

1 Présentation du service étudié

1.1 Principe général de Communauto

Communauto est un service basé dans différentes villes, principalement au Canada mais également à Paris, qui propose de la location de voitures sous deux formules :

free-floating : les utilisateurs peuvent utiliser une voiture garée dans l'espace public afin d'effectuer un trajet d'un point A à un point B, tout en restant dans la zone de service de Communauto. Cette formule est principalement destinée aux trajets quotidiens, par exemple entre le domicile et le lieu de travail.

basé-station : les utilisateurs peuvent utiliser une voiture garée dans l'une des nombreuses stations réparties dans la ville afin d'effectuer un trajet puis de ramener la voiture à la station de départ. Cette formule est plus destinée aux plus long trajets, planifiés à l'avance, par exemple dans le cadre de voyages sur le week-end.

Ces deux systèmes partagent un même abonnement ainsi qu'un tarif identique après 4 heures d'utilisation, ce qui permet des interactions entre ces deux systèmes, par le report d'utilisateur d'un système à l'autre en fonction de la disponibilité.

1.2 Fonctionnement du système et lexique utilisé

Les utilisateurs ont accès à une application sur lesquelles ils peuvent réserver une voiture. Le moment où un usager réserve une voiture est appelé *temps d'inscription*.

Pour les deux systèmes, on peut séparer le temps entre l'inscription d'une réservation et sa fin en deux : Le temps s'écoulant entre l'inscription et le début de la réservation est appelé *temps de pré-réservation*, le *temps de réservation* (et non temps de trajet) correspond lui à la durée de réservation elle-même.

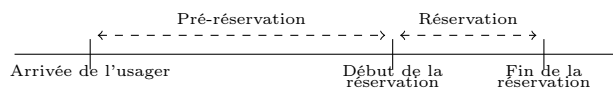


FIGURE 1 – Schéma d'une réservation

1.2.1 Système free-floating

Dans ce système, les usagers peuvent sélectionner une voiture disponible et la réserver. Une voiture réservée l'est alors pendant 30 minutes, temps pendant lequel le client peut se rendre à la voiture et la prendre. La réservation est automatiquement annulée au bout de 30 minutes si le client ne s'est pas présenté. Une fois la voiture récupérée, l'utilisateur peut la garder aussi longtemps qu'il le souhaite avant de la libérer.

Le temps de pré-réservation est ainsi inférieur à 30 minutes.

1.3 Système basé-stations

Dans ce système, les usagers se connectant à un certain instant peuvent sélectionner une voiture assignée à une station et réserver un créneau futur s'il est disponible. La réservation peut s'effectuer jusqu'à un mois à l'avance.

Il est également possible pour un usager d'annuler une réservation effectuée ou encore de rendre la voiture avant la fin annoncée de la réservation. Dans ce dernier cas, on parlera de *fin de réservation anticipée* par opposition aux *réservations menées à leur terme*.

Le temps de pré-réservation peut ainsi monter à 1 mois tandis que la durée de réservation, à la différence du free-floating, est défini dès le début même s'il peut être modifié par la suite.

2 Analyse de données

2.1 Présentation des données

Grâce à un partenariat entre Communauto et Polytechnique Montréal, j'ai eu accès aux données correspondant aux réservations de voitures effectuées dans la ville de Montréal. J'ai ainsi pu étudier les données de l'année 2021, comprenant aussi bien les réservations effectuées sur les voitures en free-floating que sur celles en basés-stations, cela correspondant à respectivement à 69% et 31% des réservations effectués.

Le système est basé sur plus de 2000 voitures, dont environ 40% sont en free-floating tandis que 60% sont des voitures basées-station réparties en près de 500 stations réparties dans la ville.

Je disposais pour ces données d'un grand nombre de champs dont ceux que j'ai principalement utilisés sont les suivant :

- la date et l'heure de début et de fin de la réservation
- la distance effectuée durant la réservation

- les points de départ et d’arrivée de la réservation. Cela correspond à la localisation de la station pour les basées-station
- l’adresse approximative des usagers
- lorsque la réservation est annulée, la date d’annulation
- le statut de la réservation, parmi active, finie prématurément, annulée ou modifiée entre autres

On retrouve dans les données principalement des réservations actives (menées à leur terme), annulées ou finies prématurément ainsi qu’un très petit nombre de réservations au statut spécial car effectuées par Communauto pour maintenance. Leur répartition est présentée figure 2.

