- ① Entropie et notations
- ②  $I_{\varphi^*}$ -divergence
- ③ Dualité
- ④ Région de confiance
- ⑤ Bibliographie
- ⑥ Annexe

# Entropie : propriétés

Initialement, mesure de l'information <sup>1</sup> dans le cas discret fini ... avec quelques propriétés

$P, Q, R$  probabilités discrètes de cardinal du support  $n$  ;  $p_k$  est la probabilité associée à  $k^e$  événement,  $k = 1, \dots, n$  ; et  $PQ$  est une loi jointe ayant pour marges  $P$  et  $Q$ .

On note  $H(P) = H(p_1, p_2, \dots, p_n)$

---

<sup>1</sup> suivant les lignes de J.-F. Bercher, LIGM, UMR 8049, ESIEE/Université Paris-Est, elles même extraites de « Méthodes d'inversion appliquées au traitement du signal et de l'image », ISBN 2746245485, J.-F. Giovannelli et J. Idier éditeurs, Hermès-Lavoisier, décembre 2013.

# Entropie : quelques propriétés

- 1 symétrie :  $H(p_1, p_2, \dots, p_n)$  ne dépend pas de l'ordre des événements
- 2  $p \mapsto H(p, 1 - p)$  continue
- 3  $H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$
- 4 récursivité (branchement) :  $H_{n+1}(p_1q_1, p_1q_2, p_2, \dots, p_n) = H_n(p_1, p_2, \dots, p_n) + p_1H_2(q_1, q_2)$
- 5 expansible :  $H_{n+1}(p_1, p_2, \dots, p_n, 0) = H_n(p_1, p_2, \dots, p_n)$
- 6 sous-additivité :  $H(PQ) \leq H(P) + H(Q)$  et additivité dans le cas indépendant :  $H(PQ) = H(P) + H(Q)$
- 7 sous-additivité conditionnelle :  
 $H(PQ | R) \leq H(P | R) + H(Q | R)$
- 8 récursivité généralisée :  
 $H_{n+1}(p_1q_1, p_1q_2, p_2, \dots, p_n) = H_n(p_1, p_2, \dots, p_n) + p_1H_2(q_1, q_2)$ .

# Entropie : quelques propriétés

- Les 4 premiers postulats - axiomes de Faddeev [**Fad56**]  
- suffisent à caractériser l'entropie de Shannon

$$H(P) = - \sum_{i=1}^n p_i \log p_i.$$

- Levé du postulat de récursivité (prop. 4) et ajout d'une exigence d'additivité : les solutions possibles incluent en particulier l'entropie de Rényi.

# Entropie : quelques propriétés

- Remplacement de récursivité (prop. 4) par récursivité générale (prop. 8) avec  $m(p_1 p_2) = m(p_1) m(p_2)$ 
  - en particulier  $m(p) = p^q$ , conduit à l'entropie d'ordre  $q$ ,  $H_q(P) = \frac{1}{2^{1-q}-1} (\sum_{i=1}^n p_i^q - 1)$  introduite par Havrda et Charvat [**Hav67**], indépendamment par Daróczy [**Dar70**], puis redécouverte dans le champ de la physique statistique par C. Tsallis [**Tsa88**].
  - Pour  $q \geq 1$ , ces entropies sont sous-additives (prop. 6), mais ne sont pas additives.
  - Dans le cas  $q = 1$ , l'entropie d'ordre  $q = 1$  n'est autre que l'entropie de Shannon.

## Entropie : quelques propriétés

- Dans l'axiomatique de Faddeev, Rényi a proposé de remplacer le postulat de récursivité (prop. 4) par la propriété d'additivité, et d'ajouter une propriété de moyenne de l'entropie, qui spécifie que l'entropie de l'union de deux distributions de probabilité incomplètes est égale à la moyenne pondérée des entropies des deux distributions.
  - Lorsque la moyenne utilisée est une moyenne arithmétique : unique solution, entropie de Shannon.
  - Lorsque la moyenne utilisée est une moyenne exponentielle : entropie de Rényi  $H_q(P) = \frac{1}{1-q} \log \sum_{i=1}^n p_i^q$ .

## Entropie : quelques propriétés

- Une autre manière (d'appréhender l'entropie de Rényi) est : l'entropie de Shannon est la moyenne arithmétique, avec les poids  $p_i$ , des informations élémentaires  $I_i = -\log p_i$  associées aux différents événements.
  - En remplaçant la moyenne arithmétique par une moyenne de Kolmogorov-Nagumo, l'entropie devient  $H_{\psi}(p_1, \dots, p_n) = \psi^{-1}(\sum p_i \psi(-\log p_i))$ .
  - Sous une condition supplémentaire d'additivité et sous  $\lim_{p \rightarrow 0} H_{\psi}(p, 1-p) = 0$ , alors cette entropie est soit l'entropie de Shannon, soit l'entropie de Rényi, avec  $q \geq 0$ .
  - À nouveau, par la règle de l'Hôpital, on retrouve l'entropie de Shannon pour  $q = 1$ .
  - Par ailleurs, pour  $q = 0$ , l'entropie de Rényi devient l'entropie de Hartley, le logarithme du nombre d'événements de probabilité non nulle.



# Entropie dans le cas continu

Dans le cas continu, on définit l'entropie de Shannon associée à une densité  $f(x) = \frac{dP}{d\mu}$  au sens Radon-Nikodym par

$$H(P) = - \int f(x) \log f(x) dx,$$

Cela correspond presque au passage la limite de l'entropie de Shannon dans le cas discret, à une constante additive tendant éventuellement vers l'infini près, voir [**Pap81**]  
-> Entropie différentielle.

