# Stage de deuxième année

*juillet 2017 - août 2017*

Alexandre Fauquette
*Promotion 2018*

*CentraleSupélec – Campus de Gif*

CentraleSupélec

*Encadrante :*
Johanne COHEN

*Laboratoire d'acceuil :*
LRI (*Laboratoire de Recherche en Informatique*)

LABORATOIRE DE RECHERCHE EN INFORMATIQUE

# Table des matières

# Table des figures

# 1 Remerciements

Tout d'abord, je souhaite remercier la direction des Etudes du campus de Gif de CentraleSupélec, et tout particulièrement Mme Christine Despretz d'avoir approuvé ce stage.

Je remercie Mme TOMASIK Joanna et M RIMMEL Arpad de m'avoir aidé à trouver ce stage.

Je remercie tout particulièrement Mme COHEN Johanne de m'avoir fait confiance pour l'exécution de ce stage et de m'avoir proposé une telle opportunité. Bien évidemment, elle m'a également apporté son soutien pendant toute la durée du stage a la fois sur le travail à effectuer et la découverte du monde de la recherche.

Ensuite, je souhaite remercier les personnes avec qui j'ai travaillé sur ce problème pour leur soutient.

Enfin, je souhaite mentionner les chercheurs, les doctorants et les stagiaires de l'équipe GALAC avec qui j'ai pue échanger tout au long de ce stage.

# 2 Introduction

Ce stage s'inscrit dans le cadre des stages de deuxième année à CentraleSupélec. Le stage de deuxième année à pour objectif la mise en pratique dans un cadre professionnel les connaissances acquises durant les deux premières années du cursus. Ainsi que la découverte du monde professionnel, au travers de son organisation hiérarchique, son fonctionnement pratique et l'organisation du travail.

Je souhaite m'orienter vers le domaine de la recherche. Il était donc important pour moi de trouver un stage dans un laboratoire ou un organisme de recherche. Et plus particulièrement dans la recherche en informatique. Durant la deuxième année, j'ai participé à un projet long sur une application des réseaux de neurones et l'apprentissage profond. Il s'agissait d'un sujet de recherche très orienté pratique, avec une grande part donnée à la programmation et aux expériences. J'ai donc voulu pour ce stage quelque chose de plus théorique afin d'élargir ma vision de la recherche en informatique.

Dans un premier temps, je présente le sujet de mon stage en explicitant son contexte. Dans un second temps, je proposerai une bref présentation des définitions des objets mathématiques manipulés, ainsi que les résultats présent dans la littérature utilisés tout au long de ce stage. Enfin, je résume les apports variés qu'a eut ce stage pour ma formation.

# 3 Contexte

## 3.1 Le laboratoire

Le LRI (Laboratoire de Recherche en Informatique) d'Orsay est situé sur le plateau du Moulon, aux côtés d'autres établissements où la recherche informatique est présente (Supélec, Polytech, Université Paris-Sud,...). Aujourd'hui, le LRI fait partie de l'Université Paris-Saclay regroupant 2 universités et 10 grandes écoles. Il constitue une unité mixte de recherche de l'Université Paris-Sud et du CNRS. Il est composé de huit équipes de recherche regroupant en tout 240 personnes dont une centaine de doctorants. Le directeur du LRI est M. Yannis MANOUSSAKIS. Outre le pôle recherche, le laboratoire est doté d'équipes technique et administrative.

FIGURE 3.1 – Façade Sud du LRI

J'y ai intégré l'équipe GALAC (Graphes, Algorithmes et Combinatoire) qui est composée de 18 membres permanents et 12 membres non-permanents. Mme COHEN Johanne est responsable de l'équipe.

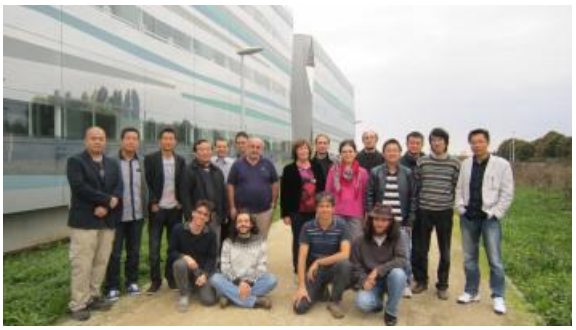

FIGURE 3.2 – Photo de groupe de l'équipe GALaC

Mais ma mission n'était pas centrée sur l'étude des graphes. En effet, il s'agissait d'un projet interdisciplinaire entre mathématiques et informatique, servant à la création d'un outil informatique dédié aux bornes stochastique XBorne [**?**]. Le projet en parti soutenu par le le PEPS MASTTODONS du CNRS qui a pour vocation d'encourager la réalisation de travaux interdisciplinaires portant sur le thème dans grands jeux de données, que ce soit des améliorations techniques, ou des études sur les impacts des "big-data" sur la société et leurs enjeux éthiques.

# 4 Travail réalisé

## 4.1 Présentation du sujet

### 4.1.1 Introduction au problème

L'idée générale à l'origine de l'article est la suivante. La plupart des algorithmes actuels résolvent des problèmes d'optimisation dans des cadres déterministes. On peut citer comme exemple classique trouver les plus court chemin dans un graphe, ou calculer la durée minimale de réalisation d'une graphe de tâche. Mais ces algorithmes partent d'une représentation simplifiée du monde. En effet, toute personne habitant dans la région Ile de France sait que la durée pour traverser le périphérique n'est pas une constante.

On pourrait approximer la durée de traversée du périphérique par une fonction du temps. Mais là encore ce serrait très réducteur, car personne ne sait exactement quand commencent les embouteillages. Une autre solution est de modéliser cette durée de trajet comme une variable aléatoire. Ainsi, traverser le périphérique parisien ne prend plus $t$ heures, mais $T$ heures , une variable aléatoire, définie par $P(T = t)$ la probabilité que la durée de traversée soit de $t$ heures. Le problème est que plus notre variable aléatoire à de valeurs possibles, plus la complexité en temps de calcul augmente, rendant impossible son calcul.