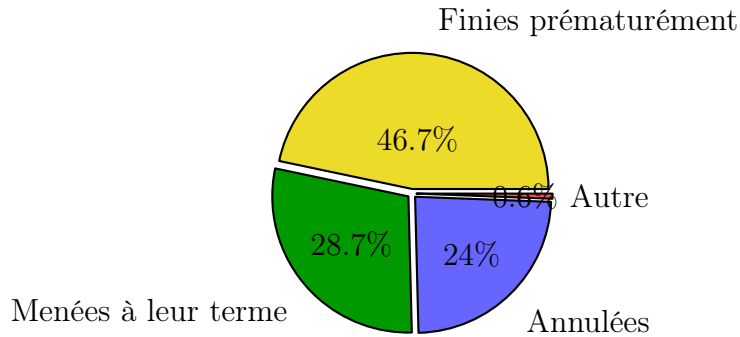


FIGURE 2 – Répartition des statuts des réservations basées-stations en 2021

2.2 Nettoyage préalable des données

Pour étudier les données et tenter de déterminer les lois de probabilités par lesquelles on peut approximer les différentes distribution, il faut commencer par nettoyer et filtrer les données :

Dans un premier temps, il s’agit ainsi d’isoler les réservations réelles, c’est à dire les réservations effectuées par les usagers e, et d’éliminer les autres, par exemple les réservations pour maintenance ou pour changer la voiture de station de rattachement, ainsi que des restes de réservations annulées, qu’en peut identifier par une distance parcourue nulle et un tarif appliqué nul .

Suivant le paramètre étudié, j’ai pu avoir à filtrer selon le statut de la réservation, afin d’éliminer les trajets annulés pour mesurer le temps de réservation effectif par exemple.

2.3 Fluctuations au cours du temps de la capacité utilisée

Je me suis intéressé au comportement des usagers, pour voir notamment quelles périodes sont les plus prisées par les usagers.

La figure 3 présente le nombre de voitures réservées tout au long de l'année afin d'avoir un premier aperçu des périodes pendant lesquelles les voitures basées-stations sont les plus utilisées.

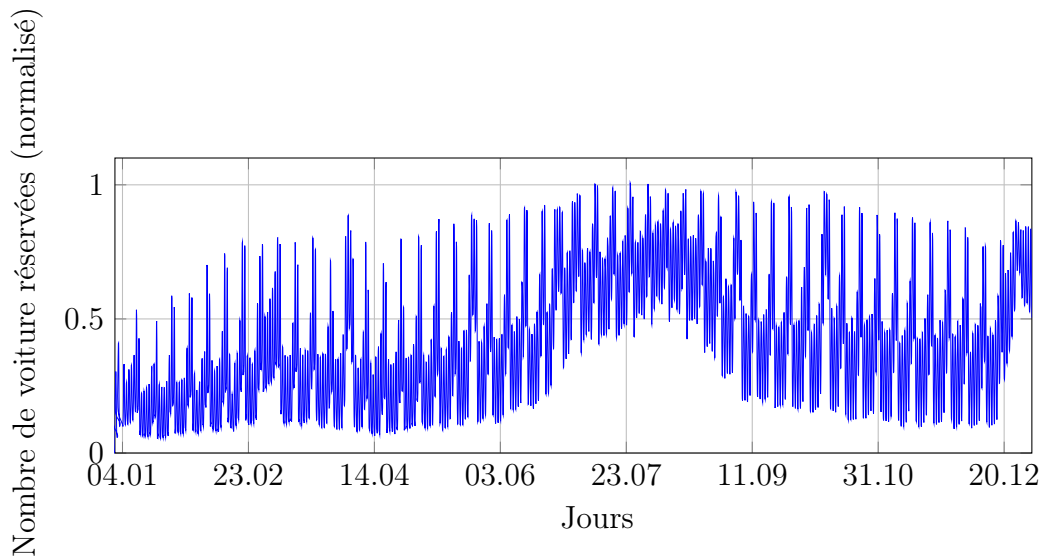


FIGURE 3 – Occupation du système au cours de l'année 2021

Ce graphe permet d'identifier les périodes les plus importantes, qui sont par ordre décroissant les vacances d'été et de Noël, où le minimum de réservation à un instant donné est particulièrement élevé, les autres vacances et week-ends prolongés puis les week-ends.

Ce graphe ne prend pas en compte les réservations ayant commencé avant 2021, ce qui explique le faible nombre de réservations début 2021 par rapport à 2021 : il y a en réalité un grand nombre de longues réservations commençant fin 2020 qui n'apparaissent pas sur ces données.