## Entropie dans le cas continu

Jaynes [**Jay63**] explicite la mesure  $\mu$  sous-jacente avec  $P \ll \mu$  : l'entropie différentielle correspondante ( $\mu$ -Entropie) est alors définie par  $H_{\mu}(P) = - \int f(x) \log \frac{f(x)}{\mu(x)} d\mu(x)$ .

Par ailleurs, on remarque que

$$\begin{aligned} H(P) &= - \int f(x) \log f(x) dx \\ &= - \int \frac{dP}{d\mu}(x) \log \frac{dP}{d\mu}(x) d\mu(x) \end{aligned}$$

Puis, on explicite la quantité d'information apportée par  $P$  par rapport à ce qui est "chargée" par  $Q$  à la place de  $\mu$  avec  $P$  et  $Q$ , deux distributions de probabilité de densités  $f(x) = \frac{dP}{d\mu}$  et  $g(x) = \frac{dQ}{d\mu}$  par rapport à une mesure commune  $\mu$  et avec  $Q \ll P$ ,

$$H(P/Q) = - \int \frac{dP}{dQ}(x) \log \frac{dP}{dQ}(x) dQ(x)$$

## Entropie dans le cas continu

$$\begin{aligned} H(P/Q) &= - \int \frac{dP}{dQ}(x) \log \frac{dP}{dQ}(x) dQ(x) \\ &= - \int \frac{dP/d\mu(x)}{dQ/d\mu(x)} \log \frac{dP/d\mu(x)}{dQ/d\mu(x)} dQ(x) \\ &= - \int \frac{dP}{d\mu}(x) \log \frac{dP/d\mu(x)}{dQ/d\mu(x)} d\mu(x) \\ &= - \int f(x) \log \frac{f(x)}{g(x)} d\mu(x). \end{aligned}$$

-> Similaire à une divergence de Kullback-Leibler au signe près [**Kul59**]. On note cette divergence par :

$$D(P/Q) = \int f(x) \log \frac{f(x)}{g(x)} d\mu(x)$$

avec la convention  $0 \log 0 = 0$ .

# Entropie dans le cas continu

- Pour  $Q$  uniforme par rapport à  $\mu$ , au signe près la divergence de Kullback  $D(P/Q)$  devient une  $\mu$ -entropie.
- Pour  $\mu$  mesure de Lebesgue  $\rightarrow$  entropie différentielle de Shannon
- Pour  $\mu$  mesure de comptage  $\rightarrow$  entropie de Shannon.

# Entropie dans le cas continu

Versions continues des entropies de Rényi et Tsallis, pour un index  $q \neq 1$ .

Soit  $f(x) = \frac{dP}{d\mu}$

- Entropie de Tsallis

$$S_{q,\mu}(P) = \frac{1}{1-q} \left( \int f(x)^q d\mu(x) - 1 \right)$$

- Entropie de Rényi

$$H_{q,\mu}(P) = \frac{1}{1-q} \log \int f(x)^q d\mu(x)$$

## Entropie dans le cas continu

Comme précédemment, on leur associe une divergence de manière naturelle. Exemple avec la divergence de Rényi. Soient  $f(x) = \frac{dP}{d\mu}$  et  $g(x) = \frac{dQ}{d\mu}$  densités de  $P$  et  $Q$  par rapport à une mesure commune  $\mu$ , avec  $Q \ll P$ . Entropie de Rényi

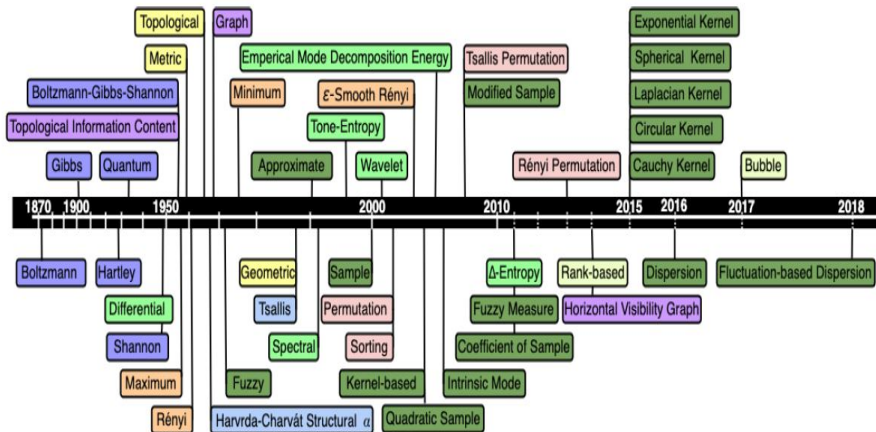
$$H_{q,\mu}(P) = \frac{1}{1-q} \log \int f(x)^q d\mu(x)$$

Divergence de Rényi

$$D_q(P/Q) = \frac{1}{1-q} \log \int f(x)^q g(x)^{1-q} d\mu(x),$$

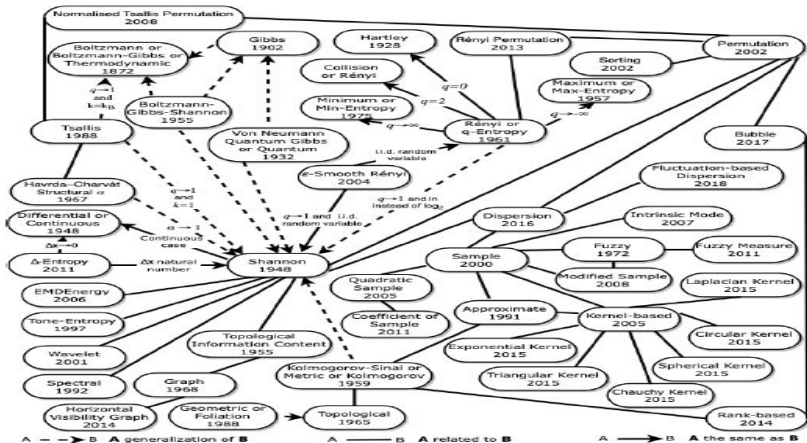
(lorsque  $q \rightarrow 1 \rightarrow$  divergence de Kullback).