La proposition faite est de diminuer le nombre de valeurs pouvant être prise par la variable aléatoire. C'est à dire proposer une variable aléatoire avec moins de valeurs tout en s'assurant que la solution de l'algorithme donnera une réponse suffisamment proche de ce qu'il aurait du donner si il avait eu toutes les valeurs. Pour plus formel, l'objectif est de chercher deux distributions qui vont encadrer notre distribution d'origine suivant la relation d'ordre convexe. Et de plus s'assurer que ces bornes supérieures et inférieures soient les plus proche possible de notre distribution initiale.

Pour résumer, le but du stage est donc de trouver une méthode algorithmique, qui à partir d'une variable aléatoire $X$ pouvant prendre $N$ valeurs nous donnera deux distribution $Y_1$ et $Y_2$ pouvant prendre $K < N$ valeurs. Telles que $Y_1 \preceq_{cx} X \preceq_{cx} Y_2$ où $\preceq_{cx}$ est la relation d'ordre convexe.

### 4.1.2 Pourquoi la relation d'ordre convexe

La relation d'ordre convexe a été choisie, car l'addition, la fonction max et min sont monotone sur cette relation d'ordre. C'est à dire que si $X \preceq_{cx} Y$, alors

$max(X, a) \preceq_{cx} max(Y, a)$. Comme un grand nombre de problèmes sont basés sur ces trois opérations, résoudre le problème pour les sous-distributions encadrant notre distribution d'origine donne un encadrement de la solution. Ainsi, si $f(X)$ est la solution d'un problème basé sur le fonction $(max, min, +)$ et que $Y_1 \preceq_{cx} X \preceq_{cx} Y_2$, alors $f(Y_1) \preceq_{cx} f(X) \preceq_{cx} (Y_2)$. On a donc un encadrement de la solution recherchée.

Trouver une sous distribution de $K$ valeurs parmi les $N$ valeurs de la distribution d'origine n'est pas compliqué. L'objectif est de trouver le meilleur encadrement. En effet, si $A \preceq_{cx} B \preceq_{cx} X$, on aura $f(A) \preceq_{cx} f(B) \preceq_{cx} f(X)$. Ainsi $B$ donnera une meilleure borne inférieure de $f(X)$ que $A$.

### 4.1.3 Propriétés de la borne convexe

Vue que tout le stage est dédié à l'étude des bornes convexe, voici une explication rapide des résultats issues de la littérature [?, ?]. Résultats que Jean-Michel nous avait résumé pour le début du stage.

La définition de base de la borne convexe est la suivante : $A \preceq_{cx} B$ lorsque pour toute fonction $\phi$ convexe, $E[\phi(A)] \leq E[\phi(B)]$.

Pour rappel, dans le cas de distribution discrète, l'espérance $E[\phi(X)] = \sum_i \phi(x_i)p_i$ où $x_i$ et $p_i$ sont les valeurs pouvant êtres prise par $X$ et la probabilité associée. Et une fonction convexe est une fonction telle que quels que soient deux points $A$ et $B$ du graphe de la fonction, le segment $[A, B]$ est entièrement situé au-dessus du graphe.

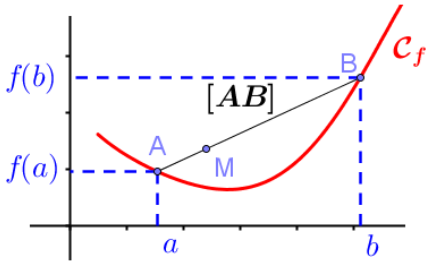


FIGURE 4.1 – Exemple de fonction convexe.

Cette définition bien que simple à exprimer n'est pas la plus aisée à manipuler. On a souvent eu recours à la propriété suivante :

$$A \preceq_{cx} B \Leftrightarrow \begin{cases} E[A] = E[B] \\ \text{et} \\ \forall d \in \mathbb{R}, E[(A-d)^+] \leq E[(B-d)^+] \end{cases}$$

La fonction $(x)^+ = max(0, x)$. On notera alors que $E[(A-d)^+] \leq E[(B-d)^+]$ peut se réécrire sous la forme suivante :

$$\sum_{a_i > d} (a_i - d)P(A = a_i) - \sum_{b_i > d} (b_i - d)P(B = b_i) \leq 0$$

Une autre propriété grandement utilisée, est que si l'on coupe la distribution $X$ en deux parties $X_1$ et $X_2$ ainsi que la distribution $Y$ en $Y_1$ et $Y_2$, on a la propriété suivante :

$$\left.\begin{array}{l} X_1 \preceq_{cx} Y_1 \\ X_2 \preceq_{cx} Y_2 \end{array}\right\} \Rightarrow X \preceq_{cx} Y$$

Comme de plus $X_2 \preceq_{cx} X_2$, on peut s'intéresser au passage $X_1$ à $Y_1$ indépendamment de ce qui arrive à $X_2$.

## 4.2 Recherche d'une borne supérieure optimale

Dans le cas de la borne supérieures, nous nous sommes restreint aux sous distributions ayant leurs valeurs incluses dans la distribution d'origine. C'est à dire qu'à partir de la distribution $X$ pouvant prendre $N$ valeurs $x_1, \cdots, x_N$ avec des probabilités non nulles $p_1, \cdots, p_N$, on cherche une distribution $Y$ ayant $K$ valeurs notées $y_1, \cdots, y_K$ associées aux probabilités $q_1, \cdots, q_K$ telles que $\{y_1, \cdots, y_K\} \subset \{x_1, \cdots, x_N\}$ et que $X \preceq_{cx} Y$. On dira que notre solution est optimale (notée $Y_{opt}$) si pour toute solution $Y$ respectant les conditions définies ci-dessus (valeurs incluses dans $X$ et borne supérieure de $X$ suivant l'ordre convexe), on a $Y_{opt} \preceq_{cx} Y$.