La figure 4 présente la capacité utilisée typique durant une semaine, obtenue en moyennant chaque jour de la semaine sur les 53 semaines de l'année 2021 et qui permet de voir plus finement ce qui se passe au cours d'une semaine.

On y observe ainsi de grandes fluctuations de l'occupation au cours de la semaine, principalement à l'échelle de la journée avec des pics chaque jour

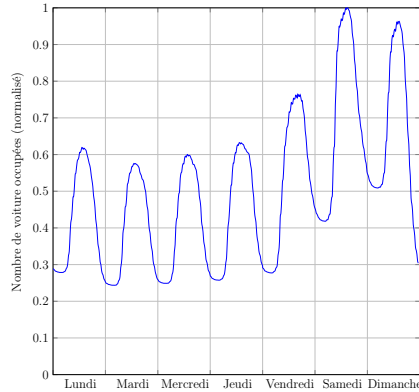


FIGURE 4 – Occupation du système au cours d'une semaine type

aux alentours de 13h et des périodes de creux durant les nuits. Cette unicité du pic est intéressante car c'est un comportement différent des voitures de free-floating, présentant un pic d'utilisation en matinée et en soirée (figure à tracer),

Par ailleurs un effet dû aux weekends est très clairement présent, tout d'abord avec des pics en journée plus hauts que durant le reste de la semaine mais aussi avec des creux plus élevés, représentant les usagers qui réservent une voiture afin de partir en week-end avec. Par ailleurs, on remarque que le vendredi fait en quelque sorte la transition entre la semaine et le week-end, avec un pic intermédiaire en milieu de journée et un creux de nuit entre vendredi et samedi plus chargée que les nuits de semaine, ce qui montre que les départs en weekends peuvent commencer dès le vendredi.

2.4 Étude du temps de réservation des usagers

2.4.1 Premières analyses

On mesure un temps de réservation moyen de 7.8h alors que le maximum de la distribution est situé à 2h et le temps médian est lui de 3.5. La moyenne du temps de réservation est en effet fortement augmentée par la présence de très longues réservations, 1% des réservations durant notamment plus de 2 semaines, soit plus de 300h.

De plus le kurtosis de cette distribution vaut 160, ce qui indique que l'on a affaire à une distribution à queue lourde, c'est à dire qu'elle possède une queue plus épaisse qu'une loi normale et donc que les événements extrêmes y sont plus probables.

2.4.2 Détermination et lissage de la distribution

Afin de déterminer des informations plus fines, il faut étudier l'allure de la distribution : Pour cela, dans le cadre du temps de réservation des usagers, le champ *res_trip_length_hr* présent dans les données correspond précisément au temps de réservation de chaque élément.

Il ne comprend supposément que des temps définies au quart d'heure près, les réservations pour les voitures basés-station se faisant sur des créneaux de la taille d'un quart d'heure. Cependant, on retrouve parmi les données certaines valeurs aberrantes, telles des réservations durant 1 ou 40 minutes par exemple. Cela correspondant à seulement 13 réservations, j'ai décidé de les enlever de l'étude du temps de réservation.

On peut alors obtenir la répartition empirique des temps de réservation (courbe bleue sur la figure 5) qui reste néanmoins très irrégulière ; on peut en effet observer une certaine périodicité d'ordre 4, certainement due aux utilisateurs, qui auront une plus grande tendance à réserver une durée entière en heures. Pour la même raison, on observe un pic de réservations d'une durée de 10 heures ainsi que des pics successifs présents pour tout les jours entiers.

Afin de pouvoir exploiter la courbe, il est nécessaire de lisser la courbe, par exemple avec un estimateur à noyau de fenêtre assez large pour enlever ces périodicités. J'ai préféré un estimateur à taille de noyau variable afin de ne pas trop écraser les données présentes à l'origine et ainsi de garder l'allure générale de la courbe, ce qui est réalisé sur la courbe rouge en figure 5.

