

# Expériences empiriques

Thierry Bazier-Matte

6 avril 2017

## Table des matières

<b>1</b>	<b>Expériences empiriques</b>	<b>2</b>
1.1	Méthodologie	2
	Noyau	2
	Fonctions d'utilité	3
	Régularisation	3
	Loi de marché	3
	Précision de la borne et quantiles d'erreur	4
	Échantillonnage	4
	Environnement de calcul	4
1.2	$n$ variable, $p$ constant	6
	Loi de marché	6
1.2.1	Erreur de généralisation	6
	Quantiles d'erreur – Figure 3	6
	Erreur de généralisation et aversion au risque – Figure 4	6
	Borne sur l'erreur – Figures 5 et 6	7
	Erreur en util et en équivalent certain	7
1.2.2	Erreur de sous optimalité	12
	$\lambda$ constant – Figure 7	12
	$\lambda$ décroissant – Figure 8	12
1.3	$n$ constant, $p$ variable	15
	Protocole d'expérience	15
	Erreur de généralisation – Figure 9	15
1.3.1	Sous optimalité	18
	Utilité espérée optimale	18
	Erreurs de sous optimalité	18
	Borne sur l'erreur de sous optimalité – Figures 12, 13 et 14	18
1.4	Ajout d'information et d'échantillons	24
	Méthodologie	24
1.4.1	Erreur de généralisation	24
	Borne de généralisation – Figure 16	24
	Régime d'ordre plus élevé – Figure 17	25
1.4.2	Erreur de sous optimalité	25

## Table des figures

1	Utilités LEU . . . . .	5
2	Utilité et utilité inverse . . . . .	5
3	Quartiles et valeurs maximales de l'erreur de généralisation . . . . .	8
4	Aversion au risque et erreur de généralisation . . . . .	9
5	Borne théorique sur l'erreur de généralisation . . . . .	10
6	Composantes de l'erreur maximale . . . . .	11
7	Borne de sous optimalité, $\lambda$ constant . . . . .	13
8	Borne de sous optimalité, $\lambda$ décroissant . . . . .	14
9	Erreur de généralisation avec ajout d'information . . . . .	17
10	Progression de $EU^*$ relatif . . . . .	19
11	Progression de la sous optimalité relative . . . . .	20
12	Borne sur l'erreur de S.O. I . . . . .	21
13	Borne sur l'erreur de S.O. II . . . . .	22
14	Borne sur l'erreur de S.O. III . . . . .	23
15	Progression de $p$ par rapport à $n$ . . . . .	24
16	à faire . . . . .	26
17	Erreur de généralisation à très haut régime . . . . .	27
18	à faire . . . . .	28
19	Progression de $\ q^*\ ^2$ . . . . .	29

## 1 Expériences empiriques

Cette section sera l'occasion de valider numériquement les garanties présentées à la [Citation needed] quant aux erreurs de généralisation et de sous optimalité inhérentes à l'algorithme d'investissement présenté dans ce mémoire.

Il va sans dire que le cadre théorique général qui a été développé jusqu'à maintenant présente plusieurs paramètres (dimensionnalité du problème, loi de marché, fonction d'utilité, noyau employé, etc.); tous les décrire représenterait une tâche titanesque, aussi certains choix devront être faits pour restreindre la quantité de paramètres étudiés; la Section 1.1 énumérera le choix fait pour chacun de ces paramètres.

Par la suite, les Sections 1.2, 1.3 et 1.4 étudieront la qualité des garanties de généralisation et de sous optimalité dans un contexte où, respectivement, la taille de l'échantillonnage augmente, la taille de l'échantillonnage est fixe mais la dimensionnalité du problème augmente et enfin, la taille de l'échantillonnage et de la dimensionnalité augmentent toutes les deux, mais à des rythmes différents.

### 1.1 Méthodologie

**Noyau** Le noyau employé dans nos expériences sera linéaire. En particulier, c'est avec un tel noyau que la dépendance entre la dimensionnalité du problème est les erreurs

de sous optimalité et de généralisation est la plus facilement caractérisable.

**Fonctions d'utilité** Chaque expérience sera conditionnée par une fonction d'utilité exponentielle Lipschitz  $LEU_\mu$  (voir Fig. 1 (p. 5) pour une description de cette famille).

Ces utilités sont idéales pour deux raisons : d'abord elles ont toutes un coefficient Lipschitz  $\gamma = 1$  ; ensuite, leur paramètre  $\mu \geq 0$  permet de quantifier facilement l'aversion au risque qu'elles convoient,  $\mu \rightarrow \infty$  correspondant à une attitude neutre au risque et  $\mu = 0$  correspondant à l'attitude extrêmement aversive où aucune utilité n'est accordée aux rendements supérieurs à zéro. Mathématiquement, les fonctions exponentielles Lipschitz sont définies par

$$LEU_\mu(r) = \begin{cases} r & r < 0 \\ \mu(1 - e^{-r/\mu}) & r \geq 0 \end{cases} . \quad (1)$$

La fonction d'utilité inverse  $LEU_\mu^{-1} : \mathcal{U} \rightarrow \mathcal{R}$ , nécessaire pour exprimer en terme de rendement équivalent les erreurs exprimées en util, est illustrée à la Fig. 2 (p. 5). On peut vérifier algébriquement que

$$LEU_\mu^{-1}(r) = \begin{cases} r & r < 0 \\ -\mu \log(1 - r/\mu) & r \geq 0 \end{cases} . \quad (2)$$

Finalement, les bornes d'erreur de généralisation et de sous optimalité, lorsqu'elles sont exprimées en équivalent certain, font intervenir l'inverse du sous-gradient de  $u^{-1}$ . Dans le cas de l'utilité LEU, celui-ci correspond tout simplement à l'inverse de la dérivée de  $LEU_\mu$  et est donc donné par

$$\left( \frac{d}{dr} LEU_\mu^{-1}(r) \right)^{-1} = \begin{cases} 1 & r < 0 \\ e^{r/\mu} & r \geq 0 \end{cases} . \quad (3)$$

**Régularisation** Sauf exception, le facteur de régularisation  $\lambda = 1/2$  sera employé au cours de toutes les expériences.

**Loi de marché** La loi de marché  $M$  sera construite en deux temps. D'abord, une loi de marché théorique  $\tilde{M} \in \mathcal{R}^{p+1 \times p+1}$  sera définie. Toutes ses marges seront décrites par des variables aléatoires Rademacher (retournant  $\pm 1$  avec probabilité  $1/2$ ). La dépendance entre les marges sera modélisée à l'aide d'une copule gaussienne dont la matrice de corrélation sera définie en début de chaque section.

Puis, à partir de cette loi de marché théorique  $\tilde{M}$ , un échantillon fini  $M \sim \tilde{M}^{5000}$  de 5000 points en sera tiré afin de former une loi de marché discrète  $M$  à partir de laquelle toutes les expériences seront réalisées. En quelque sorte,  $M$  fournira alors une approximation à  $\tilde{M}$ , mais permettra de déterminer exactement des variables comme

l'utilité hors échantillon  $EU(q)$  d'une politique  $q$  ou l'utilité espérée optimale  $EU^*$ , qu'il serait autrement impossible à déterminer théoriquement (sauf dans le cas de l'utilité neutre au risque).

**Précision de la borne et quantiles d'erreur** Les bornes sur les erreurs présentées aux Théorèmes <sup>[Citation needed]</sup> s'appliquent à tout échantillon  $\mathcal{S}_n$  avec une probabilité  $1 - \delta$ . Elles s'appliquent donc, de façon équivalente, au  $1 - \delta$ -ième quantile avec probabilité 1.

Ainsi, pour confirmer ces bornes, celles-ci seront évaluées à  $\delta = 5\%$  et le 95<sup>e</sup> percentile d'erreur sera mesuré.

**Échantillonnage** Les échantillons d'entraînement  $\mathcal{S}_n$  seront traités de façon équivalente pour toutes les sections.

Par exemple, à la Section 1.2, où c'est la taille de l'échantillon qui augmente linéairement, on tirera d'abord  $m \times \bar{n}$  réalisations de  $M$  afin d'obtenir  $m$  échantillons d'entraînement  $\mathcal{S}_{\bar{n}}$ . Puis, on exposera progressivement à l'algorithme  $n$  des  $\bar{n}$  points afin d'obtenir peu à peu une meilleure représentation de  $M$ . Les  $m$  points serviront à déterminer le 95<sup>e</sup> percentile d'erreur.

À la Section 1.3, où c'est la dimensionalité du problème qui varie, l'idée demeure la même, cette fois avec  $n$  fixe et  $p$  variable. On tire donc tout d'abord  $m \times \bar{n}$  réalisations de  $M$ . Comme chacune de ces réalisations est constituée de  $\bar{p}$  variables d'information, où  $\bar{p}$  note le nombre de marges d'information de  $M$ , on a alors qu'à présenter à l'algorithme des réalisations "incomplètes", dont seules les  $p$  premières dimensions sont connues.

Enfin, à la section 1.4, la situation est un mélange des deux précédentes, où de plus en plus de points provenant d'un même échantillon sont présentés à l'algorithme, leur dimension dévoilée progressant en fonction de  $n$ .

Toutes nos expériences disposeront de  $m = 100$  échantillons d'entraînement.

**Environnement de calcul** L'identification numériques des politiques optimales  $\hat{q}$  se fera à partir de l'implémentation CVXPY[DB16] et du solveur ECOS[DCB13]. Les calculs numériques se feront à partir de la librairie BLAS et de l'interface NUMPY.