### 4.2.1 Supprimer une valeur

Comme l'on cherche une distribution $Y$ ayant moins de valeurs possibles que $X$ la distribution d'origine, il nous faut étudier ce qui se passe lorsque l'on supprime un atome.

Le cas le plus simple est de prendre 3 valeurs de $X$ que l'on notera $a < b < c$ et leurs probabilités $p_a, p_b, p_c$. Si l'on veut que $Y$ ne contienne que $a$ et $c$ (associées aux probabilités $q_a$ et $q_c$) et soit comparable à $X$ suivant la relation d'ordre convexe, il faut remplir deux conditions :

- $q_a + q_c = p_a + p_b + p_c$ qui correspond au fait que la somme des probabilités est une constante valant 1
- $E[X] = E[Y]$ qui est une condition nécessaire pour comparer $X$ et $Y$ sur l'ordre convexe

On a donc deux équations qui permettent de trouver $q_a$ et $q_c$. Et le calcul montre que dans tous les cas, on a unicité de la solution. La solution correspond bien à une distribution (on n'a pas de probabilité sortant de l'intervalle $[0, 1]$) et supprimer la valeur centrale donne bien $X \preceq_{cx} Y$.

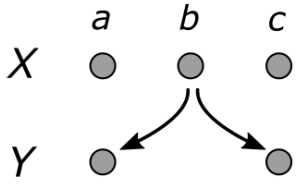


FIGURE 4.2 – Suppression d'une valeur.

On peut de plus montrer de plus que si on a quatre valeurs $a < b < c < d$, il est préférable de supprimer $b$ en répartissant sa probabilité sur $a$ et $c$ que sur $a$ et $d$. Ce qui nous permet de conclure qu'il vaut mieux supprimer une valeur en répartissant sa probabilité sur la valeur la plus proche à gauche et la valeur la plus proche à droite.

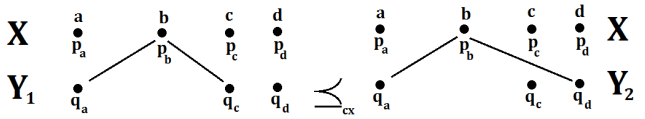


FIGURE 4.3 – Comparaison de deux suppressions possibles d'une valeur.

### 4.2.2 Première démonstration d'optimalité

A partir de cette propriété, on peut se dire qu'il ne faudrait réaliser que des suppressions de valeur sur des triplets consécutifs de valeurs. Ce qui revient à dire que si je veux supprimer une valeur je dois obligatoirement répartir sa probabilité sur la valeur située à gauche et sur celle à droite.

C'est sur cette idée que nous somme partis. Le problème c'est qu'en s'y prenant ainsi, on réalise à chaque étape une opération optimale. Mais rien ne garantie que la succession d'opérations optimale donnera une solution optimale. On a tenté pas mal de choses avant de se rendre compte qu'il ne fallait pas basée la démonstration d'optimalité sur une méthode de construction itérative de la solution. Mais supposer que l'on connait les $K$ valeurs appartenant à la solution. Puis à partir de cette information chercher les probabilités qui lui sont associées. Plus loin, on verra comment choisir ces $K$ valeurs.

Prenons un exemple graphique. La valeurs pouvant être prises sont représentées par des points sur l'axe des réels, orienté de la gauche vers la droite dans le sens croissant. En bleu sont représenté les valeurs choisies pour faire partie de la la solution $Y$. On pourra noter que dans cet exemple, la plus grande et la plus petite valeur de X font partie de la solution. Ce n'est pas un hasard, car toute distribution $Y$ telle que $X \preceq_{cx} Y$ doit contenir une valeur inférieure où égale à la valeur minimale de $X$ et une valeur supérieure ou égale à la valeur maximale de $Y$.

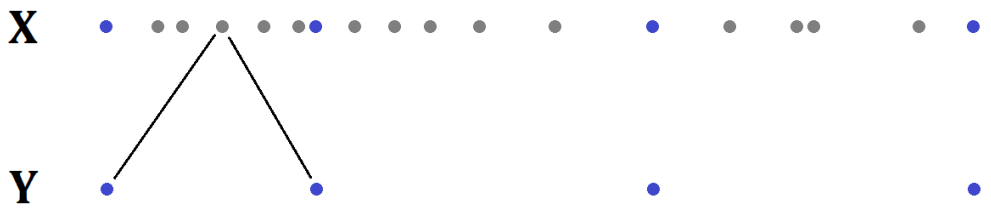


FIGURE 4.4 – Suppression optimale d'une valeur comprise entre deux autres.

Tout les points qui n'ont pas été choisit pour faire partir de la solution doivent être supprimés. Il va donc falloir distribuer leur probabilité sur les valeurs de $Y$. Et

l'on peut montrer qu'il est préférable de répartir cette probabilité sur les valeurs de $Y$ encadrant la valeur à supprimer. Comme sur le dessin, où une valeur de $X$ est reliées aux deux valeurs de $Y$ qui l'encadre.

Conclusion, pour $K$ valeurs données parmi les $N$ valeurs de $X$, la solution optimale est obtenue en répartissant les probabilités des valeurs supprimées sur les valeurs de la solution les encadrant.

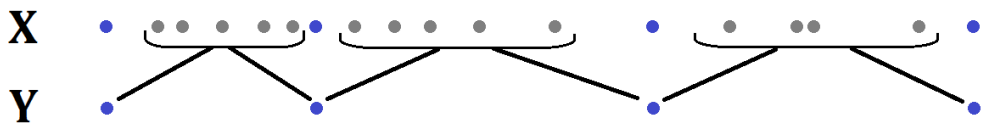


FIGURE 4.5 – Solution optimale étant donné les $K$ valeurs à conserver.