# Combien d'entropies ?



Quelques entropies selon le temps, échelle logarithmique,  
 Source : M. Ribeiro et al, *The Entropy Universe*, Entropy 2021, 23, 222

# Qui inspire quoi ?

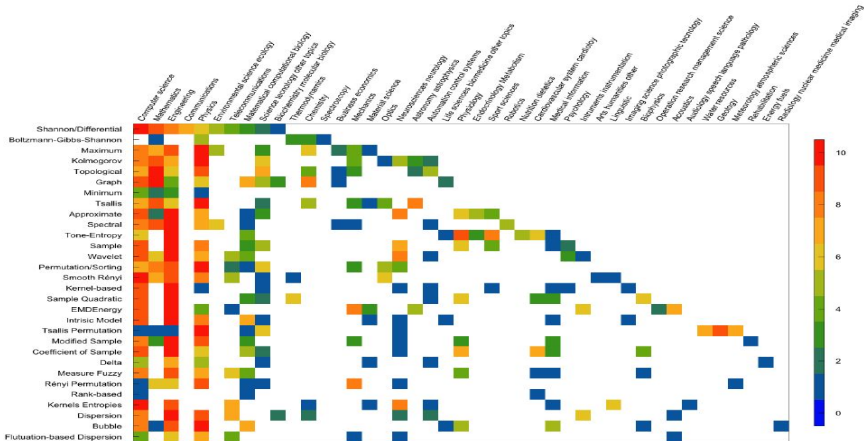


Univers des entropies temporelles

Source : M. Ribeiro et al, *The Entropy Universe*, Entropy 2021, 23, 222



# Selon les domaines



Les dix domaines qui citent le plus chaque "article source" des entropies, selon les domaines de recherche de WoS (cf en annexe selon scopus). Légende : 0 domaine de recherche le moins cité, à 10 domaine de recherche le plus cité

# $I_{\varphi^*}$ -divergence : $\varphi$

Comme vu précédemment, de très nombreuses divergences  $\rightarrow$  sous-classe de divergences convexes :

$I_{\varphi^*}$ -divergence, issues d'une fonction  $\varphi$ .

On désigne par  $\varphi$  toute fonction convexe

- $d(\varphi) = \{x \in \mathbb{R}, \varphi(x) < \infty\}$  domaine de  $\varphi$
- $\inf d(\varphi)$  et  $\sup d(\varphi)$  les points terminaux respectivement du domaine de  $\varphi$
- sa conjuguée convexe  $\varphi^*$  (transformée de Fenchel-Legendre)

$$\varphi^*(x) = \sup_{y \in \mathbb{R}} \{xy - \varphi(y)\}.$$

# $I_{\varphi^*}$ -divergence : $\varphi$

Hypothèses sur  $\varphi$ .

- 1  $\varphi$  est strictement convexe et  $d(\varphi)$  contient un voisinage de 0
- 2  $\varphi$  est deux fois différentiable sur un voisinage de 0
- 3  $\varphi(0) = 0$  et  $\varphi^{(1)}(0) = 0$
- 4  $\varphi^{(2)}(0) > 0$ , (i.e.  $\varphi$  admet un unique minimum en 0).

Ainsi,

- $\varphi^*$  convexe et s.c.i., domaine de définition  $d(\varphi^*)$  non vide
- $\varphi^{(1)}$  est inversible et  $\varphi^*(x) = x \cdot \varphi^{(1)-1}(x) - \varphi(\varphi^{(1)-1}(x))$
- $(\varphi^*)^{(1)} = \varphi^{(1)-1}$  et  $(\varphi^*)^{(2)}(0) = \frac{1}{\varphi^{(2)}(0)}$ .

# $I_{\varphi^*}$ -divergence : $\varphi$

Remarques : plusieurs notations pour  $I_{\varphi^*}$ -divergence.  
Les différences essentielles

- ① l'inversion de  $\varphi$  et  $\varphi^*$  : les hypothèses portent toujours sur  $\varphi$ , sans conséquence car se propagent à la version duale ;
- ② version renormalisée (minimum en 0 au lieu d'en 1) [**Rao91**] ;
- ③ tableau synthétique avec des noms différents (Kullback-Leibler pour l'un, Kullback-Liebler modifié pour l'autre... etc.

Mais identiques en réalité.

# Discrepancies de Cressie-Read

Sous-classe des  $I_{\varphi^*}$ -divergences : très utilisées, les Cressie-Read.