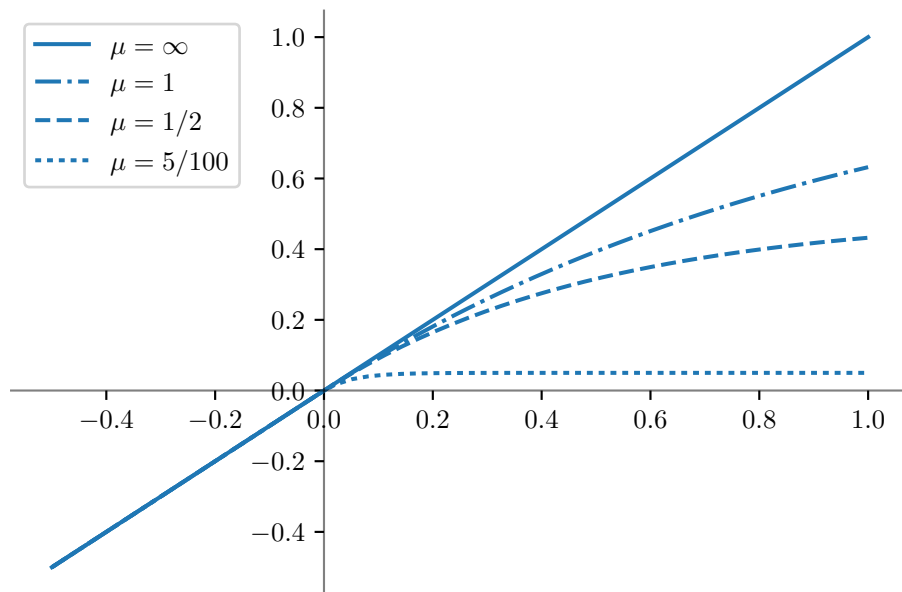


FIGURE 1 – Fonctions d'utilité exponentielles Lipschitz (LEU)

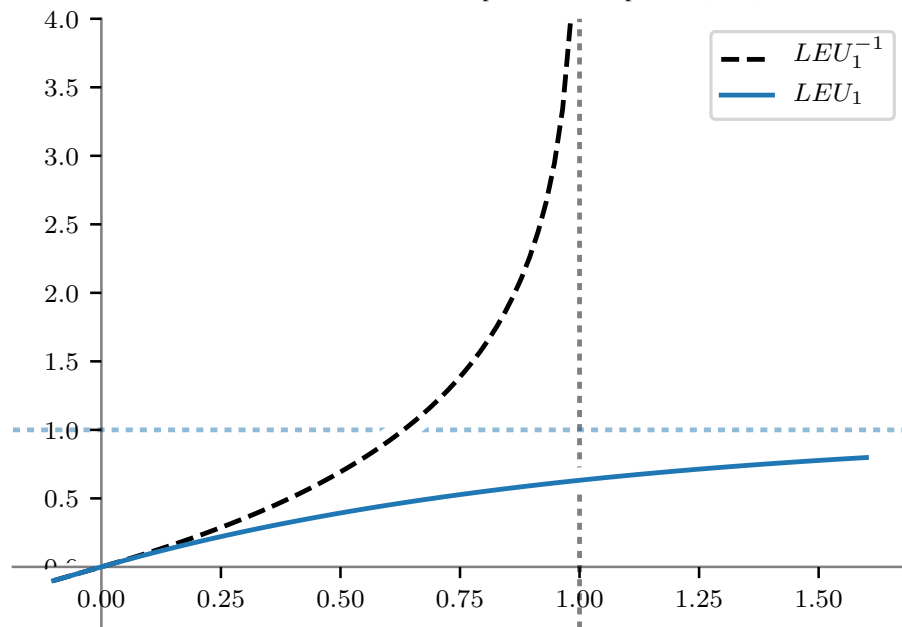


FIGURE 2 – Utilité et utilité inverse

## 1.2 $n$ variable, $p$ constant

L'objet de cette section est l'étude du cas canonique où la taille  $n$  de l'échantillon  $\mathcal{S}_n$  augmente linéairement.

**Loi de marché** Tel qu'expliqué à la Section 1.1, une loi de marché discrète  $M$  sera dérivée d'une loi théorique  $M$ . Cette loi théorique disposera ici de trois marges (deux variables d'information et une variable de rendement, toutes trois Rademacher). La loi théorique de marché  $M$  sera modélisé à partir de la matrice corrélation  $\Sigma$  donnée par

$$\Sigma = \begin{matrix} & \begin{matrix} X_1 & X_2 & R \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ R \end{matrix} & \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \end{matrix}, \quad (4)$$

où  $\rho = 1/\sqrt{2}$ , ce qui correspond à la plus grande valeur de corrélation permettant à  $\Sigma$  d'être semi-définie positive. Ainsi,  $X_1$  et  $X_2$  seront mutuellement indépendants, mais auront toutes deux une influence égale sur la réalisation de  $R$  (la corrélation entre  $X_j$  et  $R$  correspond au tau de Kendall :  $\text{Corr}(X_j, R) = \frac{2}{\pi} \arcsin(\rho) = 1/2$ . Voir [Ré13] pour des précisions).

On en déduit évidemment que  $\xi = \sqrt{2}$  et  $\bar{r} = 1$ .

### 1.2.1 Erreur de généralisation

On rappelle que l'erreur de généralisation d'une politique d'investissement  $q$  consiste à mesurer la différence entre l'utilité (resp. équivalent certain) espérée observée en échantillon avec l'utilité (resp. équivalent certain) espérée hors échantillon, ou, mathématiquement, de déterminer  $\widehat{EU}(q) - EU(q)$  (resp.  $\widehat{CE}(q) - CE(q)$ ).

**Quantiles d'erreur – Figure 3** Tout d'abord, la Fig. 3 (p. 8) indique les quantiles d'erreur de généralisation des  $m$  échantillons d'entraînement, incluant la valeur maximale et minimale pour chaque  $n$ . Nos bornes ne donnent que des garanties partielles sur les maximums ni sur les minimums, cependant il est intéressant d'observer le comportement convergent vers zéro de chacun des quantiles d'erreur.

**Erreur de généralisation et aversion au risque – Figure 4** Par ailleurs, si intuitivement on peut s'attendre à observer une relation entre l'erreur de généralisation et l'aversion au risque, la Fig. 4 (p. 9) montre qu'effectivement, une plus forte aversion au risque (caractérisée par  $\mu$ ) entraîne une erreur de généralisation plus faible, alors qu'au contraire, une faible aversion au risque entraîne une erreur de généralisation plus importante. Cette relation est importante puisqu'elle permet de généraliser les observations empiriques faites à partir d'une seule utilité à d'autres utilités. Dans les expériences suivantes, l'utilité étalon sera celle caractérisée par un coefficient  $\mu = 1$ .

**Borne sur l'erreur – Figures 5 et 6** La Fig. 5 (p. 10) permet de constater la validité des garanties théoriques offertes par l'algorithme d'investissement. On constate ici que la borne n'est pas exactement serrée, les courbes théoriques et empiriques différant d'un ordre de grandeur. Néanmoins, il faut conserver à l'idée que ces bornes sont valides pour toute distribution de marché  $M$  de dimension  $\xi \leq \sqrt{2}$  et  $\bar{r} \leq 1$  et toute courbe d'utilité  $u$  de coefficient Lipschitz 1. C'est toutefois avec cette forme particulière de  $M$  (marges Rademacher) qu'on a pu observer les bornes plus serrées.

Ceci dit, si les bornes ne sont en tant que telles pas particulièrement fortes, l'ordre  $\mathcal{O}(n^{-1/2})$  qu'elles indiquent est lui très bien respecté empiriquement et il pourrait donc être possible d'anticiper de combien l'erreur empirique peut décroître selon la taille de l'échantillonnage en interpolant tout simplement les erreurs déjà observées avec un polynôme  $\mathcal{O}(n^{-1/2})$ .

**Erreur en util et en équivalent certain** Quant à l'erreur théorique et sa borne, on constate qu'il y a en fait peu de distorsion entre le domaine de l'util et celui du rendement. Pour expliquer ce phénomène, la Fig. 6 (p. 11) décompose l'erreur de généralisation : d'une part sa partie erreur en échantillon et hors échantillon. On y observe que les valeurs obtenues en util, toutes inférieures à 0.6, entraînent une faible distorsion si on se fie à la Fig. 2 (p. 5).