### 4.2.3 Deuxième démonstration d'optimalité

La démonstration brièvement décrite précédemment est le fruit de découverte successive. Le problème était que pour la rédiger formellement, il nous a fallut accumuler les lemmes donnant une preuve très technique et difficile à lire. De plus la notion de répartition d'une probabilité n'est pas formellement définie, il s'agit plus d'une intuition.

Ce qui m'a conduit à réfléchir à une autre preuve. Cette fois ci, en étudiant la fonction

$$f_Y : d \to E[(Y-d)^+] \leq E[(X-d)^+]$$

Pour $Y$ obtenu comme décrit précédemment (en répartissant les valeurs supprimées sur les valeurs voisines appartenant à la solution), on remarque que pour toutes les valeurs de $Y$ notées $y_1, \cdots, y_K$ on a

$$f_Y(y_i) = 0 \quad \forall i \in [|1, K|]$$

On peut montrer par un calcul, que si on prend $\tilde{Y}$ une distribution qui a les mêmes valeurs que $Y$ mais pas les mêmes probabilités. Alors dire que pour deux valeurs de $Y$ successives $y_i$ et $y_{i+1}$,

$$\left.\begin{array}{r} f_{\tilde{Y}}(y_i) \geq f_Y(y_i) \\ f_{\tilde{Y}}(y_{i+1}) \geq f_Y(y_{i+1}) \end{array}\right\} \Rightarrow \forall d \in [y_i, y_{i+1}], \quad f_{\tilde{Y}}(d) \geq f_Y(d)$$

Si $Y$ n'est pas optimale, alors il existe une distribution $\tilde{Y}$ telle que :
— $\tilde{Y}$ a les mêmes valeurs que $Y$ (mais des probabilités différentes)
— $X \preceq_{cx} \tilde{Y}$
— $Y \not\preceq_{cx} \tilde{Y}$

Comme $X \preceq_{cx} \tilde{Y}$, on en déduit que $\forall d,\ f_{\tilde{Y}}(d) \geq 0$ et donc entre autre $f_{\tilde{Y}}(y_i) \geq 0$. Comme on a dit précédemment que $f_Y(y_i) = 0$ on en déduit que $f_{\tilde{Y}}(y_i) \geq f_Y(y_i)$. On peut en déduire que $\forall d \in [y_1, y_K]\ f_{\tilde{Y}}(d) \geq f_Y(d)$. Pour $d \in ]-\infty, y_1] \cup [y_K, +\infty[$ , on peut montrer en posant le calcule que $f_{\tilde{Y}}(d) = f_Y(d) = 0$.

On vient de montrer que si il existe $\tilde{Y}$ tel que $X \preceq_{cx} \tilde{Y} \preceq_{cx} Y$. Alors, on a :

$$E[(\tilde{Y}-d)^+] - E[(X-d)^+] \geq E[(Y-d)^+] - E[(X-d)^+] \Rightarrow E[(\tilde{Y}-d)^+] \geq E[(Y-d)^+]$$
$$\Rightarrow Y \preceq_{cx} \tilde{Y}$$

Les deux première conditions faisant que $Y$ n'est pas optimale créent une contradiction avec la troisièmes. On en déduit donc par l'absurde que $Y$ est optimale.

### 4.2.4 Choisir les valeurs des la solution

Maintenant que l'on sait trouver les probabilités une fois les valeurs connues, il nous reste à choisir ces valeurs. Cette fois, on ne va pas pouvoir trouver un optimum sur la borne convexe, car toutes les distribution ne seront pas comparables.

Pour s'en convaincre, prenons une distribution $X$ quelconque ayant un grand nombre de valeurs. on va comparer deux distribution $Y$ et $Z$ ayant toutes deux 3 valeurs. Comme ce sont des bornes supérieures de $X$, on a $y_1 = x_1 = z_1$ et $y_3 = x_N = z_3$. Il nous reste à choisir $y_2$ et $z_2$ sachant que $y_2 \neq z_2$. Voici le tracé de $E[(Y-d)^+] - E[(Z-d)^+]$, pour $X \sim \mathcal{U}([|1,10|])$, $z_2 = 3$ et $y_2 = 7$.

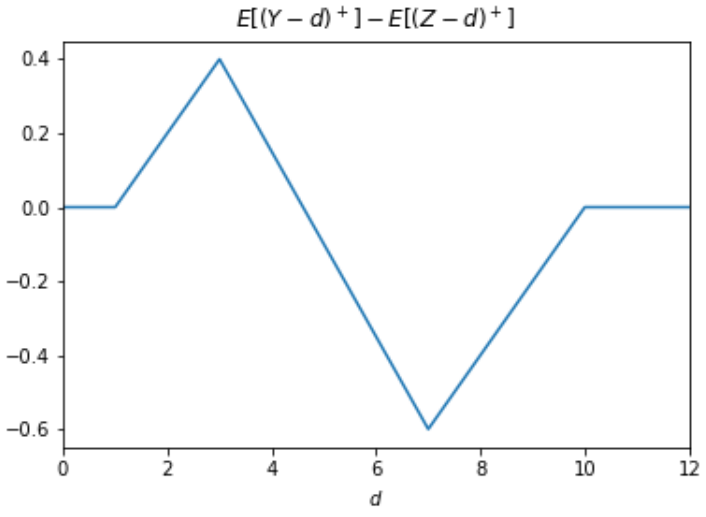


FIGURE 4.6 – Tracé de $E[(Y-d)^+] - E[(Z-d)^+]$

On peut remarquer qu'il y a une partie positive, et une partie négative. On ne peut donc pas comparer ces deux distribution suivant l'ordre convexe. Pour mieux comprendre, voici le tracé de $E[(Y-d)^+] - E[(X-d)^+]$ et $E[(Z-d)^+] - E[(X-d)^+]$. Le graphique ci-dessus ne représente rien d'autre que la différence de ces deux fonctions. On remarque que la fonction est nulle pour $d \leq x_1$ et $d \geq x_{10}$. Et que $E[(Y-d)^+] - E[(X-d)^+] = 0$ pour $d \in \{y_1, y_2, y_3\}$. De même $E[(Z-d)^+] - E[(X-d)^+] = 0$ pour $d \in \{z_1, z_2, z_3\}$.

