1. Context

- La censure est une forme de problème de données manquantes qui est commun dans l'analyse de survie. Idéalement, les deux dates de naissance et de décès d'un sujet sont connus, dans ce cas, la durée de vie est connue. Quand on estime les taux d'extinction, par exemple, on fait une seule simulation, avec le temps fixée, si ce temps de simulation est trop long, alors on obtient aucunne donnée censurée, on trouve tout à fait les données non-censurées. Pour les données non-censurées, si on utilise la méthode non-paramétrique pour estimer le taux de survie, alors ce n'est pas une bonne méthode.
- En fait, pour un essai clinique, sa duration est trop long (100 ans), cela est impossible. Alors, choisir un temps de simulation raisonnable est un problem.
- Plus, qu'est ce qui se passe quand le temps de simulation est trop grand. Ensuite, pour les données qu'on a, l'information dans les donées conteint deux types d'information, l'information qui a le coût, et l'information qui a le perte. On espère qu'on peut trouver les information valables à partir des donées. Dans le cas, si le nombre d'information obtenue est beaucoup redondant (grand), cela a beaucoup de désavantage comme suite : Gaspiller la ressource du processeur quand on lance des simulations trop longes. En particulier, pour le cas où le temps de simulation est exponentiel. Information qui est contenu dans les données obtenues est redondante, il y a beaucoup d'information non-valables. Pour chaque intervalle de temps, la probabilité de survie est calculé comme le nombre d'agents survivants divisé par le nombre de patients à risque. Les agents qui sont morts, abandonné, ou déménagent ne sont pas comptés comme «à risque», à savoir, les agents qui sont perdus sont considérés comme «censurés» et ne sont pas comptés dans le dénominateur. Quand on fait des simulation trop long, alors, on trouve toujours les données non-censurées. Tous les données obtenues sont non-censurées ce que nous voullons toujours arriver, on peut estimer exactement le taux de survie. Par contre, le temps de simulation est trop long, on gaspille beaucoup de ressource de processeur, et on attends à gagner le résultat trop long. Voilà, on doit trouver une méthod qui optimise les dynamiques du taux censuré pour les temps d'extinctions simulées.
- On demande une méthode qui est satisfaite de deux conditions : (1) apprendre le meilleur « taux censuré » dynamiquement et (2) faire une bonne balance entre l'exploration des nouvelles connaissances et l'exploitation des connaissances courants pour donner des bonnes décisions de la future.

2. IDÉE : Aprrendre le meilleur « taux censuré » dynamiquement.

Alors, on doit donner un modèle qui peut répondre bien la hyphothèse pour les simulations longes. La titre est "Simulate censored data such that most with longest time are censored".

3. ETAT DE L'ART

3.1. **Bandit model.** Multi-armed bandit :

The multi-armed bandit problem (sometimes called the K-[1] or N-armed bandit problem[2]) is a problem in which a gambler at a row of slot machines (sometimes

known as "one-armed bandits") has to decide which machines to play, how many times to play each machine and in which order to play them.[3] When played, each machine provides a random reward from a distribution specific to that machine. The objective of the gambler is to maximize the sum of rewards earned through a sequence of lever pulls. The multi-armed bandit problem models an agent that simultaneously attempts to acquire new knowledge (called "exploration") and optimize his or her decisions based on existing knowledge (called "exploitation"). The agent attempts to balance these competing tasks in order to maximize his or her total value over the period of time considered. There are many practical applications of the bandit model, for example: clinical trials investigating the effects of different experimental treatments while minimizing patient losses.

- Le problème de bandit multi-armés stochastique est un modèle important pour étudier le « tradeoff » exploration-exploitation dans l'apprentissage par renforcement. C'est une algorithme heuristique, une algorithme glouton (gourmand) qui est un choix optimum local.

3.1.1. *Bandits.*

- There are n machines
- Each machine i returns a reward $y \sim P(y; \theta_i)$

The machine's parameter θ_i is unknow.

- Let $a_t \in \{1, ..., n\}$ be the choice of machine at time t.
  Let $y_t \in R$ be the outcome with mean y_{a_t}
- A policy or strategy maps all the history to a new choice :
  $\pi : [(a_1, y_i), (a_2, y_2), ..., (a_{t-1}, y_{t-1})] \mapsto a_t$
- Problem : Find a policy π that :

$$Max\left(\sum_{t=1}^{T} y_t\right)$$

Or

$$max\left(y_T\right)$$

Or other objectives like discounted infinite horizon $max\left(\sum_{t=1}^{\infty} \gamma y_y^t\right)$

3.1.2. *Trading off exploration and exploitation.* L'Apprentissage par renforcement est une méthode d'apprentissage pour l'agent dans un environnement non-fixe avec l'objectif de maximiser le résultat final dans le long terme. Dans cet environnement, on nécessite un équilibre entre l'exploitation de connaissances actuelles et d'exploration de nouvelles connaissances à partir de ces régions inexploitées. Par exemple, le problème bandit multi-armés célèbre.