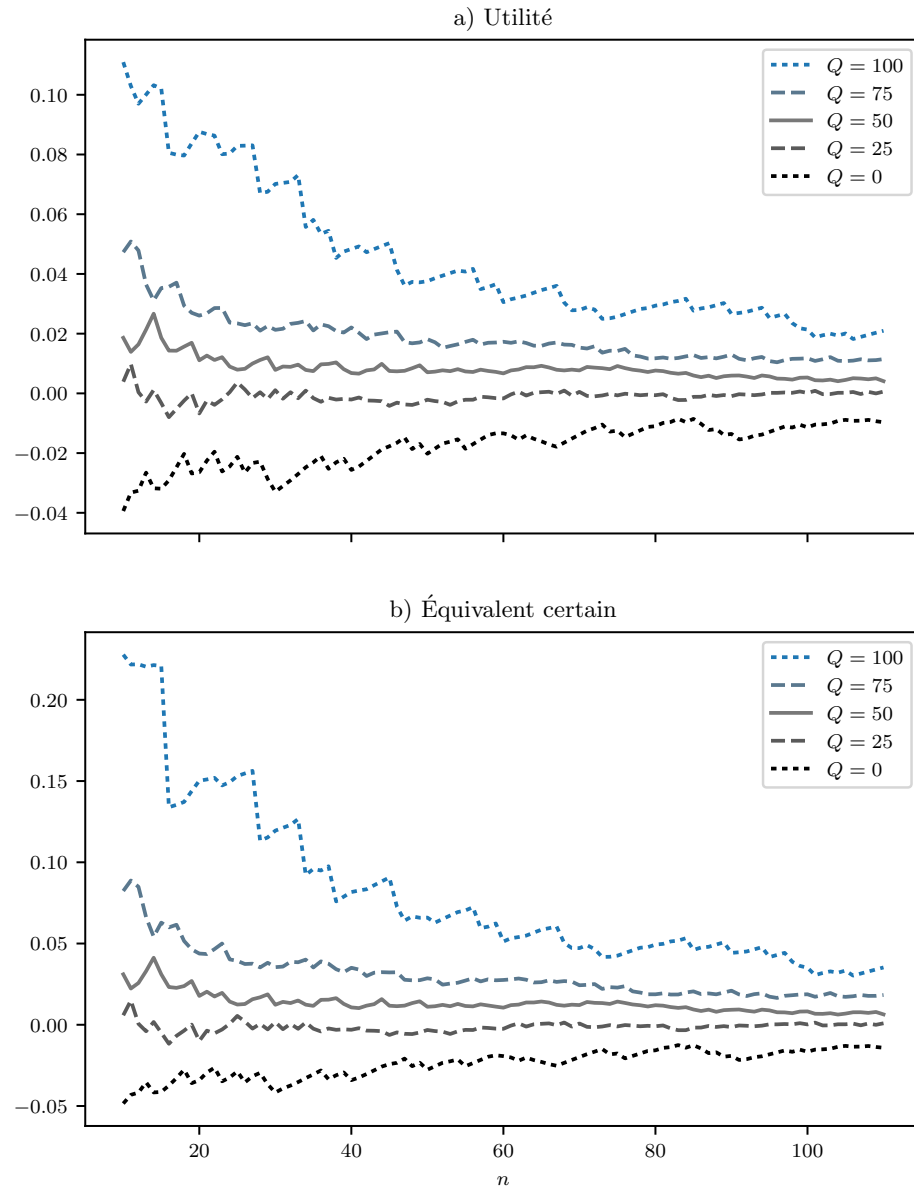


FIGURE 3 – Quartiles et valeurs maximales de l'erreur de généralisation



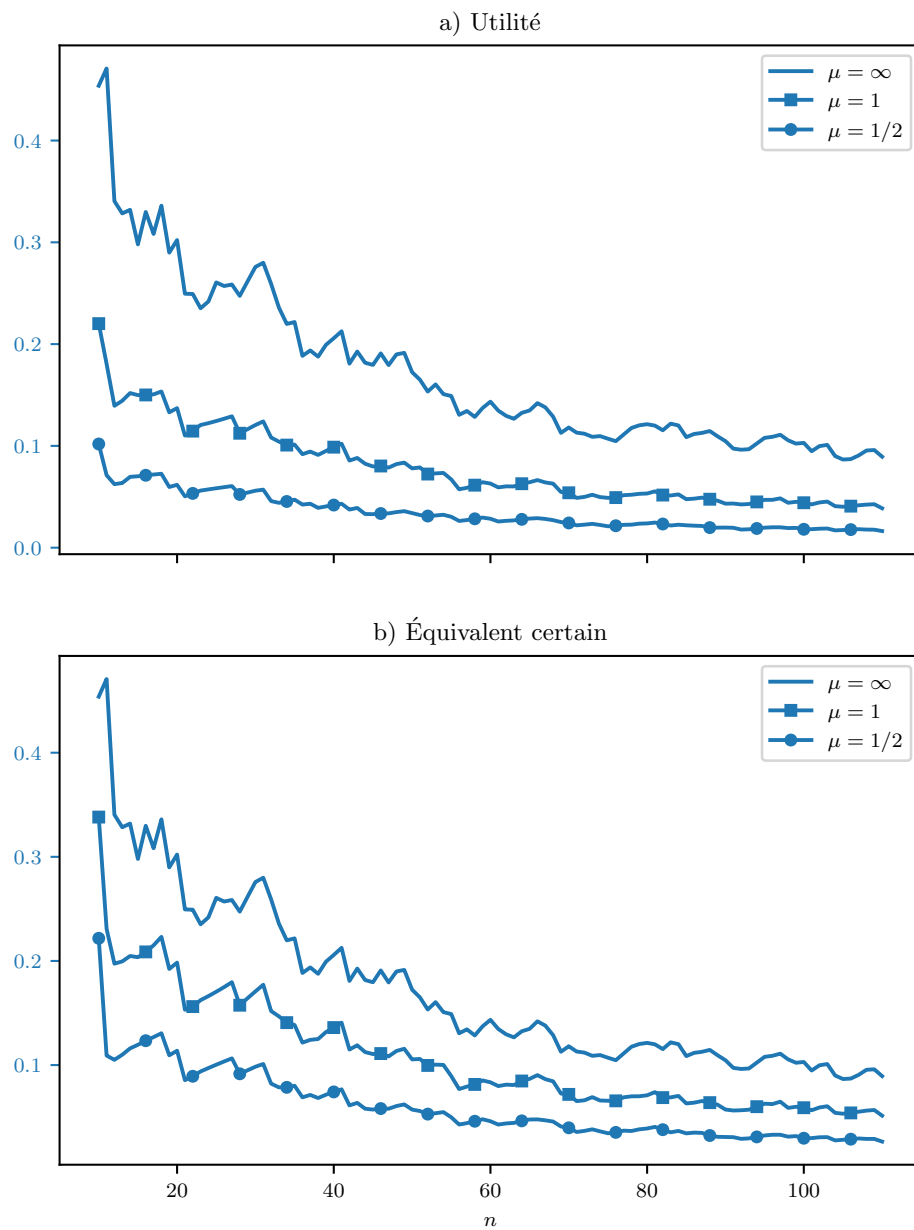


FIGURE 4 – Aversion au risque et erreur de généralisation (95<sup>e</sup> percentile)

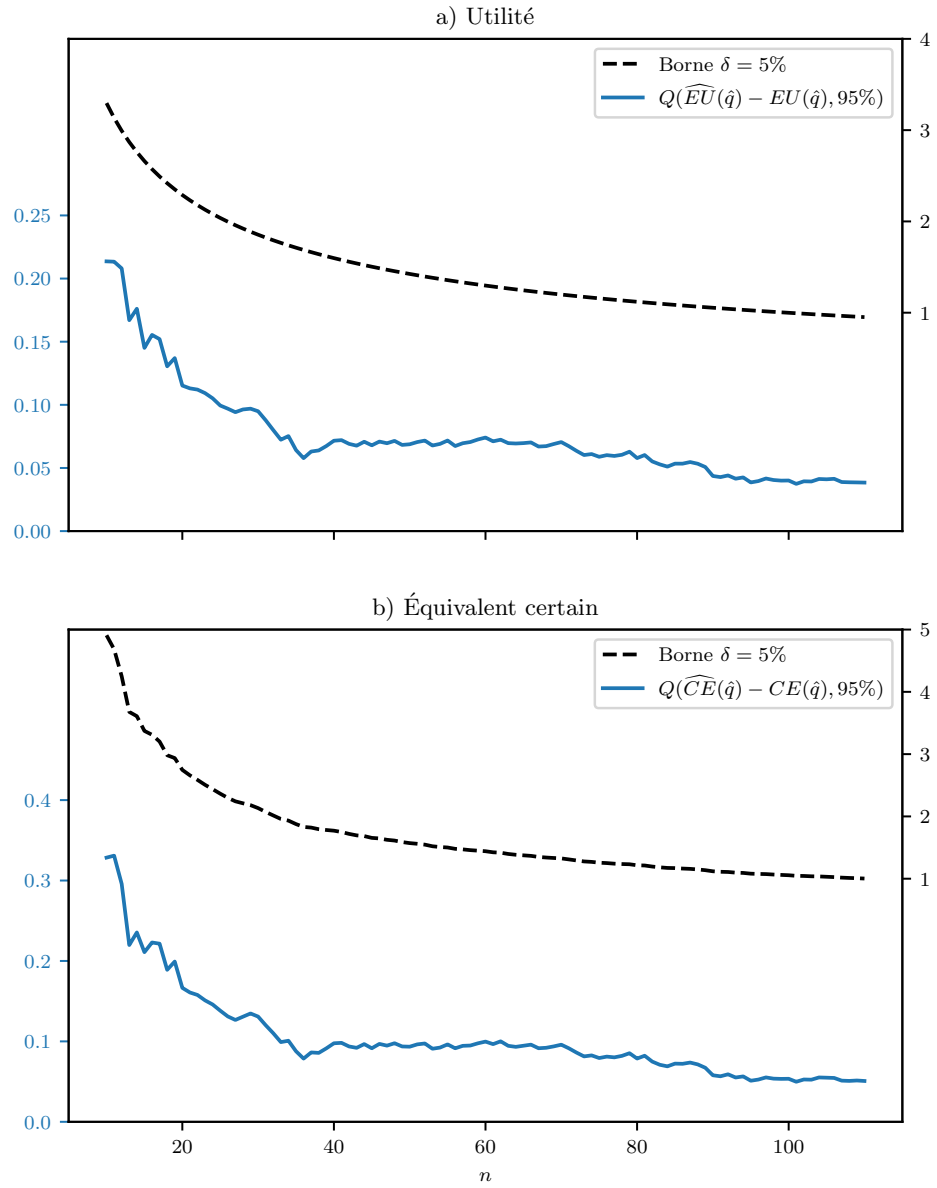


FIGURE 5 – Borne sur le 95<sup>e</sup> percentile de l'erreur de généralisation. L'axe de gauche indique la valeur de l'erreur empirique, tandis que l'axe de droite indique celle de la borne théorique.

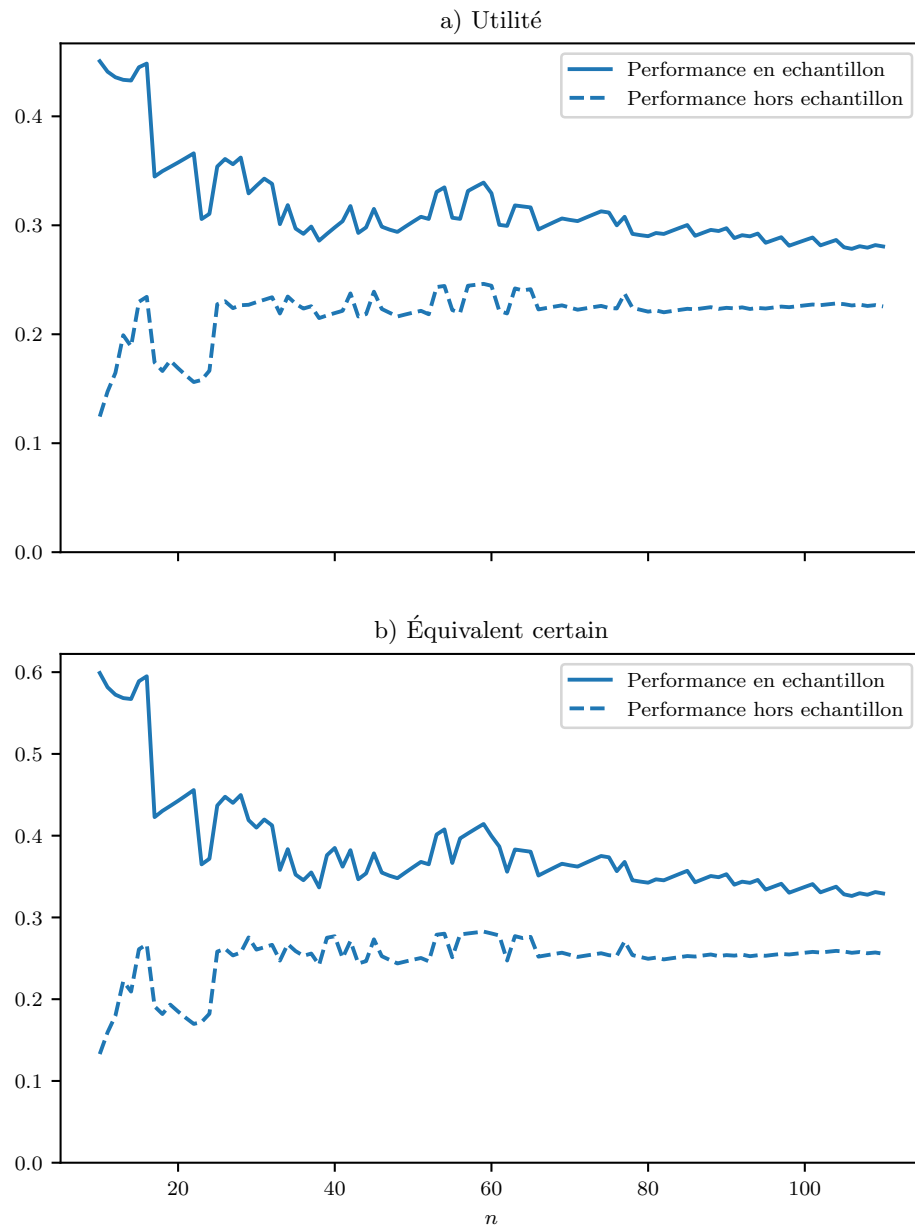


FIGURE 6 – Composantes en échantillon et hors échantillon de l'erreur maximale.