En voyant cette figure, on comprend bien que si une valeur appartient à $Y$ mais pas à $Z$, l'une va être nulle alors que l'autre serra strictement positive. Du coup, si

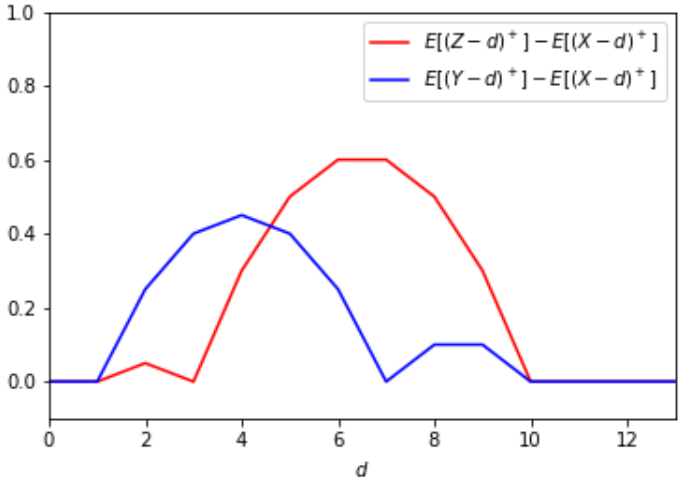


FIGURE 4.7 – Comparaison de $E[(Y-d)^+] - E[(X-d)^+]$ pour deux solutions optimales de 3 valeurs données dont une différente.

il est possible de construire des sous-distributions $Y$ et $Z$ optimale ayant des valeurs différentes, il serra exceptionnel (voire impossible) de pouvoir les comparer ces deux solution suivant la relation d'ordre convexe.

Il faut donc choisir une fonction $\phi$ convexe, qui nous permet de définir la fonction de coût suivante :

$$cout(Y) = E[\phi(Y)] - E[\phi(X)]$$

Il s'agit tout simplement de définir notre fonction objectif pour départager les solutions faisant parti ne notre front de Pareto.

Cette fonction de coût est additive. C'est à dire que

$$E[\phi(Y)] - E[\phi(X)] = \sum_{i=0}^{K-1} c_{i,i+1}$$

où $c_{i,i+1}$ correspond l'augmentation du coût due à la suppression des valeurs comprissent entre $y_i$ et $y_{i+1}$.

La solution proposée est donc de créer un graphe ayant $N$ sommets (autant de sommets que de valeurs dans $X$). Et les arêtes reliant le sommet $i$ et $j$ correspond à $c_{i,j}$, l'augmentation du coût liée à la suppression des valeurs comprises entre $x_i$ et $x_j$.

Imaginons que l'on veuille calculer le coût associé à la distribution conservant les valeurs $x_1, x_4, x_9, x_{10}$. Ce coût correspond au poids de chemin passant par les sommets $1, 4, 9, 10$. On a donc ramené notre problème à la recherche d'une plus court chemin de taille $K$. Pour cela, on peut adapter l'algorithme de Ford-Bellman [**?**, **?**] de la manière suivante : On note $OPT(i,k)$ le plus court chemin allant du sommet 1 au sommet $i$ utilisant **exactement** $k$ sommets. La relation de récurrence devient alors

$$OPT(i,k) = min_{j<i}\left(OPT(j,k-1) + c_{j,i}\right)$$

## 4.3 Recherche d'une borne inférieure optimale

### 4.3.1 Pourquoi cette recherche s'est avérée plus compliquée

La recherche de la borne inférieures s'est avérée être plus compliquée que la recherche de la borne supérieure. Pour comprendre la différence, repartons du Lemme qui a servit de point de départ à la recherche de la borne supérieure. Il s'agissait de prendre 3 valeurs et d'en supprimer une en répartissant sa probabilité sur les deux autres.

On ne peut évidement pas supprimer la valeur située au milieu, car c'est la transformation que nous avons largement utilisée dans la partie précédente. Cette transformation donne une borne supérieure et non une borne inférieure. On peut en revanche supprimer l'une des deux valeurs extrêmes.

Mais contrairement à la transformation précédente, cette manipulation n'est pas toujours possible. En effet, pour conserver l'égalité de l'espérance, il faut augmenter la probabilité de l'une des valeurs et diminuer la probabilité de l'autre valeur restante. Pour éclairer mon propos, voici une résumé graphique des 3 cas possibles de suppression d'une valeur parmi trois valeurs notées $a, b, c$ telles que $a < b < c$.

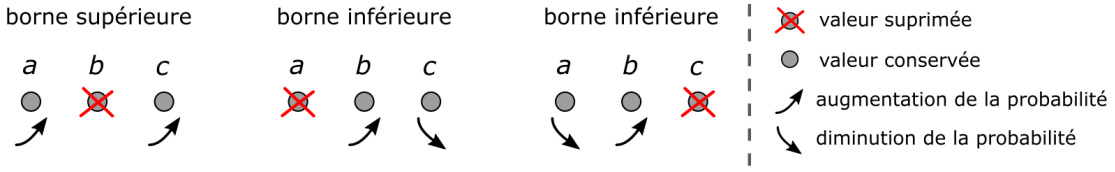


FIGURE 4.8 – Résumé graphique des propriétés liées à la suppression d'une valeur parmi trois

Comme en supprimant $a$ la valeur de la probabilité associée à $c$ diminue, il faut faire attention à ce que la probabilité de $c$ reste strictement positive malgré sa diminution. Par le calcul, on trouve que cette condition est vérifiée lorsque la valeur $M = \frac{ap_a+cp_c}{p_a+p_c}$ (un barycentre de $a$ et $c$ pondérés par leurs probabilités) est telle que $b < M$. Pour le cas de la suppression de $c$, la condition est $b > M$. Donc en fonction du signe de $M - b$, on peut supprimer $a$ ou $b$.