Dans le cas général, les Cressie-Read s'écrivent pour  $\alpha \in \mathbb{R}/\{0, 1\}$

$$\varphi_{\alpha^*}(x) = \frac{(1+x)^\alpha - \alpha x - 1}{\alpha(\alpha-1)}$$

$$\varphi_\alpha(x) = \frac{((\alpha-1)x+1)^{\frac{\alpha}{\alpha-1}} - \alpha x - 1}{\alpha}$$

# Discrepancies de Cressie-Read

Divergences	$\alpha$	$\varphi_{\alpha}(x)$ $\varphi_{\alpha^*}(x)$	$d(\varphi_{\alpha})$ $d(\varphi_{\alpha^*})$
Entropie relative	1	$e^x - 1 - x$ $(x + 1) \log(x + 1) - x$	$\mathbb{R}$ $] - 1, +\infty]$
Kullback	0	$-\log(1 - x) - x$ $x - \log(1 + x)$	$] - \infty, 1[$ $] - 1, +\infty]$
Hellinger	0.5	$x^2/(2 - x)$ $2(\sqrt{(x + 1)} - 1)^2$	$] - \infty, 2[$ $] - 1, +\infty]$
$\chi^2$	2	$x^2/2$ $x^2/2$	$\mathbb{R}$ $\mathbb{R}$

## $I_{\varphi^*}$ -divergence : $\varphi$

Introduite par Csizar [**Csi63**] [**Csi67**] pour des probabilités, et étendues par plusieurs auteurs aux mesures positives, puis aux mesures signées de masse 1 ou non, la  $\varphi^*$ -divergence associée à  $\varphi$ , appliquée à  $Q$  mesure signée et  $P$ , mesure positive, est définie par :

$$I_{\varphi^*}(Q/P) = \begin{cases} \int_{\Omega} \varphi^* \left( \frac{dQ}{dP} - 1 \right) dP & \text{si } Q \ll P \text{ et si défini} \\ +\infty & \text{sinon.} \end{cases}$$

Grace à la convexité de  $\varphi$ ,  $I_{\varphi^*}$  est convexe en  $Q$ , et si de plus  $\varphi^*$  est strictement convexe autour d'un voisinage de 0,

$$I_{\varphi^*}(Q/P) = 0 \iff Q = P.$$

# Dualité : processus empirique

On observe un  $n$ -échantillon  $X_1, \dots, X_n$  de  $\mathbb{R}^q$ , de loi  $P$ . Soit  $P_n$  la mesure empirique définie par

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

où  $\delta_{X_i}$  est la mesure de Dirac.

Dire que  $Q_n \ll P_n$  signifie donc que  $Q_n$  évolue dans  $\mathcal{P}_n$  avec

$$\mathcal{M}_n = \left\{ Q_n = \sum_{i=1}^n q_i \delta_{X_i}, q_i \in \mathbb{R} \right\}.$$



# Dualité : Contraintes

Travaux statistiques avec  $I_{\varphi^*}$  divergence relatifs à l'estimation par région de confiance d'un paramètre  $\theta$  de dimension  $q$ .

ou de tests statistiques qui reviennent également in fine a proposer une région de confiance pour ce paramètre.

# Dualité : Contraintes

Travaux différents selon la forme de  $\theta$  par rapport à  $P$ , la loi de  $X_1$ , et  $n$ .

Exemple :

- espérance d'une fonction  $f$  évoluant dans une certaines classes de fonction  $\mathcal{F} : \mathbb{E}_P(f(X_1, \theta)) = 0$  ;
- éventuellement,  $q \geq n$  ou du même ordre ;
- extension à des paramètres de contraintes fonctionnelles plus généraux  $\theta = T(P)$  avec  $T$  de  $M \subset L_{\infty}(F)$  dans  $\mathbb{R}^q$  Hadamard différentiable en  $\mathbb{P} \in M$ <sup>2</sup>

---

<sup>2</sup> tangentiellement à  $B(F, \mathbb{P})$ . On note  $T$  est HDT<sub>F</sub> –  $\mathbb{P}$ , si et seulement s'il existe une application linéaire  $dT_{\mathbb{P}} : M \rightarrow \mathbb{R}^q$ , telle que pour toutes suites  $\mu_n \rightarrow \mu \in B(F, \mathbb{P})$  et pour  $t_n \rightarrow 0$ , lorsque  $n \rightarrow \infty$  on a

$$\frac{T(\mathbb{P} + t_n \mu_n) - T(\mathbb{P})}{t_n} - dT_{\mathbb{P}} \cdot \mu \xrightarrow{n \rightarrow \infty} 0.$$

## Dualité : Contraintes

Pour l'exposé, contrainte linéaire  $\mathbb{E}_P(f(X_1, \theta)) = 0$ .

On se sert de la forme de la contrainte sous  $P$ , en exigeant le respect d'atteinte de cette contrainte pour une mesure signée  $Q_n$

$$\mathbb{E}_{Q_n}(f(X_1, \theta)) = 0, \quad f \in \mathcal{F}$$

On remarque qu'alors

$$\mathbb{E}_{Q_n}(f(X_1, \theta)) - \mathbb{E}_{P_n}(f(X_1, \theta)) = -\mathbb{E}_{P_n}(f(X_1, \theta)).$$

Une catégorie de travaux relaxe cette exigence d'atteinte de la contrainte, en demandant que  $Q_n$  évolue dans une boule autour de la contrainte.