### 1.2.2 Erreur de sous optimalité

**$\lambda$  constant – Figure 7** Contrairement à l’erreur de généralisation, l’erreur (en util) de sous optimalité  $EU(q^*) - EU(\hat{q})$  (resp.  $CE(q^*) - CE(\hat{q})$ ) dans le domaine des rendements) ne bénéficie pas d’une convergence vers zéro du fait de la présence du terme de régularisation dans l’algorithme  $\mathcal{Q}(\mathcal{S}_n)$ . En fait, tel que vu au théorème <sup>[Citation needed]</sup>, pour  $n \rightarrow \infty$ , la meilleure borne qu’on puisse avoir est proportionnelle à  $\lambda \|q^*\|^2$  (domaine des utils).

Par exemple la Fig. 7 (p. 13) illustre précisément comment les erreurs empiriques et théoriques plafonnent toutes les deux à des constantes non nulles. De plus, contrairement à la borne de généralisation, la borne de sous-optimalité se trouve à deux ordres de grandeur de l’erreur empirique. Ceci a un effet particulièrement néfaste lorsqu’on considère la borne dans le domaine des rendements où l’effet de l’utilité inverse (voir Fig. 2 (p. 5)) se fait violemment sentir : puisque l’utilité de la politique optimale est proche de la limite asymptotique  $\lim_{r \rightarrow \infty} LEU_1(r) = 1$ , l’inversion  $u^{-1}(EU^*)$  retourne une valeur très élevée.

Néanmoins, tout comme c’était le cas pour l’erreur de généralisation, si la borne de sous optimalité ne donne pas nécessairement de fortes garanties, en revanche elle suggère un ordre de convergence qui lui semble être en adéquation avec l’erreur de sous optimalité empirique maximale (ou plutôt, avec son 95<sup>e</sup> percentile d’erreur).

**$\lambda$  décroissant – Figure 8** Comme il fut discuté à la section <sup>[Citation needed]</sup>, en utilisant un facteur de régularisation  $\lambda = \mathcal{O}(n^{-1/2})$ , on peut garantir une convergence de l’erreur de sous optimalité vers zéro (voir Fig. 8 (p. 14)). Cependant, si dans ce cas-ci l’erreur empirique de sous-optimalité, qu’elle soit exprimée en util ou en rendement, semble bien converger à un rythme  $\mathcal{O}(n^{-1/2})$ , la borne théorique elle ne progresse qu’à un rythme de  $\mathcal{O}(n^{-1/4})$ , ce qui est particulièrement lent. Par contre, même une progression aussi lente permet quand même d’obtenir des garanties d’équivalent certain un peu plus raisonnables puisqu’on force alors la limite  $\lambda \|q^*\|^2$  de la borne en util à s’éloigner de la région où  $u^{-1}$  retourne des valeurs très grandes.

Par contre, si cette décroissance de  $\lambda$  entraîne une convergence de l’erreur de sous optimalité, c’est au prix de la garantie sur l’erreur de généralisation, qui est elle proportionnelle à  $\mathcal{O}(\lambda^{-1})$ .

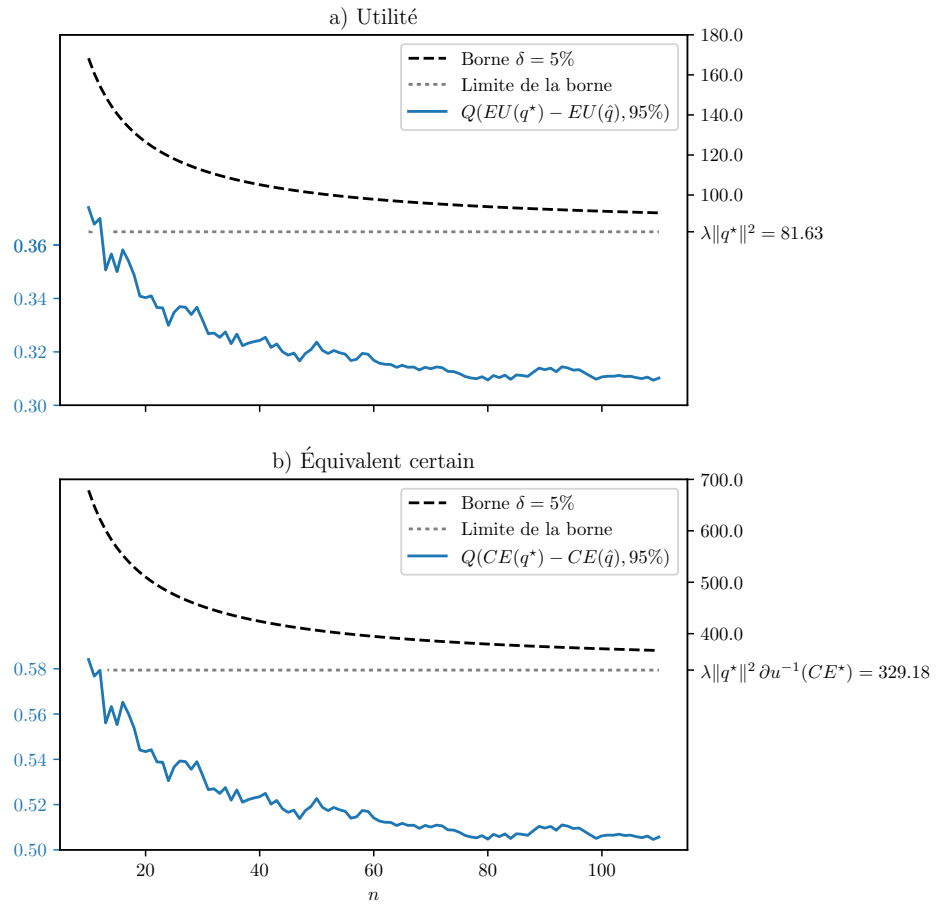
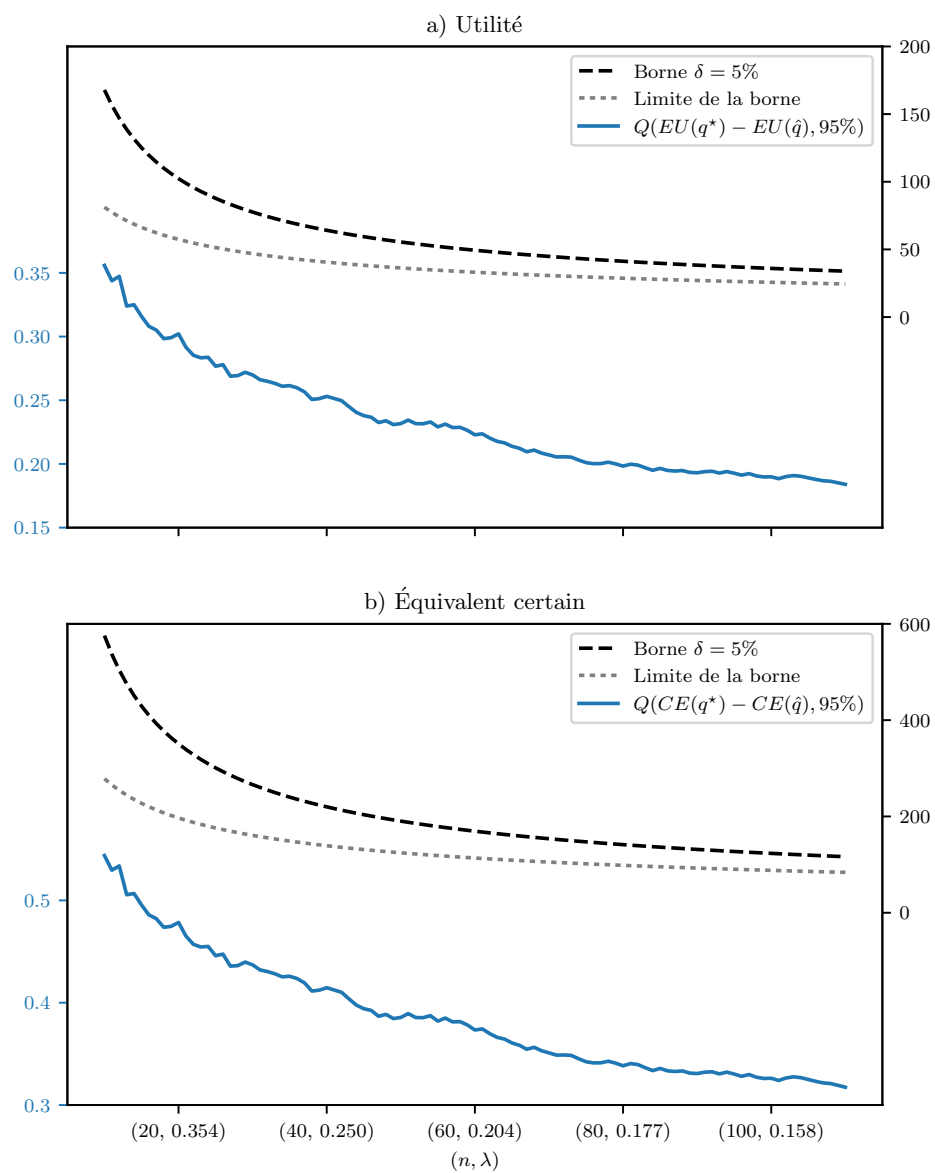


FIGURE 7 – Borne de sous optimalité,  $\lambda$  constant



### 1.3 $n$ constant, $p$ variable

On peut aussi considérer le rapport qu'entretennent les bornes de généralisation et de sous-optimalité de l'algorithme de maximisation d'utilité régularisé lorsqu'on ajoute de nouvelles informations indépendantes des précédentes, tout en conservant la taille d'échantillonnage constante.

**Protocole d'expérience** Afin de bien comprendre l'effet que peut avoir un régime en haute dimension sur les deux types d'erreur étudiées, une copule gaussienne avec marges Rademacher sera encore employée pour modéliser  $\tilde{M}$ . Cette fois cependant, cette copule disposera de  $\bar{p}$  marges d'information indépendantes dont chaque marge sera dévoilée progressivement. De plus, trois situations seront étudiées : celle où toute l'information est concentrée à la première marge, les autres étant indépendantes de  $R$ , celle où chaque marge dispose d'une corrélation de  $1/\sqrt{\bar{p}}$  avec  $R$  (information dispersée) et finalement celle où aucune information n'est présente pour déterminer  $R$ , c'est-à-dire que toutes les marges  $X_j$  sont indépendantes de  $R$ .

