Estimation d'une densité discrète sous contrainte de k-monotonie.

Jade Giguelay sous la direction de Christophe Giraud et Sylvie Huet.

Orsay - 23 juin 2016

Sommaire.

- Densités $k-$monotones discrètes.
- Estimation sous-contrainte de $k-$monotonie :
  - Propriétés.
  - Algorithme.
- Le choix de k.

Notations.

- $p^* = (p_0, p_1, \dots)$ est une densité dans $\mathbb{N}$.

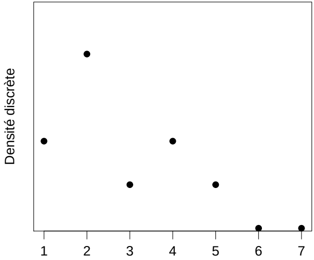
- Le maximum du support τ de p est défini par :

  $$\tau = \min\{j \in \bar{\mathbb{N}}, \forall k > j, p_k = 0\}.$$

- $X_1, .., X_n$ i.i.d de loi p de maximum de support τ (inconnu)

- $\tilde{p}_n : \tilde{p}_n(j) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i = j\}}$ est l'estimateur empirique.

- $k \geqslant 2$ est un entier fixé.

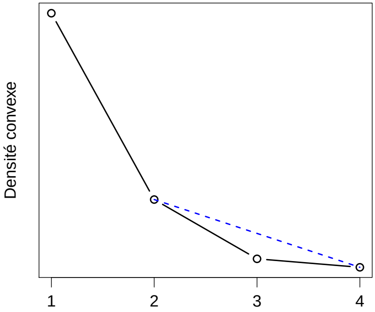


La convexité.

Définition

(Le Laplacien) $\Delta^2 p(i) = p_i - 2p_{i+1} + p_{i+2} = (p_{i+2} - p_{i+1}) - (p_{i+1} - p_i)$.

- La loi p est dite **convexe** si pour tout $i \in \mathbb{N}$, $\Delta^2 p(i) \geqslant 0$.
- Si $\Delta^2 p(i) > 0$ on dit que i est un **noeud** de p.



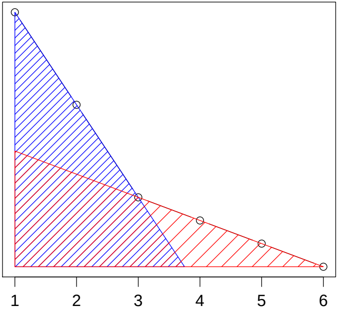
La convexité 2.

Une fonction convexe est mélange de fonctions triangulaires : si p est une fonction convexe discrète alors :

$$p(i) = \sum_{j=0}^{\infty} \frac{\Delta p(j+1)}{(j+1)(j+2)} T_j(i)$$

. Où T_τ est la distribution triangulaire de support $\{0, .., \tau\}$:

$$T_\tau(i) = \begin{cases} \frac{2(\tau+1-i)}{(\tau+1)(\tau+2)} & \text{si j} \in \{0,..,\tau\} \\ 0 & \text{sinon.} \end{cases}$$



Fonction $k-$monotone.

Une fonction p de $L^1(\mathbb{N})$ est :

- **monotone** ($1-$**monotone**) si pour tout i, $\Delta^1 p(i) := p_{i+1} - p_i \leqslant 0$
- **convexe** ($2-$**monotone**) si pour tout i,

$$\Delta^2 p(i) := p_i - 2p_{i+1} + p_{i+2} = \Delta^1 p_{i+1} - \Delta^1 p_i \geqslant 0.$$

$\rightarrow$ Pour généraliser on dira que p est :

- $3-$**monotone** si pour tout i, $\Delta^3 p(i) := \Delta^2 p_{i+1} - \Delta^2 p_i \leqslant 0$,

$\vdots$

- $k-$**monotone** si pour tout i, $\Delta^k p(i) := \Delta^{k-1} p_{i+1} - \Delta^{k-1} p_i$ est du signe de $(-1)^k$.

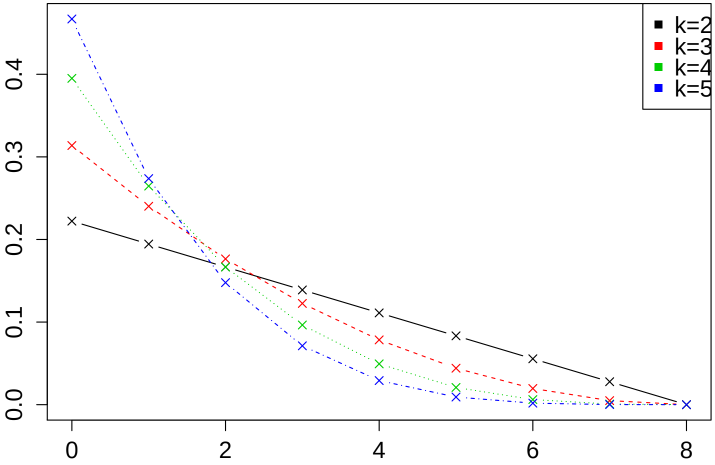
Fonction $k-$monotone.