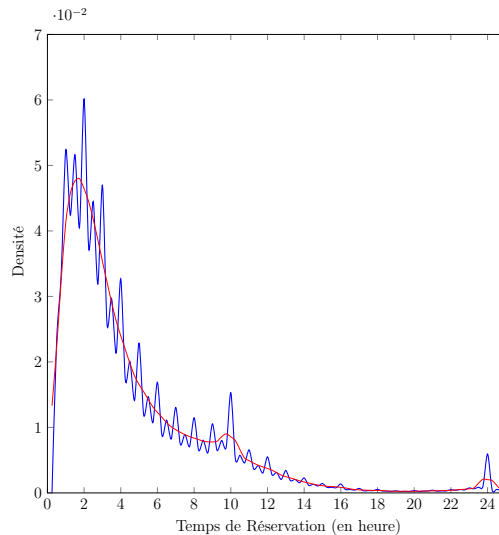


FIGURE 5 – Densité du Temps de réservation lissée

2.4.3 Détermination de la loi la mieux adaptée aux données

A partir de cette courbe lissée, j'ai cherché à déterminer par quelle loi l'approcher pour correspondre le mieux aux données et ainsi adapter le mieux le modèle. Pour cela, je me suis concentré sur des lois à queues lourdes, à savoir notamment loi gamma, loi log-normale, loi de Weibull ou encore loi de Burr.

Je les aies comparées à la distribution empirique du temps de réservation en utilisant la distance en variation totale :

Definition 1. Distance en variation totale

Soient μ, ν deux mesures de probabilité de densités respectives f et g .

On note $d_{VT}(\mu, \nu) = \int_{\mathbb{R}} |f(x) - g(x)| dx$ la distance en variation totale entre μ et ν .

Après optimisation des différents paramètres pour les différentes lois, j'ai obtenu le tableau 6 présentant la distance minimale entre chaque loi et la distribution du temps de réservation.

Loi approximante	Distance
Loi gamma $\Gamma(0.63, 70.82)$	0.23
Loi Log-normale $\log - \mathcal{N}(2.62, 0.9)$	0.17
Loi de Weibull ($k = 1.13, \lambda = 16, 7$)	0.24

FIGURE 6 – Distance entre différentes lois et la distribution du temps de réservation

Grâce à cela, il s'avère que c'est une loi log-normale de paramètres $\mu = 2.6$ et $\sigma^2 = 0.9$ qui permet la meilleure approximation du temps de réservation.

Le choix d'une loi log-normale est confirmé par l'étude des graphes quantile-quantile des différentes distributions théoriques présents en figure 2.4.3 seule la loi log-normale sélectionnée permet une bonne approximation de la queue de distribution.

Definition 2. Loi log-normale

Une variable aléatoire X suit une loi log normale de paramètres μ et σ^2 lorsque $Y = \ln(X)$ suit une loi normale de paramètres μ et σ^2 .

2.5 Étude des stations

2.5.1 Capacité des stations

Dans le cadre de cette étude, il est intéressant de regarder la capacité des stations dans lesquelles sont stationnées les voitures basées-stations ainsi que

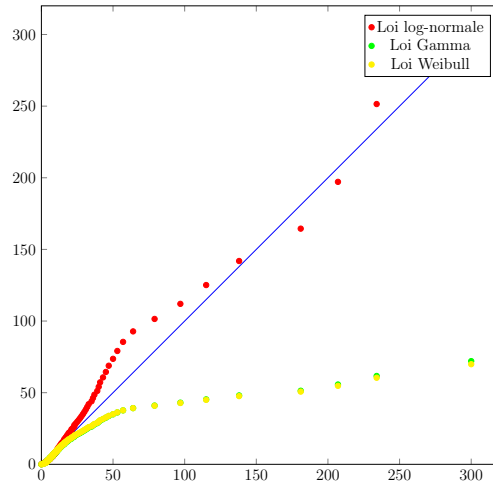


FIGURE 7 – Q-QPlot

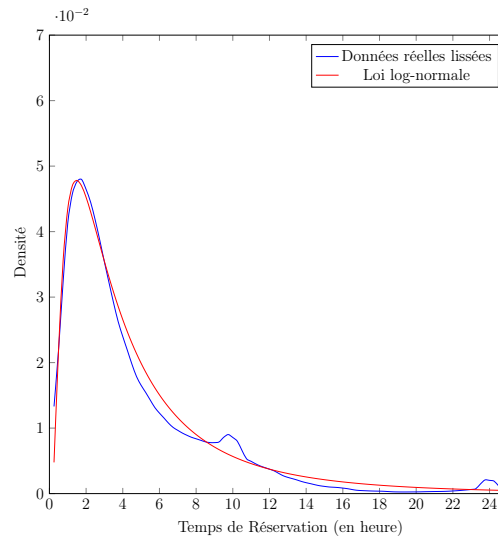


FIGURE 8 – Ajustement du temps de réservation par loi log-normale

leur répartition dans la ville.