Mathématiquement,  $\tilde{M}$  est donc décrit par une copule gaussienne à  $\bar{p} + 1$  marges Rademacher dont la matrice de corrélation est paramétrée par un vecteur de corrélation  $\rho \in \mathcal{R}^{\bar{p}}$  :

$$\Sigma = \begin{matrix} & \begin{matrix} X_1 & \cdots & X_{\bar{p}} & R \end{matrix} \\ \begin{matrix} X_1 \\ \vdots \\ X_{\bar{p}} \\ R \end{matrix} & \begin{pmatrix} \ddots & & & | \\ & I_{\bar{p} \times \bar{p}} & & \rho \\ & & \ddots & | \\ - & \rho & - & 1 \end{pmatrix} \end{matrix}. \quad (5)$$

Le cas de l'information concentrée se traduira par un vecteur de corrélation donné par

$$\rho = (1 \quad 0 \quad \cdots \quad 0), \quad (6)$$

celui de l'information dispersée par le vecteur de corrélation

$$\rho = (1/\sqrt{\bar{p}} \quad \cdots \quad 1/\sqrt{\bar{p}}), \quad (7)$$

et celui sans aucune information par le vecteur de corrélation

$$\rho = (0 \quad \cdots \quad 0). \quad (8)$$

Enfin, les expériences qui suivent fixent le nombre total de variables d'information à  $\bar{p} = 50$ .

**Erreur de généralisation – Figure 9** On a déjà remarqué que la borne de généralisation affiche une croissance  $\mathcal{O}(\xi^2)$  ce qui, dans le cas d'un noyau linéaire, devrait se traduire par une progression linéaire  $\mathcal{O}(p)$ . Ainsi, la Fig. 9 (p. 17) montre effectivement une telle progression pour les trois situations énumérées à la section précédente.

En fait, la plupart des observations qui ont été faites dans le cas où  $p$  est constant et  $n$  est variable peuvent être réutilisées. Par exemple, on remarque que la même différence d'un ordre de grandeur entre l'erreur empirique et théorique persiste à mesure qu'on dévoile de nouvelles variables d'informations  $X_j$ . Et ici encore, on perçoit une faible dilatation de valeur entre la borne exprimée en util et en rendement.

Pour ce qui concerne les trois situations d'information, bien que chacune d'entre elles affichent à peu près la même progression linéaire, la situation où toute l'information est concentrée dès  $p = 1$  entraîne d'abord une erreur nulle, qui augmente à mesure que de nouvelles variables "de bruit" sont ajoutées. On remarquera par ailleurs la forte similarité entre les courbes *Information dispersée* et *Aucune information*. En effet, comme  $\bar{p}$  est assez important, lorsque  $p = 1$  et que le signal est dispersé, le signal perçu à partir d'une seule caractéristique est très faible. Néanmoins, la courbe d'information diluée fléchit par rapport à celle de l'absence complète d'information à mesure que  $p$  augmente vers  $\bar{p}$ , conformément à l'intuition qu'on pourrait en avoir.



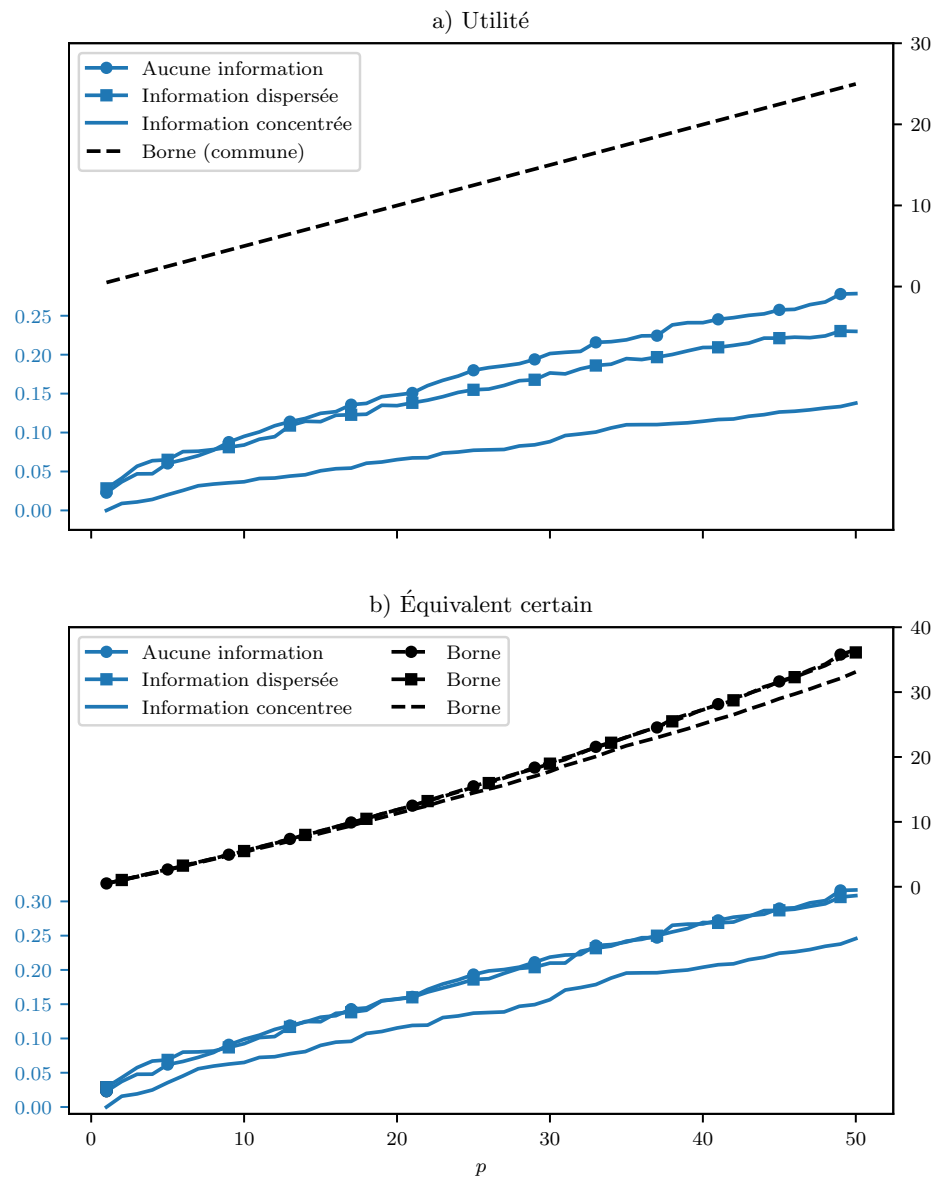


FIGURE 9 – Erreur de généralisation avec ajout d'information

### 1.3.1 Sous optimalité

**Utilité espérée optimale** Dans le cas où on ajoute de l'information, la sous optimalité, contrairement à l'erreur de généralisation, peut référer à deux types d'erreur. Soit on compare la performance hors échantillon de  $\hat{q}$  à celle de la politique optimale qui ne dispose que de  $p < \bar{p}$  variables d'information, soit à la politique optimale qui dispose de  $\bar{p}$  variables d'information nécessaires pour décrire  $M$ . Cependant, le développement théorique qui a été mené au cours de la dernière section ne s'est implicitement préoccupé que de la première situation.

Par exemple, la Fig. 10 (p. 19) indique la progression de l'utilité espérée optimale à mesure que de nouvelles variables d'information sont dévoilées. On y observe sans surprise que le cas où toute l'information est disponible dès  $p = 1$  affiche une utilité espérée optimale constante, alors qu'on a une progression à peu près linéaire lorsqu'on dévoile progressivement des variables d'information qui sont chacune faiblement corrélées à  $R$ , mais indépendantes l'une à l'autre. Enfin, aucune information se traduit inévitablement par une utilité espérée optimale nulle.

**Erreurs de sous optimalité** La Fig. 11 (p. 20) elle, indique la sous optimalité relative à mesure que de nouvelles variables d'information sont dévoilées. On y note d'abord une progression de l'erreur pour les trois situations étudiées ; de plus, la différence entre le panneau a) de la Fig. 11 et la Fig. 10 est une manifestation de la présence du facteur de régularisation constant à mesure que  $p$  augmente. Finalement, si les courbes *information dispersée* et *aucune information* subissent peu de distortion entre le domaine des utils et celui des rendements, la courbe *information concentrée* affiche une énorme sous optimalité lorsqu'elle est exprimée en rendement. Cela s'explique par le fait que  $EU^*$  est très proche de 1 (numériquement  $1 - EU^* = 1.89 \times 10^{-10}$  lorsque  $p = 50$ ), ce qui entraîne un équivalent certain de l'ordre de 10, puisque  $CE^* = -\log_e(1.89 \times 10^{-10})$ .

**Borne sur l'erreur de sous optimalité – Figures 12, 13 et 14** Ces trois figures traduisent comment la borne théorique sur l'erreur de sous optimalité se comporte pour chacune des trois situations explorées ici. Tout d'abord, on remarque que pour chacune d'elle la courbe théorique permet bel et bien de borner la courbe empirique. Cependant, comme c'était le cas pour l'erreur de généralisation, on observe à nouveau un décalage de deux ordres de grandeur entre les courbes. Cependant, le caractère *linéaire* qu'annonce la borne théorique semble se matérialiser empiriquement. Pour ce qui concerne la borne de la situation avec information concentrée dès la première variable d'information, on constate qu'exprimer la sous optimalité en terme d'équivalent certain peut donner lieu à des situations aberrantes, puisqu'on obtient en effet une borne théorique de l'ordre de  $10^{13}$ .