Définition

(Le j-ième Laplacien) $\Delta^j p(i) = \sum_{h=0}^{j} \binom{j}{h} (-1)^{j-h} p(h+i)$.

- La distribution p est $k-$monotone sur $\mathbb{N}$ si $(-1)^k \Delta^k p(i)$ est positive pour tout i.
- Si $(-1)^k \Delta^k p(i) > 0$ on dit que i est un $k-$**noeud** de p.
- $k-$monotone $\Rightarrow$ $j-$monotone strictement sur son support, $j = 1, .., k-1$.

Exemple de fonctions $k-$monotones.



Ecriture en base de spline.

On définit une famille de splines $(Q_j^k)_{j\in\mathbb{N}^*}$:

$$Q_j^k(i) = C_{j-i+k-1}^{k-1} = \frac{(j-i+k-1)\dots(j-i+1)}{(k-1)!}\mathbb{I}_{j\geqslant i} \qquad (1)$$

(Lefèvre et Loisel (2012)).

Exemple :

- $Q_j^2(i) = (j+1-i)_+$ est la triangulaire T_j.
- $Q_j^3(i) = \frac{1}{2}\left((j+1-i)_+^2 + (j+1-i)_+\right)$.
- $Q_j^4(i) = \frac{1}{6}\left((j+1-i)_+^3 - (j+1-i)_+\right)$.

Ecriture en base de splines.

Une fonction $k-$monotone s'écrit comme un mélange de Q_j^k :

Theorem

Soit p une densité de masse finie sur $\mathbb{N}$. p est $k-$monotone si et seulement si p s'écrit :

$$p(i) = \sum_{j \geqslant i} (-1)^k \Delta^k p_j Q_j^k(i) = \sum_{j \geqslant i} \tilde{\pi}_j \frac{Q_j^k(i)}{m_j}, \tag{2}$$

où $m_j = \sum_{i=0}^{j} Q_j^k(i)$.
De plus, p est une probabilité $\Leftrightarrow$ $\tilde{\pi}$ est une probabilité.

La convexité 3.

Durot et al. (2013) : Estimation de p^* sous-contrainte de convexité par méthode des moindres carrés :

$$\hat{p}_n = argmin\{||f - \tilde{p}_n||_2, f \text{ fonction convexe}\}. \tag{3}$$

- $\hat{p}_n$ est de support fini.
- $\hat{p}_n$ est de masse 1.
- Algorithme exact via Support Reduction Algorithm (Groeneboom et al., (2008))

Estimateur sous contrainte de $k-$monotonie.

L'estimateur $\hat{p}_{n,k}$ sous contrainte de $k-$monotonie est défini ainsi :

$$\hat{p}_{n,k} = argmin\{||f - \tilde{p}_n||_2, f \text{ **probabilité** } k-\text{monotone}\}. \tag{4}$$

- Il existe et il est unique.
- Son support est fini.

Propriétés de $\hat{p}_n$.

Propriété

Pour toute probabilité f $k-monotone$ $||f - \hat{p}_{n,k}||_2 \leqslant ||f - \tilde{p}_n||_2$ avec une inégalité stricte si $\tilde{p}_n$ n'est pas $k-$monotone.

Propriété

Soit p une probabilité discrète quelconque.
Pour tout $r \in [2, +\infty]$ on a $\sqrt{n}||p_{S_k} - \hat{p}_{n,k}||_r = O_P(1)$
où $p_{S_k} = argmin\{||f - p||_2, f \in S_k\}$ est le projeté orthogonal de p sur l'ensemble des probabilités $k-$monotones.

Propriété

Soit p une densité $k-$monotone de support fini. Soit $r \in \mathbb{N}$ un $k-$noeud de p. Alors avec probabilité 1 il existe un rang n_0 tel que pour tout $n \geqslant n_0$, r est un $k-$noeud de $\hat{p}_n$.

Remarque sur la masse.