3.2. **Adaptive design.**

- Article : http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2422839/

- Adaptive designs Ici, « Adaptive designs » est un désign qui a pour de maximiser le résultat obtenu et minimiser le résultat rédondant. Pendant le temps de simulation, les agents apprendent pour adapter à ses situation courentes et maximiser le résultat. Le but est non seulement d'identifier efficacement les avantages du traitement cliniques sous enquête, mais aussi pour accroître la probabilité de succès du développement clinique. La conception adaptative est définie comme une conception qui permet des adaptations aux essais et / ou des procédures statistiques de l'essai après l'ouverture sans supprimer à la validité et l'intégrité du procès [21]. Une conception adaptative est comme une conception de l'essai clinique qui utilise des données acculantes pour décider de la façon de modifier certains aspects de l'étude pour qu'il continue, sans supprimer à la validité et l'intégrité du procès [19]. Dans de nombreux cas, un dessin ou modèle adaptatif est également connu comme une conception flexible.
- The use of adaptive design methods in clinical research and development based on accrued data has become very popular due to its flexibility and efficiency. Based on adaptations applied, adaptive designs can be classified into three categories: prospective, concurrent (ad hoc), and retrospective adaptive designs. An adaptive design allows modifications made to trial and/or statistical procedures of ongoing clinical trials. Adaptive design methods in clinical research are very attractive to clinical scientists due to the following reasons. First, it reflects medical practice in real world. Second, it is ethical with respect to both efficacy and safety (toxicity) of the test treatment under investigation. Third, it is not only flexible, but also efficient in the early and late phase of clinical development. However, it is a concern whether the p-value or confidence interval regarding the treatment effect obtained after the modification is reliable or correct. In addition, it is also a concern that the use of adaptive design methods in a clinical trial may lead to a totally different trial that is unable to address scientific/medical questions that the trial is intended to answer [17,18].
- Maximizing patient survival times : The use of adaptive design methods in clinical research and development based on accrued data has become very popular due to its flexibility and efficiency. Based on adaptations applied, adaptive designs can be classified into three categories: prospective, concurrent (ad hoc), and retrospective adaptive designs. An adaptive design allows modifications made to trial and/or statistical procedures of ongoing clinical trials. Adaptive design methods in clinical research are very attractive to clinical scientists due to the following reasons. First, it reflects medical practice in real world. Second, it is ethical with respect to both efficacy and safety (toxicity) of the test treatment under investigation. Third, it is not only flexible, but also efficient in the early and late phase of clinical development. However, it is a concern whether the p-value or confidence interval regarding the treatment effect obtained after the modification is reliable or correct. In addition, it is also a concern that the use of adaptive design methods in a clinical trial may lead to a totally different trial that is unable to address scientific/medical questions that the trial is intended to answer [17,18].

3.3. **Greedy algorithm.**

Algorithm 1 Algorithm proposed by JEAN DANIEL

REPEAT DURING TS CPU time

- REPEAT DURING TS/10
  - Select Random Population of N Sub-pop
  - Run and WAIT until Extinction to compute Extinction Time //FONCTIONNER combien de fois, paramètre
- END REPEAT
- COMPUTE OPTIMAL WaitingTime to have NUMBER SIMU →OPTWAIT

REPEAT UNTIL END OF BUDGET

- RUN until OPTWAIT to cmpute

3.4. **Heuristic algorithm.**

4. Solution

4.1. JEAN DANIEL :.

- https://communities.sas.com/
- Problème : simulate censored data such that most with longest time are censored
- Title of paper: Dynamic Optimization of Censor Rate for Simulated "Extinction"
- Idée: Apprendre le meilleur censor rate dynamiquement

Censored Data :

- Donées censurées I:
  - Pendant les heures de test de T, il y a r échecs (où r peut être n'importe quel nombre de 0 à n).
  - Les temps exacts de défaillance sont $t_1, t_2, ..., t_r$.
  - Il y a $(n - r)$ cas qui sont survécus à l'ensemble du test T-heure sans défaillance.
  - T est fixé à l'avance et r est aléatoire. On ne sais pas combien échecs va se produire jusqu'à ce que le test est fini.
- Donées censurées II :
  - On observe $t_1, t_2, ..., t_r$, où r est défini à l'avance. Le test se termine au temps $t = t_r$, et $(n - r)$ unités sont survécu.
  - Rarement utilisé.

ALGO

- Input:
  - Budget de TS CPU time
  - N Sub Pop
  - NUMBER OF SIMU

4.2. **Bandit.** UTILISER Bandit Algo :

- Bandit algorithm is a framework to balance the tradeoff of Exploitation and Exploration.
  - Exploration : exploit existing knowledge
  - Exploitation : explore new knowledge from these unknow regions.

- Multi-armed bandit Algorithms : Estimate the reward of each item based on the click and impression counts.
- Contextual Bandit algorithms : Estimate the reward of each item based on a feature-based prediction model, where the context is seen as a feature vector.
- Multi Armed Bandits:
  - Arms = possible treatments
  - Arm Pulls = application of treatment to inidividual
  - Rewards = outcome of treatment
  - Objective = maximize cumulative
  - reward = maximize benefit to trial population (or find best treatment quickly)
- UniformBandit Algorithm Pull :
  - Each arm w times (uniform pulling). Return arm with best average reward.
- Non-Uniform Sampling:
  - If an arm is really bad, we should be able to eliminate it from consideration early on Idea: try to allocate more pulls to arms that appear more promising
  - Median Elimination Algorithm :
    - Median Elimination A = set of all arms
    - For i = 1 to .....
      - Pull each arm in A w i times
      - m = median of the average rewards of the arms in A
      - A = A – {arms with average reward less than m}
      - If |A| = 1 then return the arm in A Eliminates half of the arms each round.

UTILISER l'algorithm Bandit pour aprrendre le meilleur censor rate dynamiquement :

- Input
  - Etat : Extinction globale ou NON
  - Action : N sub-pop

4.3. **Greedy algorithm.**

4.4. **Heuristic algorithm.**