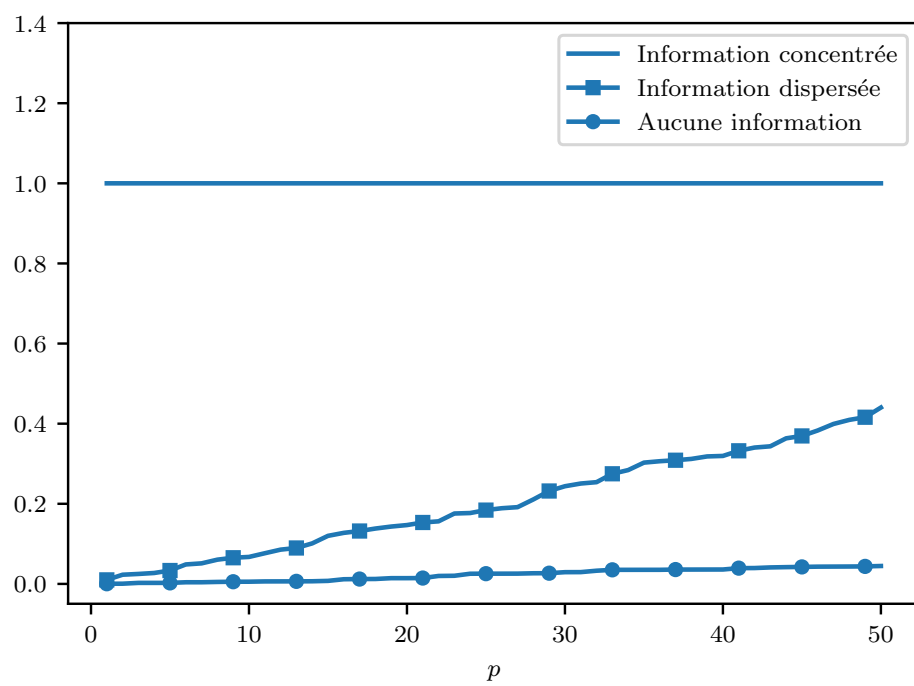


FIGURE 10 – Progression de  $EU^*$  relatif

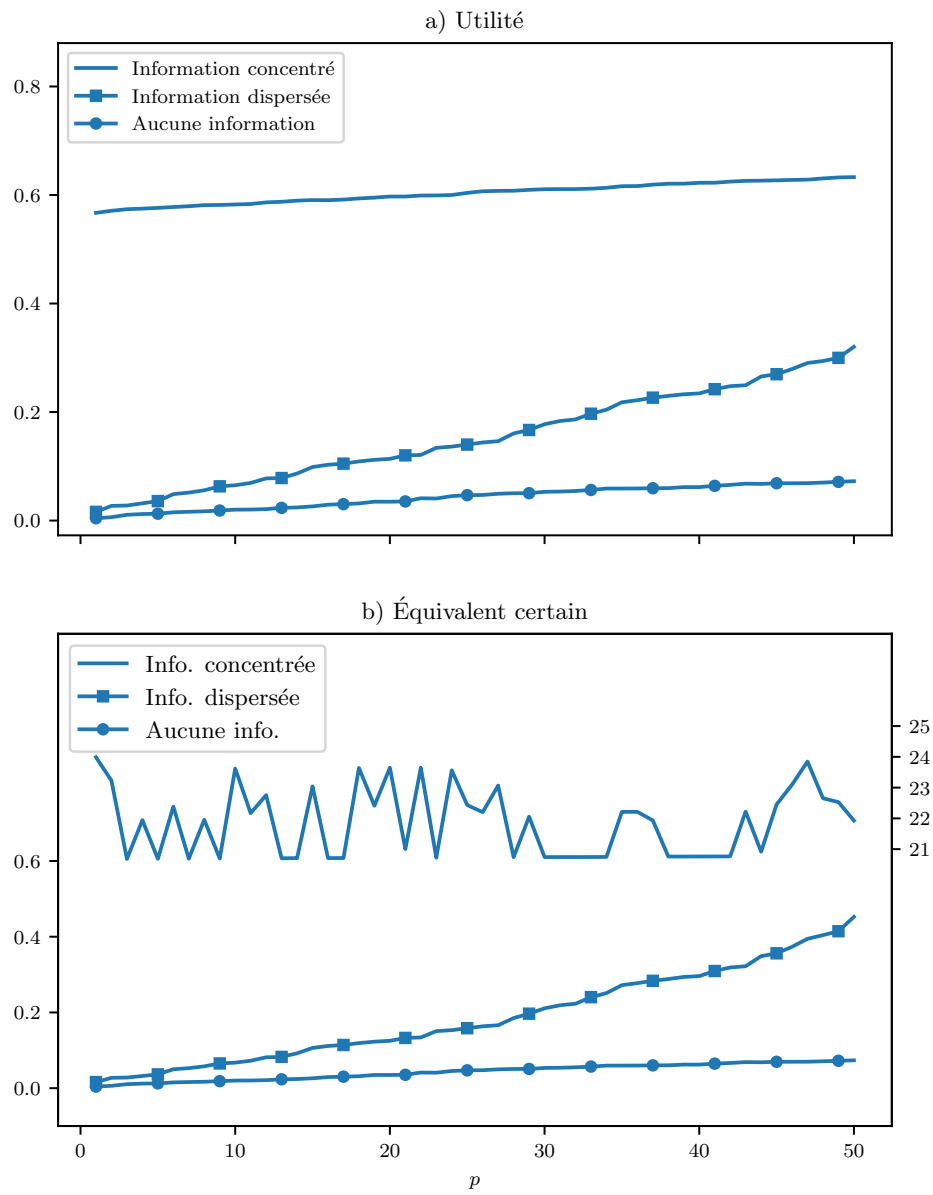


FIGURE 11 – Progression de la sous optimalité relative

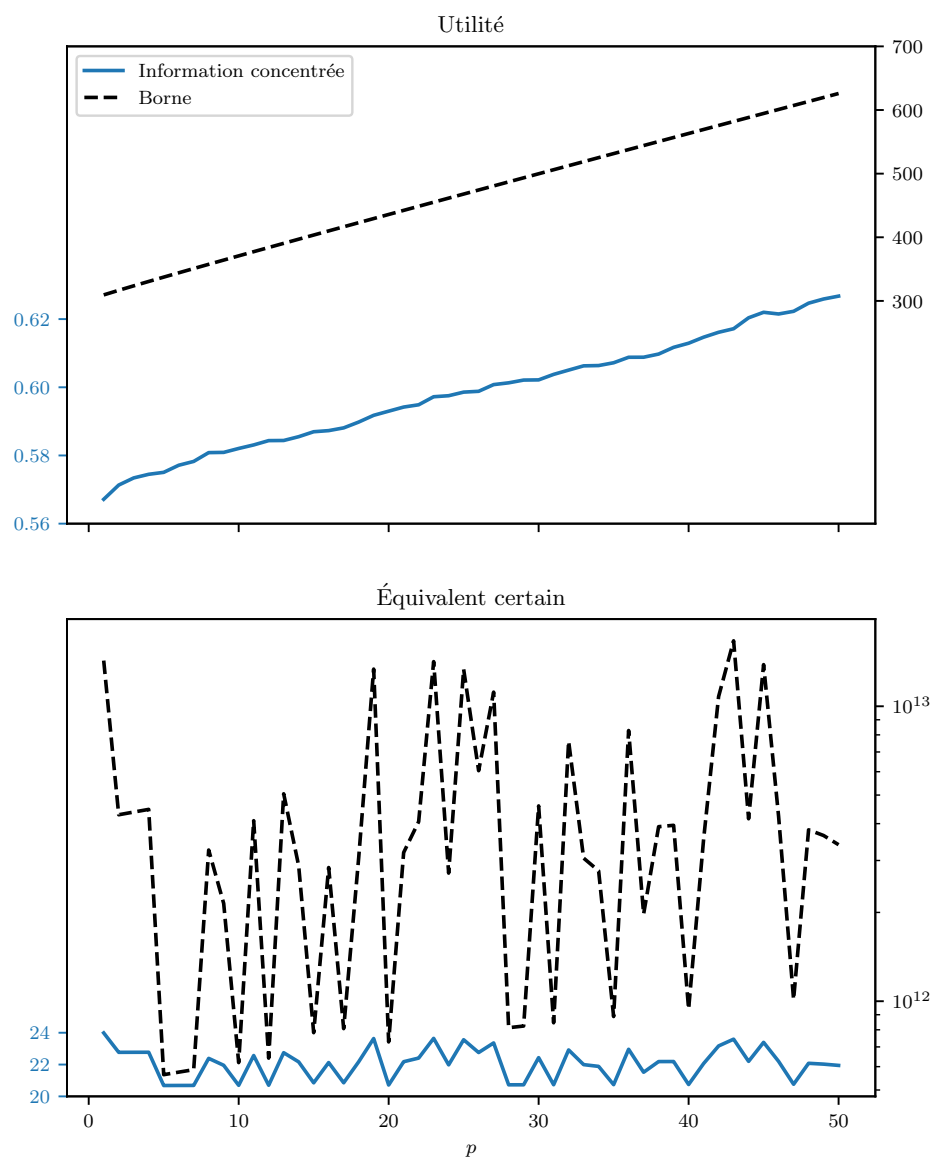


FIGURE 12 – Borne sur l'erreur de sous optimalité, information concentrée

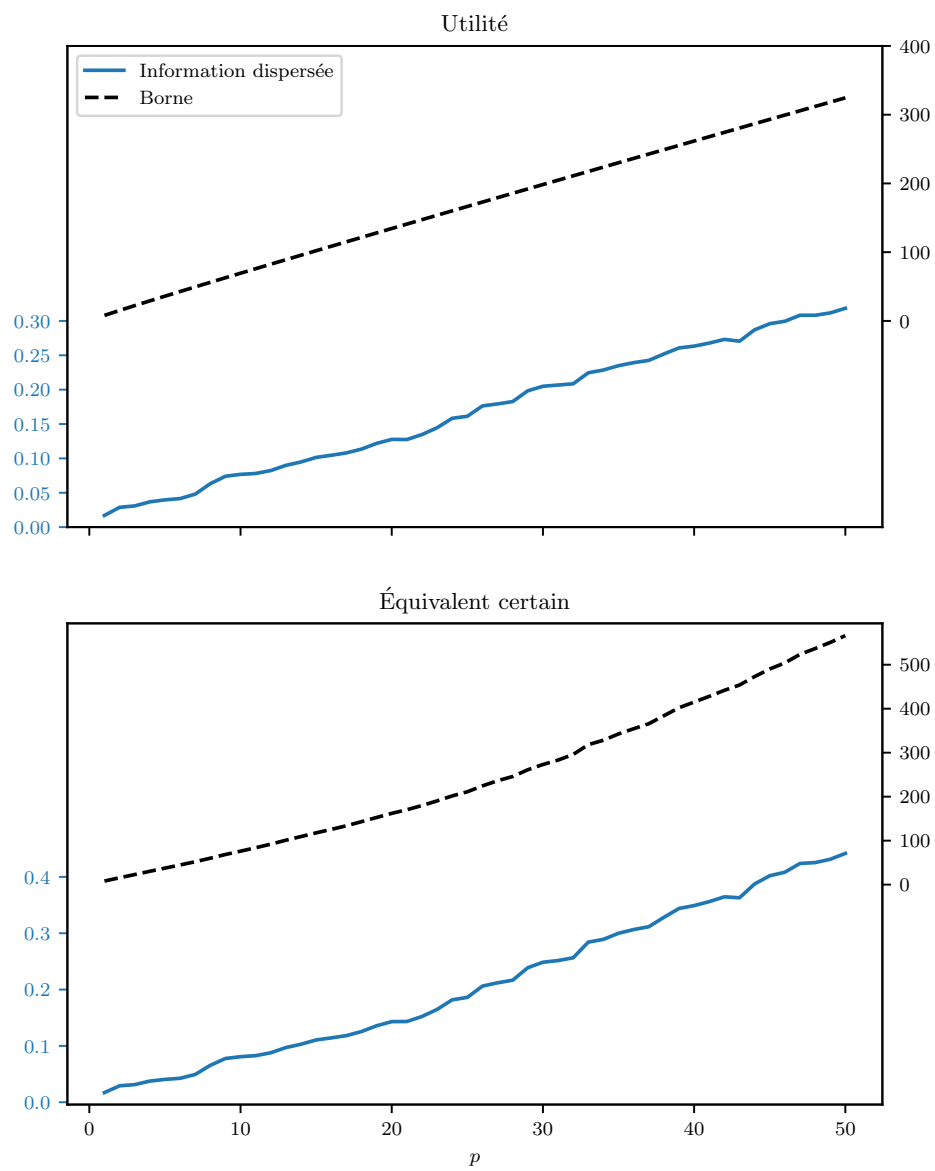


FIGURE 13 – Borne sur l'erreur de sous optimalité, information dispersée

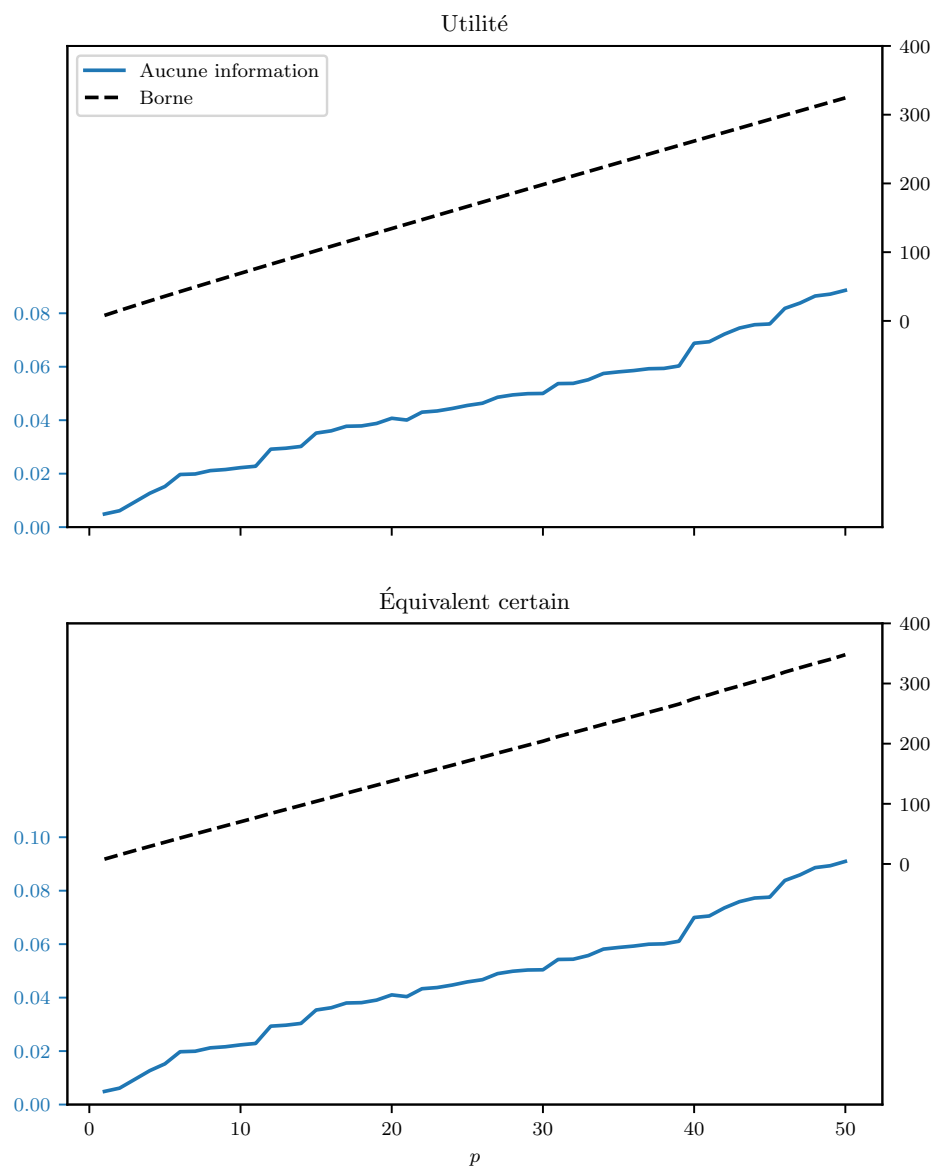


FIGURE 14 – Borne sur l'erreur de sous optimalité, aucune information

## 1.4 Ajout d'information et d'échantillons

Cette section cherche à illustrer les comportements des deux bornes lorsqu'on est en présence de régimes dynamiques en  $n$  et en  $p$ , *i.e.* lorsque  $p = (n^k)$ .

**Méthodologie** Cette fois, la loi de marché théorique  $\tilde{M}$  disposera de  $\bar{p} = 100$  variables d'information, encore liées à  $R$  à partir d'une copule gaussienne  $\Sigma$  donnée par

$$\Sigma = \begin{matrix} & \begin{matrix} X_1 & \dots & X_{\bar{p}} & R \end{matrix} \\ \begin{matrix} X_1 \\ \vdots \\ X_{\bar{p}} \\ R \end{matrix} & \begin{pmatrix} \ddots & & & | \\ & I_{\bar{p} \times \bar{p}} & & \rho \\ & & \ddots & | \\ - & \rho & - & 1 \end{pmatrix} \end{matrix}. \quad (9)$$

où cette fois,  $\rho = (1/\sqrt{100} \dots 1/\sqrt{100})$ .

Trois régimes seront étudiés : celui où  $p = \mathcal{O}(n^{1/2})$ ,  $p = \mathcal{O}(n^{3/4})$  et  $p = \mathcal{O}(n)$ . À des fins de simplicité, chacun de ces régimes sera décrit par la puissance  $k$  de  $n$ , de manière à avoir  $k = 1/2, 3/4$  et  $1$ . La figure 15 illustre précisément les trois régimes.

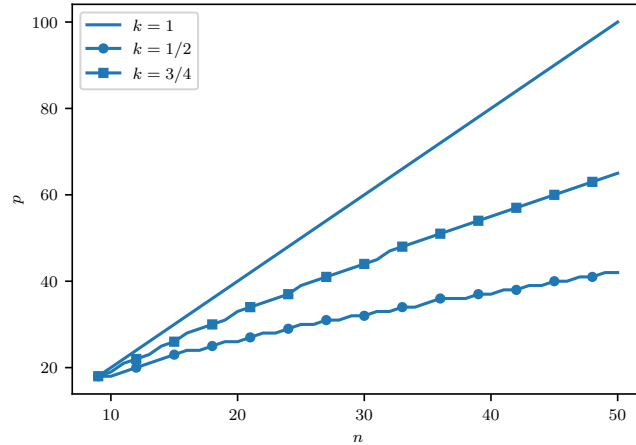


FIGURE 15 – Progression de  $p$  par rapport à  $n$

### 1.4.1 Erreur de généralisation

**Borne de généralisation – Figure 16** La Fig. 16 (p. 26) illustre enfin le problème avec la théorie présentée jusqu'à présent. Jusqu'à présent, même si les bornes théoriques dérivées n'étaient pas nécessairement serrées, elles servaient tout de même de guide pour connaître le régime dans lequel l'erreur empirique progressait. Par exemple,



lorsque  $p$  était fixe, on obtenait systématiquement une erreur de généralisation décroissant avec un