On peut également définir :

$$\hat{p}^*_{n,k} = argmin\{||f - \tilde{p}_n||_2, f \textbf{ fonction } k-\text{monotone}\}. \tag{5}$$

Lemme

Soit δ_1 la masse de Dirac en 1. Le projeté de δ_1 sur les fonctions 3−monotones est :

$$p = \frac{3}{238} Q_5^3 + \frac{1}{238} Q_6^3,$$

La masse de p est d'environ 1.14.

Un algorithme en deux temps.

Step 1) Adaptation de l'**Algorithme de Réduction de Support**, Groeneboom et al. (2008).

$L \in \mathbb{N}^*$ est fixé, on calcule l'estimateur $\hat{p}_n^L$:

$$\hat{p}_n^L = \text{argmin}\{||\tilde{p}_n - q||_2, q \in S_{k,L}\},$$

où $S_{k,L} = \{\text{probabilités } f \ k-\text{monotones, supp}(f) \subset \{0, .., L\}\}$.

Step 2) On trouve un **critère d'arrêt** : à partir de quel $L \in \mathbb{N}^*$, $\hat{p}_n = \hat{p}_n^L$?

Caractérisation de $\hat{p}_n$.

- $\forall l \in \mathbb{N}, \begin{cases} F_f^1(l) = F_f(l) = \sum_{i=0}^{l} f(i), \\ \forall j \geqslant 2, F_f^j(l) = \sum_{i=0}^{l} F_f^{j-1}(l). \end{cases}$
- $\beta(p) = \sum_{i=0}^{\infty} p(i)\big(p(i) - \tilde{p}_n(i)\big)$.

Theorem

Soit p une probabilité. Soit m_l^k la masse de Q_l^k. Il y a équivalence entre :

1. $p = \hat{p}_n$.
2. - *Pour tout $l \in \mathbb{N}$ on a l'égalité :*

$$F_{\hat{p}_n}^k(l) - F_{\tilde{p}_n}^k(l) \geqslant m_l^k \beta(\hat{p}_n) \tag{6}$$

   - *il y a égalité si l est un noeud de $\hat{p}_n$.*

Critère d'arrêt.

Résultats :

- On sait calculer un tel L en pratique pour tout $k \geqslant 3$.
- Pour les cas $k = 3$ et $k = 4$ on a un résultat plus précis :

Propriété

Soit s le maximum entre le support de p et le support de $\tilde{p}_n$. On a équivalence entre :

1. $p = \hat{p}_n$.
2. 1. $\forall l \leqslant s+1, F^k_{\hat{p}_n}(l) - F^k_{\tilde{p}_n}(l) \geqslant m^k_l \beta(\hat{p}_n)$.
   2. *Si l est un $k-$noeud de p il y a égalité dans l'inégalité précédente.*
   3. $\forall j \in \{1,..,k-1\}, F^j_p(s+1) \geqslant F^j_{\tilde{p}_n}(s+1)$

Bilan et nouveaux objectifs.

A k fixé :

- On a proposé un estimateur de p sous contrainte de $k-$monotonie.
- Il est de support fini, de masse 1.
- On sait le caractériser et le calculer en pratique.

Objectifs :

- Quantifier la concentration de $\hat{p}_k$ autour de p en fonction de k :

$$\mathbb{P}\left(||\hat{p}_k - p^*||_2^2 \leq \mathrm{BIAIS} + V(k) + f(x)\right) \geq 1 - e^{-x}.$$

- Choisir k.

Adaptation d'un argument de Chatterjee (2014).

Objectif : $\mathbb{P}\left(||\hat{p}_k - p^*||_2^2 \leq \mathrm{BIAIS} + V(k) + f(x)\right) \geq 1 - e^{-x}$.

Rappel : $||\hat{p}_k - p^*||_2^2 = O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$.

On suppose $p^* \in S_k :=$ {probabilités k-monotones}.

$$||\hat{p}_k - p||_2^2 = \underset{t}{\mathrm{argmax}} \left[\underbrace{\min_{p \in S_k, ||p-p^*||_2 \leq 2} < \tilde{p} - p^*, p - p^* >}_{M(t)} - \frac{t^2}{2} \right] := \hat{t}$$

$$M(t) = \frac{c_k t}{\sqrt{n}} \Rightarrow \hat{t} = \frac{d_k}{\sqrt{n}}$$

Nouvel Objectif : Trouver un majorant $\bar{M}$ de M de la forme $\bar{M}(t) = \frac{c_k t}{\sqrt{n}}$.

Entropie Métrique.

- (E, d) un espace métrique de fonctions à valeur réelle.
- T un sous-espace de E.
- $\epsilon > 0$.