Pour cela, j'ai regardé les voitures en activité et leur station lorsque le maximum de voitures étaient utilisées, à savoir le 24 juillet à 15h30, avec un pics à plus de 1300 voitures réservées en même temps. J'ai procédé ainsi afin de ne pas avoir des doublons dans les données, dues aux éventuelles voitures changeant de stations au gré des ajustements réalisées par Communauto.

Les graphes suivants se basent donc sur l'hypothèse que toutes les voitures étaient réservées à ce moment là, ce qui est assez proche de la réalité

car trouver une voiture le week-end, et à fortiori durant la période estivale, est très difficile. Les graphes permettent d'avoir un aperçu du système au maximum de sa capacité.

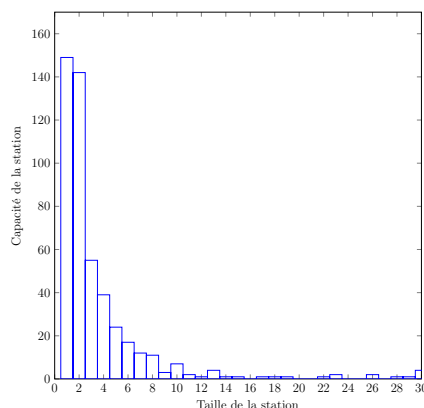


FIGURE 9 – Distribution du nombre de voitures par station

En traçant l'histogramme de la capacité(Figure 9), on observe une majorité de très petites stations : 66% des stations ayant 1 ou 2 voitures seulement. C'est explicable par la politique de Communauto d'avoir des station au plus près des usagers, multipliant ainsi la quantité de stations déployées, ainsi que par la difficulté d'obtenir des zones afin d'y réaliser de grandes stations. Il y a cependant quelques grandes stations, 5% des stations ayant plus de 10 voitures et elles représentent ainsi plus du quart des voitures basées-stations en circulation.

2.5.2 Localisation des stations

La figure(à ajouter) présent toutes les stations de communauto présentes à Montréal.

2.6 Étude de la distance de trajet

Afin d'étudier la loi de la distance parcourue, il faut dans un premier temps retirer des données les réservations ayant une distance parcourue nulle, correspondant la majorité du temps à des réservations "à vide", sans bouger la voiture. (Cela pourrait également correspondre à de très petits trajets, la distance relevée étant arrondie au kilomètre près, mais cela ne concerne qu'un faible nombre de trajets). Ces réservations correspondent à 10% du total des réservations

L'étude de la distance est après cela tout à fait similaire à celle de du temps de trajet et la distribution qui l'approche le mieux est pareillement une loi log-normale, dont les paramètres sont cette fois $\mu = 3.08$, $\sigma^2 = 1.25$. La Figure 10 présente la loi expérimentale et la loi approchée en superposition.

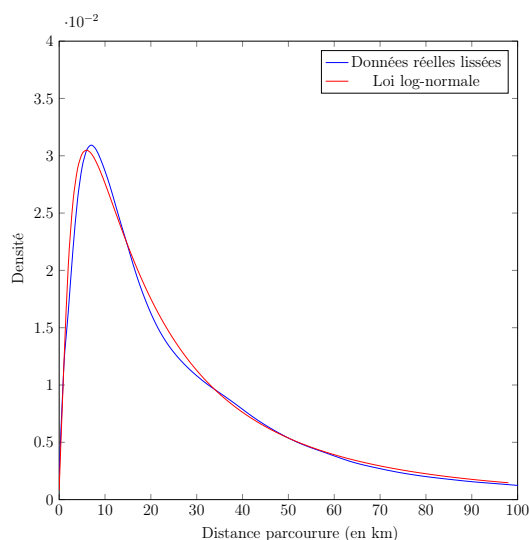


FIGURE 10 – Ajustement de la distance parcourue par une loi log-normale

2.7 Corrélation entre le temps de réservation et la distance parcourue pour obtenir une voiture

Il est que les usagers sont souvent près à traverser la ville afin d'obtenir une voiture. J'ai donc voulu vérifier cela en traçant la distribution de la distance entre la station utilisée et le domicile des usagers pour différents taux d'occupation du système.

2.8 Étude du temps d'annulation

Une autre donnée intéressante concerne l'annulation des réservations. En effet, comme montré par la figure 2 cela correspond à une grande part des données : 24% des réservations initiales étant annulées.