rythme  $\mathcal{O}(n^{-1/2})$ . De la même façon, avec  $n$  constant, ces mêmes erreurs évoluaient selon un régime  $\mathcal{O}(p)$ . On se serait donc attendu à ce qu'on obtienne un régime global d'erreur  $\mathcal{O}(p/\sqrt{n})$ . Cependant, la Fig. 16 (p. 26) démontre clairement que tel n'est pas le cas, puisque peu importe le régime, là où on se serait attendu soit au moins à une erreur constante, on observe plutôt une diminution de l'erreur de généralisation !

**Régime d'ordre plus élevé – Figure 17** Il faut effectivement attendre de voir apparaître des régimes très élevés (par exemple  $k = 3/2$ ) avant d'enfin observer une augmentation de l'erreur. Cependant, l'ordre exact  $k$  ne semble pas clair et certaines expériences laissent croire que  $k$  pourrait plutôt dépendre des paramètres du problème.

#### 1.4.2 Erreur de sous optimalité

La Figure 18 présente quant à elle l'erreur de sous optimalité pour les trois régimes étudiés et une fois encore, on constate une vive différence entre l'ordre donné par la borne théorique et ce qu'on observe en pratique. La courbe empirique  $k = 1/2$  à peu près stable laisse croire qu'on est en présence d'un régime  $\mathcal{O}(p/\sqrt{n})$ , mais la courbe  $k = 1$  vient en fait démentir cette interprétation puisque la progression paraît alors linéaire, un peu comme ce qui était observé lorsque  $n$  était constant.

De plus, la forme même des bornes théoriques permet de constater les deux régimes théoriques superposés, c'est-à-dire de la forme  $\mathcal{O}(p/n + p/\sqrt{n})$ . Par ailleurs, la présence du terme de régularisation vient une fois de plus brouiller les interprétations, puisqu'il faut conserver à l'esprit qu'un terme  $\lambda \|q^*\|^2$  est aussi compris dans la borne, et dont la progression est encore mal comprise (voir Fig. 19 (p. 29)).

## Références

- [DB16] Steven Diamond and Stephen Boyd. CVXPY : A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83) :1–5, 2016.
- [DCB13] A. Domahidi, E. Chu, and S. Boyd. ECOS : An SOCP solver for embedded systems. In *European Control Conference (ECC)*, pages 3071–3076, 2013.
- [Ré13] Bruno Rémillard. *Statistical Methods for Financial Engineering*. CRC Press, 2013.