## Dualité sur des mesures signées : Kuhn et Tucker

Si l'on suppose qu'il existe une mesure signée  $Q_n$ , dominée par  $P_n$ , respectant la contrainte et les conditions de qualification  $Qual(P_n)$  pour une fonction  $\varphi$  donnée (avec les hypothèses précédentes), alors le problème primal égale le problème dual.

$$Qual(P_n) \left\{ \begin{array}{l} \exists Q_n \ll P_n, Q_n \text{ mesure signée, } \mathbb{E}_{Q_n}(f(X_1, \theta)) = 0 \\ \inf d(\varphi^*) < \inf_{\Omega} \frac{dQ_n}{dP_n} \leq \sup_{\Omega} \frac{dQ_n}{dP_n} < \sup d(\varphi^*) \end{array} \right.$$

alors

$$\begin{aligned} (\beta_n(\theta) \quad =) \quad & \inf_{Q_n \ll P_n} \{ I_{\varphi^*}(Q_n/P_n) \} \\ & \mathbb{E}_{Q_n - P_n}(f(X_1, \theta)) = -\mathbb{E}_{P_n}(f(X_1, \theta)) \\ & = \sup_{\lambda \in \mathbb{R}^q} \{ -\lambda' \mathbb{E}_{P_n}(f(X_1, \theta)) - \mathbb{E}_{P_n}(\varphi(\lambda' f(\cdot, \theta))) \} \end{aligned}$$

# Dualité sur des mesures signées : Kuhn et Tucker

De plus si  $\varphi'$  a des limites finies à droite et à gauche aux frontières de  $d(\varphi)$ , alors le sup du problème dual est atteint en un  $\lambda_n^\diamond$  et l'inf du problème primal est atteint en une mesure signée  $Q_n^\diamond$  avec

$$Q_n^\diamond = (1 + \varphi'((\lambda_n^\diamond)^T f)) P_n.$$

On rappelle que

$$\begin{aligned} \mathcal{M}_n &= \{Q \text{ mesure signée avec } Q \ll \mathbb{P}_n\} \\ &= \left\{ Q = \sum_{i=1}^n q_i \delta_{X_i}, (q_i)_{1 \leq i \leq n} \in \mathbb{R}^n \right\}, \end{aligned}$$

Ainsi, il existe toujours une mesure signée  $Q_n \in \mathcal{M}_n$  respectant la première condition de qualification  $Qual(P_n)$ .

# Dualité sur proba : enveloppe convexe

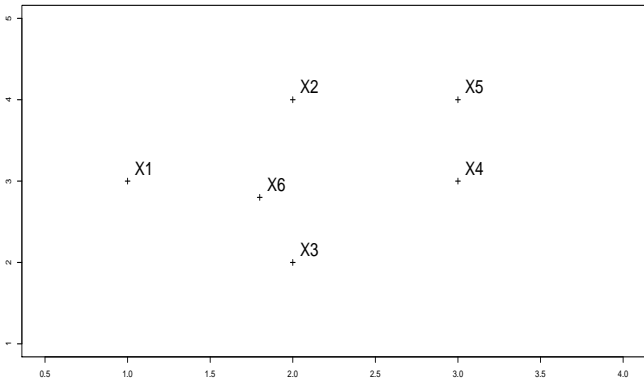
Lorsqu'on maintient la condition  $Q_n$  probabilité, on impose que le paramètre recherché soit à l'intérieur de l'enveloppe convexe des observations.

-> le problème n'a parfois pas de solution.

Pour remédier à cela, dans le cadre de la divergence de Kullback Leibler modifiée qui va chercher un optimal sur les probabilités, plusieurs études ont proposé des variantes.

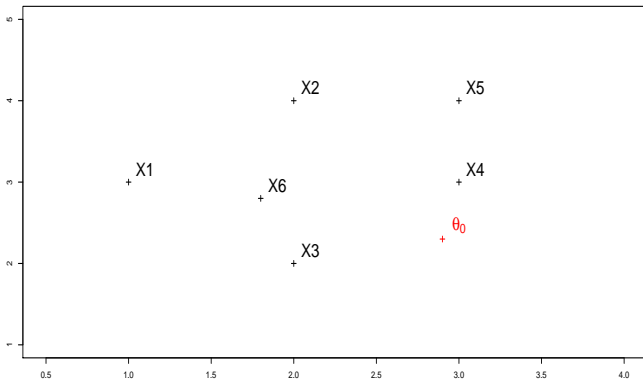
-> présentation en image pour les 1ères.

# Dualité sur proba : enveloppe convexe, Kullback



Représentation des  $X_i$ , avec  $(n, q) = (6, 2)$

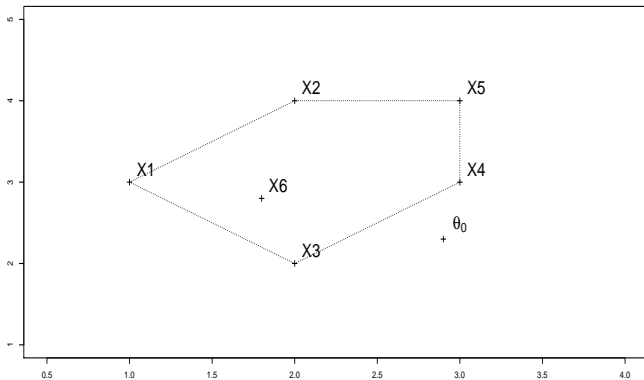
# Dualité sur proba : enveloppe convexe, Kullback



Représentation des  $X_i$  avec  $\theta_0$ ,  $(n, q) = (6, 2)$



# Dualité sur proba : enveloppe convexe, AEL



Représentation des  $X_i$  avec  $\theta_0$  et l'enveloppe convexe des  $X_i$ ,  
 $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, AEL

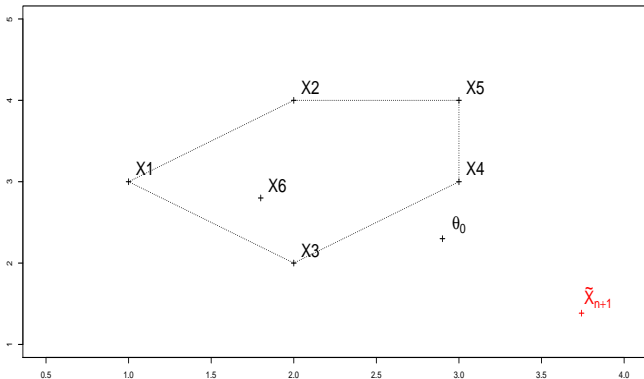


Figure: Representation of  $\tilde{X}_i$ 's, for  $i = 1, \dots, n + 1$ ,  $\theta_0$ , convex hull of the  $X_i$ 's,  $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, AEL