Il est alors intéressant de regarder quand cette annulation a lieu dans le cycle de vie de la réservation. Pour cela, on peut s'intéresser à la loi du temps s'écoulant entre l'annulation de la réservation et son début théorique. La loi empirique est tracée figure 11

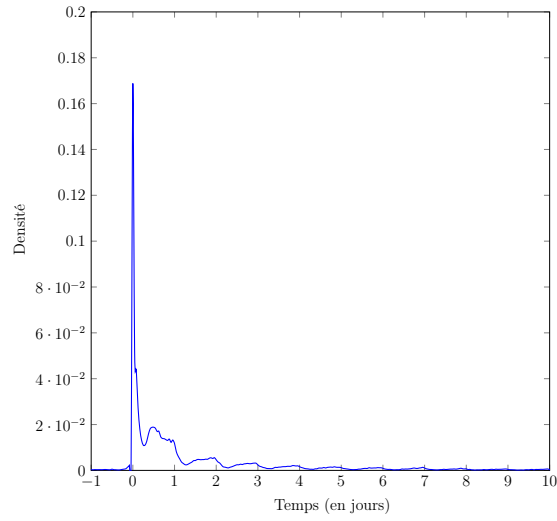


FIGURE 11 – Densité du temps entre annulation et début de la réservation

La première chose que l'on remarque est la présence d'un petit nombre de réservations annulées après leur temps de début de réservation, correspondant à environ 0.5% des trajets annulés. Ceci n'est pas une erreur des données, la réservation étant payée par les usagers pour le temps écoulé.

Ensuite la majorité des annulations ont lieu juste avant le début du trajet, comme montré par le pic observé sur les données en 0. Ainsi, les réservations annulées le sont en moyenne moins de 3 jours avant le début de la réservation : 62% étant annulées moins de 24h à l'avance.

Enfin, en regardant au temps long sur la figure 12, un léger pic apparaît à 31 jours, représentant les clients annulant leur réservations juste après l'avoir réservée.

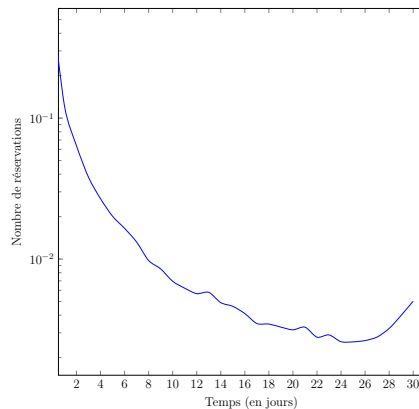


FIGURE 12 – Densité du temps entre annulation et début de la réservation

2.9 Étude du temps de pré-réservation

Les données disposent d'un champ *date d'inscription* correspondant en théorie à la date à laquelle une réservation a été inscrite dans le système, c'est à dire la date à laquelle l'utilisateur a réservé un créneau. Cependant, cette date ne correspond généralement pas à la date d'inscription réelle mais à la date de dernière action sur la réservation, que ce soit son annulation, sa modification ou bien un retour anticipé de la voiture. Ainsi les données présentant leur date d'inscription initiale sont très partielles et difficiles à identifier.

3 Simulations numériques

Afin de modéliser l'évolution

Dans le cadre du modèle utilisé, on nécessite 5 lois afin de simuler le système :

- la loi du temps d'arrivée des usagers.
- la loi du temps de pré-réservation
- la loi du temps de réservation
- la loi du temps d'annulation de la réservation (le temps entre l'inscription et l'annulation de la réservation)
- la loi du temps de retour anticipé du véhicule

Le point crucial consiste à déterminer sous quelles conditions une réservation demandée par un usager est acceptée ou non. Le critère simple est qu'une réservation est acceptée si, sur l'intervalle voulu par l'utilisateur, la capacité résiduelle due aux réservations précédente est toujours strictement positive.

En effet, étant donné qu'on considère un système où le client ne réserve pas une voiture spécifique mais une voiture d'une station, la condition assure qu'au moins une voiture ne sera pas encore utilisée au moment du début de la réservation de chaque client, voiture qui pourra ainsi être prise par le client.