Un ϵ-**crochet** de borne $(t_1,\ t_2) \in E^2$ est défini par :

$$[t_1, t_2] = \{t \in T, s_1 \leqslant t \leqslant s_2\}$$

Le nombre minimal de ϵ-crochets nécessaires pour recouvrir T est noté $N_{[.]}(\epsilon, T, d)$.

$\log(N_{[.]}(\epsilon, T, d))$ est l'**entropie à crochet** de T.

Lien entre Entropie Métrique et Vitesse de Convergence.

Objectif : Trouver un majorant $\bar{M}$ de M de la forme $\bar{M}(t) = \frac{c_k t}{\sqrt{n}}$.

Theorem

(Massart, p193). Soit $t \in [0,1]$. On suppose que :

- $\forall h \geq 2$, $\forall f \in \mathcal{F}_t$, $\mathbb{E}[|f(X_1)|^h] \leq \frac{h!t^h}{2}$.
- *Pour tout $\delta > 0$ on peut construire un crochet de $\mathcal{F}_t$, $B_\delta = \{(\underline{f}_1 - \bar{f}_1), ..., (\underline{f}_{N_\delta} - \bar{f}_{N_\delta})\}$ de longueur N_δ tel que pour tout $i \in [\![1, N_\delta]\!]$:*

$$\forall h \geq 2, \mathbb{E}[|\underline{f}_i(X_1) - \bar{f}_i(X_1)|^h] \leq \frac{h!t^h}{2}.$$

Alors :

$$\mathbb{P}\left(M(t) \leqslant \frac{27}{\sqrt{n}} \int_0^t \sqrt{H(u) \wedge n} du + \frac{4tH(t)}{n} + \frac{7t\sqrt{2x}}{\sqrt{n}} + \frac{2tx}{n} \right) \geqslant 1 - e^{-x} \quad (7)$$

Le calcul de l'entropie sur l'espace des fonctions $k-$monotones et le résultat de concentration associé.

Propriété

Soit $\delta > 0$. On a :

$$H(\delta) := N_{[.]}\left(\delta, l_2(p^*), \mathcal{F}_t\right) = \left(\frac{1}{\delta}\log(\frac{t^2}{\delta}) + o(\frac{1}{\delta})\right)\mathbb{I}_{t^2 \geq \delta}.$$

Propriété

Il existe des constantes α_k, β_k telles que pour tout $t \in]0,1[$, pour tout $x > 0$, pour tout $k \geq 2$:

$$\mathbb{P}\left(M(t) \leqslant \alpha_k t^2 + \frac{\beta_k t}{\sqrt{n}} + 2t\frac{x+2}{n}\right) \geqslant 1 - e^{-x} \tag{8}$$

Une différence avec le cas continu.

$\rightarrow$ Vitesse paramétrique.

$$\begin{aligned}\mathbb{E}\left[||\tilde{p}-p^*||_2^2\right] &= \mathbb{E}\left[\sum_{l=0}^{\infty}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{X_i=l}-p^*(l))^2\right] \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{l=0}^{\infty}(\mathbb{I}_{X_i=l}-p^*(l))^2\right] \\ &= \frac{1}{n}p^*(l)(1-p^*(l)).\end{aligned}$$

Theorem

(Gao et Wellner (2009)). Soit $\epsilon > 0$. Soit $\mathcal{F}_k([0,1])$ l'ensemble des densités (continues) k-monotones sur $[0,1]$. Il existe des constantes C, D telle que pour toute densité p^* :

$$\log N_{[.]}(\epsilon, \mathcal{F}_k([0,1])), ||.||_{2,p^*}) \leq C\frac{D^{1/k}}{\epsilon^{1/k}}.$$

Perspectives.

- Améliorer la borne supérieure $\bar{M}$ pour une meilleure dépendance en k.
- Trouver une borne inférieure.
- Choisir k via sélection de modèle.
- Application à l'estimation du nombre d'espèces dans une population.

Bibliographie.

Merci !

- Durot, Huet, Koladjo, and Robin, Computational Statistics and Data Analysis, **67**, (2013)
- Groeneboom et al., Scandinavian Journal of Statistics, **35**, (2008)
- Balabdaoui and Wellner, The Annals of Statistics, (2007)
- Lefevre and Loisel, Journal of Applied Probability, **50**, (2013)
- Chatterjee, The Annals of Statistics, **42**, (2014)
- Massart, Concentration Inequalities and Model Selection, Springer, Ecole d'Eté de Probabilités de Saint-Flour (2003).
- Gao and Wellner, On the rate of convergence of the maximum likelihood estimator of a k-monotone density, Springer, (2009).