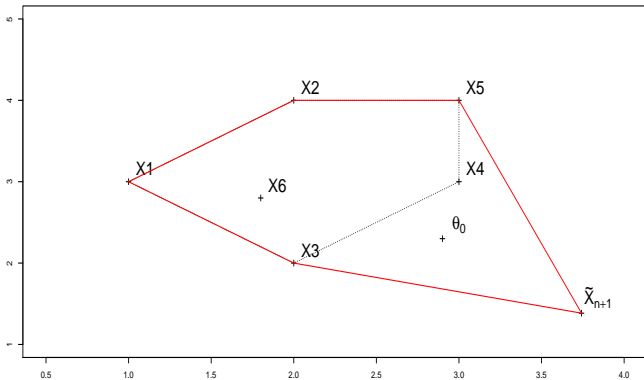
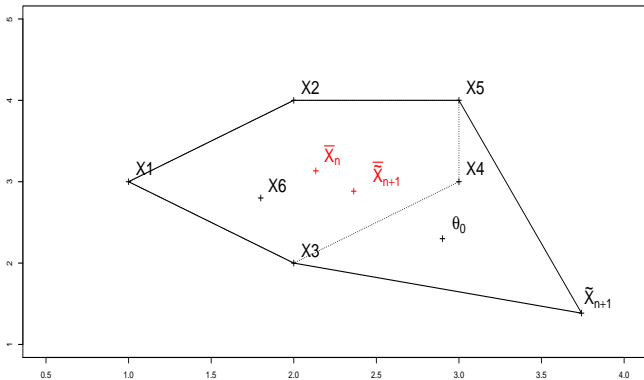


Figure: Représentation de  $\tilde{X}_i$ , pour  $i = 1, \dots, n + 1$ ,  $\theta_0$ , **enveloppe convexe des  $\tilde{X}_i$** ,  $(n, q) = (6, 2)$

## Dualité sur proba : enveloppe convexe, AEL



**Figure:** Représentation de  $\tilde{X}_i$ , pour  $i = 1, \dots, n + 1$ ,  $\theta_0$ , **enveloppe convexe** des  $\tilde{X}_i$ , et  $\tilde{X}_n$ ,  $\tilde{X}_{n+1}$ ,  $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, AEL

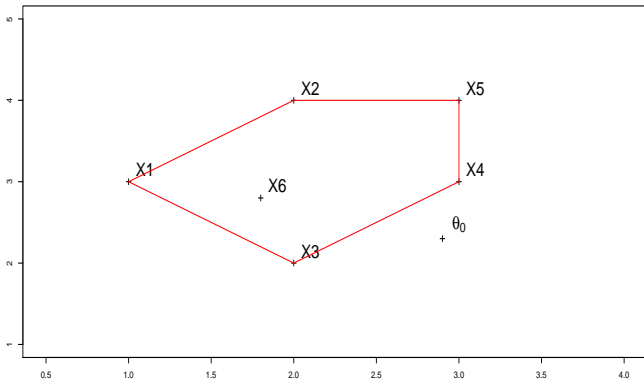
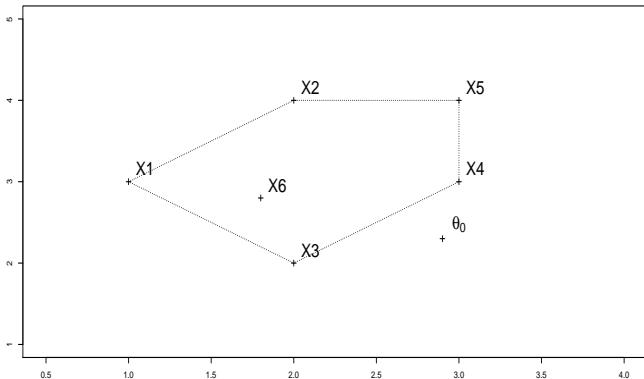


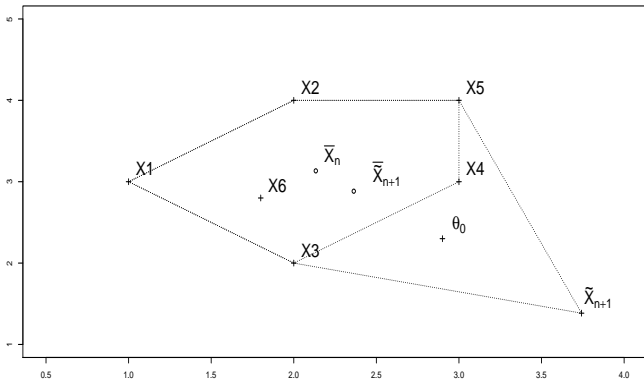
Figure: Représentation des  $X_i$ ,  $\theta_0$  et **enveloppe convexe des  $X_i$** ,  
 $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, BAEL



**Figure:** Représentation des  $X_i$ , pour  $i = 1, \dots, n$ ,  $\theta_0$  et enveloppe convexe des  $X_i$ ,  $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, BAEL



**Figure:** Représentation des  $\tilde{X}_i$ , pour  $i = 1, \dots, n + 1$ ,  $\theta_0$ ,  $\tilde{X}_{n+1}$ , convex hull of the  $X_i$ 's, et enveloppe convexe des  $\tilde{X}_i$ ,  $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, BAEL