3.1 Le modèle du "Tetris"

Afin de visualiser cette condition d'acceptabilité des réservations, on peut représenter les réservations successives comme des pièces d'une certaine longueur proportionnelle à la longueur du créneau réservé. On peut alors représenter les réservations successives comme différentes pièces tombant dans une grille de tetris infinie, l'axe des abscisses représentant l'axe temporel. Chaque réservation descend alors sur l'intervalle que le client souhaite réserver et la réservation est acceptée uniquement si ce bloc peut se retrouver en totalité en dessous de la capacité du système.

Cela est schématisé par la figure 13, présentant une capacité maximale de 4.

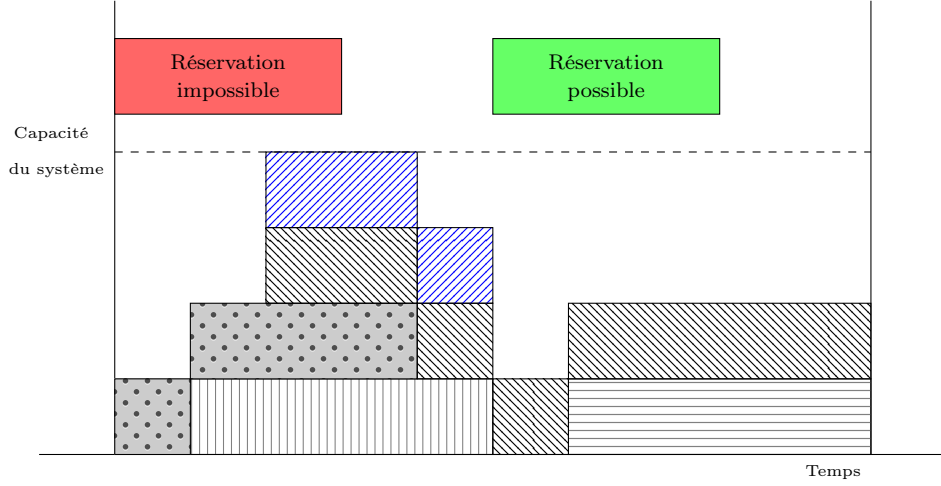


FIGURE 13 – Schéma représentant la condition d'acceptabilité d'une réservation

3.2 Modélisation mathématique

On considère pour l'instant un modèle simplifié où les usagers réservent à l'avance un créneau dans le futur et ne peuvent ni l'annuler ni le finir prématurément. Par ailleurs, comme dans les simulations, on considère que les usagers réservent un créneau et pas directement une voiture.

Definition 3. Représentation des réservations

Les tentatives de réservations peuvent être représentée par un processus de poisson marqué sur $\mathbb{R}_+^3 (t_n, p_n, \sigma_n)$ avec :

- t_n est le moment de l'inscription de la réservation, suivant un processus de point de Poisson, de paramètre λ
- p_n est la durée de pré-réservation de la réservation. Les p_n sont i.i.d
- σ_n est la durée de réservation de la réservation

La suite (p_n, σ_n) est i.i.d et (t_n) et (p_n, σ_n) sont indépendantes. Notons (p, σ) un couple de variables aléatoires ayant même loi que les (p_n, σ_n)

On ajoute à ce triplet un 4-ème nombre, ϵ_n , qui va correspondre à la validation de la réservation et vaut 1 si cette réservation ne rentre pas en conflit avec les réservations précédentes au moment de l'inscription.

Definition 4. Capacité réservée

La capacité réservée à un instant donné correspond au nombre de véhicules réservés à cet instant, en prenant en compte l'ensemble des réservations validées.

Elle vaut

$$C(t) = \sum_{n=0}^{\infty} \epsilon_n \mathbb{1}_{\{t_n + p_n < t < t_n + p_n + \sigma_n\}} \quad (1)$$

Definition 5. Capacité prévue

La capacité prévue en t depuis l'instant t^* correspond à la prévision de la capacité en t_1 à partir des informations disponibles en t_0 , c'est à dire sans tenir compte des inscriptions ultérieures à t^* .

Elle vaut alors

$$C(t|t^*) = \sum_{n=0}^{\infty} \epsilon_n \mathbb{1}_{\{t_n < t^*\}} \mathbb{1}_{\{t_n + p_n < t < t_n + p_n + \sigma_n\}} \quad (2)$$

Theorem 1. *Acceptation des réservations*

La condition d'acceptabilité de la n -ième réservation se traduit par

$$\forall t \in [t_n + p_n, t_n + p_n + \sigma_n] \quad C(t|t_n) < C_{\max} \quad (3)$$

L'expression de ϵ_n est alors :

$$\epsilon_n = \mathbb{1}_{\left\{ \sup_{t \in [t_n + p_n, t_n + p_n + \sigma_n]} C(t|t_n) < C_{\max} \right\}} \quad (4)$$

3.3 Le cas avec capacité infinie

Dans ce cas, toutes les réservations demandées sont acceptées dans le système, ainsi le ϵ_n introduit précédemment n'a plus d'utilité.