Conclusion on n'est pas sure qu'il serra possible de supprimer tous les atomes. Dans la recherche de la borne supérieure, on avait ajouté une valeur pour montrer qu'il valait mieux s'intéresser à des triplets de valeurs proche. Dans la cas de la borne inférieure, on se retrouve avec plusieurs configurations possibles donc certaines non comparables sur l'ordre convexe. Ce qui rend la tâche bien plus compliquée.

### 4.3.2 Une nouvelle approche

Comme supprimer des valeurs semble compliqué du fait de la possibilité de créer sur des probabilités négatives, une autre approche utilisée est la fusion de valeurs. L'idée est assez simple. Si on a un distribution prenant deux valeurs $a$ et $b$. Alors on peut proposer une nouvelle distribution comprenant une seule valeur $x = \frac{ap_a+bp_b}{p_a+p_b}$

associée à la probabilité $p_x = p_a + p_b$. On a bien conservation de l'espérance, et le calcul montre que cette nouvelle distribution est bien inférieure à la précédente suivant l'ordre convexe. Avec cette transformation, il n'y a plus de problème de probabilités négatives. En revanche, à chaque fois, on supprime deux valeurs et on en crée une nouvelle.

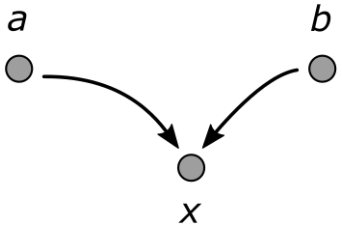


FIGURE 4.9 – Illustration d'une fusion de deux valeurs.

La première observation que l'on peut faire, c'est que le résultat final est indépendant de l'ordre dans lequel on a réalisé les fusions. Cela se comprend assez bien, car si on fusionne $k$ valeurs $x_1, \cdots, x_k$ en une unique valeur $y$, il faut conserver la somme des probabilités, ce qui implique que $p_y = \sum_i p_{x_i}$. De plus pour avoir un résultat qui soit une inférieur à la distribution d'origine suivant la borne convexe, il faut que l'espérance soit conservée. Ce qui implique que $yp_y = \sum_i x_i p_{x_i}$, laissant ainsi une seule valeur possible pour $y$.

On peut donc voire le problème d'une autre manière. Au lieu de voire la réduction du nombre de valeurs comme des fusions successives, on peut le foire comme un partitionnement de l'ensemble de nos valeurs de départ. Ainsi, si on veut passer de $N$ à $K$ valeurs, il va valoir attribuer à chaque une des $N$ valeurs une étiquette appartenant à l'ensemble $\{1, \cdots, K\}$. De manière à fusionner ensemble toutes les valeurs ayant la même étiquette.

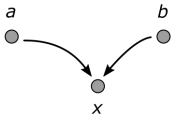


FIGURE 4.10 – Illustration d'un étiquetage d'une distribution et de son résultat.

A partir de là, il faut trouver des règles permettant d'éliminer le plus possible de partitionnement en montrant que certains sont supérieurs à d'autre suivant la borne convexe. La proposition qui a été faite, est la suivante, les sous ensembles contiennent uniquement des valeurs consécutifs. C'est à dire que sur l'exemple cidessous, la distribution à droite est meilleur que celle de gauche. Car à gauche, il y a une valeur étiqueté 2 au milieu des valeurs étiquetés 1.

Pour rappel, dire qu'une distribution $Y$ est meilleure qu'une distribution $Z$ revient à dire que $Z \preceq_{cx} Y$. Car ce sont toutes deux des distribution inférieures à la distribution d'origine suivant la borne convexe. Ce qui veut dire que $Z \preceq_{cx} Y \preceq_{cx} X$. On voit donc que $Y$ donne une meilleure borne inférieure que $Z$.

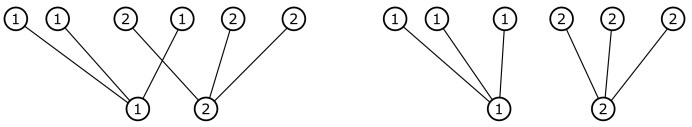

FIGURE 4.11 – Exemple d'un étiquetage ne respectant pas la règle proposé et d'un étiquetage la respectant.

Pour étudier ceci, l'idée est de montrer que si la règle n'est pas respectée, on peut proposer une meilleure solution. Si une distribution ne respecte pas la règle, c'est que l'on peut trouver deux étiquettes (ici les étiquettes 1 et 2) telles qu'en suivant l'ordre croissant, on va avoir au moins un passage de l'étiquette 1 à l'étiquette 2 et un passage de l'étiquette 2 à l'étiquette 1. Un exemple d'étiquetage ne respectant pas la règle est donné juste en dessous. Nous allons nous appuyer dessus pour expliquer comment exhiber une meilleure solution.

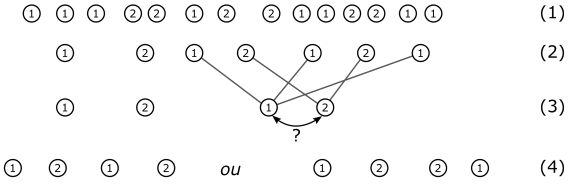


FIGURE 4.12 – Illustration des étapes permettant de réduire le problème à un problème à 4 valeurs.

Pour simplifier le problème, on réalise la fusion des valeurs consécutive ayant la même étiquette. On se retrouve alors des valeurs alternant entre 1 et 2. On peut donc avoir les combinaison $1, 2, 1$ que l'on traitera à part. puis les combinaison $1, 2, 1, 2$ puis $1, 2, 1, 2, 1$ puis $1, 2, 1, 2, 1, 2$ et ainsi de suite. Ce qui correspond à la ligne (2)

On va ramener le problème au cas à 4 valeurs. Pour cela, on conserve la première valeur étiqueté 1 et la première étiqueté2. Puis on fusionne toutes les autres valeurs étiquetées 1 ensemble, puis de même pour celles étiquetées 2. Ce qui correspond à la ligne (3) Seul problème, on ne peut pas savoir si la fusion de valeurs étiquetés 2 donnera un valeur plus petite ou plus grande que celle donnée par la fusion des valeurs étiquetés 1. On se retrouve donc avec deux cas correspondant à la ligne (4).