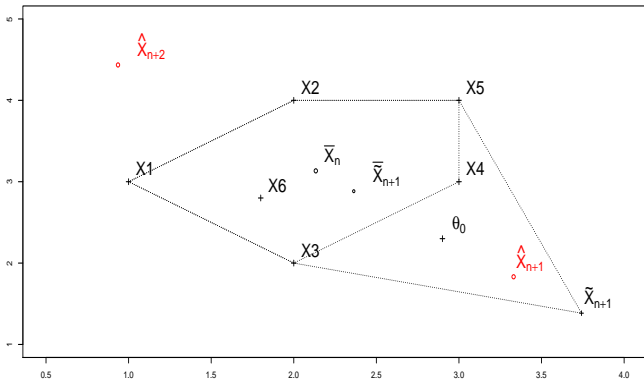


Figure: Représentation des  $\tilde{X}_i$ , pour  $i = 1, \dots, n + 2$ ,  $s = 1.8$ ,  $\theta_0$ ,  
enveloppe convexe des  $\tilde{X}_i$ , et des  $X_i$ , et de  $\bar{X}_n, \bar{X}_{n+1}$ ,  $(n, q) = (6, 2)$



# Dualité sur proba : enveloppe convexe, BAEL

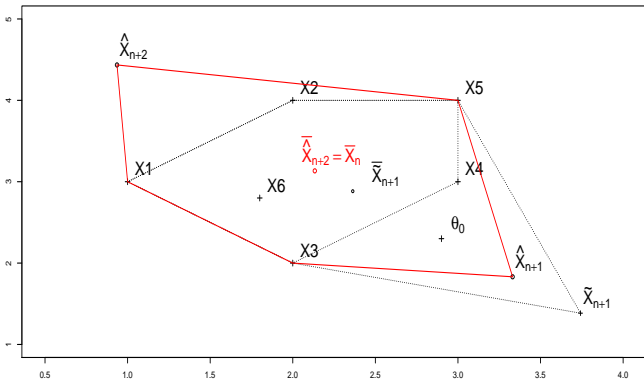


Figure: Représentation des  $\tilde{X}_i$ , pour  $i = 1, \dots, n + 2$ ,  $\theta_0$ , et enveloppe convexe des  $\tilde{X}_i$ ,  $\tilde{X}_i$ ,  $X_i$ , et  $\bar{X}_n, \tilde{X}_{n+1}, \bar{X}_{n+2}$ ,  $(n, q) = (6, 2)$

# Dualité sur proba : enveloppe convexe, pénalisation

Autres études relaxent la contrainte en modifiant la divergence de Kullback.

- (PEL) Revient à pénaliser le problème dual avec un terme correspondant d'une part à la contrainte et d'autre part à l'inverse de la matrice de covariance empirique (plus précisément d'une version modifiée de cette dernière pour existence).
- (PELAM) Cette procédure, permet d'envisager la méthode en haute dimension  $q \geq n$  à l'aide de modifications de la matrice de covariance empirique (plusieurs procédures existent par ailleurs dans le domaine de l'estimation de matrices de covariance en haute dimension).

# Dualité sur proba : enveloppe convexe, full space et combinaison

Autres voies :

- (EELFS) On applique la procédure standard mais en recherchant un  $h_n^{-1}(\theta)$ , avec  $h_n$  qui est une transformée bijective de l'espace des  $\theta$  pour lesquels il existe une probabilité empirique  $Q_n$  atteignant la contrainte, prenant ses valeurs dans  $\mathbb{R}^q$ . Cette fonction est telle que l'enveloppe convexe des points contient  $h_n^{-1}(\theta)$ .
- (Quasi-EL). Combinaison de  $\varphi^*$  : le critère à optimiser devient une combinaison convexe de deux  $I_{\varphi^*}$  divergences -  $\chi^2$  et Kullback. Revient à pénaliser sur le pb primal, contrairement aux autres méthodes de pénalisation qui pénalisent le critère dual. Allie les avantages des deux divergences.

## Dualité sur les mesures signées : poids négatifs ?

Pour tout  $0 < \alpha < 1$ , on définit une boule empirique de l'espace des mesures signées dominées par  $P_n$ , centrée en  $P_n$ , de rayon  $r = \frac{\chi_q^2(1-\alpha)}{2}$  :

$$\mathcal{M}_n(\alpha) = \left\{ Q_n \ll P_n, nI_{\varphi^*}(Q_n, P_n) \leq \frac{\chi_q^2(1-\alpha)}{2} \right\}$$

Soit  $Q_n(= (q_i)_{i=1,\dots,n})$  une probabilité de  $\mathcal{M}_n(\alpha)$ .

On définit

- $I_n = \{i, q_i > \frac{1}{n}\}$
- $I_n^+ = \{i, 0 < q_i \leq \frac{1}{n}\}$
- $I_n^- = \{i, q_i \leq 0\}$

de cardinaux respectifs  $N, N^+, N^-$ .