Le nombre de voitures réservées à l'instant t , c'est à dire ayant fini la période de pré-réservation mais pas encore celle de réservation proprement dite vaut alors :

$$LCt) = \sum_{n=0}^{\infty} \mathbb{1}_{\{t_n + p_n \leq t < t_n + p_n + \sigma_n\}} \quad (5)$$

En notant $B_t = \{(u, x, y) \in (\mathbb{R}^+)^3 \mid u + x < t < u + x + y\}$ et \mathcal{N} le processus de Poisson marqué d'intensité $\mu(du, dx, dy) = \lambda du \nu(dx, dy)$ représentant l'arrivée des usagers, cela donne :

$$L(t) = \mathcal{N}(\mathbb{1}_{B_t}) \quad (6)$$

Theorem 2.

$$\forall t \quad \mathbb{E}[L(t)] = \mathbb{E}[\min((t - p)^+, \sigma)] \quad (7)$$

Démonstration.

$$\begin{aligned}
\mathbb{E}[L(t)] &= \mathbb{E}[\mathcal{N}(\mathbb{1}_{B_t})] \\
&= \mu(\mathbb{1}_{B_t}) \\
&= \int_{(\mathbb{R}^+)^3} \mathbb{1}_{B_t}(u, x, y) \mu(du, dx, dy) \\
&= \lambda \int_{(\mathbb{R}^+)^2} \int_{\mathbb{R}^+} \mathbb{1}_{\{u+x < t < u+x+y\}} du \nu(dx, dy) \\
&= \lambda \int_{(\mathbb{R}^+)^2} \int_{\mathbb{R}^+} \mathbb{1}_{\{u < t-x < u+y\}} du \nu(dx, dy) \\
&= \lambda \int_{(\mathbb{R}^+)^2} \int_{(t-x-y)^+}^{(t-x)^+} du \nu(dx, dy) \\
&= \lambda \int_{(\mathbb{R}^+)^2} ((t-x)^+ - (t-x-y)^+) \nu(dx, dy) \\
&= \lambda \int_{(\mathbb{R}^+)^2} \min((t-x)^+, y) \nu(dx, dy) \\
&= \mathbb{E}[\min((t-p)^+, \sigma)]
\end{aligned}$$

□

3.3.1 Limite fluide

Considérons désormais un processus similaire, où le taux d'arrivé des clients λ est remplacé par λN . On note alors le processus $C_N(t)$, le processus d'arrivée des usagers \mathcal{N}_N et on définit le processus normalisé par

$$\bar{C}_N(t) = \frac{C_N(t)}{N}$$

Definition 6. Limite fluide

On appelle limite fluide de $(C(t))$ la limite de $\bar{C}_N(t)$ pour N tendant vers l'infini.

Theorem 3. *Limite du processus normalisé*

Le processus renormalisé tend en loi vers $\lambda \mathbb{E}[\min((t-p)^+, \sigma)]$,

Démonstration.

$$\begin{aligned}
C_N(t) &= \mathcal{N}_N(B_t) \\
&= \int_{(\mathbb{R}^+)^3} \mathbb{1}_{\{B_t\}}(u, x, y) \mathcal{N}_N(du, dx, dy) \\
&= \int_{(\mathbb{R}^+)^3} \mathbb{1}_{\{B_t\}}(u, x, y) (\mathcal{N}_N(du, dx, dy) - \lambda N du \nu(dx, dy)) \\
&\quad + \int_{(\mathbb{R}^+)^3} \mathbb{1}_{\{B_t\}}(u, x, y) \lambda N du \nu(dx, dy)
\end{aligned}$$

Notons $M_N(t) = \int_{(\mathbb{R}^+)^3} \mathbb{1}_{\{B_t\}}(u, x, y) (\mathcal{N}_N(du, dx, dy) - \lambda N du \nu(dx, dy))$,
de sorte que $C_N(t) = M_N(t) + \lambda \mathbb{E}[C(t)]$
 $(M_N(t))$ est une martingale de processus croissant associé $\lambda \mathbb{E}[C(t)]$

□