Le deuxième cas est le plus simple, car on peut montrer par des raisonnements géométrique (que je n'explique pas ici pour me focaliser sur le raisonnement général) qu'un étiquetage $1, 2, 2, 1$ est soit inférieur suivant l'ordre convexe à l'étiquetage $1, 2, 2, 2$ soit inférieur à $1, 1, 1, 2$. Supposons que dans notre cas ce soit l'étiquetage $1, 2, 2, 2$ qui est supérieur à l'étiquetage $1, 2, 2, 1$ (le raisonnement est exactement le même quand c'est $1, 1, 1, 2$ qui est supérieur à $1, 2, 2, 1$).

On peut remonter les étapes qui nous avaient permis de réduire notre problème à seulement 4 valeurs, et ainsi obtenir un meilleur étiquetage. Ci dessous se trouve l'illustration de la méthode utilisée pour expliciter une meilleure solution. A la ligne (1) se trouve la solution une fois que toutes les fusion on étaient faites. Avec à gauche le cas ne respectant pas la règle, et à droite un cas supérieur à celui de gauche. On

peut constater que les fusions à droites et à gauches entre la ligne (2) et la ligne (4) sont les mêmes. On a donc exactement les même valeurs et probabilités sur ces lignes. La seule chose qui est modifiée, c'est l'étiquetage de ces valeurs.

On peut voire le dessin si dessous comme les étapes de la création d'une meilleure solution. On par de la proposition faite en ligne (4) à gauche. On réalise le simplifications menant à la ligne (2) à gauche. A partir de là on propose une meilleur solution pour cette ligne que l'on place à droite. Puis on modifie l'étiquetage de la colonne de droite pour que l'étiquetage situé sur la ligne (4) à droite corresponde bien à celui proposé à la ligne (2).

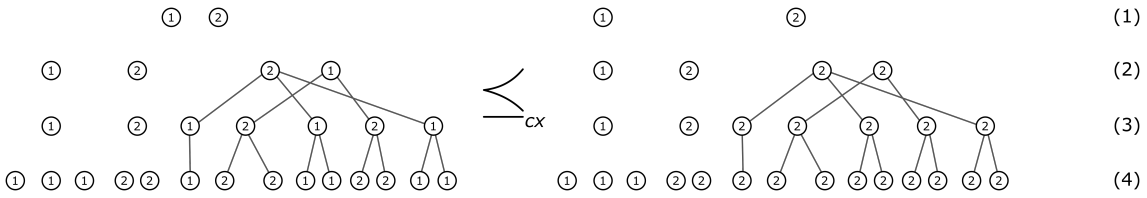


FIGURE 4.13 – Étapes de la création d'une meilleure solution.

Il faut maintenant traiter l'autre cas. Celui où l'étiquetage final de notre simplification est $1, 2, 1, 2$. Pour beaucoup de distribution, l'étiquetage $1, 1, 2, 2$ est supérieur. Le problème, c'est que l'on peut trouver certaines distributions avec de forts écarts entres leurs différentes probabilités telle que $1, 1, 2, 2$ ne soit pas comparable avec $1, 2, 1, 2$. Par contre il est strictement impossible que $1, 2, 1, 2$ soit supérieur à $1, 1, 2, 2$.

Mais ce ne sont pas les seuls étiquetages possibles. En effet, on peut aussi avoir l'étiquetage $1, 1, 1, 2$ et $1, 2, 2, 2$. Mais là encore, on peut construire des distributions telles que le résultat de leur étiquetage $1, 2, 1, 2$ ne soit pas comparable suivant l'ordre convexe avec le résultat des étiquetage $1, 1, 2, 2$ ou $1, 1, 1, 2$ ou $1, 2, 2, 2$.

La question qui reste ouverte, est la suivante. Existe-t-il une fonction convexe $\phi$ telle que $E[\phi(Bad)] \geq E[\phi(Good_i)]$ pour $i \in \{1, 2, 3\}$. Où $Bad$ est la distribution obtenue par l'étiquetage $1, 2, 1, 2$. Et $Good_1, Good_2, Good_3$ les distributions obtenues par les étiquetages $1, 1, 2, 2$ ; $1, 1, 1, 2$ et $1, 2, 2, 2$.

J'ai cherché par ordinateur une telle fonction $\phi$ sans résultats. J'ai l'impression que

$$E[\phi(Bad)] \geq E[\phi(Good_1)] \Rightarrow E[\phi(Bad)] \leq E[\phi(Good_2)]$$

Mais je n'ai pas réussi à prouver ce résultat.

Si on arrive à le montrer, on pourra dire que pour toute fonction $\phi$ il existe un étiquetage parmi $1, 1, 2, 2$ ; $1, 1, 1, 2$ et $1, 2, 2, 2$ qui sera meilleur que l'étiquetage $1, 2, 1, 2$. On peut donc trouver une meilleur solution en utilisant le même résultat que pour l'étiquetage $1, 2, 2, 1$.

### 4.3.3 Adapter l'algorithme

Si le résultat précédent venait à être confirmé, il serrait alors possible d'adapter l'algorithme de la borne supérieure à celui de la borne inférieur. Cette fois si, un

arrête allant du sommet $i$ vers le sommet $j$ correspondrait à la fusion de toutes les valeurs comprissent entre entre $x_i$ inclus et $x_j$ exclus. Il faut ajouter un sommet à notre graphe, le sommet $N+1$ pour pouvoir inclure le sommet $x_N$ dans une des fusion. Et bien entendu modifier les poids des arêtes pour qu'elles correspondent à la variation de $E[\phi(X)]$ lors de la fusion des valeurs.