# Dualité sur les mesures signées : poids négatifs ?

Alors, sous les conditions sur  $\varphi$  précédentes, pour tout  $r > 0$ ,

- $\forall i \in I_n, \quad \frac{1}{n} \leq q_i \leq \frac{\varphi^{*-1}(r)+1}{n}$
- Pour  $r < \frac{1}{2}$ ,  $\forall i \in I_n^+, \quad \frac{1-\sqrt{2r}}{n} \leq q_i \leq \frac{1}{n}$
- Pour  $r \geq \frac{1}{2}$ ,

$$\max_{i \in I_n^+} q_i \geq \frac{1}{n} \max \left\{ \left(1 - \sqrt{\frac{2r}{N^+}}\right); 0 \right\}$$

$$\frac{(N^+ - 2r)^2}{4N^+n^2} \leq \sum_{i \in I_n^+} q_i^2$$

et

$$\forall i \in I_n^- \quad \frac{1 - \sqrt{2r}}{n} \leq q_i \leq 0$$

$$N^- \leq 2r.$$

# Dualité sur les mesures signées de masse unitaire

Travailler sur les mesures signées non qualibrées à 1 permet de s'extraire de la sous-couverture dûes à l'enveloppe convexe des points.

Cette approche permet de conserver la dualité, et d'établir la convergence du critère vers l'analogue du problème dual sur le critère non empirique.

# Région de confiance

Si l'on avait  $\mathbb{E}_{P_n}(f(X_1, \theta)) = 0$ , on aurait alors  $Q_n^\diamond = P_n$  et la quantité  $\beta_n(\theta)$  serait nulle.

On recherche  $Q_n$  telle que la  $\beta_n(\theta)$  soit petite. Ainsi, on considérant que tous les  $\theta$  satisfaisant cette demande nous intéressent pour former la région de confiance, on en tire la forme de cette région.

$$C_n(\eta) = \{ \theta \in \mathbb{R}^q \mid \exists Q_n \ll P_n, \mathbb{E}_{Q_n}(f(X_1, \theta)) = 0 \text{ et } nI_{\varphi^*}(Q_n/P_n) \leq \eta \},$$

avec  $\mathbb{P}(\theta \in C_n(\eta)) = 1 - \alpha + o(1)$ .

Ainsi

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\theta \notin C_n(\eta)) &= \lim_{n \rightarrow \infty} \mathbb{P}(n\beta_n(\theta) \geq \eta) \\ &= \alpha. \end{aligned}$$

# Bibliographie I

Michel Broniatowski and Amor Keziou.  
Divergences and duality for estimation and test under moment condition model.  
*Journal of Statistical Planning and Inference*, 142(9):2554–2573, 2012.

Patrice Bertail, Emmanuelle Gautherat, and Hugo Harari-Kermadec.  
Empirical  $\phi^*$ -Divergence Minimizers for Hadamard Differentiable Functionals.  
*ESAIM: Proceedings and Surveys*, 51:212–231, 2014.  
10.1051/proc/201551012.

D. Faddeev.  
On the concept of entropy of a finite probabilistic scheme.  
*Uspekhi Mat. Nauk*, 11(1):227–231, 1956.  
(en russe).

J. Havrda and F. S. Charvát.  
Quantification method of classification processes. Concept of structural  $\alpha$ -entropy.  
*Kybernetika*, 3:30–35, 1967.

Z. Daróczy.  
Generalized information functions.  
*Information and Control*, 16(1):36–51, March 1970.

C. Tsallis.  
Possible generalization of Boltzmann-Gibbs statistics.  
*Journal of Statistical Physics*, 52(1):479–487, July 1988.



# Bibliographie II

A. Papoulis.

Maximum entropy and spectral estimation: A review.

*IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1176–1186, December 1981.

E. T. Jaynes.

Information theory and statistical mechanics.

In K. W. Ford, editor, *1962 Brandeis Summer Institute in Theoretical Physics*, volume 3, pages 182–218. W. A. Benjamin Inc., New York, 1963.

Reprinted in: E. T. Jaynes: *Papers on Probability, Statistics and Statistical Physics*, ed. R. D. Rosencrantz, Synthèse Library, Vol. 138, Reidel, 1983.

S. Kullback.

*Information Theory and Statistics*.

John Wiley & Sons, New York, 1959.

Republished by Dover Publications, 1997.

M. M. Rao and Z. D. Ren.

*Theory of Orlicz Spaces*.

Marcel Dekker, New York, 1991.

I. Csiszár.

Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markowschen Ketten.

*Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:85–108, 1963.

# Bibliographie III

I. Csiszár.

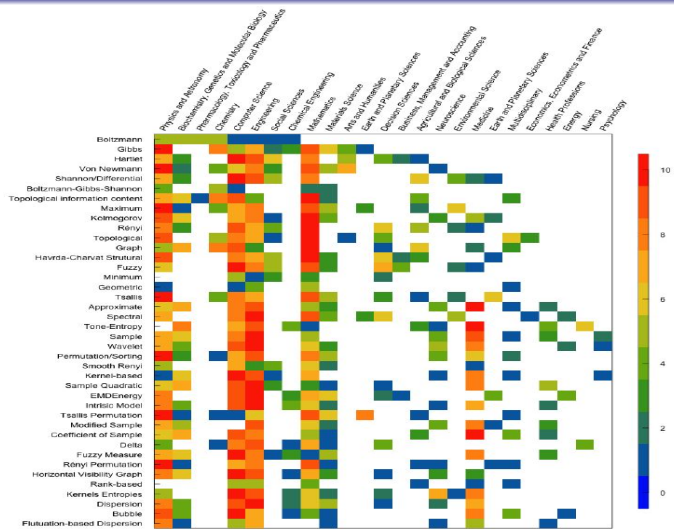
On topology properties of f-divergences.

*Studia Sci. Math. Hungar.*, 2:329–339, 1967.

F. Liese and I. Vajda.

*Convex Statistical Distances*, volume 95.

BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987.



Les dix

domaines selon "Documents by subject area" de Scopus. Légende : 0 domaine , à 10 domaines

Source : M. Ribeiro et al, *The Entropy Universe*, Entropy 2021, 23, 222