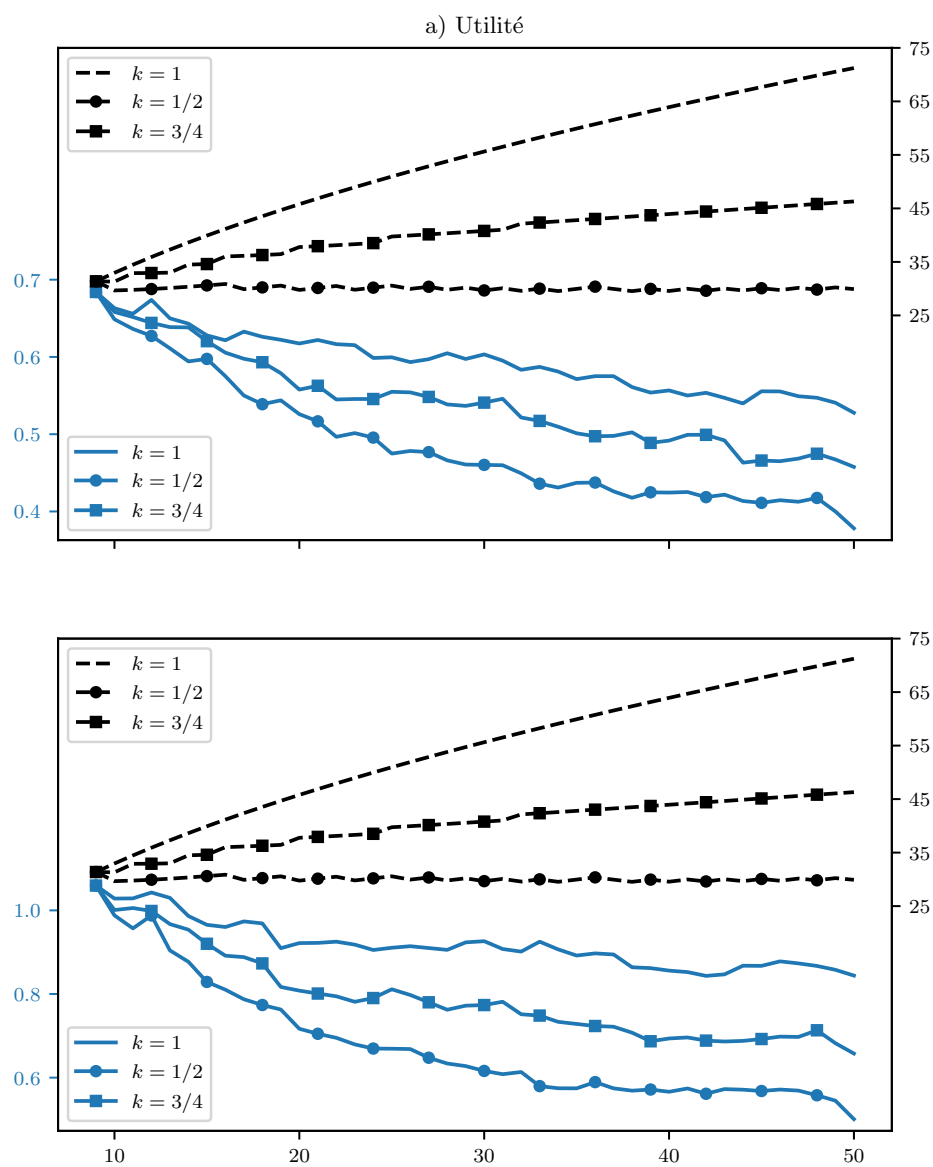


FIGURE 16 – Borne sur l'erreur de généralisation,  $n$  et  $p$  variables

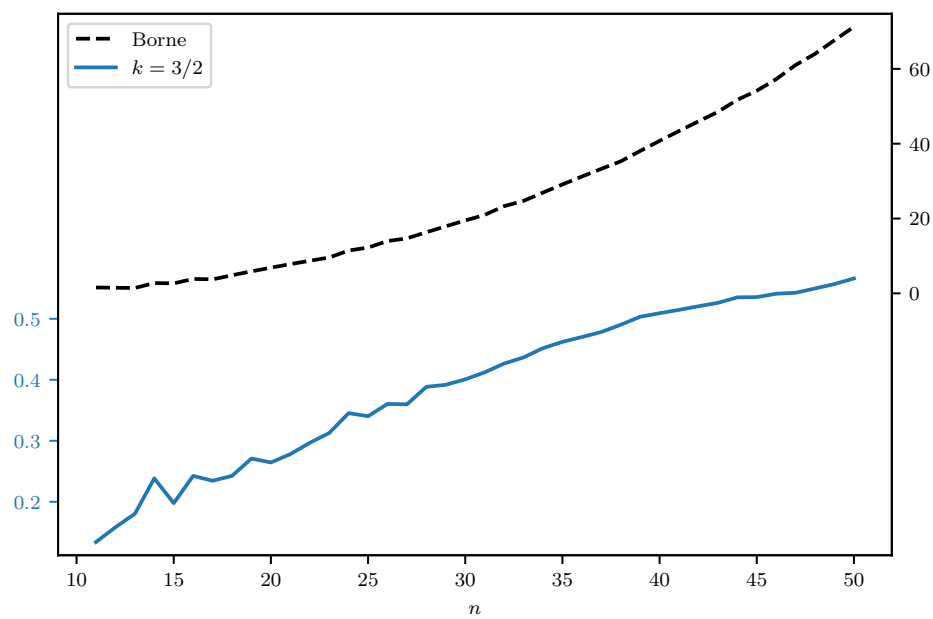


FIGURE 17 – Erreur de généralisation à très haut régime

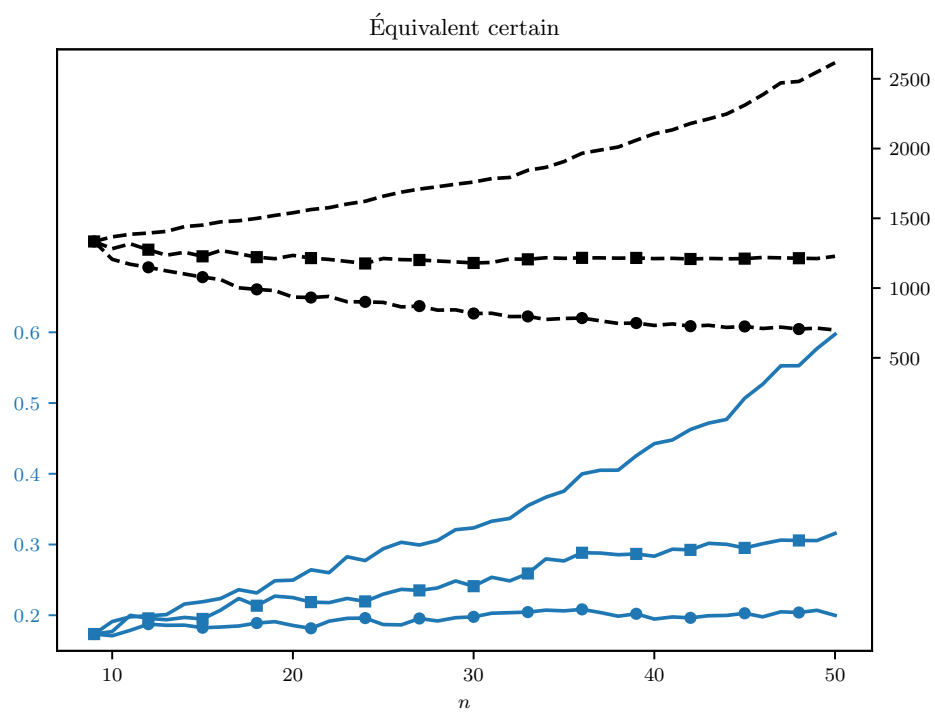
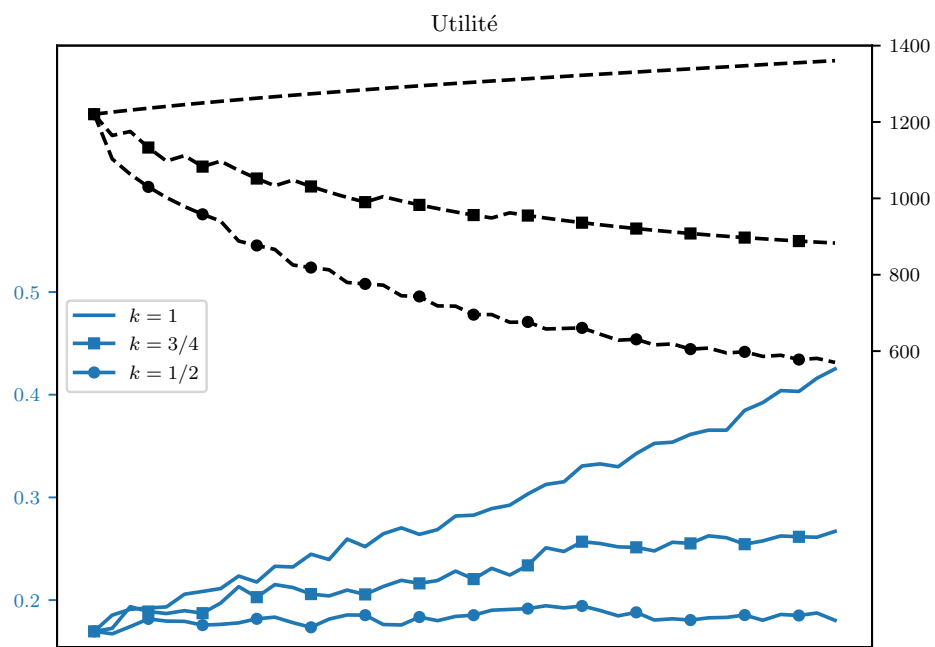


FIGURE 18 – Borne sur l'erreur de sous-optimalité,  $n$  et  $p$  variables

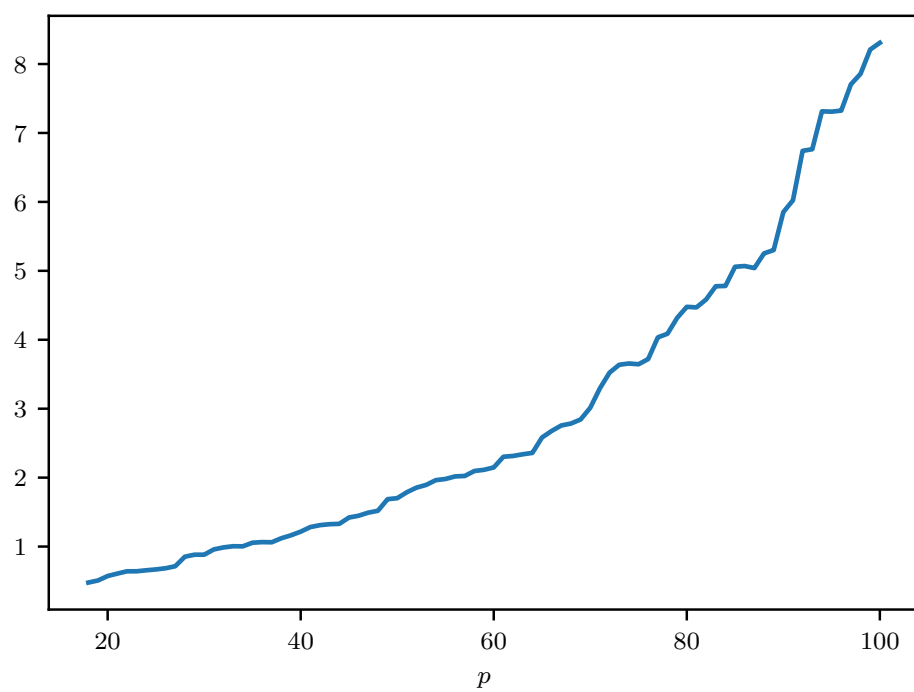


FIGURE 19 – Progression de  $\|q^*\|^2$