Ci dessous, se trouve illustré une relation entre le plus court chemin dans notre graphe, l'étiquetage des valeurs et la sous distribution associée.

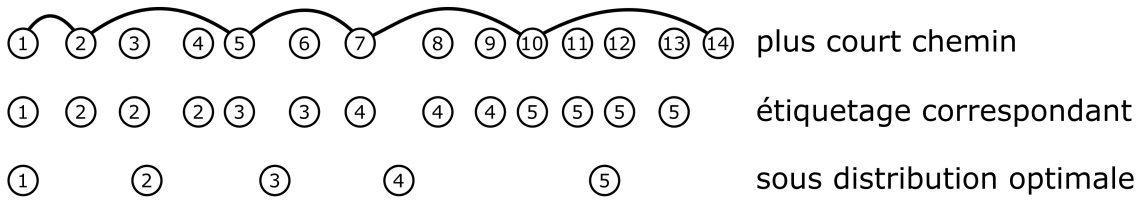


FIGURE 4.14 – Illustration sur un exemple des trois représentations différentes décrivant la même solution

# 5 Analyse des résultats obtenus

Les résultat majeur de ce stage est l'algorithme permettant de trouver une borne supérieur suivant l'ordre convexe, et l'accélération du calcul de la matrice d'adjacence du graphe utilisé pour cela. Permettant de passer d'une complexité en $O(N^3)$ pour la méthode naïve à une complexité en $O(N^2)$ pour le calcul du poids des arêtes. Ce qui fait que ce n'est plus le calcul du poids des arêtes qui domine la complexité, mais bien l'algorithme de Ford-Belman adapté aux chemins de taille fixé. Qui est en $O(KN^2)$.

Une limitation de ce résultat est l'ensemble des valeurs pouvant êtres prisent par la borne supérieure. Car on s'est limité à des bornes supérieures prenant leurs valeurs dans celles de la distribution d'origine. On pourrait peut être levé cette limitation en ajoutant des valeurs de probabilité nulles. Ce qui ne modifie pas la distribution, mais augment le nombre de valeurs pouvant être prise par la borne supérieure.

La démonstration de la borne inférieure résiste encore. Mais l'absence de contre exemple lors de tests sur ordinateur est encourageant pour continuer dans cette direction.

Il reste à vérifier que le choix de cet relation d'ordre est plus pertinent que celui d'une autre relation d'ordre. Comme la relation strong étudiée qui avait fait l'objet d'une étude similaire par la même équipe de chercheurs.

# 6 Apports

## 6.1 Apports théorique

Ce stage m'a permis d'élargir mes connaissances scientifiques. Les notions d'ordre stochastiques ainsi que de relation d'ordre partiels m'étaient jusque là inconnues.

En plus du sujet lui même, j'ai appris pas mal de chose en cherchant à résoudre ce problème. J'ai par exemple approfondi mes connaissance sur le Lagrangien et la programmation linéaire.

C'est aussi l'occasion de voire comment acquérir de nouvelles connaissances. Car il a fallut rapidement comprendre les propriétés des bornes convexes pour pouvoir commencer à travailler. Chose que je n'avais encore jamais fait au paravent, c'est de vérifier lemme après lemme les propriétés que Jean-Michel Fourneau nous avait proposait.

## 6.2 Apports pratique

J'ai aussi eu la possibilité de découvrir de nouveaux outils. Avec entre autre la bibliothèque python sympy [**?**] qui permet de faire du calcul symbolique sur python, et qui m'a épargné pas mal de temps et d'erreur de calcul. L'écriture de preuve et d'algorithme en latex. En apprenant notamment à bien poser la définition des lemmes, ainsi que quelques astuces de mise en page. Par exemple les balises
proof, les références à d'autres parties du document, ainsi que les citations bibliographiques.

## 6.3 Apports professionnel

En dehors des apports pédagogiques classiques, qui pourraient faire l'objet de cours, ou de formations, ce stage à aussi été l'occasion de me familiariser avec le milieu de la recherche, en discutant avec les doctorants de leurs travaux. En discutant avec les chercheurs de leurs problèmes de tout les jours, ainsi que de comment bien préparer sa thèse.

# 7 Conclusions et recommandations

Lors de ce stage j'ai étudié un problème et proposé un algorithme de résolution. J'ai cherché à la fois à optimiser la complexité de l'algorithme, et à rendre sa démonstration la plus naturelle possible.

Ce stage a été très formateur, il m'a permis de gagner en autonomie. J'y ai aussi acquis des connaissances en statistique et en optimisation.

J'ai également pu travailler dans un environnement stimulant et découvrir le domaine de la recherche, ce qui m'a permis de confirmer mon projet professionnel et mon choix de suivre en plus du cursus Supélec un Master orienté recherche avec l'ENS Paris-Saclay.

# Bibliographie


[1] Richard Bellman. On a routing problem. *Quarterly of applied mathematics*, 16(1) :87–90, 1958.

[2] Lester R Ford Jr. Network flow theory. Technical report, RAND CORP. SANTA MONICA CA, 1956.

[3] Jean-Michel Fourneau, Youssef Ait El Mahjoub, Franck Quessette, and Dimitris Vekris. Xborne 2016 : A brief introduction. In Tadeusz Czachórski, Erol Gelenbe, Krzysztof Grochla, and Ricardo Lent, editors, *Computer and Information Sciences - 31st International Symposium, ISCIS 2016, Kraków, Poland, October 27-28, 2016, Proceedings*, volume 659 of *Communications in Computer and Information Science*, pages 134–141. Springer, 2016.

[4] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. Sympy : symbolic computing in python. *PeerJ Computer Science*, 3 :e103, January 2017.

[5] A. Muller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley, New York, NY, 2002.

[6] M. Shaked and J. G. Shantikumar. *Stochastic Orders and their Applications*. Academic Press, San Diego, CA, 1994.